Perceived Quality in the Automotive Industry: A Neural Network Based Assessment of Split-lines

vorgelegt von

M. Sc.

Serena Simone Striegel

von der Fakultät V - Verkehrs- und Maschinensysteme

der Technischen Universität Berlin

zur Erlangung des akademischen Grades

Doktor der Ingenieurwissenschaften

- Dr.-Ing. -

genehmigte Dissertation

Promotionsausschuss:

Vorsitzender:	Prof. Dr.-Ing. Henning Meyer
Gutachter:	Prof. Dr.-Ing. Matthias Rötting
Gutachter:	Prof. Dr. Claus-Christian Carbon

Tag der wissenschaftlichen Aussprache: 03. Februar 2021

Berlin 2021

Meiner Familie

“Das was aus Bestandteilen so zusammengesetzt ist,
dass es ein einheitliches Ganzes bildet,
nicht nach Art eines Haufens,
sondern wie eine Silbe,
das ist offenbar mehr als bloß die Summe seiner Bestandteile.”

Aristoteles, Metaphysik

ACKNOWLEDGEMENTS

Die vorliegende Arbeit ist in Zusammenhang mit meiner Tätigkeit bei der BMW Group im Bereich Absicherung Gesamtfahrzeug entstanden. Ich bin stolz in meiner Arbeit ein Stück Forschung in die Industrie gebracht zu haben und einen Teil zur Weiterentwicklung und Transformation der Industrie beigetragen zu haben.

Ein besonderer Dank geht nach Berlin zu Herrn Prof. Rötting von der technischen Universität. Seine vielseitige Unterstützung und Betreuung lieferten einen wesentlichen Beitrag zur Erstellung der Arbeit. Die vielen fachlichen Gespräche und Diskussionen lieferten einen wertvollen Beitrag im Rahmen der Arbeit.

Weiterhin möchte ich Herrn Prof. Carbon von der Otto-Friedrich Universität Bamberg danken. Durch seine Mitwirkung konnten unterschiedliche und abwechslungsreiche Ansichten in die Arbeit einfließen. An vielen Stellen sind die technischen Betrachtungen und Ergebnisse um psychologische Komponenten erweitert und ergänzt, was einen besonderen Blickwinkel in die Arbeit bringt.

Daniel Zielinski gilt ein besonderer Dank für die fachliche Betreuung meiner Promotionsarbeit bei der BMW Group. Die hilfreiche Unterstützung, die wertvollen Ratschläge, die kreativen Diskussionen aber auch die motivierende Zusammenarbeit waren ein wichtiger Bestandteil während der Promotionszeit.

Zusätzlich möchte ich allen Kollegen und Kolleginnen für Ihre Ideen, für ihre Unterstützung und für die Umsetzung unterschiedlicher Beiträge bei der Arbeit danken: Bernhard, Nicolas, Matthias, Peter, Dominik, Sasha, Linh und Vanessa.

Meinen Eltern, Peter und Simone, möchte ich auf besonderer Weise danken. Sie sind ein wichtiger Bestandteil und Stütze in meinem Leben und so auch während der Promotion. Ich danke meiner Mama für jedes motivierende Wort, für die vielen Zusprüche und für all die liebevolle Unterstützung. Meinem Papa danke ich für jedes ermutigende Wort, für seinen Glauben in meine Fähigkeiten und für sein absolutes Vertrauen in mich.

Meiner gesamten Familie möchte ich von Herzen für all die Unterstützung, für jeden Ratschlag und für jede Ablenkung danken, die sie mir während der Promotionszeit geschenkt haben: Carlo, Anika, Thiago, Mateo, Elias, Opa Hermann, Michael und Sarah.

Meinen Freunden möchte ich von Herzen für Ihre Geduld und Unterstützung danken: Michaela, Nela, Max, Franzi, Lou, Thomas, Yvonne, Benedikt, Hendrik, Klemens, Nina, Manuela, Florian, Coco, Jonas, Katrin, Anne, Desiree, Martin – und alle die hier keinen Platz mehr finden.

The opinion expressed in this work are those of the author. They do not purport to the opinion of the BMW Group.

For better legibility, gender-specific language was omitted and the generic masculine was used without intention to discriminate against the gender of any study subject.

ABSTRACT

The automotive industry is currently experiencing a volatile market environment. To be successful in this environment, it is essential to retain existing customers and to attract new customers. A comparison of competitor automobiles demonstrates a similarity in technical features. Product features such as design and quality play a decisive role in the success of a company and in convincing customers of its products (Carbon and Jakesch 2013; Esch 2013; Falk, Quattelbaum, and Schmitt 2010; Fontana, Giannini, and Meirana 1999). Both factors differ from traditional, objective assessments of products. The consideration of design and quality involves the research area of perceived quality. Perceived quality is a concept that describes the communication of the product quality and the subject's perception. The definition of perceived quality is that the entirety is more than the sum of the individual parts (Steenkamp 1990; Zeithaml 1988a). To stand out from the competition in the area of perceived quality and to remain resource-efficient, it is necessary to understand the customer's perception, their attitudes, and their needs and requirements. As these are highly subjective and depend on experience, knowledge, and expectations (Carbon 2019), the collection, interpretation, and assessment of design and quality at an objective level is challenging (Carbon and Jakesch 2013). Given this problem, some research methods were developed, which evaluate design and quality of products concerning perceived quality. The central element of these methods is the decomposition of a product into its individual components. This procedure contradicts the definition of perceived quality. Based on this gap in research, the aim of this thesis is to develop an objective assessment method for perceived quality. Split-lines are the object of investigation. The attributes gap, flush, and parallelism were identified, thus they represent design and a quality feature. Additionally, the development of an objective assessment method included psychological foundations from the field of perception. Four research aspects were defined in the thesis to derive an objective evaluation method for the perceived split-line quality. The first aspect examined the industrial evaluation process of visual split-line quality. A processual deficit was identified, as the current visual split-line quality was derived from split-lines' functional requirements. To eliminate this deficit, an evaluation level with four areas, Global Quality, Global Design, Local Quality, and Local Design was developed and integrated into the current process. In the second aspect, parameters influencing perception related to the visual split-line quality were investigated. Two parameter categories were derived: human-related and product-related categories. In the third aspect, the influence of the global product context on the visual split-line quality was investigated. In detail, the perception effects resulting from the combination of split-line attributes in the global product context were considered. In the fourth aspect of the thesis, the previous results were incorporated into developing a method for the objective evaluation of the visual split-line quality. Since the modeling of correlations and rules for a high level of perceived split-line quality depends on different variables and the complexity increases drastically with the capture of all variables, a neural network was used to map these correlations. By using the neural network, the split-line quality as an entirety was assessed. Thus, a method is developed, which assesses the perceived split-line quality concerning the definition of perceived quality.

ZUSAMMENFASSUNG

Die Automobilindustrie befindet sich derzeit in einem volatilen Marktumfeld und die Bindung und Gewinnung von Kunden gegenüber dem Wettbewerb steigt in Bedeutung. Der Wettbewerbsvergleich zeigt eine große Ähnlichkeit von technischen Produktmerkmalen, wie beispielsweise die Beschleunigung. Um dennoch Kunden von eigenen Produkten zu überzeugen, bilden Produktmerkmale wie Design und Qualität entscheidende Erfolgsfaktoren (Carbon and Jakesch 2013; Esch 2013; Falk, Quattelbaum, and Schmitt 2010; Fontana, Giannini, and Meirana 1999). Beide Faktoren unterscheiden sich von der klassischen, objektiven Bewertung von Produkten. Die Betrachtung von Design und Qualität sind im Forschungsbereich von Perceived Quality (wahrgenommene Qualität) enthalten. Perceived Quality beschreibt die Produktqualität wahrgenommen durch ein Subjekt. Die Definition von Perceived Quality ist, dass die Gesamtheit mehr ist als die Summe der Einzelteile (Steenkamp 1990; Zeithaml 1988a). Um sich im Design und der Qualität, vom Wettbewerb abzuheben und trotzdem ressourceneffizient agieren zu können, ist es notwendig die Kundenwünsche und -anforderungen zu kennen. Eine Herausforderung sind die Erhebung, die Auslegung und die Bewertung von Design und Qualität auf einem objektiven Level (Carbon and Jakesch 2013). Da Design und Qualität stark subjektiv, von Erfahrungen, von Wissen und von Erwartungen (Carbon 2019) abhängen. Unterschiedliche Methoden zeigen für die Bewertung von Perceived Quality eine Produktzerlegung in Einzelkomponenten auf. Die Einzelkomponenten werden soweit heruntergebrochen, bis diese mit traditionellen Methoden bewertbar sind. Anschließend an die Bewertung werden die Einzelergebnisse aufsummiert und ergeben das Perceived Quality Level. Diese Vorgehensweise widerspricht jedoch der Definition von Perceived Quality. Das Ziel der Doktorarbeit war die Entwicklung einer objektiven Bewertungsmethode anhand der Definition von Perceived Quality. Die Methode wurde am Beispiel der Spaltmaßqualität von Breite, Versatz und Parallelität von Fahrzeugen hergeleitet. Zur Herleitung einer objektiven Bewertungsmethode der wahrgenommenen Spaltmaßqualität wurden vier Aspekte in der Arbeit bearbeitet. Der erste Aspekt war die Untersuchung des industriellen Bewertungsprozess der visuellen Spaltmaßqualität. Ein Ergebnis der Untersuchung war, dass die visuelle Bewertung der Spaltmaßqualität von funktionellen Anforderungen abgeleitet werden. Um dieses Defizit aufzuheben, wurde ein Bewertungslevel mit vier Bereichen Globale Qualität, Globales Design, Lokale Qualität und Lokales Design entwickelt und in den derzeitigen Prozess integriert. Im zweiten Aspekt der Arbeit wurden Einflussparameter in Bezug auf Perceived Quality untersucht. Dafür wurden zwei Parameterkategorien definiert, welche in mensch-zentriert und produkt-zentriert unterteilen wurden. Der dritte Aspekt untersuchte den Einfluss des Gesamtkontextes auf die visuelle Spaltmaßqualität. Die Erkenntnisse wurden in einer weiteren Studie detailliert und zusätzlich sind die Einflüsse von Spaltmaßkombinationen auf die Wahrnehmung der Qualität untersucht. Im vierten Aspekt wurde eine Methode zur objektivierten Bewertung der visuellen Spaltmaßqualität entwickelt. Da die Abbildung der Zusammenhänge und Regeln, welche für eine gute wahrgenommene Spaltmaßqualität ausschlaggebend sind, von unterschiedlichen Variablen abhängen und da bei der Erfassung aller Variablen die Komplexität zunimmt, wurde, um diese Zusammenhänge abzubilden, ein Neuronales Netz entwickelt. Durch die Verwendung des Neuronalen Netzes konnte die Spaltmaßqualität gesamtheitlich bewertet werden und die Definition von Perceived Quality eingehalten werden.

FIGURES

TABLES

NOMENCLATURE

ACADE	Application for Computer-Aided Design for Emotional Impression
AGI	Artificial General Intelligence
ANI	Artificial Narrow Intelligence
AR	Augmented Reality
ASI	Artificial Super Intelligence
BPM	Business Process Management
BPR	Business Process Reengineering
cGAN	Conditional generative adversarial network
DMAIC	Define-Measure-Analyse-Improve-Control
EFQM	European Foundation for Quality Management
fbx	Filmbox
FEM	Finite element method
GPU	Graphic processor unit
HDRI	High Dynamic Range Image
jt	Jupiter Tessilation
MS	Microsoft
NMSE	Normalized mean square error
PDCA	Plan-Do-Check-Act
PDM	Product-Data-Management
ReLU	Rectified Linear Unit
SIPOC	Supplier, Input, Process, Output, Customer
SOP	Start of Production
tanh	Hyperbolic tangent
TCA	Traditional Conjoint Analysis
TPQ	Technical-Based perceived quality

TPU	Tensor processor unit
TQM	Total Quality Management
VPQ	Value-Based perceived quality
VQA	Visual Quality Appearance

TABLE OF CONTENT

1. INTRODUCTION

Products provide the connection between companies and customers. Even in the current turbulent and volatile market environment, this connection remains valid. Uncertain factors such as globalization or political circumstances complicate this connection and companies are forced to assert themselves in the market with new and innovative products. Caused by similar factors like globalization, customers can choose from a wide range of products from different markets, which in turn creates additional competitive pressure on companies. In order to survive in the market and to be able to convince by products and services, it is an essential advantage for companies to know their customers, their needs and expectations (Schmitt et al. 2015).

The automotive industry faces exactly these circumstances. Additional to the large number of competitors, the offerings of competitors are similar in many technical solutions, for example such as the realized power of an engine. To convince customers of own products, companies are forced to distinguish themselves by further product features. It turns out that product-external characteristics such as brand or service (Homer 2008) are purchasing-critical for customers. A strategy for differentiation from the competition is the automobile design and its quality (Carbon 2015a). As determined in various studies and research, design and quality convert to an important purchasing criteria (Carbon and Jakesch 2013; Esch 2013; Falk, Quattelbaum, and Schmitt 2010; Fontana, Giannini, and Meirana 1999). Through individual designs and expressive design language, a product character is assigned that fascinates customers or customers even identify with this design (Creusen and Schoormans 2004). In addition, a perfect fit of an automobile exterior demonstrates technical perfection and a high level of craftmanship. In turn, this competence of perfect execution indicates another distinguishing feature from competitors. Therefore, the appearance of an automobile, by individual designs and perfect execution, is an opportunity to differentiate from competitors and to excite customers. Consequentially, design and quality are essential for a long-term market success of companies (Leder and Carbon 2005). It is crucial to understand the own product and the importance of individual characteristics and their mutual visual effects. The discussion of these questions is the aim of perceived quality, which is anchored in different research disciplines, such as engineering, psychology and marketing.

However, the already challenging objective to design products with a high level of customer satisfaction alone is subjected to high-cost pressure. Due to resource-related restrictions, companies are forced to achieve the highest possible customer satisfaction at low costs. Since an automobile is a complex product with a multitude of linked features. For example, the design of an automobile is directly linked to its quality. Complex designs often require challenging manufacturing technologies. In contrary, if an elaborate design is poorly realized, it may not convince customers and decrease the payment disposition (Schmitt et al. 2015). New and innovative manufacturing technologies increase customer expectations regarding an automobile quality realization. Consequently, realizing an innovative design and a high level of quality often results in high costs (Stylidis, Wickman, and Söderberg 2018). Facing the cost pressure, the realization of such designs in high quality performance is highly difficult.

One product feature that contributes to the visual appearance of automobiles and fits in the research scope of perceived quality are split-lines. Split-lines are necessary due to different functions, like opening a door. Additionally, their composition influences on the one hand the design and on the other hand the perceived quality of the automobile. The global and design-strategic location of split-lines has a significant influence on the visual appearance. The global design of split-lines also influences the subsequent manufacturability and thus again the visual appearance of the automobile. In this context, the described relationship between design and quality is evident. Different methods have been developed to understand the cause-effect relationships, to structure the product complexity and to measure the subjective perception of products. The assessment process is one common problem with these methods. A product is decomposed into its individual characteristics and then evaluated by using numerical methods. For example, the visual quality of individual split-line attributes are assessed using discrete numerical values. This procedure contradicts the perception process. The split-line attribute deviations do not occur individually and are therefore not perceived individually. The numerical limitation cannot be recorded in a visual assessment. This decomposition contradicts the human perception, thus human perception does not necessarily reflect the physical world (Carbon 2015b). Carbon (2015b) summarizes that human perception is mostly reliable and objective, but is not necessarily veridical or valid. Based on the fundamentals of perception, considerations of product or quality perception are defined by the term perceived quality. Its definition is that perceived quality is more than the sum of the individual attributes (Zeithaml 1988a). Therefore, current industry- and research-related assessment processes are not suitable for the visual split-line assessment. In order to assess product features with respect to perceived quality, it is necessary to understand and to consider human perception within these usually engineering-based assessment processes (Carbon 2019). However, including human perception and deriving an objective assessment method for subjective product features represents a major challenge (Carbon 2016). An important factor is to know and understand customer requirements and deriving an approach to transfer these requirements into products to reach high customer satisfaction or fascination (Carbon and Jakesch 2013). Fundamental to this approach are the consideration of following relations:

(1) Product features are very closely related to each other and influence each other,
(2) their visual effects and correlations are unknown and thus represent a factor of unawareness,
(3) parameters that determine the subjective evaluation of product features are unidentified,
(4) an objective assessment of these attributes, which correspond with the perception process and includes the product context, is important.

The aim of this thesis is to derive an assessment method which pursues the definition of perceived quality by using a four-stage approach. First, the process of assessing split-lines regrading technical and visual requirements in industry was investigated. Therefore, a method was defined and applied for identifying weaknesses and improving the assessment process of split-lines. Second, parameters which influence the visual perception process of split-lines were identified. The identified parameters were analyzed in terms of how they are perceived with respect to design and quality and how they affect an assessment of design and quality regarding perceived quality. It is crucial to identify the parameters that influence the evaluation of split-lines and to define them in sufficient detail to enable a reliable evaluation. Guidelines were derived based on the research results, which should be chosen in an evaluation process of split-lines regarding perceived quality. Third, based on the current research methods the context of split-lines in the field of perceived quality related to design and quality was examined. As investigation method a conjoint analysis was applied, which aims to disclose the correlations between split-lines as design and as quality attributes. In addition, the influence of split-line regarding perceived quality was investigated. In a further investigation, the split-line influence regarding perceived quality was examined by a perceptual study. Individual split-line attributes were variated and combined. The resulting combinations were tested for their influence on perceived quality. This approach allows a fundamental analysis of the split-lines in the perceived quality context with a focus on perception, importance and connections. Fourth, the research gap of an objective assessment method of split-lines was investigated. Machine learning methods were considered to provide an assessment of split-lines based on images and related to the definition of perceived quality. This investigation was based on the previous derived parameters, which effect the split-line assessment. The approach for applying machine learning methods for a subjective assessment was presented.

In summary, the objectives of the thesis are:

(1) Process analysis of the current split-line assessment process and derivation of a process improvement.
(2) Identification of the factors and parameters which effect the assessment of split-lines in perceived quality and deducing guidelines for a reasonable visual assessment.
(3) Investigation of the context of split-lines and perceived quality regarding the global product concept and regarding relations between split-lines and split-line attributes.
(4) Derivation of an objective assessment method by using a machine learning approach.

2. THEORETICAL BACKGROUND

The theoretical background is split into five sections. The first section introduces the discipline of tolerance management. Also, the split-line attributes used in the thesis are introduced. The second section refers to rendering and ray-tracing methods, which were used in the conducted studies and for the generation of the database for the machine learning approach. Within the third chapter, qualitative, quantitative, mixed research models and conjoint analysis methods are presented. The theoretical background of perceived quality regarding different research disciplines is presented in the fourth section. For this purpose, a brief introduction to perception is obtained. Subsequently, different perceived quality assessment methods are presented, which are based on marketing, psychology, and engineering. Afterward, these methods are compared with each other by discussing similarities and differences. Based on the current methods, the definition of perceived quality in this thesis is derived. In the fifth section, an introduction to the functionality and applications of machine learning is given. Additionally, different machine learning approaches are presented. The focus lies on convolutional neural networks, which are suitable for application in the field of machine vision. Besides, the concept of Siamese networks is presented. Afterwards, methods for analyzing a network regarding importance features are presented. These methods are summarized with the term of explainable AI methods. A summary of current machine learning methods that assess product quality concerning classification or regression tasks is presented in the following section. Finally, the methods and the research questions of the thesis are presented.

2.1. Tolerance Management and Split-lines

Until the 18th century, the definition of tolerances of components was less relevant for industry (Bohn and Hetsch 2013b). Once series production was introduced and thus interchangeability of components had to be guaranteed, the definition of tolerances for components was integrated into the product development process (Bohn and Hetsch 2013b, 2013c, 2013a). For series production, components need to be replaceable. To achieve a component substitution, components need to be similar or at least similar to a certain extent (Bohn and Hetsch 2013b). The realization of this flexibility is the aim of a tolerance management process. Consequentially, tolerance management is defined as a sub-process of the product development process (Bohn and Hetsch 2013b). The central aim is to fulfill product functions by using management methods to achieve the lowest possible manufacturing costs through an optimal tolerance concept (Bohn and Hetsch 2013b). Tolerances impact the function, the producibility, the costs, and the product appearance (Bohn and Hetsch 2013b, 2013c, 2013a).

Accordingly, the tolerance management process within the automobile industry represents a major part of the development process. The tolerance management process includes tolerating individual parts up to the entire product (Bohn and Hetsch 2013b). Within the automobile industry, tolerances between exterior and interior components are defined as split-lines (Forslund and Söderberg 2008; Wagersten et al. 2013; Wickman et al. 2014). Split-lines describe the passage between two components. Their geometrical variation is defined by tolerances, which comply with product functions (Forslund and Söderberg 2008; Wagersten et al. 2013; Wickman et al. 2014). The permitted variations are defined in a concept derived in a split-line or tolerance concept by the design and construction department and submitted to the production department (Bohn and Hetsch 2013b). Figure 1 shows exemplarily split-lines of an automobile exterior (BMW Group 2018).

Figure 1. Exemplary Split-lines in context of an automobile, visualised by BMW Group, Department Total Vehicle Validation, 2018.

2.1.1. Function of Tolerances

Within the product design, tolerances comply with different functions (Bohn and Hetsch 2013b; Esch 2013; Forslund and Söderberg 2008; Juster et al. 2001), which are categorized in technical and aesthetic functions (Backhaus, Erichson, and Weiber 2015; Bohn and Hetsch 2013b; Wickman and Söderberg 2004). The aesthetic function aims to fulfill a high visual quality standard and is therefore essential for satisfying the customer requirements (Bohn and Hetsch 2013b; Wagersten et al. 2013). To achieve high visual quality, split-lines must be as small and parallel as possible. Components have to provide additional functions, such as opening and closing doors (Bohn and Hetsch 2013b). Therefore, tolerances and split-lines are categorized based on technical functions: the joining function of components, the kinematics, and the density function (Bohn and Hetsch 2013b). The joining function of components defines the position concerning other components (Bohn and Hetsch 2013b). Therefore, the three-dimensional orientation of individual components and the surfaces for joining other components are described. The function of kinematics considers three-dimensional movements of components, such as locking and unlocking the doors. Components with kinematic function often affect other functions such as aesthetical or density functions (Bohn and Hetsch 2013b). The third function considers the density of products or assemblies. Density should protect against moisture, water, and noise (Bohn and Hetsch 2013b). Besides, components must fulfill high aesthetical functions, especially in the exterior.

In summary, tolerances and split-lines have to comply different functions, which affect each other. Their definition is complicated, depending on the number of components involved and their functions. The assessment of the functions is related to different criteria. Table 1 illustrates different functions and their assessment criteria.

Table 1. Functions of tolerances, adapted from Bohn (2013).

Function	Assessment criteria	Method
Aesthetical	Split-line attributes (e.g. gap, flush, parallelism) Surface (e.g. character lines)	Optical definition and transfer to numerical limits
Joining	Position, orientation and surfaces	Numerical values, design and calculation
Kinematics	Position and orientation for the maximal movement	Numerical values, design and calculation
Density	Position and dimension compression	Numerical values, design and calculation

Different calculation and simulation methods are used in the design process to define tolerance limitations concerning their requirement fulfillment. These are mainly used to design technical functions. An evaluation of the optical function fulfillment usually is performed afterward. The possible calculation methods simulate either stiff components or elastic components. Finite element methods (FEM) represent complicated approaches and are used to calculate elastic components (Bohn and Hetsch 2013b).

On the contrary to FEM, traditional calculation methods are:

(1) The analytical best-worst calculation,
(2) the analytical statistical calculation, and
(3) the numerical statistical calculation. (Bohn and Hetsch 2013b)

Within the analytical best-worst calculation, the tolerances' worst extreme values are summed up (Bohn and Hetsch 2013b). The result is then analyzed for functional fulfillment of requirements. The analytical statistical calculation uses the root-sum-square method, where normally distributed individual tolerances are added, squared, and then the root is taken (Bohn and Hetsch 2013b). Both methods only work in simple geometric contexts (Bohn and Hetsch 2013b). Numerical methods are used for more complicated geometric component contexts. A frequently used method is a Monte-Carlo simulation. Here, random individual tolerances are formed and then summed up (Bohn and Hetsch 2013b). This process occurs several times, whereby statistical characteristic values can be determined (Bohn and Hetsch 2013b; Ehlert et al. 2016). First, the tolerance chain of a split-line is defined based on a CAD model. Since each component includes a component tolerance, and each connection between two components is defined with an assembly tolerance, the combination of both tolerance types is included in a tolerance chain. This definition can be used, for example, to calculate the resulting derivation for the split-line tolerance using a Monte-Carlo simulation. In turn, these simulation results are high in detail that the deviations are determined on each split-line attribute, i.e., gap, flush, and parallelism. Based on these numerical values, maximal and minimal values, which were defined by limiting positions, can be derived. Based on the limiting positions, the fulfillment of, i.e., technical requirements, can be verified.

2.1.2. Types of Split-line Attributes

Due to process fluctuations, the realization of products includes tolerances, which affect the visual appearance of split-lines between components. Therefore, tolerance concepts for split-lines, which include the definition of acceptable tolerance derivations, were deduced within the product development stage. As a result, split-lines are spread on the exterior and are integral elements of automobiles.

The definition of split-lines regarding optical requirements is sensible. Experts define, often supported by virtual methods, complex split-line concepts (Forslund and Söderberg 2008; Kajiya 1986; Ungemach, Mantwill, and Rund 2008; Wagersten et al. 2013; Wickman and Söderberg 2007). Within these concepts mainly three split-line attributes are considered: gap, flush, and parallelism (see Figure 2).

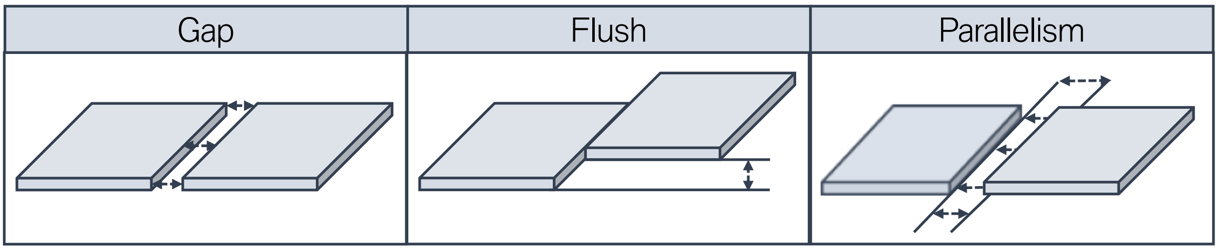


Figure 2. Geometrical variation of split-line attributes gap, flush and parallelism, adapted from Forslund, Dagman, and Söderberg (2006).

Gap is defined as the width between two components. Flush defines the offset between two components. Parallelism describes the rotation of two components towards each other. The resulting split-line appearance of the rotation is called in industry A-shaped split-line.

2.2. Rendering and Ray Tracing

The use of virtual methods in the product development process increased due to the high and rising cost pressure. Therefore, rendering methods have been established within the automobile design's evaluation process (Juster et al. 2001; Wickman and Söderberg 2007). These methods have various advantages compared to hardware prototypes. Their application offers a high degree of flexibility and enables fast and cost-effective visualizations of product concepts (Rix, Haas, and Teixeira 1995). The implementation of virtual methods is also performed during the assessment of visual split-line quality.

Rendering procedures are available in various computer software. Their objective is to picture a three-dimensional scene in a two-dimensional image (McKesson 2012). Using these methods, the product structure is transferred to polygons, which are the basis to create visualizations. The virtual image aims to represent a simulation of a scene that is as realistic as possible. The following input variables are required for a scene: the model of the object represented in polygons, the definition of materials, light source, and camera settings (Haines and Akenie-Loeler Tomas 2019; McKesson 2012; Richard 2005; Semmo 2016; Veach 1997).

Typically, three tasks are performed to create renderings. First, based on the camera position, the visible objects are defined. Second, virtual surfaces are defined by using material settings. Third, the lightning is calculated within the scene (Haines and Akenie-Loeler Tomas 2019; McKesson 2012; Richard 2005). The critical difference in rendering approaches is the method to calculate the image. For visualization of a created scene, mainly four different rendering methods are available:

(1) rasterization,
(2) ray casting,
(3) ray tracing or
(4) path tracing.

Rasterization is the digitalization of a model using simple shapes like points, lines, or polygons (Chen 2010). This method is a simple implementation and is used to display virtual geometry. A 3D scene and vectors are transformed into a 2D image, including pixels. Ray casting is a more advanced rendering method than rasterization. A ray for each pixel of a camera perspective is calculated within the virtual space. If the ray hits an object in the space, the color value for this pixel is determined (Chen 2010). Due to the simple implementation of this method, the bouncing of a ray is not further calculated, and therefore, shadows are not displayable. A more advanced method of ray casting, which includes the calculation of shadows, is ray tracing. Using the ray tracing method, a ray is sent from a viewpoint (the camera) into the scene. In doing so, intersections with objects are recursively calculated. Bounces between reflections and intersections with objects or refractions are considered in the calculation (Chen 2010). The reverse approach of calculating the ray is time-optimized compared to a ray foreword calculation (Chen 2010). The resulting image illustrates shading, refraction, and hard shadows of objects. The most advanced method is path tracing, which is also based on the ray tracing method. A ray is traced back from a camera in space. If it hits an object, it is refracted and pursued until it hits an object again. The bouncing happens until the ray hits the light source, or the ray darkens after a sufficiently long time. In retrospect, the pixel value is then calculated from the information obtained by the ray. The difference to ray tracing is that only one beam is tracked in space, and thus this beam acts like the light in reality (Veach 1997). Accordingly, the size and position of the light are decisive for path tracing methods. The path tracing result contains the same possibilities as ray tracing and color bleeding, ambient occlusion, global illumination, and soft shadows. The difference between ray tracing and path tracing is illustrated in Figure 3 (BMW Group 2018; Jensen 1996; Jensen and Buhler 2002).The left image illustrates the ray tracing method. Typically for this method are hard shadows. In the right image, the path tracing method was used. There, soft shadows, color bleeding, ambient occlusion, and global illumination are visible.

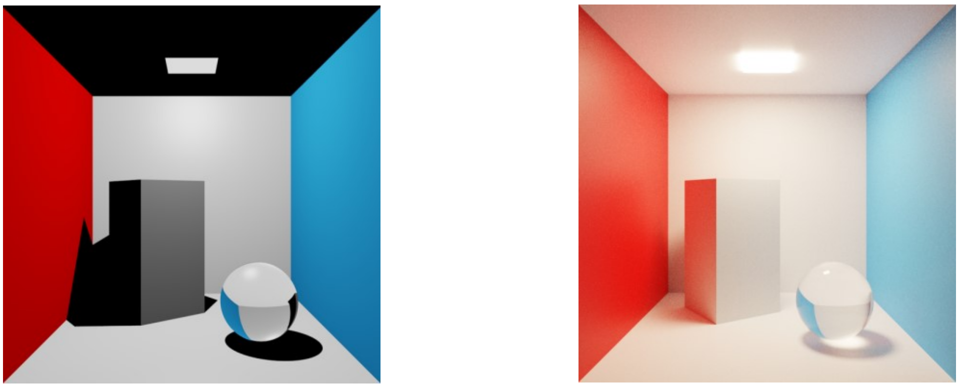


Figure 3. Comparison of ray tracing (left) and path tracing (right), rendered by BMW Group (2018).

A high contribution to a realistic image is made by the texture assigned to the object. Depending on the definition of the texture, the reflection behavior of the simulated light beam is influenced (Chen 2010). The process of assigning a texture to a pixel is texture mapping. Textures can, for example, be obtained using real images (Chen 2010).

2.3. Research Design and Methods

Empirical science considers the validity of hypotheses and theories within reality (Reinders and Ditton 2011). This investigation can either be a deduction or an induction. A deduction is an analysis, which extrapolates from the general to a particular case, or from the entire to the individual or from the abstract to the concrete (Reinders and Ditton 2011). An induction extrapolates from particular cases to general cases, or from the individual to the entire or from the concrete to the abstract (Reinders and Ditton 2011). Deductions are typically performed by quantitative research methods, whereas inductions are usually conducted using qualitative research methods (Reinders and Ditton 2011).

2.3.1. Qualitative and Quantitative Research

Qualitative and quantitative empirical research methods are based on assumptions and axioms (Schumann 2018). Qualitative research methods have their fundamentals within humanistic psychology, whereas quantitative research methods are derived by natural science (Schumann 2018). Within qualitative research methods, the focus lies on the human as individuum (Schumann 2018).Therefore, regularities within an individuum's behavior are deduced. The aim is to understand a human and their perception within their entirety. A limitation within qualitative research methods is the assumption that mapping reality is impossible (Schumann 2018). In contrast, quantitative research methods consider humans as a research object (Schumann 2018).The acquired data is used to verify general valid hypotheses, according to which humans behave. Besides, the acquired data is used for analyzing the derivation of characteristics within the considered entirety. Qualitative research has the basic assumption and the ideal to picture the reality one-to-one. Qualitative and quantitative research methods are incompatible within their basic assumptions, but the use of single methods from the contrasting research method can be performed to a certain extent (Schumann 2018). The difference between qualitative and quantitative research methods are presented in Table 2.

Table 2. Comparison of qualitative and quantitative research methods, combined and adapted from Schumann (2018).

Qualitative	Quantitative
Induction	Deduction
Holistic approach	Elementary approach
Objective: descriptive, understanding	Objective: causal relations
Data in need of interpretation	Numerical data
Theoretical generalization	Statistical generalization
Quality criterion: validity	Quality criterium: objectivity, reliability, validity
Natural environment	Laboratory

Qualitative research methods often deal with more research questions (Reinders and Ditton 2011). A characteristic of qualitative research methods is an open and flexible approach, i.e., realized in an unstructured observation or a qualitative interview (Reinders and Ditton 2011). A general approach for dealing with qualitative research methods is presented in Figure 4.

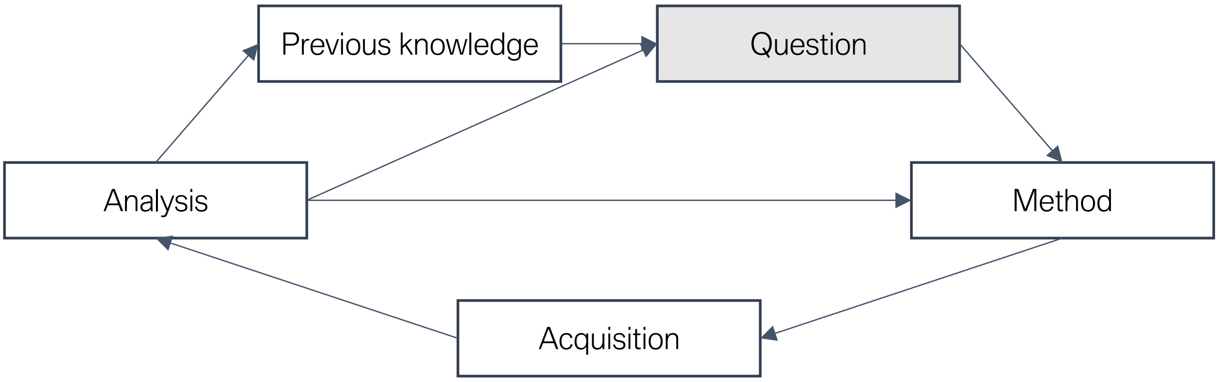


Figure 4. Approach of qualitative research methods, adapted from Reinders and Ditton (2011).

The research question is the initial starting point of qualitative research methods. The choice of methods is based on the posed questions in a second step (Reinders and Ditton 2011). Afterward, data acquisition is performed by applying the chosen method (Reinders and Ditton 2011). The analysis step results can either create new knowledge, hypotheses or affect the posed questions or affect the applied method in further acquisitions (Reinders and Ditton 2011). An advantage of qualitative research methods is that the acquired data is firstly processed and analyzed during the survey (Reinders and Ditton 2011). Therefore, early information on previous studies is used within upcoming surveys to increase the information level (Reinders and Ditton 2011). This feedback results in a dynamic research character of qualitative research methods (Reinders and Ditton 2011). Qualitative research methods demonstrate two specific characteristics. Firstly, qualitative research methods are more open, explorative, and pursue given research questions (Reinders and Ditton 2011). Secondly, this research character enables a fast combination of a question, applied method, and first results (Reinders and Ditton 2011).

Qualitative quality criteria are defined for assurance of the research quality (Himme 2009b; Mayring 2007, 2010). These criteria are documentation of the procedure, argumentative interpretation, rule-based conduction, proximity to the object, communicative validation, and triangulation. The criterion of the procedure documentation requires the planning, implementation, and analysis of a study. The criterion of argumentative interpretation forces that interpretations need to be justified. The rule-based conduction criterion defines a structured and systematic approach within the study despite its flexible qualitative research methods. The criterion of proximity to the object requests a consideration within realistic and natural environments. The results of qualitative methods depend heavily on the realization of this criterion. The communicative validation criterion evaluates the results by discussing them with the participants after the analysis. The last criterion, the triangulation, compares, and evaluates results gained by different methods but for the same research question. This criterion aims to compare different results to examine a research question from a different point of view.

Quantitative research methods examine theories or hypotheses for their validity (Reinders and Ditton 2011). The approach of quantitative research methods is structured and systematic (Reinders and Ditton 2011). The approach to conducting a quantitative research method is presented in Figure 5.

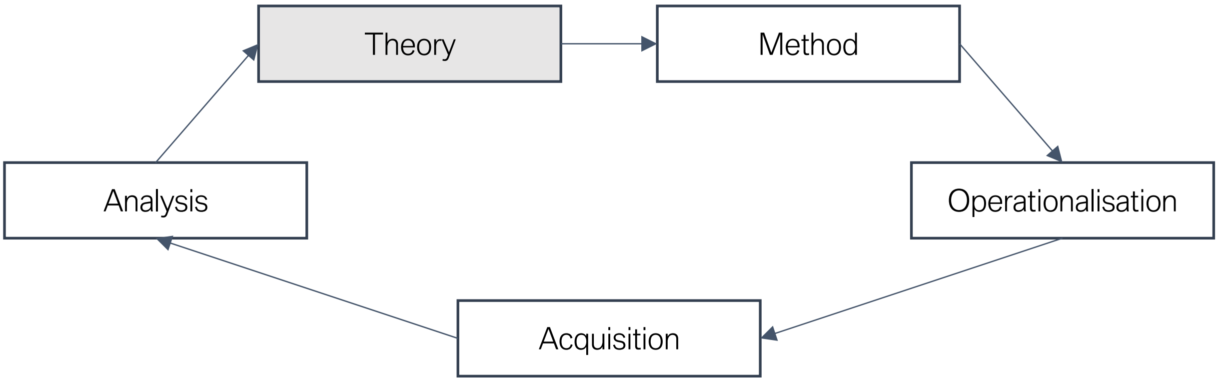


Figure 5. The approach of quantitative research methods, adapted from Reinders and Ditton (2011).

The theory or formed hypotheses are the basic consideration object of a quantitative research method (Reinders and Ditton 2011). Based on the theory, a suitable method is selected depending on the verification purpose (Reinders and Ditton 2011). The characteristics, which are included in the theory aspect, are operationalized within the third step (Reinders and Ditton 2011). The operationalization is performed to transform the characteristics into a measurable value (Reinders and Ditton 2011). Therefore, a suitable scale has to be chosen. The types of scale levels can be characterized by nominal, ordinal and metrical or cardinal scale (Greving 2009). These scale levels differentiate regarding their measurable capacity (Greving 2009). Subsequent to the operationalization the data acquisition is performed (Reinders and Ditton 2011). The data analysis step is conducted by using statistic methods. As a result of the analysis step, the validity of the initial theory or hypotheses are tested as true or false (Reinders and Ditton 2011).

The quality of quantitative research methods is evaluated by using three criteria: objectivity, reliability, and validity (Himme 2009b). The objectivity criterion is a general valid criterion for scientific research (Himme 2009b). Research results have to be independent of the situation and the investigator (Himme 2009b). The criterion of reliability considers the dependability and the stability of the study (Himme 2009b). This criterion is achieved if comparable results are obtained within a repletion test by having relatively constant conditions (Himme 2009b). The criterion of validity considers if a quantitative study has measured the right characteristics (Himme 2009b). Within an experiment, the validity criterion is separated into internal and external validity (Himme 2009b). Internal validity is achieved if the dependent variable's variation depends exclusively on the modifications of the independent variable (Himme 2009b). An experiment has external validity if the results are transformable to population, situation, and variable validity (Himme 2009b).

A mix of qualitative and quantitative methods is called mixed methods (Johnson, Onwuegbuzie, and Turner 2007). A typical combination of the methods is within elicitation or the analysis phase (Johnson, Onwuegbuzie, and Turner 2007). Four common types of mixed research models are triangulation, embedded, explanatory, and exploratory (Borrego, Douglas, and Amelink 2009).

2.3.2. Conjoint Analysis

Pioneers in the development and research of conjoint analysis are Green and Srinivasan (1990). Conjoint analysis is applied in various research disciplines and primarily used in marketing questions or product pricing targets (Hillig 2006). The aim of Conjoint analysis is the elicitation of part-worth utilities for single attributes of a product (Himme 2009a). Therefore, two methods can be applied: compositional and decompositional approaches (Sattler 2006). By using compositional methods, participants are directly asked about their preferred product attributes (Himme 2009a). By using decompositional methods, the entire product is assessed by participants, and by using statistical methods, the value of single product attributes is calculated (Himme 2009a). Conjoint analysis methods are decompositional methods (Himme 2009a). Conjoint analysis methods are defined as an approach to measure individual attribute preferences of entire products. Therefore, an advantage is the practical assessment task, as participants analyze and assess holistic products or concepts and chose those who fit their expectations most (Green and Srinivasan 1990; Hillig 2006). An overview of methods for analyzing product preferences is presented in Figure 6.

Compositional	Hybrid	Decompositional
Self-explicated methods	Adaptive conjoint analysis Golden conjoint	Traditional conjoint analysis Choice-based conjoint analysis Bridging conjoint analysis Limit conjoint analysis

Figure 6. Overview of different models for calculating product preferences, adapted from Himme (2009a).

Self-explicated methods are standard for compositional models (Himme 2009a). These methods usually perform two steps. Participants are asked to choose the best and worst product feature characteristics (Himme 2009a). In the second step, the feature characteristics of a product are weighted (Himme 2009a). Hybrid methods combine the compositional and decompositional methods (Himme 2009a). The adaptive conjoint analysis method is a hybrid variant (Himme 2009a). Within this method, participants are asked in a compositional survey part about the relevance and importance of a product's features and characteristics (Himme 2009a). Within the decompositional survey part, the product variants with different combinations of feature characteristics are assessed (Himme 2009a). The advantage of adaptive models is that essential product attributes are asked in detail, and irrelevant product attributes are neglected by the algorithm (Backhaus et al. 2016; Backhaus, Erichson, and Weiber 2015). A recent hybrid conjoint analysis method is the golden conjoint (Himme 2009a; Sattler 2006). In the first compositional survey part, participants are asked about their most important product attributes (Himme 2009a). In the second step, the mentioned attributes are weighted using a slider bar (Himme 2009a). In the final part of the conventional survey, unimportant attributes are eliminated by the participants (Himme 2009a). The results of the compositional survey part are used to generate choice-sets (Himme 2009a). These choice-sets are pareto-optimal and include no dominant stimuli (Himme 2009a). Within the decompositional survey part, the choice-sets are assessed by the participants (Himme 2009a). Based on the ranking of the choice-sets, the weighting of the attribute (Himme 2009a). The traditional conjoint analysis (TCA) was the first conjoint analysis approach and established the basis for further developed methods (Backhaus et al. 2016). The following premise underlay the TCA: customer preferences towards products are achieved by ranking product alternatives. Accordingly, a product alternative is a package of attributes, and each attribute had different levels. This theory is a multi-attribute method, and participants make a weighing decision, where they latently evaluate product attributes (Green and Srinivasan 1990; Hillig 2006). The Choice-based conjoint analysis method is based on the assumption that participants chose utility-maximizing, and therefore, the contribution of product characteristics is calculated (Backhaus et al. 2016; Himme 2009a). Hierarchical conjoint analysis methods are applied within complex product decisions with countless product characteristics

(Himme 2009a). Therefore, a decision is firstly made on an abstract level, and participants are based on their choices guided through more detailed levels (Himme 2009a; Sattler 2006). A similar method is the bridging conjoint analysis, which uses sub designs to picture complex products (Himme 2009a). By using the limit conjoint analysis, the traditional conjoint analysis is extended, and the purchase intention is requested (Himme 2009a).

The approach of conjoint analysis is based on the TCA and is divided into five steps (Backhaus et al. 2016; Backhaus, Erichson, and Weiber 2015), which are:

(1) Definition of product attributes
(2) Stimuli
(3) Assessment of stimuli
(4) Calculation of utility values
(5) Aggregation of values

The product attributes are selected within the first step, which is analyzed by using the conjoint analysis. The selection of product attributes is based on seven criteria: relevant, influenceable, independent, feasible, compensatory relation, no exclusion criteria, and limited (Backhaus et al. 2016; Backhaus, Erichson, and Weiber 2015). By considering the first criterion, product attributes are chosen, which are assumed to affect a product's value. Within the conjoint analysis, a stimulus is defined as a product combination by several attributes (Backhaus et al. 2016; Backhaus, Erichson, and Weiber 2015). The criterion influenceable prefers attributes for the conjoint analysis, which can be variated (Backhaus et al. 2016; Backhaus, Erichson, and Weiber 2015). The criterion independent suggests attributes, which can be individually variated (Backhaus et al. 2016; Backhaus, Erichson, and Weiber 2015). This criterion is primarily essential for the adaptive conjoint analysis (Backhaus et al. 2016; Backhaus, Erichson, and Weiber 2015). The criterion feasible focuses on these attributes, which are technically producible (Backhaus et al. 2016; Backhaus, Erichson, and Weiber 2015). The criterion of compensatory relations between attributes relates to the value of a product by accumulating each attribute, which is substitutable (Backhaus et al. 2016; Backhaus, Erichson, and Weiber 2015). The criterion of no exclusion defines that the chosen attributes are not supposed to be mutually precluding (Backhaus et al. 2016; Backhaus, Erichson, and Weiber 2015). The last criterion considers the limited number of attributes. Within conjoint studies, the number of attributes and product variants can exponentially increase, and therefore, this attribute considers the limitation of attributes and their levels (Backhaus et al. 2016; Backhaus, Erichson, and Weiber 2015).

Based on the defined attributes, several product alternatives are created in the second step. The product alternatives result from different combinations of the defined attributes and their levels (Backhaus et al. 2016; Backhaus, Erichson, and Weiber 2015). Within a subsequent step three, the assessment of the stimuli by participants is conducted. The necessary data is acquired in the next step (Backhaus et al. 2016; Backhaus, Erichson, and Weiber 2015). The calculation of the utility values and the importance values of the product attributes are performed in the fourth step (Förster and Kreuz 2003). The aggregation of the results is performed in the fifth step of the TCA analysis. The fact that participants continually evaluate exclusive products with different versions of the variants means that the global product context and the evaluation's relations remain the same. A disadvantage of this approach is the limited scope for consideration. The study design of conjoint analyses can quickly gain in complexity and the number of resulting product variants. For this reason, only a few product attributes and their levels can be considered.

Within conjoint analysis methods, the product remains in its entirety. Therefore, three advantages of the conjoint method are presented (Förster and Kreuz 2003):

(1) A reliable method to map the decision-making process and determine the benefit of product attributes,
(2) Precise knowledge about the individual attribute and its characteristic,
(3) The analysis provides an accurate result regarding the pricing of a product.

2.4. PERCEIVED QUALITY

Quality is more than just fulfilling the technical requirements of a product (Zeithaml 1988a). Quality is multidimensional and defined by many influencing factors. Garvin (1984) specified eight levels of quality: performance, features, reliability, conformance, durability, serviceability, aesthetics, and perceived quality (Garvin 1984). This framework, especially the two dimensions, aesthetic and perceived quality, was classified as highly subjective. The explanation of these dimensions' subjective assessment was based on a personal assessment of how a product was individually perceived through sensory perception (Garvin 1984). Based on this research, the perceived quality was deduced in detail by Zeithaml (1988a). With the focus on the character of perceived quality, four aspects of perceived quality were presented:

(1) Difference between perceived and objective quality,
(2) Higher abstraction level,
(3) Global assessment and
(4) Experience based perception (Zeithaml 1988a).

Both research contributions demonstrate the importance of perceived quality. Besides, they also point to the subjectivity, the abstraction level, and the globality of their evaluation.

Perceived quality appeared in various research disciplines, such as marketing, engineering, and psychology. A central aspect of perceived quality in each discipline was the subjective assessment of product value regarding product attributes by customers (Chi 2009; Freeman, Dart, and Dart 2008; Oude Ophuis and Van Trijp 1995; Snoj, Korda, and Mumel 2013; Stylidis et al. 2015; Tsiotsou 2014). Often within marketing perspectives, a link between prize, product, and brand image was created. Therefore, a practical implementation of perceived quality correlates in quantifiable success for companies, and this relationship was called prize-perceived quality (Lichtenstein and Burton 1989; Snoj, Korda, and Mumel 2013; Tsiotsou 2014; Völckner and Hofmann 2007). In contrast, engineering perspectives focused on product execution and strategies to realize customer-related attributes, and some concepts were presented by Falk, Quattelbaum, and Schmitt (2010); Liu (2003); Steenkamp (1990); Stylidis et al. (2015). Psychological-based approaches often concentrated on finding interdependencies between perceived quality and customer satisfaction (Gotlieb, Grewal, and Brown 1994) or investigated sensory influences in detail (Carbon and Jakesch 2013).

2.4.1. PERCEPTION IN PRODUCT DESIGN

Human perception is optimized for daily situations to make fast, efficient, and focused decisions (Carbon 2015b). Perception is highly subjective and does not always comply with reality (Carbon 2015b; Raab, Unger, and Unger 2016; Zimbardo 1978). It is not essential to record the actual reality of an object, but it is essential to correctly, and target-oriented interpret tan situation (Carbon 2015b). Therefore, the understanding of human perception is essential concerning engineering science. Thus, a technically perfect design does not necessarily correlate with perception as a perfect object. For example, the definition of split-lines via a numerical limitation of deviations may be technically correct, but optical perfection is not necessarily achieved by using this procedure. In design, different theories of perception are regarded as the design basis for product design (Raab, Unger, and Unger 2016). As a basis for perceived quality in the technical environment, the following section is a brief introduction of perception basics.

The necessary prerequisites of the perceptual apparatus need to be clarified. A comparison of the use of different senses shows that visual perception is given a major priority (Bream 2004). The human eye has two systems with different functions: rods and cones (Zimbardo 1978). The cones are active within lighted situations and are responsible for seeing colors and visual acuity (Zimbardo 1978). The rods are active within dusk, react sensibly regarding less brightness, and respond to black, white, and grey (Zimbardo 1978). Rods and cones are present on the lowest layer of the retina (Zimbardo 1978). To receive light, the light has to pass through different layers, nerve fibers, and blood vessels (Zimbardo 1978). The concentration of cones is high in the fovea area, and their concentration decreases when getting to the periphery of the retina (Zimbardo 1978). Rods are present at the entire retina, without the area of the fovea (Zimbardo 1978). Both receptors are connected to bipolar cells, which are synapses with ganglion cells (Zimbardo 1978). The optic nerve consists of axons of these ganglion cells (Zimbardo 1978). The optic nerve connects to the corpus geniculatum laterale of the thalamus and connects the occipital lobe (Zimbardo 1978).

Nevertheless, human visual perception is, like all visual systems, restricted. The perception of visible light for humans is between a wavelength of 380 – 780 nm (Carbon 2015b). Besides, the brain loses information by visual leakage and a filter mechanism. Many external influences evidence the filtering mechanism of the brain. The brain prevents an overflow of stimuli by blocking supposedly unimportant information (Carbon 2014; Chollet 2018).

The direction of perception was often discussed. Navon demonstrated a global to local process (Navon 1981): A forest was perceived first and afterward single trees. If the same individual elements are present, the global appearance is perceived first (Navon 1981, 2003). Gregory introduced a top-down theory (Gregory 1997, 2005; Ramachandran and Gregory 1991). Based on this theory, information reaches the brain, but the majority is filtered. The arriving information is processed and actively reconstructed by using knowledge and experience. The top-down process principle was that hypotheses were tested through perception influenced by knowledge and personal experience. One assumption is that visual illusions are an incorrect question of the brain.

In addition to the previous perception concepts, there are other perceptual effects, such as the blind spots or saccades, which affect human perception (Carbon 2014, 2015b). Blind spots occur in the area of the retina of the eye, where light receptors are missing (Zimbardo 1978). If information hits the blind spot, information is lost, but subsequently, the input is interpolated, and an overall image is created (Carbon 2014, 2015b). A similar interpolation occurs in saccades, which appear during fast eye movements. During these movements, the perception is interrupted. Nevertheless, the exposure is filled, and the perceived snapshots are interpolated, creating a seamless experience (Carbon 2014, 2015b).

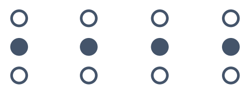
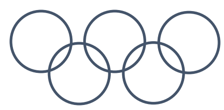
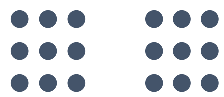
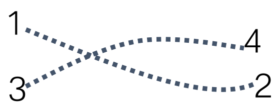
The incoming information is processed further based on influence factors (Raab, Unger, and Unger 2016). This further processing is manipulated by experience, prior knowledge, and expectations (Carbon 2015b). Various experiments demonstrated that information influences perception in daily situations, called priming. Priming is defined by triggering a particular perception by targeted advanced information, thus objects appeal to a higher probability in a specific context (Carbon 2015b). Further theories identified expectations to have a proven influence on perception. The importance of an object's context was demonstrated by Bruner and Minturn (1955). Within a study, they wrote less precise a combination of the letter "B" and the number "13". They had verified that participants perceive either the letter or the number, depending on the context. In conclusion, the product context is essential for perception and, therefore, for assessing perceived quality.

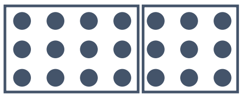
Additionally, an equally good influence of prior knowledge on perception was demonstrated by an experiment, which included an image displaying a rat or a man (Bugelski and Alampay 1961).

Within these experiments, the evidence was provided that combined perception and comparison with previous knowledge and expectations support daily decisions and helps to recognize things under challenging conditions (Carbon 2015b). Furthermore, perception aims to recognize patterns including a meaning. In structures or images, an urge is to recognize the design, which corresponds to a reward of the visual system and, at the same time, counteracts cognitive tiredness (Muth and Carbon 2013). If an object is imprinted in perception under different conditions, mental representations arise. These representations support perception, and even if objects are exposed to a strong distortion or noise, they can be identified (Carbon 2015b). Despite the many influencing factors on perception, these factors are consistent for every individuum. Therefore, these conditions result in a universal space of perception shaped by expectations, prior knowledge, and experience (Carbon 2015b). In summary, the influencing factors are subjectivity, selectivity, and activity (Kroeber-Riel and Gröppel-Klein 2013; Raab, Unger, and Unger 2016). Subjectivity reflects on the individual perception process of persons and their reality. Selectivity summarizes the filtering factors, distortion factors, and falsification factors that influence the information which needs to proceed. The activity includes the conscious reception and processing of information (Raab, Unger, and Unger 2016).

When considering the appearance of products, such as automobiles, theories of perception become relevant and the basics of the perception process. One theory is the gestalt theory, which was introduced by Wertheimer (1923), Koffka (1937) and Köhler (1967). The theory's basis is that the entirety is more than the sum of its parts (Zimbardo 1978). Based on this, individuals strive to recognize clear patterns embedded in an overall context (Raab, Unger, and Unger 2016). The recognition of these relationships can create a kind of reward and joy for the perceiver (Carbon 2015b; Muth and Carbon 2013). The order, conciseness, harmony, or meaningful forms within objects are therefore incentives that are sought in an overall concept (Raab, Unger, and Unger 2016). Based on these theories, Wertheimer deduced six Gestalt laws (Wertheimer 1923), presented in Table 3.

Table 3. Laws of Gestalt based on the Gestalt theory.

Law of gestalt	Explanation	Examples
Law of similarity	Objects of comparable shape, colour are perceived as a group.	
Law of pragnanz	Forms are perceived as simple as possible	
Law of proximity	Objects which are close to each other are perceives as group.	
Law of continuity	Points or dotted lines are perceived, as are they combines by a line.	
Law of closure	Interruptions in objects are perceived as entirety and spaces are interpolated.	

Law of common regions	Elements or objects with the same purpose are perceived as a belonging together.	

The law of similarity demonstrates that objects with a similar shape are perceived as a group. This law can occur within images or music as well. Table 3 demonstrates the law of similarity by different colored dots, which were perceived line-wise. The law of Prägnanz is demonstrated by using the Olympic sign, which is perceived as five cycles. Other wordings of this law are the law of good figure or the law of simplicity. This law refers to the fact that forms are perceived as thoroughly as possible. The law of proximity describes a grouped perception of objects that are close to each other. Due to the closeness of objects, they are perceived as belonging together. This law is demonstrated in Table 3 by 18 dots, which are perceived as two groups of 9 dots. The continuity law explains that points or dotted lines appear as if they were connected to smooth lines. The lines are perceived as continuous lines rather than sharp angles. This law is illustrated by using the dots in a line. These are perceived from 1 to 2 and from 3 to 4. Due to line-wise perception, these are perceived continuously. The closure law describes that even through interrupted lines, a Gestalt of an object is perceived as continuous and complete. Interrupted lines or figures are interpolated, and the entirety is perceived, which is demonstrated in Table 3, having a rectangle and a triangle. According to the law of common regions, objects with the same purpose are perceived as belonging together. For example, when elements move in the same direction, they are perceived as belonging together. The law of common regions is demonstrated by using a border around the elements. The elements within a boarder are perceived as a group.

Some of these laws are used within an automobile design to achieve an attractive appearance. An example, within the interior, operating buttons, which belong to the same function, are grouped, for example, to operate the radio. Applying the law of proximity, users understand the connection and the shared function of the buttons. The split-lines of the exterior are to align with other component edges. Designs at the rear lights often fulfill the law of continuity. The split-lines are continuous and ideally have no angles, which creates a homogeneous impression.

Furthermore, perception is more schematically than image-based (Carbon 2015b). An example of the abstract perception and filtering mechanism was discussed by Gregory (2005) and Gimini (2016a, 2016b). In the project Velocipedia, Gimini illustrated bizarre bicycle drawings, which people have created out of memory (Gimini 2016a, 2016b). The bizarre sketches explained that the brain filters irrelevant information, such as the individual components for a bicycle, and creates a sketch. Besides, special functions were assigned to these sketches and made the object more abstract (Gregory 2005). Recognizing a bicycle is simple but drawing a truthful sketch from memory is complicated. This effect can be referred to as recall and recognition. For humans, recognition is more accessible than the reproduction of objects (recall).

A concept for translating psychology theories to product design was introduced by Carbon (2019). Fundamental psychology theories were presented, which considered aspects of perception in product design to meet customer requirements and reach a high customer satisfaction level. This concept – the so-called Psychology of Design - defined eleven conceptual moves. They support creating a product design and are closely related to customer requirements, safe usage, and loved design (Carbon 2019). These conceptual moves were derived from personal experience, designers, and technicians. The power of associations to assess aesthetics was considered by Fechner and translated by Ortlieb and Ku (2020). Within these studies, two concepts of perceiving aesthetics were presented: the Aesthetics from Above and the Aesthetics form Below (Ortlieb and Ku 2020). The focus of Fechner's work was the Aesthetics from below. This research's novelty lay in the consideration of previous knowledge and experience within perception and the evaluation of beauty. The association of an object attaches meaning to it and associates an impression. This impression is highly important for evaluating an object as beautiful (Ortlieb and Ku 2020). Therefore, the Aesthetics of Below is crucial for product design. Thus, customers' association affects the perceived pleasure of products, which needs to be understood within the product development process. For realizing a psychological turn in product design, the following conceptual moves are necessary (Carbon 2019):

(1) Understanding the importance of perception related to a hypothesis-driven process, which cannot be assessed objective.
(2) Understanding the effects of design on prediction, which impact conscious actions.
(3) Understanding Zeitgeist, in terms of a cognitive mechanism regarding changes and drivers, triggers of change.
(4) Understanding the underlying reasons and arguments for assessing design.
(5) Understanding the relation of cognition and emotions shaped and triggered by the body.
(6) Understanding the product context of design.
(7) Transforming products from a functional design to a design with an own personality.
(8) Enabling a modulation of affordance to meet flexible affordances related to culture, knowledge, experience and expectations.
(9) Understanding the impact of analogies to support customers understanding product functions.
(10) Transforming the product assessment to a multisensory consideration and respecting a natural perception of product designs.
(11) Including the Gestalt theories within products for a successful design.

The first conceptual move is related to the perception process, which had a highly subjective character. Products and their operation depend on a person's expectations, knowledge, and hypothesis testing (Carbon 2019). These variables differ between persons, and therefore, the perceived reality of an individual is always subjective. The Gestalt theory established the second conceptual move. This move's objective was to understand how product designs were perceived and which predictions were triggered using different designs. By understanding the effects of design on predictions, product designs, which ended up in incorrect handling, were avoided (Carbon 2019). A changeable factor of design due to time was considered in the third conceptual move. The introduction of Zeitgeist within design demonstrated that a physical and mental change occurs, which needed to be considered for a successful design in the future (Carbon 2019). Within the fourth conceptual move, the reasons for assessing design were investigated. Based on the familiarity of a design, the product's assessment differed, especially regarding innovativeness (Carbon and Leder 2005). Product familiarity needed to be considered for assessing a product design. An assumption was that by doing so, assessments gain the power of prediction (Carbon 2019). The relation between body and soul, which were identified to impact the perception of the design, was considered in the fifth conceptual move. Body and soul were connected, and therefore, a separate consideration is unreasonable. Within the sixth conceptual move, the importance of the product context regarding product design perception was derived. Products were designed to fulfill requirements, which were different related to the scenario. Therefore, product designs stood in context to the scenario-specific requirements and were assessed within this context (Carbon 2019). The seventh conceptual move demanded to design products with a unique personality. By creating a product personality, customers quickly and implicitly understand product usage in different situations. Additionally, the usage of product personality is a strategy to differentiate between competitors.

The conceptual move for creating a product personality includes the process knowledge of building a personality, understanding its impact, the comparability of personalities, and the effects of strategic personality changes (Carbon 2019). Culture, knowledge, experience, expectations, and task-related factors have a proven effect on product perception—the variability of these factors needed to be considered in product design. The eighth conceptual move solves the variability of affordance. The idea for handling the variability was a modular structure of affordance (Carbon 2019). Within the ninth conceptual move, analogies were presented which support the understanding of product functions. Analogies within product design establish product functions, which were more intuitive and resulted in a good product experience (Carbon 2019). The combination of senses in product design was presented in the tenth conceptual move. The multisensory view on product design a more aesthetic and natural product design resulted (Carbon 2019). Finally, the eleventh conceptual move demonstrated the importance of the Gestalt theory in product design, whereby the definition of perceived quality was respected. In its entirety, a product's design was a success factor for pleasure design (Carbon 2019).

In summary, the influence of experience and distortion of reality creates an individual perception of the outside world (Carbon 2014). Returning to the perceived quality assessment, the perception and internal and external influences play a central role. To achieve a high level of perceived quality or to assess perceived quality, the perception process and influences on perception are essential to consider.

2.4.2. Perceived Quality in Marketing

Based on the definition of perceived quality and foundations of perception, the methods and different research disciplines are presented following (Figure 9). An early developed marketing-based approach was defined by Oude Ophuis and Van Trijp (1995). A customer-integrated approach dealing with perceived quality from a marketing viewpoint was presented. This method's basis was the perception process of customers and the relationship of perceived quality to product, place, and person. Also, the perception process of perceived quality was investigated and explained in detail. The main contribution was the abstract and multidimensional estimation of perceived quality, which cannot be captured and evaluated by consumers and results in an evaluation of indirect quality characteristics. A support for companies was a developed quality guidance, introducing three steps: identify customer-based quality assessments, demerge quality evaluations, and transfer customer requirements, including the knowledge of intrinsic quality attributes.

An equally marketing-based approach, provided by Chi (2009), investigated the relationship between perceived quality and brand loyalty. The attempt to investigate the relationship was realized by four study-related research questions dealing with

(1) Influences of brand awareness and perceived quality,
(2) effects of perceived quality and brand loyalty,
(3) meditational effects of perceived quality on the relationship of brand awareness and purchase intension, and
(4) meditational effects of brand loyalty on the relationship between brand awareness and purchase intention.

The results showed a significant and positive relation between brand awareness or loyalty and perceived quality. Perceived quality and brand loyalty were identified as mediator roles between brand awareness and purchase. The primary goal was to establish the correlations of the four categories more precisely. Thus, a detailed consideration of individual product attributes regarding the four research questions was not further addressed. A positive correlation regarding a purchase intention was determined with brand identity, brand recall, action, and effective loyalty based on the general research scope.

The research scope of Snoj, Korda, and Mumel (2013) was perceived quality, perceived risk, and perceived product value. The main aspects of perceived value: customers assess product value related to their expertise, knowledge of buying and using, product value is assessed by customers subjectively and cannot be assessed objectively by companies, perceived value is a multidimensional concept. Snoj, Korda, and Mumel (2013) identified and criticized approaches of simplifying perceived value as a combination of perceived quality and price. Based on a discussion of various concepts of perceived quality, Snoj, Korda, and Mumel (2013) followed the definition of Zeithaml (1988a) that the entity of a product is more than the sum of its part. Perceived sacrifices were defined as product aspects that negatively affect perceived value, which can be the price of a product and the non-monetary aspects like waiting or searching costs. Based on the fundamental research of the three research disciplines, Snoj, Korda, and Mumel (2013) investigated their relationship and focused on four aspects dealing with perceived value, quality, and risk, which were (Snoj, Korda, and Mumel 2013):

(1) perceived risk and perceived value,
(2) dependencies between variables,
(3) disparities between the study groups and
(4) the disparities between the theoretical basics and the study results.

The main contribution was the derived triangle relationship and the individual relationships between perceived quality, perceived risk, and perceived value, described by Cronbach's alpha. Significant results were the adverse effects of perceived quality on perceived risks, perceived risks on perceived value, and perceived quality's positive effects on perceived value. Like the Chi (2009) approach, Snoj, Korda, and Mumel (2013) conducted a general research scope of a product. The product was considered in its entity without detailed consideration of individual product attributes.

2.4.3. Perceived Quality in Psychology

A marketing and psychological approach was presented by Steenkamp (1990). This approach contributed to a detailed understanding of perceived quality and their relating influences. He discussed different definitions of perceived quality and identified a lack of considering psychological and customer behavior within these definitions. Therefore, he derived a new definition of perceived quality, which included three main commitments (Steenkamp 1990): Personal preferences are involved during a perceived quality evaluation, perceived quality is neither objective nor completely subjective, the subject-object relation is characterized by comparison of competing products, personal and situational, and the entirety of a product defines perceived quality. Based on the presented definition of perceived quality, a conceptual model for describing the quality perception process was developed. Figure 7 illustrates the model.

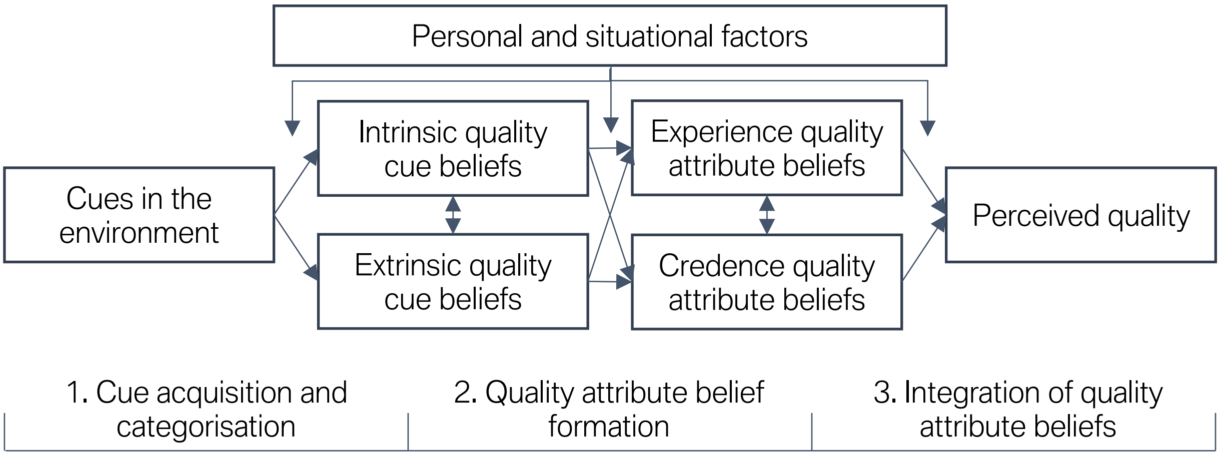


Figure 7. Conceptual model to describe the quality perception process, adapted from Steenkamp (1990).

Steenkamp's derived model (Steenkamp 1990) included three levels: cue acquisition and categorization, quality attribute belief formation, and integration of quality attribute beliefs. Within the first stage, customers were overflowed by product quality cues. Based on the cognitive system, these cues were filtered, acquired, and categorized. Within the second stage of the conceptual model, the acquired and categorized cues were assessed based on individual experience and credence. Within this stage, single attributes and combinations of attributes were evaluated concerning perceived quality. This evaluation was defined as the transition to the third stage. Within the third stage, the evaluations of single and combined attributes were compiled. The developed model demonstrated an initial process explanation of how customers perceive quality. The method is differentiated to further methods by including psychological aspects within the perception process.

A psychological concept, which sheds light on the relationship between perceived quality, satisfaction, and behavioral intentions, was presented by Gotlieb, Grewal, and Brown (1994). Primary research questions were the investigation of a complementary or divergent relation between perceived quality and customer satisfaction. Firstly, the context and effect chain were analyzed concerning disconfirmation up to the behavioral intention. Afterward, it was discussed whether the theories resemble the developed process up to the behavioral intention. Subsequently, a process and influence diagram were developed, which showed the relationship, whether positive or negative and in the exact order.

Lastly, an also psychological-based concept was presented by Carbon and Jakesch (2013). Contrary to the previous models, the model focused on the influence of haptic product characteristics on perceived quality.

2.4.4. Perceived Quality in Engineering

An engineering-based approach was introduced by Stylidis et al. (Stylidis et al. 2015; Stylidis, Wickman, and Söderberg 2015a, 2018; Stylidis et al. 2015). The contribution was a perceived quality framework primary for the premium automobile industry. Product quality value was defined by Value-based perceived quality (VPQ), which included the total product experience and had subsections, e.g., brand value. A substage of VPQ was Technical-based perceived quality (TPQ). This stage considered and assessed product attributes related to sensory modalities. Figure 8 shows the perceived quality framework, including the stages VPQ and TPQ.

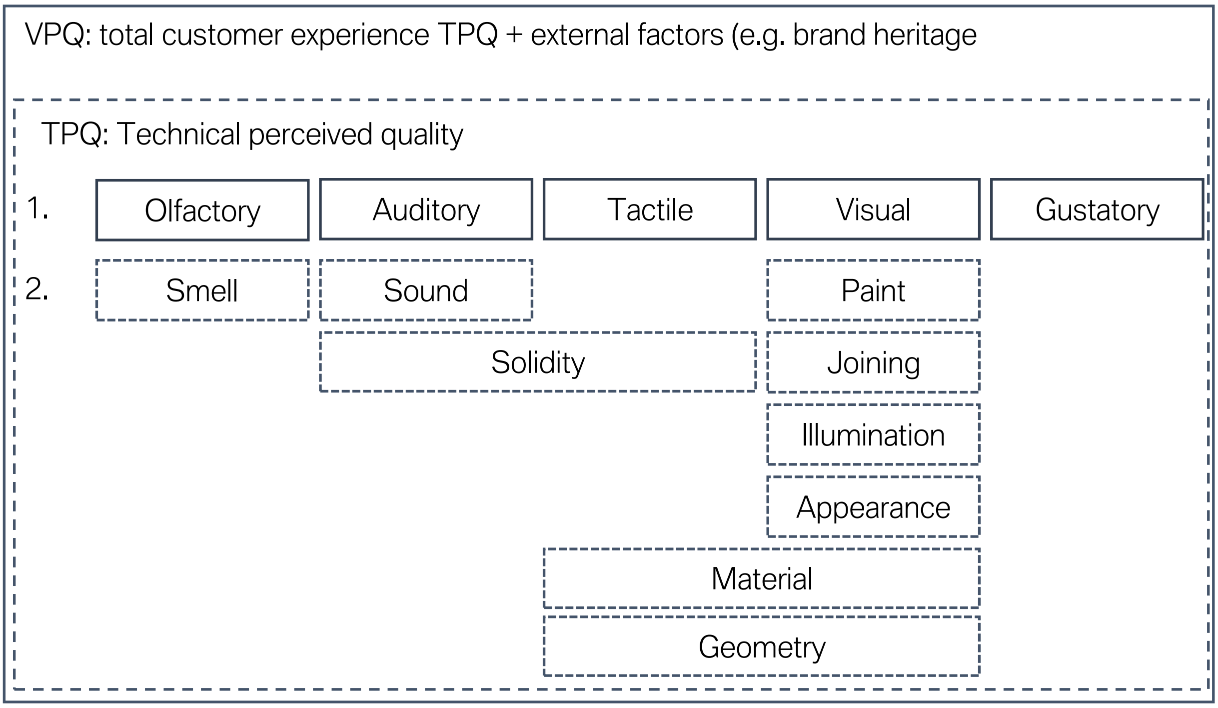


Figure 8. Perceived Quality Framework, adapted from Stylidis et al. (Stylidis et al. 2015; Stylidis, Wickman, and Söderberg 2015a, 2018; Stylidis et al. 2015).

TPQ is defined by different levels. The first level included primary senses assigned to olfactory, auditory, tactile, visual, and gustatory quality. The second stage defined sensory modalities, e.g., smell, sound, or geometry quality. Within the third level, ground attributes were defined and considered. For consideration of those attributes, the framework used ranking methods to weigh the attributes (Stylidis, Wickman, and Söderberg 2018). Given by the decomposition, in the lowest product level, attributes were weighted by ranking methods. By accumulating the results per level, a corresponding value for each primary sense was developed.

Zöller and Wartzack (2017) introduced a product design, which was called Application for Computer-Aided Design for Emotional Impression (ACADE), and demonstrated a method for matching subjective customer requirements to product features. Contrary to other methods, this method demonstrated an approach for designing a product close to customer requirements instead of assessing a final product. Therefore, the method was designed in three stages. The first stage was developed to investigate the product context by using four levels. Firstly, the product context was defined. Secondly, a characterization of the product was conducted. A varying and visualizing of the product in its characteristics was conducted in levels three and four. These steps aimed to develop a product profile. Parallel to this process, a user context was established, which resulted in a personal profile. In the third step, the product and personal profiles were placed on top of each other. The product design was based on personal preferences. Through this procedure, target-oriented customer-oriented product development could be realized. This approach went beyond assessing product attributes regarding perceived quality and showed similarities to Kansei Engineering (Nagamachi 2016). The described procedure objectively captures customer perception and translates it into a product.

Lieb, Quattelbaum, and Schmitt (2008) presented an equally engineering-based approach and included a perception-based assessment of product quality. The relationship between the term perceived quality and product-customer-relationship was explained in detail:

(1) Perceived quality influences the opinion of the customer and, thus, his satisfaction.
(2) The perception of product quality is subjective.
(3) Perceived quality often includes a comparison of several similar products.
(4) Major influencing factors on perceived quality are usage, aesthetics, environmental influences, brand, and reputation.

In order to improve the understanding and handling of perceived quality, this research introduced a three-step method, which included the levels “Identification,” “Quantification,” and “Qualification.”

Similar to the methods of Lieb, Quattelbaum, and Schmitt (2008), Falk, Quattelbaum, and Schmitt (2010) aimed to introduce a customer-centered concept to identify relevant product features and to present an objective evaluation method. A top-down method, similar to PQF, was presented. The method decomposed a product into individual characteristics by reducing the degree of abstraction. The including steps are overall impression, perception cluster, quality attribute, descriptor, and technical parameter. The developed method was applicable within the whole supply chain, and a product evaluation by concrete facts was possible.

2.4.5. Comparison of the Methods

The presented methods demonstrated various approaches to objectify perceived quality and to match customer requirements for a high level of customer satisfaction. Figure 9 shows the presented methods and approaches for assessing perceived quality.

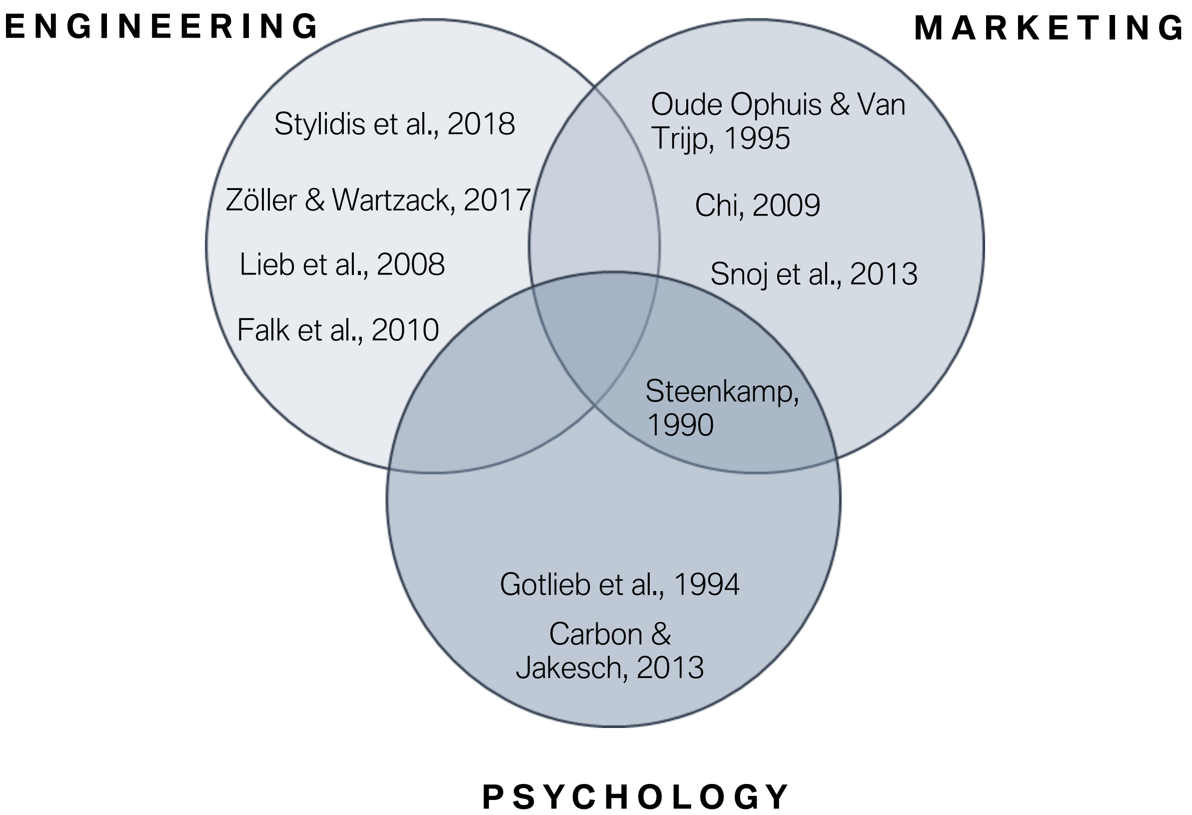


Figure 9. Exemplary perceived quality methods structured regarding the research area.

These four criteria were applied to compare the methods: Research discipline, definition of perceived quality, product context and relations, perceived quality measurement method.

These criteria were derived by the four research hypotheses of the thesis. The first criterium is related to the research discipline. Afterwards, the second criterium reviews the given definitions within the approaches towards the concept of perceived quality. The third criterion considered the product context and relations of attributes regarding the perceived quality or considers the strategy to realize high customer satisfaction. As a perceived quality assessment method is subjective, the fourth criterion considered the availability of a method, and if any exists, the procedure is briefly described. Table 4 presents an overview of the eight presented methods and the four introduced criteria.

Table 4. Comparison of perceived quality methods.

	Research discipline	Definition	Context & relations	Assessment method
Stylidis et al. 2015; Stylidis, Wickman, and Söderberg 2015a, 2018; Stylidis et al. 2015	Engineering/ perceived quality	A product is divided into VPQ and TPQ attributes. VPQ: customer experience & external factors TPQ: technical product aspects	Product-Based, structuring in Visual, Feel, Sound and Smell Quality	Definition of a Perceived Quality Framework and attribute ranking
Zöller and Wartzack 2017	Engineering/ customer satisfaction	Perceived quality as a feature for reaching a high level of customer satisfaction, categorization of objective and subjective product attributes.	Cross individual product attributes.	ACADE, deducing product profile and user profile, matching via product attributes
Lieb, Quattelbaum, and Schmitt 2008	Engineering/ perceived quality	Four statements: 1. Influenced by satisfaction 2. Perceived quality is subjective 3. Comparison of products 4. Different influencing factors	Consideration of usage, aesthetics, environmental influences, brand and company's reputation	Dealing with perceived quality by three steps: Identification, Quantification and Qualification
Falk, Quattelbaum, and Schmitt 2010	Engineering/ perceived quality	Perceived quality is assessed by a cognitive & emotional comparison of a customer's conscious and unconscious experiences.	All sensory stimuli, entire product development process and supply-chain	Approach to assess by product decomposition.

Oude Ophuis and Van Trijp 1995	Marketing/ perceived quality	Based on Aaker (Aaker 1991) overall quality perception by customer, also superiority of product or service	Intrinsic or extrinsic quality cues: bridging the gap between designed quality & perceived quality	Quality guidelines: 1. Identification 2. Separation: perceptions on intrinsic quality and quality attributes 3. Translation
Chi 2009	Marketing/ perceived quality	consumer subjective assessment on product quality, and is influenced by previous experiences and feelings	Perceived quality in relation to brand awareness, brand loyalty & purchase intention effect	
Snoj, Korda, and Mumel 2013	Marketing/ perceived quality	Based on Zeithaml (Zeithaml 1988b): perceived quality is defined in three levels: objective or actual quality, higher abstraction level, global assessment	Correlations between perceived quality, perceived risk and perceived product value	
Steenkamp 1990	Marketing& Psychology/ perceived quality	1. perceived quality involves preference. 2. neither objective nor wholly subjective: i) product evaluated in comparison to others ii) personal iii) situational 3. perceived quality does not reside in the acquisition of the product but in its consumption.	Product perception process including psychological influences and parameters.	theoretical investigation of the quality perception process → process visualization
Gotlieb, Grewal, and Brown 1994	Psychology/ perceived quality	Based on Zeithaml (Zeithaml 1988b): consumer's appraisal of a product's overall excellence or superiority	Focusing on confirming expectations, perceived quality, satisfaction, perceived situational control, and behavioral intentions	

Carbon and Jakesch 2013	Psychology/ perceived quality	perceived quality reveals the opinion about the product quality independent of actual physical qualities by perceiver	Process-based investigation of haptic perception.	Functional, stage-based model of haptic aesthetic processing and relates it to real-world design issues

The term perceived quality was established in all methods, whereby further terms such as perceived value (Snoj, Korda, and Mumel 2013), haptic aesthetics (Carbon and Jakesch 2013), or customer satisfaction (Gotlieb, Grewal, and Brown 1994; Oude Ophuis and Van Trijp 1995; Zöller and Wartzack 2017) were related to perceived quality. The second criteria provided an identical consumer-centric perspective on the perceived quality of every research approach. Some of the methods linked in their definition to further perceived quality definitions, e.g., Zeithaml (1988a) or Shocker and Aaker (2006). Other methods introduced a more detailed definition for perceived quality, which contained more aspects (Steenkamp 1990). A similarity was the highly subjective nature of perceived quality and the individual and experience-based assessment by consumers. A somewhat differently consideration was the third criterium, which considers the context and relations of attributes related to perceived quality. Whereby Stylidis et al. (Stylidis et al. 2015; Stylidis, Wickman, and Söderberg 2015a, 2018; Stylidis et al. 2015) had a broad product scope on perceived quality, Falk, Quattelbaum, and Schmitt (2010) and Lieb, Quattelbaum, and Schmitt (2008) defined perceived quality attributes across the supply chain. Zöller and Wartzack defined perceived quality as an input factor for customer satisfaction (Zöller and Wartzack 2017). Steenkamp (1990) deduced an abstract perceived quality level without considering individual product functions. Snoj, Korda, and Mumel (2013), Gotlieb, Grewal, and Brown (1994) and Chi (2009) considered different perceived quality relationships with additional features. The features were company-related, such as brand names or product-specific or, in the third case, customer-specific, such as satisfaction. Carbon and Jakesch (2013) considered the perception process, influences, and expands the strong visual perception of perceived quality to include influences of haptics. The last criterium considered an assessment method for perceived quality of a product. Four of the presented methods provided a perceived quality assessment method for products; these concepts were introduced by Stylidis et al. (Stylidis et al. 2015; Stylidis, Wickman, and Söderberg 2015a, 2018; Stylidis et al. 2015), Falk, Quattelbaum, and Schmitt (2010), Oude Ophuis and Van Trijp (1995), Zöller and Wartzack (2017), and Carbon and Jakesch (2013). The first three methods were similar and aiming to decompose the product into individual functions, attributes or characteristics. The complexity was reduced, and the individual product functions were objectively evaluated. Subsequently, the individually evaluated functions were accumulated and resulted in a total perceived quality value. As a disadvantage, the related individual functions were individually assessed, and relations between characteristics remain unnoticed. Zöller and Wartzack (2017) described a more holistic approach without decomposing a product. Within this method, the product context and user context were elaborated, and in conclusion, the customer requirements and product attributes were matched. Carbon and Jakesch (2013) examined haptic effects by subdividing the perception process into three levels. The product remained in its entirety, but a quantifiable value for perceived quality was not derived. As the comparison concerning the criteria indicated, some presented perceived quality methods were similar in several aspects across disciplines but differed in the scope of perceived quality and perceived quality assessment methods. In summary, the technical-based approaches decomposed a product to assess perceived quality, whereas marketing-based approaches considered the relation to customers and psychological-based approaches considered the perception process.

2.4.6. Definition Of Perceived Quality in Relation to Split-lines

A consideration of perceived quality definitions of various research approaches identified a lot of similar aspects. In all definitions, the subjective perception of product quality was underlined and highlighted. Subjective criteria often exist in the methods next to the objective product or quality attributes. A distinguishing feature is the abstraction level of perceived quality. A definition of perceived quality by Zeithaml (1988a) illustrated that the perception of the entity is more than just the sum of the individual product elements.

Additionally, Steenkamp (1990) defined perceived quality as the entirety of a product. Both definitions are strongly related to the basis of design theory that the whole is more than just the individual parts' sum. All three definitions raised perceived quality and the shape of objects to an abstract level. Within this thesis, the term perceived quality is examined in detail and the context of an automobile. The assessment of perceived quality is considered at different levels. A holistic consideration of an automobile represents the highest level. Within detailed levels, product features, such as split-lines, regarding perceived quality, are investigated. The scope of this thesis is the consideration of the perceived quality of split-lines. Perceived quality is, therefore, characterized in this thesis by the following statements.

(1) It is subjective and therefore individually perceived,
(2) It stands in context to product design,
(3) It is not independently considerable,
(4) It contributes to the entire product perception and
(5) It depends on personal and product-based factors.

2.5. MACHINE LEARNING

Artificial intelligence is defined as the attempt to create computer programs, which execute intellectual tasks usually performed by humans (Chollet 2018; McCarthy, Minsky, Rochester, & Shannon 1955; Minsky 1982). The research field of artificial intelligence includes learning methods and technologies to perceive input signals or to move. In comparison to human intelligence, Artificial Intelligence is divided into three levels (Kaplan and Haenlein 2019):

(1) Artificial Narrow Intelligence (ANI)
(2) Artificial General Intelligence (AGI)
(3) Artificial Super Intelligence (ASI)

Various functions differentiate Artificial Intelligence systems: Cognitive Intelligence, Emotional Intelligence, Social Intelligence, and Artistic Creativity (Kaplan and Haenlein 2019). Figure 10 shows the research discipline of artificial intelligence and further sub-disciplines.

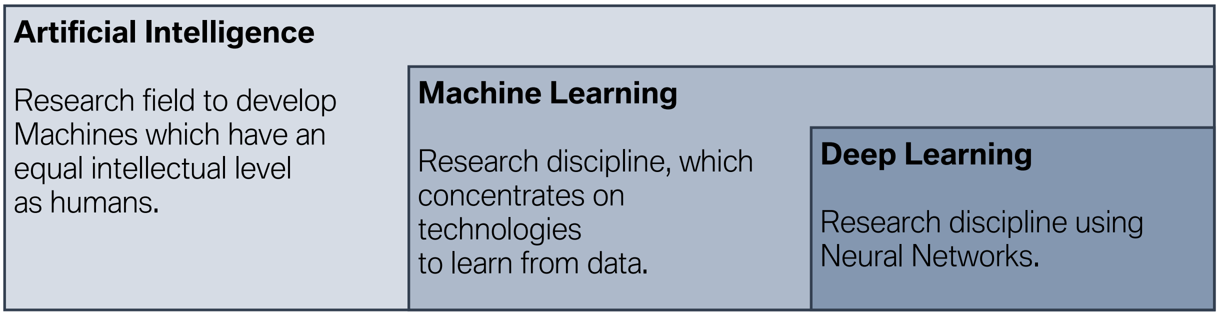


Figure 10. Context of Artificial Intelligence, Machine Learning and Deep Learning, adapted from Chollet (2018).

As illustrated in Figure 10, machine learning is a sub-discipline of Artificial Intelligence (Chollet 2018). It deals with computer learning techniques, which significantly differ from traditional programming approaches. Machine learning is defined as a programming technique in which complex problems are solved by computers using a high amount of input data and their corresponding answers (Chollet 2018; Eraslan et al. 2019; Kaplan and Haenlein 2019). Subsequently, unknown rules and relationships are extracted and learned by the program itself (Chollet 2018; Eraslan et al. 2019; Kaplan and Haenlein 2019). As a result, a trained model can predict answers on a new data set using the developed algorithms. The approach of machine learning stands in contrast to traditional programming approaches. Traditional programs calculate input data by using defined rules to get answers.

For machine learning issues, three aspects are significant to solve (Chollet 2018; Eraslan et al. 2019; Kaplan and Haenlein 2019; Lecun, Bengio, and Hinton 2015):

(1) the input data (i.e., pictures for visual tasks),
(2) examples of the output (i.e., animals of different classes) and
(3) a method to evaluate the performance of a model (i.e., a calculation of the derivation between original and predicted output).

In summary, machine learning programs are “trained rather than explicitly programmed” (Chollet 2018, 5). The central task of machine learning is to find other or further representations of data to approximate rules, which match inputs to correct outputs. The main challenge is the extraction of rules, which are associated with features. The process of extracting them is often called feature engineering (Lecun, Bengio, and Hinton 2015).

Machine learning approaches differ in the input and output data and the type of problem they solve (Chollet 2018; Kaplan and Haenlein 2019). The approaches are commonly defined in three approaches: Supervised learning, unsupervised learning, and reinforcement learning. An overview of these training approaches is presented in Figure 11.

Unsupervised Learning	**Supervised Learning**	**Reinforcement Learning**
Clustering Algorithms Dimension Reduction Anomaly Detection	Classification Algorithms Regression Algorithms	Agent Systems Action Strategy

Figure 11. Overview of different machine learning algorithms, summarized from Lanquillon (2019).

Supervised learning is present if the learning data set is represented by input data combined with labels, often determined by experts or humans (Chollet 2018; Eraslan et al. 2019; Kaplan and Haenlein 2019; Lecun, Bengio, and Hinton 2015; Schmidhuber 2015). Classification or regression algorithms represent supervised learning methods (Chollet 2018; Eraslan et al. 2019). Examples of methods using supervised learning approaches are image classification or regression for predicting purchase decisions. Unsupervised learning aims to find data transformations without getting information about the output (Chollet 2018; Eraslan et al. 2019; Kaplan and Haenlein 2019; Lecun, Bengio, and Hinton 2015). An example of using unsupervised learning is clustering, i.e., for grouping customer data. In contrast to supervised learning and unsupervised learning methods, a third learning approach is called reinforcement learning. It has an optimization request besides the challenge's classification to reach a maximum or minimum in combination with various decisions (Chollet 2018; Kaplan and Haenlein 2019; Schmidhuber 2015).

Besides machine learning approaches, models, which are used for training, are necessary to introduce. There are different types of models developed within research, which can perform one or more machine learning approaches. The most common models are decision trees, support vector machines, linear regression, logistic regression neural networks, and convolutional neural networks for image classification with a focus on unsupervised and supervised learning algorithms. The categorization of these models to a learning approach is presented in Figure 12.

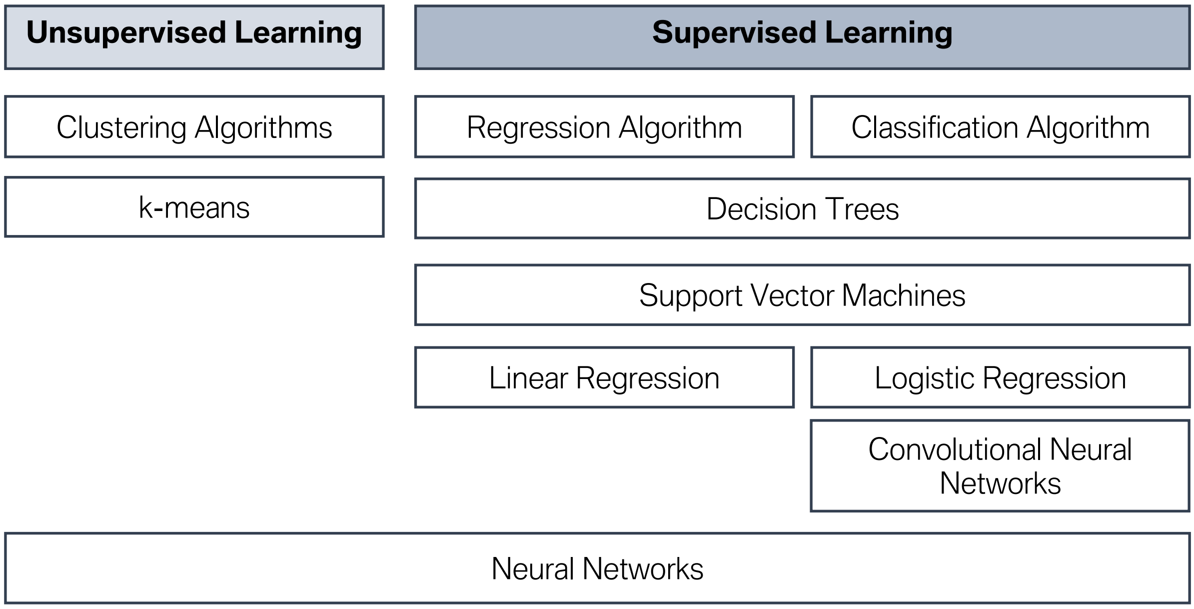


Figure 12. Categorisation of machine learning approaches.

2.5.1. Decision Trees

Decision trees are a machine learning method to document a decision-making process based on past knowledge (Ben-Gal et al. 2014). Decision trees are non-parametric algorithms with a fast calculation time (Richter 2019a). The decision trees' approach uses an amount of attributes/pair values to classify an object (Beierle and Kern-Isberner 2019). Besides the ability of classification, decision trees can also perform a regression algorithm (Richter 2019a). The architecture of a decision tree is related to Boolean function (Beierle and Kern-Isberner 2019). A truth-value marks a tree's leaves; each node is marked by an attribute (Beierle and Kern-Isberner 2019). A classification task is performed by going through a tree's different nodes (Beierle and Kern-Isberner 2019). The training of a decision tree is performed by the architecture modeled (Figure 13), and the training data listed in the form of a table (Table 5) (Kotsiantis 2013).

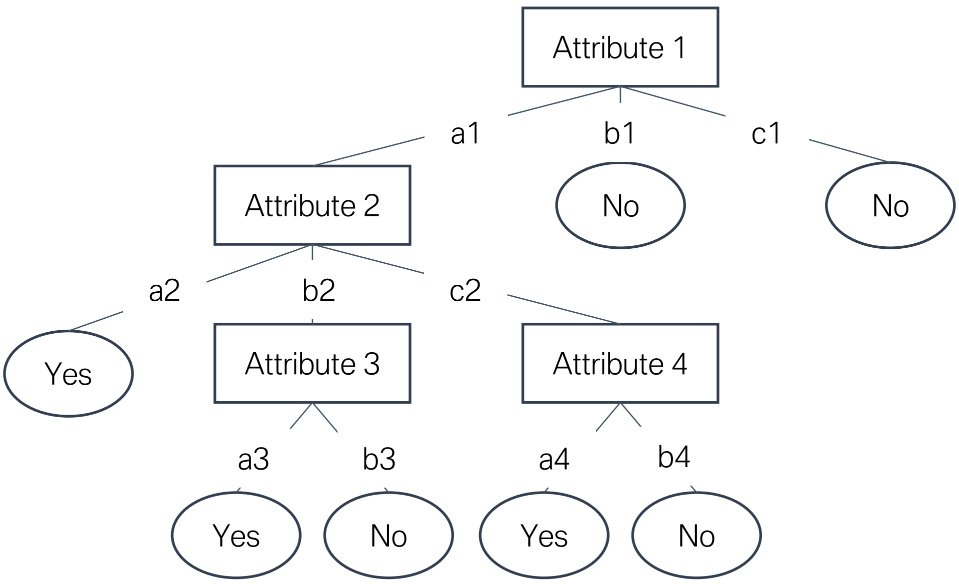


Figure 13. Example of a decision tree architecture, adapted from Kotsiantis (2013).

The training set is provided within a table, which is exemplary presented in Table 5.

Table 5. Training data set for a decision tree machine learning approach, adapted from Kotsiantis (2013).

Attribute 1	Attribute 2	Attribute 3	Attribute 4	Class
a1	a2	a3	a4	Yes
a1	a2	a3	b4	Yes
a1	b2	a3	a4	Yes
a1	b2	b3	b4	No
a1	c2	a3	a4	Yes
a1	c2	a3	b4	No
b1	b2	b3	b4	No
c1	b2	b3	b4	No

Various iterations perform one training cycle. A single iteration uses a defined path through the tree, which is described by the attributes (Attribute 1-4, Figure 13) and the connections (a1-a4, b1-b4, c1, Figure 13) (Kotsiantis 2013).

By using this general approach, different decision tree variants are developed, which are for example binary trees, top-down induction of decision trees (TDIDT), ID3, C4.5, CART (Gray 1990; Kotsiantis 2013).

2.5.2. Support Vector Machines

Support Vector Machines (SVM) represents a machine learning method, which can be applied within classification or regression tasks (Lanquillon 2019). An example of SVM's application is the classification of the fraudulent usage of credit cards by using a massive amount of information (Noble 2006). For realizing an SVM algorithm, four concepts are decisive: the separating hyperplane, the maximum-margin hyperplane, the soft margin, and the kernel function (Noble 2006). The separating hyperplane is a mathematical function, which divides two bundles of points by a function. The distance between the divided bundles is equal (Figure 14 a) (Cortes and Vapnik 1995; Noble 2006). The mathematical function, which separates the bundles by having the maximal distance to both bundles, is called maximum-margin hyperplane (Figure 14 a) (Noble 2006). This hyperplane results in the optimal hyperplane with an optimal margin (Cortes and Vapnik 1995). These functions are used by having a problem, which is separable by using a linear function. The soft margin is applied if single points of one bundle are within the other bundle. The soft margin is applied to handle such a situation, which allows some data points of one bundle to push through the margin without affecting the result (Figure 14 b)) (Noble 2006). The kernel function is used to separate data sets with more than one dimension (Noble 2006). The fundamental concept of the kernel function is to add another dimension to the data by squaring the original expression values (Figure 14 c),d)) (Cortes and Vapnik 1995; Noble 2006). In summary, the kernel function is a mathematical trick that enables running an SVM algorithm on one-dimensional data by transforming data of low dimensions in a higher-dimensional space (Noble 2006). By using a suitable kernel function, it is possible to separate any data set linearly and to handle nonlinear problems (Lanquillon 2019; Noble 2006; Schölkopf et al. 2002). The presented concepts of SVM are presented in Figure 14.

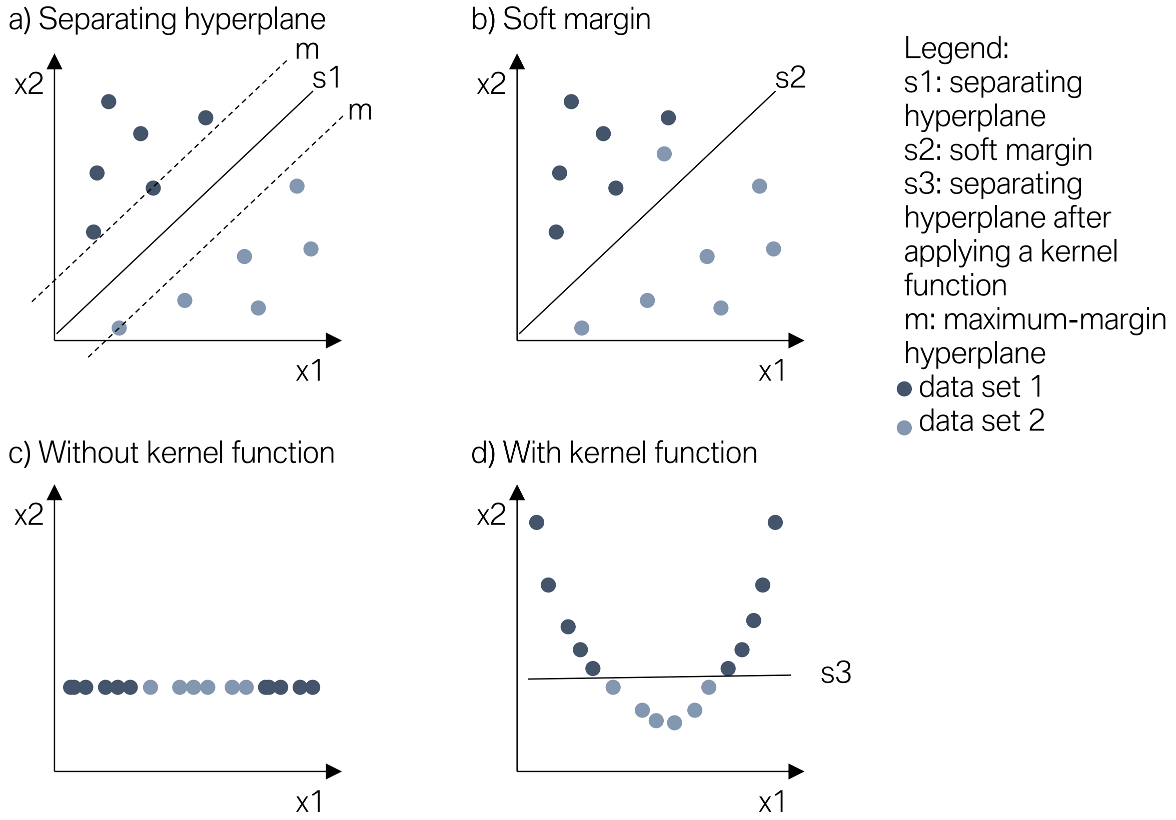


Figure 14. Concepts of SMVs. a) separating two bundles of data points with a separating hyperplane and showing the maximum-margin hyperplane. b) showing the concept of soft margin. c) one dimensional data set. d) transformed data set by the kernel function and defining a separating hyperplane, adapted from (Noble 2006),(Cortes and Vapnik 1995).

Within a basic form, SVMs are applicable for two different variables. The SVM algorithm has to be extended to a multi-class SVM for using more than two variables (Noble 2006). Therefore, the two possibilities are presented. A first possibility is a one-versus-rest approach, where each class is differentiated from each (Cortes and Vapnik 1995; Noble 2006). The second possibility is a one-versus-one classification, where respectively, two classes are distinguished (Cortes and Vapnik 1995; Noble 2006). Some advantages of SVMs are a high prediction accuracy, a low tendency for overfitting, efficient classification of new objects and compact models (Burges 1998; Cortes and Vapnik 1995; Cristianini and Shawne-Taylor 2000; Joachims 1998). Disadvantages of SVMs can belong to training time, complex implementations, and trained models are hard to evaluate (Burges 1998; Cortes and Vapnik 1995; Cristianini and Shawne-Taylor 2000; Joachims 1998).

2.5.3. Linear Regression

Linear regression has its fundamentals within statistics (Richter 2019b). A metrical target variable is calculated using further metrical attributes when using a linear regression model (Lanquillon 2019). Within the primary case, the relation between the target variable and other explanatory variables is described by a linear function or a more dimensional space by a linear hyperplane (Lanquillon 2019). A linear regression model within a two-dimensional space is presented in Figure 15.

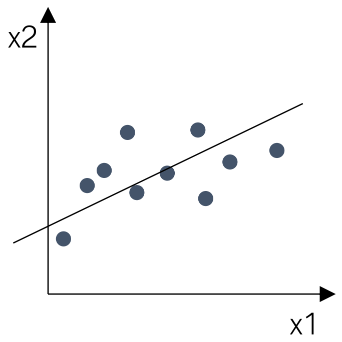


Figure 15. Linear Regression within a two-dimensional space, adapted from Shalev-Shwartz and Shai (2014).

Linear regression models can also cover non-linear function, the term linear stands for the coefficients in the function and do not limit the degree of features (Lanquillon 2019; Richter 2019b; Shalev-Shwartz and Ben-David 2014). An advantage of linear regression models is that the parameters of a linear function can be estimated by using the least square method (Shalev-Shwartz and Ben-David 2014).

2.5.4. Logistic Regression

Logistic regression is an algorithm for classification tasks, which is a predictive analysis algorithm using probability (Lanquillon 2019). The primary use of logistic regression models is to calculate the probability of a binary event related to different explanatory variables (Lanquillon 2019). Therefore, logistic regression uses the sigmoid function to calculate the class probability of inputs (Dreiseitl and Ohno-Machado 2002). The sigmoid function is presented in Figure 16.

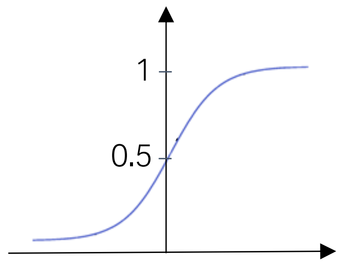


Figure 16. Sigmoid function, used in logistic regression algorithms, adapted from Shalev-Shwartz and Shai (2014).

The basic function of a logistic regression model is (Park 2013)

$$logit(y) = \alpha + \beta X. \quad (1)$$

The underlying function is binominal, and the parameters are estimated using the maximum likelihood method (Lanquillon 2019; Park 2013; Shalev-Shwartz and Ben-David 2014). A likelihood is a probability to predict the dependent variable using the values of the independent variables (Park 2013). The advantage of logistic regression models is a low risk of overfitting the network using a variable selection method (Dreiseitl and Ohno-Machado 2002). Nevertheless, this decreased risk goes hand in hand with reduced flexibility of logistic regression models (Dreiseitl and Ohno-Machado 2002).

2.5.5. K-Means

The k-means algorithm was developed and presented by MacQueen (1967). The k-means algorithm is an approach for clustering input data and is therefore allocated to the unsupervised learning methods (Lanquillon 2019; Wagstaff et al. 2001). The algorithm clusters input data points to a priori defined number of clusters (Lanquillon 2019). Therefore, the algorithm minimizes the distance between the cluster center and the single points of the data set (Lanquillon 2019). The beginning is a random mapping that is iteratively optimized using four steps (Lanquillon 2019; Richter 2019b; Shalev-Shwartz and Ben-David 2014). In the first step, the cluster centers are places in the space of data sets (Kodinariya and Makwana 2013). In the second step, the single data points are mapped to one cluster center, to which the distance from the point to the center is minimal (Lanquillon 2019). In a third step, the center's value is calculated by the mean value of all distances (Lanquillon 2019). In a fourth step, the cluster centers are newly located (Kodinariya and Makwana 2013). This procedure is repeated until the centers are optimal and do not move (Kodinariya and Makwana 2013). The optimum results after a finite number of iterations and the algorithm converge at a local minimum (Lanquillon 2019). Therefore, the algorithm is typically performed several times using several different values for the number of clusters and different starting values (Lanquillon 2019). The cluster center's data point is smaller by having more defined cluster numbers (Lanquillon 2019). Therefore, the number of cluster centers must be limited to provide overfitting of the algorithm (Lanquillon 2019).

2.5.6. Neural Networks

Within standard machine learning methods, the extraction of essential features is performed manually, and for many problems, this effort exceeds feasible limits. Therefore, a subdiscipline of machine learning was developed: deep learning (cf. Figure 10). The effort of deep learning methods is that feature engineering is performed automatically by the method itself (Lecun, Bengio, and Hinton 2015). Thereby, deep learning methods can map complex and multidimensional data and solve more scientific and industrial problems. (Deep) neural networks represent the idea of deep learning. Neurobiology and the abstract concept of a human brain's neurons inspired their name, but neural networks do not aim to represent the human brain (Chollet 2018). Within a neural network architecture, a neuron displays the fundamental element (Husain, Dellen, and Torras 2017). Each layer within a neural network includes several neurons (Husain, Dellen, and Torras 2017). The combined neurons form the output of one layer l and the input of the subsequent layer $l+1$ (Husain, Dellen, and Torras 2017). The relation of output and input layers are defined by,

$$a^{l+1} = f\left(W^l a^l + b^l\right), \tag{2}$$

where W^l is the weight matrix, l the layer, b^l the bias and f the activation function for layer l provided by a^l (Husain, Dellen, and Torras 2017). The training of a neural network implies to find a set for W and b with a minimal cost function (Husain, Dellen, and Torras 2017). Figure 17 shows the approach of defining an optimal set of parameters for W and b.

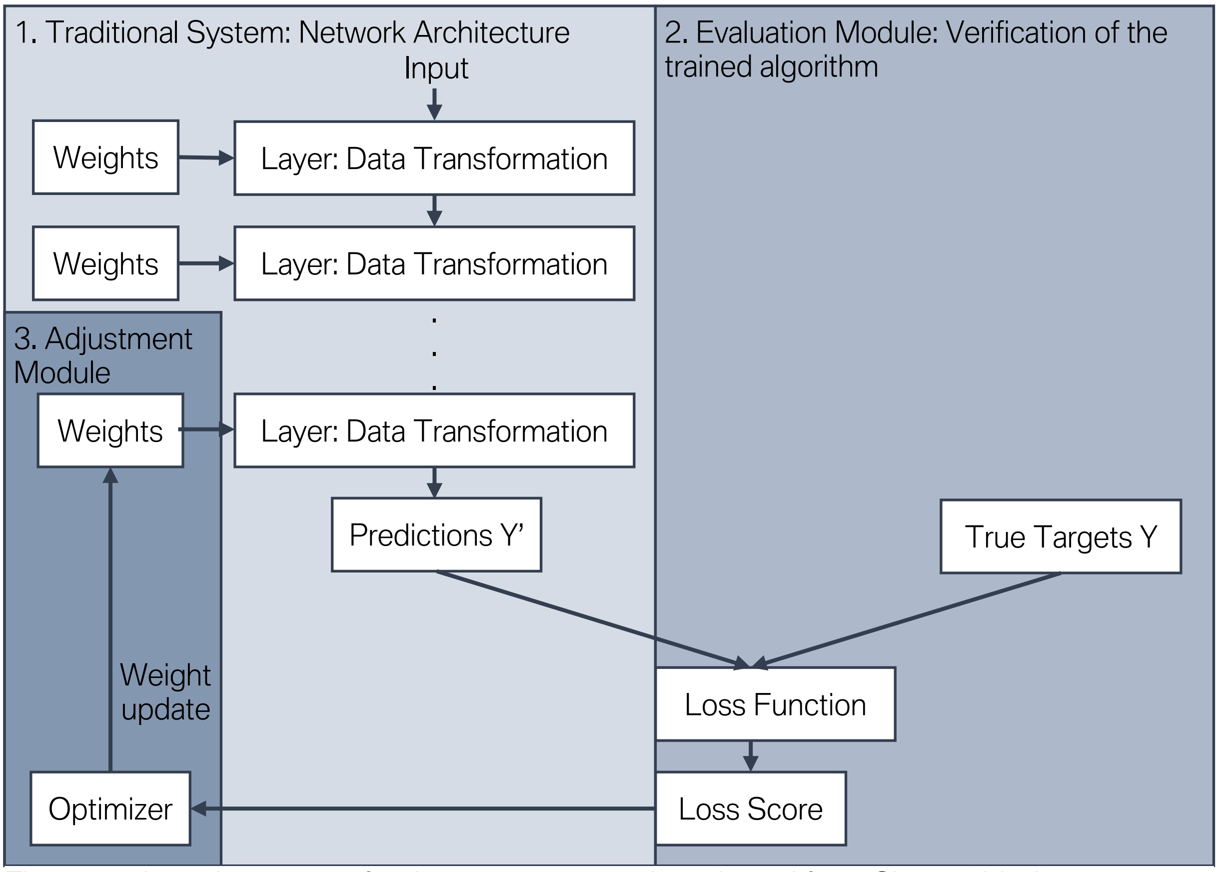


Figure 17. Learning system for deep neural networks, adapted from Chollet (2018).

Within this process three subsystems are essential: the traditional system, the evaluation module, and the adjustment module. Within the first traditional system, the architecture of a neural network is defined by neurons, layers and weights. The evaluation system considers the calculation of the loss function, which is the accuracy of the network predictions related to the original targets y. In this case, the loss function is calculated to measure the derivation between original target and predicted output. For one training sample, the loss function is defined as,

$$J(W, b;\, y', y) = \frac{1}{2} \|y' - y\|^2. \tag{3}$$

As a base for the subsequent adjustment system, the result of the loss function is considered. In combination with the results, the model weights are modified to get a higher prediction accuracy by reducing the loss function due to updated weights. This process is called backpropagation and is an integral approach of neural networks (Chollet 2018; LeCun and Bengio 1995). For this purpose, an algorithm (i.e., gradient descent) is used to compute the gradients concerning the loss function (Chollet 2018; Eraslan et al. 2019; Husain, Dellen, and Torras 2017).

The difference between deep learning and machine learning approaches is automatizing feature engineering, which aims to identify layers with a more useful representation of input data to predict an output. The advantages of neural networks are summarized (Chollet 2018):

(1) Simplicity: Given the automation of Deep Neural Network's feature engineering, the engineering requirements are reduced, and end-to-end solutions are achievable.
(2) Scalability: High level of parallelism is possible due to the usage of graphics processing units (GPU) or tensor processing unit (TPU).
(3) Versatility and reusability: Approaches are recyclable, and for example, a trained network for image classification can be reused for a video processing network or smaller datasets.

2.5.7. Convolutional Neural Networks

Convolutional neural networks (CNN) are up to many scientific and industrial fields, such as autonomous driving (Ciresan et al. 2012; Lecun, Bengio, and Hinton 2015). Besides classification problems, convolutional neural networks can perform object detection, multi-object detection, and image segmentation problems. Within an object detection algorithm, certain classes were locally predicted within an image (Chollet 2018). Similarly, multi-object detection algorithms identify various classes within an image (Chollet 2018). Image segmentation problems assign individual pixels to a particular class (Chollet 2018).

Common concepts of convolutional neural networks include three concepts (LeCun and Bengio 1995): local receptive fields, shared weights, and spatial or temporal subsampling. Figure 18 shows the architecture and the processing of a convolutional neural network by sketching convolutions and subsampling.

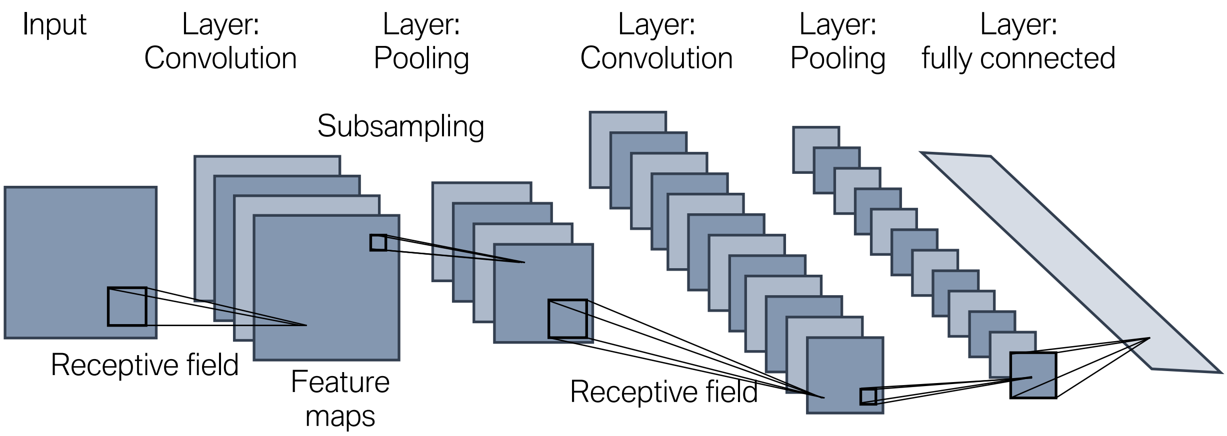


Figure 18. Architecture of convolutional neural networks, adapted from LeCun and Bengio (1995).

Local receptive fields (Figure 18) can extract basic features, i.e., edges combined by higher layers to display more abstract patterns or total images. Besides, different receptive fields within one image perceive features using a detector with identical weights over the entire image (shared weights). The result of calculating an image with a receptive field is called a feature map. In conclusion, various feature maps (Figure 18) with different weights are included within a convolutional layer (LeCun and Bengio 1995). Since a feature is detected and extracted, the local position is irrelevant unless a connection to other features is available. Within the next step, a subsampling of the feature maps is performed to reduce the image size and increase network robustness. A typical arrangement of convolutional neural networks is convolution, and subsampling layers are in the form of a pyramid, displaying an increasing number of feature maps by having a reduced spatial resolution (Chollet 2018; LeCun and Bengio 1995).

Based on the presented common concepts of convolutional neural networks, the following section focuses on the different types of layers. The layers within a convolutional neural network alternate between convolutional layers, pooling layers, and concludes with fully connected layers

(Chollet 2018). Independently of the layer type, all neurons of one layer have similar input weights (shared weights, cf. 2.5.6). Within the convolutional layer, an activation function is used to detect local patterns. As activation function is binary step function, linear activation function or non-linear activation function are selectable (Eraslan et al. 2019). Linear functions have limitations regarding image and speech recognition (Lecun, Bengio, and Hinton 2015). Therefore, non-linear activation functions were applied within convolutional neural networks, such as the sigmoid or the Rectified Linear Unit (ReLU) function. The ReLU function especially demonstrated good results in machine vision (Ramachandran, Zoph, and Le 2017; Schmidt-Hieber 2017).

Additional non-linear activation functions are hyperbolic tangent (tanh), leaky ReLU, parametric ReLU, softmax, and swish (Lecun, Bengio, and Hinton 2015). Compared to the other activation functions, the ReLU function trains faster than, for example, the tanh function (Lecun, Bengio, and Hinton 2015). The output of convolutional layers is feature maps (Goodfellow, Bengio 2015), which are processed further. Pooling layers realize a subsampling of the feature maps by applying a filter. For this filter, the max pool function is often applied (Chollet 2018). The last layers are typically fully-connected, which are also called dense layers. These layers differentiate from convolutional layers by the fact that they learn global patterns (Chollet 2018). In preparation for the dense layer, the convolutional layers' multidimensional outputs are converted into a one-dimensional vector. This step is called flatten (Hinton, Sutskever, and Salakhutdinov 2014). As activation function for the output layer of a classification problem, the sigmoid or the softmax function are implementable. Based on the number of output variables, one of these functions is used. The sigmoid function is applied by having two output labels, and the softmax function is applied by having more than two labels (Chollet 2018). Additionally, within the final layer, a loss function and an optimizer are defined. The loss function, or cost function, is defined as the function which has to be minimized within the training (Goodfellow, Bengio, and Courville 2016). The initial problem indicates the choice of loss function and is directly linked to the activation function of the output layer (Goodfellow, Bengio, and Courville 2016). Regression problems often use a linear activation function and a mean-squared error loss function (Chollet 2018). Binary classification problems are often modeled using a sigmoid activation function and a binary cross-entropy as a loss function (Chollet 2018). Multi-class classification problems are often solved by applying the softmax activation function and a categorical cross-entropy as a loss function (Chollet 2018). Cross-entropy is a method based on information theory that calculates the distance between prediction and correct label (Chollet 2018).

Several network architectures were developed in research, which is, i.e., VGG-16 (Simonyan and Zisserman 2014), ResNet (He et al. 2016), Inception (Szegedy et al. 2015), or AlexNet (Krizhevsky, Sutskever, and Hinton 2012). These architectures generally include the presented layers within their architecture, but the specific combination, selection of local receptive fields, shared weights, and spatial or temporal subsampling are different.

2.5.8. Siamese Networks

A dual-input of two VGG-16 networks was used within the thesis, which was combined directly after the weighted layers. For the implementation of this network architecture, the concept of Siamese networks was exploited. Initially, Siamese networks were introduced by Bromley et al. (1993) and were used to identify and verify signatures. The basic idea was to use two identical subnetworks, which were combined at their output. Additionally, the subnetworks had identical architecture and had identical shared weights (Zagoruyko and Komodakis 2015). Each branch got one input, and this was processed by several convolutional ReLU and pooling layers (Zagoruyko and Komodakis 2015). Subsequently, the outputs of the subnetworks were combined. These were the input for the decision network. This type of network was represented by fully-connected and ReLU layers (Zagoruyko and Komodakis 2015). The derivation between both outputs provided information about the similarity of their features (Bromley et al. 1993). The architecture of Siamese neural networks is presented in Figure 19.

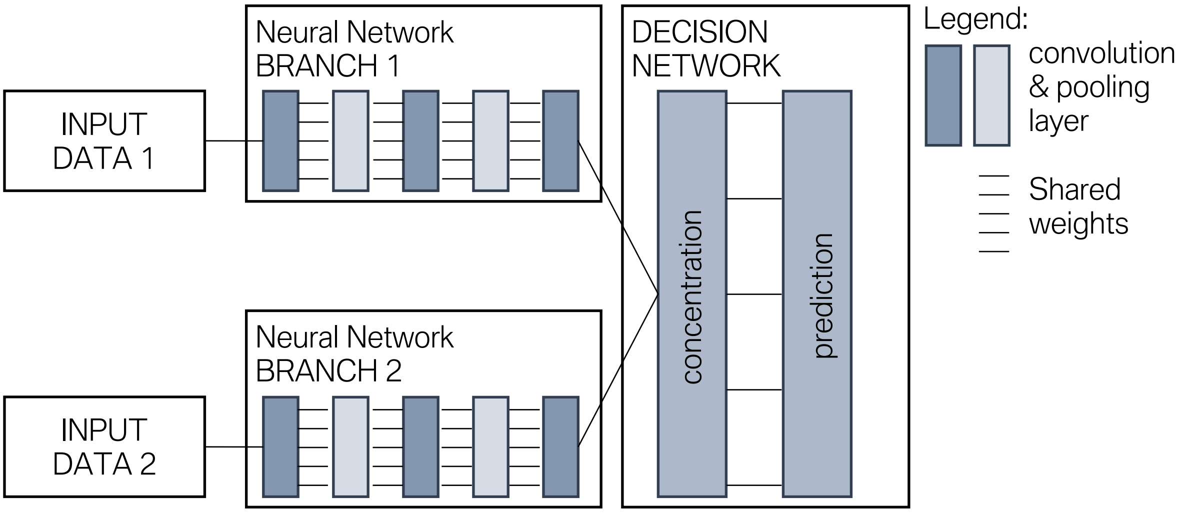


Figure 19. Architecture of Siamese networks, adapted from Bromley et al. (1993).

Based on this idea, various studies applied Siamese networks on several levels, for example, to compare image patches (Zagoruyko and Komodakis 2015), to improve one-shot learning (Koch 2015), to reidentify humans (Varior, Haloi, and Wang 2016), and using a VGG-16 for easy visual tracking (Li and Zhang 2019). Siamese networks were used to compare two data inputs towards their similarity or create a kind of memory (Koch 2015; Li and Zhang 2019; Varior, Haloi, and Wang 2016; Zagoruyko and Komodakis 2015). In the present research, Siamese networks were primarily used to increase the robustness and the reliability of the network.

2.5.9. Explainable AI

Massive expenditure on machine learning and convolutional neural networks had required evaluation techniques at the same time. Therefore, methods were derived from gaining insights into the network relevant features. Given that convolutional neural networks optimize weights and thus the feature maps themselves, it was essential to decode the black boxes and get an insight into them (Chollet 2018). Explainable AI methods were developed to realize decoding. The idea of explainable AI methods is presented in Figure 20.

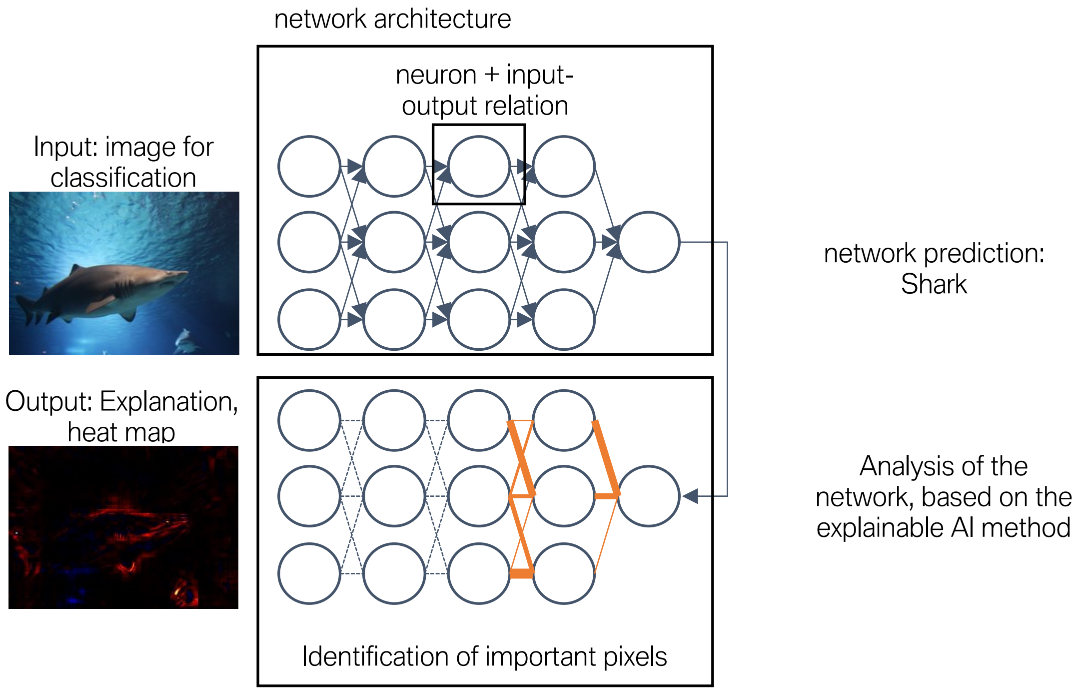


Figure 20. Approach of Explainable AI methods for visualizing decision relevant pixels, images from Chollet (2018).

The central objective of all methods is the derivation of heatmaps, which highlight relevant pixels. The initial step is to predict an input image by a trained neural network. The second step is to analyze the network architecture after a prediction. Thereby, a necessary pixel can be derived from three different methods. Thus, explainable AI methods are categorized concerning those three methods: attribution, function, and signal (Kindermans et al. 2017). Table 6 shows the three categories, an explanation, and examples of the developed methods.

Table 6. Categorization of explainable AI methods, adapted from Kindermans et al. (2017).

Method	Explanation	Examples
Attribution	Contribution of the signal dimensions to the output through the layers (Kindermans et al. 2017)	Deep Taylor (Müller et al. 2016), Layer-wise Relevance Propagation (Bach et al. 2015)
Signal	Activities are traced back to the input pixel and it is shown which part of the input caused which activation of the features maps (Kindermans et al. 2017; Zeiler and Fergus 2014)	Guided Backpropagation (Springenberg et al. 2014)
Function	Investigation of the functions the model uses to predict the output. By making small changes to the input it is possible to see how the output changes. (Kindermans et al. 2017)	SmoothGrad (Smilkov et al. 2017), Integrated Gradient (Sundararajan, Taly, and Yan 2017)

Attribution-based methods are based on the signal dimension and the contribution of different layers concerning the output layer (Kindermans et al. 2017). The relevance of inputs is calculated by individual contributions of the output using the backpropagation (Kindermans et al. 2017). Two methods within attribution-based methods were Deep Tylor and Layer-wise Relevance Propagation (Kindermans et al. 2017).

Deep Taylor is an explainable AI decomposition method introduced by Müller et al. (2016). The objective was to derive a heatmap for extracting feature relevant pixels by tracking an input image through the non-linear structure of a network. The basic idea was a divide-and-conquer method, which split the trained functions and the weights of a network. These split functions were applied either for local image pixels or within a particular abstraction level. Additionally, this research's approach deduced a top-down dependency between neurons and their particular relevance, which depended again on other neurons and their relevance.

The second approach of attribution-based methods was presented by Bach et al. (2015), which is called Layer-wise Relevance Propagation (LRP) and considers non-linear classifiers of a network (Bach et al. 2015). LRP demonstrates an unsupervised process during which individual feature maps were disassembled to discover relevancies for a single pixel. Additionally, the method also used the last neuron. Therefore, the relevancies were calculated for the entire network architecture from the output layer to the input layer by adding each pixel's relevance. The result was a heatmap highlighting areas with high relevance and areas with low or zero relevance for the actual prediction.

The second method category is signal-based explainable AI methods. The input patterns are analyzed, which causes a triggering of the feature maps (Zeiler and Fergus 2014). Therefore, the network signal network is calculated, which causes an activation within feature maps (Kindermans et al. 2017).

A signal-based explainable AI method is Guided Backpropagation, developed by Springenberg et al. (2014). This method used a deconvolutional approach for underlining prediction-relevant pixels: Guided Backpropagation used the idea of Zeiler & Fergus (2014) for implementing a deconvolutional network to visualize relevant parts of an image. Neurons and their contribution are analyzed within high-level feature maps, and a reassembled image is formed based on the importance of a neuron—a mapping result, which displays decision-relevant areas of the input image.

Function-based methods are the third explainable AI methods. The operations are identified, which a trained network uses (Kindermans et al. 2017). By applying function-based methods, the network is changed, and based on changes; the functions can be extracted. The generated heatmaps visualize the functions used for classification. Due to the non-linear structure of deep neural networks, the function extraction is only approximate (Kindermans et al. 2017).

The SmoothGrad method was deduced by Smilkov et al. (2017) and implemented the idea of derived gradient-based sensitivity maps. For calculating these maps, the noise was added to a particular input image, and this additional noise had a smoothing effect, which contributed to highlighting relevant pixels. These effects were enhanced by enriching the training with images of random noises.

The integrated Gradient method was presented by Sundararajan, Taly, and Yan (2017). This method distinguished between the further explained methods because of its simple and easy implementation. For generating a heatmap based on relevant pixels or areas, the method calculates the invariance of gradients in connection with sensitivity techniques.

2.6. Machine Learning Approaches in Perceived Quality

Machine learning methods for quality assessments have been successfully tested and applied in different industrial areas. These methods range from the textile industry to wood processing and food quality checks. In this thesis, six different methods and research approaches were considered.

Chang et al. (1997) presented an approach for cork quality classification in 1998. The advantage of the machine learning-based classification was standardization and process automation. A combined approach of morphological filtering, contour extraction, feature extraction, and neural networks was chosen for implementation. Two two-dimensional images of a cork top and a cork bottom were used as input data. The top image had 288×170 pixels, and the bottom image had 288×93 pixels. As a result, a rejection error of the machine learning method of 6.7 % was presented, rated as excellent by experts. The long computing time was identified as a weakness of the method. For further research potential and an application in the running production, a real-time prediction of the methods had to be reached (Joongho Chang et al. 1997).

Another use of neural networks for quality control was developed by Bahlmann, Heidemann, and Ritter (1999) in 1999. Their object was textile seams, whereby the human-based evaluation process was to be replaced by a standardized and time-optimized process. The neural network was defined by an unsupervised learning approach combined with feature vectors (Kohonen 1982). The neural network training data was acquired based on construction, whereby a camera and the illumination could be optimally adjusted and varied. As a result of the data acquisition, 2D images with 80×512 pixels could be obtained. The classification result of the neural network was according to expert-based evaluations, and the network reached a normalized mean square error (NMSE) of 0.13. The evaluation time could be reduced by 29 s using the neural network for quality assessment (Bahlmann, Heidemann, and Ritter 1999).

Edwards et al. (1999) had achieved a further use of neural networks for quality control in the paper industry. One quality parameter in the paper production process was the paper curl. The examination of this parameter was complicated, as the paper curl could only be inspected beside the running production process. The idea was to predict paper quality using ten parameters. An advantage of a neural network-based prediction was the saving of quality control during running production. As a machine learning method, multilayer perception neural networks were used. The classification performance was evaluated using a confusion matrix, and the prediction results showed an accuracy of approximately 80.0 % (Edwards et al. 1999).

Another application of a neural network quality assessment was developed by Packianather and Drake (2000) in the wood industry. The application was to classify 13 - 17 different wood errors, e.g., holes or rotten knots. Traditionally, a stressful expert-based assessment was used with short assessment intervals. Evidenced by further literature was an expert-based accuracy of wood error detection between 55 % and 68 %. For training the neural network, two-dimensional images of 512 × 512 pixels in greyscale were used. As a first result, the network scored 86.5 %, which was a higher accuracy compared to the expert-based assessment. Additionally, the prediction time of the neural network was between 0.16 - 0.11 s per image. In summary, the trained neural network performed better than an expert-based assessment due to the accuracy and the time (Packianather and Drake 2000).

Pan et al. (2017) presented a study in the automotive environment, which evaluated the aesthetics of automobiles in customer satisfaction. In a first step, a conditional generative adversarial network (cGAN) was trained to create downscaled images of automobiles by filtering design elements, e.g., body type. A second stage was designed to classify the cGAN-generated images. Therefore, a Siamese network was used for a customer-centered evaluation. With the network's assistance, a prediction in how a particular customer perceives a specific automobile design. The prediction result was evaluated in a third stage using guides backpropagation to visualize design relevant features on the automobile. The network accuracy was concerning four different design attributes (appealing, sporty, luxurious, innovative) between 67.3 % and 75.4 % (Pan et al. 2017).

Gussen, Ellerich, and Schmitt (2019) presented a methodology for combining objective and subjective surface quality assessments. The approach was to collect objective, measurable parameters of surfaces and conduct a study for a subjective assessment of similar surfaces. Subsequently, the collected data will be combined by using machine learning methods. This procedure was presented as an idea and a future approach. In detail, surfaces will be analyzed by using a visual, tactile, and acoustic sensor. In parallel, the similar surfaces will be assessed by conducting a study and asking participants to assess the surfaces' perceived quality on a scale between 1 and 7. Consequentially, different machine learning types will be tested to perform a regression, classification, or clustering. Therefore, linear and logistic regression, super vector machines, k-nearest neighbors, decision trees, neural networks, k-means, and expectation-maximization will be investigated. Gussen, Ellerich, and Schmitt (2019) presented an idea and an approach for implementation; first results and an implementation of the machine learning methods are missing.

The Table 7 presents a summary of the presented methods.

Table 7. Comparison of methods using machine learning for quality assessments.

	Industry/ Object	Method	Data	Result
Chang et al. 1997	Cork quality classification	Mixed method, using morphological filtering, contour extraction, fuzzy-Neural Network	2D cork images, (top & bottom)	rejection error 6.7 %
Bahlmann, Heidemann, and Ritter 1999	Quality classification of textile seams	Neural Network: Kohonen Map and feature vectors	2D images 80×512 pixels	NMSE 0.13
Edwards et al. 1999	Prediction of paper curl quality	multilayer perceptron (MLP) Neural Network	10 different production-based parameters	accuracy 80.0 %
Packianather and Drake 2000	Classification of wood errors	Neural Network	2D images in greyscale	accuracy 86.5 %

Pan et al. 2017	Classification of automobile designs for certain customer	cGAN, Siamese network, guided backpropagation	2D data of cars	Accuracy: appealing: 67.3 %, sporty: 75.1 % luxurious: 75.1 % innovative: 75.4 %
Gussen, Ellerich, and Schmitt 2019	Visual, haptic and acoustic quality of surfaces	Regression, classification and clustering	Sensory data and study results	No results presented

3. SUMMARY AND RESEARCH QUESTIONS

Perceived quality in research includes aspects of marketing, psychology, and engineering. Especially in the automotive industry, perceived quality is a distinguishing feature between competitors. As a result, strategies for realizing a high level of perceived quality in products gain interest. The assessment of perceived quality demonstrates several difficulties. When analyzing the definition of perceived quality, one difficulty becomes inevitable: the entire product's perception is a crucial factor (Zeithaml 1988a).

Current research studies introduced methods for designing products with a high perceived quality level or for assessing perceived quality. Within these methods, visual, haptic, acoustic, and olfactory senses were considered (Stylidis, Wickman, and Söderberg 2018). A common approach for assessing a product regarding perceived quality is considering the sensory organs and using a product decomposition to identify individual attributes. Subsequently, these attributes were assessed, and the results were cumulated for a total perceived quality value. Figure 21 generally presents the decomposition approach.

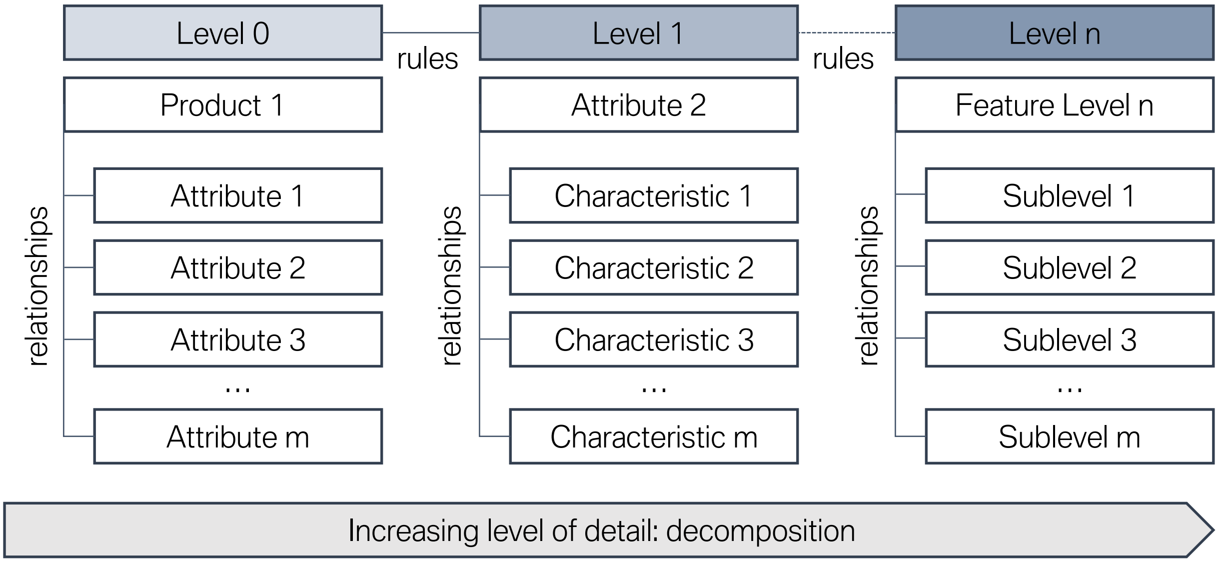


Figure 21. Product decomposition, rules between levels and relationships between attributes.

This assessment approach shows a discrepancy in the definition of perceived quality. Especially methods based in psychology identified this discrepancy. Carbon (2019) developed eleven conceptual moves, which support product design and a higher perceived quality level with psychology aspects. An aspect within the conceptual moves was the complexity of product perception, i.e., the effects of experience or expectations (Carbon 2019). In summary, the conceptual moves provided requirements for assessing a product, which is significant in developing a perceived quality assessment method.

The development of a perceived quality assessment method in this thesis is conducted exemplarily by automobile split-lines. Split-lines are, besides the functional necessity, a design, and a quality feature. Accordingly, the realization of split-lines has an impact on the perceived quality of an automobile. Current methods for assessing split-line quality use a decomposition approach, which cannot include the complexity of the perception process. Therefore, the visual assessment of split-lines regarding perceived quality is used for applying the developed method in this thesis.

The research question is: Which perceived quality assessment method is concerning the definition of perceived quality, which means to include the global product context, relations, and influencing factors?

Based on this question, the following questions were derived:

I. How is the current process of assessing split-lines, and does this process need to be expanded to assess visual requirements?
II. Which influencing factors are present during the split-lines assessment, and how have they to be defined for an improved assessment?
III. What effect do the holistic product context and relations between split-lines and their attributes affect the perception of the visual quality of split-lines?
IV. How can machine learning approaches offer an improved assessment method for split-line perceived quality?

The holistic research method for considering the four research questions is shown in Figure 22.

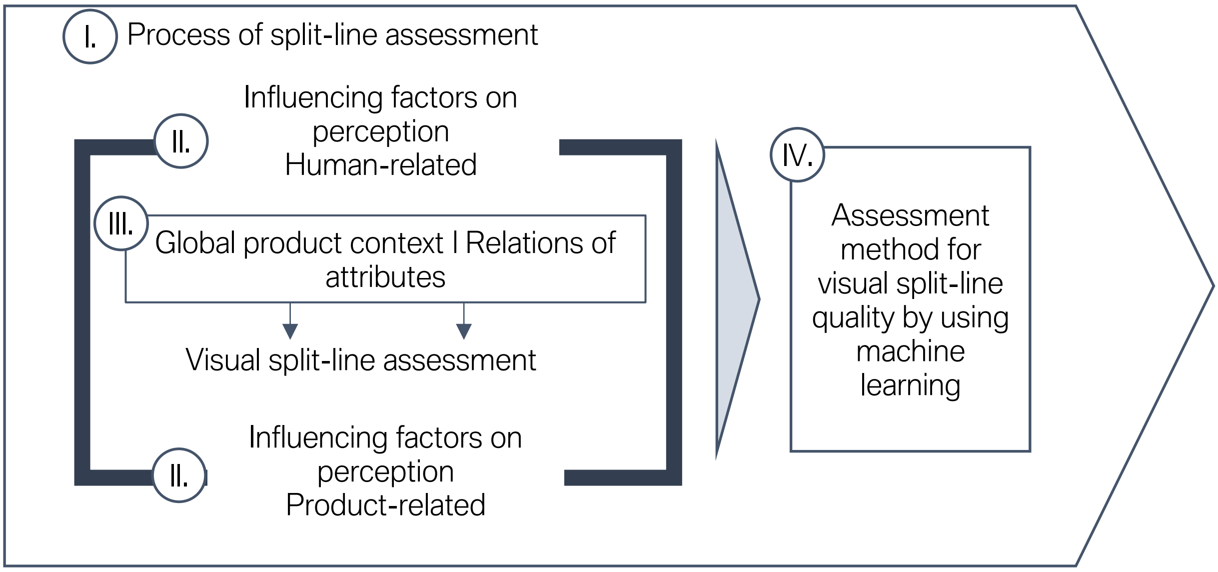


Figure 22. Research approach used to develop the research questions in the thesis.

The first question investigates the current assessment process of split-lines. As presented in the theoretical background (cf. chapter 2.1), split-lines fulfill functional and aesthetic requirements. The literature review demonstrated methods for defining tolerance ranges for single split-line attributes with different levels of complexity. These methods focus on the definition and assessment of functional split-line requirements. Additionally, the presented methods lack the definition and consideration of the assessment process within the industry. Interactions and relations between the assessment of functional and aesthetical split-line requirements are less considered. As a result, there is a risk that functional split-line requirements will be given preference in the evaluation process. The current assessment process of split-lines has to be identified, analyzed, and improved to develop an objective assessment method for aesthetical split-line requirements based on perceived quality. Without identifying the current process, the development of an improved split-line assessment of aesthetical requirements neglects industrial needs. Therefore, the first research question deals with a process analysis with the following objectives:

(1) identify and visualize the current split-line assessment process,
(2) analyze the assessment of functional and aesthetical split-line requirements,
(3) identify methods and tools for functional and aesthetical split-line assessments,
(4) identify weaknesses in the assessment of split-lines with a focus on aesthetical requirements,
(5) deduce an improved assessment split-line assessment process and define methods and tools to assess functional and aesthetical split-line requirements.

After analyzing the current split-line assessment process in the industry and deducing an improved process, the visual assessment of split-lines is investigated. The current and human-based assessment of split-lines shows the following weaknesses:

(1) the parameters, which affect a split-line assessment are unknown,
(2) effects of the product context are disregarded,
(3) split-line attributes are individually assessed, and relations, which appear in reality, are ignored,
(4) the reproduction of an assessment result by the same person can lead to different results,
(5) the assessment result is subjective and depends on the operating person.

The first identified weakness is unknown parameters, which affect the assessment of split-lines. The assessment of split-line perceived quality is performed in virtual and real environments by experts. Two parameter categories are decisive for a visual split-line assessment in reality as well as in virtual environments. Firstly, the assessment of split-lines is affected by the expert. The taken position and perspective during an expert assessment have a direct influence on the visible parts of an automobile. Therefore, relevant assessment positions and perspectives summarized by the term "human-related parameters" are essential within a split-line assessment. Mainly by using virtual environments, the human-related parameters are estimated (Ilhan and Çelik 2017; İlhan, Ünal, and Altınok 2017) in research and industry. An estimation of these parameters contains the risk that visualizations are created with an incorrect camera position and perspective, leading to a different assessment result of the split-line perceived quality. The thesis aims to investigate human-related parameters, which minimize this risk of an unrealistic representation of split-lines. Besides the human-related parameter, the product-related parameter also affects the assessment of split-lines. The varnish of an automobile has a major impact on the perception of split-lines. Different colors and surface structures have a strong influence on human perception (Yoto et al. 2007). The assessment results of split-lines of an automobile with dark paint probably turn out differently than the assessment results of split-lines of an automobile with light paint. The effects of different automobile varnish on human perception of split-lines are also considered in the second research question. Additionally, virtual environments for split-line assessments offer many possibilities to visualize split-line situations of automobiles. To guarantee a virtual environment, which matches a realistic assessment situation, the knowledge of human- and product-related parameters, and how they affect split-line perceived quality are essential and necessary.

The second identified weakness addresses the investigation of the product context effect on the perceived split-line quality. This identified weakness is part of the third research question. Split-lines are spread all over an automobile. Depending on their position, split-lines are better or more difficult to recognize from different views. The different split-line attributes can also be recognized differently in different levels of detail, depending on their position on the automobile and the chosen view. The assessment position and perspective chosen by a human are examined in the previous research question. The effect of a split-line position on an automobile on the perception of split-line attributes is not investigated. For designing an automobile with high visual quality, it is crucial to know and understand which split-line attributes are perceived dominantly in different automobile areas.

The third weakness considers relations and combinations of attributes and their effect on the split-line perceived quality. Currently, split-line attributes are separately considered. This means that the split-line quality is defined by considering gap, flush, and parallelism individually and independently. In reality, these attributes appear mostly in combination and rarely alone. In the current assessment process, the risk results that other split-line situations are considered and evaluated than those that occur in reality. Besides, it is unknown what influence combinations of split-line attributes have on the perception of quality. The investigation of the third weakness is part of the third research question.

The identified weaknesses four and five are addressed in the fourth research question. As presented in the section above, the methods for assessing perceived quality often show a decomposition approach for assessing perceived quality. These methods stand in contrast to the definition of perceived quality. Using a decompositional approach, product relations, for example, relations between split-line attributes, are often neglected. Therefore, a need for a method is identified that evaluates the perceived quality of split-lines on a higher level and includes attribute relations. Based on the various relations of different product attributes, the definition of a method for assessing perceived quality is challenging. Machine learning methods provide a possibility to extract unknown rules and relations in data sets (cf. chapter 2.5). Especially deep learning methods, for example, neural networks, were successfully used within different industries for assessing visual quality in products (Bahlmann, Heidemann, and Ritter 1999; Edwards et al. 1999; Gussen, Ellerich, and Schmitt 2019; Joongho Chang et al. 1997; Kohonen 1982; Packianather and Drake 2000; Pan et al. 2017).

Compared to traditional machine learning methods, neural networks have the advantage of being able to perform the feature extraction of decision-relevant information by themselves. Since the relations are challenging to identify when assessing the split-line quality, it makes sense to use a neural network to classify the quality of split-lines rather than to use a traditional machine learning approach. Additionally, a trained machine learning algorithm reproduces an identical classification result for the same data input. This possibility addresses the fourth identified weakness of a human-based split-line assessment: the lack of reproducibility. Because human perception is characterized by subjectivity, selectivity, and activity (Raab, Unger, and Unger 2016), a risk exists that one person's assessment results are different in two situations. The usage of a trained machine learning algorithm can provide reproducible results. By using a machine learning algorithm to assess the split-line quality, it is reasonable to use a supervised learning approach. Supervised learning approaches require training data with a label. Thereby the algorithm can be taught the quality classes directly. This procedure offers the possibility to objectify the labels and, therefore, the assessment result. The objectification can be achieved by several test persons labeling the data concerning each quality class. As a result, the training data is objectified. Therefore, the fifth weakness can be reduced by using a machine learning approach for assessing the split-line quality. The application of a trained algorithm is originally always a black box. The relevant features extracted by the algorithm itself are not transparent to the operator. Therefore, there is a risk of not being able to evaluate the trained algorithm. In the worst case, the extracted and relevant features are the wrong ones. An essential step is to evaluate a trained algorithm. Research methods were derived from visualizing relevant features of a trained neural network (cf. Chapter 2.5.9). Therefore, the trained algorithm for classifying the visual split-line quality is evaluated by explainable AI methods.

4. PROCESS OF SPLIT-LINE ASSESSMENTS

Currently, technical requirements are primarily the basis for the visual assessment of split-lines. For example Stoll and Paetzold (2008), Wittmann (2011), and Ungemach et al. (2008) introduced virtual methods for a visual assessment of tolerances in the industry. The basis for the virtual assessment methods was deformed CAD geometries (Stoll and Paetzold 2008), simulated assembly and geometry tolerance derivations (Wittmann 2011), or tolerance derivations based on FEM simulation results (Ungemach, Mantwill, and Rund 2008). These methods are aimed at optimizing the tolerance ranges for components or products. A common approach of these methods was ray tracings or renderings for an additional visual assessment. The subsequent assessment of the tolerances and the product appearance was not supported within these methods and was, therefore, expert-based and highly subjective.

The first research question considered the process of split-line assessment in an industry with a focus on technical and visual requirements. The research question addresses two aspects, the investigation of the current split-line assessment process and identifying a process improvement. These two aspects are considered in this chapter. For the investigation, both aspects are transferred into questions, which are answered within this chapter:

(1) How is the current process of assessing split-lines described in the industry?
(2) How is the interaction between the assessment of technical and aesthetical split-line requirements in the current process?
(3) Which methods and tools should be used to support an assessment of technical and aesthetical split-line requirements?
(4) Is there an improvement potential regarding the aesthetical assessment of split-lines?
(5) How has a balanced process for assessing technical and aesthetical split-line requirements to be designed?

The research method, the results of the analysis of the split-line assessment process, and the proof of concept were published by the author (S. Striegel et al. 2017).

4.1. METHOD

For investigating the first research question, a new method was derived following process improvement methods. Based on their approach, the method for analyzing the current process of visual split-line assessment was derived. The resulting method is presented in Figure 23.

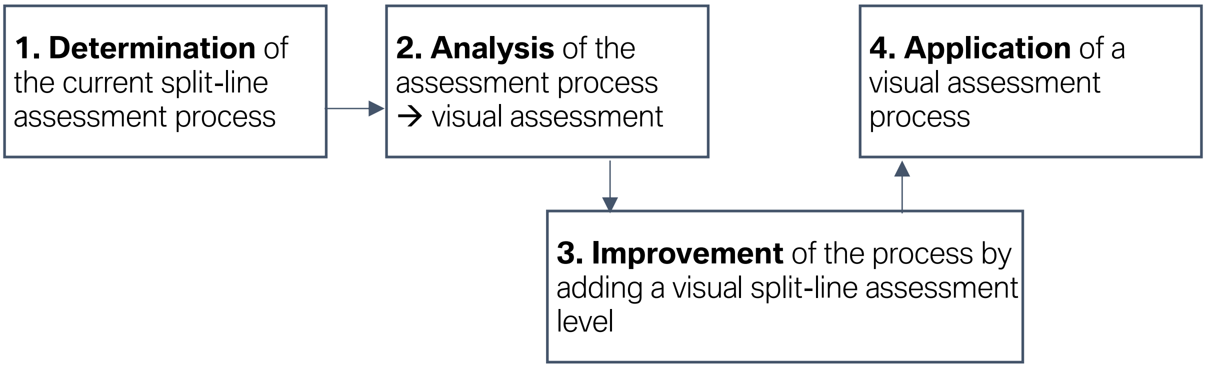


Figure 23. Method for analysing and improving the current visual split-line assessment process.

Within the first step of the process analysis, the current process of assessing the split-lines is derived. Based on the results of the process identification, the next step is to analyze the process. The focus of the process analysis lies in the visual split-line assessment process. Missing assessment levels are deduced and within the next step of process improvement, these process weaknesses are improved. Additionally, the improvement step includes an identification of necessary assessment tools and methods. Within the last step of the process analysis, the new and improved process is applied and tested within an industrial application.

4.2. RESULTS

Perceived quality offers companies to differentiate from competitors. Therefore, the design and the quality of products gain in importance (Carbon and Jakesch 2013). For realizing an assessment method, the current process of assessing the split-line quality needed to be deduced, analyzed, and optimized. Therefore, the scope and the importance of this process analysis is based on the arguments presented in the introduction. Similarly, the chances of an assessment process considering the visual quality of split-lines lie in the improvement of perceived quality, which can result in a distinguishing feature and was presented within the introduction, as well.

4.2.1. DETERMINATION OF THE CURRENT ASSESSMENT PROCESS

The process of defining and assessing split-lines is based in an early stage of the product development process. This process aims to calculate the split-line quality by using tolerance management methods and tools. Priority is given to the construction and verification of the technical requirements, such as tightness. The process of reviewing the technical requirements is shown in Figure 24.

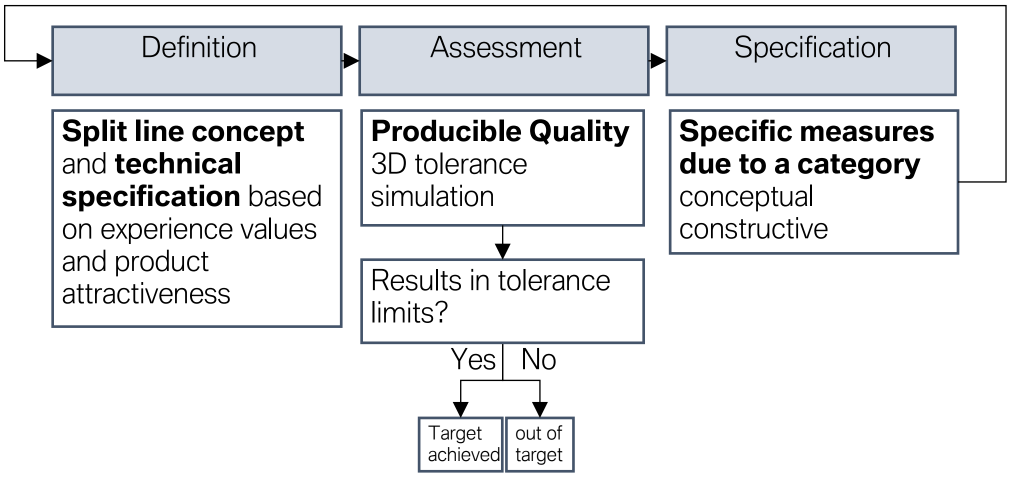


Figure 24. Identified split-line assessment process in industry.

Within an initial process step, the definition of split-lines is based on internal expertise and comparison to competitors. The assessment of these initial split-line definitions is verified based on tolerance management methods, i.e., Monte-Carlo simulations. As a result of these methods, upper and lower limiting positions are derived, resulting from form and assembly tolerances of components within the tolerance chain (Ehlert et al. 2016). The limiting positions were derived for each split-line attribute, gap, flush, and parallelism separately (cf. section 2.1.1). Subsequently, the resulting tolerance limits were categorized into critical and non-critical (cf. Figure 24). Non-critical results are not further considered. Thus, the technical requirements are fulfilled with these split-line results. Conceptual or constructive specifications modify the critical simulation results. In conclusion to the application of specifications and the adjustment of the split-line concept, a new tolerance simulation iteration follows. This process is repeated until the split-line results are in target and fulfill the technical requirements. During this simulation, the technical design is examined to ensure the production of automobiles in reality. Therefore, the consideration of these aspects is summarized by the term of Producible Quality.

4.2.2. ANALYSIS OF THE DERIVED PROCESS

Producible Quality is defined by a functional investigation of split-lines, i.e., the tightness. The fulfillment of these requirements is measured by using tolerance management tools. The problem definition of Producible Quality is to derive numerical values for upper and lower limiting positions (Ehlert et al. 2016) of split-line attributes and has an objective assessment character. A weakness of the current process is that visual verifications of the split-lines remain unnoticed based on numerical simulations and their results. In the current process, a visual assessment is performed at a later stage using hardware prototypes. Accordingly, changes in the product are within a late stage of the product development process and are costly and time-consuming. The visual requirements need to be assessed at an early stage of the development process to improve the perceived quality and be integrated into the current process. Therefore, a new assessment level, which was called Refinement, was deduced. For analyzing the relation between both levels, a matrix, which is presented in Figure 25, was derived.

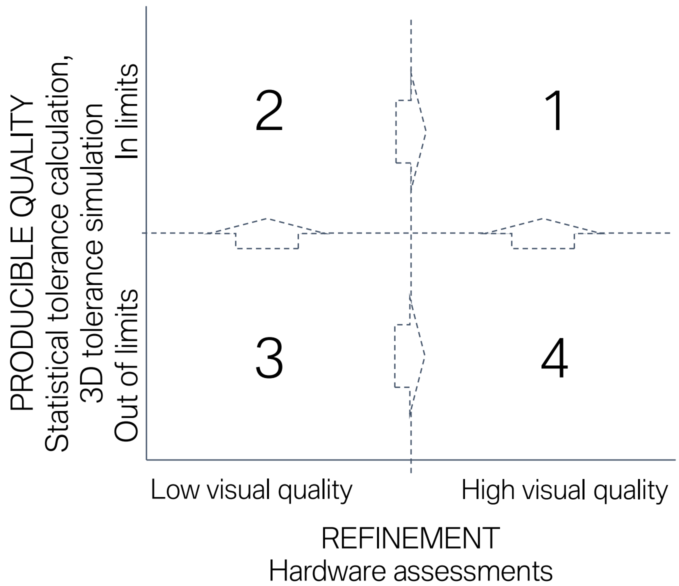


Figure 25. Derived matrix for categorisation of Producible Quality and Refinement.

The first category was achieved by having the level Producible Quality in limits and the Refinement level with high visual quality. This category represented a high level of perceived split-line quality. Tolerance calculations or simulations resulted in limits, which correspond to technical requirements. Experts assessed the Refinement level represented by visualizations with a high level of quality. Therefore, design and tolerance processes should aim to reach this category.

The second category differed from the first category in the Refinement level. Producible Quality assessments resulted within limits, and technical requirements were fulfilled. In contrast, the Refinement level was evaluated by experts with low visual quality. Therefore, the Refinement level had to be optimized to reach higher visual quality.

The third category represented producible quality out of limits and Refinement with a low visual quality. Split-line assessments resulting in this category were optimized regarding both categories. Technical, as well as visual requirements were not fulfilled. Strategies had to be deduced, which include the improvement of both levels.

Producible Quality results defined the fourth category out of limits and Refinement with high visual quality results. Technical requirements were not realized, and the concept had to be enhanced. Thus, the visual quality was assessed with high quality, and a strategy was performed to optimize tolerance limits in a tolerance chain without changing the Refinement level results.

This matrix offered an opportunity to structure split-line assessment results concerning the levels of Producible Quality and Refinement. This analysis concluded that without an early assessment of the Refinement level, visual quality issues were identified at a late stage of the product development process. For a process improvement, this matrix needed to be detailed to assess the visual split-line quality and improve the level of perceived quality.

The visual assessment of split-lines, within the Refinement level, has a design and a quality aspect. Both aspects contribute to the level of perceived quality. To assess a product concerning the level of perceived quality, both aspects need to be considered. With regard to the perception aspects of the assessment process, Navon (1981, 2003) demonstrated a global-to-local effect, which was called global precedence. Based on this research, the assessment of split-lines was divided into a local and a global assessment level. A product-perception matrix was derived, considering both product levels and both perception levels. This matrix is shown in Figure 26.

	Quality	**Design**
Global	Focus on split-line design and quality on a global level.	Visual product appearance, includes entire design attributes
Local	Quality assessment of individual split-lines	Individual design attributes in a local investigation scope

Figure 26. Deduced product and perception matrix for analysing quality and design issues.

The focus was on the perception of several split-lines and their visual interaction within the global quality level. The harmony of their split-line arrangement and the split-line run is considered. The global design level considers the harmony of the product. Therefore, the interaction between several attributes, including the split-lines, is focused. The local quality level investigates the visual result of the technical execution of split-lines. Local Design includes the assessment of several design elements.

4.2.3. Improvement of the derived Process

Based on the analysis level results, the Refinement level had to be integrated into the current process of assessing split-line quality. For the integration of this level, dependencies between Producible Quality and Refinement were pointed out.

The Refinement level considered the visual appearance of a product related to the product-perception matrix. Therefore, Refinement aims to define a high-quality split-line design. The Refinement level is mainly assessed by using hardware prototypes. An additional process improvement is the use of ray tracings (Kajiya 1986) in the Refinement level and an early stage of the product development process. Thus, hardware prototypes are cost-intensive and have a temporal offset for the visual quality assessment. A summary of both levels and the dependencies between both levels are presented in Figure 27.

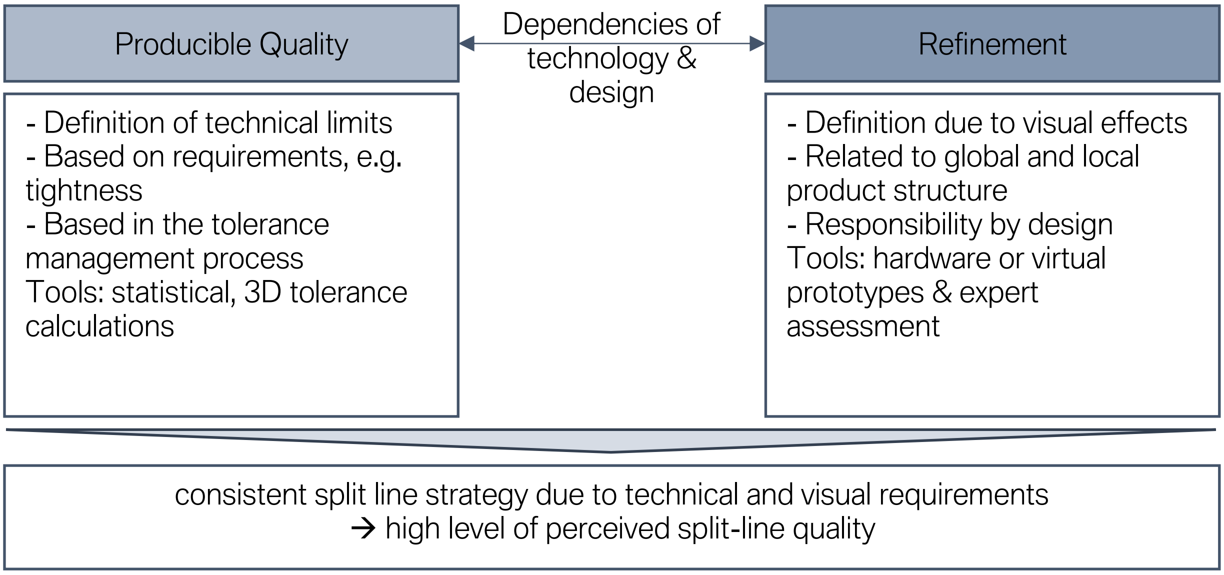


Figure 27. Identification of dependencies between the levels Producible Quality and Refinement.

As a result of this process step, the responsibilities of both levels need to be defined. Therefore, the process responsibility of the Producible Quality level lies in the tolerance management department of a company; thus, this level is related to traditional tolerance management tasks (Bohn and Hetsch 2013b). Compared to Producible Quality, the evaluation of Refinement is intensely subjective and is, therefore, provided by designers.

The process improvement included a synchronization between Refinement and Producible Quality level and to integrate the classification matrix. Therefore, the current assessment process of split-lines was expanded by adding a fourth stage to the initial three-stage process. The four-stage process shows this additional assessment level (see Figure 28).

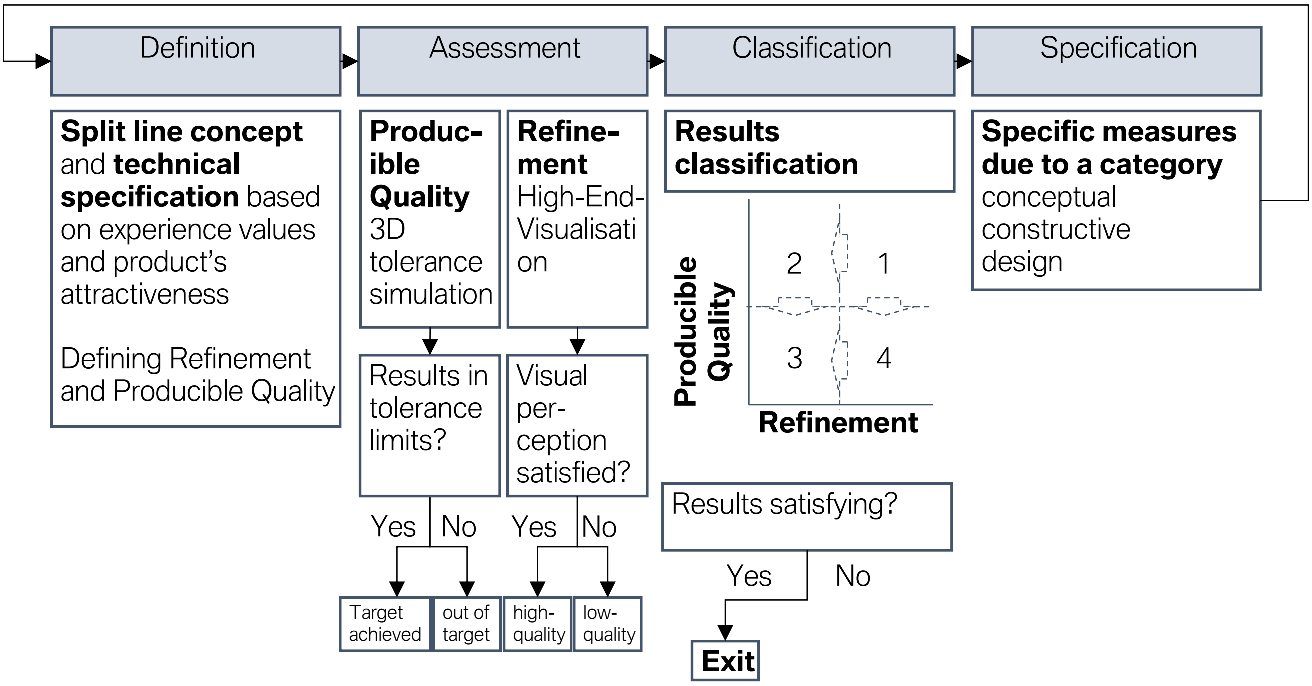


Figure 28. Evolution of the split-line assessment process by implementing the derived Refinement level.

The improved process started with a similar process step as the former process: defining an initial split-line concept derived by experience, previous construction limits, and competitor designs. An addition within this level was assessing this initial concept by experts of the Producible Quality and Refinement level. Within Producible Quality, familiar tolerance management tools were used to assess the concept. Within the Refinement level, the split-line concept was visualized by using raytracing or rendering methods. Subsequently, the visualizations were assessed due to the assessment purpose. This purpose was further specified by using the product-perception matrix.

Consequently, the results from both levels were categorized using the developed Producible Quality-Refinement matrix. The resulting category was decisive for further considerations and product improvements regarding perceived quality—results within categories 2 to 4 needed to be improved. Based on the category, strategies were deduced to improve either the Producible Quality, the Refinement, or both levels. With the adjusted split-line concept, the process started again. This process is repeated until results were classified in category 1 of the Producible Quality-refinement matrix. This category is achieved with a high split-line quality, which fulfills technical and visual requirements.

4.2.4. Industrial Application

The improved process was integrated into the automobile industry at an early stage of product development. The consideration included the split-line between the chassis and the tailgate based on the Producible Quality and Refinement levels. By using the derived Producible Quality-Refinement matrix, the split-line was classified in category 2. The Producible Quality level is characterized in limit but a Refinement level with low visual quality. Therefore, a strategy was derived from increasing the Refinement level. The derived strategy was the additional painting of components within an inner level of the automobile body. The increase in product quality is shown in Figure 29.

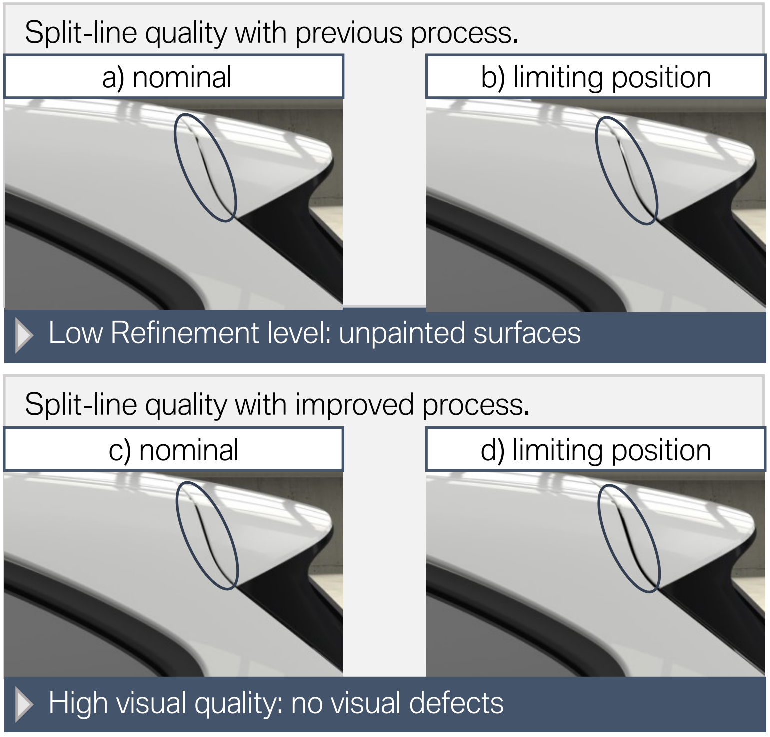


Figure 29. Resulting split-line quality by using the traditional and the improved split-line assessment process.

Figure 29 shows the trunk lid of an BMW 2 series (SOP 2014) and the tolerances within a nominal and upper limiting position. The images a) and b) in Figure 29 represent the visual quality of the split-line with the previous process, which solely had the Producible Quality level. The visualized quality defect are visible noticeable unpainted surfaces at the trunk lid. Based on the statistical methods and without considering the Refinement level, the unpainted surfaces were not identified. By performing the improved process, the unpainted surfaces were detected within the Refinement level. The local quality level was improved by conducting a painting measurement. The improved quality is visualized in Figure 29 c) and d).

4.3. SUMMARY AND CONTRIBUTION OF RESEARCH QUESTION 1

The first research question dealt with analyzing the current split-line quality assessments process and the deduction of a process improvement. Therefore, an approach for analyzing and improving the visual split-line assessment process was derived based on process improvement methods. In an initial step, the current assessment process of split-lines in the automotive industry was resumed. This analysis is also the answer to the first asked question at the beginning of the chapter. Based on this current process, the interaction between the technical and visual assessment process of split-lines was conducted (the second question of research question one). Therefore, the levels of Producible Quality and Refinement were identified, which were necessary for a high level of perceived quality regarding split-lines. The levels were analyzed regarding their applied tools (the third question of the research question). The producible quality level demonstrated a solid base of methods, which were primary tolerance management tools.

On the contrary, the Refinement level demonstrated improvement potential, which was defined by the fourth question of the first research question. The product-perception matrix was developed, including Global Quality, Global Design, Local Quality, and Local Design, to support the Refinement level. This level, including the product-perception matrix, were implemented within the current assessment process of split-lines. The interaction of both levels, the Producible Quality, and the Refinement level were included in a process improvement, and the fifth question of the first research question was answered. For evaluation of the new process, an industrial application was demonstrated. This application improved the level of perceived quality by using photorealistic visualizations in an early stage of the product development process. This analysis demonstrated a frontloading and a higher product maturity in an early product development phase.

The conducted process analysis of the first research question supports state-of-the-art methods within visual assessments of optimized results. The assessment of the product visualization of these methods can be supported by using the developed product-perception matrix within the Refinement level. By using this matrix, the presented methods are extended. Therefore, the improved process is the next step of these methods and no substitutional method.

Carbon (2019) deduced eleven conceptual moves, which support the design of products by psychological aspects. These aspects can be realized and integrated within the improved process. These conceptual moves can be monitored within the Refinement level. This level demonstrated an initial approach to integrate psychological aspects within the visual assessment of split-lines. The product-perception matrix was deduced to include psychological aspects. This matrix demonstrated a methodical tool for categorizing visual assessment tasks for split-lines and can also be used to assess products or attributes regarding design and quality.

The integration of a Refinement level within the split-lines assessment process demonstrated a processual implementation for capturing the level of perceived quality. Various methods demonstrated an assessment method for the level of perceived quality of a product. Besides these assessment methods, the improved process demonstrates integrating these methods within the product development process. Besides evaluating technical product requirements, a level was developed, focusing on evaluating perceived quality requirements.

5. HUMAN-RELATED PARAMETERS

A method to define human-related parameters was introduced by Ilhan and Çelik (2017); İlhan, Ünal, and Altınok (2017). The method presented activities, which affect perceived quality, from the concept process to the final product. Within these activities, assessment positions were estimated in a workshop. The concluding assessment position was a distance of 1.5 m to 0.5 m and a height of 1.8 m. A limitation of the assessment position introduced by Ilhan and Çelik (2017); İlhan, Ünal, and Altınok (2017) is that the trunk lid was focused and the investigation was on the local joint quality without differentiating the evaluation level regarding a global and local investigation. The scope of the presented tracking study was the entire automobile, and therefore, the assessment positions and perspectives were derived independently of a specific component.

The second research question demanded factors that influence the perception and the assessments of split-lines regarding perceived quality. There are three parameters identified in a perceived quality product assessment: human, product, and environment. Figure 30 presents the context of the three parameters categories human, product, and environment.

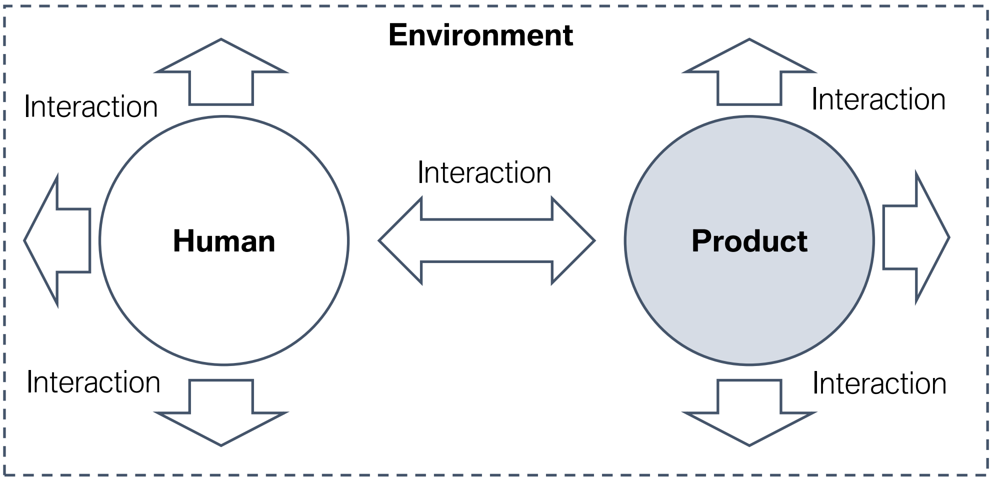


Figure 30. Illustration of the context of product, human and environment during an assessment.

The subject of this thesis and the second research question are the human- and product-related parameters. The consideration of both parameter types is divided into two chapters. Based on the research question, the following questions, which will be answered in the following two chapters, were derived:

(1) Which methods are suitable for deriving human-related parameters?
(2) Which are the variables for human-related parameters?
(3) How are these parameters for a visual split-line assessment used within the industry?

The human-related parameters determined the position and the perspective during an assessment. A person intrinsically chose both parameters. A definition of human-related parameters regarding the visual split-line quality assessment, implemented in the thesis, is:

Human-related parameters are those, which an individual chooses intrinsically, intuitively, and independently in a visual assessment of products.

The research method, the approach and the results of the human-related parameters were published by the author (Striegel and Zielinski 2018)

5.1. Method

Experts and laypersons intuitively choose the human-related parameters during a split-line assessment. Previous knowledge about these parameters is rare. Instead, there is the question of how human-related parameters are chosen within an assessment. Therefore, the chosen research method was a qualitative research method. The approach was an observation of individuals during their split-line assessment. Each participant was tracked during their assessment. The general method for identifying and analyzing human-related parameters within a visual split-line assessment is presented in Figure 31.

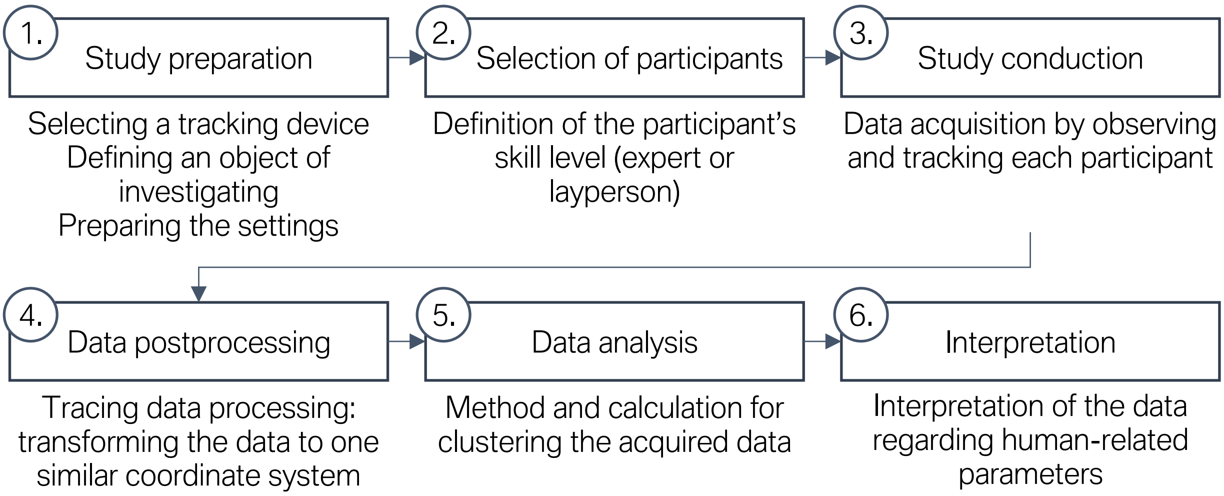


Figure 31. Derived method for analysing human-related parameters in context of the visual assessment of split-line quality.

The first step of the method is the study preparation. Within this step, organizational tasks are arranged. These tasks are selecting a tracking device based on the data that wants to be acquired, defining the object of investigation, and preparing the setting, i.e., the environment and supportive tools. This first step forms the basis for further consideration. The second step considers the selection of participants. Within the focus on the visual split-line assessment, the participants can be experts or laypersons in the field of split-line assessment. The third step is the study conduction, and the data is acquired by tracking the participants and having an observer for further information. The fourth step considers the processing of the tracked data. Based on the selected tracing data, the personal data and the investigated object's data have to be synchronized. The subsequent step is the data analysis step. Within this step, the acquired and processed data is analyzed regarding the human-related parameters. This fifth step is the basis for the concluding interpretation of the data.

5.2. APPROACH

Within an assessment of products, the identified human-related parameters position and perspective are essential for the perception. Experts in the industrial quality assessment were trained to use the parameters intuitively. Therefore, assessments were, to some extent, standardized, but a direct request would be inefficient since there were no explicit specifications of these parameters. The idea of the study was to acquire the data indirectly by tracking experts during their visual assessment of split-lines. In doing so, the results were a reflection of the assessment procedure of hardware prototypes. This approach has two advantages. Firstly, the assessment process is recorded and therefore archived and digitalized. Secondly, the acquired data is the basis for an analysis of the assessment process to derive perception influencing parameters.

5.2.1. STUDY DESIGN

The investigation object of the study was a BMW 5 series sedan (SOP 2017). The study settings were as follows: A participant assessed an automobile regarding the visual split-line quality. The study settings are presented in Figure 32.

Task:	**Tolerances**:	**Participants**:	**Acquisition**:
Split-lines design & quality assessment → product-perception matrix	gap, flush, parallelism, symmetry	14 experts in assessing split-lines	tracking function of MS Hololens

Figure 32. Study design for investigating human-related parameters assessing the split-line quality of an automobile.

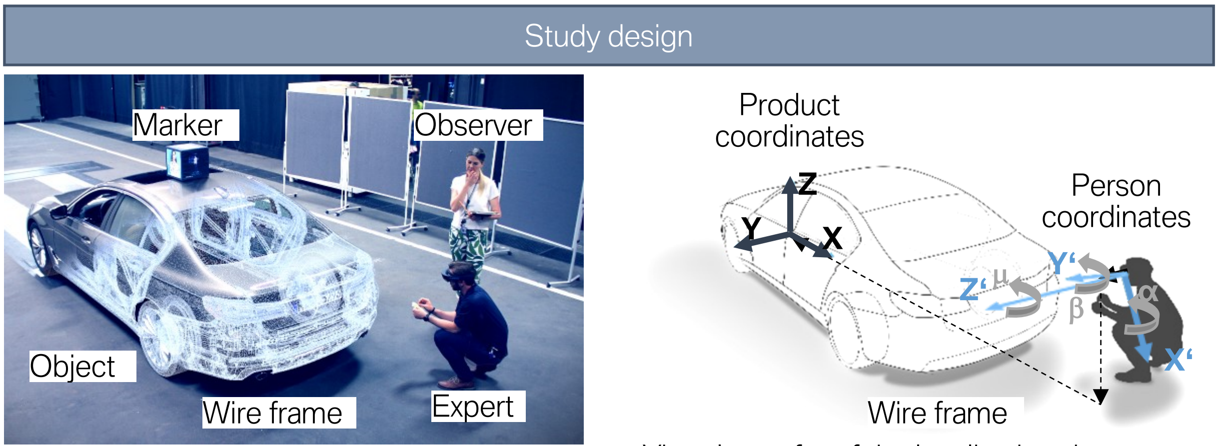
Figure 32 visualizes the study settings. A participant's position (P1) and the viewing perspective (V1) were tracked. The participants' assessment task was to evaluate the visual split-line quality regarding the global and local level of the developed product-perception matrix (cf. Figure 26). Therefore, the split-line attributes gap, flush, parallelism, and symmetry were assessed. The attributes were identified related to their industrial assessment relevance. In addition to the previous attributes, the attribute symmetry was appended. This attribute considers an equal technical execution of mirrored split-lines, for example, at the left and right front light. When considering parts of an automobile, this attribute can often not be considered.

5.2.2. PARTICIPANTS

14 participants (female=1) joined the tracking-based study. These were experts in the field of visual split-line assessments. The age of the participants was between 28 and 55 years. Their level of experience was between 1 and 15 years.

5.2.3. Apparatus & Stimulus

As a tracking device, the Augmented Reality (AR) device from Microsoft (MS), the HoloLens (Microsoft 2020), was utilized. The MS HoloLens uses inside-out tracking for self-localization within the environment by default (Microsoft 2020). This inside-out tracking function was modified and used for the study related tracking. A script-based request was set six times per second. Within this request, the current position and the corresponding perspective were located. As the tracking was placed in the three-dimensional environment, the position was tracked by 3D coordinates. These coordinates were analyzed by transforming them in the distance and height related to the object under investigation. The tracking of the perspective was subjected to the MS HoloLens function. Indeed, the MS HoloLens used to track the normal vector in viewing direction. For the study, the assumption of the participants' straight viewing direction was made, so the normal vector tracking was a simplification of eye-tracking—the output of the perspective tracking as a vector. For using the MS HoloLens, two previous requirements had to be implemented. Firstly, a virtual model (wireframe) of the 5 series was necessary to calculate the parameters concerning the object. Secondly, to overlap exactly the virtual model with the hardware model, a marker was necessary. The marker had the additional function to ensure continuous and consistent tracking during the study. Figure 33 presents the hardware and virtual study design.



Wire frame → necessary for localization
Tracking information: coordinates (X, Y, Z), vector (α, β, μ)
→ Translation and rotation

Virtual transfer of the localization data. Implementation of product-based and person-based coordinate systems.

Figure 33. Concept of the tracking approach within the study and showing the approach for tracking a person related to the automobile using the MS HoloLens as tracking device

Within Figure 33 the same situation is visualized. In the left image an actual assessing situation is presented. The right image shows the same assessment situation transferred to the virtual scene. The virtual scene visualizes a central element of the study: the coordination transformation. During the assessment, an observer attended the study for documentation of special conspicuous features. The study setting was spacious, with 450 m^2 to ensure full freedom of movement and simulate a typical discovery procedure. Additionally, within the study settings, the area lights were unique lights to simulate daylight. Initially, each participant was shortly briefed outside the study area and prepared for the set-up of the tracking device. The average assessment time per participant was 25 min.

5.2.4. DATA PROCESSING

Since the MS HoloLens and the investigated automobile had different coordinate systems, data processing was necessary. The acquired data was proceeded by using coordinate transformations. Therefore, translations were performed by using a value t related to the x-, y- or z-dimensions, and rotations were conducted by the rotation angles per dimension (α, β, μ) (Arens et al. 2013; Hackbusch, Schwarz, and Zeidler 2003; Wörle, Rumpf, and Erven 1994). The applied three-dimensional translation matrices are represented in the Appendix A.

The rotation was three-dimensional and implemented around each axis, and the translation was performed by a linear adjustment of a three-dimensional coordinate point. In doing so, the acquired person-based data was transformed into the product-based coordinate system. Using this procedure, a relation between the tracked participant and the product results. This transformed data was the input value for the data analysis.

5.2.5. DATA ANALYSIS

The data analysis aimed to support the identification of three different values: frequently assessed areas, position, and perspective. For analyzing the data, a heatmap of the acquired and transformed data was derived. The steps are for the data analysis, and the heatmap deduction are presented in Figure 34.

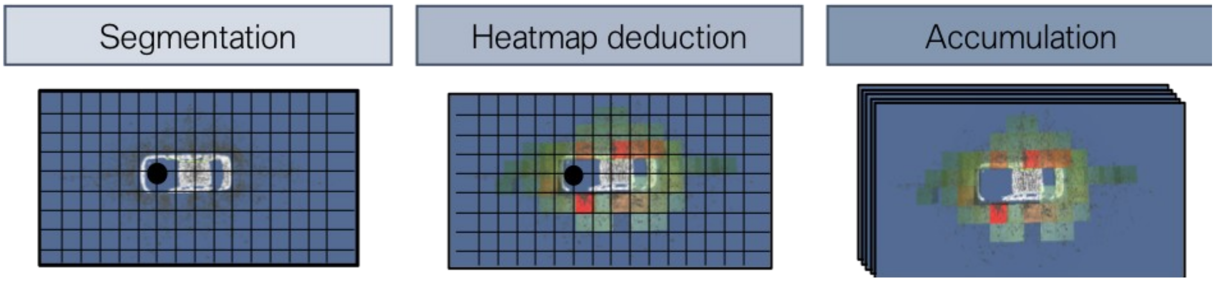


Figure 34. Approach for analysing the acquired data and for deducing human-related assessment parameters.

The input was the transferred data, which consisted of three-dimensional space coordinates of one participant. Within the first step, the space was segmented in similar cubes with the extent of 0.8×0.8×0.8 m and a volume of 0.51 m^3. This segmentation was the basis for further consideration. The heatmap deduction was performed by calculation of the tracked positions within each cube. The assessment intense was calculated using the tracking quantity. For each participant, a single heat map was deducted by using the following procedure:

(1) Counting the number n of tracking coordinates t per cube c with 0.51 m^3:

$$n_c = \sum_{k=0}^{t} t_k \tag{8}$$

(2) Identifying for all cubes c the maximum and minimum number n tracking coordinates t:

$$n_{max} = max\{n_1, \dots, n_c\} \tag{9}$$

(3) Calculating the color scale s per cube c:

$$s_c = \frac{n_c}{n_{max}} \tag{10}$$

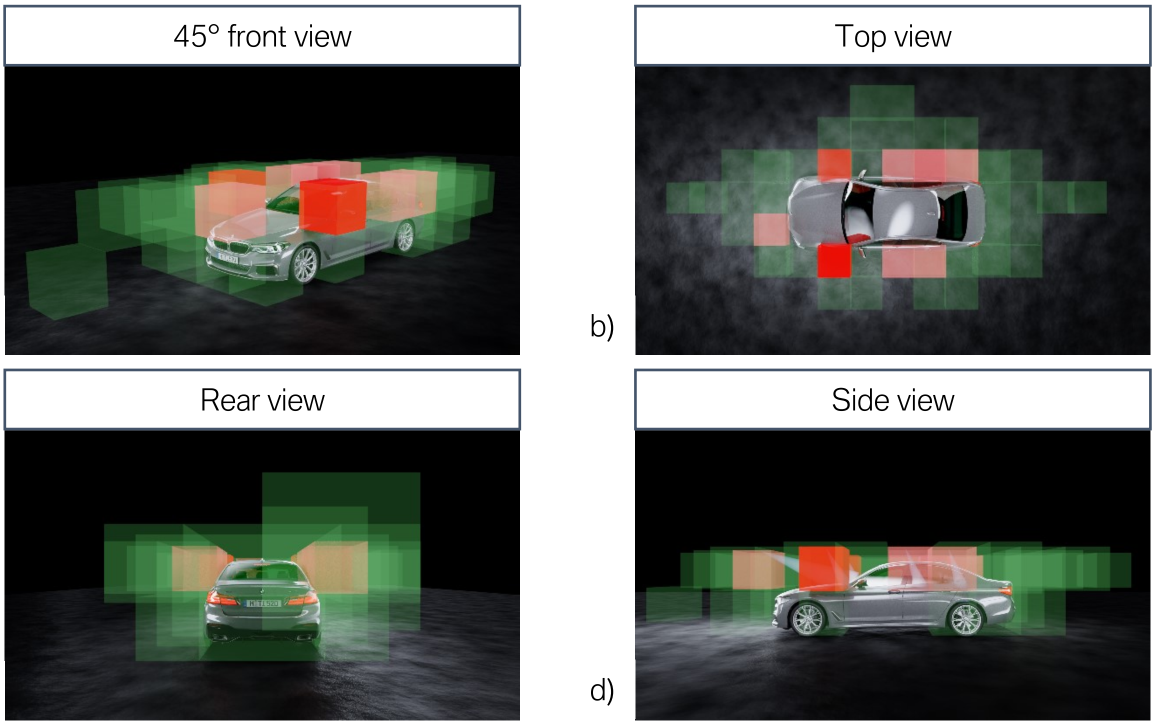
(4) Color scale:
Red: $s_c = 1$
Green: $0 < s_c < 1$
Transparent: $s_c = 0$

For a derivation of a total heatmap of all participants, a similar procedure was performed. Initial to the procedure, the three-dimensional space coordinates of one participant were accumulated to one data input. Subsequently, the segmentation and heatmap deduction was applied.

5.3. RESULTS

The results are presented using the deduced overall heatmap regarding the three values frequently assessed areas, position and perspective. The deduced heatmap in four different perspectives is presented in Figure 35. Each cube has the extent of 0.8×0.8×0.8 m.

The cubes were derived by using the calculation method presented in chapter 5.2.5. For presenting the results of the tracking study, these were categorized within frequently assessed areas, assessment position and perspective.



a) b) c) d)

Figure 35. Resulting heatmaps of the human-related parameters during a split-line assessment. a) visualizing the assessment cubes from a 45° front view, b) visualizing the assessment cubes from a top view, c) visualizing the assessment cubes from a rear view, c) visualizing the assessment cubes from a side view.

5.3.1. FREQUENTLY ASSESSED AREAS

Within the visualized heatmaps, the frequently assessed areas were identified by the different colored cubes. According to the heatmaps, the cubes were mainly symmetrical to the x-axis (cf. Figure 35b)). The evaluation intensity of the attributes on the driver's and passenger's side was accordingly the same. Red-marked cubes were areas with a high or long assessment occurrences of the participants. These cubes were identified at the a-pillar on the driver and the passenger side, at the doors on both sides, the fuel flap, and the front flap.

Furthermore, the red cubes were close to the automobile, which indicates a local quality assessment. Green highlighted cubes classified assessment positions with low frequency. These areas were typically at a greater distance to the assessed object and indicated a global product quality assessment. Green cubes were primary at the front-end and the rear-end of the automobile. In summary, the frequently assessed areas were identified and assigned to the study's global and local assessment tasks.

5.3.2. Position

The assessment positions were defined by the distance of a person related to the product and the height. Within the tracking study, the distance and height were characterized by three-dimensional coordinates in the person-based system. The assessment distance and an assessment height concerning the product were identified by transferring the tracked coordinates into the product-based coordinate system. The distance was defined in x- or y-dimension within the product-based coordinate system. In contrast, the height was described in the z-dimension of the product coordinate system. Consequently, the tracked three-dimensional coordinate of participants was first transferred into the product-based coordinate system. Secondly, the x- and y-dimension parts were translated into the assessment distance. Thirdly, the z-dimension parts of the coordinates were defining the assessment height. In doing so, the frequently received assessment distances were up to 0.8 m (cf. Figure 35b)).

Further assessment distances were up to 3.2 m (cf. Figure 35b)). During the tracking study, the assessment height was mainly chosen at the eye level of 1.6 m (cf. Figure 35c)-d)). The derived cubes have their center at a height of 1.6 m. Therefore, the upper cube starts at the height of 1.2 m and ends at 2.0 m. The lower cube starts at a height of 1.2 m and ends at the height of 0.4 m. The chosen height was constant and independent of the assessment distance. An exception was during the assessment of the front end. Within this assessment distance, a lower assessment height was chosen, which was up to 0.8 m.

5.3.3. Perspective

As a third category, the perspective during an assessment was considered. The parameter perspective was defined by the normal vector of the MS HoloLens. Similar to the deduction of the position, the derivation of the parameter perspective required a coordinate transformation from the person-based system into the product-based system. In doing so, a data correction was required. The participants interacted with their surroundings, e.g., with the observer, and therefore, the perspectives which did not focus on the automobile were deleted. Due to the simplification of eye-tracking, a precise definition of the field of vision is not possible. The perspective complied with the head and body rotation related to one area.

5.3.4. Standardized Human-related Parameters

The conducted study supported the derivation of human-related assessment parameters identified by frequently assessed areas, assessment positions, and perspectives. Based on the three-dimensional assessment environment, the parameter position was characterized by three-dimensional coordinates. The parameter perspective was defined by a position-corresponding vector in the three-dimensional environment. The assessment procedure of experts was tracked and therefore virtualized and documented. This documentation was an accompanying effect, as assessments were, in the meantime, experience-based and intuitive. The main advantage provides the deduction of standardized split-line assessment parameters. Therefore, the results were summarized, and a parameter recommendation was deduced per assessment level. The results were maintained to be categorized by the global and local assessment levels. Global assessments include the product context and, therefore, further product attributes. In order to perform a global product assessment, an assessment distance up to 3.2. m, and an assessment height of 1.6 m were deducted. The perspective in a global quality assessment was accompanied by head and body rotations to perceive the whole product context. Within local quality assessments, the focus lay on individual product attributes. Therefore, the assessment distance was closer to the automobile and up to 0.8 m. Similar to a global assessment, the height was constant at 1.6 m. The deduced parameter is presented in Table 8.

Table 8. Summary of the human-related parameters with respect to the assessment level.

Frequently assessed areas	Assessment level	Position-distance	Position-height	Perspective
A-pillar	Local Quality	0.8 m	1.6 m	Focused: details
Fuel flap	Local Quality	0.8 m	1.6 m	Focused: details
Doors	Local Quality	0.8 m	1.6 m	Focused: details
	Global Design	1.6 - 2.4 m	1.6 m	Head & body rotations: holistic
Front flap	Local Quality	0.8 m	1.6 m	Focused: details
Front end	Global Design	3.2 m	0.8 – 1.6 m	Head & body rotations: holistic
Rear end	Global Design	3.2 m	1.6 m	Head & body rotations: holistic

The application was exemplarily demonstrated at the most frequently assessed area to apply the derived standardized human-related parameters (third question research question 2). As the most frequently assessed area, a-pillar was identified. Within this area, the assessment scope were Local Quality attributes of split-lines. This area was assessed by using two assessment distances and heights. The first assessment distance was 0.8 m and a height of 1.6 m. This perspective demonstrated a front view of the a-pillar, shown in Figure 36 on the left side. The second assessment distance was 0.8 m and a height of 1.6 m at a more frontal view of the driver side. The results of this second view are presented in Figure 36 on the right side.

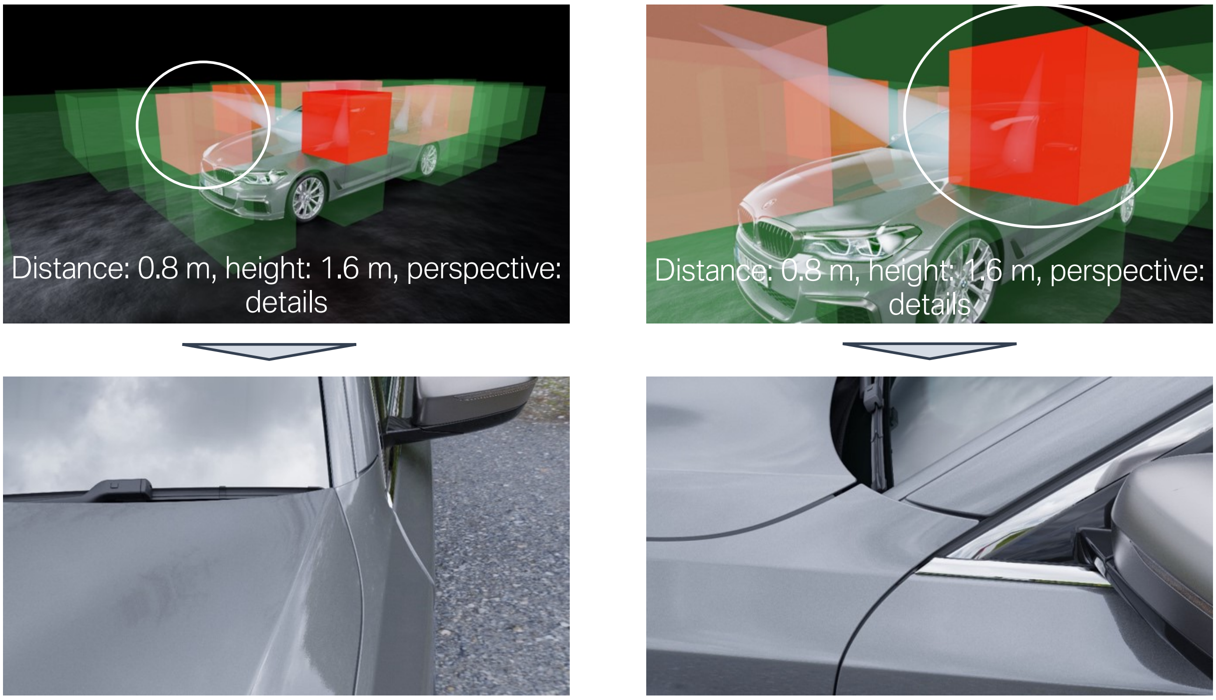


Figure 36. Application of the derived human-related parameters within an industrial application.

6. PRODUCT-RELATED PARAMETERS

In addition to the human-related parameters, product-related parameters were identified as influencing parameters in assessing perceived quality, which represents the third research question (cf. Figure 30). As product-related parameters, color and reflection were identified. Regarding the product-related parameter investigation, the following questions were derived by the second research question and extend the research questions of the human-related parameters:

(1) Which method is suitable for the investigation of product-related parameters?
(2) Which are the variables for product-related parameters?
(3) Which values can be derived for Global Quality, Local Quality, Global Design, and Local Design assessment tasks related to human and product parameters?

Similar to the definition of the human-related parameters, a definition of product-related parameters was deduced. The definition of product-related parameters influenced the perception of individual product attributes or features. For example, reflections were perceived differently with light automobile paint than with dark paint. This category of parameters results from the following definition:

Product-related parameters are those, which are extrinsically assigned to the object before the assessment.

6.1. General Research Method

Compared to the human-related parameters, which were intuitively captured during an assessment, the product-related parameters were defined before an assessment. The investigation is a mixed-method using technical analysis and a quantitative research method. The investigation method for the empirical study was based on an experimental design. The entire approach of the investigation is presented in Figure 37.

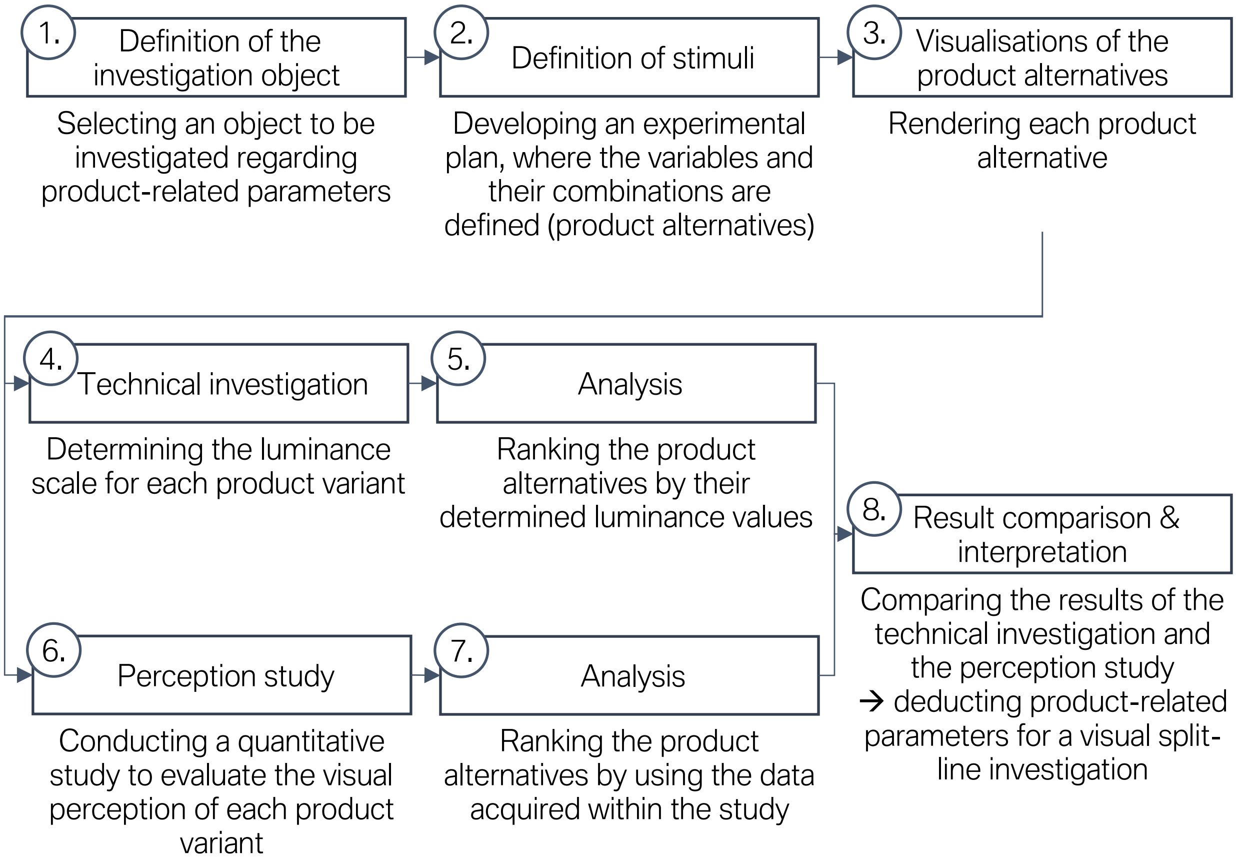


Figure 37. Method for investigating product-related parameters in context of the visual assessment of split-line quality.

Within the first step, the object for the investigation of product-related parameters is defined. Based on the object, the variables which affect the product appearance are identified. For each variable, the split-line attribute and their characteristic are defined. A product alternative, or a stimulus, is represented by a combination of split-line attributes and their characteristic. The product alternatives are visualized using rendering or ray tracing methods (section 2.2). The rendered visualizations are in photorealistic quality to provide a realistic assessment. The method for generating the visualizations is presented in section 6.2. A technical investigation and a perception study are conducted to provide an evaluation of the product-related parameter results. The technical investigation uses the luminance scale to analyze the renderings regarding the visual assessment of split-lines. The luminance scale values of the product alternatives are ranked to get a best-to-worst order of the alternatives. Analogically, the perception study aims to deduce an order of the product alternatives. The perception study is an experimental method for deducing such an order, and participants assess the product alternatives related to different visual split-line assessment tasks. Therefore, the study is implemented by using a browser-based application. The method for designing and implementing the application is presented in Appendix B. In a final step, the results of the technical and of the perception study are compared.

A product-related parameter set for different visual split-line assessment tasks is recommended based on both results and comparison. Additionally, the results of both approaches can be used to verify whether the technical investigation is a substitutional approach for the perception study.

6.2. METHOD FOR GENERATING PHOTOREALISTIC VISUALIZATIONS

Within the thesis, photorealistic visualizations were used in various tasks, i.e., to illustrate the product and the stimuli for assessing the product-related parameters. The photorealistic visualizations were created with the software Blender (version 2.79b) (Blender Foundation 2019a), based on a path tracing method. The approach for generating the visualizations is shown in Figure 38.

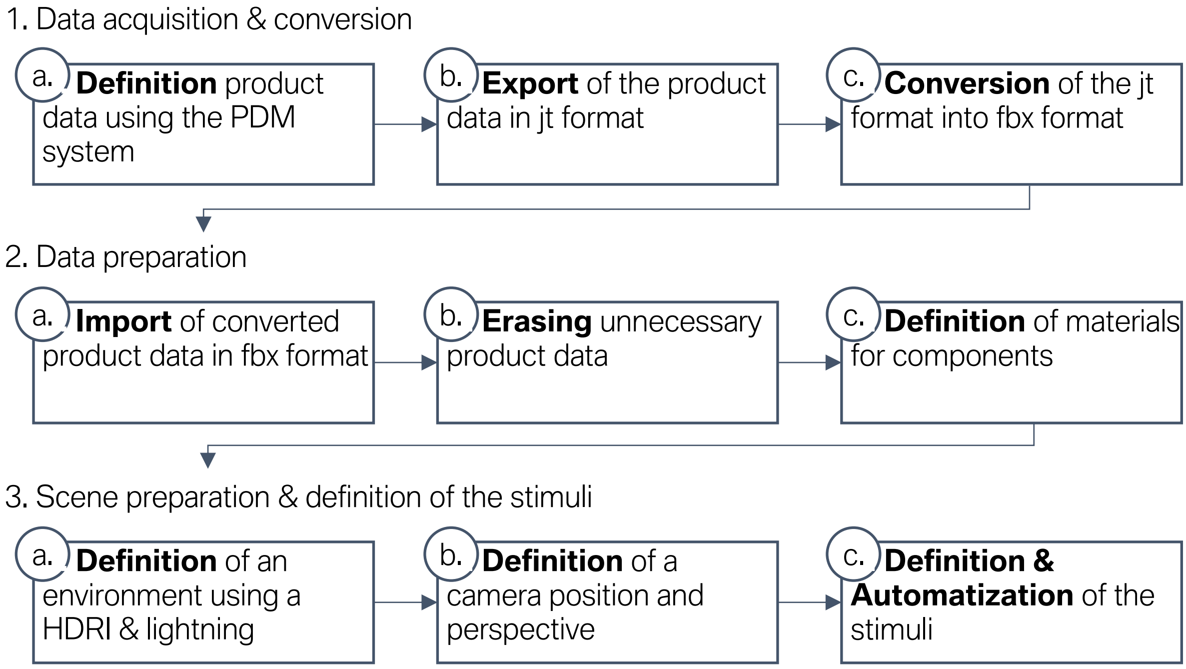


Figure 38. Approach for generating photorealistic visualisations by using the software Blender.

The approach for generating the visualizations was divided into three sub-processes. The first process included data acquisition and the conversion of the data format. Within the first step, the relevant product data was defined using the product-data-management (PDM) system (cf. Figure 38, 1a). Within the thesis, the BMW 5 series was used as an investigation object. Therefore, the product data of this automobile was extracted in the first step. Via a visualization tool, i.e., Teamcenter visualization by Siemens, the second step was performed, and the product data was saved in jupiter tessellation (jt) format (cf. Figure 38, 1b). For realizing a high quality within the visualizations, the data format needed to be converted in the third step into filmbox (fbx) format (cf. Figure 38, 1c). Within this converted data format, information regarding the surface normal was maintained, and an even surface resulted. The second sub-process included data preparation with the software Blender (Blender Foundation 2019). The first step of this sub-process was importing the converted fbx product data (cf. Figure 38, 2a). Cleaning the product data and erasing unnecessary product data was the next step (cf. Figure 38, 2b). For example, the thesis considered the exterior of the automobile. Therefore, the interior data and internal product data, i.e., the engine compartment, was removed. Within the third step of the second sub-process, various materials were assigned to the components (cf. Figure 38, 2c). For example, the windscreen and the side windows were defined as glass, whereas the front flap or the chassis were defined as painted components.

The third sub-process was individually adapted concerning each research hypotheses. In general, the sub-process started with the definition of a high dynamic range image (HDRI) (cf. Figure 38, 3a). These images were the background of the visualization, and the environment the product was placed. The type of HDRI was decisive for the subsequent visualization result, i.e., the reflections and the lighting were affected. An advantage of using an HDRI for lighting settings is a realistic and smooth illumination within the scenery. Subsequently, the camera settings were defined in the third sub-process (cf. Figure 38, 3b). The third sub-process step was manipulating parameters for creating the stimuli (cf. Figure 38, 3c). In this research question of identifying product-related parameters, two different parameters were variated, which were color and reflection. The variable color was variated in red, blue, yellow, black, white, and grey. The variable reflection was modified in metallic, regular, and matte.

The use of photorealistic visualizations had various advantages compared to the use of real images. The advantages were: easy manipulation of the stimuli, simple control of the settings, i.e., camera, lighting, HDRI, the color of the automobile, uniform adjustment of all settings in each visualization, and flexible generation of further visualizations.

The method for designing a Browser based-survey is in the Appendix B.

6.3. APPROACH

Two approaches conducted the approach of investigating the product-related parameters. The generating of the stimuli was identically for both approaches. The conduction of the technical and the perception study was separately performed and is therefore presented individually.

6.3.1. STUDY DESIGN

Like the previous studies, the investigation object for the product-related parameters was a BMW 5 series sedan. The scope of the investigation of the product-related parameters was a global and local product assessment regarding the design and quality of the automobile. Consequently, the product-related parameters were categorized into four tasks, dealing with Global Quality, Global Design, Local Quality, and Local Design. The categories were based on the product-perception matrix, which is presented in Figure 26. As investigation method was an approach chosen, which used a technical and a perception study. In the second step, the results of both investigation methods were compared. An additional objective of this investigation was to find a substitutional method for the perception study.

6.3.2. APPARATUS & STIMULUS

The data for both methods were photorealistic product visualizations, which were generated by the method presented in the Appendix B. The investigation of product-related parameters required other visualizations compared to the previous studies. The stimuli within this investigation were the variables color and reflection. Therefore, the third sub-process of the data generation method was manually modified for each stimulus. As HDRI, a sunny day was used, which was similar to the simulated lighting setting, to investigate the human-related parameters. The camera setting was defined in a more general setting and related to a global assessment derived by Striegel and Zielinski (2018) in the tracking study for the human-related parameter. Therefore, a distance of 3.2 m and a height of 1.6 m were selected. The perspective was a sidewise and frontal view of the entire automobile. For generating the stimuli, the parameters color and reflection were variated. The parameter color was varied in six levels. For a comprehensive investigation, the colors red, blue, yellow, black, white, and grey were considered. The RGB values of the color variables were: red (255, 0, 0), blue (0, 0, 255), yellow (255, 255, 0), black (0, 0, 0), white (255, 255, 255) and grey (0, 0, 0) with 35 % transparency. The variations of the reflection variable were metallic, regular, and matte. The metallic variation was rendered using a second transparent color with additional metallic flakes (Dumont-Bècle et al. 2001). The second color was transparent and glossy for realizing the regular reflection variant (Dumont-Bècle et al. 2001). No additional color created the matte variations. The resulting variants are presented in Figure 39.

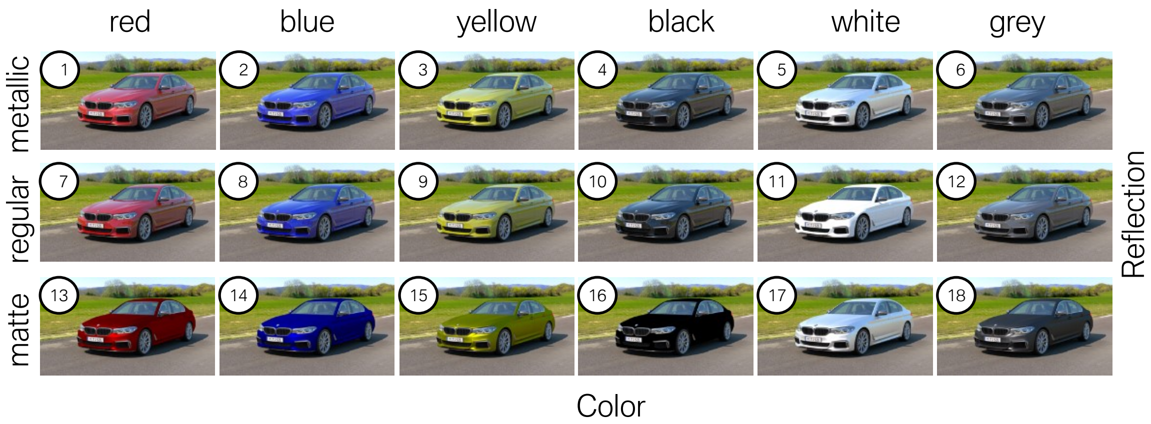


Figure 39. Resulting variants by a full factorial study design of the parameters color and reflection.

The resulting products had slight differences in the color saturation due to the second paint for simulating the reflection levels metallic and regular. The saturation of these variants decreased compared to the mate variants. The study design was full factorial, and therefore, 18 product variants resulted.

6.3.3. Technical Investigation Method

For deducting standardized assessment parameters, two investigation methods were applied. The first investigation method was a technical consideration of the variable color and reflection. The technical investigation method was deduced by the industrial approach of assessing product designs and split-lines. Within the industrial assessment approach, only two levels were considered. The considered levels were the Global Design and the Local Quality levels based on the product-perception matrix. The basic idea for the technical investigation was to use the luminance scale. Within the software Blender, the luminance scale was defined by the light intensity of images, surfaces, squares, or different elements (Blender Foundation 2019c, 2019d, 2019e). Using the luminance scale, visual effects, such as character lines or split-lines, can be identified. Due to the variables color and reflection, the product variants were different regarding the luminance values. The assumption for a subsequent ranking was: the higher the difference between the maximum and minimum luminance value, the better the variant was suitable for an assessment within the category. The approach to measuring the luminance scale is illustrated in Figure 40.

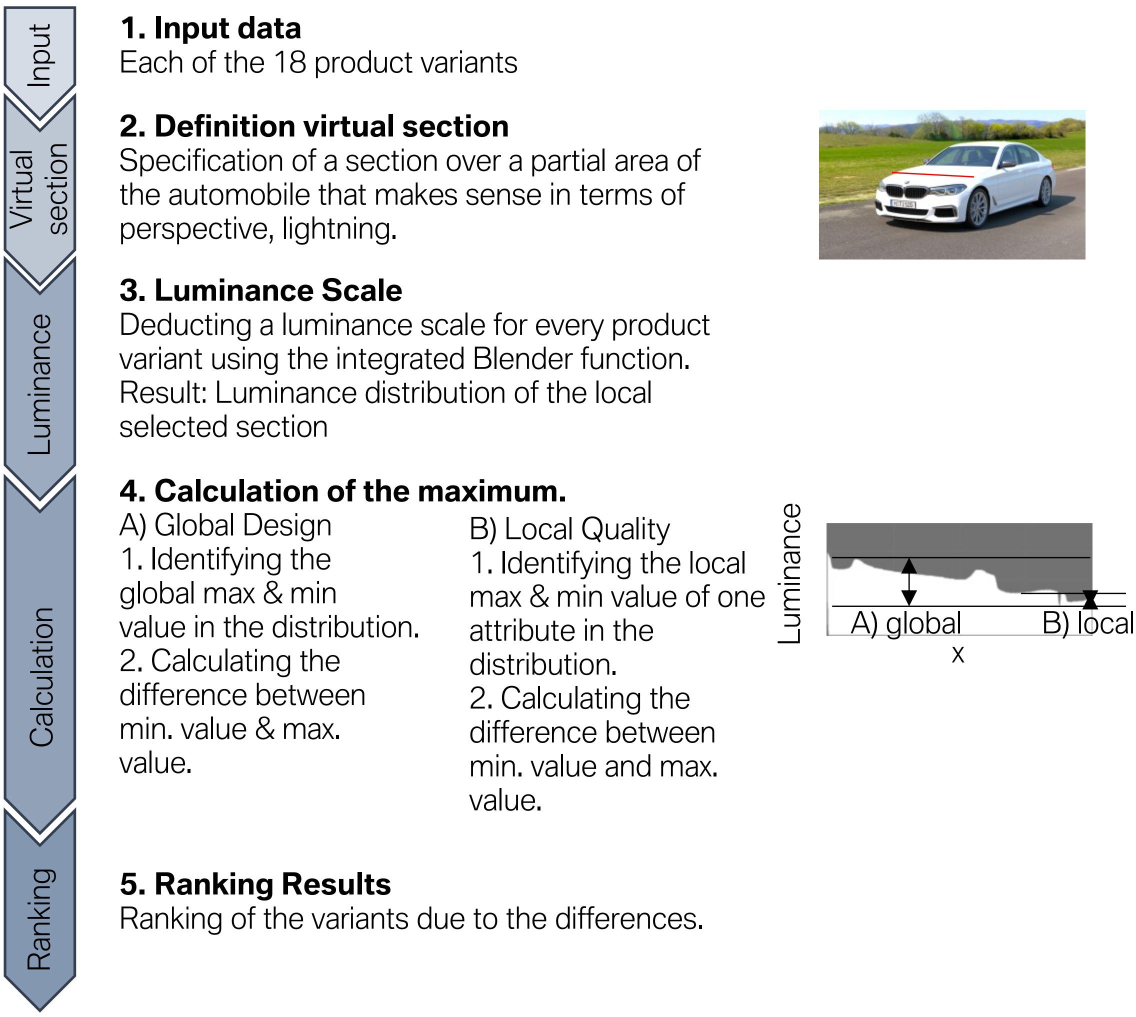


Figure 40. Approach for the calculation of the variants ranking in the technical investigation.

The input data for deducing the luminance ranking were the 18 photorealistic product variants (cf. Figure 39). Within a second step, a virtual section was defined. The defined section was at a similar position within each variant. Within the investigation, the virtual section was placed at the front flap. The luminance was deduced in a third step, related to the virtual section. A Blender integrated function was applied for deducing the luminance at a specific object or line (Blender Foundation 2019b). The luminance at the virtual section was deduced in the form of a luminance distribution for each variant. The distribution had on the x-axis the section length x and the y-axis the luminance value from 0 to 1. Based on the distributions, an analysis regarding the levels of Global Design and Local Quality was conducted. The difference between the maximum and minimum luminance value was calculated for analyzing and ranking the variants. A differentiation between the Global Design and Local Quality consideration was conducted. The differentiation was performed by distinguishing between a local and global minimum and maximum values within the luminance scales. In the Global Design consideration, the global minimum and maximum values of the entire distribution were identified. Within the Local Quality investigation, the local minimum and maximum values of one specific quality attribute were considered. The investigation was performed using the local split-line attribute at the front flap.

6.3.4. PERCEPTION STUDY

The second applied investigation method for considering product-related parameters was a perceptual study. The study's basis was the identical rendered data, which was used in the technical investigation method (cf. Figure 39). The study's scope was to determine the perceptual assessment possibilities of the derived product-related parameters regarding the four levels of the product-perception matrix (cf. Figure 26).

Participants

In summary, the study participants were 33 (female = 15) non-professionals in the field of design or quality assessments but with experience in the automotive industry. The age of the participants was between 23 and 59 years. The experience of the experts in the field of split-line assessments was between 3 and 14 years.

Perception study Questions

In a first step for generating the apparatus and stimulus of the perception study method, the product-perception matrix derived levels were translated into a study question. The resulting study questions are presented in Table 9.

Table 9. Study questions deduced by the perception product matrix.

PPM level	Question	Definition
Global Quality	How precisely can the automobile split-line’s harmony be assessed?	The question asks for a global and a holistic assessment of the automobile’s split-line design due to the different variants.
Global Design	How precisely can the automobile’s appearance and its gestalt be assessed?	Assessment of the variants regarding the entire product appearance including individual design attributes and their context.
Local Quality	How precisely can the course of the split-lines be assessed with regard to its geometric quality (focus on details of individual split-lines)?	The assessment of the variants with regard to split-line details and individual split-line attributes.
Local Design	How precisely can character lines, lighting design, component design and other design features be assessed?	Question scope was, by which product variants design details can be assessed best.

Within the first question, the 18 product variants were examined concerning global split-line quality. Consequently, the focus lay on the product context, the integration, and the combination between split-lines. The aim was to identify the product-related parameter combinations by which the Global Quality context was assessed best. The Global Design level within the study had an abstract and comprehensive scope. The first impression of the entire product and, therefore, the automobile's design and appearance regarding the parameters color and reflection was investigated by the study question. In comparison to this question, the Local Quality level in the third question, the variables color, and reflection were evaluated concerning their reliability of a detailed split-line assessment. The focus lay on individual split-lines and their assessment regarding the product-related parameters. Conclusively, the fourth study question considered the Local Design level. Within this level, the product-related parameters were assessed regarding their assessment possibilities of individual design attributes. Individual design attributes were specified in detail by, i.e., character lines, lighting design, and component design. In contrast to the Local Quality Level, the scope of this question was the parameter context regarding further design attributes than split-lines. Since split-lines were defined as design and quality attributes, their product-related parameters were considered compared to further design attributes.

For conducting the study, a web-based application was used, which was established using the method presented in Appendix B. The application was an HTML-script based on SurveyJS. The structure of the study was split into five sections. Section one gave a short introduction and study details. By using study-included information, each participant had the same knowledge base. Subsequently, the parameter variants of the first question regarding the Global Quality were presented in the second section. The following sections three to five considered the other three study questions. A metric scale level, which had levels between 0 % and 100 %, was used to assess the product variants. This functionality of the assessment slider was implemented by using JavaScript and the library SurveyJS. The order of the variants was randomly varied within the four questions and between the participants to prevent sequence effects in the results. The results of the study were downloaded for each participant automatically.

In the beginning, the participants were introduced, and the background information of the study was presented. The study was conducted with a similar 17" monitor, which had full HD (1920×1080 pixels) and a 16:9 aspect ratio. For each study iteration, the processing time was approximately 30 minutes per participant for the four questions (Table 9).

Data analysis

The results of the perception study were prepared by using descriptive statistical methods. For deriving a ranking order of all product variants for each survey question, the mean value was conducted. Therefore, the following approach for deriving the ranking order was conducted.

The first step was to calculate the mean value for each product alternative v_{aj}. For this calculation following formula was used

$$v_{aj} = \frac{\sum_{j=1}^{n} x_j}{x_j}, \tag{11}$$

with the product alternative a_i (for $i = 1, \dots, 18$) of each participant x_j (for $j = 1, \dots n$ and $n = numbre\ of\ participants$). The resulting mean values v_{aj} were ranked from the highest to the lowest mean value for each study question.

6.4. Results

The results of the technical investigation and the perception study are presented separately. For each investigation, the ranking of the product alternatives based on each analysis method was derived and is presented in this section. Subsequently, a comparison of the results was conducted, and a parameter set for the product-related parameters were derived.

6.4.1. Technical Investigation Method

The technical investigation method was conducted by deriving a ranking of the different product variants. Therefore, the approach presented in chapter 6.3.3 was used to analyze the luminance scale. Concerning the product-perception matrix, a difference between a Global Design and a Local Quality assessment was made. The results for the Global Design and the Local Quality levels are presented in Table 10. Within this table, the calculated luminance differences and the rankings for both levels are illustrated.

Table 10. Resulting technical ranking of the product variants.

Product variant	Reflection	Color	Global luminance difference	Global Design Rank	Local luminance difference	Local Quality Rank
1	Metallic	Red	0.48	8	0.12	8
2	Metallic	Blue	0.50	6	0.09	10
3	Metallic	Yellow	0.34	11	0.25	2
4	Metallic	Black	0.53	3	0.07	12
5	Metallic	White	0.31	13	0.23	3
6	Metallic	Grey	0.53	2	0.10	9
7	Regular	Red	0.50	7	0.13	7
8	Regular	Blue	0.52	4	0.07	12
9	Regular	Yellow	0.36	9	0.29	1
10	Regular	Black	0.55	1	0.08	11
11	Regular	White	0.06	15	0.06	16
12	Regular	Grey	0.50	5	0.21	5
13	Matte	Red	0.05	16	0.04	17
14	Matte	Blue	0.05	17	0.07	15
15	Matte	Yellow	0.31	12	0.14	6

16	Matte	Black	0.04	18	0.07	14
17	Matte	White	0.34	10	0.21	4
18	Matte	Grey	0.14	14	0.03	18

The variants of regular black, metallic grey and metallic black were ranked best for assessing the Global Design level. Metallic grey and metallic black reached the same luminance difference value and were sharing their ranking position. The worst-ranked variants to assess the Global Design level were the matte variants and the regular white variant. The matte variants and bright variants, such as white or yellow, were ranked worst for assessing the Global Design level. On the contrary, dark colors with a regular or metallic reflection were ranked best for assessing the Global Design level. Based on the technical investigation, the best product-related parameters for assessment tasks at a Global Design level were for color black or grey and for reflection, metallic or regular.

Within the Local Quality level, the top-ranked color and reflection variants were regular yellow, metallic yellow, and metallic white. The assessment task of the Local Quality level was a detailed split-line assessment. Due to the shadows and the resulting dark-colored split-line shape, the contrast of a bright-colored product was ranked higher. Thus, bright colors such as yellow, white, or grey were suitable parameters for assessing the Local Quality level. The reflection parameter did slightly influence the ranking results. Mostly, the dark color variants were ranked worse. In summary, the recommendation for assessing split-lines is to choose a bright color, such as white, yellow, or grey, with any reflection variables metallic, regular, or matte. The best-ranked product variants inclusive their luminance derivation for Global Design and Local Quality levels, were presented in Figure 41.

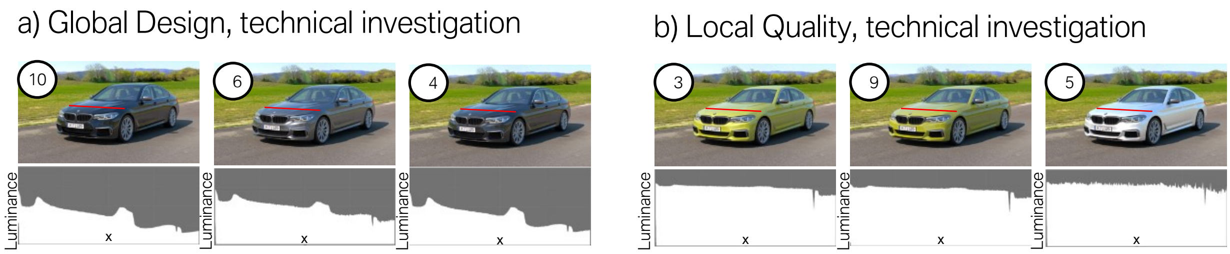


Figure 41. The three best rated product variants in the technical investigation.

The three best-ranked variants of the Global Design Level show each a luminance distribution with a characteristic value increase at the beginning, in the middle, and at the end (cf. Figure 41a)). The variants of the Local Quality Level presented a similar distribution for all three variants (cf. Figure 41b)). Their distribution showed a constant level of luminance without high differences between luminance values. A small luminance value was represented solely in the area of a split-line. Nevertheless, these three distributions differed from the three distributions of the Global Design Level.

6.4.2. Perception study

The collected data analysis was for each study question by calculating the mean value per parameter combination. On this basis, a ranking was deducted. The order sequence was from the best-assessed variant to the worst assessed variant. Therefore, the color and reflection variants which provide an optimal assessment basis for each study question was identified. The following Figure 42 presents the mean values per parameter variant and for each level.

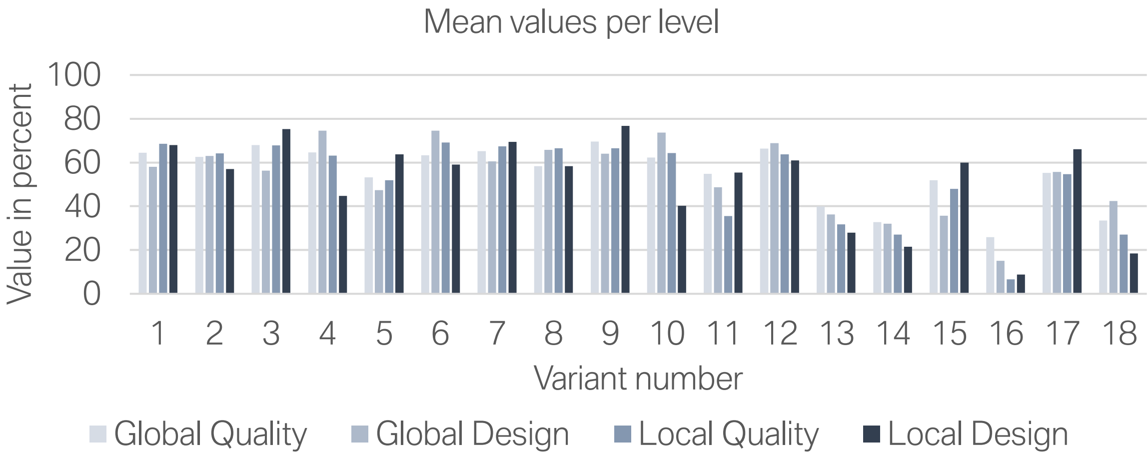


Figure 42. Overview of the study results of all product variants and the four study questions.

Based on the calculated mean values the ranking of the parameters variants was deduced for each level. The resulting ranking is presented in Table 11.

Table 11. Resulting rankings for the product variants by the perception study.

Product variant	Reflection	Color	Global Quality Rank	Global Design Rank	Local Quality Rank	Local Design Rank
1	Metallic	Red	6	9	4	2
2	Metallic	Blue	8	7	11	8
3	Metallic	Yellow	2	10	2	3
4	Metallic	Black	5	1	13	10
5	Metallic	White	13	13	6	12
6	Metallic	Grey	7	2	9	1
7	Regular	Red	4	8	3	4
8	Regular	Blue	10	5	10	5
9	Regular	Yellow	1	6	1	6
10	Regular	Black	9	3	14	7

11	Regular	White	12	12	12	14
12	Regular	Grey	3	4	7	9
13	Matte	Red	15	15	15	15
14	Matte	Blue	17	17	16	16
15	Matte	Yellow	14	16	8	13
16	Matte	Black	18	18	18	18
17	Matte	White	11	11	5	11
18	Matte	Grey	16	9	17	17

Within the Global Quality level, the variants of regular yellow, metallic yellow, and regular grey were identified as product-related parameters that support an assessment. On the contrary, the parameter variants matte grey, matte blue, and matte black were assessed as variants, which do not support Global Quality assessment. For assessing the level of Global Quality, brighter colors in combination with a metallic or regular reflection level were deduced as suitable product-related parameters.

For assessing the level Global Design, the following varnish variants were assessed to be supportive: metallic black, metallic grey, and regular black. The variants identified to be unsupportive within an assessment of the Global Design level: matte yellow, matte blue, and matte black. A product-related parameter set recommendation is a black or grey color in combination with a metallic or regular reflection.

The results of the Local Quality level demonstrated a ranking of the following best parameter variants: regular yellow, metallic yellow, and regular red. The parameter variants which scored poorly were matte blue, matte grey, and matte black. For assessing products and product details in the Local Quality level, a bright color, e.g., yellow, and a metallic or regular reflection level was supportively perceived and is therefore recommended.

Individual design attributes, such as front lights or character lines, were assessed within the Local Design level. The following parameters were identified within the study, which supported the assessment: metallic grey, metallic red, and metallic yellow. The parameter variants identified for not supporting the visual assessment were: matte blue, matte grey, and matte black. These parameter variants were similar to the Local Quality assessment. Equally to the other levels, the matte reflection parameters were assessed worse, and therefore, this reflection parameter is not suitable for one of the four research questions. A recommendable color and reflection parameters for an assessment within the Local Design level are the color grey or yellow and a metallic reflection. The best assessed parameter variants are illustrated in Figure 43.

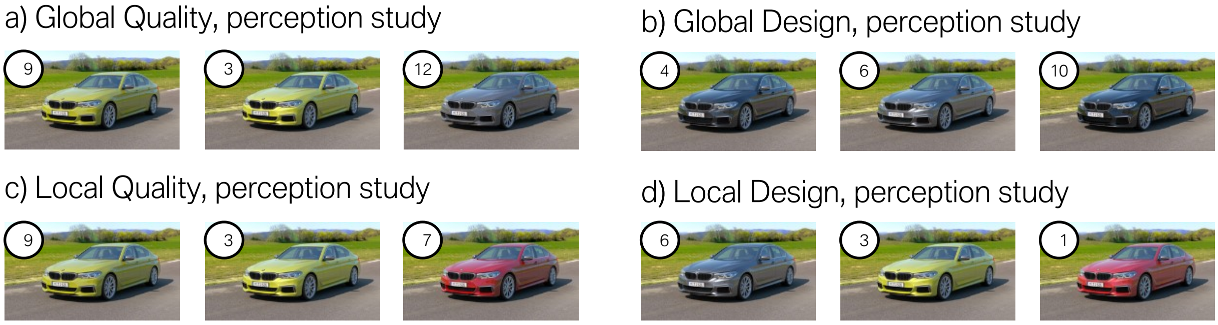


Figure 43. The three best ranked product variants in the perception study.

For the Global and Local Quality level, the color and reflection of the first two parameter combinations were similar. In both levels, the color yellow and the reflection parameters regular were ranked best. Besides, the color parameter yellow and the reflection parameters metallic rated as second best in both levels. Therefore, a yellow color and regular or metallic reflection are supportive and reasonable assessment parameters for a Global or Local Quality assessment.

For the Local and Global Design levels, the metallic grey parameter combination was ranked on position one and two. Therefore, this combination is supportive and reasonable for a Design assessment.

6.4.3. Standardized Product-related Parameters

The third research question considered parameters, which influence the assessment of perceived quality. In order to accomplish this requirement, human-related and product-related parameters were identified. To generate a consistent assessment basis, the parameters that support a visual product assessment were examined. Due to the multi-layered product assessment, this was divided into four explicit levels: Global Quality, Global Design, Local Quality, and Local Design. For the product-related parameters, these four areas were examined concerning the parameters color and reflection. The aim was to identify the parameter combination that supports the corresponding level in a visual assessment. The identified product-related parameters from both investigations favorable for the levels are illustrated in the following Table 12.

Table 12. Summary of technical investigation and perception study.

Level	Technical investigation	Visualization	Perception study	Visualization
Global Quality	color: - reflection: -		color: yellow reflection: regular	
Global Design	color: black reflection: regular		color: black reflection: metallic	
Local Quality	color: yellow reflection: regular		color: yellow reflection: regular	
Local Design	color: - reflection: -		color: grey reflection: metallic	

The levels of Global Quality and Local Design were not considered within the technical investigation method. In summary, for an assessment of Local Quality or Global Quality, the regular yellow parameters are reasonable parameters based on both investigations. For a Global Design assessment based on the technical investigation, the metallic black, based on the perception, study the regular black parameters. Both variants were perceived similarly, and therefore one of them can be chosen. For the Local Design assessment, the metallic grey combination is suggested. Since this combination was also very similar to the metallic and regular black, these can also be summarized. Consequently, the metallic grey variant is suitable for a Global and Local Design assessment. Besides, this variant is recommended for an overall assessment of all four areas. Additionally, this combination is also highly listed in the Global and Local Quality rankings.

6.5. SUMMARY AND CONTRIBUTION OF RESEARCH QUESTION 2

The scope of the previous two chapters was research question two, which focused on the consideration of parameters affecting the assessment process of split-lines regarding perceived quality. Thereby, three-parameter categories were identified: human-related, product-related, and environment-related. Environment-based parameters had effects on the product appearance based on Zeitgeist (Carbon 2010, 2020). Environment-based parameters were not considered within the thesis. These parameters are abstract and follow time aspects, trends, and current events. Within the static consideration in this thesis, these parameters were independent of the considered product and, also, not influenceable. The thesis focuses on the human and product-related parameters. Within this thesis, the human-related and product-related parameters were investigated by using different study types. The human-related parameters were conducted using an empirical study using a qualitative research method (the first question research question two). The participants were observed and tracked during an assessment of split-line design and quality. In doing so, assessment patterns were tracked and subsequently analyzed. As a result, the human-related parameters position and perspective were deduced (the second question research question two). The research method for the product-related parameters was a mixed approach, using a technical, analytical investigation, and a perception study (the fourth question research question two). The product-related parameters were identified as color and reflection (the fifth question research question two). The parameter color was variated in six levels: red, blue, yellow, black, white, and grey. The parameter reflection had three characteristics: metallic, regular, and matte. For the selected full factorial study design, 18 variants were analyzed. The parameters were conducted by using a technical investigation method and a perception study. The technical investigation was based on the luminance analysis of an automobile section. By analyzing the luminance derivation of the parameter combinations, a ranking of the combinations was derived. The assumption of the analysis was, based on a high difference between maximum and minimum luminance value within a derivation, the reflection and the automobile appearance are perceived well, and therefore, the combination is supportive for a visual assessment. The perception study was based on a perceptual study. The parameter combinations were assessed regarding a Global Quality, Global Design, Local Quality, and Local Design level. In conclusion, the results of both investigations were compared, and the similarities and differences were discussed. In summary, the results of both investigations were mainly similar. The study results for the human- and product-related parameters are presented in Table 13. The results were structured by Global Quality, Global Design, Local Quality and Local Design (the sixth question research question two).

Table 13. Derived human-related and product-related parameters for a visual assessment of split-line quality.

	Human-related Parameters		Product-related Parameters	
	Position	Perspective	Color	Reflection
Global Quality	Distance: 3.2 m Height: 1.6 m	Wide angle	Technical: - Perception: yellow	Technical: - Perception: regular
Global Design	Distance: 3.2 m Height: 1.6 m	Wide angle	Technical: black Perception: black	Technical: regular Perception: metallic
Local Quality	Distance: 0.8 m Height: 1.6 m	Small angle	Technical: yellow Perception: yellow	Technical: regular Perception: regular
Local Design	Distance: 0.8 m Height: 1.6 m	Small angle	Technical: - Perception: grey	Technical: - Perception: metallic

Related to the assessment level, Table 13presented reasonable human-related and product-related parameters. The human-related parameters were structured related to the assessment level, either global or local. The global assessment position was 3.2 m for distance, 1.6 m for height with a broad perspective angle. The local assessment parameters were defined as distance with 0.8 m, height with 1.6 m, and a small perspective angle. The recommendation for product-related parameters is metallic grey within a combined assessment of design and quality at global and local levels. This parameter combination was ranked in the first half of the ranking based on the perception study.

Similarly, the technical investigation revealed a metallic grey variant for a Global Design assessment. Since the color yellow is perceived differently, this color can be a disruptive factor. The color grey is, therefore, more subtle and is not intensely perceived. Therefore, the metallic grey parameter presents a good fit by assessing quality and design in combination.

6.5.1. Human-related Parameters

The derivation of the human-related parameters demonstrated a frequently assessed area of the a-pillar. An explanation for this frequent assessment is a complex product concept. Three components, the side panel, the doors, and the front flap, were combined within this area. Due to the many components, the split-lines had a sophisticated design. Based on this sophisticated design, these split-lines' tolerances were high and required an increased assessment effort.

Further conspicuity is the differentiation between the assessment distances, which are explainable by a global and local quality assessment. A global quality assessment includes the global product shape and the context of design elements. Therefore, a greater assessment distance was necessary. In conclusion, an assessment distance of 3.2 m was chosen for global product assessments. Additionally, the greater assessment distance was mainly identified at the front- and rear-end of the automobile. In these areas, many designs and quality attributes stand in contrast and define an automobile's appearance primarily. Therefore, the front- and rear-end were assessed concerning global and local product quality.

The presented method by Ilhan et al. (Ilhan and Çelik 2017; İlhan, Ünal, and Altınok 2017) demonstrated related results to the tracking study, and the estimated values can be used as a first concept. The study by Ilhan et al. (Ilhan and Çelik 2017; İlhan, Ünal, and Altınok 2017) represented the following assessment position: a distance of 1.5 m to 0.5 m and a height of 1.8 m. The tracking-based study results in the thesis show a distance of 3.2 m to 0.8 m and a height of 1.6 m regarding the perception level. Within the tracking study, the results were in detail, executed, and more specific. Additionally, besides the investigation of the assessment position, the tracking-based study in the thesis derived an assessment perspective, which is an additional distinctive feature compared to the estimated values by Ilhan and Çelik (2017); İlhan, Ünal, and Altınok (2017). The derived human-based parameters contribute to various virtual assessment methods. The different assessment positions and perspectives identified a distinguishing feature between the global and local quality assessment. For example, Wickman and Söderberg (2004) studied differences in hardware and virtual assessment tasks. A possible improvement could be achieved by implementing the derived human-related parameters within the study. In conclusion, the deduced position and perspective apply to various virtual research contributions.

A limitation of this method was the derivation of the perspective—the analysis of the perspectives performed by a simplified eye tracking method. The actual eye movement was reduced to the normal vector of the tracking device. A detailed consideration of the parameter perspective by using an eye-tracking device provides further research potential. Based on the scope of the study, the simplified consideration was acceptable to deduce standardized human-related assessment parameters.

6.5.2. Product-related Parameters

The product-related parameters were derived by two investigation methods, a technical investigation, and a perception study. An advantage of this approach is that the results of both investigations can be compared. Both investigation methods derived a ranking of the product-related parameters color and reflection. The parameter color was variated in both investigations in red, blue, yellow, black, white, and grey. The reflection parameter was variated in metallic, regular, and matte. By combining the parameters color and reflection, 18 product variants were generated. The ranking of these product variants using the technical investigation method and perception study a comparison is possible.

A similarity between the technical investigation and the perception study results was the ranking of the matte product variants. Within both investigations, the matte reflection parameter was ranked in the last positions. Additionally, the combinations of matte reflection with dark colors, e.g., black or blue, were rated worse. Considering the luminance derivations of the matte product variants, a less distinctive distribution with a constant low luminance value occurred (cf. Figure 44).

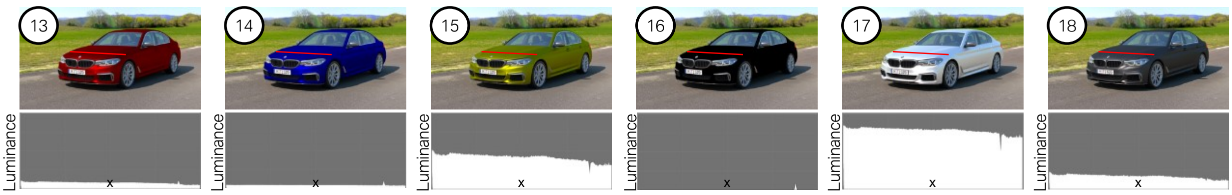


Figure 44. Comparison of the luminance scale of matte product variants using the technical investigation method.

The colors red, blue, black, and grey presented a distribution with a low luminance scale. Based on this low scale, the context of reflections due to the product shape was imperceptible. This result was demonstrated by the technical investigation and the perception study. Within the technical investigation, the calculated difference value was low, and therefore, the resulting ranking position was at the end. Similarly, within the perception study, the ranking positions of these variants were at the end. Thus, both investigations demonstrated that the reflection matte is no reasonable parameter for assessing the four levels.

Further correspondence of the results from both investigations occurred concerning the Local Quality level. The aim of this level is a detailed quality assessment of individual split-line attributes. In both studies, the color yellow was identified as a suitable assessment parameter for the Local Quality level. Furthermore, the perception study revealed that bright colors were preferred in the split-line assessment. Both investigations identified metallic or regular for the parameter reflection to be favorable for a detailed split-line assessment.

The technical investigation rankings and the perception study demonstrated a similarity across almost all positions for the Global Design and the Local Quality level. The rankings of the product variants overlap or deviate by one position. Some exceptions existed, in which the ranking deviate at most four positions between both investigation methods. An example of such a deviation is the metallic red product variant. This variant was on position eight in the technical investigation and on position four in the perception study. Nevertheless, both rankings were very similar and strengthened the validity of the results. Therefore, a conclusion is that the technical investigation with the luminance value assistant is a suitable approximation to the perception study. For first investigations in different areas, the technical investigation can be used as approximation and simulation.

Similar high comparability of the product rankings is provided by the technical investigation and the perception study for the Global Design level. For instance, both investigations presented regular black or metallic grey as favorable product-related parameters at a Global Design level. The product variants regular black and metallic grey indicate good substitutability of both product variants based on their similar luminance distribution. Therefore, the technical investigation indicates a reliable approximation for the perception study.

A compromise for a combined assessment of all levels is the grey metallic or grey regular parameter combination. The color grey and the metallic reflection was ranked in the first part of the ranking table. Exaptation is the Local Quality assessment. Within this level, the grey metallic variant was ranked in the ninth position. Nevertheless, the metallic grey combination demonstrated compromise within the product-related parameter set.

The investigation of the product-related parameters was conducted by a technical investigation and a perception study. Both the study scope of analyzing product-related parameters and the mixed-method approach were new concepts. Therefore, no comparable methods in literature were identified. However, the product-related parameter investigation results supported various virtual assessment methods with Design or Quality scope. For example, the consistent method introduced by Ilhan et al. (Ilhan and Çelik 2017; İlhan, Ünal, and Altınok 2017) to evaluate perceived quality could be enhanced by applying reasonable product-related parameters for each assessment task. Equally, the results of the comparison of the virtual and hardware-based quality perception introduced by Wickman and Söderberg (2004) might be reviewed by applying the parameters per assessment task. The assessment difference between virtual and hardware might be reduced related to an improvement in the parameters. In conclusion, the investigation of the product-related attributes introduced unconsidered influencing factors. Due to the less existing methods, the investigation for these parameters required basic research. Thereby, a different perception of the parameters color and reflection related to a different assessment task was demonstrated. Additionally, a recommendation of a suitable parameter set was deduced.

7. GLOBAL PRODUCT CONTEXT IN PERCEIVED QUALITY

After analyzing the split-line assessment process and the analyses of human-related and product-related parameters, the visual assessment of split-lines regarding perceived quality was further investigated within the third research question. A comparable method concerning the investigation of the global product context and relations between attributes was introduced by Dagman, Söderberg, and Wickman (2004). Their objective was to analyze different product attributes of Visual Quality Appearance (VQA) of an automobile exterior. For an investigation, different VQA attributes were derived in design, color, feeling, VQA of split-lines, material, brand, size, rims, details and chrome. The perception level of the considered attributes ranged between local, e.g. rims or chrome, and global, e.g. design or brand. The assessment of these attributes was performed by industry experts and customers during an interview. As a result, the VQA of split-lines were perceived as an important attribute for both industry and customers. Additionally, an assessment gap and perception asymmetry between experts and customers were identified. The importance of split-lines was higher ranked by experts than by customers. The research scope of Dagman, Söderberg, and Wickman (2004) was the consideration of split-line attribute in relation other attributes. For a detailed investigation of the global product context, the third research question was divided into two studies. The first study considered the importance of the global product context and the second study the perception of variations of split-line attributes. For each study, questions were derived, which are answered in the following two chapters. The derived questions for the investigation of the global product context are:

(1) How are different split-line attributes perceived in different areas of an automobile?
(2) Which split-line attributes designed with smaller tolerances at which area?
(3) How is the impact of the global product appearance influencing the split-line perception?

The importance of a global context and relations of attributes regarding the perception of products were derived by Carbon's conceptual moves (Carbon 2019). For example, the sixth conceptual move addressed the importance of understanding the product context for deducing a harmonious design (Carbon 2019). Within this thesis, the global context of split-lines in product design and the relations between split-line attributes were investigated. A summary and the contribution of the third research question is presented in chapter 8.4.

7.1. General Research Method

The first study investigated the influence of split-lines on product design. Therefore, a conjoint analysis was chosen as the investigation method. The advantages of conjoint analyses are the decompositional approach, where a product is assessed within its entirety. Using this decompositional approach, participants can assess a whole product alternative and not explicitly assess individual product attributes. Therefore, the assessment process is realistic for participants. Based on the variants of conjoint analysis methods, the traditional approach was selected. This TCA approach was selected based on two arguments. Firstly, the traditional split-line assessment process in the industry considers each tolerance attribute individually. In reality, tolerance attributes occur in combination. For providing a realistic assessment situation for participants, the holistic approach of TCAs was suitable. The second argument is that the objective was to investigate the exact influence of individual split-line attributes within a global context. This research intention has an explorative character with less previous information. A product decomposition into the individual attributes was hardly possible. The more reasonable approach was to maintain a holistic product assessment. The method for conducting the conjoint analysis is presented in Figure 37.

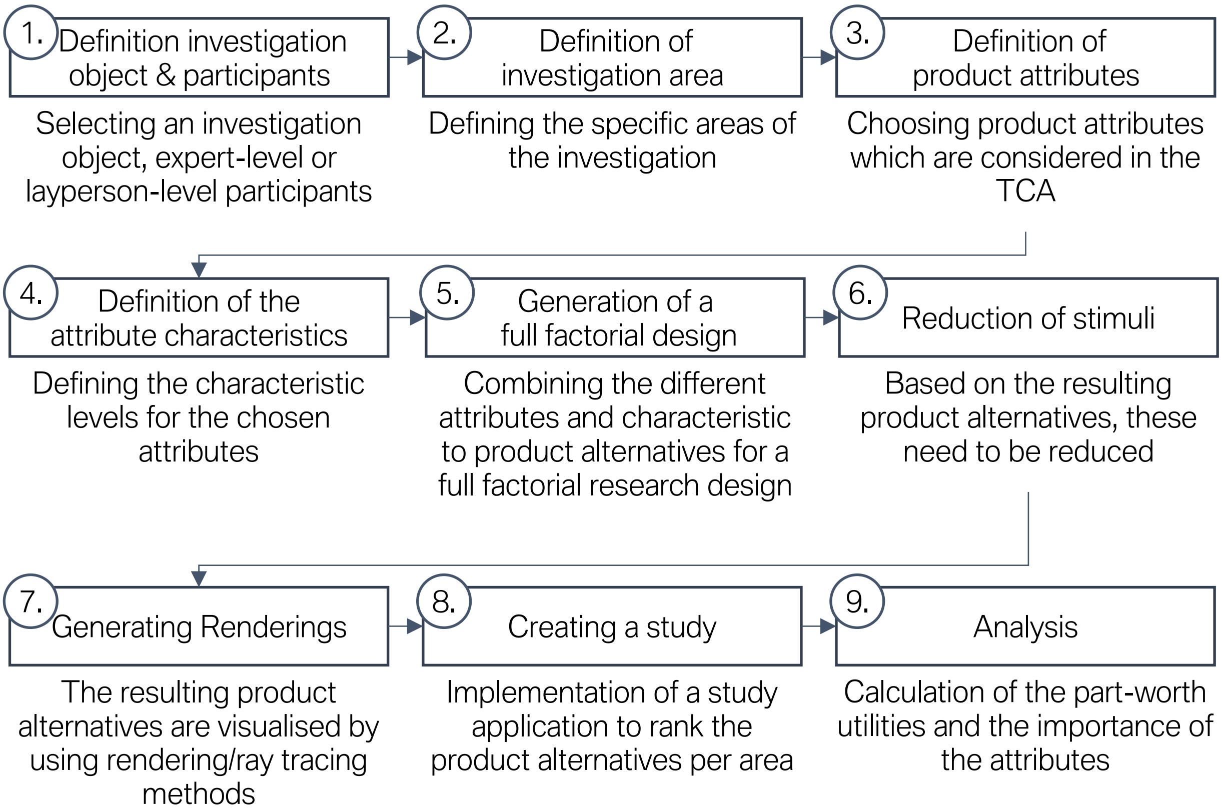


Figure 45. Approach for investigating the effects of the global product context on visual split-line assessments.

The first step is to choose an investigation object and the participants' knowledge level, which can be an expert or layperson level. Based on the chosen object, the specific investigation areas need to be defined. These areas are chosen by different split-line shapes to picture a high variance between the chosen areas. The third step is the definition of product attributes, which are to be considered within the TCA. These attributes are defined by the criteria of relevance, feasibility, independence, compensability, and several attributes (Guckelsberger and Unger 1999). The relevance criterion demands attributes, which have an impact on the purchase decision of customers. The feasibility criterion requires such attributes, which are modifiable by the company. The autonomy of the selected attributes is essential based on the independence criterion. The compensability criterion excludes the necessary presence of an attribute. The limited number of attributes is defined within the fifth criterion. Based on the TCA's defined attributes, the level of attribute characteristics is defined in the fourth step of the investigation approach. The resulting attributes and their characteristic levels are combined in the fifth step of the approach, and a full factorial design of product alternatives (stimuli) results. Thus, the number of product alternatives needs to be limited to a practicable quantity; the full factorial design must be reduced in the sixth step. The seventh step considers the generating of photorealistic visualizations of the remaining product alternatives. The rendering method is presented in chapter 7.2. The study conduction is supported by a browser-based application developed within the eight-step, and the method is presented in chapter Appendix B. The ninth step of the approach includes the study conduction and data collection. The collected data is analyzed by the TCA approach. Thereby, the part-worth utilities and the importance of the attributes are calculated

7.2. METHOD FOR GENERATING PHOTOREALISTIC VISUALIZATIONS

The generation of the photorealistic visualizations is related to the approach presented in chapter 6.2. The sub-process one (data acquisition and conversion) and the sub-process two (data preparation) is similar in both methods. The third sub-process is different between both generation processes. The generation of the stimuli is within this study the manipulation of three split-line attributes. Therefore, an automatization was developed to generate product alternatives. The manipulation of the split-line attributes was performed based on the Blender integrated add-in, called Animation nodes. This add-in represents a visual programming language. The approach of the automatized stimuli generation is shown in Figure 46.

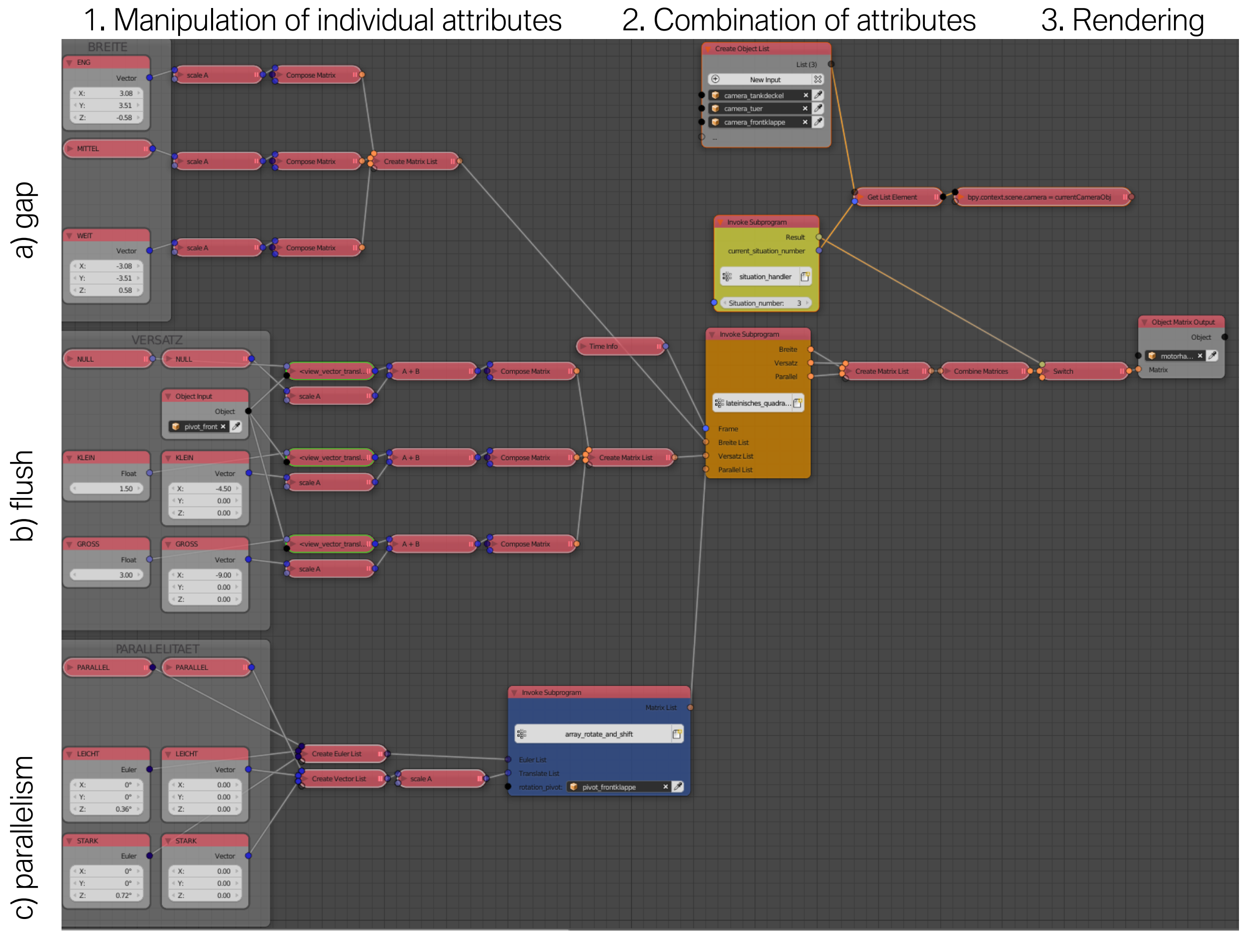


Figure 46. Approach for automatically generating the rendered stimuli by using Blender nodes.

This method was applied in the selected areas of the automobile. The approach for automatically generating the stimuli was similar for each area. Therefore, the approach is explained, considering the area of the front flap. The script was designed from the left to the right side. Within the first step of the script, the number and the value of the individual attributes were varied. The three considered split-line attributes gap, flush, and parallelism was arranged below each other on the left side. Within the second step of the script, the individual split-line attributes were combined, and a full factorial design resulted. The third step of the script defined the automatic start of the rendering process and the automatic saving in a user-defined folder.

7.3. APPROACH

The following three different automobile areas were considered to investigate the third research question: the fuel flap, the doors, and the front flap. These areas are shown in Figure 47. Within each area, the split-line quality of the Refinement level was conducted in a global and holistic product consideration. Due to the different split-line design executions, the three areas were selected.

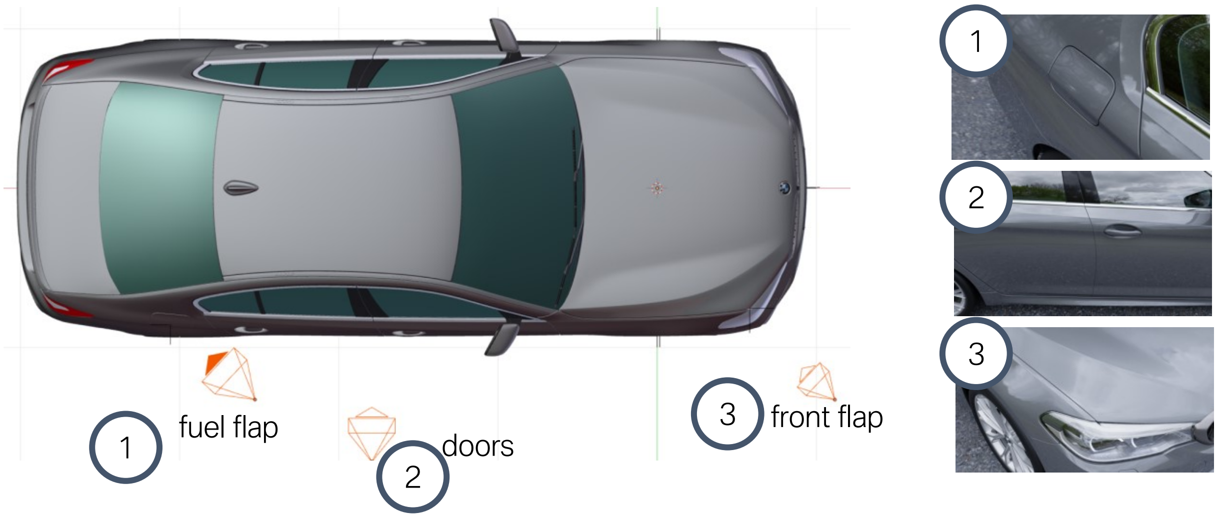


Figure 47. The three chosen areas for the investigation: fuel flap, doors and front flap.

As a first investigation area, the fuel flap was chosen. This split-line design at the fuel flap was oval-shaped, with a character line on the surface, and included two components. For the second investigation area, the doors were considered. A vertical profile characterized this split-line design with a long straight shape, but with additional local design features, i.e., the chromium-plated board. An additional design feature is included at the doors: a chromium-plated board (design feature) combined with a split-line (quality feature). This combination ensured a complicated assessment situation: local design and quality features within a global product context. The third investigation area was the front flap. A horizontal, straight split-line characterized this area with several character lines on the surface of the front flap. Classifying the front light using the product-perception matrix is defined as a local design feature that impacts the global product appearance. Therefore, a similar assessment situation resulted in the front flap and at the doors: a local quality and design assessment of the split-line in the global product context.

The investigation method for analyzing the influence of the product design on split-line assessment was a conjoint analysis. The method for collecting the data was based on the methods presented in the previous section. The detailed method was divided into three steps: participants, apparatus & stimulus, and procedure.

7.3.1. PARTICIPANTS

A number of 34 participants joined the study, which were non-professionals in the field of split-line assessment. The focus group was selected based on potential customers of the automotive premium segment. The age of the participants was between 18 and 60 years.

7.3.2. APPARATUS & STIMULUS

The study was conducted using photorealistic visualization, which was generated by the method presented in chapter 6.2. The third-subprocess (cf. Figure 38), which was the scene preparation, and the definition of the stimuli were individually adapted for the conjoint study. The HDRI for the conjoint study was a panoramic image with cloudy scenery. For defining a realistic camera setting, the distance, the height, and the perspective of the virtual camera were deduced by a study performed by Striegel and Zielinski (2018). Accordingly, the camera's following parameters were chosen: distance up to 0.8 m, the height of 1.6 m eye level, and a perspective with a focus on details. The individual split-line attributes were variated in three levels: perfect, medium, drastic. The level entirely complied with the product construction without tolerance derivations. The second level was defined by a medium split-line tolerance derivation, which was realized by a slight variation from the initial construction. The greatest manipulation of the split-line tolerance was investigated by the attribute level drastic.

The tolerance derivations per split-line attribute and manipulation level were transferred into a vector for generating photorealistic visualizations. These vectors represented a translation for the attributes gap and flush in mm and a rotation for the attribute parallelism in degree. The vectors for each area, attribute, and manipulation level are presented in Table 14.

Table 14. Translation and rotation vectors for realizing the manipulation level.

Area	Attribute	Perfect	Medium	Drastic
Fuel Flap	Gap (mm)	$\vec{G}_{P1} = \begin{pmatrix} 1.005 \\ 1.005 \\ 1.005 \end{pmatrix}$	$\vec{G}_{S1} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$	$\vec{G}_{B1} = \begin{pmatrix} 0.995 \\ 0.995 \\ 0.995 \end{pmatrix}$
	Flush (mm)	$\vec{F}_{P1} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$	$\vec{F}_{S1} = \begin{pmatrix} 0 \\ 1.45 \\ 0 \end{pmatrix}$	$\vec{F}_{B1} = \begin{pmatrix} 0 \\ 2.9 \\ 0 \end{pmatrix}$
	Parallelism (°)	$\vec{P}_{P1} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$	$\vec{P}_{S1} = \begin{pmatrix} 0 \\ 1.2 \\ 0 \end{pmatrix}$	$\vec{P}_{B1} = \begin{pmatrix} 0 \\ 2.4 \\ 0 \end{pmatrix}$
Doors	Gap (mm)	$\vec{G}_{P2} = \begin{pmatrix} -4 \\ 0 \\ 0 \end{pmatrix}$	$\vec{G}_{S2} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$	$\vec{G}_{B2} = \begin{pmatrix} 4 \\ 0 \\ 0 \end{pmatrix}$
	Flush (mm)	$\vec{F}_{P2} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$	$\vec{F}_{S2} = \begin{pmatrix} 0.069 \\ 1.658 \\ 3.117 \end{pmatrix}$	$\vec{F}_{B2} = \begin{pmatrix} 0.138 \\ 3.316 \\ 6.234 \end{pmatrix}$
	Parallelism (°)	$\vec{P}_{P2} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$	$\vec{P}_{S2} = \begin{pmatrix} 0 \\ -0.29 \\ 0 \end{pmatrix}$	$\vec{P}_{B2} = \begin{pmatrix} 0 \\ -0.58 \\ 0 \end{pmatrix}$
Front Flap	Gap (mm)	$\vec{G}_{P3} = \begin{pmatrix} 3.08 \\ 3.51 \\ -0.58 \end{pmatrix}$	$\vec{G}_{S3} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$	$\vec{G}_{B3} = \begin{pmatrix} -3.08 \\ -3.51 \\ 0.58 \end{pmatrix}$
	Flush (mm)	$\vec{F}_{P3} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$	$\vec{F}_{S3} = \begin{pmatrix} -3.072 \\ 0.263 \\ 0.377 \end{pmatrix}$	$\vec{F}_{B3} = \begin{pmatrix} -6.144 \\ 0.527 \\ 0.753 \end{pmatrix}$
	Parallelism (°)	$\vec{P}_{P3} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$	$\vec{P}_{S3} = \begin{pmatrix} 0 \\ 0 \\ 0.36 \end{pmatrix}$	$\vec{P}_{B3} = \begin{pmatrix} 0 \\ 0 \\ 0.72 \end{pmatrix}$

A risk within in conjoint studies is a quick increasement of the product alternatives, which participants need to evaluate. Product alternatives were defined as the combination of attributes with different manipulation levels. Green and Srinivasan (1990) recommended a number of 20 different product alternatives. Based on experience, these alternatives represent a manageable size for participants to rank. This number was exceeded rapidly since the number of attributes multiplied with the level. Considering a full-factorial design of the conducted study with three attributes and three characteristics, the resulting number of product alternatives was 27. Therefore, a systematic selection of stimuli was necessary. One method for reducing the alternatives considering a symmetrical study design was the Latin square method (Backhaus, Erichson, and Weiber 2015). This reduction method was used because of the symmetrical study design. Each split-line attribute had an equal number of manipulation levels. In addition to the condition of a symmetrical study design, the Latin square method was applicable because the attributes and the levels each amounted to three. According to the application of the Latin square method, the product alternatives were reduced to nine. The advantage of reduced study design was that each attribute and level was represented in a product alternative. The conjoint analysis calculates the importance and utility value for each attribute and level. Therefore, concluding the analysis, each combination of a full factorial design can be calculated. The reduced study design and the alternatives are presented in Figure 48.

1 Gap: perfect Flush: perfect Parallelism: perfect	2 Gap: medium Flush: perfect Parallelism: medium	3 Gap: drastic Flush: perfect Parallelism: drastic
4 Gap: perfect Flush: medium Parallelism: medium	5 Gap: medium Flush: medium Parallelism: drastic	6 Gap: drastic Flush: medium Parallelism: perfect
7 Gap: perfect Flush: drastic Parallelism: drastic	8 Gap: medium Flush: drastic Parallelism: perfect	9 Gap: drastic Flush: drastic Parallelism: medium

Figure 48. Resulting and reduced study design after applying the Latin square method.

Within the method for generating the visualizations, these product alternatives were rendered automatically. This automatization was defined as the first step of the script (cf. (cf. Figure 46). The combination of the vectors was performed in the second step and resulted in a full factorial design. The reduction of the product alternatives was performed in the third stage of the script.

Within the conjoint analysis, assessing the resulting nine product alternatives was done by ranking the alternatives (Clancy and Garsen 1970). As described in Appendix B, the study used for acquiring the data was a browser-based application with two study sections, having a general and a study-related part. The design of this application is generally explained in Appendix B. The general part addressed the question about an individuum's automobile property. If any was available, the segment of the automobile was asked. The study-related part was adjusted regarding the ranking method of the conjoint analysis. Within the conjoint analysis, a drag-and-drop technique was used. The adjustment was conducted within the second step of the

JavaScript code, which performed the study functions. The approach and illustration of this technique is shown in Figure 49.

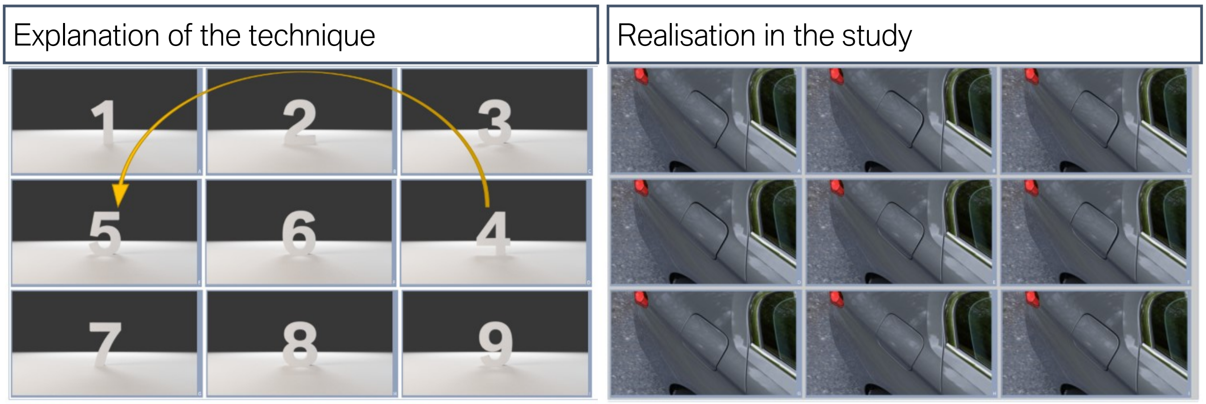


Figure 49. Realised drag-and-drop technique for ranking the product alternatives within the browser-based study application.

The ranking of the product alternatives was executed regarding the three areas: the fuel flap, the doors, and the front flap. The order of these three areas was similar between the participants. The order of the visualizations per area was randomly sampled to avoid order effects in the data.

Each participant was separately questioned. Therefore, a short introduction was presented. An introduction was included in the browser-based application to guarantee similar information for each participant. The average processing time was 15 min per participant. The study was performed using a 17'' monitor with a display resolution of full HD (1920×1080 pixels) and a 16:9 aspect ratio. The acquired data analysis was performed by using the traditional conjoint analysis approach, which deduces the ranking of the alternatives, the utility values, and the attribute importance.

7.3.3. Data Analysis

The chosen conjoint analysis type was a TCA. The calculation of the part-worth utilities and the importance value was performed by using the approach of Guckelsberger & Unger (1999). Therefore, the total utility value y_i of a stimulus is calculated by using the additive model [40]:

$$y_i = \mu + \sum_{k=1}^{K} \sum_{j=1}^{m_k} \beta_{k_j} * \sigma_{k_j} \tag{12}$$

y_i, which is the estimated total utility value of the stimulus i,

β_{k_j}, which is the part-worth utility value of an attribute m with the level k, σ_{k_j} is a binary value of 1 or 0, which depends if the stimulus i, an attribute m and level k is available

μ is the basis utility value

The idea of the estimation of the part-worth utilities is that the individual values of stimuli differ from the basis utility value μ (Guckelsberger and Unger 1999). Therefore, the least square method can be used for estimating the part-worth utilities (Guckelsberger and Unger 1999). This approach results in a minimization problem, where the total utility value of a product alternative is estimated, so that the value differs as little as possible from the real ranking position (Guckelsberger and Unger 1999).

$$\sum_{i=1}^{n}(r_i - y_i)^2 \ \rightarrow min \tag{13}$$

r_i is the ranking position of a stimulus i

By using the calculation of the total utility value y_i (1) the following minimization problem results (Guckelsberger and Unger 1999).

$$\sum_{i=1}^{n}(r_i - \mu + \sum_{k=1}^{K}\sum_{j=1}^{m_k}\beta_{k_j} * \sigma_{k_j})^2 \ \rightarrow min \tag{14}$$

For calculation of this problem, the following approach is used (Guckelsberger and Unger 1999).

In a first step of the approach, the basis utility value μ is calculated (Guckelsberger and Unger 1999).

$$\mu = \frac{1}{n}\sum_{i=1}^{n} r_i \tag{15}$$

n is the number of levels,

r_i is the ranking position of a stimulus i

In a second step, the average ranking position μ_{k_j} is calculated for each attribute m with each level k (Guckelsberger and Unger 1999). Therefore, the each ranking position, in which an attribute with an level appears, is summed and divided by the number of ranking positions (Guckelsberger and Unger 1999). In a third step, the part-worth utility values β_{K_j} are calculated by subtracting basis utility value μ from the the average ranking position μ_{k_j} (Guckelsberger and Unger 1999).

$$\beta_{K_j} = \mu_{K_j} - \mu \tag{16}$$

In a fourth step, the resulting part-worth utility value is transformed to reach values of a positive value $\beta`_{K_j}$.

$$\beta`_{K_j} = \beta_{K_j} - min\beta_k \tag{17}$$

Within a fifth step, this positive value $\beta`_{K_j}$ is normalised to a part-worth utility value $\beta``_{K_j}$ to be comparable to different importance values:

$$\beta``_{K_j} = \frac{\beta`_{K_j}}{\sum_{K=1}^{k}\max\beta`_{K_j}} \tag{18}$$

The normalized total utility value y'_i of a product alternative is the sum of the normalized part-worth utility values $\beta``_{K_j}$.

The importance value of an attribute is calculated in three further steps. In the first step, the level range s_m of an attribute m is calculated by subtracting the minimal part-worth utility value $\beta_{k\ min}$ from the maximal part-worth utility value $\beta_{k\ max}$ (Backhaus et al. 2016).

$$s_m = \beta_{k\,max} - \beta_{k\,min} \tag{19}$$

In the second step, the level range s_m of each attribute (m=1, …, n) are summed up (Backhaus et al. 2016).

$$s_{ges} = s_{m1} + \cdots + s_{mn} \tag{20}$$

In the third step, the attribute importance per attribute W_m is calculated by dividing the individual level range s_m by the total level range s_{ges} (Backhaus et al. 2016).

$$W_m = \frac{s_m}{s_{ges}} \tag{21}$$

7.4. Results

The conjoint analysis subject was an investigation of the influence of the global product context on the perception of split-lines. The results are presented regarding the ranking of the alternatives, the part-worth utilities, and the importance values for the areas fuel flap, doors, and front flap.

7.4.1. Ranking of the alternatives

The first result of the study was the ranking of the visualizations concerning the three considered areas. This ranking of the product alternatives was calculated by the average ranking choices of the 34 individual participants, which is illustrated in Figure 50.

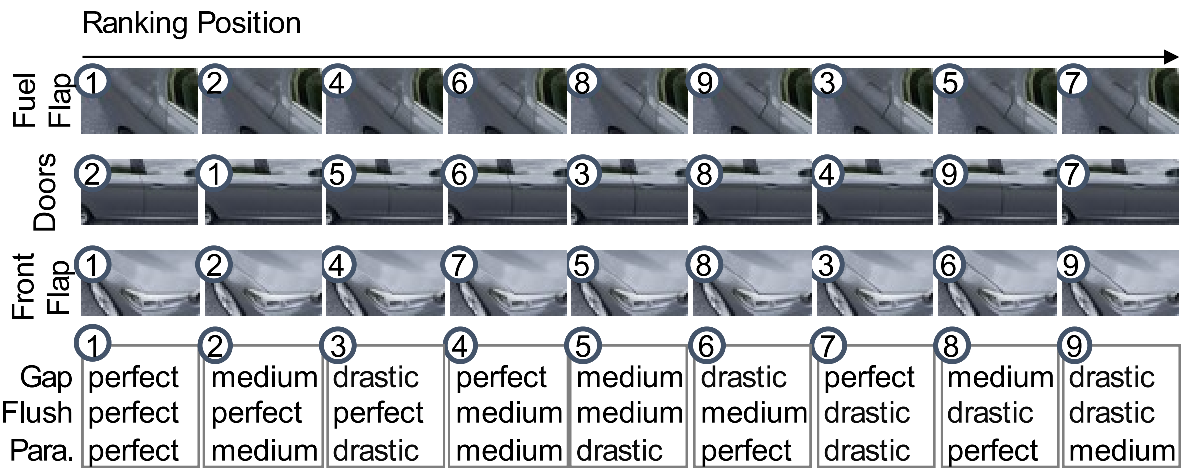


Figure 50. Ranking result per area of the split-line quality assessment within the conjoint analysis.

The figure is structured as follows: the best rating is placed on the left side. The averaged visualization ratings get worse when going right. Besides the visualization rankings, explanations are presented concerning the stimuli at the button. The top-ranking within the perspective fuel flap had the characteristics gap perfect, flush perfect, and parallelism perfect. Also, these characteristics were ranked for the top position within the perspective front flap. However, the results of these two perspectives stand in contrast to the perspective doors, which had the best ranking for the stimuli gap medium, flush: perfect, parallelism medium. In comparison, the three rankings of the perspectives show slight differences. One variation was the top-ranking position. Whereas the perspectives fuel flap and front flap concluded a small gap with no flush and perfect parallel split-line level, the perspective doors ensured as the top-ranking position a split-line with medium gap and parallelism but with no flush. The perspectives fuel flap and front flap revealed further similarities in the ranking positions two, three, and seven. An almost entirely different ranking demonstrated the perspective doors with the perspective fuel flap. There were two positions, four and nine, similarly ranked in the perspectives of fuel flap and doors. In comparison to the perspective front flap, the sixth position was similarly listed in the perspective doors.

In addition to the first results, these rankings are also the basis for the conjoint analysis and the calculation of the utility values and the attribute importance.

7.4.2. Part-worth Utility value

Additional results of the conjoint analysis and the including calculation were the part-worth utility values for each tolerance attribute and their level. Due to the reduced study design, not each alternative was queried by the participants. Nevertheless, the reduction by the Latin square method's assistance allowed the calculation of every alternative utility value in conclusion. Therefore, the individual utility values for each alternative were calculated based on the individual alternative rankings. The utility values were standardized for subsequent comparison of the different product areas. The results of the different standardized utility values are presented in Figure 51.

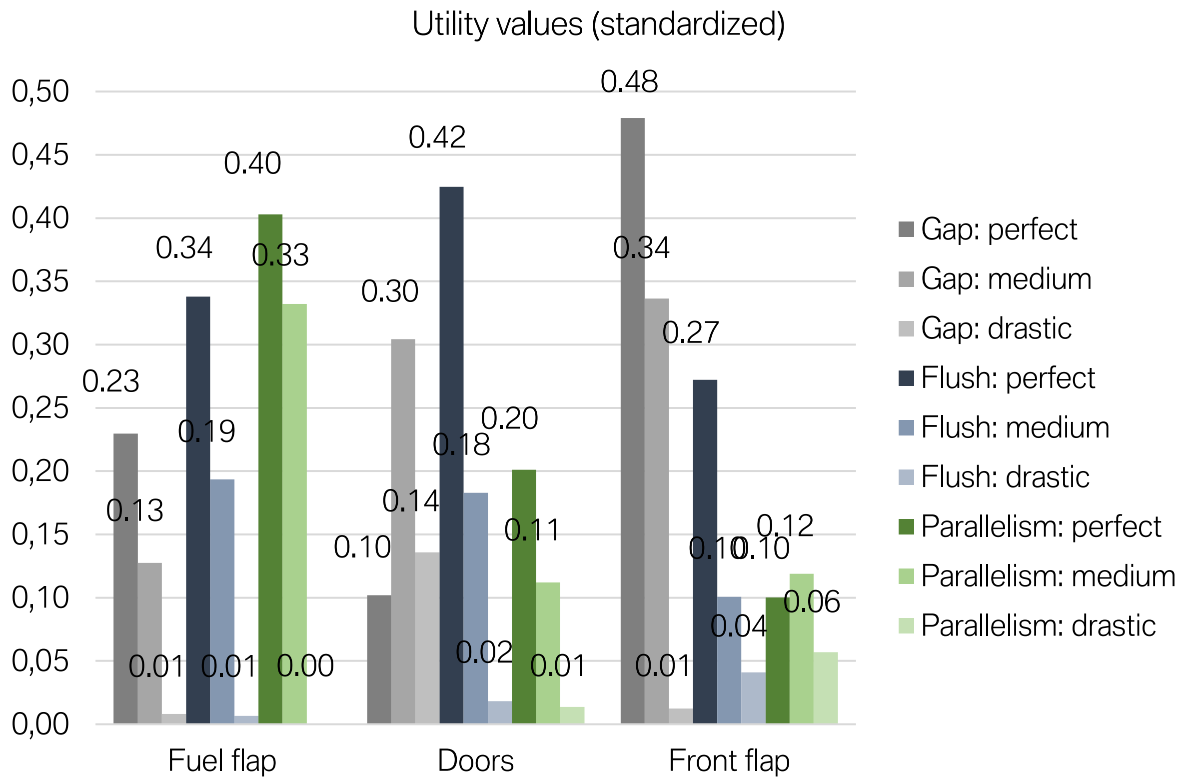


Figure 51. Resulting part-worth utility values per investigated area using the conjoint analysis.

The figure is structured by the three perspectives fuel flap, doors, and front flap. Within these perspectives, the tolerance attributes gap, flush, and parallelism and their levels are presented. The tolerance attribute parallelism demonstrated the highest part-worth utility value by reaching 0.40 and the lowest value by having 0.00 at the fuel flap. The split-line attribute combination with the highest utility value for the front flap was gap, flush, and parallelism with perfect execution of all three attributes. In contrast to the fuel flap, the calculation within the area doors presented the tolerance attribute flush with a high difference of the highest utility value, reaching 0.42, and the lowest value reaching 0.01. This difference was an indicator of a high importance of this attribute at the area doors. Within the doors' area, the attribute combination with medium gap, perfect flush, and perfect parallelism gained the highest utility value. The third area front flap contrasted both previous results. The front flap results presented the tolerance attribute gap with the most considerable difference of the highest utility value 0.48, and the lowest value 0.01. Therefore, the highest utility value was reached with the split-line combination of perfect gap, perfect flush, and medium parallelism.

7.4.3. Attribute Importance

A primary objective of conjoint analysis is the derivation of the attribute importance. In the analysis, each attribute importance concerning the three perspectives was calculated based on the above presented standardized utility values.

The attribute importance values of the fuel flap perspective reached 24 % for the attribute gap of 34 % for flush, and the highest value got the attribute parallelism by having 42 %.

In contrast, the importance values within the perspective doors achieved 35 % for the attribute gap, 43 % for the attribute flush, and 22 % for the third attribute parallelism. The third area front flap reached the importance values of the attribute gap of 53 %, the attribute flush attained 29 %, and the attribute parallelism obtained 18 %. Similar to evaluating the utility values, the evaluation of the attribute importance indicated different priorities regarding the three perspectives. The essential attribute within the perspective fuel flap was parallelism, whereas the attribute flush was influential in the second perspective doors, and the attribute gap identified the highest importance value considering the front flap. A summarizing illustration of the calculated attribute importance is presented in Figure 52.

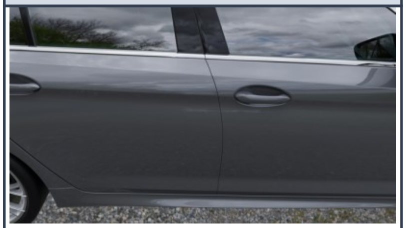
Fuel Flap	Doors	Front Flap
		
1. Parallelism: 42 % 2. Flush: 34 % 3. Gap: 24 %	1. Flush: 43 % 2. Gap: 35 % 3. Parallelism: 22 %	1. Gap: 53 % 2. Flush: 29 % 3. Parallelism: 18 %

Figure 52. Importance values of the split-line attributes per area using the conjoint analysis.

In summary, the conjoint study was conducted to gain information about customer's product attribute prioritization gap, flush and parallelism. A resulting statement is that for all three areas the different tolerance attributes were decisive and the customer's importance ranking allowed different attributes to be considered important in each of the three areas.

8. PERCEPTION OF SPLIT-LINE ATTRIBUTES

The third research question was divided into two parts. The first part was investigating the global product context on the assessment of split-lines regarding perceived quality. The second part considered the relation between split-line attributes regarding perceived quality. The objective of this part was a deep dive into the investigated areas and a detailed investigation of the effect of combinations of split-line attributes. Therefore, the following questions were investigated within this research part:

(1) How do split-line attributes affect each other concerning perceived quality?
(2) Which impact has the area of consideration and the combination of split-line attributes on the perception of quality?
(3) How do the results of both studies affect each other?
(4) Which attribute is dominant, what are conspicuousness of attribute levels, what effects have attribute combinations, and what patterns are in the diagrams?
(5) Which design guidelines can be derived for a high level of perceived quality regarding the split-line attributes?

This research part focuses on deducing basic knowledge of split-line perception. Research and industry have a research gap regarding the perception of an individual or combined split-line attributes. The third research question focuses on closing the gap and deducing new knowledge. Therefore, recommendations regarding the design and tolerancing of split-line attributes are derived by using both research parts.

8.1. METHOD

The research of effects of split-line attribute combinations on perceived quality is performed based on a quantitative research approach. Therefore, an experimental study for considering theses effects is used. The study is related to the previous research using a TCA. The method for the consideration of relations between split-line attributes is presented in Figure 53.

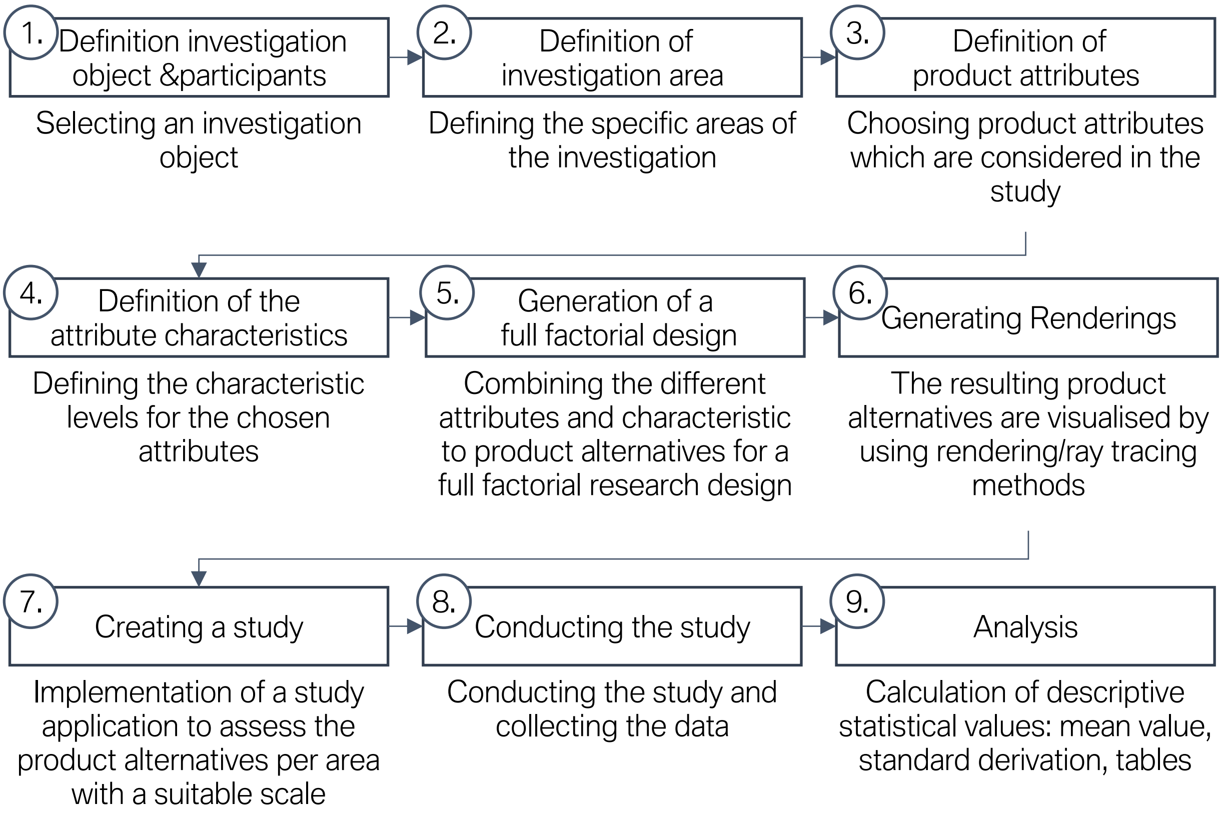


Figure 53. Method for investigating split-line attribute relations on perceived quality.

The first five steps of the approach for investigating split-line attribute relations on perceived quality are similar to the research approach in chapter 7.1. The first step is to define an investigation object and the knowledge level of the participants. Within the second step, the areas for investigation are defined in detail. Steps three and four focus on the selection of attributes and their characteristic levels. Based on this selection, a full factorial design is derived. The resulting product alternatives of the full factorial design are visualized by using ray tracing methods. This visualization process is similar to the process described in chapter 7.2. Similarly, the process step of implementing a browser-based study application is related to Appendix B. The eight-step is the procedure of the study by using the product alternatives and the browser-based study application. Within the ninth step, the collected data is analyzed using descriptive statistical methods, which are the calculation of mean value and standard derivation. Additionally, for each area, a table for an easy visual comparison of the product alternatives is derived.

8.2. APPROACH

The experimental study to analyze the effect of relations between split-line attributes was based on the conjoint analysis approach. Similarly, the experimental study focused on the three automobile areas: fuel flap, doors and front flap (Figure 47). The method for the perceptual study is further explained in the following section, by presenting the participants, apparatus & stimulus and the procedure.

8.2.1. PARTICIPANTS

The participants of the study were non-professionals in the field of split-line quality assessment. In comparison to experts, non-professionals perceive split-line quality differently. Due to a daily confrontation, experts have different experiences and can receive detailed knowledge about split-lines. Thus, a difference between the perception process exists. For companies, the perception of customers is decisive, and therefore, non-professionals had participated. The number of participants was 40, and the participants' average age was 23,9 years and a range of 18 to 51. 32 participants were female.

8.2.2. APPARATUS & STIMULUS

The perceptual study was performed by using photorealistic visualizations of the BMW 5 series sedan. The method for generating the visualizations was presented in chapter 6.2. The third sub-process was adopted for the perceptual study. The parameters like HDRI, camera settings, and automatic stimuli generation was identical to the conjoint study. The investigated variables were the split-line attributes gap, flush, and parallelism in the area of the fuel flap, the doors, and the front flap. Compared to the conjoint study, the attributes were variated in five levels. The range of the levels was based on a previous split-line concept. The five levels were specified regarding: perfect, slight, medium, high, and drastic. The levels increased linearly from perfect to drastic. Similar to the conjoint study, manipulating the split-line attributes was performed by using a translation and rotation vector to relocate components. The translation vector affected the split-line attributes gap and flush. The rotation vector influenced the level of the split-line attribute parallelism. The numerical translation vector was in mm, and the values in the rotation vector were defined in degree. The study design was full factorial, and therefore, the split-line attributes and their levels were combined. The total number of product alternatives was 125. The numerical values for each split-line attribute and the respective manipulation level are presented in Table 15.

Table 15. Levels and the translation and rotation manipulations.

Area	Attribute	Perfect	Slight	Medium	High	Drastic
Fuel Flap	Gap (mm)	$\vec{G}_{P1} = \begin{pmatrix}1.005\\1.005\\1.005\end{pmatrix}$	$\vec{G}_{P1} = \begin{pmatrix}1.0025\\1.0025\\1.0025\end{pmatrix}$	$\vec{G}_{S1} = \begin{pmatrix}1\\1\\1\end{pmatrix}$	$\vec{G}_{P1} = \begin{pmatrix}0.9975\\0.9975\\0.9975\end{pmatrix}$	$\vec{G}_{B1} = \begin{pmatrix}0.995\\0.995\\0.995\end{pmatrix}$
	Flush (mm)	$\vec{F}_{P1} = \begin{pmatrix}0\\0\\0\end{pmatrix}$	$\vec{F}_{P1} = \begin{pmatrix}0\\0.73\\0\end{pmatrix}$	$\vec{F}_{S1} = \begin{pmatrix}0\\1.45\\0\end{pmatrix}$	$\vec{F}_{P1} = \begin{pmatrix}0\\2.18\\0\end{pmatrix}$	$\vec{F}_{B1} = \begin{pmatrix}0\\2.9\\0\end{pmatrix}$
	Parallelism (°)	$\vec{P}_{P1} = \begin{pmatrix}0\\0\\0\end{pmatrix}$	$\vec{P}_{P1} = \begin{pmatrix}0\\0.6\\0\end{pmatrix}$	$\vec{P}_{S1} = \begin{pmatrix}0\\1.2\\0\end{pmatrix}$	$\vec{P}_{P1} = \begin{pmatrix}0\\1.8\\0\end{pmatrix}$	$\vec{P}_{B1} = \begin{pmatrix}0\\2.4\\0\end{pmatrix}$
Doors	Gap (mm)	$\vec{G}_{P2} = \begin{pmatrix}-4\\0\\0\end{pmatrix}$	$\vec{G}_{P2} = \begin{pmatrix}-2\\0\\0\end{pmatrix}$	$\vec{G}_{S2} = \begin{pmatrix}0\\0\\0\end{pmatrix}$	$\vec{G}_{P2} = \begin{pmatrix}2\\0\\0\end{pmatrix}$	$\vec{G}_{B2} = \begin{pmatrix}4\\0\\0\end{pmatrix}$
	Flush (mm)	$\vec{F}_{P2} = \begin{pmatrix}0\\0\\0\end{pmatrix}$	$\vec{F}_{P2} = \begin{pmatrix}0.035\\0.829\\1.559\end{pmatrix}$	$\vec{F}_{S2} = \begin{pmatrix}0.069\\1.658\\3{,}117\end{pmatrix}$	$\vec{F}_{P2} = \begin{pmatrix}0.1035\\2.487\\4.676\end{pmatrix}$	$\vec{F}_{B2} = \begin{pmatrix}0.138\\3.316\\6.234\end{pmatrix}$
	Parallelism (°)	$\vec{P}_{P2} = \begin{pmatrix}0\\0\\0\end{pmatrix}$	$\vec{P}_{P2} = \begin{pmatrix}0\\0.145\\0\end{pmatrix}$	$\vec{P}_{S2} = \begin{pmatrix}0\\0.29\\0\end{pmatrix}$	$\vec{P}_{P2} = \begin{pmatrix}0\\0.44\\0\end{pmatrix}$	$\vec{P}_{B2} = \begin{pmatrix}0\\0{,}58\\0\end{pmatrix}$
Front Flap	Gap (mm)	$\vec{G}_{P3} = \begin{pmatrix}3.08\\3.51\\-0.58\end{pmatrix}$	$\vec{G}_{P3} = \begin{pmatrix}1.54\\1.755\\-0.29\end{pmatrix}$	$\vec{G}_{S3} = \begin{pmatrix}0\\0\\0\end{pmatrix}$	$\vec{G}_{P3} = \begin{pmatrix}-1.54\\-1.755\\0.29\end{pmatrix}$	$\vec{G}_{B3} = \begin{pmatrix}-3.08\\-3.51\\0.58\end{pmatrix}$
	Flush (mm)	$\vec{F}_{P3} = \begin{pmatrix}0\\0\\0\end{pmatrix}$	$\vec{F}_{P3} = \begin{pmatrix}-1.536\\0.132\\0.189\end{pmatrix}$	$\vec{F}_{S3} = \begin{pmatrix}-3.072\\0.263\\0.377\end{pmatrix}$	$\vec{F}_{P3} = \begin{pmatrix}-4.608\\0.395\\0.566\end{pmatrix}$	$\vec{F}_{B3} = \begin{pmatrix}-6.144\\0.527\\0.753\end{pmatrix}$
	Parallelism (°)	$\vec{P}_{P3} = \begin{pmatrix}0\\0\\0\end{pmatrix}$	$\vec{P}_{P3} = \begin{pmatrix}0\\0\\0.18\end{pmatrix}$	$\vec{P}_{S3} = \begin{pmatrix}0\\0\\0.36\end{pmatrix}$	$\vec{P}_{P3} = \begin{pmatrix}0\\0\\0.54\end{pmatrix}$	$\vec{P}_{B3} = \begin{pmatrix}0\\0\\0.72\end{pmatrix}$

The realization of the perceptual study was supported by using a browser-based application related to the presented method in Appendix B. The study structure was divided into a general section and a study-related section. Within the general section, the age, the gender, the current automobile, and segment were questioned. The study-related section was clustered concerning the three areas, the fuel flap, the doors, and the front flap. The order of these three areas remained similar between the participants. The photorealistic visualizations within the three areas were randomly varied to avoid sequence errors within the data. Like the conjoint study, the second step for designing the browser-based application needed to be customized for the perceptual study. The second step is considered the measurement technique of the study. Each product alternative was assessed by using a seven-point Likert scale (Joshi et al. 2015). The Likert scale is metric and ranges traditionally between strongly disagree to strongly agree (Joshi et al. 2015). This scale was adapted within the study. The seven-point scale Likert scale levels were ranged between extremely likely and not likely at all. The assessment of a product alternative within the study is presented Figure 54.

Please assess the perceived quality of the split-lines at the fuel flap.

1. Please assess (1: extremely likely, 7: not likely at all).

1	2	3	4	5	6	7
○	○	○	○	○	○	○

Page 1 of 126

Previous Next

Figure 54. Visualisation of the browser-based study design for assessing the split-line quality and considering split-line attribute relations.

The study was conducted for each participant separately. In the beginning, information about the study was provided within the browser-based application. A participant needed approximately 25 minutes to complete the study. The used monitor size was 17'' with full HD (1920×1080 pixels) and a 16:9 aspect ratio.

8.2.3. Data Analysis

The collected data was proceeded by using descriptive-statistical methods. For each investigated area, the mean value and the standard derivation was calculated for each product alternative. The formula for the calculation of the mean value m_a per product alternative a is

$$m_a = \frac{1}{n}\left(\sum_{i=1}^{n} b_i\right), \tag{22}$$

Where n is the number of participants, b_i is the resulting assessment of a product alternative related to the 7-point likert scale. Based on the mean value m_a calculation, the standard derivation sd_a for each product alternative a is calculated using following formula,

$$sd_a = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(b_i - m_a)^2}. \tag{23}$$

8.3. RESULTS

The results were presented in the following section related to each investigated area. For each area, the best and worst assessed product alternative and the mean value of the results were calculated. Additionally, the assessment ranges of single attributes by a constant perfect level of the other attributes is presented.

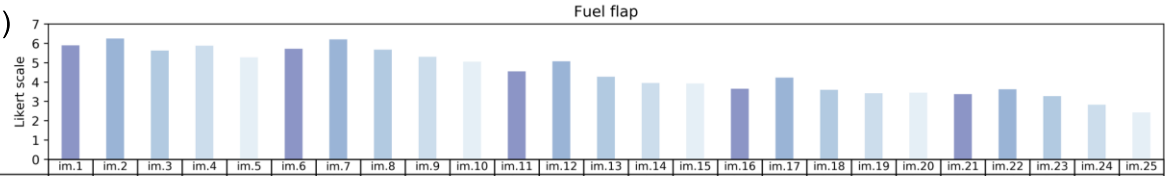
8.3.1. FUEL FLAP

Within the area of the fuel flap, the assessment scale of the results ranged from 6.2 to 1.7. The best assessed attribute variants at the fuel flap were images 2 and 7, the worst image was number 124. A summary of the best and the worst combinations is followed:

- Image 7: gap slight
flush perfect
parallelism perfect
- Image 2: gap slight
flush slight
parallelism perfect
- Image 124: gap high
flush drastic
parallelism drastic

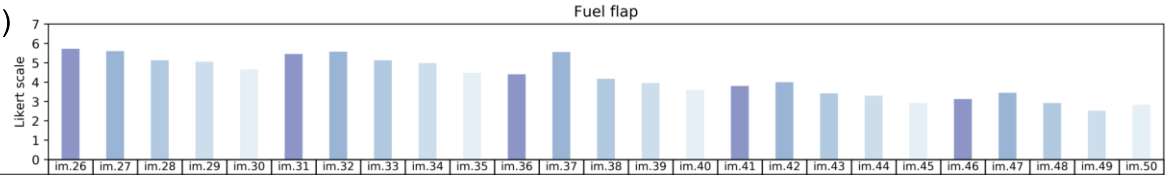
These attribute variants reached a mean value of 6.2 and a standard derivation of 1.1 for image 2 and 1.0 for image 7. The worst assessed image reached a mean value of 1.7 and a standard derivation of 1.2. The assessment range of the single attribute gap by a constant perfect level of flush and parallelism was between 6.2 and 5.3 of the mean value. For the attribute flush and a constant perfect level of gap and parallelism, the mean values ranged between 5.9 and 3.4. The mean value of the attribute parallelism and constant perfect levels of gap and flush variated between 5.9 and 2.1. The widest range within the assessment and of a variation in the single attribute demonstrated the attribute parallelism. The results of all 125 attribute variants are presented in Figure 55.

a)



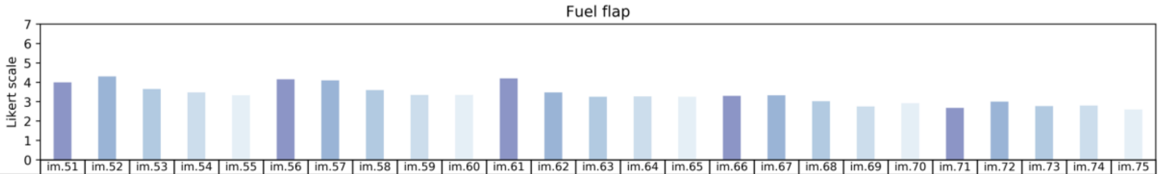
	im.1	im.2	im.3	im.4	im.5	im.6	im.7	im.8	im.9	im.10	im.11	im.12	im.13	im.14	im.15	im.16	im.17	im.18	im.19	im.20	im.21	im.22	im.23	im.24	im.25
mean	5.9	6.2	5.6	5.9	5.3	5.7	6.2	5.7	5.3	5.0	4.5	5.1	4.3	4.0	3.9	3.6	4.2	3.6	3.4	3.5	3.4	3.6	3.3	2.8	2.4
sd	1.5	1.1	1.5	1.2	1.4	1.4	1.0	1.4	1.6	1.2	1.6	1.3	1.4	1.6	1.2	1.3	1.7	1.4	1.5	1.4	1.4	1.6	1.3	1.3	1.2
gap	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
flush	1	1	1	1	1	2	2	2	2	2	3	3	3	3	3	4	4	4	4	4	5	5	5	5	5
parallelism	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

b)



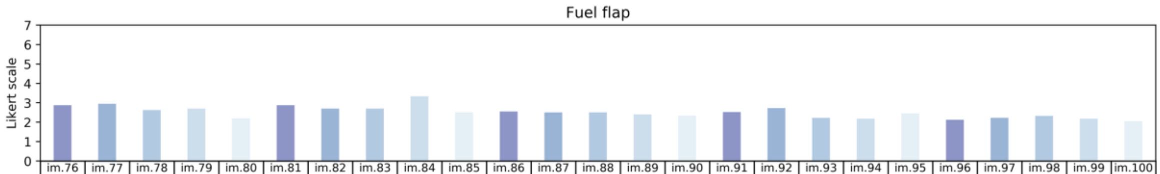
	im.26	im.27	im.28	im.29	im.30	im.31	im.32	im.33	im.34	im.35	im.36	im.37	im.38	im.39	im.40	im.41	im.42	im.43	im.44	im.45	im.46	im.47	im.48	im.49	im.50
mean	5.7	5.6	5.1	5.0	4.7	5.5	5.6	5.1	5.0	4.5	4.4	5.5	4.2	4.0	3.6	3.8	4.0	3.4	3.3	2.9	3.1	3.5	2.9	2.5	2.8
sd	1.3	1.5	1.4	1.3	1.5	1.4	1.5	1.4	1.4	1.7	1.4	1.4	1.5	1.5	1.4	1.5	1.4	1.2	1.5	1.3	1.3	1.6	1.3	1.5	1.4
gap	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
flush	1	1	1	1	1	2	2	2	2	2	3	3	3	3	3	4	4	4	4	4	5	5	5	5	5
parallelism	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2

c)



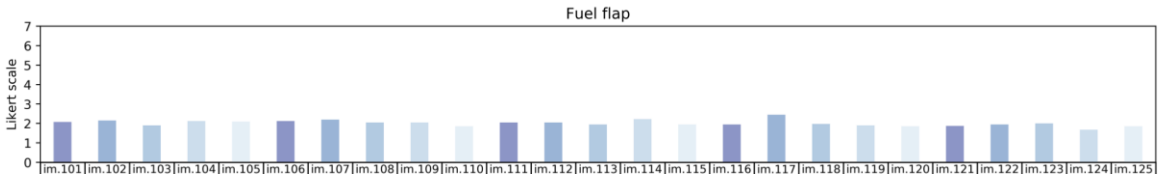
	im.51	im.52	im.53	im.54	im.55	im.56	im.57	im.58	im.59	im.60	im.61	im.62	im.63	im.64	im.65	im.66	im.67	im.68	im.69	im.70	im.71	im.72	im.73	im.74	im.75
mean	4.0	4.3	3.6	3.5	3.3	4.2	4.1	3.6	3.4	3.4	4.2	3.5	3.2	3.3	3.2	3.3	3.3	3.0	2.8	2.9	2.7	3.0	2.8	2.8	2.6
sd	1.8	1.6	1.5	1.5	1.5	1.7	1.7	1.6	1.4	1.6	1.9	1.7	1.4	1.4	1.3	1.1	1.5	1.4	1.5	1.6	1.2	1.4	1.2	1.6	1.4
gap	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
flush	1	1	1	1	1	2	2	2	2	2	3	3	3	3	3	4	4	4	4	4	5	5	5	5	5
parallelism	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3

d)



	im.76	im.77	im.78	im.79	im.80	im.81	im.82	im.83	im.84	im.85	im.86	im.87	im.88	im.89	im.90	im.91	im.92	im.93	im.94	im.95	im.96	im.97	im.98	im.99	im.100
mean	2.9	3.0	2.6	2.7	2.2	2.9	2.7	2.7	3.3	2.5	2.5	2.5	2.5	2.4	2.3	2.5	2.7	2.2	2.2	2.5	2.1	2.2	2.3	2.2	2.0
sd	1.4	1.7	1.4	1.5	1.2	1.5	1.4	1.2	1.9	1.3	1.3	1.3	1.3	1.3	1.4	1.6	1.4	1.3	1.2	1.3	1.1	1.3	1.4	1.2	1.4
gap	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
flush	1	1	1	1	1	2	2	2	2	2	3	3	3	3	3	4	4	4	4	4	5	5	5	5	5
parallelism	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4

e)



	im.101	im.102	im.103	im.104	im.105	im.106	im.107	im.108	im.109	im.110	im.111	im.112	im.113	im.114	im.115	im.116	im.117	im.118	im.119	im.120	im.121	im.122	im.123	im.124	im.125
mean	2.1	2.1	1.9	2.1	2.1	2.1	2.2	2.0	2.0	1.9	2.0	2.0	1.9	2.2	1.9	1.9	2.5	2.0	1.9	1.9	1.9	1.9	2.0	1.7	1.9
sd	1.2	1.3	0.9	1.2	1.4	1.3	1.2	1.2	1.2	1.1	1.2	1.3	1.3	1.3	1.2	1.3	1.4	1.3	1.1	1.1	1.3	1.2	1.2	1.2	1.1
gap	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
flush	1	1	1	1	1	2	2	2	2	2	3	3	3	3	3	4	4	4	4	4	5	5	5	5	5
parallelism	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5

Figure 55. Resulting split-line quality assessments of the area fuel flap.

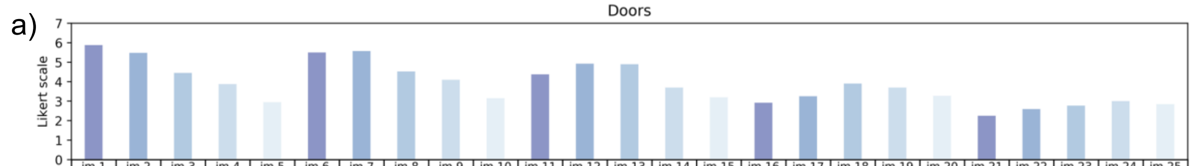
8.3.2. DOORS

At the area doors the assessment range was between 6.0 and 1.9. The best assessed attribute variants were image 26 and 27. The worst assessed image was image 122. Following the attribute combinations were:

- Image 26: gap slight
  flush perfect
  parallelism perfect
- Image 27: gap slight
  flush slight
  parallelism perfect
- Image 122: gap high
  flush drastic
  parallelism drastic

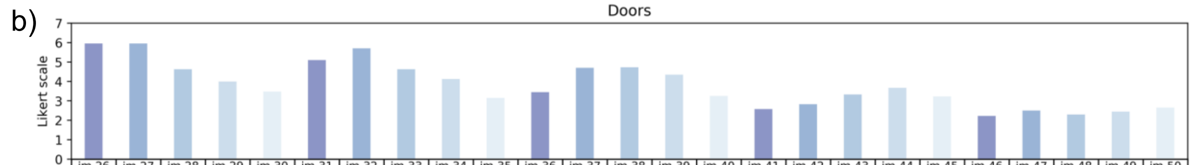
The standard derivations of the best-assessed images were 1.4 for image 26 and 1.1 for image 27. Image 122, with the worst ranking, had a standard derivation of 1.3. The single attribute gap with a constant perfect level of the attributes flush and parallelism variated between 5.9 and 3.0. The attribute flush assessment range by a constant perfect level of the attributes gap and parallelism was between 5.9 and 2.2. The third attribute parallelism by constant perfect gap and flush levels variated between 5.9 and 3.2. Figure 56 presents the different results of the 125 attribute variants.

a)



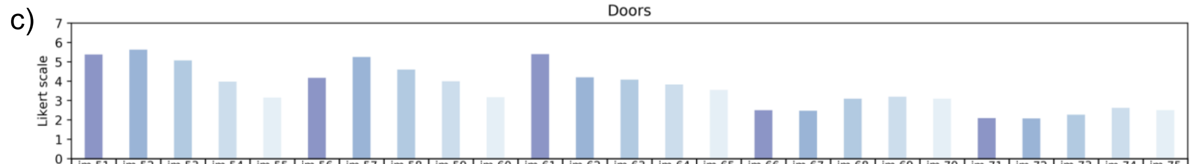
	im.1	im.2	im.3	im.4	im.5	im.6	im.7	im.8	im.9	im.10	im.11	im.12	im.13	im.14	im.15	im.16	im.17	im.18	im.19	im.20	im.21	im.22	im.23	im.24	im.25
mean	5.9	5.5	4.5	3.9	3.0	5.5	5.6	4.5	4.1	3.1	4.4	4.9	4.9	3.7	3.2	2.9	3.2	3.9	3.7	3.3	2.2	2.6	2.8	3.0	2.9
sd	1.6	1.2	1.3	1.2	1.4	1.6	1.3	1.5	1.6	1.4	2.0	1.6	1.3	1.3	1.1	1.7	1.9	1.7	1.6	1.5	1.3	1.4	1.5	1.6	1.6
gap	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
flush	1	1	1	1	1	2	2	2	2	2	3	3	3	3	3	4	4	4	4	4	5	5	5	5	5
parallelism	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

b)



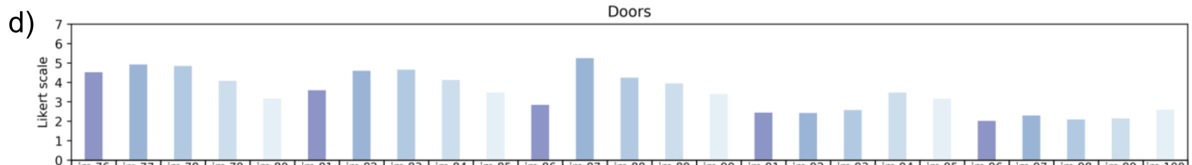
	im.26	im.27	im.28	im.29	im.30	im.31	im.32	im.33	im.34	im.35	im.36	im.37	im.38	im.39	im.40	im.41	im.42	im.43	im.44	im.45	im.46	im.47	im.48	im.49	im.50
mean	6.0	6.0	4.6	4.0	3.5	5.1	5.7	4.6	4.1	3.1	3.5	4.7	4.7	4.3	3.2	2.6	2.8	3.3	3.7	3.2	2.2	2.5	2.3	2.5	2.6
sd	1.4	1.1	1.2	1.3	1.5	1.7	1.4	1.1	1.5	1.4	1.6	1.7	1.5	1.3	1.4	1.5	1.6	1.5	1.5	1.5	1.5	1.3	1.3	1.4	1.5
gap	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
flush	1	1	1	1	1	2	2	2	2	2	3	3	3	3	3	4	4	4	4	4	5	5	5	5	5
parallelism	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2

c)



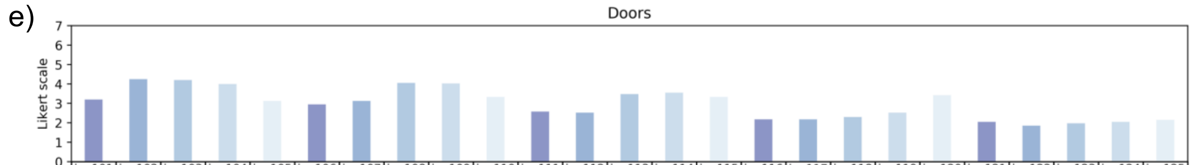
	im.51	im.52	im.53	im.54	im.55	im.56	im.57	im.58	im.59	im.60	im.61	im.62	im.63	im.64	im.65	im.66	im.67	im.68	im.69	im.70	im.71	im.72	im.73	im.74	im.75
mean	5.4	5.6	5.1	4.0	3.1	4.2	5.2	4.6	4.0	3.2	5.4	4.2	4.1	3.8	3.5	2.5	2.5	3.1	3.2	3.1	2.1	2.1	2.3	2.6	2.5
sd	1.5	1.4	1.2	1.3	1.5	2.0	1.3	1.5	1.5	1.4	1.6	1.7	1.5	1.4	1.5	1.6	1.4	1.3	1.4	1.6	1.4	1.1	1.4	1.4	1.3
gap	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
flush	1	1	1	1	1	2	2	2	2	2	3	3	3	3	3	4	4	4	4	4	5	5	5	5	5
parallelism	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3

d)



	im.76	im.77	im.78	im.79	im.80	im.81	im.82	im.83	im.84	im.85	im.86	im.87	im.88	im.89	im.90	im.91	im.92	im.93	im.94	im.95	im.96	im.97	im.98	im.99	im.100
mean	4.5	4.9	4.8	4.1	3.2	3.6	4.6	4.7	4.1	3.5	2.9	5.2	4.2	4.0	3.4	2.5	2.4	2.6	3.5	3.1	2.0	2.3	2.1	2.1	2.6
sd	1.9	1.6	1.2	1.4	1.3	2.0	1.6	1.3	1.5	1.3	1.5	1.4	1.3	1.4	1.4	1.3	1.5	1.2	1.6	1.4	1.2	1.2	1.2	1.3	1.4
gap	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
flush	1	1	1	1	1	2	2	2	2	2	3	3	3	3	3	4	4	4	4	4	5	5	5	5	5
parallelism	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4

e)



	im.101	im.102	im.103	im.104	im.105	im.106	im.107	im.108	im.109	im.110	im.111	im.112	im.113	im.114	im.115	im.116	im.117	im.118	im.119	im.120	im.121	im.122	im.123	im.124	im.125
mean	3.2	4.2	4.2	4.0	3.1	3.0	3.1	4.0	4.0	3.3	2.6	2.5	3.5	3.5	3.3	2.2	2.2	2.3	2.5	3.4	2.0	1.9	2.0	2.0	2.1
sd	1.8	1.6	1.4	1.6	1.4	1.6	1.6	1.6	1.4	1.3	1.5	1.3	1.5	1.5	1.4	1.2	1.3	1.2	1.2	1.3	1.4	1.0	1.1	1.3	1.4
gap	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
flush	1	1	1	1	1	2	2	2	2	2	3	3	3	3	3	4	4	4	4	4	5	5	5	5	5
parallelism	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5

Figure 56. Resulting split-line quality assessments of the area doors.

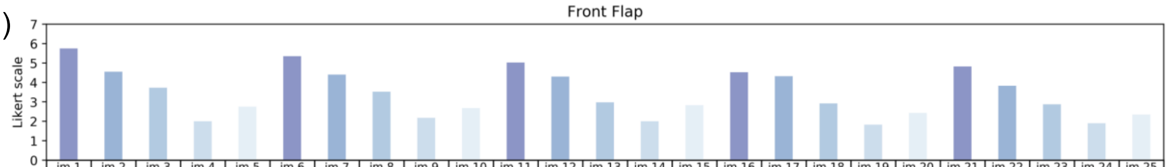
8.3.3. Front Flap

At the area front flap, the best and worst assessed attribute variants were represented in image 51 and image 19 and the attribute levels were:

- Image 51: gap perfect
  flush perfect
  parallelism medium
- Image 19: gap high
  flush high
  parallelism perfect

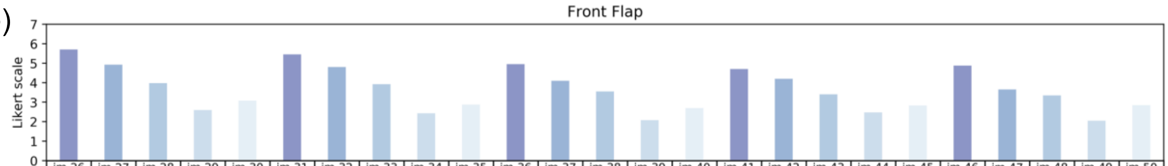
Image 51 was assessed best with a mean value of 6.0 and a standard derivation of 1.6. In contrast, image 19 was assessed worst with a mean value of 1.8 and a standard derivation of 1.0 regarding the PQ impression. The single-level variation of the attribute gap with constant level flush and parallelism was 5.8 and 2.0. A single variation of the attribute flush and constant gap and parallelism resulted in an assessment range of 5.8 to 4.8 for the mean value. Considering the variation range of the attribute parallelism with a constant level of gap and flush resulted between 6.0 and 5.8. Figure 57 illustrates the results of the 125 attribute variants at the area of the front flap.

a)



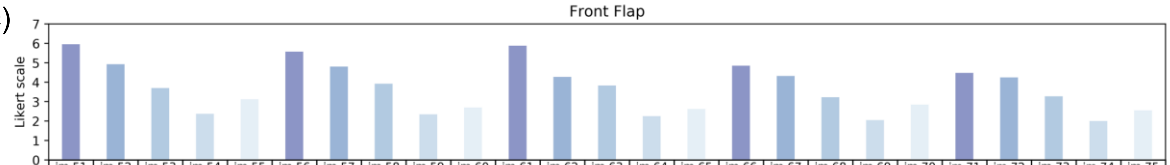
	im.1	im.2	im.3	im.4	im.5	im.6	im.7	im.8	im.9	im.10	im.11	im.12	im.13	im.14	im.15	im.16	im.17	im.18	im.19	im.20	im.21	im.22	im.23	im.24	im.25
mean	5.8	4.5	3.7	2.0	2.8	5.3	4.4	3.5	2.2	2.7	5.0	4.3	3.0	2.0	2.8	4.5	4.3	2.9	1.8	2.4	4.8	3.8	2.9	1.9	2.4
sd	1.6	1.7	1.4	1.0	1.4	1.6	1.3	1.5	1.3	1.3	1.7	1.4	1.4	1.0	1.4	1.7	1.5	1.5	1.0	1.2	1.6	1.8	1.4	1.1	1.4
gap	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
flush	1	1	1	1	1	2	2	2	2	2	3	3	3	3	3	4	4	4	4	4	5	5	5	5	5
parallelism	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

b)



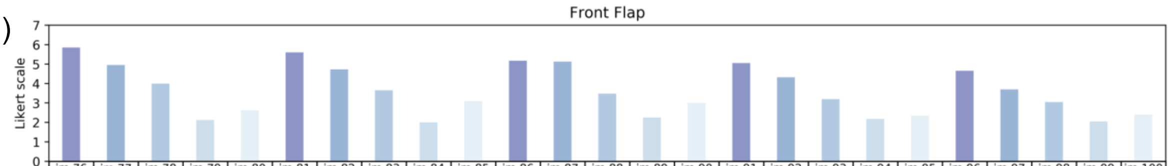
	im.26	im.27	im.28	im.29	im.30	im.31	im.32	im.33	im.34	im.35	im.36	im.37	im.38	im.39	im.40	im.41	im.42	im.43	im.44	im.45	im.46	im.47	im.48	im.49	im.50
mean	5.7	4.9	4.0	2.6	3.1	5.5	4.8	3.9	2.4	2.9	5.0	4.1	3.5	2.1	2.7	4.7	4.2	3.4	2.5	2.8	4.9	3.6	3.4	2.0	2.9
sd	1.9	1.5	1.3	1.6	1.2	1.9	1.7	1.7	1.6	1.2	1.8	1.8	1.6	1.1	1.2	2.0	1.5	1.5	1.6	1.2	1.7	1.8	1.4	1.2	1.6
gap	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
flush	1	1	1	1	1	2	2	2	2	2	3	3	3	3	3	4	4	4	4	4	5	5	5	5	5
parallelism	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2

c)



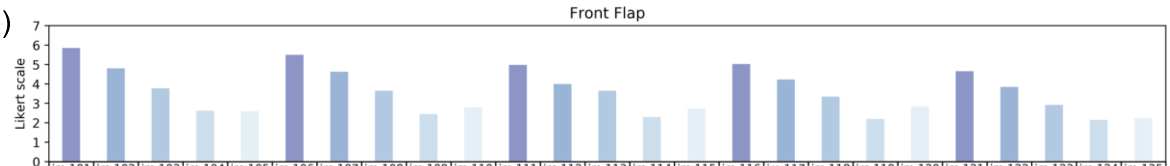
	im.51	im.52	im.53	im.54	im.55	im.56	im.57	im.58	im.59	im.60	im.61	im.62	im.63	im.64	im.65	im.66	im.67	im.68	im.69	im.70	im.71	im.72	im.73	im.74	im.75
mean	6.0	4.9	3.7	2.4	3.1	5.6	4.8	3.9	2.4	2.7	5.9	4.3	3.8	2.2	2.6	4.8	4.3	3.2	2.0	2.9	4.5	4.2	3.3	2.0	2.5
sd	1.6	1.5	1.6	1.5	1.4	1.5	1.5	1.6	1.4	1.5	1.7	1.6	1.6	1.0	1.1	1.9	1.5	1.3	1.0	1.5	1.9	1.7	1.4	1.3	1.4
gap	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
flush	1	1	1	1	1	2	2	2	2	2	3	3	3	3	3	4	4	4	4	4	5	5	5	5	5
parallelism	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3

d)



	im.76	im.77	im.78	im.79	im.80	im.81	im.82	im.83	im.84	im.85	im.86	im.87	im.88	im.89	im.90	im.91	im.92	im.93	im.94	im.95	im.96	im.97	im.98	im.99	im.100
mean	5.8	5.0	4.0	2.1	2.6	5.6	4.7	3.6	2.0	3.1	5.2	5.1	3.5	2.2	3.0	5.0	4.3	3.2	2.2	2.4	4.7	3.7	3.0	2.0	2.4
sd	1.7	1.6	1.6	1.2	1.4	1.7	1.6	1.5	1.0	1.4	1.8	1.5	1.6	1.3	1.3	1.6	1.5	1.5	1.3	1.3	1.8	1.7	1.2	1.2	1.2
gap	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
flush	1	1	1	1	1	2	2	2	2	2	3	3	3	3	3	4	4	4	4	4	5	5	5	5	5
parallelism	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4

e)



	im.101	im.102	im.103	im.104	im.105	im.106	im.107	im.108	im.109	im.110	im.111	im.112	im.113	im.114	im.115	im.116	im.117	im.118	im.119	im.120	im.121	im.122	im.123	im.124	im.125
mean	5.8	4.8	3.8	2.6	2.6	5.5	4.6	3.6	2.5	2.8	5.0	4.0	3.6	2.3	2.7	5.0	4.2	3.4	2.2	2.9	4.7	3.9	2.9	2.1	2.2
sd	1.9	1.9	1.6	1.5	1.3	1.9	1.8	1.3	1.3	1.5	1.9	1.7	1.4	1.4	1.3	1.8	1.8	1.7	1.1	1.5	1.9	1.7	1.3	1.1	1.1
gap	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
flush	1	1	1	1	1	2	2	2	2	2	3	3	3	3	3	4	4	4	4	4	5	5	5	5	5
parallelism	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5

Figure 57. Resulting split-line quality assessments of the area front flap.

8.4. SUMMARY AND CONTRIBUTION OF RESEARCH QUESTION 3

The results of the previous two chapters reflected the third research question, which considered effects on the assessment of perceived quality based on the global product context and relations between split-line attributes. Therefore, two studies were performed, which investigated both the global product context and relations between attributes. For both studies, three different automobile areas were analyzed: the fuel flap, the doors, and the front flap. As stimuli, the split-line attributes gap, flush, and parallelism were variated in three levels in the first study and five in the second study. The first study was a conjoint analysis, which investigated effects on perceived quality regarding the product context. Based on the study results, the three attributes were differently perceived (the first question of the third research question). At the fuel flap, the attribute parallelism was perceived dominantly. At the doors, the attribute flush was dominant and worse assessed. At the front flap, the third attribute, the gap, was dominantly perceived and assessed worse (the second question of the third research question). In the second study, the relations between attributes were investigated. The second study results confirmed partially the results of the conjoint study (the sixth question of the third research question). In the fuel flap area, the attribute parallelism and the combination of the attributes parallelism and flush were poorly assessed. The attribute flush had bad results within the levels medium, high, and drastic at the doors. The attribute gap was dominantly perceived at the front flap and demonstrated by levels with medium, high, and drastic bad results in assessing the perceived quality level (the fifth question of the third research question).

8.4.1. GLOBAL PRODUCT CONTEXT ON SPLIT-LINE ATTRIBUTES

A central finding of the conjoint analysis was the varying prioritization of split-line tolerance attributes according to the importance ranking of customers, which resulted from the first question of research question three. The three considered attributes, gap, flush, and parallelism, achieved different importance values for each perspective and automobile area. An explanation for the different attribute prioritization regarding the area is the global product context. The global product context affects the appearance of split-lines, and therefore, different split-line attributes were important in different automobile areas (question three of the research question three). For example, design attributes such as the chromium-plated border influence the perception of tolerance attributes at the area of the doors, and the split-line attribute flush was highlighted. Additionally, the superordinate shape of particular split-lines was essential for product appearance and assessment. Following an example of the chromium-plated border on the doors: the chromium-plated highlighted the transition of the doors in z-dimension. When both doors were not optimally positioned and show a misalignment, this quality imperfection is directly and visually readable by the narrow line. Additionally, this imperfection is highlighted by the chromed border, which reinforces the poor visual quality impression.

Another effect was the shape of split-lines, which were three-dimensionally designed and affected the global product design. The three-dimensional design of split-lines was included with the perspectives of fuel flap and front flap. The split line shape at the fuel flap was oval, its three-dimensional orientation within the product shape was mostly in the z-axis of the automobile, and the side lengths of the split-line were relatively short. The short side length of this particular split-line enabled a comparison of distances between two positions on the split-line. In conclusion and provided by the comparison, the split-line attribute parallelism was perceived dominant. Therefore, participants ranked parallelism with a high importance value. Besides, the split-line at the front flap had a complex shape, including a three-dimensional turn. This split-line was long and oriented in x-direction. Additionally, front light as a design characteristic influenced the global product structure. The long shape of the split-line complicated the assessment of the attribute parallelism. Moreover, the selected perspective from the top impeded the assessment of the flush within the z-dimension. In conclusion, the attribute gap was highly rated due to the visual conspicuity.

8.4.2. Perception of Split-line Attributes

With a focus on the influence of split-line attributes on perceived quality, the discussion is organized by the individual areas and followed by comparing the conjoint study results. The areas fuel flap, doors, and front flap are analyzed by the following four categories (question seven of research question four):

(1) Dominant attribute,
(2) conspicuousness of attribute levels
(3) Effects of attribute combinations
(4) Pattern in the result diagrams.

Fuel flap

The fuel flap area results demonstrated a decrease in the quality assessments within the high and drastic levels of the attribute parallelism. The results indicated a high-quality perception within the levels perfect and slight of the attributes parallelism and flush. By having the levels high and drastic of the attribute flush, the quality perception declined. Indeed, the most intense effect on quality perception has the attribute parallelism. A dominant character was identified in the results within the levels high and drastic of the attribute parallelism. Dealing with the high and drastic levels of parallelism, the attribute levels of gap and flush had almost no effects on the quality perception, and the mean values were between 3.3 and 1.7. Therefore, the dominant attribute within the area fuel flap was identified as parallelism. The level medium of parallelism demonstrated first deteriorations in the results and worse quality perception. This attribute is intensively perceived within the levels high and drastic. Thus, the attribute parallelism is identified as a dominant attribute with a strong influence on perceived quality.

The effects of the attribute gap demonstrated conspicuousness. The attribute gap was often assessed better within the level slightly than within level perfect. Therefore, an assumption is that a slightly wider tolerance of the attribute gap is perceived in higher quality than a split-line with a narrow execution of the attribute gap. An explanation of this better assessment of a wider gap is the related opening function of the fuel flap. A tight execution of the function could assume an impairment of this function. Based on the results, the tolerance of the attribute gap can be expanded in the area of the fuel flap.

The effects of attribute combinations were split into two cases. Firstly, the attribute combinations of gap and flush by a constant perfect to medium attribute level of parallelism. Secondly, combining all three attributes by having a constant high to drastic attribute level of parallelism. Within the first consideration, the effects of the attribute gap, combined with flush, demonstrated a decrease in quality perception. The decrease was continuous, whereby the effects of the attribute flush were more effective. The second consideration demonstrated dominant visual effects of the attribute parallelism in the levels high and drastic.

The constant perfect to medium levels of the attribute parallelism was similarly perceived at the fuel flap. Additionally, the high to drastic levels of the attribute parallelism demonstrated a similar pattern.

Within the conjoint analysis, the attribute parallelism was identified with the highest importance. This result is similar to the derivation of the dominant attribute in the perceptual study. Similarly, the attribute gap was identified within the conjoint analysis as the third essential feature. A result of the perceptual study was the higher quality perception with a wider split-line attribute gap. In summary, the results of the conjoint analysis were confirmed by the perceptual study.

Based on the discussion following design guidelines were derived at the fuel flap (question eight of research question three). The attribute parallelism demonstrated dominant and high influence on visual perception as of a drastic level. Therefore, this attribute is sensible, and the tolerance of this attribute needs to be strict. Otherwise, the attribute gap was perceived worse at the perfect level than at the slight level. Thus, the tolerance range of this attribute can be expanded.

Doors

At the area of the doors, the attribute flush demonstrated a dominant visual influence on the perceived quality level. Compared to the single attribute influences, the attribute flush demonstrated the most comprehensive assessment range between 5.9 and 2.2. An explanation for the intense perception and worse assessment of the attribute levels is the global product context. At the areas of the doors, the chromium-plated border highlighted the side windows. A flush of the doors was also transferred to the chrome border, and the flush was, therefore, more prominent. Besides the visual influence of the single attribute flush, the combination of flush and parallelism demonstrated worse assessment results. Therefore, a worse execution of the attributes flush and parallelism is additionally dominant perceived.

Conspicuousness was identified for the attribute gap. Within several attribute combinations, e.g., image 77, 78, 87, a more robust distinct attribute level of gap was assessed better, and therefore, a more exhaustive execution of the attribute gap at the area doors was perceived with higher quality impression. This assessment of the attribute gap occurred frequently. This conspicuousness is explainable by the function of the doors, similar to the function of the fuel flap. A too narrow execution is perceived that the function of opening can no longer be fulfilled.

Mainly, the combinations at the area doors were assessed worse. The combinations of the attribute flush and parallelism and the flush and gap combinations demonstrated a quality perception decrease. Otherwise, within a high level of the attribute flush, the quality perception was worse assessed, but both other attributes demonstrated almost no other effects.

The results of the conjoint analyses presented flush as an essential split-line attribute. Within the perceptual study, this attribute was identified as being dominant perceived. However, a more specified result was the worse perception of flush combinations with gap and parallelism. Both the perceptual study results and the results of the conjoint study indicate a better assessment of a wide gap. In the conjoint study, this better assessment can be seen from the higher utility value of the medium gap level (utility value = 0.30, cf. Figure 51) than the utility value of the perfect gap level (utility value = 0.10, cf. Figure 51).

In conclusion, the design of the split-line attributes at the area of the doors to realize a high level of perceived quality were the following ones (question eight of research question three). The attribute flush demonstrated a dominant visual effect, especially within the higher levels. Therefore, the split-line concept should strictly limit this attribute. Additionally, the combination of flush with gap or parallelism was perceived with low visual quality. These combinations need to be precisely limited. The attribute gap demonstrated a higher quality assessment within the levels of slight and medium. The perfect execution was assessed worse than the levels slight a medium. Therefore, this attribute limits can be widened in the split-line concept.

Front flap

Within the area front flap, the single attribute gap was assessed with wide range between 5.8 and 2.0. The other attribute assessment ranges were between 5.8 and 4.8 for flush and 6.0 and

5.8 for parallelism. The assessment differences in the levels perfect to drastic of the attributes flush and parallelism were minimal and barely noticeable. Therefore, the attribute gap range was broad and indicated a dominant perception of the attribute gap. The best assessment result was demonstrated within image 51, which included a medium level of the attribute parallelism. This result suggests that the effects of parallelism are minimal. Additionally, the attributes flush and parallelism demonstrated in combination with the attribute gap no effects. The assessment results were almost identical between the perfect and drastic levels of the attributes flush and parallelism.

The attribute gap was rated better at the drastic level than at the high level. This assessment result occurred in all attribute combinations independent of the levels. The function cannot explain the conspicuous feature as in the other two areas. Although the front flap also has an opening function, this conspicuousness occurred at both extreme levels high and drastic. Therefore, one possibility would be that the drastic execution of the attribute is regarded as necessary. Since the variation was correspondingly large, the observer assumes that such a design is necessary.

The diagrams demonstrated a highly similar assessment pattern. The first five diagram bars were repeated in the different combinations and levels. The first five diagram bars represented the five levels of the attribute gap with perfect levels of the attributes flush and parallelism. The repletion of the five bars was independently noticeable of the other attribute levels and is, therefore, another indicator for the dominant visual perception of the attribute gap.

Similarly, the conjoint analysis results presented the attribute gap at the front flap as the essential attribute. Within the perceptual study, the dominance of the gap attribute is explicitly noticeable. Within the conjoint study, the attribute gap was assessed by an importance value of 53%, which was the highest importance value of all areas and attributes. Therefore, the attribute gap was assessed as an essential attribute.

Consequently, the front flap area is highly influenced by the attribute gap (question eight of research question three). Therefore, this attribute has to be strictly dimensioned and limited within the split-line concept. A recommendation is a limiting of this attribute between the perfect to the medium level. The levels high and drastic should be excluded related to the low-quality assessments.

The results indicated a different dominant attribute per area. Additionally, the combinations of attributes demonstrated the worst assessment results, primarily in the area of the doors. These findings stand in contrast to present design approaches. Therefore, the results were presented in design guidelines, which are summarized in the following Figure 58 (question eight of research question three).

_ Parallelism is dominant perceived
_ Tolerance limits of the attribute gap can be widened
_ Flush demonstrates a linear perception, except the high and drastic levels of parallelism

_ Flush and combinations with flush were dominant perceived
_ Tolerance limits of the attribute gap can be widened
_ independent of the other attributes, flush demonstrates worse quality assessments in drastic level

_ Gap is dominant perceived
_ Tolerance limits of the attribute gap can be widened
_ Other attributes demonstrate less further effects on quality assessment

Figure 58. Summary of important guidelines for designing and tolerating split-line attributes per considered area.

The conducted study presented detailed information regarding the perception and quality assessment of split-lines. Within the different areas, different attributes were perceived dominant. Furthermore, the consideration of attribute combinations indicated their importance and effects on perceived quality. Therefore, a recommendation is the consideration of attribute combinations within the split-line concept. The complexity and the multitude of possible combinations are problematic. Consideration of a complex assessment by applying the classic split-line concept turns out to be extremely difficult. Additionally, the subjective assessment within the study is recognizable in the standard derivation. The standard derivation includes information on the dispersion of the results. Therefore, a small standard derivation indicates a homogenous assessment, whereas a high standard derivation signifies diffuse assessments. A high standard derivation is an indicator of subjective perception. Within the results, the higher levels of the attributes often demonstrated a higher standard derivation. This higher standard derivation directed a disagreement on the results. Therefore, the development of a standardized and objective assessment tool is recommended.

9. OBJECTIVE PERCEIVED QUALITY METHOD

The use of neural networks in the industry for quality assessment was numerous (cf. chapter 2.6). The presented methods were based in different industrial sectors, from textile seam production to the food and wood industry. Accordingly, these methods presented approaches to classify quality errors and to increase the process by, for example, faster detection times. Thus, a differentiation of these methods and the presented method is the level of complexity. The assessment of perceived quality and split-line quality is fuzzier than the current assessment tasks, i.e., detecting wood errors by Packianather and Drake (2000).

Therefore, the fourth research question demands an objective assessment method for visual split-line quality. Based on perceived quality, the focus has to be an assessment method, which considers the entire product. The focus on the entirety of a product was an issue in the previous research hypotheses, and these results form the basis of the subsequent investigation. The fundamental assessment approach is a machine learning method. Therefore, this research section considers an entire process for developing a machine learning-based assessment approach. The development is supported questions:

(1) How can a general machine learning approach for assessing visual quality features be defined?
(2) How can the product complexity concerning human-related, product-related parameters, and relations between split-line attributes be considered in a method?
(3) How can the results be evaluated due to the not observable algorithm of the machine learning method?
(4) How flexible is the trained machine learning approach due to the database quality?

The fourth research question was considered by developing a general approach. In the second step, the general approach was applied in the split-line assessment process. Data generation, data labeling, network architecture selection, implementation, and training were performed based on the developed approach. Additionally, the machine learning algorithm was evaluated using different explainable AI methods and was applied to a database with different image quality.

9.1. METHOD

Various industry and research methods for the deduction of a machine learning-based assessment method are presented in section 2.6. These methods presented approaches for assessing quality in different industry sections. Most of these methods identified error patterns based on the occurrence of defects. This classification differs from the assessment of split-lines. The perceived visual quality of split-lines has a more fuzzy and complex classification factor. Due to the subjective perception of individuals, the classification of split-line quality can be inconclusive. Therefore, a new approach was developed. The new method's specification was the inclusion of the global product context, influences of split-line attributes, human- and product-related parameters needed. These requirements resulted in a high level of complexity, where rules and relations were unknown. Therefore, the application of machine learning methods was identified as a strategy for developing an assessment method. Figure 59 shows the developed approach.

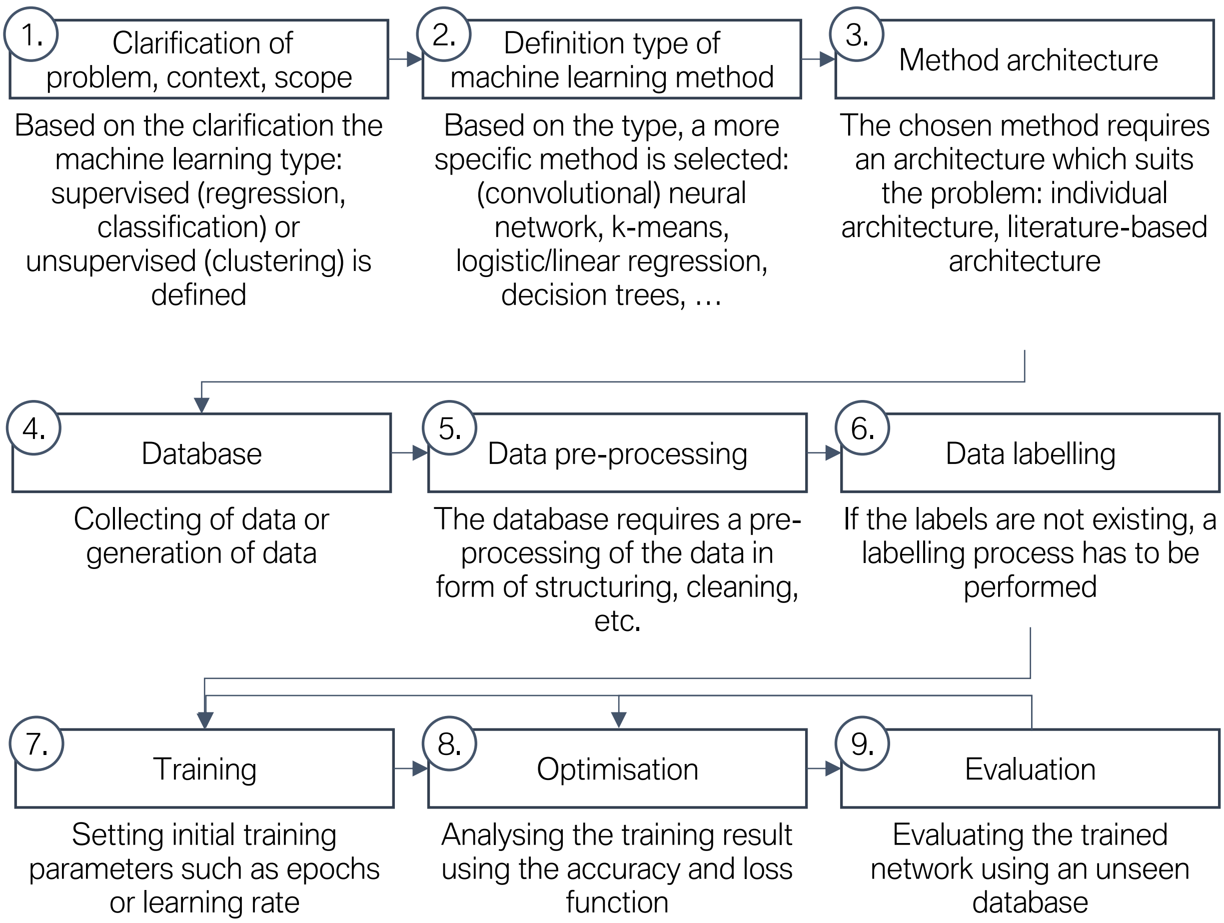


Figure 59. Derived method to apply machine learning methods for visual quality assessment.

The developed method is divided into three sub-processes: the definition of a machine learning method and architecture, the data preparation, and the training of a model. The first step of the first sub-process considers the analysis and the clarification of the problem, which should be solved using a machine learning approach. This step is essential for selecting a machine learning type, whether supervised learning with the regression or classification or an unsupervised learning approach with clustering is suitable. Based on this selection, the type of machine learning method is defined in the second step. The kind of machine learning type already restricts the applicable method. Additionally, by using the method, the specification, whether a machine learning or a deep learning approach is used, is more precisely. Based on the chosen method, the particular architecture of the model can be defined in the third step. Based on the analyzed problem in step one, the particular method architecture can be available, a new architecture, or a combination or a modification of an existing one. With defining this third step, the second sub-process the data preparation starts. The first step of the second sub-process considers the database. The database is strongly influenced by the initial problem and the defined machine learning type, method, and architecture. For example, by using a convolutional neural network, the database consists typically of images. Within the industry, a database can often not be available, or the database is disordered. If the database is not available, the generation or reinforcement of a dataset must be performed using synthetical data. For example, synthetical data can be photorealistic visualizations generated by using the approach presented in section 6.2. If a database is partially available or the database is not sufficient, synthetical data can reinforce the base. The following step two of the second sub-process considers a pre-processing of the data. Thus, the database can be disordered, or incorrect or unimportant data can be included; the database needs to be structured and cleaned. An advantage of using synthetical data can be reduced conduction of this step. The third step of the data preparation sub-process is the labeling of the data. The labeling step is essential and directly affects the training results. A lousy quality database or a badly labeled database can hardly obtain a good training result and an excellently trained machine learning algorithm. Additionally, the labeling is dependent on the machine learning type and the chosen method. A clustering method does not need labeling because it is an unsupervised approach. The supervised learning methods regression and classification require labeling, but the label differs between both methods. Whereas a regression method needs a continuous variable, a classification method needs a discrete variable as a label. Subsequently to this step, the second sub-process is completed, and the sub-process of training the model is performed. The first step of the third sub-process is to define the training. Therefore, the database needs to be divided into three bundles: training set, test set, and validation set. The training and test set is used for the training epochs, and the validation set is used for the evaluation of the trained model.

Additionally, training parameters, such as the number of epochs or learning rates, are initially defined. The second step considers the training output and analyses the accuracy, the loss function, and progress within the epochs. Based on this analysis, the training parameters are redefined, and new training is performed. This loop is executed until the training results fit the expectations, or no further improvement is achieved. The trained model is evaluated within the third step of the third sub-process. The evaluation is performed by using the validation set and by using explainable AI methods. By using the validation set, the confusion matrix is derived. This matrix displays the true labels towards the predicted labels of the network. With the aid of this matrix, the prediction accuracy can be derived and analyzed. The explainable AI methods can visualize the essential characteristics for a prediction within the data.

Approach

The classification of the visual split-line quality by using a machine learning approach is procedure by using the presented method. Therefore, the three sub-processes with the nine steps are presented in the following sections.

9.1.1. Clarification of Machine learning Type and Definition of Method

The problem definition in this thesis was a visual assessment of split-line quality. The assessment context was split-lines within the product context. The chosen machine learning method was supervised learning with a classification method. The supervised classification task was binary and differentiated the split-line quality in high and low visual quality. Based on the complex rules and relations of the visual assessment of split-lines, a convolutional neural network was chosen as a machine learning approach. The advantage of choosing a deep learning method was the automatic extraction of features by the neural network. For analyzing and implementing a convolutional neural network, an investigation area of the automobile was identified. For continuous consideration, the BMW 5 series sedan was investigated. Based on the deduction of the human-related parameters (Striegel and Zielinski 2018), the most frequently assessed area was the a-pillar at the driver and co-driver side. Based on the high assessment interest of the a-pillar, this area was chosen as an assessment example for the convolutional neural network. By choosing the a-pillar, a specialty within the assessment occurred. A finding of the empirical study for analyzing human-related parameters was assessing the a-pillar by using two different viewpoints (cf. Figure 35). The first viewpoint was at the side of the automobile, and the second viewpoint was at the front. Both viewpoints had a distance of 0.8 m and a height of 1.6 m, derived by human-related parameters (Striegel and Zielinski 2018), and were chosen as camera settings. The perspective was a small angle and focused on details related to the local assessment scope of the a-pillar. Additionally, based on the investigation of product-related parameters (section 6.4), the parameters color and reflection were defined by grey and metallic. Since the assessment of the split-line included further design details, the metallic grey parameter was chosen. Examples of the resulting images are presented in Figure 60.

Side viewpoint Front viewpoint

Figure 60. Resulting perspectives for the deep learning-based assessment of visual split-line quality at the a-pillar.

9.1.2. Network Architecture

The training of the neural network was a supervised approach with binary image classification. The basic architecture of the convolutional neural network was a VGG-16. The architecture was developed by Simonyan and Zisserman (2014). The major development within this network architecture was small convolutional filters, or receptive fields, which realize 3×3 filters. By using such small receptive fields in succession without having a subsampling layer, two main advantages arose. For example, to realize a 7×7 receptive field, three 3×3 filters in a row had to be executed. As a result, three non-linear layers instead of one non-linear layer were formed, which results in more representations of the input. Secondly, the number of parameters by implementing a 3×3filter was decreased related to a 7×7 filter. The concept of small receptive fields was realized by various numbers of weighted layers, having 11, 13, 16, and 19 (Simonyan and Zisserman 2014). Figure 61 presents VGG-16 network architecture, which demonstrates 16 convolutional layers.

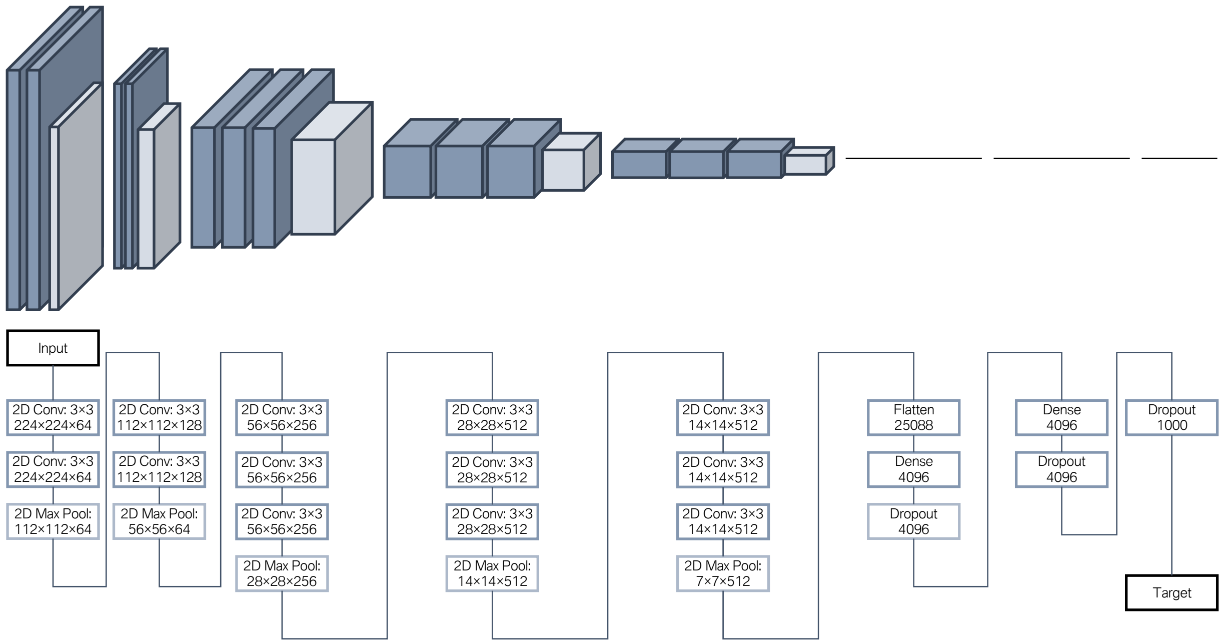


Figure 61. Architecture of a VGG-19 network (Simonyan and Zisserman 2014).

The input is an image with 224×224 pixels in three channels, having RGB. This image was firstly proceeded by a convolutional block. This block had two convolutional layers with 3×3 receptive fields and 64 channels and a pooling layer using maxpool. The second convolutional block had the same layers with the same receptive fields but with 128 channels. The third convolutional block had three convolutional layers with 3×3 receptive fields and a maxpool layer, with each layer having 256 channels. The fourth and fifth convolutional block had similar layers and receptive fields like the third block but with 512 channels. By concluding the fifth convolutional block, a flatten layer resulted.

Since two viewpoints were relevant for the quality assessment at the a-pillar (cf. Figure 60), the VGG-16 network was extended to two input branches. For this extension, the concept of Siamese networks was adapted. Thus, the network architecture had two input branches, which were converged at the end for an overall assessment. Each input branch had an image with a side view or a front view. The aim was to achieve a robust assessment by extending the system to two strands. The resulting architecture is presented in Figure 62.

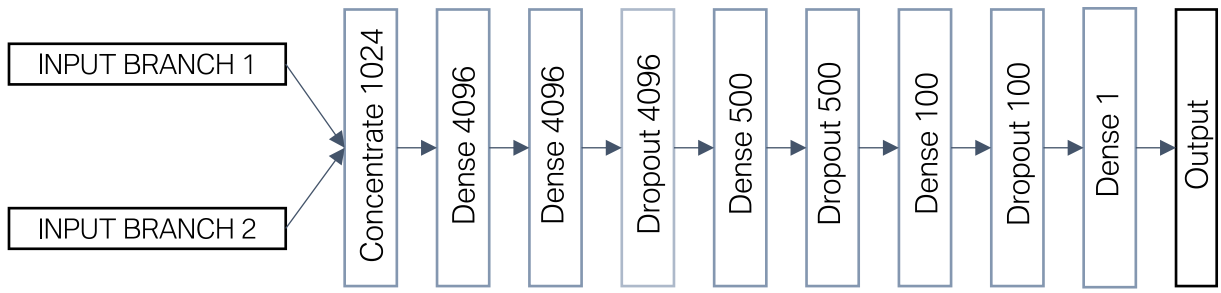


Figure 62. Developed network architecture for visual quality assessment of split-lines.

Due to the successful application of the ReLU activation function within similar problems, this activation function was chosen for the hidden layers. The Siamese branches were concentrated after their 512 channels flatten layer in a concentration layer of 1024 channels. The concentration layer was followed by two dense layers and a dropout layer. Afterward, a dense and dropout layer was repeated two times. By using a dense layer with one channel, the final output was calculated. Since the network was implemented to predict two classes, the output layer was realized by a sigmoid function. As optimizer, "Adam" with a binary crossentropy loss function was chosen. For avoiding overfitting three dropout layers were manually implemented (Hinton, Sutskever, and Salakhutdinov 2014) in conclusion of the dense layers.

9.1.3. Database and Pre-Processing

The basis for training neural networks is large amounts of data. For implementing a supervised learning approach, neural networks require much data in the form of an input and a corresponding output in the form of a label. Thereby, the input data format can differ between single values, vectors, images, or videos.

The database, including the images and labels, needed to be generated within the thesis. Therefore, two strategies for generating a database existed: (1) using photos of automobiles or (2) using virtual generated images based on renderings or ray tracings. Using photos of automobiles was complicated due to three aspects. Firstly, the data acquisition took a long time related to the number of images and the long production period of an automobile. Secondly, the pictured split-line quality was difficult to reproduce. The split-line quality was presented by the combination of split-line attributes gap, flush, and parallelism. For presenting different split-line qualities, the single opportunity was to manipulate the components of automobiles, which was effortful and cost intense. Thirdly, the feature extraction, on which the neural networks are based, was very complicated and sensitive to variables. Disturbing variables, such as example reflection or the environment, were difficult to control by using photos. The disturbing variables should first be controlled and not variated to focus the neural network on the split-line attributes. The approach of using virtual generated images included the advantages of manageable image generation duration, flexible manipulation of components, and, therefore, quality and reasonable control of disrupting factors. Therefore, the database was generated by using photorealistic visualizations. The generation of these visualizations was based on the method presented in chapter 6.1.1. The first two sub-processes were identical to the presented method. The data acquisition and conversion and the data preparation were performed by using the presented steps. The third sub-process of the data generation method was modified. The camera settings were defined due to Striegel and Zielinski (2018), which were a distance of 0.8 m and a height of 1.6 m.

For generating various split-line situations, the attributes gap, flush and parallelism were automatically and randomly variated. Therefore, the components were translationally and rotationally manipulated. Figure 64 illustrates the components which effected the split-line attributes.

Figure 64. Component manipulation regarding translation and rotation for defining and generating the deep learning database.

The considered components for the manipulation were the front flap, the side panel, the side frame, and the front door. Therefore, three split-lines were influenced by translation and rotation of the four components. In doing so, the split-line attributes gap, flush, and parallelism were defined. The manipulation of the components was randomly and automatically created. Therefore, the Blender integrated node function was used for a script-based automatization (Blender Foundation 2019a). Each component manipulation represented a split-line situation, with different tolerance derivations and combinations of the split-line attributes. For every split-line situation, a specific and unique identification number was generated. The identification number, the translation, and rotation manipulation information were saved in a CSV file to identify and reproduce the images in a later stage. For each split-line situation, two different visualization qualities were generated. The first quality was a full HD resolution with a photorealistic rendering. These visualizations were used for the labelling. The second quality type was images with a resolution of 224×224 pixels and a rendering quality. These images were used for network training. The differences between both qualities are illustrated in Figure 63.

Full HD resolution – labelling quality 224×224 resolution – training quality

Figure 63. Comparison of labelling and training visualisation quality.

On the left side are the full HD quality visualizations for subsequent labelling presented. For each split-line situation, both images were generated. On the right side, the training data quality is represented. The reduced resolution was implemented due to the used network architecture, which had a maximal resolution of input data of 224×224 pixels. The training data set was augmented and multiplied. In order to get the necessary amount of data, a multiplication was implemented in the data generation. Therefore, one split-line situation was visualized 100 times. With the 100 resulting visualizations, the camera position and the focus point varied. The camera position moved within 10 images from the side view to the front view. The camera moved down one centimeter after 10 visualizations and then again started from the side view and ended at the front view. The focus point of the camera was slightly varied between the individual visualizations to achieve further data augmentation. Accordingly, one split-line situation was visualized 102 times. Two images were rendered for the data labeling, and 100 visualizations were rendered for the network training. In summary, 1.001 different split-line situations were simulated and subsequently rendered. Consequently, the labeling database included 2.002

images, and the training database included 100.100 images. The process of data generation is presented in the following Figure 65.

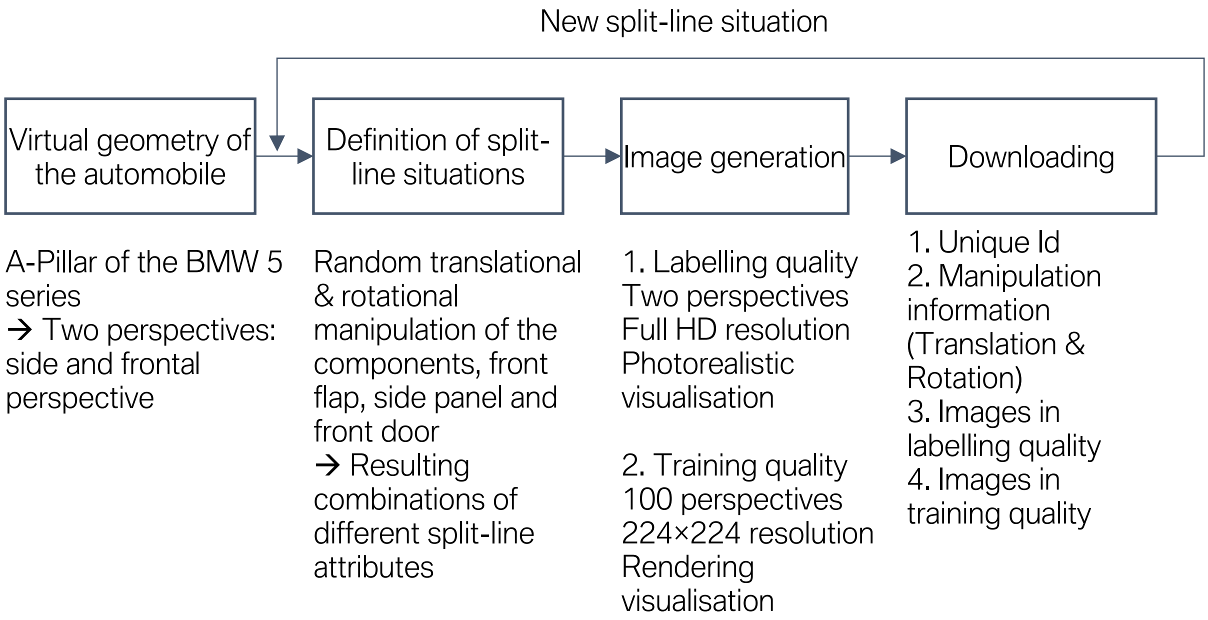


Figure 65. Derived method for an automatized data generation process.

Based on the synthetically generated dataset, a pre-processing in form of cleaning or sorting the dataset was not necessary. The pre-processing of the data was performed by dividing the dataset in three data bundles: training, test and validation.

9.1.4. Data Labelling

Subsequently to the generation of the dataset, the data labeling followed. Therefore, the 1.001 split-line situations, which included two images per situation, had to be assessed. For fast and easy labeling, an application was developed. The labeling application was designed for Mac OS, with a 15-inch and full HD resolution display. Within the application, the rendered images were randomly uploaded. The images were assessed using the arrow buttons. By using the space key, the perspective switched to the second viewpoint per split-line situation. The image order between the assessments was randomly variated to prevent sequence effects in the results. Additionally, between two split-line situations, a short interruption by a black screen was implemented. This short interruption prohibited the direct comparison of two split-line situations. The assessment of the participant represented a cumulated label for both perspectives of a split-line situation, which was automatically downloaded. Following Figure 66 illustrates the images within the application and the assessment.

Figure 66. Labelling application for the high and low visual quality levels.

The whole data set was assessed by five experts in the field of split-line quality assessments. These experts were supposed to reflect customer opinion. Based on their expertise, a more consistent labeling result was assumed. An essential aspect of the labeling was that the first impression of the split-line quality had to be assessed.

The deduction of the final labels was the mean value of the assessments of the five experts. For example, one situation was assessed in high visual quality three times, and in low visual quality two times, the resulting label was high visual quality. In conclusion, the data set included 281 split-line situations of high visual split-line quality and 720 situations of low visual split-line quality. Since a balanced data set was necessary for successful network training, this in-balance was corrected with 441 new split-line situations of high visual split-line quality. In conclusion, in 1442 different split-line situations with 100 visualizations per situation were simulated and rendered. The entire data set was 144.200 images large.

9.1.5. Training and Optimization

The training iterations were performed on a remote system running with Ubuntu 16.04.5 LTS (GNU/Linux 4.4.0-116-generic x86 64). The environment included Python version 3.6, and Keras (Chollet 2018) was running on top of TensorFlow (Abadi et al. 2016). As support for the coding, Jupyter notebooks (Kluyver et al. 2016) were used.

For all training iterations the data basis was split into three bundles: train, test and validation. The train dada set included 80% of the database, the test and the validation set respectively included 10% of the database. The data splitting in the three bundles is illustrated in Figure 67.

Data basis

Train 80%	Test 10%	Valid 10%

Figure 67. Data splitting ratio for training the neural network.

Additionally, a pre-processing of the train data was performed to enlarge the data basis and to remain a robust network. To not influence the form and quality of the split-lines, only the following three methods were used: zooming, flipping, shifting. Methods such as sheering were avoided.

Two strategies performed the training of the network. One strategy included the whole database with the above-illustrated data slitting (version 1). A second strategy was data cleaning. Therefore, the labels with triple labeling by the experts were soft and could confuse the network prediction. The second strategy was removing the triple labeled data to remain a more significant data set. In doing so, the data set was reduced to 460 split-line situations in high visual quality and 460 split-line situations in low visual quality. In summary, the whole database included 920 split-line situations and per situation 100 images (version 2). Thus, the database was reduced by 36,2 % related to the initial data set. Within the training of the deduced network architecture, both data set versions were separately used. Finally, the following data volumes resulted for each database version:

- Version 1: Train (80%) 1154 split-line situations → 115.400 images
  Test (10%) 144 split-line situations → 14.400 images
  Valid (10%) 144 split-line situations → 14.400 images
- Version 2: Train (80%) 736 split-line situations → 73.600 images
  Test (10%) 96 split-line situations → 9.600 images
  Valid (10%) 96 split-line situations → 9.600 images

Both database versions were trained by using following trainings approach illustrated in Figure 68.

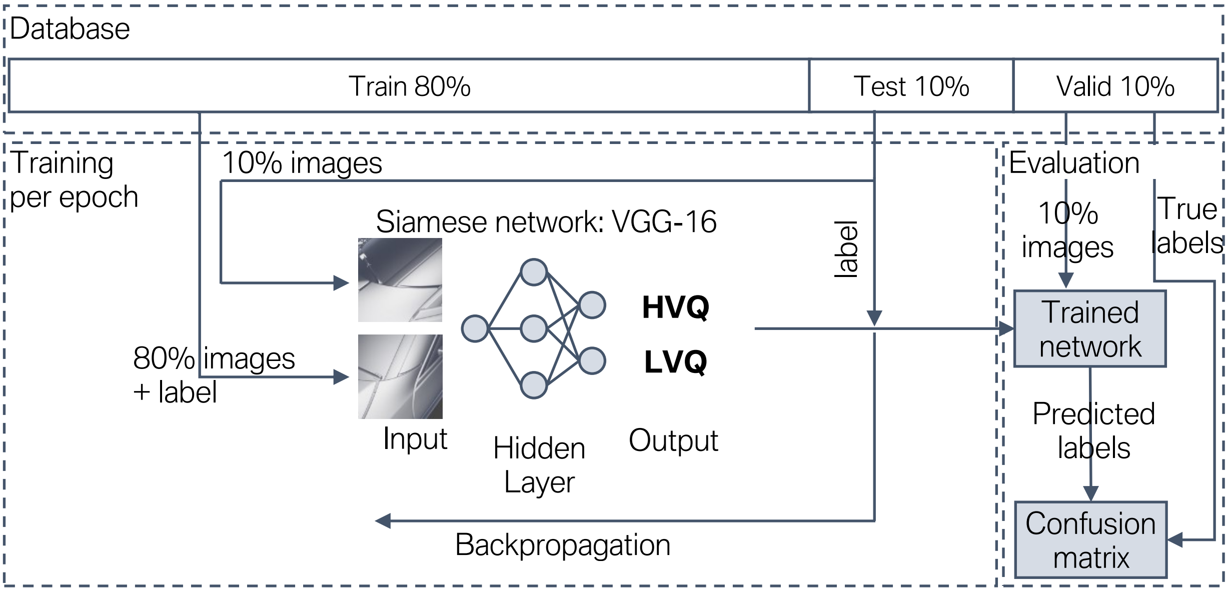


Figure 68. Visualised training and validation approach using the VGG-16 network with a Siamese network architecture for assessing the visual split-line quality.

The train dataset was used for the training loop. Within the training loop per each epoch the following three steps were executed (Chollet 2018):

(1) Run the network based on the generated images to obtain the predictions in form of labels.
(2) Calculate the loss function in form of the mismatch between the true and the predicted labels.
(3) Update all trainable parameters by using the gradient of the loss function (backpropagation).

The first step was performed to train the parameters by using the training dataset, including the labels. After the training, the resulting network performance was tested by using the test dataset. Therefore, the images were predicted to gain a label. This predicted label was compared to the true label in step two. For the comparison of both labels, a loss function was calculated. Based on the gradient of the loss function, the parameters changes were estimated and through a backpropagation in step three applied. This training loop was iterated per epoch. After the defined number of epochs, the training loop was interrupted. A learning rate of 0.00001 was chosen for the training. The result of the training loop was a trained network model. Within the evaluation loop, the performance of the trained network was computed by deriving the confusion matrix.

9.1.6. EVALUATION

The developed network architecture was trained by using two data set versions. Therefore, the results are presented for version 1 and subsequently for version 2. For verification of the trained network, three methods were applied: traditional variables, explainable AI, and the transferability of the trained network. Firstly, for each network, traditional variables were considered. These were training iterations and parameters, and the results, which were presented by the accuracy, the error function, and the confusion matrices. The accuracy and the model loss function illustrate the performance of a model during the training process at the end of each epoch. The confusion matrices of both networks were derived by new and unseen data included in the validation data set. Thereby, the new and unseen data quality complied with the training data quality. For the deduction of the confusion matrices, the trained neural network model was tested, regarding how well the feature extraction and classification worked with new data. The comparison of the predicted label and the true label was visualized in the form of a matrix. The structure of a confusion matrix is presented in Figure 69.

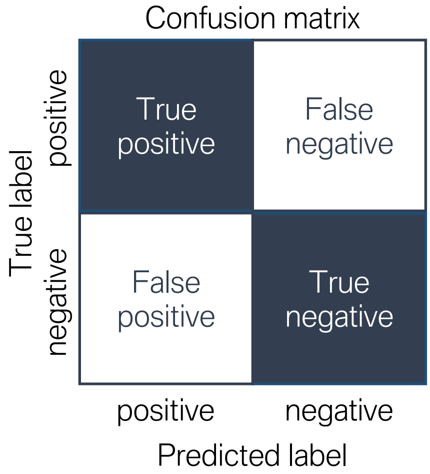


Figure 69. Concept of the confusion matrix for evaluating the trained model.

The confusion matrix structure was 2×2 categories: true positive, true negative, false positive, and false negative. The true positive predictions were those, which were correctly predicted as the true label. Within the study, the true positive predictions were those, which were initially and predicted by the network as low visual quality. The true negative predictions were the predictions that were correctly predicted as negative. Within the thesis, these predictions were the initial and predicted images with a high visual quality label. The false-positive category defined those labels which truly negative but predicted by the network as positive. Within the thesis, this category included truly high visual quality images but by the network wrongly classified as low visual quality. Finally, the fourth category was false negative predictions. Within the thesis, this category was formed by the images, which were truly low visual quality, but predicted by the network as high visual quality.

The second verification method was Explainable AI. These methods were presented in section 2.5.9. and represented methods for visualizing decision-relevant features in images. The approach of explainable AI methods is presented in Figure 20. Within this thesis, the trained networks were verified by using five of these methods. The five methods were Deep Taylor, Layer-wise Relevance Propagation, SmoothGrad, Integrated Gradient, and Guided Backpropagation. These five methods represented selecting the most popular methods from attribution-, signal- and function-based explainable AI approaches. These methods were applied

For both trained networks with random images from both classes. The approach to apply these methods are presented in Figure 70.

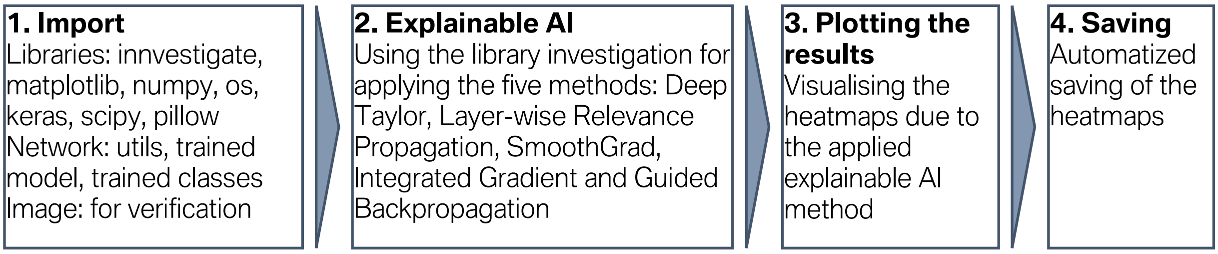


Figure 70. Method for verification the trained neural network by using explainable AI methods.

Within the first step, the necessary libraries, the network variables, and the image were imported. A necessary library was the innvestigated library, which includes the different explainable AI methods and was provided by (Alber et al. 2018). Additionally, the import included the trained network, the trained classes, and the image which needs to be analyzed. The second step was to apply the explainable AI methods to the trained network and image. The application was supported by the innvestigated library (Alber et al. 2018) and enabled an application by merely defining the method and the network. The resulting heatmaps were visualized and plotted within the third step, which was downloaded in the fourth step. For each trained network or each image, this method needed to be restarted. Based on the resulting heat maps of the five methods, the relevant features of a network for predicting the classes were identified. Based on these features, differences between the classes and the trained networks were recognized.

The third verification method, the trained network with database 2 was tested regarding its transferability to photorealistic images and real photos. The database for both images and photos were collected. A database for the photorealistic visualizations was the already generated labelling data set used. The database for real photos was collected within the plant. The photos were taken inside and outside the plant to verify different kinds of reflections. The transferability of the network was evaluated by comparing the predicted label and the real label. Besides, the explainable AI method Deep Taylor was applied to the trained network with the photorealistic visualizations and the real photos. Therefore, the method described in Figure 70 was applied by using as import image either the photorealistic visualization or the real photo, the trained network with database version 2, and the Deep Taylor explainable AI method.

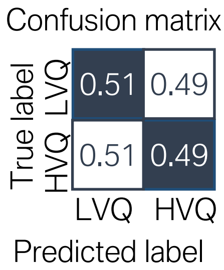
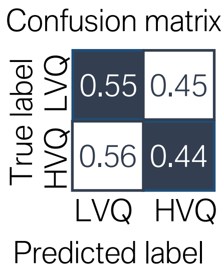
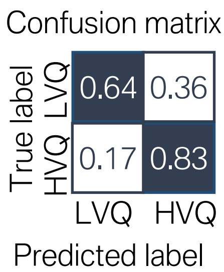
9.2. RESULTS

The results of the developed method for classifying the visual split-line quality are presented in this chapter. Therefore, the results are introduced by presenting traditional variables of database versions 1 and 2 and by presenting the heatmaps of the explainable AI methods. Additionally, the trained models were tested regarding their transferability to photorealistic visualizations and real photos.

9.2.1. RESULTS OF DATABASE VERSION 1

For each training iteration, the input image size was 224×224 pixels, and a triple channel for RGB. The batch size was 64 and the training was implemented by having check-files after each training epoch. The number of trained epochs was 50, 70, and 80. The following Table 16 presents for each training iteration the confusion matrix.

Table 16. Training results for database version 1 by variating the training epochs.

Image size, channels	Batch size	Epochs	Confusion matrix	Ranking
224×224, 3	64	50		3
224×224, 3	64	70		2
224×224, 3	64	80		1

The confusion matrix result of the training with 50 epochs demonstrated a prediction of 51 % for true positives and 49 % for the prediction of true negatives. This result of the confusion matrix indicated a poor performance of the trained network. The confusion matrix of the network with 70 epochs demonstrated a similar result of the confusion matrix. Related to the confusion matrix of 50 epochs, the true positives and the true negatives were approximately 50. In the third training iteration with 80 epochs, the results demonstrated a confusion matrix for the true positives and the true negatives above 50 %. Thus, this training iteration was further considered. In doing so, the accuracy and the loss function was analyzed. Figure 71 illustrated the training progression of the network with 80 epochs.

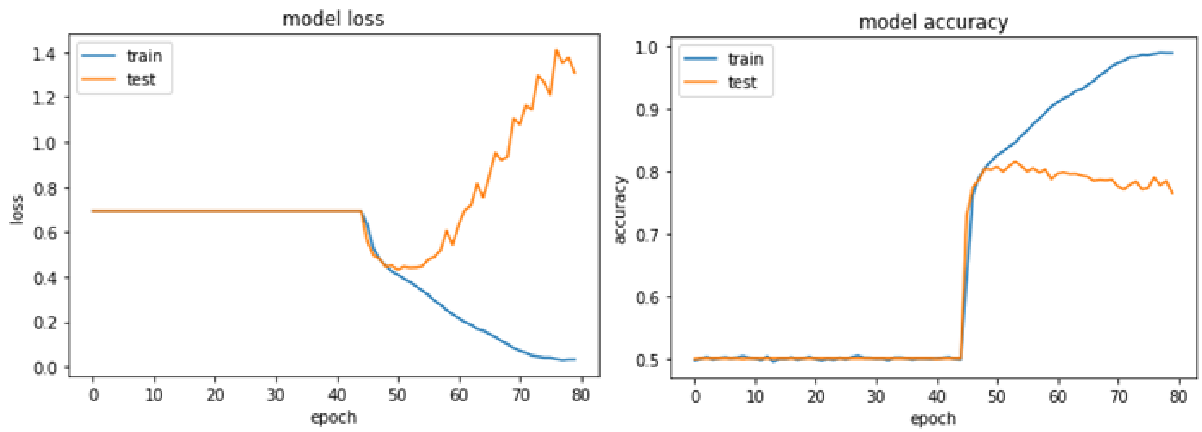


Figure 71. Model results of the training, test loss function and accuracy per epoch for database version 1.

The model loss function and the model accuracy demonstrated a tendency for overfitting as of epoch 53. Overfitting is indicted by a derivation of the train and test curve. Within both diagrams, the derivation is identifiable. By implementing the check-files the training results of the previous epochs were saved. Therefore, the trained model from epoch 53 was available. The accuracy, model loss function and confusion matrix for this iteration are presented in Table 17.

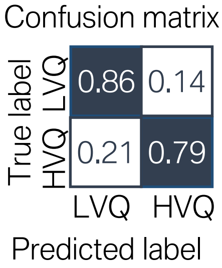
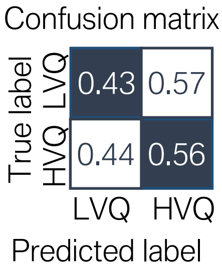
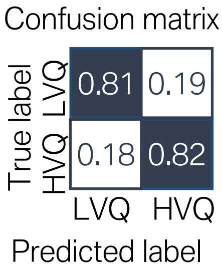
Table 17. Best trained model with database version 1.

Image size, channels	Batch size	Epochs	Confusion matrix.	Loss function, accuracy
224×224, 3	64	53	Confusion matrix True label LVQ: 0.81 0.19 True label HVQ: 0.22 0.78 LVQ HVQ Predicted label	0.51, 0.79

9.2.2. Results of Database Version 2

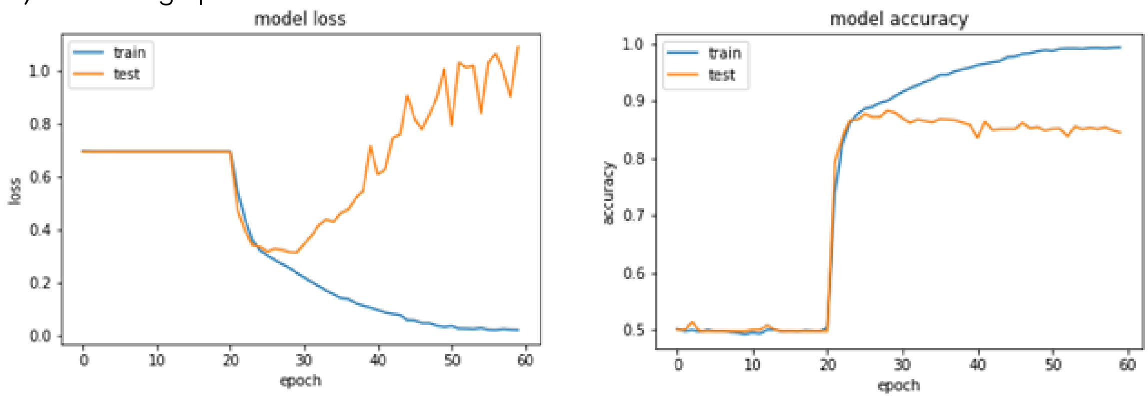
The training parameters of the second data set, which was defined as version 2, were similar to the version 1 data set: image size was 224×224 pixels, a triple channel for RGB, batch size was 64 and check-files after each training epoch. The training epochs were 30, 40, and 80 epochs. Table 18 presents the training results.

Table 18. Training results for database version 2 by variating the training epochs.

Image size, channels	Batch size	Epochs	Confusion matrix	Ranking
224×224, 3	64	30		1
224×224, 3	64	40		3
224×224, 3	64	60		1

The model was trained by using 40 epochs demonstrated a confusion matrix with true positive predictions of 43 % and true negative predictions of 56 %. On contrary, the training results of epoch 30 and 60 were very similar related to the confusion matrices. Both confusion matrices demonstrated a true positive prediction between 81 % and 86 % and a true negative prediction between 79 % and 82 %. Therefore, both trained models were further investigated. The detailed investigation of both training iterations included the training diagrams based on the model loss and accuracy. Both diagrams for both training iterations were presented in Figure 72.

a) 60 training epochs



b) 30 training epochs

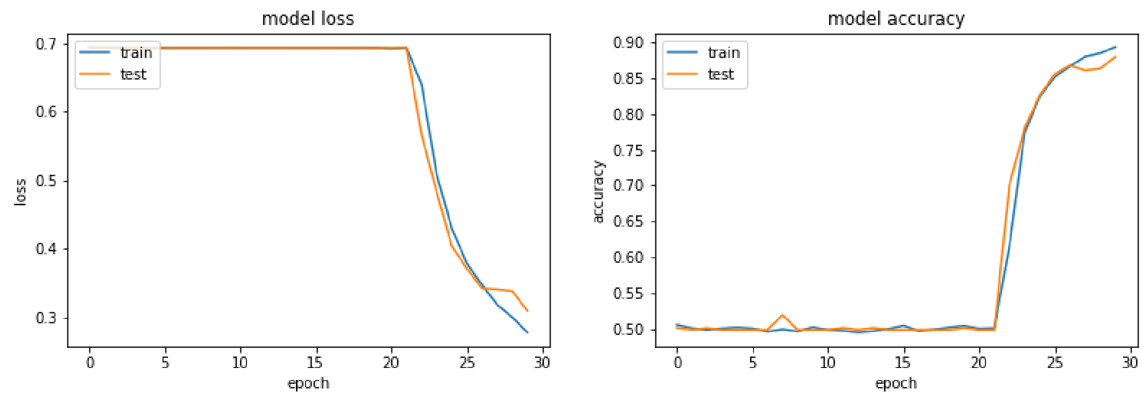
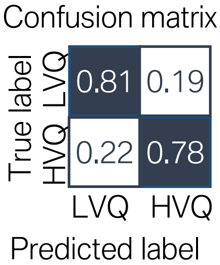
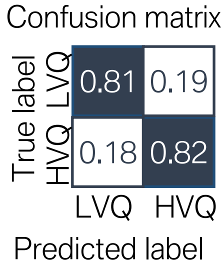


Figure 72. Model results of training, test loss function and accuracy per epoch for database version 2.

Figure 72 a) visualize the model loss, and the model accuracy of 60 epochs is presented. The model loss of 60 epochs was 1.32, and the accuracy was 0.82. The results present a high accuracy but with an additionally high loss function. Within successfully trained models, the loss function is low. Furthermore, the diagram shows a discrepancy between the train and test curves. Therefore, the overfitting of the model is recognizable. In summary, the trained model of 60 epochs demonstrated a good performance of the confusion matrix, but the loss function was high, and overfitting of the model was identified. Figure 72 b) shows the training results in model loss function and model accuracy with 30 epochs. The resulting value of loss function was 0.41 and is therefore low. Also, the accuracy was similar high as the accuracy of 60 epochs and reached a value of 0.82. The training and testing curves were close to each other in the diagram, which represents a well-fitted model. In summary, the model with 30 training epochs indicated a well-trained network, based on the results. Accuracy, loss function, diagram, and confusion matrix confirmed a well-trained model. Therefore, the model with the 30 training epochs was further considered.

After the results of the data sets of versions 1 and 2 were presented, a comparison of the best trained models for each version follows in Table 19.

Table 19. Comparison of best trained models of data set version 1 and 2.

Data set version 1, 53 epochs			Data set version 2, 30 epochs		
Accuracy	Loss function	Confusion matrix	Accuracy	Loss function	Confusion matrix
0.79	0.51		0.82	0.41	

The network with data set version 1 trained most successfully with 53 epochs. Thereby, the model accuracy was 0.79, and the model loss was 0.51. In contrast, the training results of the data set version 2 demonstrated a model accuracy of 0.82 and a model loss function of 0.41. Both model accuracy and loss confirmed a better training result for the data set of version 2. The model accuracy increased by 3 % points between version 1 to 2. Additionally, the loss function result decreased by 0.1 from the data set version 1 to 2. Moreover, the confusion matrices of both versions demonstrated a better-trained model of the data set version 2. The true positive values were 0.81 by training with the data set version 1 and 0.86 by training with data set version 2. Consequently, within the model of data set version 2, the prediction of the true positives was improved by 5 %. The true negatives were predicted with the trained model of the data set version 1 by 0.78. The model trained with the data set of version 2 predicted the true negative by 0.79, which signified an increase of 1%. In addition to the accuracy and model loss results, the confusion matrix demonstrated a better prediction of new data for the data set of version 2. In conclusion, the data set version 2 performed better training and resulted in a well-fitted model. In this regard, it should be noted that the data set of version 2 had 36.2 % less data than version 1. Nevertheless, the training result was better than that of version 1 in all four factors considered.

9.2.3. Explainable AI Methods

Within this section, the explainable AI methods were applied to verify both trained models. Therefore, images of the high visual quality and low visual quality class were analyzed by using different explainable AI methods. The applied explainable AI methods were Deep Taylor, Layer-wise Relevance Propagation, SmoothGrad, Integrated Gradient, and Guided Backpropagation. Due to the Siamese network architecture, two images each for one branch were tested by the five methods. The two images differ in the camera settings related to the camera distance, height, and focus point. The several explainable AI results are firstly presented regarding the data basis and the class. Secondly, the results of both trained models are compared and discussed. Figure 73 illustrates the applied explainable AI methods for the data basis version 1 and the class of high visual quality.

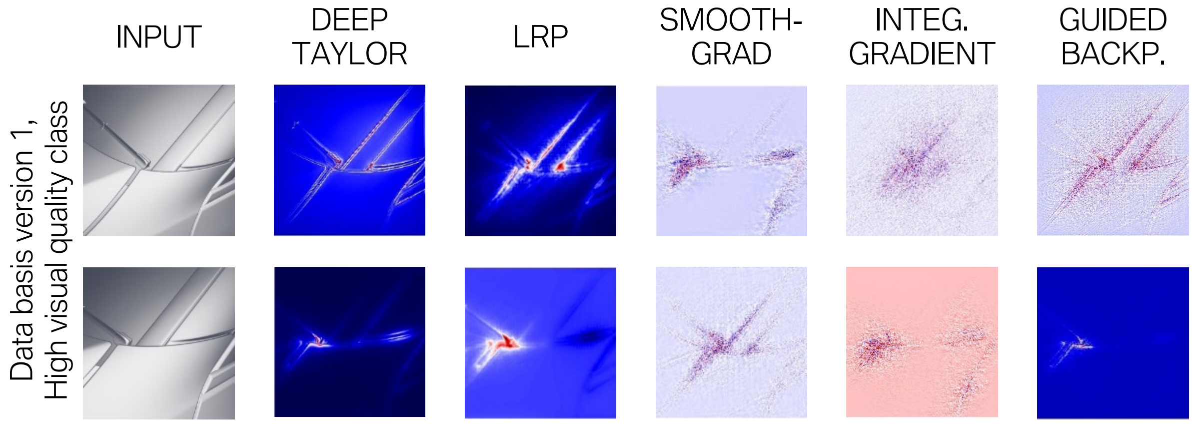


Figure 73. Visualized heatmaps of the Explainable AI methods for the high visual quality class of trained network with data set version 1.

The applied explainable AI methods demonstrate for the trained network of data set version 1, for the most part, the same highlighting of the relevant areas. Input image one, displayed on top, shows a similar highlighting of the split-lines at the side frame and side panel in the methods Deep Taylor, LRP, and Guided Backpropagation. Smoothgrad and Integrated Gradient methods highlighted the split-line of the front flap and the side panel. Compared to the first image, the second input image reveals a different heatmap of the area per applied method. This input highlights the same split-line between the side panel and front flap as a relevant area. In summary, the resulting heatmaps of the input branch one and input branch two differed. The methods highlighted per branch similar areas. The relevant areas were the split-line between the front flap and side panel and the split-line between the side frame to the side panel.

The second class of low visual quality for the trained network with the data split of version 1 is presented in Figure 74.

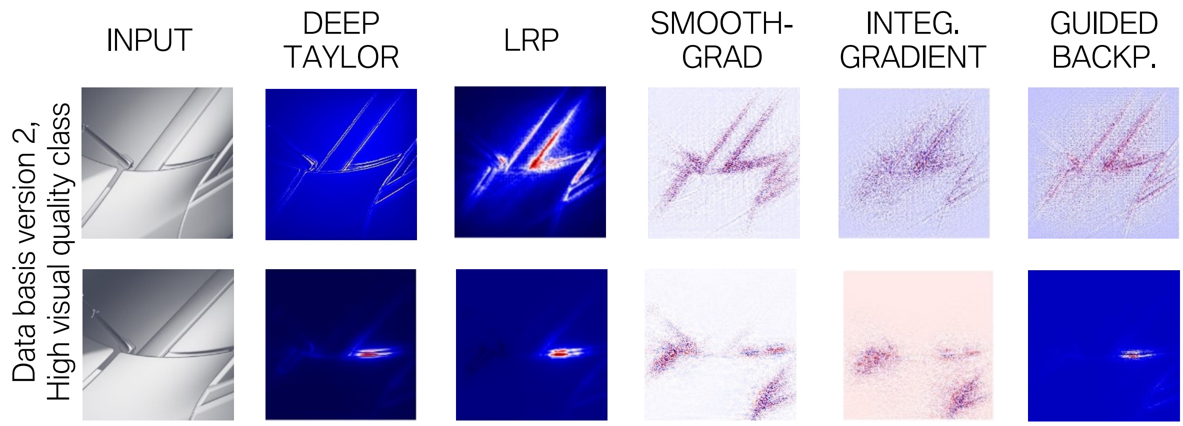


Figure 74. Visualized heatmaps of the Explainable AI methods for the low visual quality class of trained network with data set version 1.

Input branch of image one demonstrates similar heat maps for all five methods. The highlighted areas are the split-lines between the front flap and side panel, between the side frame and side panel, and between the side panel and front door. The resulting heatmaps of the second input branch of the low visual quality class demonstrate similar heatmaps of Deep Taylor, LRP, and guided backpropagation. The highlighted area within these heatmaps is limited on the split-line between the side panel and side frame. The heatmaps of the methods smoothgrad and integrated gradient illustrate highlights at the split-lines between the side panel and side frame and additionally, between the front door and side panel and between front flap and side panel. Compared to the explainable AI heatmaps of the high visual quality class, more areas were highlighted with a higher intensity. The visualizations of the different explainable AI methods indicate that the learned feature of the neural network between the high visual quality class and the low visual quality class was the area's intensity. This feature is also consistent with a classical split-line assessment. A strongly varying split-line pattern, characterized by high levels of the attributes gap flush and parallelism, reflects the low visual quality.

Subsequently, the trained network with the data basis version 2 was investigated by the five explainable AL methods. Therefore, a representative image of both high and low visual quality categories was tested in combination with the trained model. The resulting heatmaps of the explainable AI methods for the data basis version 2 and the first class, the high visual quality, are presented in Figure 75.

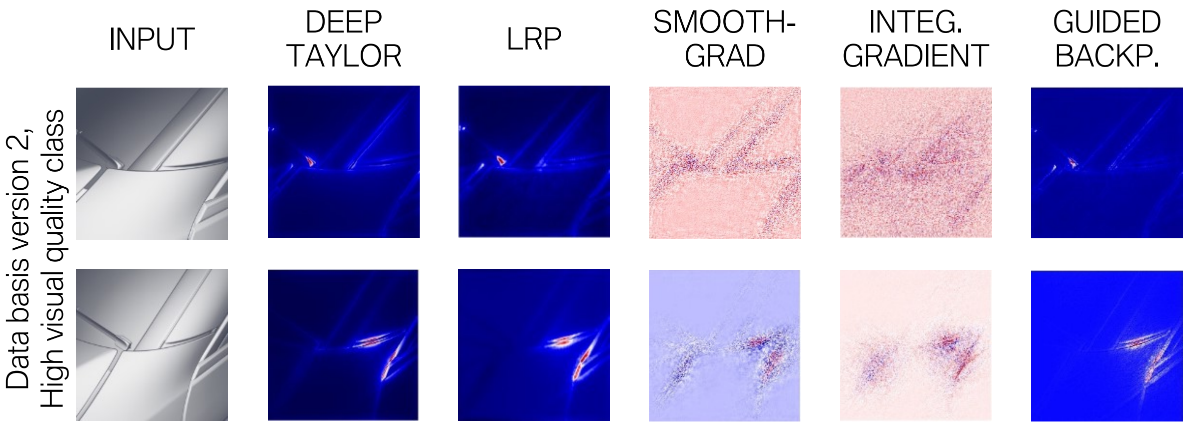


Figure 75. Visualized heatmaps of the Explainable AI methods for the high visual quality class of trained network with data set version 2.

The resulting heatmaps of the first input branch demonstrate a similar impression for all applied explainable AI methods. The highlighted areas were the split-lines at the side frame and between the side panel and the front door. The windscreen wiper is slightly highlighted within the heatmaps of the methods Deep Taylor, LRP, and guided backpropagation. Thus, the relevance for the network contrasts with the actual learning objective of the split-line assessment. Since this area is not exclusively marked, the relevance is to be seen in combination with the other highlighted areas. The heatmaps of the second input branch display similar prediction-relevant areas: These areas were the split-lines between the side frame and the side panel and between the front door and the side panel. Additionally, in the heatmaps of the methods smoothgrad and integrated gradient, the split-line between the front flap and the side panel was highlighted in combination with the previous ones. The windscreen wiper was not highlighted in the heat maps of the five applied methods in this input branch. Therefore, the second input branch did not extract the windscreen wiper as a relevant feature. In summary, the heatmaps of the input branches one and two were slightly different. In the first input branch, the side frame split-lines and the split-line between the front door and the side panel were relevant. Whereas, the heatmaps of the second input branch illustrated the split-lines of the side panel as relevant features for the high visual quality class.

In a fourth step, the trained model of data basis version 2 and the second class of low visual quality were examined using the five explainable AI methods. The resulting heatmaps are illustrated in Figure 76.

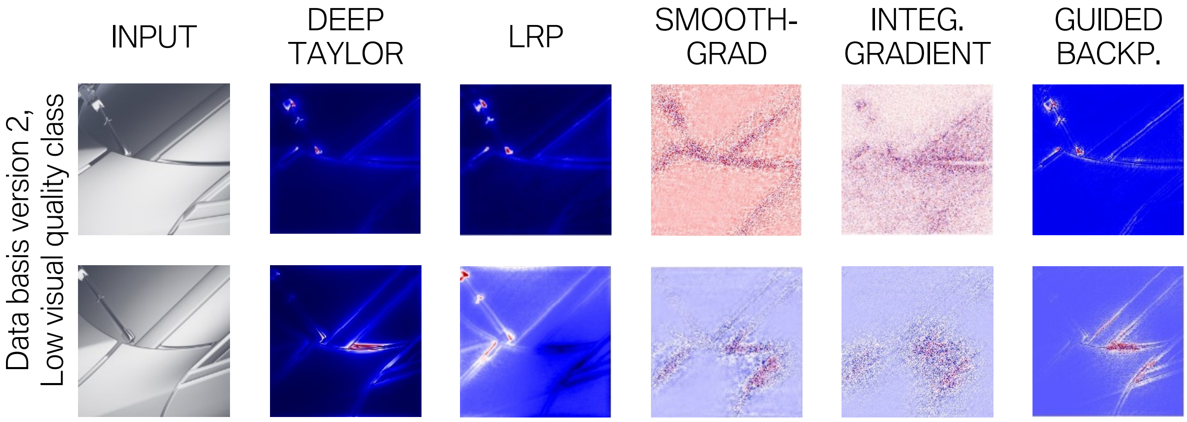


Figure 76. Visualized heatmaps of the Explainable AI methods for the low visual quality class of trained network with data set version 2.

The heatmaps of the five explainable AI methods illustrate similar visualizations of the first input branch. The five methods highlight the windscreen wiper, the side frame split-lines, and slightly the split-line between the side panel and side frame. The area of the windscreen wiper stands out. Compared to the previous heatmaps of the high visual quality class, the wiper is even more highlighted. Therefore, the relevance of the windscreen wiper for the classification of the low visual quality category is even stronger than for the high visual quality category. In addition to the strong marking of the windshield wiper, split-lines at the side frame and side panel are also relevant. The comparison with the previous high visual quality class also reveals a similar highlighting of the areas, but with a stronger intensity. The applied explainable AI methods to the second input branch show very similar heat maps between each other. Highlighted as highly relevant is the split-line between the side frame and the side panel. In addition to this strongly marked area, the split-line from the front door to the side panel is also marked. A comparison of these heatmaps to the heatmaps of the high visual quality class presents that similar areas were marked. The difference between the classes is also the intensity of the highlighting. The high visual quality class illustrates slighter highlights of the areas than the low visual quality class. Therefore, this trained network implies that the extracted and learned features are different variations and combinations of the split-line attributes gap, flush, and parallelism. Thus, the developed heatmaps suggest that the right features have been recognized and trained by the network.

The explainable AI methods which were applied to both trained models, with data basis version 1 and data basis version 2, demonstrated the relevance of the split-lines for the classification related to the two categories high and low visual quality.

The visualized heatmaps presented that the intensity of the variation of the three split-lines was necessary for the classification. In both networks, the split-lines were marked more intense and more comprehensive in the class of low visual quality. In heatmaps of the trained network with the data basis version 2, the windscreen wiper was also marked and, therefore, classification relevance. This feature extraction was not expected or wanted and should, therefore, be eliminated. A strategy is a variation of the windscreen wiper position to simulate the unimportance. Also, the heat maps developed are a point image of individual images. By further investigation of input images using the explainable AI methods, the verification can be reviewed again.

9.2.4. Transferability to Photorealistic renderings

Both trained models of data set versions 1 and 2 were tested towards their transferability to photorealistic renderings of different split-line situations. Two images represented each split-line situation. The renderings were created by using the Software Blender and were inspired by the labelling image quality. Compared to the visualizations for the labelling, the pixels were reduced to 224×224 pixels based on the network input format. Both trained models were tested by using four split-line situations in photorealistic visualizations. The true class of these four images were based on expert labelling. The visualizations and their true class are presented in Figure 77.

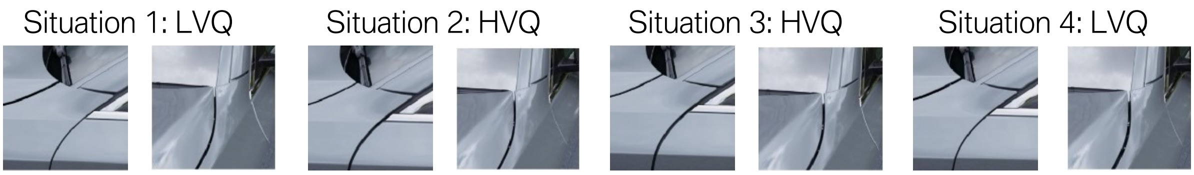


Figure 77. Visualisation of the photorealistic image rendering quality for testing the model transferability.

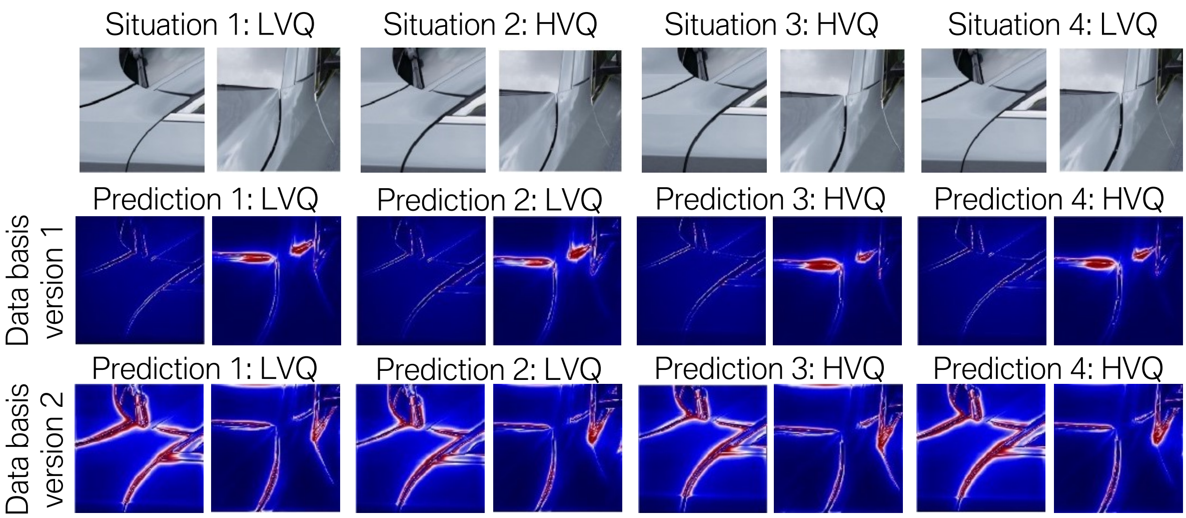
These four split-line situations were given as input sizes to both trained models. The classification predictions of both models were presented in Table 20.

Table 20. Prediction results of trained models with data set version 1 and 2.

Network model	Situation 1: low visual quality	Situation 2: high visual quality	Situation 3: high visual quality	Situation 4: low visual quality
Data basis version 1, 53 epochs	Low visual quality	Low visual quality	High visual quality	High visual quality
Data basis version 2, 30 epochs	Low visual quality	Low visual quality	High visual quality	High visual quality

Both trained networks predicted a similar class for each split-line situation in photorealistic quality. Situation 1 was predicted as low visual quality, which corresponds with the true label. The second situation was predicted as low visual quality by both trained networks. This prediction contrasts with the true label, which was high visual quality. Both trained networks correctly predicted the third split-line situation. The true and the predicted label was high visual quality. Finally, the fourth split-line situation was wrongly predicted by the networks. The true label was low visual quality, and both networks predicted the class as high visual quality.

In conclusion, the prediction of the trained networks demonstrated 50 % correct predictions. However, this percentage reflects only a sample. The four situations tested indicate transferability, but the study should be expanded to increase research potential. Additionally, the photorealistic data represented a new data format, which was unseen and unknown by the trained model. A transferability to the new image quality is difficult for the network because now disturbing factors like reflection or background are added. For a deeper understanding of both model predictions, the Deep Taylor explainable AI method is applied each of the four split-line situations (cf. Figure 78).



The explainable AI heatmaps of the trained model with data set version 1 demonstrated similar

Figure 78. Resulting heatmaps of the Explainable AI method Deep Taylor for evaluating the transferability of the photorealistic renderings.

results. For each split-line situation, the right image was more informative. The low visual quality class predictions identified the split-line between side frame and side panel and the area above the front flap as necessary. Similar areas were highlighted within the low visual quality class within the heatmaps, but with less intensity. The left image per split-line situation illustrated similar highlighted areas with low intensity. Therefore, the split-lines between the side panel and front flap and side panel and front door were important. The resulting heatmaps by using the Deep Taylor method for the network trained with data basis 2 illustrated different important areas compared to the heatmaps of the network trained with data basis 1. The left images of all split-line situations highlighted the split-lines between the front flap and the side panel and the side panel and the front door with high intensity. The windscreen wiper was also identified as an essential area, similar to the heatmaps of the tested renderings. Additionally, the split-line between the side panel and the side frame was highlighted. The right images demonstrated a low intensity of the split-line between the front flap and the side panel. A second intensely highlighted area was the split-line between the door and the side mirror. Based on the different highlighted areas of both networks, a conclusion is that the networks extracted different features for the split-line quality classification. Nevertheless, the differences within the heatmaps between the two classes high and low visual quality were slight for both trained networks. The derived heatmaps of the renderings showed a difference in the intensity but show a very similar pattern.

9.2.5. Transferability to Real Images

Based on the better training results of the network with data basis version 2, this network was investigated regarding the prediction performance of real automobile images. For consideration, two data sets were used. The first data set include the automobile model, which was in the training data set. The second data set included other automobile models. The camera positing was readjusted in the same way the rendering camera position was placed. In doing so, the images of actual cars and their split-lines in two situations are presented in Figure 79 resulted.

Figure 79. Images of different automobile models for testing the transferability capabilities of the trained model regarding real images.

All displayed images correspond to a high visual quality class. The images demonstrate the different illustration quality of the automotive images. Compared to the renderings and the photorealistic images, an increase of reflections and high variance of product-based parameters is identifiable. Besides, the geometry of the two new automobile models is another unknown factor for the trained model. In summary, these illustrations represent a high complexity and new variables for the trained model. In this evaluation, the focus was more on principle transferability. The trained model identifies split-lines as a relevant feature in these input images with the new display quality. Nevertheless, the prediction results of the trained network with data basis version 2 were for the images of situation 1 for all three automobiles high visual quality. Therefore, the trained model predicted the situation 1 images independent of the automobile model correctly. The situation 2 images of the three automobile models were predicted as low visual quality, which represents a wrong prediction. For a better understanding, the explainable AI method Deep Taylor was applied for the real images. The resulting heatmaps are presented in Figure 80.

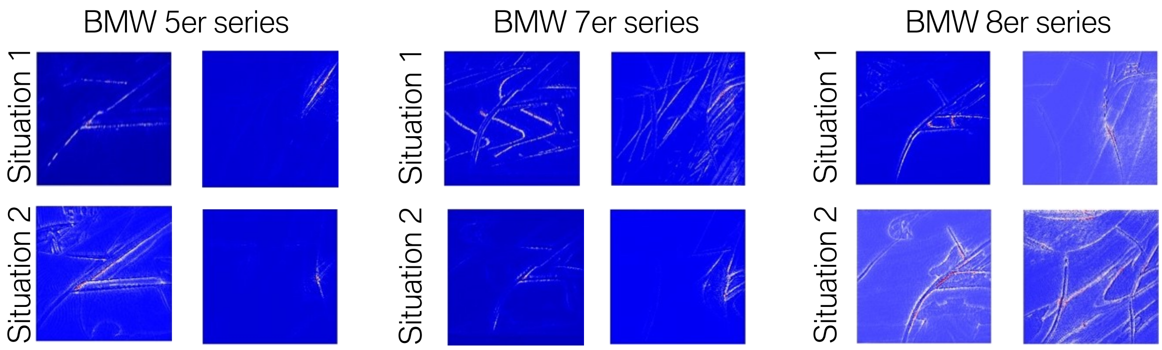


Figure 80. Resulting heatmaps of the Explainable AI method Deep Taylor of the trained model by testing the transferability regarding real images.

The resulting heatmaps are illustrated per automobile and image type, situation 1 and 2, different heatmaps. The 5 series' heatmaps, which was analogical to the training data set, illustrated similar visualizations for both image situations. Based on the left images of situation 1, the focus of the network was the chromium-plated border at the front door and the side mirror. Slightly highlighted areas were available at the split-line between the front door and the side panel. The right heatmap of the 5 series model had slightly marked the split-line between the side panel and the front door. Between the situation 1 and 2 heatmaps, a small difference in the highlighting intensity is noticeable. The heatmaps of the 7 series highlighted different areas and features between the inside and the outside images. Within the situation 1 images, the reflections were intensely marked in both heatmaps. Within the situation 2 images, slight highlights were marked: the split-line between the front door and the side panel on the left image. On the right image, the chromium-plated border was slightly highlighted. Despite the strong reflections, the situation 1 image was correctly classified by the trained model. Nevertheless, it is not entirely conclusive whether the correct characteristics were used for the classification. The heatmaps of the BMW 8 series illustrated differences between the situation 1 and 2 images. Within the situation 1 images, the left heatmap marked the split-line between the front door and the side panel as a relevant area. On the right image, the split-lines were as well marked as important. Within the situation 2 image, the reflections and the split-lines were intensely highlighted. Similar to the inside image of the 7er series, the reflections were selected as a relevant feature.

The results of the real images classification of the trained model are as expected. Correct classification was only partly possible. By applying the explainable AI method Deep Taylor, the relevant features could be identified. The split-lines were always of importance. The recognition and the focus on the split-lines shows a first transfer possibility of the network to real data. Additionally, the applied explainable AI method shows that the reflections influence the classification of the trained model.

9.3. Summary and Contribution of Research Question 4

The fourth research question focused on an objective assessment method for the level of perceived quality. Therefore, an objective assessment method was deduced to implement convolutional neural networks for visual assessments of perceived quality (the first question of the fourth research question). Based on the high subjective character of assessing perceived quality, the developed convolutional neural network dealt with a more sophisticated classification task (the second question of the fourth research question). The developed method was a convolutional neural network with a supervised classification task, having two classes: high and low visual quality. As network architecture, a VGG-16 model with a Siamese approach was chosen. The database and labeling of both classes were synthetically generated within the method by using product ray tracings and a self-developed labeling tool. The convolutional neural network training was performed iteratively, and the best performance of the network was realized by having 30 epochs and a learning rate of 0.00001. With these parameters, the network realized accuracy of 0.82, and it was able to predict correctly 79 % of the high visual quality class and 86 % of the low visual quality class (the third question of the fourth research question). Additionally, evaluations using explainable AI methods were used to comprehensively visualize the extracted features by the convolutional neural network (the third question of the fourth research question). The heatmaps were derived by five different explainable AI methods: Deep Taylor, Layer-wise Relevance Propagation, SmoothGrad, Integrated Gradient, and Guided Backpropagation. The derived heatmaps illustrated the split-lines as a decision important feature for both trained networks. In a further investigation, the trained network was tested regarding the prediction results of images with a different visualization quality, which was in photorealistic visualizations and real photos. Their results imply that a trained network transfer regarding this image quality was hardly possible (the fourth question of the fourth research question).

A related approach was developed by Pan et al. (2017). They presented a sophisticated approach using a generative adversarial network for an exclusive assessment of automobile designs regarding the character. This method used synthetic data, which was created by the neural network. The network then evaluates the generated data. This procedure is not feasible with this split-line evaluation. According to perception basics, a synthetic generation of the split-line data without additional labeling can result in a falsified database. Also, there is a risk that the data generated by the network will not correspond to the split-line variations that occur during production. Therefore, the presented method does not apply to data generation, labeling, used network architecture, and classification scope.

As convolutional neural networks can perform the human-related task of recognizing images and assign them to a category, this similarity between a convolutional neural network for classification and human perception was used in the thesis. Scientific findings indicated that the layers of the neural network perform different functions. Whereas the first hidden layers recognized points or lines, the deeper layers recognized connections and abstract general concepts. For example, the Velocipedia (Gimini 2016a, 2016b; Gregory 2005) was cited to explain the data proceeding of convolutional neural networks (Chollet 2018) compared to human perception. Similarities between human perception and the classification of a convolutional neural network were identified because humans and neural networks recognize the global concept in images (Chollet 2018). Individual elements were filtered in perception, making it, for example, difficult for many people to sketch a bicycle from memory. The process of identifying objects or images is called recognition (cf. chapter 2.4.1). By applying a convolutional neural network, following process improvements in the visual assessment of split-lines regarding perceived quality were derived (the second question of the fourth research question):

(1) handling attribute complexity,
(2) global quality assessment,
(3) visual quality assessment without product decomposition,
(4) constant and reproducible assessment result and
(5) objective assessment through accumulated subjective assessments.

Since convolutional neural networks can handle image data representing any split-line attribute combinations, including the global automobile context and relations to other split-line attributes, the method is suitable for capturing and handling greater quality complexity. The visual assessment was also based on images, which were represented on a global level without further processing effort. Therefore, the assessment using a convolutional neural network is authentic and direct. Additionally, using a labeling method unfiltered, and direct customer assessment can be included in the assessment method. Further, a product decomposition, like other methods is using, is irrelevant, which leads to a process simplification. Furthermore, a successfully trained algorithm reflects a continuous and reproducible assessment process. Assessments are no longer subject to any individual emotions, experience, or situations and were hence a stable basis for a split-line perceived quality assessment. The fifth improvement potential was an objective assessment process. A more objective process is established through a multi-stage quality assessment and the derivation of a label concerning the average assessment result.

An often-discussed disadvantage of neural networks as a quality assessment tool is the prediction model, which is a black box. Within the model, the weights and parameters were defined. Nevertheless, the definition of parameters and weights of a network layer cannot be viewed or influenced due to the automated process. An approach to decode this black box is explainable AI methods. These methods displayed network-relevant pixels through a heat map. This visualization offered the possibility to check whether an algorithm had extracted the decisive features for each class. These methods were used as an evaluation method of the trained models within the thesis. The limitations of the thesis regarding the application of a neural network are:

(1) transferability to photorealistic and real images,
(2) transferability to further automobiles,
(3) transferability to further quality attributes,
(4) the number of training classes,
(5) the network architecture.

As presented in chapters 9.2.4 and 9.2.5, a transfer and a prediction of the visual split-line quality with photorealistic and real images was partially possible with the trained model. The databases consisted of CAD-quality images that exclude factors like reflection, color, and background information about the automobile. In comparison, the photorealistic and real images display these factors. Using a trained model based on CAD-quality images, a prediction of photorealistic and real images is difficult. The factors affect the images on a high level, which prevents a robust classification. For a robust application of a trained model for photorealistic or real images, the database needs to include these kinds of images. Another limitation of the trained model is the transferability regarding other automobile types. The trained area was the a-pillar of a BMW 5 series sedan. The design of the a-pillar is different in different automobiles. This great variety of the design of a-pillars needs to be included in the database for realizing a robust classification model. Also, the transfer of the trained model for further attributes, such as, i.e., the design of front lights, is not possible. For realizing such a model, the database needs to be filled up with examples: In all three transfer tasks, the database must be established according to the classification task. The realized model has a narrow stage, which is, by definition, an expert model in one domain but cannot translate to further tasks (Davenport et al. 2019). Based on the narrow level of the neural network, a transfer is limited possible. The number of trained classes is flexible and can be increased by splitting the database or completing the database with additional classes. Therefore, the derived approach of using a neural network for quality assessments can be used. A fifth limitation is the network architecture. By using two input branches, the flexibility regarding the input perspectives is reduced. Optimization of the network would be the reduction to one input branch. In doing so, the input perspectives would be reduced and separately considered. By having one input branch of the neural network, the training of different split-lines is more flexible.

State-of-the-art methods in research considering and assessing the perceived quality of products were often using a product decomposition. Complex products were segmented into individual attributes, and the considerable scope was contracted until an attribute was assessable with standard quantification methods, e.g., ranking methods. By summarizing the results of individual attributes, a result of the entire perceived quality value was generated. Comparing these methods to the definition of perceived quality, a disregard attracts: the product context, relations were neglected. Also, influencing parameters were partly included but current methods, but they lacked in a consistent approach. The state-of-the-art assessment process is presented in Figure 28.

An approach that attends to use an implementation of this traditional assessment is presented by Stylidis, Wickman, and Söderberg (2018). The deduced framework was based on a definition of perceived quality (Stylidis, Wickman, and Söderberg 2015b) and an attribute structuring using the definition (Stylidis et al. 2015). This approach presented a holistic method to investigate any product. Accordingly, these methods aimed to model and correctly represent the rules and relationships of single product attributes concerning perceived quality. Due to the high product complexity and the sophisticated perception process, the modeling is difficult, and the risk of incorrectly modeled attribute relationships arises.

The idea of machine learning methods, equally neural networks, is to extract rules and relationships using data. With the background of the subjective assessment character of perceived quality and the unknown rules and relationships between the perception of attributes, machine learning methods have potential. The idea of using a neural network-based assessment process is presented in Figure 21.

An approach that attends to use an implementation of this traditional assessment is presented by Stylidis, Wickman, and Söderberg (2018). The deduced framework was based on a definition of perceived quality (Stylidis, Wickman, and Söderberg 2015b) and an attribute structuring using the definition (Stylidis et al. 2015). This approach presented a holistic method to investigate any product. Accordingly, these methods aimed to model and correctly represent the rules and relationships of single product attributes in relation to perceived quality. Due to the high product complexity and the sophisticated perception process, the modeling is difficult, and a risk of incorrectly modeled attribute relationships arises.

The idea of machine learning methods, equally neural networks, is to extract rules and relationships by using data. With the background of the subjective assessment character of perceived quality and the unknown rules and relationships between the perception of attributes, the application of machine learning methods has potential. The idea of using a neural network-based assessment process is presented in Figure 81.

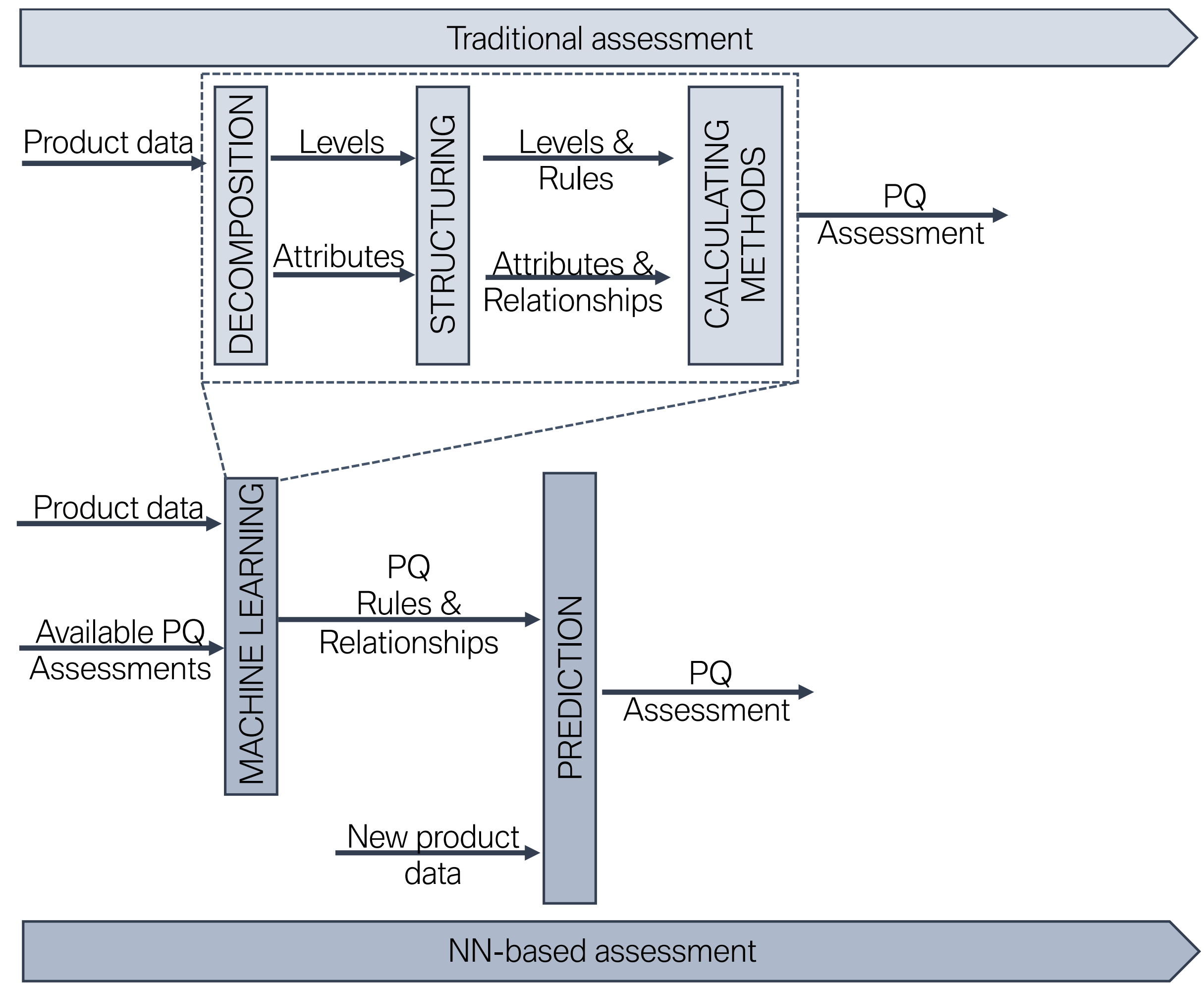


Figure 81. Traditional perceived quality assessment process versus machine learning based perceived quality assessment.

Accordingly, the presented process rules and relationships were extracted by the neural network in an algorithm. In a second step, the new data can be classified by using the algorithm. The critical success factor for implementing a neural network is to classify categories by using trained features correctly. Features represent characteristics that are significant for a category. For example, for a neural network classifying cats and dogs, a cat's whiskers are a distinguishing feature between both categories and a significant characteristic for the category cat. Consequently, a neural network trained to classify perceived quality categories of split-lines deduces from the available input data significant characteristics for each category, which significantly distinguishes from other categories. Thus, the data has to represent any class, and characteristic features have to be maintained. The feature extraction is an uncertainty in the training of neural networks. Indeed, the extracted features per category are not observable. A feature investigation of convolutional neural networks is possible in the form of a test data set or in the form of explainable AI methods.

10. DISCUSSION

Due to industrial and scientific interest, the research discipline of perceived quality gained in importance. Various methods and approaches to assess the level of perceived quality of a product were developed. The thesis developed a new method for assessing the perceived quality level of a product by addressing four research hypotheses, which were:

I. To improve perceived quality, the process of assessing split-lines needs to be investigated and expanded to an assessment of visual requirements.
II. Identification and definition of parameters that affects the visual assessment of split-line attributes.
III. The global product context and relations between split-lines and their attributes are important for the perception of split-lines regarding their visual quality.
IV. The application of machine learning approaches offers an improved assessment method of the global product and relations between attributes equally pursuing the definition of perceived quality.

The discussion focuses on five aspects: the advantage of the underlying approach, the inclusion of psychological aspects, the complementary aspect of the individual research results, the level of automatization, standardization and reusability of the method, and the limitation of the method.

The first aspect discusses the advantage of the research approach. A literature review deduced the research gap of a method for assessing the perceived quality of products by respecting its definition. Based on the derived research gap, the general need was divided into four research questions. In doing so, the first advantage is the flexibility of the method and the results. The approach is usable as a holistic method, or the individual results can be implemented separately in various applications. For example, a possible application lies within the early stage of the product development process. Using the holistic approach, product features, which have a perception-based assessment character, can be designed and limited by including relevant psychological basics and an objective assessment method. Nevertheless, another application is the consideration of individual research results. For example, the derived human- and product-related parameters can be applied in different assessments of perceived quality attributes. Their validity is independent of the application of a split-line assessment. The second advantage is the consistency of the approach. Within all four research questions, the investigation object was similar. Additionally, the results of previous questions were included in subsequent proceedings. As mentioned above, the general approach is independent of the split-line consideration and can be applied to further visual quality attributes or further investigation objects. For example, the fitness of different automobile materials or the split-line assessment of a phone could also be the initial problem for applying the developed method. Therefore, the third advantage is the independence of the developed method regarding the investigation object or the visual assessment topic.

The second aspect, which is discussed, is the inclusion of psychological aspects. A disregard of perception in product assessments was discussed by Carbon (2019) within the Psychology of Design. Eleven conceptual moves were deduced, which support the design of products with a high level of satisfaction. The derived method of the thesis is discussed regarding a supporting character to some of the conceptual moves. Using a neural network for assessing visual quality, the derived method accompanies the sixth and eleventh conceptual move. The sixth conceptual move demonstrated the importance of including the psychological context within product design (Carbon 2019). The trained neural network was based on images showing the split-line in the context of the a-pillar and different design features, for example, the side mirror. Also, the global product context regarding a visual assessment of split-lines was investigated within the third research question. The results of this consideration were included in the neural network-based assessment method. The eleventh conceptual move was evolved based on the Gestalt theory (Carbon 2019). Therefore, this conceptual move deduced the necessity of considering the Gestalt theory within the product design. By using convolutional neural networks, the entirety of a product can be captured and assessed. Therefore, Gestalt principles can be regarded within the use of neural networks for visual quality assessment. Conceptual moves which were critical realized within the thesis were primarily conceptual move two and ten. The second conceptual move considered the understanding of how design affects prediction (Carbon 2019). By using neural networks, these relations and rules were modeled by the algorithm but remain unknown. Therefore, the application of neural networks does not support a consideration of the process between designs and the predictions they trigger, and this conceptual move is not promoted by using the thesis approach. The tenth conceptual move demonstrated the importance of a multisensory view on product design (Carbon 2019). Using convolutional neural networks causes an isolated consideration regarding visual quality features. Additionally, current applied neural networks were defined as narrow AI (Kaplan and Haenlein 2019). Within a narrow AI level, trained features and relation cannot be transferred by the neural network to other similar problems. For example, visual and optical requirements regarding product design are not easily grasped using standard machine learning methods. A cross-disciplinary application for measuring product design in a multidimensional view requires additional research effort. A conceptual move that was not realized by the current neural network of the thesis but is possible to model using machine learning is the third move. Within this move, the consideration of Zeitgeist within product design is mentioned. A possible task using machine learning could be predicting future design by representing the Zeitgeist of former design within the training data for the machine learning algorithm. The developed method using a neural network for a visual assessment of split-lines regarding perceived quality considers the sixth and eleventh conceptual moves but demonstrated limitations regarding the second and tenth move.

The third discussion aspect addresses the complementary function of individual research results or the entire method. The developed machine learning-based assessment method can support various existing approaches based on the mainly decompositional methods for assessing perceived quality in research and industry. For example, the PQF method of Stylidis et al. (2015); Stylidis, Wickman, and Söderberg (2015a, 2018); Stylidis et al. (2015) demonstrates a holistic approach for assessing the entire perceived quality level of a product. The applied method has a decompositional character and could be validated using the developed machine learning-based method in this thesis. This application would have an additional advantage. In applying two different approaches, the resulting perceived quality results of a similar product can be compared. Based on this comparison, the direct impact of unknown rules and relations within the product is identifiable. The derived method does not need to replace the traditional definition of split-lines and can have a supportive function. A combination of the machine learning-based method within the traditional definition and assessment process of split-lines is related to derived process improvement of the first research question. A result of this process improvement was the need of an assessment method for the Refinement level besides the solid toolbox for assessing the Producible Quality level. A reinforcement of the Refinement level using the machine learning assessment method demonstrates such a combination of different assessment methods. Therefore, the derived approach and the generated results can either be radically applied or gently integrated within existing processes.

The fourth discussion aspect focuses on the automatization, standardization, and reusability level of the derived approach and the results. The possibility to reduce manual operations in the method is considered as an automatization level. The standardization aspect considers the new method for assessing perceived quality. The reusability of the method discusses how different results can be transferred to other problems. Automatization of the method is hardly possible. The conducted surveys need manual design and conduction. The analysis of the collected data and the calculation of various descriptive key values can be partly automated. Similarly, the automatization of the machine learning approach is problematic; thus, the identification of a suitable machine learning approach, a suitable model, or the training and optimization needs human intelligence. The new method for assessing the split-line quality using a neural network has the advantage of a standardized process and assessment result. Expert's assessment supported the traditional process of assessing split-line quality. These assessments fluctuated between one expert based on the current situation and between several experts based on different expertise levels. The developed method, which supported the split-line assessment process by a neural network, enables repeatability of assessment results, and the assessment results do not longer differ between several experts. Therefore, a standardized assessment process with a consistent and repeatable assessment approach is realized by the developed method. The level of reusability of the results depends on the question. The used approach of the first research question is reusable for further process improvements, whereas the results are split-line specific, and the process is partly dependent on the company. The same applies to identified parameters, the global product context, the relations between split-lines, and the developed machine learning-based assessment method. For example, the human-related and product-related parameters are reusable within further automobile assessments. A transferability of the parameters to further products is partly possible if the product is comparable. Nevertheless, the developed methods are generally reusable for further investigations.

The last discussion aspect is the limitation of the research results. The developed machine learning algorithm is related to the narrow AI level. Machine learning algorithms are defined as narrow AI if they are an expert system for one specific task (Kaplan and Haenlein 2019). Narrow AI systems are not able to transfer knowledge or lessons learned to a new task. Every machine application, which is deduced in the last years, is a narrow AI system. Therefore, applying the trained model of the thesis is an expert for assessing the split-line quality at the specific area of the a-pillar. A direct transfer of this knowledge to another automobile area or another product is not possible. Additionally, the more sophisticated task to transfer the knowledge of a high-quality split-line level into spoken words is not possible. Indeed, there are machine learning algorithms that understand spoken words, and there are algorithms to classify quality, but both systems cannot communicate with each other without human intelligence. A realization of such a system with a human-related intelligence level is called general AI (Kaplan and Haenlein 2019).

11. CONCLUSION

The thesis focused on the development of new approaches for assessing the visual split-line regarding perceived quality. By focusing on one quality attribute, the perceived quality assessment process was comprehensively analyzed, and potentials in the form of standardization and objectivation were deduced. Due to the standardization and implementation of a neural network-based assessment of split-lines, a possibility for a more objective assessment of split-lines was introduced. Based on these findings, four further research potentials were derived regarding the application of neural networks or different machine learning approaches:

(1) Training with virtual images and predicting real images,
(2) Transferability of quality
(3) Further product attributes
(4) Further network architectures

The first research potential exists in the field of transfer learning. An application in the design process is the definition and training of split-line quality in the concept phase and applying the trained algorithm in production. The advantage is the consistent use of machine learning in defining objectives and the verification of the objectives. A problem is the data format. In the concept phase, the data exists in a virtual format. Nevertheless, in production, the data would be real images of the automobile. Transfer learning methods would be a possible solution to enable the neural network to predict real images. The primary training has to be done with the virtual training data in the concept phase. Additionally, when having the hardware prototypes, these data could be added to the training data set.

A second research potential lies in the optimization of the training effort. The training of a neural network required many resources. For industrialization, the process has to be optimized. Therefore, transfer learning is as well a solution. The acquired knowledge of a neural network has to be extended by adding new split-line categories. To this end, already trained primary layers of the network can be reused in transfer learning. The effort of training new quality categories of split-line attributes is reduced. The parameters, such as data quantity, had to be identified to implement the transfer learning. Additionally, a fitting network architecture had to be deduced.

The third research potential is a consideration of using a neural network to assess further product attributes. The feature extraction needs to be investigated. Similar to the split-lines, the attribute must differ within the categories so that the network can extract the category-relevant features. Furthermore, it has to be investigated which attributes can be classified with a neural network. For example, acoustic quality features, such as squeak and rattle, can be assessed by a classification of the frequencies. The modeling of the design process has to be considered with caution. Thus, this process is characterized by creativity. The research of neural networks is currently not, in the long run, to transfer the human creative process to a machine. Such a challenge requires a much more complicated process. Pan et al. (2017) have shown the first possibility of using neural networks. They were able to classify the appearance of an automobile into the categories “sporty”, “appealing”, “innovative,” or “luxurious”.

The implemented network was a VGG-16 architecture with Siamese input branches. Besides applying different automobile areas for a split-line quality assessment, the presented area of the a-pillar could be validated using another network architecture, for example, using the AlexNet (Krizhevsky, Sutskever, and Hinton 2012) or new network architecture.

The neural network-based assessment method for visual quality features can be implemented within an early phase of the product development process. Based on this method, quality standards and visual quality features can be defined and objectively measured by the neural network. By transferring the trained network into the production of a product, a standardized and consistent assessment process of visual quality features can be realized. The advantage of using the trained network are reduction of assessment effort, consistent assessment results, and an assessment of each product.

The use of neural networks demonstrated an increase in process standardization and automation in various industrial areas. The present thesis results presented the possibility of objectifying a strongly perception-based assessment process and including the entirety of a product appearance. Thus, these investigations hopefully enable further research in the use of neural networks in the area of perceived quality.

A. APPENDIX A

The applied three-dimensional translation matrix was:

$$\begin{pmatrix} x' \\ y' \\ z' \\ 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & t_x \\ 0 & 1 & 0 & t_y \\ 0 & 0 & 1 & t_z \\ 0 & 0 & 0 & 1 \end{pmatrix} * \begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix} \tag{4}$$

The applied three-dimensional rotation matrix around the x-axis was:

$$\begin{pmatrix} x' \\ y' \\ z' \\ 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & cos\alpha & -sin\alpha & 0 \\ 0 & sin\alpha & cos\alpha & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} * \begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix} \tag{5}$$

The applied three-dimensional rotation matrix around the y-axis was:

$$\begin{pmatrix} x' \\ y' \\ z' \\ 1 \end{pmatrix} = \begin{pmatrix} cos\beta & 0 & sin\beta & 0 \\ 0 & 1 & 0 & 0 \\ -sin\beta & 0 & cos\beta & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} * \begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix} \tag{6}$$

The applied three-dimensional rotation matrix around the z-axis was:

$$\begin{pmatrix} x' \\ y' \\ z' \\ 1 \end{pmatrix} = \begin{pmatrix} cos\mu & -sin\mu & 0 & 0 \\ sin\mu & cos\mu & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} * \begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix} \tag{7}$$

B. APPENDIX B

The studies in chapter 6.3, 7.3, 8.2 were conducted using a browser-based application. This application was implemented by using HTML and JavaScript. Additionally, the SurveyJS library (OÜ and Baltic 2020) was used for the basic study design and operating elements within the study. For each study, the general approach for preparation is shown in Figure 82.

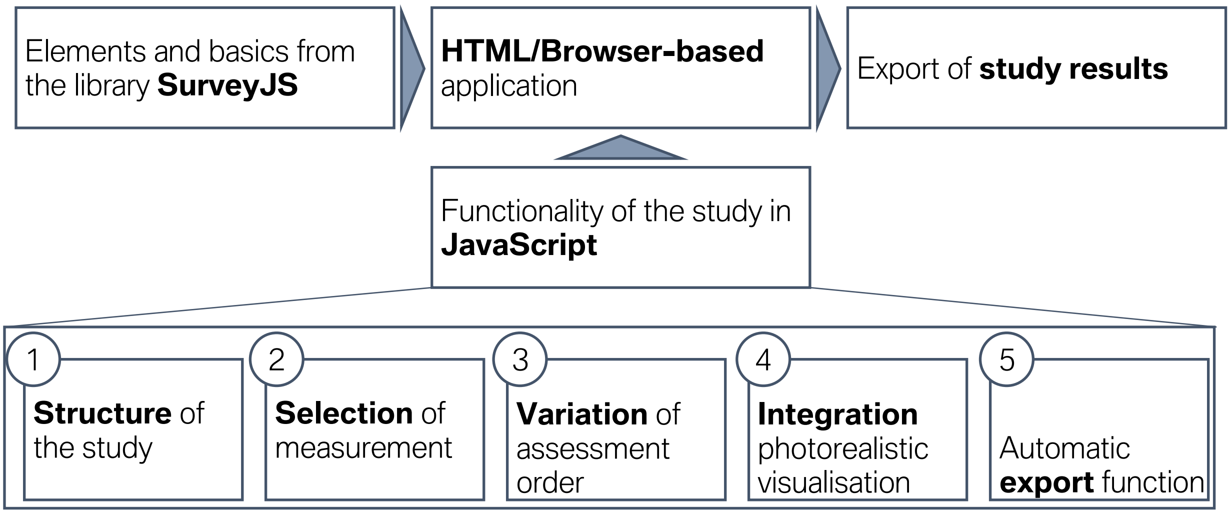


Figure 82. Method for creating a browser-based study application by using html, javascript and css.

The central document for the browser-based study application was the HTML script. Within this script, the appearance and the study functionalities were included. The approach for defining a manipulating, the HTML script was supported with elements from the library SurveyJS (OÜ and Baltic 2020) and these functions were based in JavaScript code. The functionality of the JavaScript code was divided in five steps. The first step described the structure of the study. Within this step, different study structure was defined. The structure of both studies were a general question and a study related section. The second step was the selection of a measurement technique. This step was closely related to the chosen scale level in the study. This step was a different in the studies related to the research hypotheses: the perception study of product-related parameters used a slide bar, the conjoint study used a drag-and-drop feature to rank the product alternatives and the perceptual study used a seven point Likert scale level (Joshi et al. 2015). The third step in the JavaScript code considered a random variation of the image order in both study types and with both measurement techniques. This step was performed to avoid sequence effects within the data. Within the fourth step of the JavaScript code, the generated photorealistic visualizations (cf. chapter 6.2) were integrated into the study. The integration was performed by using a unique image description. The fifth step considered an automated export of the study results. Within the JavaScript code, the download of the results was performed after each participant and by answering the last study question. The results were saved in the download folder of the computer.

The design of the general HTML application was that all required documents, the html script, the visualizations, the elements of the SurveyJS library and the JavaScript code needed to be stored in one folder.

REFERENCES

Aaker, David a. 1991. “Managing Brand Equity.” Journal of Marketing.

Abadi, Martín et al. 2016. “TensorFlow : A System for Large-Scale Machine Learning This Paper Is Included in the Proceedings of the TensorFlow : A System for Large-Scale Machine Learning.” In USENIX Symposium on Operating Systems Design and Implementation, Savannah.

Alber, Maximilian et al. 2018. “INNvestigate Neural Networks!” arXiv:1808.04260 [cs.LG]. http://arxiv.org/abs/1808.04260.

Arens, Tilo et al. 2013. Grundwissen Mathematikstudium Grundwissen Mathematikstudium. Berlin, Heidelberg: Springer Berlin Heidelberg. http://link.springer.com/10.1007/978-3-8274-2309-2.

Bach, Sebastian et al. 2015. “Analyzing Classifiers: Fisher Vectors and Deep Neural Networks.” http://arxiv.org/abs/1512.00172.

Backhaus, Klaus, Bernd Erichson, Wulff Plinke, and Rolf Weiber. 2016. Multivariate Analysemethoden. 14th ed. Berlin, Heidelberg: Springer Berlin Heidelberg. http://link.springer.com/10.1007/978-3-662-46076-4.

Backhaus, Klaus, Bernd Erichson, and Rolf Weiber. 2015. Fortgeschrittene Multivariate Analysemethoden. Berlin, Heidelberg: Springer Berlin Heidelberg. http://link.springer.com/10.1007/978-3-662-46087-0.

Bahlmann, Claus, Gunther Heidemann, and Helge Ritter. 1999. “Artificial Neural Networks for Automated Quality Control of Textile Seams.” Pattern Recognition 32(6): 1049–60.

Beierle, C., and G. Kern-Isberner. 2019. “Maschinelles Lernen.” In Methoden Wissensbasierter Systeme, Wiesbaden: Springer Fachmedien Wiesbaden GmbH.

Ben-Gal, Irad, Alexandra Dana, Niv Shkolnik, and Gonen Singer. 2014. “Efficient Construction of Decision Trees by the Dual Information Distance Method.” Quality Technology & Quantitative Management 11(1): 133–47.

Blender Foundation. 2019a. “Blender - a 3D Modelling and Rendering Package.” https://www.blender.org/.

———. 2019b. “Blender 2.81 Manual - Display Mode.” https://docs.blender.org/manual/en/dev/video_editing/preview/display_mode.html?highlight=luminance.

———. 2019c. “Blender 2.81 Manual - Glossary.” https://docs.blender.org/manual/en/dev/glossary/index.html?highlight=luminance#term-luminance.

———. 2019d. “Blender 2.81 Manual - Introduction.” https://docs.blender.org/manual/en/dev/editors/image/introduction.html?highlight=luminance.

———. 2019e. “Blender 2.81 Manual - Visual Effects.” https://docs.blender.org/manual/en/dev/grease_pencil/visual_effects/glow.html?highlight=luminance.

BMW Group. 2018. Department Total Vehicle Validation.

Bohn, Martin, and Klaus Hetsch. 2013a. “Einführung Zum Toleranzmanagement.” In Maßhaltige Kunststoff-Formteile, München: Carl Hanser Verlag GmbH & Co. KG, 33–44. http://www.hanser-elibrary.com/doi/abs/10.3139/9783446436893.003.

———. 2013b. Toleranzmanagement Im Automobilbau. München: Carl Hanser Verlag.

———. 2013c. “Toleranzmanagement Im Entwicklungsprozess.” In Toleranzmanagement Im Automobilbau, München: Carl Hanser Verlag GmbH & Co. KG, 95–98. http://www.hanser-elibrary.com/doi/abs/10.3139/9783446435445.005.

Borrego, Maura, Elliot P Douglas, and Catherine T Amelink. 2009. “Quantitative , Qualitative , and Mixed Research Methods in Engineering Education Division of Student Affairs.” Journal of Engineering Education 98(1): 53–66. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.1020.1245&rep=rep1&type=pdf.

Bream, Harald. 2004. Die Macht Der Farben. Wirtschaft. München: Wirtschaftsverlag Langen Mller / Herbig.

Bromley, Jane et al. 1993. “Signature Verification Using a ‘Siamese’ Time Delay Neural Network.” In Proceedings of the 6th International Conference on Neural Information Processing Systems,.

Bruner, Jerome S., and A. Leigh Minturn. 1955. “Perceptual Identification and Perceptual Organization.” The Journal of General Psychology 53(1): 21–28.

Bugelski, B. R., and D. A. Alampay. 1961. “The Role of Frequency in Developing Perceptual Sets.” Canadian journal of psychology 15(4): 205–11.

Burges, C.J.C. 1998. “A Tutorial on Support Vector Machines for Pattern Recognition.” Data Mining and Knowledge Discovery 2(2): 121–67.

Carbon, Claus Christian. 2010. “The Cycle of Preference: Long-Term Dynamics of Aesthetic Appreciation.” Acta Psychologica 134(2): 233–44. http://dx.doi.org/10.1016/j.actpsy.2010.02.004.

———. 2014. “Understanding Human Perception by Human-Made Illusions.” Frontiers in Human Neuroscience 8: 1–6. http://journal.frontiersin.org/article/10.3389/fnhum.2014.00566/abstract.

———. 2015a. “Predicting Preferences for Innovative Design: The ‘Repeated Evaluation Technique’ (RET).” GfK Marketing Intelligence Review 7(2): 34–39.

———. 2015b. “Wahrnehmungspsychologie.” In Psychologie: Eine Einführung in Ihre Grundlagen Und Anwendungsfelder, eds. Astrid Schütz, Matthias Brand, and Herbert Selg. Lautenbacher, Stefan, 73–85.

———. 2016. “Creating a Framework for Holistic Assessment of Aesthetics: A Response to Nilsson and Axelsson (2015) on Attributes of Aesthetic Quality of Textile Quality.” Perceptual and Motor Skills 122(1): 96–100.

———. 2019. “Psychology of Design.” Design Science 5: e26. https://www.cambridge.org/core/product/identifier/S2053470119000258/type/journal_article.

———. 2020. “The Umwelt for Creative Design Addressing Existential Global Challenges.” Computational and Cognitive Models of Creative Design: 1–12.

Carbon, Claus Christian, and Martina Jakesch. 2013. “A Model for Haptic Aesthetic Processing and Its Implications for Design.” Proceedings of the IEEE 101(9): 2123–33.

Carbon, Claus Christian, and Helmut Leder. 2005. “The Repeated Evaluation Technique (RET). A Method to Capture Dynamic Effects of Innovativeness and Attractiveness.” Applied Cognitive Psychology 19(5): 587–601.

Chen, Jim X. 2010. “Rendering in Computer Graphics.” Wiley Interdisciplinary Reviews: Computational Statistics 2(1): 120–26.

Chi, Hsin Kuang. 2009. “The Impact of Brand Awareness on Consumer Purchase Intention : The Mediating Effect of Perceived Quality and Brand Loyalty.” The Journal of International Management Studies 4(1): 135–44.

Chollet, F. 2018. Deep Learning with Python. USA/Conneticut: Manning Publications Co. https://books.google.de/books?id=ouVcDwAAQBAJ.

Ciresan, Dan, Ueli Meier, Jonathan Masci, and Jürgen Schmidhuber. 2012. “Multi-Column Deep Neural Network for Traffic Sign Classification.” Neural networks : the official journal of the International Neural Network Society 32.

Clancy, K. J., and R. Garsen. 1970. “Why Some Scales Predict Better.” Journal of Advertising Research 5: 33–35.

Cortes, Corinna, and Vladimir Vapnik. 1995. “Support-Vector Networks.” Machine Leaming 297: 273–97.

Creusen, Mariëlle, and Jan Schoormans. 2004. “The Different Roles of Product Appearance in Consumer Choice.” Journal of Product Innovation Management 22(1): 63–81.

Cristianini, N., and J Shawne-Taylor. 2000. An Introduction to Support Vector Machines Machines and Other Kernel-Based Learning Learning Methods. Cambridge University Press.

Dagman, Andreas, Rikard Söderberg, and Casper Wickman. 2004. “A Study of Customers’ and the Automotive Industry ’ s Attitude Regarding Visual Quality Appearance of Split Lines.” In 4th International Conference on Advanced Engineering Design AED’ 2004, , 1–8.

Davenport, Thomas, Abhijit Guha, Dhruv Grewal, and Timna Bressgott. 2019. “How Artificial Intelligence Will Change the Future of Marketing.” Journal of the Academy of Marketing Science. http://dx.doi.org/10.1007/s11747-019-00696-0.

Dreiseitl, Stephan, and Lucila Ohno-Machado. 2002. “Logistic Regression and Artificial Neural Network Classification Models: A Methodology Review.” Journal of Biomedical Informatics 35(5–6): 352–59.

Dumont-Bècle, Patricia et al. 2001. “Multi-Texturing Approach for Paint Appearance Simulation on Virtual Vehicles.” Driving Simulation Conference (DSC) 2001: 11.

Edwards, Peter J et al. 1999. “The Application of Neural Networks to the Papermaking Industry.” IEEE TRANSACTIONS ON NEURAL NETWORKS 10(6): 1456–64.

Ehlert, Matthias, Robert Hofmann, Andreas Stockinger, and Sandro Wartzack. 2016. “On the Evolvement of Limiting Positions as a Result of the Interaction of Tolerances and Compliant Systems.” Procedia CIRP 43: 76–81. http://dx.doi.org/10.1016/j.procir.2016.02.150.

Eraslan, Gökcen et al. 2019. “Deep Learning: New Computational Modelling Techniques for

Genomics." Nature Reviews Genetics. https://doi.org/10.1038/s41576-019-0122-6.

Esch, Franz-Rudolf. 2013. Strategie Und Technik Des Automobilmarketing. ed. Franz-Rudolf Esch. Wiesbaden: Springer Fachmedien Wiesbaden. http://link.springer.com/10.1007/978-3-8349-3831-2.

Falk, Björn, Bastian Quattelbaum, and Robert Schmitt. 2010. "Product Quality from the Customers' Perspective – Systematic Elicitation and Deployment of Perceived Quality Information." : 211–22.

Fontana, M., F. Giannini, and M. Meirana. 1999. "A Free Form Feature Taxonomy." Computer Graphics Forum 18(3): 107–18. http://doi.wiley.com/10.1111/1467-8659.00332.

Forslund, K., and R. Söderberg. 2008. "Optical Quality as a Product Attribute - A Descriptive Study from the Automotive Industry." Proceedings of NordDesign 2008 Conference (1): 79–88.

Forslund, K, A Dagman, and R Söderberg. 2006. "Visual Sensitivity : Communicating Poor Quality." In Proceedings DESIGN 2006, ed. D. Marjanovic. Dubrovnik, Croatia, 713–20.

Förster, Anja, and Peter Kreuz. 2003. Marketing Trends - Ideen Und Konzepte Für Ihren Markterfolg.

Freeman, Kim D, Jack Dart, and Jack Dart. 2008. "Measuring the Perceived Quality of Professional Business Services Measuring the Perceived Quality of Professional Business Services." Journal of Professional Services Marketing (June 2013): 37–41.

Garvin, David A. 1984. "Product Quality: An Important Strategic Weapon." Business Horizons 27(3): 40–43.

Gimini, Gianluca. 2016a. "Velocipedia." Tecnoscienza Italien Journal of Science & Technology Studies 1.

———. 2016b. "Velocipedia." Tecnoscienza Italien Journal of Science & Technology Studies 2.

Goodfellow, Ian, Yoshua Bengio, and Aaron Courville. 2016. Deep Learning. MIT Press. http://www.deeplearningbook.org.

Gotlieb, Jerry B, Dhruv Grewal, and Stephen W Brown. 1994. "Consumer Satisfaction and Perceived Quality : Complementary or Divergent Constructs ?" Journal of Applied Psychology 79(6): 875–85.

Gray, N.A:B. 1990. "Capturing Knowledge Trough Top-Down Induction of Decision Trees." IEEE Expert.

Green, P., and V. Srinivasan. 1990. "Conjoint Analysis in Marketing: New Developments With Implications for Reserach and Practice." Conjoint Analysis in Marketing 54(4): 3–19.

Gregory, R. L. 2005. "Images of Mind in Brain." Word and Image 21.

Gregory, R L. 1997. "Knowledge in Perception and Illusion." Philosophical Transactions: Biological Sciences 352(1358): 1121–28.

Greving, Bert. 2009. "Messen Und Skalieren von Sachverhalten." In Methodik Der Empirischen Forschung, eds. Sönke Albers et al. Wiesbaden: Gabler, 65–78. http://link.springer.com/10.1007/978-3-8349-9121-8_5.

Guckelsberger, Ulli, and Fritz Unger. 1999. Statistik in Der Betriebswirtschaftslehre.

Wiesbaden: Gabler Verlag. http://link.springer.com/10.1007/978-3-663-05931-8.

Gussen, Lars C., Max Ellerich, and Robert H. Schmitt. 2019. "Prediction of Perceived Quality through the Development of a Robot-Supported Multisensory Measuring System." In Procedia CIRP, Elsevier B.V., 368–73. https://doi.org/10.1016/j.procir.2019.04.206.

Hackbusch, W., H. R. Schwarz, and E. Zeidler. 2003. Teubner-Taschenbuch der Mathematik Teubner-Taschenbuch Der Mathematik. ed. E. Zeidler. Wiesbaden: Vieweg+Teubner Verlag. http://link.springer.com/10.1007/978-3-322-96781-7.

Haines, Eric, and Akenie-Loeler Tomas. 2019. Ray Tracing Gems. eds. Eric Haines and Tomas Akenine-Möller. Berkeley, CA: Apress. http://link.springer.com/10.1007/978-1-4842-4427-2.

He, Kaiming, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. "Deep Residual Learning for Image Recognition." IEEE Xplore: 770–78. http://joi.jlc.jst.go.jp/JST.JSTAGE/cl/2003.428?from=CrossRef.

Hillig, Thomas. 2006. "Verfahrensvarianten Der Conjoint-Analyse Zur Prognose von Kaufentscheidungen - Eine Monte-Carlo-Simulation." Technische Universität Berlin.

Himme, Alexander. 2009a. "Conjoint-Analysen." In Methodik Der Empirischen Forschung, eds. Sönke Albers et al. Wiesbaden: Gabler Verlag, 283–98. http://link.springer.com/10.1007/978-3-322-96406-9_19.

———. 2009b. "Gütekriterien Der Messung: Reliabilität, Validität Und Generalisierbarkeit." In Methodik Der Empirischen Forschung, eds. Sönke Albers et al. Wiesbaden: Gabler, 375–90. http://link.springer.com/10.1007/978-3-8349-9121-8_25.

Hinton, Nitish Srivastava Geoffrey, Alex Krizhevsky Ilya Sutskever, and Ruslan Salakhutdinov. 2014. "Dropout: A Simple Way to Prevent Neural Networks from Overfitting." Journal of Machine Learning Research 15: 1929–58.

Homer, Pamela Miles. 2008. "Perceived Quality and Image: When All Is Not 'Rosy.'" Journal of Business Research 61(7): 715–23. https://linkinghub.elsevier.com/retrieve/pii/S0148296307002652.

Husain, Farzad, Babette Dellen, and Carme Torras. 2017. Handbook of Neural Computation Scene Understanding Using Deep Learning. 1st ed. Elsevier Inc. http://dx.doi.org/10.1016/B978-0-12-811318-9.00020-X.

Ian Goodfellow, Yoshua Bengio, Aaron Courville. 2015. "Deep Learning Book." Deep Learning.

İlhan, E.Ü., C.U. Ünal, and Y.E. Altınok. 2017. "Perceived Quality and Craftsmanship: An Engineering Approach to Trunk Lid Cutline." In Advances in Affective and Pleasurable Design, eds. W. Chung and Shin C.S. Springer International Publishing, 543–53.

Ilhan, Emel Ünlükal, and Hafize Çelik. 2017. "Perceived Quality and Craftsmanship---Common Terminology Proposal for Automotive World with Guidance of Kansei Engineering." In Advances in Affective and Pleasurable Design, eds. WonJoon Chung and Cliff Sungsoo Shin. Cham: Springer International Publishing, 391–99.

Jensen, Henrik Wann. 1996. "Global Illumination Using Photon Maps." In Proceedings of the Seventh Eurographics Workshop on Rendering, Springer Vienna, 21–30. http://link.springer.com/10.1007/978-3-7091-7484-5_3.

Jensen, Henrik Wann, and Juan Buhler. 2002. "A Rapid Hierarchical Rendering Technique for

Translucent Materials." ACM Transactions on Graphics 21(3): 576–81. https://dl.acm.org/doi/10.1145/566654.566619.

Joachims, Thorsten. 1998. "Text Categorization with Support Vector Machines: Learning with Many Relevant Features." In European Conference on Machine Learning, Springer, Berlin, Heidelberg, 137–42. http://link.springer.com/10.1007/BFb0026683.

Johnson, R. Burke, Anthony J. Onwuegbuzie, and Lisa A. Turner. 2007. "Toward a Definition of Mixed Methods Research." Journal of Mixed Methods Research 1(2): 112–33. http://journals.sagepub.com/doi/10.1177/1558689806298224.

Joongho Chang et al. 1997. "Cork Quality Classification System Using a Unified Image Processing and Fuzzy-Neural Network Methodology." IEEE Transactions on Neural Networks 8(4): 964–74. https://ieeexplore.ieee.org/document/595897/.

Joshi, Ankur, Saket Kale, Satish Chandel, and D. Pal. 2015. "Likert Scale: Explored and Explained." British Journal of Applied Science & Technology 7(4): 396–403.

Juster, N P et al. 2001. "Visualising the Impact of Tolerances on Cosmetic Product Quality." In INTERNATIONAL CONFERENCE ON ENGINEERING DESIGN,.

Kajiya, James T. 1986. "The Rendering Equation." ACM SIGGRAPH Computer Graphics 20(4): 143–50. https://dl.acm.org/doi/10.1145/15886.15902.

Kaplan, Andreas, and Michael Haenlein. 2019. "Siri, Siri, in My Hand: Who's the Fairest in the Land? On the Interpretations, Illustrations, and Implications of Artificial Intelligence." Business Horizons 62(1): 15–25. https://doi.org/10.1016/j.bushor.2018.08.004.

Kindermans, Pieter-Jan et al. 2017. "Learning How to Explain Neural Networks: PatternNet and PatternAttribution." : 1–12. http://arxiv.org/abs/1705.05598.

Kluyver, Thomas et al. 2016. "Jupyter Notebooks – a Publishing Format for Reproducible Computational Workflows." In 20th International Conference on Electronic Publishing,.

Koch, Gregory. 2015. "Siamese Neural Networks for One-Shot Image Recognition." Proceedings of the 32nd International Conference on Machine Learning 37.

Kodinariya, Trupti M, and Prashant R Makwana. 2013. "Review on Determining Number of Cluster in K-Means Clustering." International Journal of Advance Research in Computer Science and Management Studies 1(6): 2321–7782.

Koffka, Kurt. 1937. "Principles of Gestalt Psychology." Zeitschrift für Sozialforschung 6(1): 195–97.

Köhler, Wolfgang. 1967. "Gestalt Psychology." Psychologische Forschung 31.

Kohonen, Teuvo. 1982. "Self-Organized Formation of Topologically Correct Feature Maps." Biological Cybernetics 69: 59–69.

Kotsiantis, S B. 2013. "Decision Trees : A Recent Overview." Artificial Intelligence Review 39: 261–83.

Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. 2012. "ImageNet Classification with Deep Convolutional Neural Networks Alex." In NIPS'12 Proceedings of the 25th International Conference on Neural Information Processing Systems, , 1097–1105.

Kroeber-Riel, Werner, and Andrea Gröppel-Klein. 2013. Konsumentenverhalten. München: Franz Vahlen.

Lanquillon, Carsten. 2019. “Grundzüge Des Maschinellen Lernens.” In Blockchain Und Maschinelles Lernen, eds. S. Schacht and C. Lanquillon. Springer-Verlag, 89–142.

LeCun, Y, and Y Bengio. 1995. “Convolutional Networks for Images, Speech, and Time-Series.” The Handbook of Brain Theory and Neural Networks. https://www.researchgate.net/publication/2453996.

Lecun, Yann, Yoshua Bengio, and Geoffrey Hinton. 2015. “Deep Learning.” Nature 521(7553): 436–44.

Leder, Helmut, and Claus Christian Carbon. 2005. “Dimensions in Appreciation of Car Interior Design.” Applied Cognitive Psychology 19(5): 603–18.

Li, Yuhong, and Xiaofan Zhang. 2019. “SiamVGG: Visual Tracking Using Deeper Siamese Networks.” arXiv:1902.02804v2: 1–17.

Lichtenstein, Donald R, and Scot Burton. 1989. “The Relationship Between Perceived and Objective Price-Quality.” Journal of Marketing Reserach XXVI(November): 429–43.

Lieb, Helmut, Bastian Quattelbaum, and Robert Schmitt. 2008. “Perceived Quality as a Key Factor for Strategic Change in Product Development.” In 2008 IEEE International Engineering Management Conference, Estoril, Portugal: IEEE, 1–5. http://ieeexplore.ieee.org/document/4617994/.

Liu, Yili. 2003. “Engineering Aesthetics and Aesthetic Ergonomics: Theoretical Foundations and a Dual-Process Research Methodology.” Ergonomics 46(13–14): 1273–92.

MacQueen, J. B. 1967. “Some Methods for Classification and Analysis of Multivariate Observations.” In Proceedings of the Fifth Symposium on Math, Statistics, and Probability, Berkeley, 281–297.

Mayring, Philipp. 2007. “Designs in Qualitativ Orientierter Forschung.” Journal für Psychologie 15(2): 1–10. http://www.journal-fuer-psychologie.de/index.php/jfp/article/view/127.

———. 2010. “Design.” In Handbuch Qualitative Forschung in Der Psychologie, eds. G. Mey and K. Mruck. Springer Fachmedien Wiesbaden GmbH, 225–37.

McCarthy, J., Minsky, M. L., Rochester, N., & Shannon, C. E. 1955. “A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence.” AI Magazine 27(4).

McKesson, Jason L. 2012. Learning Modern 3D Graphics Programming Learning Modern 3D Graphics Programming. http://alfonse.bitbucket.org/oldtut/.

Microsoft. 2020. “Microsoft HoloLens.” https://www.microsoft.com/de-de/hololens (February 15, 2020).

Minsky, M. L. 1982. “Semantic Information Processing.” MIT Press 438.

Müller, Klaus-Robert et al. 2016. “Explaining Nonlinear Classification Decisions with Deep Taylor Decomposition.” Pattern Recognition 65: 211–22. http://dx.doi.org/10.1016/j.patcog.2016.11.008.

Muth, Claudia, and Claus Christian Carbon. 2013. “The Aesthetic Aha: On the Pleasure of Having Insights into Gestalt.” Acta Psychologica 144(1): 25–30. http://dx.doi.org/10.1016/j.actpsy.2013.05.001.

Nagamachi, Mitsuo. 2016. Kansei/Affective Engineering. 1. Edition. ed. Mitsuo Nagamachi. CRC Press. https://www.taylorfrancis.com/books/9781439821343.

Navon, David. 1981. “The Forest Revisited: More on Global Precedence.” Psychological Research 43: 1–32.

———. 2003. “What Does a Compound Letter Tell the Psychologist ’ s Mind ?” Acta Psychologica 114: 273–309.

Noble, William S. 2006. “What Is a Support Vector Machine?” NATURE BIOTECHNOLOGY 24(12): 1565–67.

Ortlieb, Stefan A, and Werner A Ku. 2020. “Fechner (1866): The Aesthetic Association Principle — A Commented Translation.” i-Perception 11(3).

OÜ, and Devsoft Baltic. 2020. “SurveyJS Library Overview.” SurveyJS Library. https://surveyjs.io/Documentation/Library (February 8, 2020).

Oude Ophuis, Peter A M, and Hans C M Van Trijp. 1995. “PERCEIVED QUALITY : A MARKET DRIVEN AND CONSUMER ORIENTED APPROACH.” Food Quality and Preference 3293(94).

Packianather, M S, and P R Drake. 2000. “Neural Networks for Classifying Images of Wood Veneer . Part 2.” Journal of Advanced Manufacturing Technology 16: 424–33.

Pan, Yanxin et al. 2017. “Deep Design: Product Aesthetics for Heterogeneous Markets.” Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining Part F1296(1): 1961–70.

Park, Hyeoun Ae. 2013. “An Introduction to Logistic Regression: From Basic Concepts to Interpretation with Particular Attention to Nursing Domain.” Journal of Korean Academy of Nursing 43(2): 154–64.

Raab, Gerhard, Alexander Unger, and Fritz Unger. 2016. Marktpsychologie Marktpsychologie. Wiesbaden: Springer Fachmedien Wiesbaden. http://link.springer.com/10.1007/978-3-658-02067-5.

Ramachandran, Prajit, Barret Zoph, and Quoc V. Le. 2017. “Searching for Activation Functions.” : 1–13. http://arxiv.org/abs/1710.05941.

Ramachandran, V. S., and R. L Gregory. 1991. “Perceptual Filling in of Artificially Induced Scotomas in Human Vision.” Nature 350(6320): 699–702.

Reinders, Heinz, and Hartmut Ditton. 2011. “Überblick Forschungsmethoden.” In Empirische Bildungsforschung, Wiesbaden: VS Verlag für Sozialwissenschaften, 45–51. http://link.springer.com/10.1007/978-3-531-93015-2_3.

Richard, Szeliski. 2005. 35 European Journal of Immunology Computer Vision: Algorithms and Applications.

Richter, Stefan. 2019a. Statistisches und maschinelles Lernen Statistisches Und Maschinelles Lernen. Berlin, Heidelberg: Springer Berlin Heidelberg. http://link.springer.com/10.1007/978-3-662-59354-7.

———. 2019b. Statistisches Und Maschinelles Lernen. Berlin, Heidelberg: Springer Berlin Heidelberg. http://link.springer.com/10.1007/978-3-662-59354-7.

Rix, Joachim, Stefan Haas, and Jose Teixeira. 1995. Virtual Prototyping IFIP-The International Federation for Information Processing. eds. Joachim Rix, Stefan Haas, and Jose Teixeira. https://link.springer.com/content/pdf/10.1007%2F978-0-387-34904-6.pdf.

Sattler, H. 2006. “Methoden Zur Messung von Präferenzen Für Innovationen.” Zeitschrift für

Betriebswirtschaftlich Forschung 54(6): 154–76.

Schmidhuber, Jürgen. 2015. “Deep Learning in Neural Networks: An Overview.” Neural Networks 61: 85–117. http://dx.doi.org/10.1016/j.neunet.2014.09.003.

Schmidt-Hieber, Johannes. 2017. “Nonparametric Regression Using Deep Neural Networks with ReLU Activation Function.” http://arxiv.org/abs/1708.06633.

Schmitt, Robert, Björn Falk, Sebastian Stiller, and Verena Heinrichs. 2015. “Human Factors in Production Systems Motives, Methods and Beyond User Diversity.” In Advances in Production Technology, ed. C. Brecher. , 187–99.

Schölkopf, Bernhard, Alexander J. Smola, Bernhard Scholkopf, and Francis Bach. 2002. Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond. MIT Press.

Schumann, Siegfried. 2018. Quantitative und qualitative empirische Forschung Quantitative Und Qualitative Empirische Forschung. Wiesbaden: Springer Fachmedien Wiesbaden. http://link.springer.com/10.1007/978-3-658-17834-5.

Semmo, Amir. 2016. “Design and Implementation of Non-Photorealistic Rendering Techniques for 3D Geospatial Data.” University of Potsdam, Germany. https://publishup.uni-potsdam.de/opus4-ubp/frontdoor/index/index/docId/9952.

Shalev-Shwartz, Shai, and Shai Ben-David. 2014. Understanding Machine Learning. Cambridge: Cambridge University Press. http://ebooks.cambridge.org/ref/id/CBO9781107298019.

Shocker, Allan D., and David A. Aaker. 2006. “Managing Brand Equity.” Journal of Marketing Research.

Simonyan, Karen, and Andrew Zisserman. 2014. “Very Deep Convolutional Networks for Large-Scale Image Recognition.” : 1–14. http://arxiv.org/abs/1409.1556.

Smilkov, Daniel et al. 2017. “SmoothGrad: Removing Noise by Adding Noise.” http://arxiv.org/abs/1706.03825.

Snoj, Boris, Aleksandra Pisnik Korda, and Damijan Mumel. 2013. “The Relationships among Perceived Quality , Perceived Risk and Perceived Product Value.” Journal of Product & Brand Management.

Springenberg, Jost Tobias, Alexey Dosovitskiy, Thomas Brox, and Martin Riedmiller. 2014. “Striving for Simplicity: The All Convolutional Net.” : 1–14. http://arxiv.org/abs/1412.6806.

Steenkamp, E M. 1990. “Conceptual Model of the Quality Perception Process.” Journal of Business Research 333: 309–33.

Stoll, T., and K. Paetzold. 2008. “Gap and Flush Visualization of Virtual Nonideal Prototypes.” In Proceedings DESIGN 2008, the 10th International Design Conference, , 681–88.

Striegel, Serena, Benjamin Schleich, Daniel Zielinski, and Sandro Wartzack. 2017. “Automotive Premium Quality Improvement by High-End-Visualization.” In Volume 2: Advanced Manufacturing, American Society of Mechanical Engineers. http://proceedings.asmedigitalcollection.asme.org/proceeding.aspx?doi=10.1115/IMECE2017-71218.

Striegel, Serena Simone, and Daniel Zielinski. 2018. “Perceived Quality of the Split-Line Design and Quality.” In Proceedings of International Design Conference, DESIGN, Dubrovnik, Croatia, 453–64.

Stylidis, K., J. Landahl, C. Wickman, and H. Johannesson. 2015. “Structuring Perceived Quality Attributes for Use in the Design Process.” In 20th International Conference on Engineering Design (ICED 15), Milan, 1–10.

Stylidis, K., Casper Wickman, and Rikard Söderberg. 2015a. “Defining Perceived Quality in the Automotive Industry: An Engineering Approach.” In Procedia CIRP, , 165–70.

———. 2015b. “Defining Perceived Quality in the Automotive Industry: An Engineering Approach.” Procedia CIRP 36: 165–70. http://dx.doi.org/10.1016/j.procir.2015.01.076.

———. 2018. “Perceived Quality Attributes Framework and Ranking Method.” engrXiv. https://engrxiv.org/p5ytr/.

Stylidis, Konstantinos et al. 2015. “Structuring Perceived Quality Attributes for Use in the Design Process.” In 20th International Conference on Engineering Design (ICED 15), Milan, 1–10.

Sundararajan, Mukund, Ankur Taly, and Qiqi Yan. 2017. “Axiomatic Attribution for Deep Networks.” http://arxiv.org/abs/1703.01365.

Szegedy, Christian et al. 2015. “Going Deeper with Convolutions.” In IEEE Xplore, , 62–63.

Tsiotsou, Rodoula H. 2014. “Perceived Quality Levels and Their Relation to Involvement , Satisfaction , and Purchase Intentions.” Marketing Bulletin (November).

Ungemach, Georg, Frank Mantwill, and Michael Rund. 2008. “Visualisation of Component Deformation for Virtual Quality Assessment.” ATZ worldwide 110(6): 76–79. https://doi.org/10.1007/BF03225019.

Varior, Rahul Rama, Mrinal Haloi, and Gang Wang. 2016. “Gated Siamese Convolutional Neural Network Architecture for Human Re-Identificatio.” arXiv: 1–18.

Veach, Eric. 1997. “Robust Monte Carlo Methods for Light Transport Simulation :: Chapter 9: Multiple Importance Sampling.” Stanford. http://graphics.stanford.edu/papers/veach_thesis/.

Völckner, Franziska, and Julian Hofmann. 2007. “The Price-Perceived Quality Relationship : A Meta-Analytic Review and Assessment of Its Determinants.” Market Lett: 181–96.

Wagersten, Ola, Casper Wickman, Lars Lindkvist, and Rikard Söderberg. 2013. “Towards Non-FEA-Based Deformation Methods for Evaluating Perceived Quality of Split-Lines.” Journal of Engineering Design 24(9): 623–39.

Wagstaff, Kiri, Claire Cardie, Seth Rogers, and Stefan Schrödl. 2001. “Constrained K-Means Clustering with Background Knowledge.” International Conference on Machine Learning ICML pages: 577–84. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.90.4624&rep=rep1&type=pdf.

Wertheimer, Max. 1923. “Untersuchungen Zur Lehre von Der Gestalt.” Zeitschrift für Psychologie und ihre Grenzwissenschaften 4.

Wickman, Casper, and Rikard Söderberg. 2004. “Comparison of Non-Nominal Geometry Models Represented in Physical Versus Virtual Environments.” Journal of Computing and Information Science in Engineering 4(3): 171.

———. 2007. “Perception of Gap and Flush in Virtual Environments.” Journal of Engineering

Design 18(2): 175–93.

Wickman, Casper, Ola Wagersten, Karin Forslund, and Rikard Söderberg. 2014. “Influence of Rigid and Non-Rigid Variation Simulations When Assessing Perceived Quality of Split-Lines.” Journal of Engineering Design 25(1–3): 1–24.

Wittmann, Stefan. 2011. “Verfahren Zur Simulation Und Analyse Der Auswirkungen Toleranzbedingter Bauteilabweichungen.” Universität Erlangen-Nürnberg. http://www.opus.ub.uni-erlangen.de/opus/volltexte/2011/2404.

Wörle, Helmut, Hans-Joachim Rumpf, and Joachim Erven. 1994. Taschenbuch Der Mathematik. Oldenbourg Wissenschaftsverlag. https://www.degruyter.com/view/title/307565.

Yoto, Ai, Tetsuo Katsuura, Koichi Iwanaga, and Yoshihiro Shimomura. 2007. “Effects of Object Color Stimuli on Human Brain Activities in Perception and Attention Referred to EEG Alpha Band Response.” Journal of Physiological Anthropology 26(3).

Zagoruyko, Sergey, and Nikos Komodakis. 2015. “Learning to Compare Image Patches via Convolutional Neural Networks.” Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition 07-12-June(i): 4353–61.

Zeiler, Matthew D., and Rob Fergus. 2014. “Visualizing and Understanding Convolutional Networks.” Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 8689 LNCS(PART 1): 818–33.

Zeithaml, Valarie A. 1988a. “Consumer Perceptions of Price, Quality, and Value: A Means-End Model and Synthesis of Evidence.” Journal of Marketing.

———. 1988b. “Of Consumer Perceptions A Means-End Value : Quality , and Model Synthesis of Evidence.” Journal of Marketing 52(3): 2–22.

Zimbardo, Philip G. 1978. Lehrbuch Der Psychologie. eds. Wilhelm F. Angermeier, Johannes C. Brengelmann, and Th. J. Thiekötter. Berlin, Heidelberg: Springer Berlin Heidelberg. http://link.springer.com/10.1007/978-3-662-08326-0.

Zöller, Susan Gretchen, and Sandro Wartzack. 2017. “Emotional Engineering Volume 5.” In Emotional Engineering, Vol.5, ed. Shuichi Fukuda. , 1–197.