Approximating JPEG 2000 wavelet representation through deep neural networks for remote sensing image scene classification

Akshara Preethy Byju, Gencer Sumbul, Begüm Demir, Lorenzo Bruzzone

Akshara Preethy Byju, Gencer Sumbul, Begüm Demir, Lorenzo Bruzzone, "Approximating JPEG 2000 wavelet representation through deep neural networks for remote sensing image scene classification," Proc. SPIE 11155, Image and Signal Processing for Remote Sensing XXV, 111550S (15 October 2019); doi: 10.1117/12.2534643

Event: SPIE Remote Sensing, 2019, Strasbourg, France

Approximating JPEG 2000 Wavelet Representation through Deep Neural Networks for Remote Sensing Image Scene Classification

Akshara Preethy Byju[1], Gencer Sumbul[2], Begüm Demir[2], and Lorenzo Bruzzone[1]

[1]University of Trento, Via Sommarive, 18 I-38123, Povo, Trento, Italy
[2]Technische Universitat Berlin, Einsteinufer 17, 10587, Berlin, Germany

ABSTRACT

This paper presents a novel approach based on the direct use of deep neural networks to approximate wavelet sub-bands for remote sensing (RS) image scene classification in the JPEG 2000 compressed domain. The proposed approach consists of two main steps. The first step aims to approximate the finer level wavelet sub-bands. To this end, we introduce a novel Deep Neural Network approach that utilizes the coarser level binary decoded wavelet sub-bands to approximate the finer level wavelet sub-bands (the image itself) through a series of deconvolutional layers. The second step aims to describe the high-level semantic content of the approximated wavelet sub-bands and to perform scene classification based on the learnt descriptors. This is achieved by: i) a series of convolutional layers for the extraction of descriptors which models the approximated sub-bands; and ii) fully connected layers for the RS image scene classification. Then, we introduce a loss function that allows to learn the parameters of both steps in an end-to-end trainable and unified neural network. The proposed approach requires only the coarser level wavelet sub-bands as input and thus minimizes the amount of decompression applied to the compressed RS images. Experimental results show the effectiveness of the proposed approach in terms of classification accuracy and reduced computational time when compared to the conventional use of Convolutional Neural Networks within the JPEG 2000 compressed domain.

Keywords: Scene classification in compressed domain, JPEG 2000, deep neural network, remote sensing

1. INTRODUCTION

Advances in satellite technology have resulted in significant growth in the volume of remote sensing (RS) data [1]. Accordingly, classification of RS image scenes, which are usually achieved by direct supervised classification of each image in the archive, has received extensive attention in RS [2, 3]. Scene classification aims to assign label to each scene of the RS images and has its wide range of applications in land use/land cover (LULC) classification, disaster management, etc. Conventional RS image scene classification/retrieval approaches rely on several low-level features such as shape, texture, spectral information [4, 5]. Recent advances in deep learning show that Convolutional Neural Networks (CNNs) lead to very high scene classification performance due to their high capability to model high-level semantic content of RS images [6, 7]. In recent years, CNN architectures such as GoogLeNet [8], CaffeNet [9] have shown to achieve state-of-the-art classification performance for RS images [10]. In [11], three strategies i.e., using pre-training, fine-tuning and using CNN as feature extractor were explored to RS scene classification problems. In [12], a scale invariant CNN was introduced to avoid discriminative information loss during scene classification that are usually obtained while using fixed-scale images. In [13], a two-tunnel CNN approach was introduced to perform scene classification for multi-source RS images. To reduce the storage size of the RS image archives, RS images are usually stored in compressed format. However, most of the existing RS image scene classification approaches in deep learning requires full decompression of images

Further author information: (Send correspondence to Akshara Preethy Byju)
Akshara Preethy Byju, e-mail: akshara.preethybyju@unitn.it
Gencer Sumbul: gencer.suembuel@tu-berlin.de
Begüm Demir, e-mail: demir@tu-berlin.de
Lorenzo Bruzzone, email: lorenzo.bruzzone@unitn.it

Image and Signal Processing for Remote Sensing XXV, edited by Lorenzo Bruzzone, Proc. of SPIE
Vol. 11155, 111550S · © 2019 SPIE · CCC code: 0277-786X/19/$21 · doi: 10.1117/12.2534643

since CNNs do not directly operate on the compressed streams of RS images [10-14]. Thus, decoding of the images is required, which is computationally demanding and time-consuming for operational RS applications on large-scale RS image archives.

Although several compression algorithms were introduced in RS such as Differential Pulse Code Modulation (DPCM), Adaptive DPCM (ADPCM), Joint Photographic Experts Group (JPEG) etc., JPEG 2000 algorithm [15, 16] has gained increasing popularity due to its multiresolution paradigm, high scalability and compression ratio. As an example, Sentinel-2 and PRISMA images are compressed with JPEG 2000 algorithm before storing them into the archives. In this paper, we mainly focus on the RS images compressed with JPEG 2000 algorithm because of its wide usage in operational RS applications. In computer vision and pattern recognition, there are few works that perform image classification in JPEG 2000 compressed domain [17, 18, 19]. However, according to our knowledge, there are no works on scene classification of compressed images that benefits from DNNs in RS.

To address the above-mentioned limitations in the existing approaches for RS image scene classification, we present a novel approach that benefits from Deep Neural Networks (DNNs) to perform scene classification of compressed RS images. The proposed approach aims to minimize the amount of decompression applied to RS images. We assume that images are compressed with the JPEG 2000 algorithm. To achieve an efficient scene classification at a fast computational rate, the proposed approach consists of two steps: i) approximating wavelet sub-bands or image; and ii) feature extraction and classification of the approximated wavelet sub-band. The proposed approach begins with approximating finer (highest) wavelet resolution sub-bands of the reversible biorthogonal filter used in the JPEG 2000 from the coarsest (lowest) resolution wavelet sub-band. To achieve this, the proposed approach uses a series of deconvolutional layers for which the wavelet sub-bands are approximated. Then, the high-level semantic content of the approximated wavelet sub-bands are learnt through a sequence of convolutional layers and finally image classification is performed. Accordingly, the proposed approach utilized the multiresolution paradigm within the JPEG 2000 compression algorithm to achieve an efficient scene classification in a time-efficient manner. Experimental results performed on a benchmark archive shows the effectiveness of the proposed approach.

The rest of this paper is organized as follows: Section 2 presents an overview of the JPEG 2000 algorithm, where Section 3 introduces the proposed approach. Section 4 describes the considered benchmark archive and provides the experimental results. Finally, Section 5 draws the conclusion of the work.

2. JPEG 2000 OVERVIEW

An overview of the JPEG 2000 compression algorithm is given in Figure 1. Initially the input image is split into several small non-overlapping rectangular blocks called *tiles*. Each *tile* from each spectral band of an image is then transformed using *Discrete Wavelet Transform (DWT)*. Successive dyadic wavelet decomposition transforms each tile component into 3-high pass (high frequency) sub bands and 1-low pass (low frequency) sub band thus, allowing for a multiresolution paradigm within the framework. JPEG 2000 compression standard supports both lossless as well as lossy coding techniques. Each of these sub bands are further sub-divided into non-overlapping blocks called *precincts* and each *precinct* is further divided as *code-blocks*. Each code-block is usually of the size 32-by-32 or 64-by-64 pixels.

Entropy coding is divided into two steps in the standard JPEG 2000 compression algorithm [16]. In *Tier-1* encoding, each code-block of each component is entropy coded using the *Embedded Block Code with Optimal Truncation (EBCOT)* that is the coding technique used in JPEG 2000 compression algorithm. For the arithmetic coding, the code blocks are represented using bit planes and these bit planes are coded from Most Significant Bitplane (MSB) to Least Significant Bitplane (LSB) using three coding passes: Significance Propagation pass, Magnitude Refinement pass and clean-up pass. Then, further a context-sensitive binary arithmetic coder encodes the obtained bits obtained after the three coding passes. This is the *Tier-1* encoding in JPEG 2000 compression algorithm. Thus, the algorithm supports bit depth scalability to any range (which is another added advantage).

In *Tier-2* encoding, the bit-stream obtained from the arithmetic coder is organized as *packets* and *layers*. *Packets* contain the bit stream organization of the code blocks and have a packet header that saves the location of the bit stream related to a particular *precincts*, where *layers* contain information about all the packets that

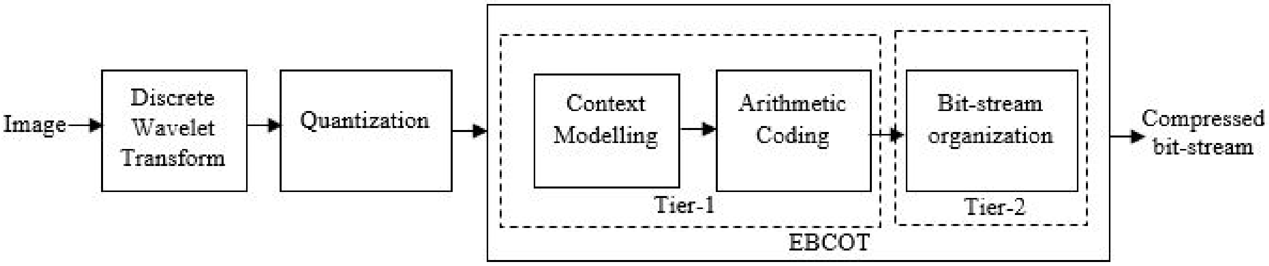

Figure 1. A general block scheme of JPEG 2000 compression algorithm.

are arranged into bit streams. *Packets* and *Layers* can be arranged using resolution, layer, component or spatial location of the image.

3. PROPOSED DNN APPROACH IN JPEG 2000 COMPRESSED DOMAIN

Let $\mathbf{X} = \{X_1, X_2, ..., X_N\}$ be an archive that consists of N compressed images, where X_i represents the i-th image. We assume that each image in $\mathbf{X}$ is compressed with JPEG 2000 algorithm. Given an image $X_i \in \mathbf{X}$, the aim of the proposed approach is to map the image to a class $y_j \in \mathbf{Y}$, where $\mathbf{Y}$ represents a set of target classes with $|\mathbf{Y}| = C$. The proposed approach initially decodes the code-streams associated to the coarsest (lowest resolution) wavelet sub-band for all the N images in the archive. The coarsest level wavelet sub-band provides the global information associated to an image. Accordingly, the proposed approach approximates the finer level sub-bands (or image itself) that provides local detailed information through a sequence of m deconvolutional layers, where m represents the number of wavelet decomposition levels used to compress a given image in the considered archive $\mathbf{X}$. Based on the approximated sub-bands, the proposed approach characterizes the features using convolutional layers with Rectified Linear Unit (ReLU) activation and two fully-connected layers to obtain the classification scores. The proposed approach reduces the amount of decompression, which is required to obtain m-th level wavelet sub-band, through approximation and performs scene classification at a reduced computational time. The proposed approach includes two steps: i) approximation of wavelet sub-bands; and ii) characterization of descriptors and scene classification. Figure 2 shows the block scheme of the proposed approach. Each step is explained in the following sub-sections.

3.1 Approximating Wavelet Sub-Bands

This step aims to approximate finer level wavelet sub-band or approximated image through a series of deconvolution layers from coarsest level wavelet sub-band. To this end, the proposed approach takes as input k^{th} binary decoded wavelet sub-bands to approximate finer wavelet sub-bands at level $(m-1)$ or the fully approximated image at level m. We can consider two scenarios while approximating wavelet sub-bands. In the first scenario, the code-stream associated with the coarsest wavelet resolution sub-band can be obtained without any decoding and approximate image through $(m-1)$ deconvolutional layers. In the second scenario, the coarsest level wavelet sub-band can be used to decode finer level wavelet sub-band to $(m-1)$ levels, which can be further used to approximate the image. When compared to the second scenario, first scenario uses only the coarsest level information to approximated the finer level wavelet sub-band or image and thus significantly improves the computational time. However, the performance of the proposed approach is expected to reduce as we use only the coarsest level information when compared to scenario 2 where decoding is performed before approximation. Let G^l be the coarsest wavelet sub-band at resolution l and A^{l-1} be the approximated wavelet sub-band at wavelet resolution level $(l-1)$. Then the proposed approach approximates the wavelet sub-band at resolution $(l-1)$ as:

$$A^{l-1} = (G^l * F^l)^T \tag{1}$$

where F^l represents the learnt deconvolution filter coefficients at level l. The choice of approximating finer resolution wavelet sub-bands using deconvolutional layers fits within the multi-resolution paradigm inherent

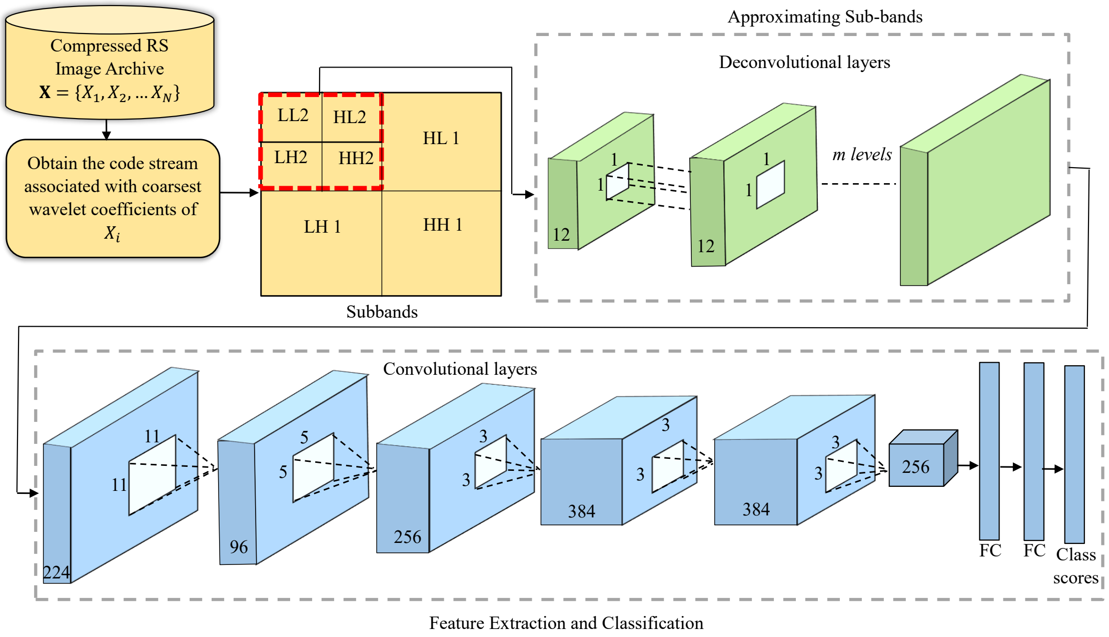
Compressed RS Image Archive
$\mathbf{X} = \{X_1, X_2, \ldots X_N\}$
Obtain the code stream associated with coarsest wavelet coefficients of X_i
LL2
HL2
LH2
HH2
HL 1
LH 1
HH 1
Subbands
Approximating Sub-bands
Deconvolutional layers
1
1
12
1
1
12
m levels
Convolutional layers
11
11
224
5
5
96
3
3
256
3
3
384
3
3
384
256
FC
FC
Class scores
Feature Extraction and Classification

Figure 2. Block scheme of the proposed DNN approach in the compressed domain.

within the JPEG 2000 compression algorithm. Two scenarios can be considered while approximating wavelet sub-bands.

The first scenario (Scenario 1) is devoted to approximation using only the coarsest level wavelet sub-band. In this scenario, the proposed approach initially takes the code-streams associated with the coarsest level wavelet sub-bands and approximates finer level wavelet sub-band or image. Figure 3 demonstrates the block scheme through which the image-level information is approximated from the coarsest level wavelet sub-band. In the figure, that the coarsest level wavelet sub-band of size 32 × 32 is considered to approximate an image of size 256 × 256. Here we directly take the code-streams associated with the coarsest level wavelet sub-band without performing further decoding of the images and therefore reduces the computational complexity. We utilize the global scale information obtained through this coarsest level sub-band to approximate the images.

The second scenario (Scenario 2) is devoted to approximation using partially-decoded finer level wavelet sub-band. In this scenario, we take the code-streams associated with the coarsest level wavelet sub-band and partially decode finer level wavelet sub-band. This finer level wavelet sub-band is used to approximate the image and then perform classification. Figure 4 shows the block scheme of this approach, where initially a 32 × 32 wavelet sub-band is considered to decode a finer level wavelet sub-band of size 64 × 64, which is further utilized to approximate the image of size 256 × 256 through two deconvolution levels. The decoded 64 × 64 wavelet sub-band holds finer detailed information and thus through approximating the image from this finer level wavelet sub-band, we can minimize the information loss. However, in this case, the computational time required to partially decode the finer level wavelet sub-band is increased, when compared to the previous scenario.

3.2 Characterization of Descriptors and Scene Classification

To obtain the features from the proposed architecture we select a CNN architecture based on AlexNet [17] that consists of five convolutional layers with 11×11, 5×5, 3×3, 3×3 and 3×3 filter sizes, respectively. There are two fully connected (FC) layers and one classification layer to obtain the class scores. To all the convolutional layers, we add zero padding and also ReLU activation. The first, second and fifth layers are followed by max-pooling layers. For the proposed approach, the weights are initialized randomly and is trained from scratch. To the best of our knowledge, no pretrained models are available for our aim and the model is trained from scratch. The

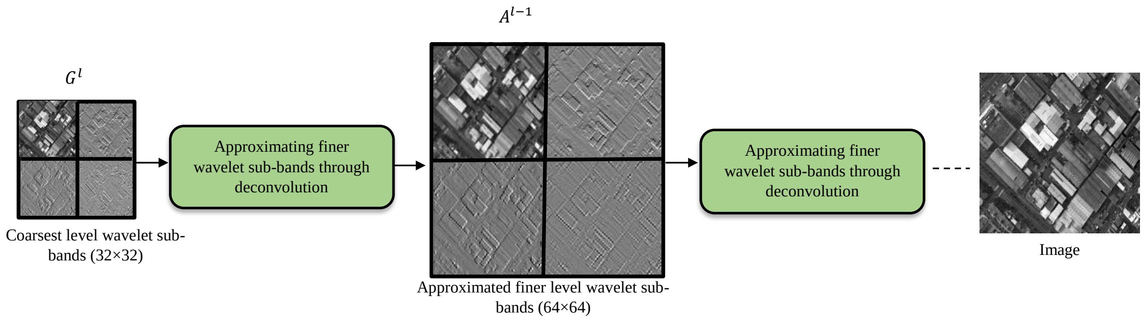
G^l
Coarsest level wavelet sub-bands (32×32)
Approximating finer wavelet sub-bands through deconvolution
A^{l-1}
Approximated finer level wavelet sub-bands (64×64)
Approximating finer wavelet sub-bands through deconvolution
Image

Figure 3. Proposed scheme when approximation is achieved by using partially decoded wavelet sub-bands.

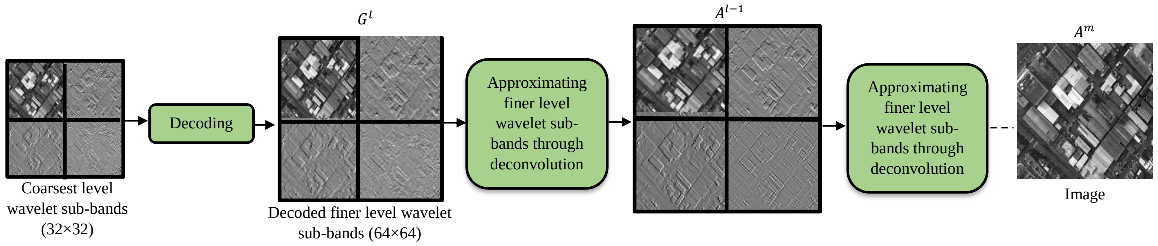
Coarsest level wavelet sub-bands (32×32)
Decoding
G^l
Decoded finer level wavelet sub-bands (64×64)
Approximating finer level wavelet sub-bands through deconvolution
A^{l-1}
Approximating finer level wavelet sub-bands through deconvolution
A^m
Image

Figure 4. Proposed scheme when when partial decoding is considered.

total loss $\mathcal{L}_{total}$ of the proposed approach is estimated using the sum of m approximation losses ($\mathcal{L}^l_{approximation}$) and a classification loss ($\mathcal{L}_{classification}$) as:

$$\mathcal{L}_{total} = \mathcal{L}_{classification} + \sum_{l=1}^{m} \mathcal{L}^l_{approximation} \tag{2}$$

$\mathcal{L}^l_{approximation}$ at level l is estimated using the Mean Squared Error (MSE) between the approximated (A^l) and decoded (D^l) wavelet sub-bands as:

$$\mathcal{L}^l_{approximation} = \sum_{i=1}^{A}\sum_{j=1}^{B} ||A^l(w[i,j]) - D^l(w[i,j])||^2 \tag{3}$$

$\mathcal{L}_{classification}$ is obtained from the cross entropy loss estimated between the original class label (y) and predicted class label (z^*) for the wavelet sub-band $w[i,j]$ of size $A \times B$ as follows:

$$\mathcal{L}_{classification} = \sum_{i=1}^{C} y_i log z^*_i \tag{4}$$

4. EXPERIMENTAL RESULTS

To assess the effectiveness of the proposed approach, a widely used NWPU-RESISC45 [20] benchmark archive that is broadly categorized into 45 different classes with 700 images each is used (see Figure 2). Each image in the benchmark archive that are of size 256 × 256 pixels with varying spatial resolution between 0.2 m and 30 m per pixel. The images are categorized into single class-labels as: airplane, airport, baseball diamond, basketball court,

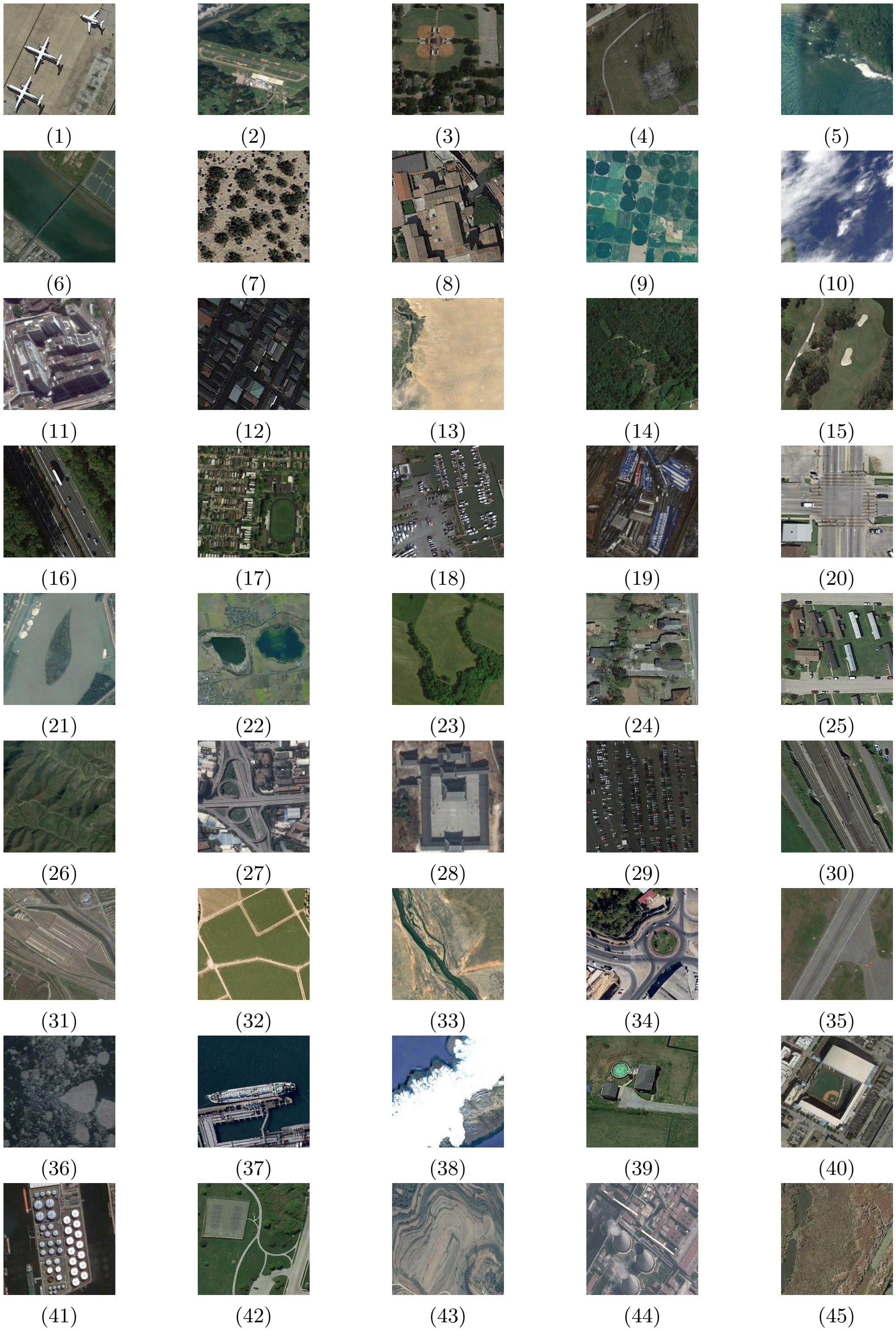

Figure 5. Example of images with their single class-labels in the NWPU-RESISC45 archive. (1) airplane (2) airport (3) baseball diamond (4) basketball court (5) beach (6) bridge (7) chaparral (8) church (9) circular farmland (10) cloud (11) commercial area (12) dense residential (13) desert (14) forest (15) freeway (16) golf course (17) ground track field (18) harbor (19) industrial area (20) intersection (21) island (22) lake (23) meadow (24) medium residential (25) mobile home park (26) mountain (27) overpass (28) palace (29) parking lot (30) railway (31) railway station (32) rectangular farmland (33) river (34) roundabout (35) runway(36) sea ice (37) ship (38) snowberg (39) sparse residential (40) stadium (41) storage tank (42) tennis court (43) terrace (44) thermal power station and (45) wetland.

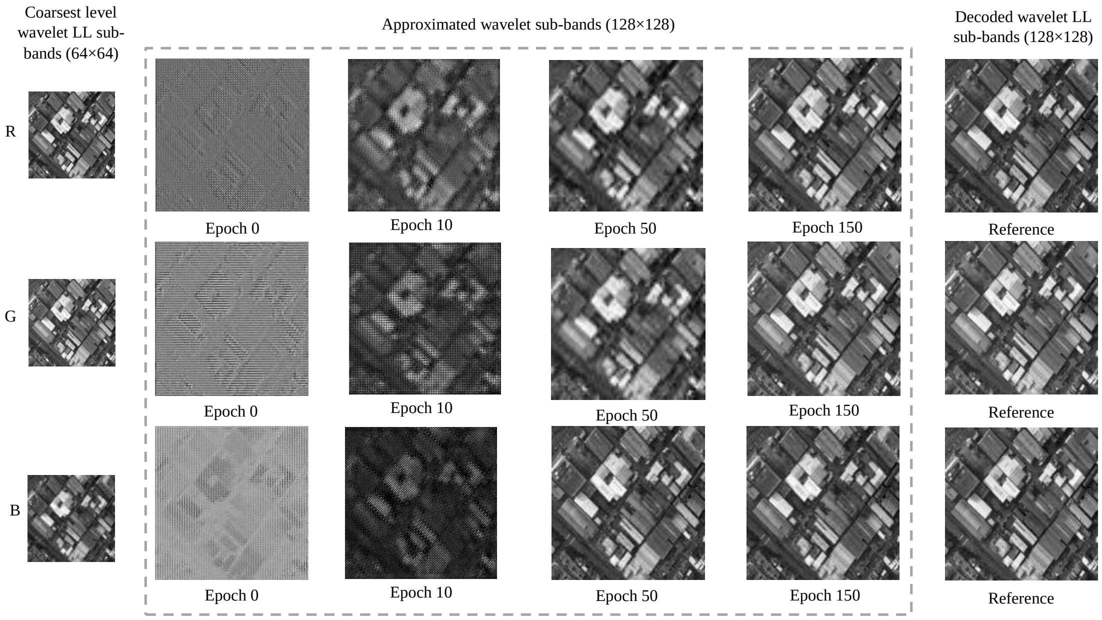

Figure 6. Qualitative results of the sub-band approximations for the LL wavelet sub-band.

beach, bridge, chaparral, church, circular farmland, cloud, commercial area, dense residential, desert, forest, freeway, golf course, ground track field, harbor, industrial area, intersection, island, lake, meadow, medium residential, mobile home park, mountain, overpass, palace, parking lot, railway, railway station, rectangular farmland, river, roundabout, runway, sea ice, ship, snowberg, sparse residential, stadium, storage tank, tennis court, terrace, thermal power station and wetland. To perform the experiments, all the images were compressed using JPEG 2000 algorithm. We considered three level wavelet decompositions of the images in the archive due to the code-block size constraint. When JPEG 2000 compression algorithm is considered, the minimum code-block size should be 32×32 or 64×64 pixel. Considering the image pixel size of the considered archive, a 32×32 pixel code-block is achieved after a three level wavelet decomposition (*m=3*). The number of deconvolutional layers is selected as 3 with 12 filters with size 1×1. From our experiments, we found the optimal learning rate for the model to be 0.001. To analyze the performance of the proposed approach, we evaluate (i) classification accuracy; and (ii) required computational time.

Thus, in our experiments we compared the proposed approach with: (i) a standard scene classification approach using DNNs (which requires full decompression of images); and (ii) a standard scene classification approach without using DNNs performing any decompression. The performance of the proposed approach was evaluated for three different cases: (i) using the coarsest level wavelet sub-band (32×32) we aim to obtain the approximated image (256×256); (ii) using the coarsest level wavelet sub-band we aim to obtain the approximated finest level wavelet sub-band (128×128); and (iii) using the finer level decoded wavelet sub-band (64×64), we aim to obtain the approximated image (256×256). Figure 3 shows the sub-band approximations of a given partially decoded LL sub-band (64×64) to approximate the finest wavelet sub-band (128×128). Table 1 shows the comparison of classification accuracies and computational time (in seconds) required for the proposed DNN and standard CNN approach. The classification accuracy (81.11%) obtained by the standard CNN on fully decompressed images is comparable to the standard CNN on the images without any decompression with a lower computational cost. The performance of the proposed approach when a coarsest level wavelet sub-band (32×32) is used to approximate the image-level information reduces to 73.27%. However, the classification phase the proposed approach requires only 5.71 seconds which is one-third required for the standard CNN that requires full decompression. Thus, one can observe that the proposed approach achieves a comparable performance when compared to the standard CNN with reduced computational time. In the second case, the

Table 1. Comparison of the experimental results obtained by the proposed DNN and the standard CNN.

Methods		**Accuracy (%)**	**Computational Time (in seconds)**	
		Test Accuracy	**Training**	**Test**
Standard CNN	Fully decompressed (256×256)	81.11	7479.69	306.24
	Without any decompression (32×32)	54.01	314.20	0.12
Proposed DNN	Approximating 3 wavelet decomposition levels (32×32) →(64×64) →(128×128) →(256×256)	73.27	8770.76	5.71
	Approximating 2 wavelet decomposition levels (32×32) →(64×64) →(128×128)	74.05	6739.87	5.68
	Approximating 2 wavelet decomposition levels (64×64) →(128×128) →(256×256)	80.09	7430.20	106.51

classification accuracy achieved by the proposed approach that is partially decoded up to finer level wavelet sub-band (64×64) to approximate the finest level wavelet sub-band (128×128) is 74.05% with a computational time of 5.68 seconds. From this, we can observe that the proposed approach achieves higher classification accuracy and lower computational when compared to the approach where the image-level information is approximated. In addition, there is also a gain in terms of computational time by 2030.89 seconds with a computational time of 5.68 seconds. Finally, if we decode finer level wavelet sub-band (64×64) to approximate image-level information the classification accuracy achieved is 80.09%, which is similar to that obtained by standard CNN approach which requires full decompression with the reduced computational time of 106.51 seconds. It is proven by the results that when the amount of required decoding time is reduced for the compressed images (using the coarsest level wavelet sub-bands) we achieve a significant reduction in the computational cost when compared to the standard CNN approach with a comparable classification accuracy during classification. Thus, approximating wavelet sub-band information and performing scene classification makes the proposed approach computationally efficient when compared to the standard CNN approaches which require full decompression of the images.

5. CONCLUSION

In this paper, we presented a novel approach that approximates wavelet sub-bands to perform scene classification in the JPEG 2000 compressed domain. The proposed approach aims at reducing the decoding time required to perform classification of images from compressed RS image archives. To this end, the proposed approach initially approximates the finer resolution wavelet sub-bands from the code-streams associated with the coarser resolution wavelet sub-bands through a series of deconvolutional layers. The number of deconvolutional layers depends on the number of wavelet decomposition levels that are used to compress the RS images. The proposed model learns the high level features associated with the approximated wavelet sub-bands through several convolutional layers. Finally, scene classification is performed using the finer features obtained from the high resolution wavelet sub-bands in an unified framework. In our experiments we analysed two different cases where (i) the images were approximated from the coarsest level wavelet sub-band and (ii) the images were approximated after performing decoding to a finer level wavelet sub-band. Experiments performed on a benchmark archive show that the proposed approach results with similar accuracies and reduced computational time during classification when compared to the state-of-the-art approaches (which work with fully decoded images). In addition, through our experiments we show that approximating finer level feature obtained after partial decoding results in similar classification accuracies when compared to the fully decoded images. As a future development, we plan to perform scene classification when the compression is achieved by deep neural networks.

References

[1] Chaudhuri, B., Demir, B., Chaudhuri, S., and Bruzzone, L., "Multilabel Remote Sensing Image Retrieval Using a Semisupervised Graph-Theoretic Method," *IEEE Transactions on Geoscience and Remote Sensing* **56**(2), 1144-1158 (2018).

[2] Chen, Y., Jiang, H., Li, C., Jia, X., and Ghamisi, P., "Deep feature extraction and classification of hyperspectral images based on convolutional neural networks," *IEEE Transactions on Geoscience and Remote Sensing* **54**(10), 6232-6251 (2016).

[3] LeCun, Y., Bengio, Y., and Hinton, G., "Deep learning," *Nature* **521**, 436-444 (2015).

[4] Yang, Y. and Newsam, S., "Geographic image retrieval using local invariant features," *IEEE Transactions on Geoscience and Remote Sensing* **51**(2), 818-832 (2013).

[5] Musci, M., QueirozFeitosa, R., Costa, G. A. O. P. and Velloso, M. L.F., "Assessment of binary coding techniques for texture characterization in remote sensing imagery," *IEEE Geoscience and Remote Sensing Letters* **10**(6), 1607-1611 (2013).

[6] Liu, Y., Zhong, Y., and Qin, Q., "Scene classification based on multiscale convolutional neural network," *IEEE Transactions on Geoscience and Remote Sensing* **56**(12), 7109-7121 (2018).

[7] Rezaee, M., Mahdianpari, M., Zhang, Y., and Salehi, B., "Deep convolutional neural network for complex wetland classification using optical remote sensing imagery," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* **11**(9), 3030-3039 (2018).

[8] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., and Rabinovich, A., "Going deeper with convolutions," in [*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*], 1-9 (2015).

[9] Jia, Y., Shelhamer, E., Donahue, J., Karayev, S., Long, J., Girshick, R., Guadarrama, S., and Darrell, T., "Caffe: Convolutional architecture for fast feature embedding," In [*Proceedings of the 22nd ACM international conference on Multimedia*], 675-678 (2014).

[10] Castelluccio, M., Poggi, G., Sansone, C., and Verdoliva, L., "Land use classification in remote sensing images by convolutional neural networks," in [*arXiv preprint arXiv:1508.00092*] (2015).

[11] Nogueira, K., Penatti, O. A., and dos Santos, J. A., "Towards better exploiting convolutional neural networks for remote sensing scene classification," *Pattern Recognition* **61**, 539-556 (2017).

[12] Liu, Y., Zhong, Y., and Qin, Q., "Scene classification based on multiscale convolutional neural network," *IEEE Transactions on Geoscience and Remote Sensing* **56**(12), 7109-7121 (2018).

[13] Xu, X., Li, W., Ran, Q., Du, Q., Gao L., and Zhang, B., "Multisource remote sensing data classification based on convolutional neural network," *IEEE Transactions on Geoscience and Remote Sensing* **56**(2), 937-949 (2017).

[14] Ma, X., Wang, H., and Geng, J., "Spectral-spatial classification of hyperspectral image based on deep auto-encoder," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* **9**(9), 4073-4085 (2016).

[15] Zhou, S., Deng, C., and Zhao, B., "Remote Sensing Image Compression : A Review," in [*IEEE International Conference on Multimedia Big Data*], 406-410 (2015).

[16] Skodras, A., Christopoulos, C., and Ebrahimi, T., "The JPEG 2000 Still Image Compression Standard," *IEEE Signal Processing Magazine* **18**(5), 36-58 (2001).

[17] Krizhevsky, A., Sutskever, I., and Hinton, G. E., "Imagenet classification with deep convolutional neural networks," in [*Advances in neural information processing systems*], 1097-1105 (2012).

[18] Descampe, A., De Vleeschouwer, C., Vandergheynst, P., and Macq, B., Scalable feature extraction for coarse-to-fine JPEG 2000 image classification., IEEE Transactions on Image Processing **20**(9), 2636-2649 (2011).

[19] Jiang, J., Guo, B. F., and Ipson, S., Shape-based image retrieval for JPEG-2000 compressed image databases, Multimedia Tools and Applications **29**(2), 93-108 (2006).

[20] Cheng, G., Han, J., and Lu, X., "Remote sensing image scene classification: Benchmark and state of the art," *Proceedings of the IEEE* **105**(10), 1865-1883 (2017).