This is a peer-reviewed, accepted author manuscript of the following conference paper: Aburaed, N., Alkhatib, M. Q., Marshall, S., Zabalza, J., & Al Ahmad, H. (2023). Bayesian Hybrid Loss for Hyperspectral SISR Using 3D Wide Residual CNN. In 2023 IEEE International Conference on Image Processing (ICIP) IEEE. https://doi.org/10.1109/icip49359.2023.10221995

BAYESIAN HYBRID LOSS FOR HYPERSPECTRAL SISR USING 3D WIDE RESIDUAL CNN

Nour Aburaed[†⋆], Mohammed Q. Alkhatib[⋆], Stephen Marshall[†], Jaime Zabalza[†], Hussain Al Ahmad[⋆]

[⋆] Department of Electronic and Electrical Engineering, University of Strathclyde, UK
[†]College of Engineering and IT, University of Dubai, UAE

ABSTRACT

Hyperspectral Imagery (HSI) has great importance in industrial remote sensing applications, such as geological exploration and soil mapping. HSI has high spectral resolution, which gives each object a unique spectral response, making them easily identifiable. Nonetheless, their spatial resolution is compromised due to sensor limitation, which hinders utilizing HSI to their full potential. This paper deals with the spatial enhancement of HSI using Single Image Super Resolution (SISR) approaches. One of the main challenges in this area of research is preserving the spectral signature of HSI while improving the spatial resolution simultaneously. To tackle this challenge, we propose a 3D Wide Residual Convolutional Neural Network (3D-WRCNN) model that effectively utilizes the principle of wide activation to enhance feature propagation throughout the network. Residual connections are also deployed to boost image reconstruction and information sharing between the layers to reduce overfitting. Furthermore, this study incorporates and demonstrates the usage of Bayesian-optimized hybrid loss function to further improve the performance of the 3D-WRCNN. The quantitative and qualitative evaluation indicate that the proposed approach prevails over other state-of-the-art approaches. The implementation of the proposed model is provided in this repository: `https://github.com/NourO93/SISR_Library`

***Index Terms*—** Hyperspectral, SISR, 3D CNN, hybrid loss function, Bayesian optimization

1. INTRODUCTION

The amount of information an image provides is expressed as "image resolution". Multispectral Imagery (MSI) are images with high spatial resolution, while Hyperspectral Imagery (HSI) have high spectral resolution. Due to sensor limitations, remote sensing images cannot be captured with high spatial resolution and high spectral resolution simultaneously. Hence, efforts have been exerted to enhance either the spectral resolution of MSI or the spatial resolution of HSI. The scope of this paper focuses on the latter, also known as HSI Super Resolution (HSI-SR). The taxonomy of HSI-SR can be broadly viewed as Fusion methods, and Single Image Super Resolution (SISR) methods. Fusion approaches, such as pansharpening [1], matrix factorization [2], and tensor-based approaches [3], assume that for a given HSI, an MSI can be captured of the same scene with high geolocation accuracy, which is not a practical assumption. Furthermore, these approaches often require knowledge about the Point Spread Function (PSF) of the sensor, which is not always attainable. Therefore, despite their excellent performance, Fusion methods are constrained by these assumptions, in addition to their high computational complexity. Conversely, SISR approaches do not require auxiliary MSI. This offers convenience, but it also introduces the challenge of dealing with a highly non-linear, ill-posed problem.

The earliest known SISR method is interpolation, such as bilinear and bicubic interpolation [4]. Even though they are no longer used as standalone approaches nowadays, they are still used within other more sophisticated approaches and Machine Learning (ML) algorithms, such as Convolutional Neural Networks (CNNs). CNNs revolutionized the field of image processing after ImageNet breakthrough in 2014 [5]. Ever since then, many SR CNNs have been developed that achieve decent performance on MSI [6, 7, 8]. Several studies established that networks that are designed for MSI-SR cannot be used directly for HSI mainly because they operate in 2D and, thus, fail to preserve the spectral fidelity of HSI [9]. This problem can be mitigated with 3D CNNs, as several studies have demonstrated [10, 11, 12]. Nonetheless, researchers still strive to overcome some of the most commonly faced challenges in the field of HSI-SISR, such as spectral distortions, overfitting due to limited dataset size. The purpose of this study is to mitigate those challenges by proposing a novel 3D residual CNN that utilizes the principles of wide activation convolution [13] to maximize spatial enhancement and minimize spectral loss. Combined with a hybrid loss function, the proposed approach surpasses other state-of-the-art algorithms. The main contributions are as follows:

1. We present an alternative approach to the hybrid spectral-spatial loss function, which we previously proposed in [14], by utilizing Bayesian Optimization Algorithm (BOA) to tune its parameters.

2. We architect a 3D Wide Residual CNN (3D-WRCNN) and integrate it with the Bayesian-optimized hybrid spectral-spatial loss function.

3. We demonstrate the advantage of the proposed approach on Pavia University dataset against other approaches quantitatively in terms of Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index Measurement (SSIM), and Spectral Angle Mapper (SAM).

The rest of the paper is organized as follows: Section 2 explains the proposed 3D-WRCNN along with the BOA hybrid loss function, Section 3 demonstrates and analyzes the advantage of 3D-WRCNN against other approaches in addition to the advantage added by the hybrid loss function, finally, Section 4 concludes this paper and states the future direction of this study.

2. METHODOLOGY

2.1. Problem Formulation

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

$$\mathbf{X} = D\mathbf{G}\mathbf{Y} + \mathcal{E}, \tag{1}$$

where $m << M$ and $n << N$. 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 bicubic interpolation as a downsampling operation. This is a common approach for generating LR-HSI according to [15].

HR-HSI can be estimated by minimizing the the loss function L between $\mathbf{Y}$ and the estimated HR-HSI $\widehat{\mathbf{Y}} = \mathbf{F}(\mathbf{X}, \varphi)$, over all bands B, as follows:

$$\widehat{\varphi} = \operatorname{argmin} L(\mathbf{Y}, F(\mathbf{X}, \varphi)) \tag{2}$$

where F is the HSI-SISR model with φ and $\widehat{\varphi}$ denoting the initial and updated parameters of the model, respectively. The complexity of HSI cube makes this a highly non-linear optimization problem, which will be solved using the proposed 3D-WRCNN.

2.2. Proposed 3D-WRCNN

The first step in the proposed network is to interpolate the LR-HSI by the required scale factor before it propagates through the network. The network overall consists of 6 convolution layers of varying sizes. The network deploys wide activation principle; that is, the amount of extracted features is expanded by as much as possible before applying ReLU. The sizes of each convolution layers are indicated in Figure 1. All the convolution layers operate in 3D in order to extract spatial features while providing spectral context simultaneously. This is considered a 3D deep CNN, and it may suffer from vanishing gradient problem and over-parametrization, especially when the dataset size is limited. This issue can be mitigated by introducing residual connections to the architecture. Two forward residual connections are added to the network. Nonetheless, this does not cancel the possibility of overfitting. It has been demonstrated that ReLU activation function can hinder the propagation of spectral features throughout the network [13]. Therefore, wide activation strategy is adapted. The 6 convolutional layers are not followed by ReLU in orde to expand the features. This boosts the performance without compromising the computational complexity. ReLU function is only added at the end of the network before producing the final output.

Batch normalization, and pixel shuffling operations have been avoided. Batch normalization has been shown to cause spatial degradation for SISR applications [16]. As for pixel shuffling, while it does not cause spatial degradation, it distorts the spectral signature of HSI [17]. Additionally, pooling is a double-edged sword that can be beneficial for SISR if used within an encoder-decoder architecture [18]. Since the proposed network is not very deep and does not follow encoder-decoder topology, pooling layer is avoided.

2.3. Bayesian Optimized Hybrid Spectral-Spatial Loss Function

Charbonier loss function shown in Equation 3 was proven to be one of the best loss functions to quantify the error between $\mathbf{Y}$ and $\widehat{\mathbf{Y}}$ [14]. However, it captures spatial errors only and does not consider the spectral context. The latter can be facilitated with Cosine Similarity (CS) loss function shown in Equation 4. This function calculates the error between the Ground Truth (GT) HSI vector $\mathbf{y}$ and the estimated HR-HSI vector $\hat{\mathbf{y}}$ at pixel location (i, j). Using either a spatial loss function or a spectral one restricts the quality of the enhanced HSI. For example, if Charbonier is used by itself, spectral distortions will potentially be introduced to the final outcome. The hybrid spectral-spatial loss function combines both aspects to boost spatial resolution while preserving the spectral one. The loss function is shown in Equation 5, where $\alpha, \beta \in [0, 1]$, and $\beta = 1 - \alpha$. Charbonier has a lower bound but it has no upper bound, such that $0 \leq L_{Ch} < \infty$. On the other hand, CS is a bounded function, such that $-1 \leq L_{CS} \leq 1$, which can be changed to $0 \leq L_{CS} + 1 \leq 2$ so that it would share the same lower bound with L_{Ch}. Note that ϵ in Equation 3 is a small value, typically 10^{-3}.

$$L_{Ch}(\mathbf{Y}, \hat{\mathbf{Y}}) = \frac{1}{MNB} \sum_{k=1}^{B} \sum_{i=1}^{M} \sum_{j=1}^{N} \sqrt{(\mathbf{Y}_{(i,j,k)} - \hat{\mathbf{Y}}_{(i,j,k)})^2 + \epsilon^2} \tag{3}$$

$$L_{CS}(\mathbf{y}, \hat{\mathbf{y}}) = -\frac{\sum_{i=1}^{M} \sum_{j=1}^{N} \mathbf{y}_i \hat{\mathbf{y}}_i}{\sqrt{\sum_{i=1}^{M} \sum_{j=1}^{N} \mathbf{y}_i^2} \sqrt{\sum_{i=1}^{M} \sum_{j=1}^{N} \hat{\mathbf{y}}_i^2}} \tag{4}$$

$$Loss = \beta L_{Ch} + \alpha(1 + L_{CS}) \tag{5}$$

α is a parameter that must be tuned in order to obtain the best possible result from the loss function. Ideally, PSNR and SSIM must be maximized, and SAM must be minimized.

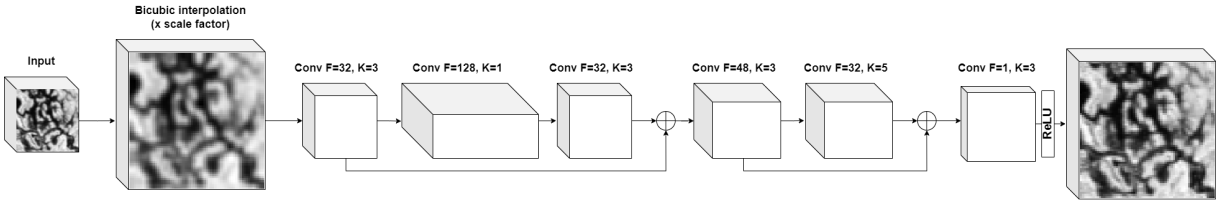

Fig. 1: Overall architecture of the proposed network with the indicated filter (F) and kernel (K) sizes, and ReLU location.

This paper focuses on maximizing PSNR as a case study. The proposed 3D-WRCNN can be considered as a computationally expensive black box objective function, with the goal of maximizing its predicted PSNR. Thus, BOA can be used a probabilistic framework to find the α that offers the global maximum. The fundamental idea behind BOA is to create a rough surrogate model of the objective function, $f(\alpha)$, and then use the model to decide on the following point to evaluate. BOA creates a method that can identify the ideal value of a non-convex function with a manageable number of evaluations since it makes use of all the data from prior assessments of $f(\alpha)$. Gaussian Process (GP) is the most commonly used probabilistic model for BOA due to its flexibility [19].

The prior over the functions and the acquisition function are the two key decisions that must be taken during the optimization process. While the acquisition function is used to find the next best point for evaluation, most likely to reduce uncertainty in the function's possible values, the prior conveys assumptions or provides information about the function being optimized. The GP denoted $G(x, y)$ is fit onto $x = \alpha$ and $y = f(\alpha)$, such that $G(x, y)$ serves as the surrogate model for $f(\alpha)$. The acquisition function then uses the GP to predict how $f(\alpha)$ varies with α in order to identify which value leads to the largest $G(x, y)$. The most commonly used acquisition function is Upper Confidence Bound (UCB) [19], which is defined as:

$$UCB(\alpha^*) = \mu(\alpha^*) + \gamma^{1/2}\sigma(\alpha^*), \tag{6}$$

where $\mu(\alpha^*)$ represents exploitation regions of a specific value α^*, $\sigma(\alpha^*)$ represents exploration regions, and γ is a parameters that balances exploitation and exploration. Often large values of γ are encouraged. In this research $\gamma = 2.0$, such that it provides balance between good values and unexplored areas. This process is iterated 50 times at maximum. Early stopper strategy is adopted, such that the optimization stops if α does not improve PSNR for 10 consecutive iterations.

3. EXPERIMENTAL ANALYSIS

3.1. Proposed 3D-WRCNN Performance

The proposed 3D-WRCNN is compared against other networks; namely 3D Super Resolution CNN (3D-SRCNN) [11], 3D Robust UNET (3D-RUNET) [18], and 3D Full CNN (3D-FCNN) [20], in addition to the traditional bicubic interpolation. All networks are trained using Mean Squared Error (MSE) loss function and Adam optimizer over 1000 epochs with 10^{-5} learning rate. To ensure fairness of comparison, training and testing procedures for all networks are done within the same environment, with each training repeated 5 times. The average results are then computed. Table 1 reveals that the proposed 3D-WRCNN outperforms all the aforementioned approaches in terms of all quantitative metrics across scale factors $\times 2$ and $\times 4$. It is worth mentioning that, with the exception of 3D-WRCNN, all approaches show sharp deterioration in SAM when moving from $\times 2$ to $\times 4$. For example, in the case of 3D-SRCNN, which showed the second best overall performance, the difference in SAM between $\times 2$ and $\times 4$ is $2.06°$. On the other hand, the deterioration for 3D-WRCNN is only by an amount of $1.52°$. This demonstrates the resilience of 3D-WRCNN with high scale factors. An interesting case to note is 3D-RUNET. This network previously showed decent performance on Botswana and Salinas datasets [18], which are bigger than Pavia University. It is a very deep network with a large number of parameters, so it overfits when trained over a small dataset.

Due to space limitation, only the qualitative results of scale factor $\times 2$ are displayed. Figure 2 shows visual comparisons between the proposed 3D-WRCNN and other approaches. The top row shows a side-by-side comparisons between the predicted results of each network (b)-(f) and the GT HSI (a). The bottom row shows a visualization of the Root MSE (RMSE) between each predicted result and the GT. Figure 2(k) produced from 3D-WRCNN indicates that it contains the least errors. The area indicatd by the yellow box shows the most distinctive differences between all the outcomes. Additionally, Figure 3 shows a plot of the spectral signature of a random pixel as generated by the proposed 3D-WRCNN and the other approaches. It can be seen that the spectral signature from 3D-WRCNN is closer to the GT one.

Table 1: Comparison between the proposed 3D-WRCNN and other approaches on Pavia University dataset.

Method	PSNR (dB)		SSIM		SAM(°)	
	$\times 2$	$\times 4$	$\times 2$	$\times 4$	$\times 2$	$\times 4$
Bicubic	29.52 ± 0.12	24.86 ± 0.15	0.846 ± 0.020	0.641 ± 0.021	5.56 ± 0.28	8.56 ± 0.29
[20]	31.79 ± 0.09	27.31 ± 0.13	0.928 ± 0.011	0.737 ± 0.015	5.03 ± 0.19	7.87 ± 0.26
[11]	31.86 ± 0.13	27.75 ± 0.25	0.930 ± 0.012	0.767 ± 0.019	4.74 ± 0.04	6.80 ± 0.09
[18]	29.88 ± 0.26	27.10 ± 0.33	0.891 ± 0.021	0.723 ± 0.025	6.97 ± 0.25	8.85 ± 0.27
Proposed	$\mathbf{32.07 \pm 0.08}$	$\mathbf{28.60 \pm 0.14}$	$\mathbf{0.934 \pm 0.002}$	$\mathbf{0.787 \pm 0.014}$	$\mathbf{4.34 \pm 0.11}$	$\mathbf{5.86 \pm 0.13}$

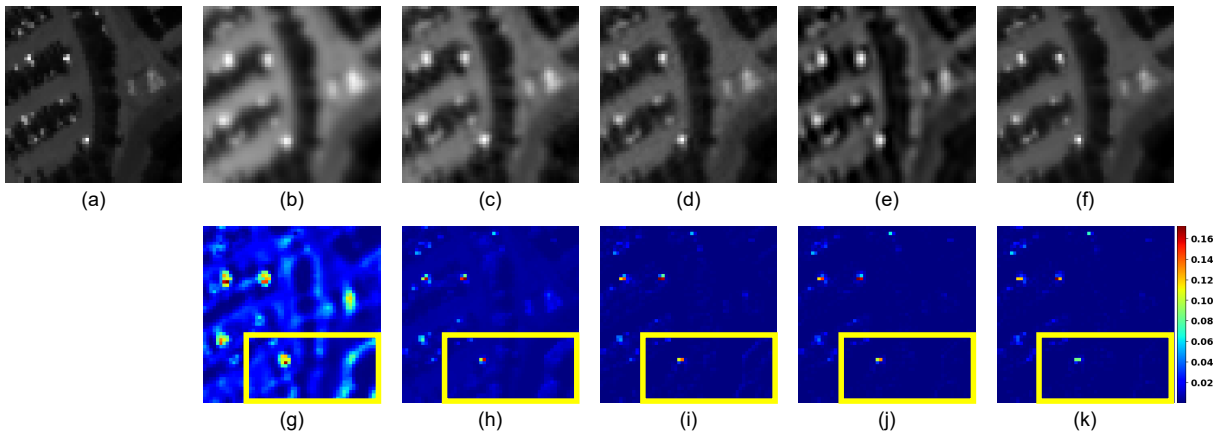

Fig. 2: The top row shows qualitative comparison of the images predicted by each approach (b) bicubic, (c) 3D-RUNET, (d) 3D-FCNN, (e) 3D-SRCNN, and (f) 3D-WRCNN, compared to the groundtruth image(a) for scale factor $\times 2$. the bottom row shows a visualization of the RMSE between each predicted result and the GT; (g) bicubic, (h) 3D-RUNET, (i) 3D-FCNN, (j) 3D-SRCNN, and (k) 3D-WRCNN.

3.2. Proposed Loss Function Performance

BOA is used to optimize α of the proposed hybrid spectral-spatial loss function. It is expected that a value of α that provides nearly equal weights between spectral and spatial features would yields the best possible results. Table 2 shows the results for scale factor $\times 2$. Overall, the results obtained from the hybrid loss function for all iterations are better than the standard loss MSE results. Iteration 14 of $\alpha = 0.597461$ lead to the highest PSNR and SSIM values. Even though the value of SAM is relatively low, it is not the absolute minimum, as the lowest value is obtained in iterations 12 and 13. This insinuates that maximizing PSNR does not necessarily minimize SAM.

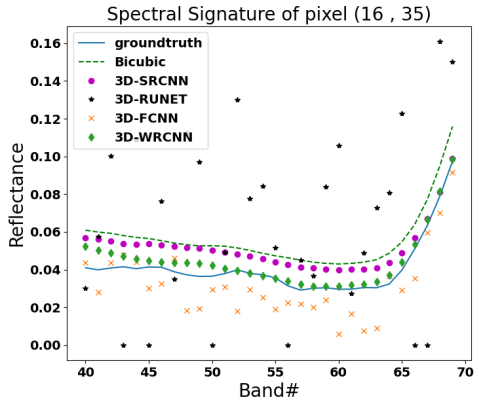

Fig. 3: Spectral signature of a random pixel at location (16, 35) produced by different methods. 3D-WRCNN follows the GT pattern more closely than other methods.

4. CONCLUSION

In this study, 3D-WRCNN was developed, trained, and evaluated in terms of PSNR, SSIM, and SAM. Qualitative and quantitative results demonstrated by the proposed network show its superiority against other approaches. Furthermore, the network is integrated with a Bayesian-optimized hybrid spectral-spatial loss function to further boost its performance in terms of maximizing PSNR. The future direction of this study invovled optimizing SSIM and SAM concurrently, while maximizing PSNR. Furthermore, data scarcity problem must be addressed in order to avoid network overfitting, potentially by using Transfer Learning or Generative Adversarial Networks [21].

Table 2: BOA of α over 15 iterations for 3D-WRCNN with scale factor $\times 2$.

Iteration	α	PSNR (dB)	SSIM	SAM (°)
0	0.970327	32.62	0.937	4.03
1	0.592509	33.02	0.937	4.07
2	0.975416	32.36	0.936	4.05
3	0.743334	32.94	0.936	4.06
4	0.727956	33.02	0.937	4.06
5	0.743356	32.94	0.936	4.07
6	0.740545	32.98	0.937	4.06
7	0.589236	33.05	0.937	4.06
8	0.585616	33.04	0.937	4.06
9	0.581712	33.05	0.937	4.05
10	0.577735	32.99	0.937	4.05
11	0.723736	32.97	0.937	4.06
12	0.732175	33.08	0.938	4.03
13	0.735368	33.04	0.937	4.03
14	**0.597461**	**33.09**	**0.938**	**4.04**

5. REFERENCES

[1] L. Loncan, L. B. de Almeida, and J. M. Bioucas-Dias *et al.*, "Hyperspectral Pansharpening: A Review," *IEEE Geoscience and Remote Sensing Magazine*, vol. 3, no. 3, pp. 27–46, 2015.

[2] H. Li, X. Feng, and D. Zhai *et al.*, "Self-Supervised Robust Deep Matrix Factorization for Hyperspectral Unmixing," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–14, 2022.

[3] M. Ding, X. Fu, and T. Huang *et al.*, "Hyperspectral Super-Resolution via Interpretable Block-Term Tensor Modeling," *IEEE Journal of Selected Topics in Signal Processing*, vol. 15, no. 3, pp. 641–656, 2021.

[4] 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.

[5] J. Deng, W. Dong, and R. Socher *et al.*, "ImageNet: A Large-scale Hierarchical Image Database," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2009, pp. 248–255.

[6] C. Dong, C. C. Loy, and K. He *et al.*, "Image Super-Resolution Using Deep Convolutional Networks," *CoRR*, vol. abs/1501.00092, 2015.

[7] J. Kim, J. K. Lee, and K. M. Lee, "Accurate Image Super-Resolution Using Very Deep Convolutional Networks," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 1646–1654.

[8] C. Ledig, L. Theis, and F. Huszàr *et al.*, "Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 105–114.

[9] N. Aburaed, M. Q. Alkhatib, and S. Marshall *et al.*, "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.

[10] S. Mei, X. Yuan, and J. Ji *et al.*, "Hyperspectral Image Super-Resolution Via Convolutional Neural Network," in *IEEE International Conference on Image Processing (ICIP)*, 2017, pp. 4297–4301.

[11] N. Aburaed, M. Q. Alkhatib, and S. Marshall *et al.*, "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.

[12] X. Dou, C. Li, and Q. Shi *et al.*, "Super-Resolution for Hyperspectral Remote Sensing Images Based on the 3D Attention-SRGAN Network," *Remote Sensing*, vol. 12, no. 7, 2020.

[13] J. Yu, Y. Fan, and J. Yang *et al.*, "Wide Activation for Efficient and Accurate Image Super-Resolution," *CoRR*, vol. abs/1808.08718, 2018.

[14] N. Aburaed, M. Q. Alkhatib, and S. Marshall *et al.*, "A Comparative Study of Loss Functions for Hyperspectral SISR," in *2022 30th European Signal Processing Conference, EUSIPCO*, in press.

[15] L. Wang, T. Bi, and Y. Shi, "A Frequency-Separated 3D-CNN for Hyperspectral Image Super-Resolution," *IEEE Access*, vol. 8, pp. 86367–86379, 2020.

[16] B. Lim, S. Son, and H. Kim *et al.*, "Enhanced Deep Residual Networks for Single Image Super-Resolution," in *IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2017, pp. 1132–1140.

[17] C. Chen, Y. Wang, and N. Zhang *et al.*, "A Review of Hyperspectral Image Super-Resolution Based on Deep Learning," *Remote Sensing*, vol. 15, no. 11, 2023.

[18] N. Aburaed, M. Q. Alkhatib, and S. Marshall *et al.*, "SISR of Hyperspectral Remote Sensing Imagery Using 3D Encoder-Decoder RUNet Architecture," in *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, 2022, pp. 1516–1519.

[19] P. I. Frazier, "A Tutorial on Bayesian Optimization," *arXiv*, vol. abs/1807.02811, 2018, stat.ML.

[20] S. Mei, X. Yuan, and J. Ji *et al.*, "Hyperspectral Image Spatial Super-Resolution via 3D Full Convolutional Neural Network," *Remote Sensing*, vol. 9, no. 11, 2017.

[21] L. Alzubaidi, J. Bai, and A. Al-Sabaawi *et al.*, "A Survey on Deep Learning Tools Dealing with Data Scarcity: Definitions, Challenges, Solutions, Tips, and Applications," *Journal of Big Data*, vol. 10, no. 1, pp. 46, Apr 2023.