Attention-infused 3D-SRCNN for Hyperspectral Image Super Resolution

Nour Aburaed*, Mohammed Q. Alkhatib†, Stephen Marshall‡, Jaime Zabalza§, Hussain Al Ahmad¶
*†¶ College of Engineering and IT, University of Dubai, UAE
*‡§ Department of Electronic and Electrical Engineering, University of Strathclyde, UK
Email: *nour.aburaed,†mqalkhatib@ieee.org
*nour.aburaed, ‡j.zabalza, §stephen.marshall@strath.ac.uk, ¶halahmad@ud.ac.ae

***Abstract*—In this paper, a new approach for Hyperspectral Image Super Resolution (HSI-SR) using a combination of 3D Convolutional Neural Networks (3DCNNs) and an attention mechanism, mainly Squeeze-and-Excitation (SE), is proposed. The devised method aims to generate High Resolution HSI (HR-HSI) from a single Low Resolution HSI (LR-HSI), an approach known as Single Image SR (SISR), which is a challenging task in remote sensing applications. 3D-SRCNN is utilized to extract and learn the spatial and spectral features of the input image, while the SE mechanism is employed to enhance the network's ability to model the interdependencies between the spectral bands. The proposed model is evaluated on ROSIS Pavia University and AVIRIS Botswana HSI datasets. Experimental results demonstrate that the proposed 3D-SE-SRCNN outperforms other methods in terms of both quantitative metrics and visual quality. The implementation of the proposed model is provided in this repository: https://github.com/NourO93/SISR_Library**

***Index Terms*—Hyperspectral, super resolution, convolutional neural networks, squeeze and excitation**

I. Introduction

Hyperspectral Imaging (HSI) has witnessed a great increase in popularity since the 1990s. Unlike Multispectral Images (MSI), panchromatic images (PAN), and natural images (i.e. RGB), HSI capture hundreds of narrow and contiguous spectral bands across the electromagnetic spectrum, typically from the visible to the near-infrared range. Each pixel in an HSI represents a spectrum of reflectance values at different wavelengths, providing a unique fingerprint of the materials in the scene. Thus, HSI provide an advantage because they contain information not only about the intensity of light, but also about its wavelength. This enables researchers to study and analyze the spectral properties of a scene, which eases image processing tasks, such as classification and object detection. Consequently, HSI is used in a wide range of applications [1], such as remote sensing, geology, agriculture, and medical imaging. For example, in remote sensing, HSI can be used to study vegetation health, mineral mapping, and Land Cover Land Use (LCLU).

However, the spatial resolution of HSI is often limited due to sensor tradeoff that allows capturing images either with high spectral resolution or high spatial resolution. As a result, HSI suffer from spectral mixing due to their low spatial resolution, which hinders utilizing the full potential of what HSI can offer. HSI Super Resolution (HSI-SR) aims to overcome this limitation by generating a High Resolution HSI (HR-HSI) from one or more Low Resolution (LR) images. This is achieved by exploiting the spatial correlation between different spectral bands and using advanced image processing techniques, such as Deep Learning, particularly Convolutional Neural Networks (CNNs). In the context of remote sensing, HSI-SR is a critical tool for enhancing the spatial resolution of HSI, enabling researchers to extract more detailed and accurate information from their data. Thus, researches constantly strive to enhance HSI while preserving their spectral signature from any distortion.

This paper deals with HSI spatial enhancement through Single Image Super Resolution (SISR). This is a notoriously ill-posed problem considering that it attempts to construct an HR-HSI from a single LR-HSI. 3D-SRCNN, which was previously devised in [2], has proven its efficiency in enhancing HSI while minimizing spectral distortions. It is also more lightweight compared to other CNNs that serve the same purpose. However, the network only utilizes 3D convolution, and does not fully exploit the spectral-spatial characteristics that can aid in enhancing HSI. Squeeze-and-Excitation (SE) [3] mechanism has not been explored in the context of HSI-SISR, despite being utilized for HSI classification. This study explores the use of SE in HSI-SISR through injecting it in 3D-SRCNN to exploit band dependencies, effectively creating 3D-SE-SRCNN. The proposed network is evaluated using Peak Signal-to-Noise Ratio (PSNR), Structure Similarity Index Measurement (SSIM), and Spectral Angle Mapper (SAM), in addition to qualitative inspection. The rest of the paper is organized as follows: Section II discusses the existing work in the literature, Section III presents the mathematical framework of the targeted problem, Section IV explains the proposed methodology in details along with the dataset utilized in this study, Section V demonstrates and analyzes the results, finally, Section VI summarizes and concludes the paper.

II. Related Work

According to [4], HSI-SR is broadly categorized into two classes: Fusion and SISR. Fusion methods often require auxiliary information, such as a supplementary MSI or PAN that is precisely co-registered with the LR-HSI, which is impractical. On the other hand, SISR methods do not require

extra information, which offers convenience. Nonetheless, this convenience comes at a cost, as the amount of available information to reconstruct HR-HSI is very limited, which makes HSI-SISR a notoriously ill-posed problem and one of the most popular open problems. The earliest SISR method dates back to Bicubic interpolation [5]. Although it is now considered outdated, it is still commonly used in commercial software and as an initial step in many advanced algorithms [6]. Since 2014, CNNs dominated the field of image processing. Research evidence shows that 2D-CNNs are not an adequate solution to process HSI, as they ignore the spectral aspect of this type of images. 3D-CNNs overcome this problem, and this was demonstrated in 2017 when Mei et al [7]. Devised Fully Connected CNN (FCNN) to enhance the spatial resolution of HSI, which showed promising results in terms of PSNR, SSIM, and SAM. In 2021, the idea of extending traditional 2D-CNNs that exhibit good performance on MSI to 3D was proposed. The simplest known SISR network is the Super Resolution CNN (SRCNN). The network architecture is modified by adjusting the filters' sizes to reduce the artifacts around image borders that result from padding. The network is then extended to 3D, which shows better performance than 3D-FCNN. Another 3D network called 3D-RUNet was devised in [8], which utilizes encoder-decoder architecture to reconstruct HR-HSI from LR-HSI. However, the network's depth renders it vulnerable to overfitting, particularly given the limited size of the publicly available HSI datasets. Thus, a correct balance between network depth and dataset size must be achieved. Other works in this area include [9]–[12].

Some research studies argue that 3D operations alone are not enough to exploit the spectral-spatial correlations of HSI [13]. This can be solved by combining 2D-3D operations to extract both spatial and spectral details. This can be seen in HSI classification CNNs, especially the ones that utilize attention mechanism. One way of implementing attention mechanism is known as Squeeze-and-Excitation (SE), which has proven to be efficient in HSI classification [14]. So far, this mechanism has not been explored in HSI-SISR. As 3D-SRCNN is a lightweight network that still has room for improvement, embedding SE into 3D-SRCNN while adding 2D elements into the network is an approach worthy of exploration.

III. Problem Statement

For a groundtruth HR-HSI denoted $\mathbf{Y} \in \mathbb{R}^{M \times N \times C}$, LR-HSI denoted $\mathbf{X} \in \mathbb{R}^{m \times n \times C}$ is defined as follows:

$$\mathbf{X} = \mathbf{DGY} + \mathcal{E}, \tag{1}$$

where $m << M$ and $n << N$. $\mathbf{D}$ is the downsampling operation, $\mathbf{G}$ is the blurring kernel, and $\mathcal{E}$ is the additive noise. In this study, LR-HSI is generated synthetically by applying Gaussian blur and using nearest neighbor interpolation as a downsampling operation. This is a common approach for generating LR-HSI according to [10].

HR-HSI can be estimated by minimizing the Forbenius norm of the difference between $\mathbf{Y}$ and the estimated HR-HSI denoted $\widehat{\mathbf{Y}}$ over all bands C, as follows:

$$\widehat{\mathbf{Y}}_k = \underset{\mathbf{Y}_k}{\operatorname{argmin}} \|\mathbf{DGY}_k - \mathbf{X}_k\|_F^2, \quad k = 1, \ldots, C \tag{2}$$

The complexity of an HSI cube makes this a highly non-linear optimization problem, which will be solved using the proposed model explained in the next section.

IV. Methodology

The architecture of the proposed model is shown in Figure 1. It comprises two main parts; 3D-SRCNN and the SE block. the details of these parts are explained in the next subsections.

A. 3D-SRCNN

3D convolution spans all three directions of an image; height, width, and bands. Therefore, it is an adequate solution to accommodate spectral context for HSI, as standard 2D-CNNs fail to preserve spectral fidelity. 3D convolution at position (x, y, z) can be expressed with the following equation:

$$F_{(x,y,z)} = ReLU\left(\sum_{i=1}^{M}\sum_{j=1}^{N}\sum_{k=1}^{C} K_{(i,j,k)} X_{(x+i,y+j,z+k)} + b\right) \tag{3}$$

where $F_{(x,y,z)}$ is the output feature, $X_{(x+i,y+j,z+k)}$ is the input that includes the original pixel and the neighboring pixels within the offset range (i, j, k), $K_{(i,j,k)}$ is the weight at location (i, j, k) that corresponds to the input, b is the bias, and f is the activation function. According to the literature, ReLU is the most suitable activation function for CNNs [15]. Convolution causes dimensionality reduction if the input image is not padded. This can be rectified using Transpose Convolution (TC) layer, also known as deconvolution layer, which is described as follows:

$$F'_{(x,y,z)} = ReLU\left(\sum_{i=1}^{M}\sum_{j=1}^{N}\sum_{k=1}^{C} X_{(i,j,k)} H_{(x+i,y+j,z+k)} + b\right) \tag{4}$$

The input in this case is used in place of the kernel, and it is convolved with a grid $\mathbf{H}$ of the desired size, larger than the input, where the known input values are spread across the grid and the values in between are set to zero.

The architecture of 3D-SRCNN is simplistic and consists of 3 layers:

- Patch extraction: which extracts features from $\mathbf{X}$, and it is represented by Equation 3.
- Non-linear mapping: which increases the resolution of the extracted features $\mathbf{F}$ by utilizing TC as in Equation 4.
- Reconstruction: which is a convolution layer that constructs the final $\tilde{\mathbf{Y}}$.

The overall equation of 3D-SRCNN that maps $\mathbf{X}$ to $\tilde{\mathbf{Y}}$ is described as follows:

$$\tilde{\mathbf{Y}} = SRCNN(\mathbf{X}) = F(F'(F(\mathbf{X}))) \tag{5}$$

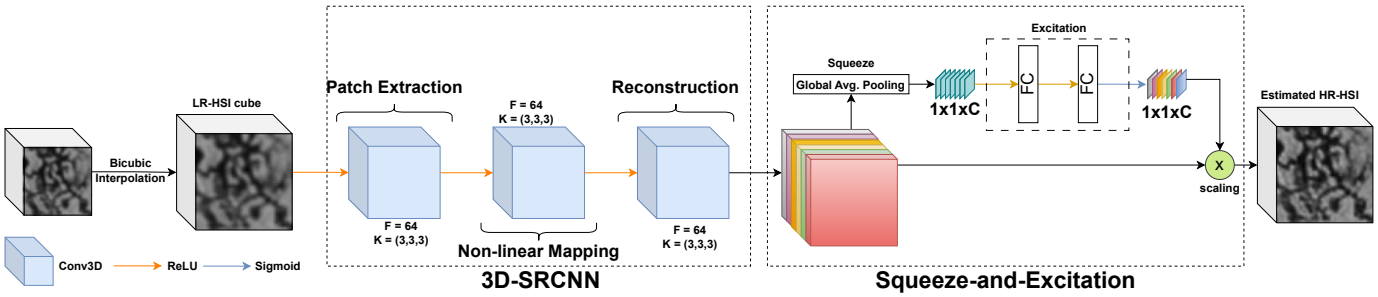

Fig. 1: Proposed 3D-SE-SRCNN for LR-HSI enhancement. The network consists of three 3D convolution layers, followed by an SE block.

B. Squeeze-and-Excitation

Squeeze-and-Excitation (SE) is a module that allows neural network models to selectively focus on informative features by adaptively recalibrating the feature map band-wise. It was first introduced by Hu et al. in 2018 [3]. The SE module consists of two main components: a squeeze operation and an excitation operation. The squeeze operation reduces the number of bands in the feature map by computing global average pooling, which aggregates spatial information across each band. The excitation operation then learns to weigh each band based on its importance, using a small neural network with a gating mechanism. Specifically, it applies a sigmoid activation function to the output of a fully connected layer that takes the squeezed features as input. This produces a band-wise scaling factor that can be applied to the original feature map, enhancing the informative bands and suppressing the less informative ones.

For any given transformation F_{tr} maps the input feature map $\tilde{\mathbf{Y}}_\mathbf{c} \in \mathbb{R}^{M \times M}$ of a particular band C to the descriptor z_c. The Squeeze procedure, denoted $F_{sq}(.)$, uses global average pooling, which converts $\tilde{\mathbf{Y}}$ to a column vector of size $1 \times 1 \times C$. The squeeze function is thus defined as:

$$z_c = F_{sq(\tilde{Y}_c)} = \frac{1}{M \times N} \sum_{i=1}^{M} \sum_{j=1}^{N} \tilde{Y}_c(i,j) \tag{6}$$

The excitation procedure is used to automatically determine the significance of each feature, amplifying those that have a bigger impact reconstructing the details of the HSI while suppressing insignificant features. The excitation function can be expressed as:

$$s = F_{ex(z,W)} = \sigma(g(z,W)) = \sigma(W_2 ReLU(W_1 z)) \tag{7}$$

where σ is the Sigmoid activation function, $W_1 \in \mathbb{R}^{\frac{C}{r} \times C}$ and $W_2 \in \mathbb{R}^{C \times \frac{C}{r}}$ are the two fully connected layers, W_1 is the dimensionality reduction layer with a dimensionality reduction ratio of r. The Sigmoid function suppresses the final output of the excitation process to a value between zero and one. The resulting output s is scaled to match the expected final height, width, and number of bands to obtain the final $\widehat{\mathbf{Y}}$.

V. EXPERIMENTS AND ANALYSIS

A. Hyperspectral Data and Implementation Details

The first dataset used in this study is Pavia University dataset, which is a commonly used HSI dataset in the remote sensing and machine learning communities [16]. It was collected by the Reflective Optics System Imaging Spectrometer (ROSIS) sensor over Pavia, Italy, and contains 103 bands with a spatial resolution of 1.3 meters per pixel.

The second dataset is Botswana, which is another widely used HSI dataset that was collected by the NASA EO-1 satellite using Hyperion sensor in 2001 [16]. This dataset consists of 1476×256 pixels with 242 spectral bands with a spatial resolution of 30 meters. Uncalibrated and noisy bands that cover water absorption features were removed, keeping only 145 bands that will be included in this study.

Since training CNNs requires a large amount of data, the datasets are divided into patches of 64×64 to increase the number of images for training and testing. The number of resulting patches for training, validation, and testing is 37, 4, and 4, respectively for Pavia University, and 76, 8, and 8, respectively, for Botswana. is Additionally, each patch is degraded using Gaussian blur and down-scaled by the required scale factor using nearest neighbor [17], [18], as explained in Section III. The resulting patch is considered as the Low Resolution HSI (LR-HSI), which will be enhanced and compared to the original ground truth HSI. In this study, all the experiments are performed on scale factors $\times 2$ and $\times 4$.

B. Results

The devised network 3D-SE-SRCNN seen in Figure 1 was developed, trained, and tested using Python's Tensorflow library. The performance is compared against Bicubic [5], 3D-FCNN [7] and 3D-SRCNN [2] in terms of PSNR, SSIM, and SAM. PSNR and SSIM measure the spatial quality of the image, while SAM measures the spectral fidelity. PSNR should be as high as possible, while SSIM should be close to 1, and SAM should be close to 0. All the networks have been trained and tested within the same environment to ensure fairness of comparison. Table I shows a summary of the testing results obtained from all the aforementioned networks. For Pavia University dataset, the proposed network shows superiority in terms of PSNR, SSIM, and SAM across scale factors $\times 2$

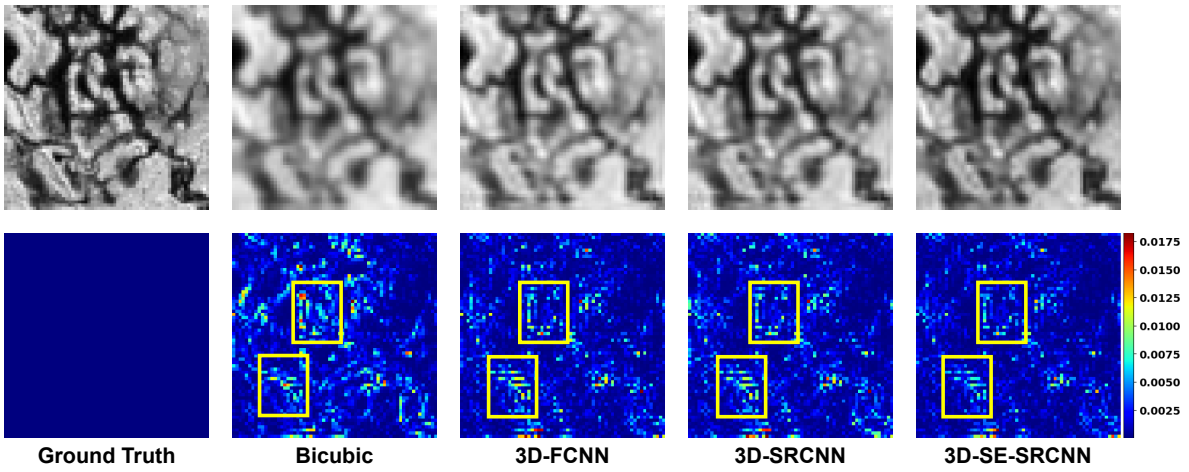

Fig. 2: Top row: Qualitative comparison of the images predicted by each approach. Bottom row: Visualization of the MSE images between each predicted result and the ground truth.

and $\times 4$. A similar observation can be made for Botswana dataset, which shows more impressive results compared to Pavia University due to the former being relatively bigger than the latter. Due to space limitation, only Botswana results are demonstrated in Figure 2. The top row shows the results of constructing band 50 using the proposed model and other benchmark methods, while the bottom row shows a visualization of the Mean Squared Error (MSE) between the predicted image the ground truth one. It can be observed that 3D-SE-SRCNN shows less errors and rectified some of the artifacts seen in the results obtained from other methods. Figure 3 shows a plot of the spectral signature of the pixel at position (61, 63), which was chosen randomly from Botswana dataset. For clarity, signatures of spectral response between bands 50 to 80 are shown. The plot clearly indicates that the signature produced by 3D-SE-SRCNN closely follows the ground truth more accurately than other methods. All of the quantitative and qualitative results prove that the proposed network was successful at enhancing HSI visually while simultaneously preserving the spectral fidelity. This demonstrates that SE mechanism indeed boosts the results of HSI-SISR networks, which is the goal of this study.

VI. Conclusion

In conclusion, this paper proposes a 3D-SE-SRCNN as an SISR technique for the purpose of enhancing HSI. The experimental results have demonstrated the effectiveness of the proposed method in generating high-quality HSI from a single LR-HSI. This approach has surpassed other methods used in this research in terms of PSNR, SSIM, SAM, as well as visual quality, indicating the potential of the proposed method for enhancing remote sensing applications. The future direction of this research includes exploring the use of different network architectures other than 3D-SRCNN to further improve the performance of the proposed method. Additionally, it is worth exploring the performance of 3D-SE-SRCNN on different degradation models in order to enhance HSI in a blind manner.

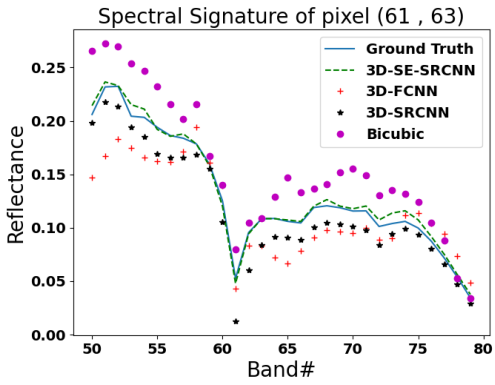

Fig. 3: Spectral signature of a pixel taken randomly at position (61, 63) from Botswana data cube. 3D-SE-SRCNN signature follows the Ground Truth one more closely compared to other methods.

TABLE I: Results summary of various methods' performance on enhancing Pavia University and Botswana datasets in terms of PSNR (dB), SSIM, and SAM (deg).

Method	PaviaU		Botswana	
	x2	x4	x2	x4
Bicubic	29.266 0.846 5.563	24.857 0.6412 8.578	32.831 0.873 3.075	29.595 0.743 4.743
3D-FCNN	30.940 0.916 7.607	25.526 0.696 9.667	34.716 0.915 1.985	31.106 0.802 3.139
3D-SRCNN	31.534 0.923 5.789	25.542 0.695 9.032	35.17 0.922 1.812	31.236 0.802 3.320
3D-SE-SRCNN	**31.794** **0.927** **5.557**	**25.641** **0.700** **8.922**	**35.546** **0.925** **1.621**	**31.335** **0.805** **3.032**

REFERENCES

[1] A. Signoroni, M. Savardi, A. Baronio, and S. Benini, "Deep learning meets hyperspectral image analysis: A multidisciplinary review," *Journal of Imaging*, vol. 5, no. 5, p. 52, 2019.

[2] N. Aburaed, M. Q. Alkhatib, S. Marshall, J. Zabalza, and H. Al Ahmad, "3d expansion of srcnn for spatial enhancement of hyperspectral remote sensing images," in *2021 4th International Conference on Signal Processing and Information Security (ICSPIS)*, 2021, pp. 9–12.

[3] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018, pp. 7132–7141.

[4] N. Aburaed, M. Q. Alkhatib, S. Marshall, J. Zabalza, and H. Al Ahmad, "A review of spatial enhancement of hyperspectral remote sensing imaging techniques," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 16, pp. 2275–2300, 2023.

[5] R. Keys, "Cubic convolution interpolation for digital image processing," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 29, no. 6, pp. 1153–1160, 1981.

[6] J. Liu, Z. Wu, L. Xiao, and X.-J. Wu, "Model inspired autoencoder for unsupervised hyperspectral image super-resolution," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–12, 2022.

[7] S. Mei, X. Yuan, J. Ji, Y. Zhang, S. Wan, and Q. Du, "Hyperspectral image spatial super-resolution via 3d full convolutional neural network," *Remote Sensing*, vol. 9, no. 11, p. 1139, 2017.

[8] N. Aburaed, M. Q. Alkhatib, S. Marshall, J. Zabalza, and H. A. Ahmad, "Sisr of hyperspectral remote sensing imagery using 3d encoder-decoder runet architecture," in *IGARSS 2022 - 2022 IEEE International Geoscience and Remote Sensing Symposium*, 2022, pp. 1516–1519.

[9] N. Aburaed, M. Q. Alkhatib, S. Marshall, J. Zabalza, and H. Al Ahmad, "A comparative study of loss functions for hyperspectral SISR," in *2022 30th European Signal Processing Conference, EUSIPCO*, in press.

[10] L. Wang, T. Bi, and Y. Shi, "A frequency-separated 3d-cnn for hyperspectral image super-resolution," *IEEE Access*, vol. 8, pp. 86 367–86 379, 2020.

[11] H. Su, H. Jin, and C. Sun, "Deep pansharpening via 3d spectral super-resolution network and discrepancy-based gradient transfer," *Remote Sensing*, vol. 14, no. 17, 2022. [Online]. Available: https://www.mdpi.com/2072-4292/14/17/4250

[12] Z. Gong, N. Wang, D. Cheng, X. Jiang, J. Xin, X. Yang, and X. Gao, "Learning deep resonant prior for hyperspectral image super-resolution," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–14, 2022.

[13] Q. Li, Q. Wang, and X. Li, "Mixed 2D/3D convolutional network for hyperspectral image super-resolution," *Remote Sensing*, vol. 12, no. 10, 2020.

[14] M. E. Asker, "Hyperspectral image classification method based on squeeze-and-excitation networks, depthwise separable convolution and multibranch feature fusion," *Earth Science Informatics*, Mar 2023. [Online]. Available: https://doi.org/10.1007/s12145-023-00982-0

[15] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016.

[16] M. Graña, M. Veganzons, and B. Ayerdi, "Hyperspectral remote sensing scenes." [Online]. Available: http://www.ehu.eus//ccwintco//index.php//Hyperspectral_Remote_Sensing_Scenes

[17] N. Yokoya, C. Grohnfeldt, and J. Chanussot, "Hyperspectral and multispectral data fusion: A comparative review of the recent literature," *IEEE Geoscience and Remote Sensing Magazine*, vol. 5, no. 2, pp. 29–56, 2017.

[18] N. Yokoya, T. Yairi, and A. Iwasaki, "Coupled nonnegative matrix factorization unmixing for hyperspectral and multispectral data fusion," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 50, no. 2, pp. 528–537, 2012.