ABSTRACT A critical aim of pansharpening is to fuse coherent spatial and spectral features from panchromatic and multispectral images respectively. This study proposes deep siamese network based pansharpening model as a two-stage framework in a multiscale setting. In the first stage, a siamese network learns a common feature space between panchromatic and multispectral bands. The second stage follows by fusing the output feature maps of the siamese network. The parameters of these two stages are shared across scales (across scales). The spectral information is preserved by adding appropriate skip connections from input multispectral image. Multi-level network parameters sharing mechanism in pyramidal reconstruction of pansharpened image, better preserves spatial and spectral details simultaneously. Experimental work carried out using deep siamese network in multi-scale setting (to obtain inter-band similarity among different sensor data) outperforms several latest pansharpening methods.

INDEX TERMS Pansharpening, image fusion, deep-learning, siamese networks, remote sensing, depth-of-field.

I. INTRODUCTION
Remote sensing imaging systems mainly provide two types of images; one enriched in spectral information while other with high spatial information. Owing to certain technological limitations and costs, satellite sensors lack to capture images with high spatio-spectral information. Multi spectral (MS) images are multiple low resolution color bands. The mono channel panchromatic images (PAN), carry high resolution spatial information while lacking spectral colors. It is desirable to fuse low resolution multi spectral (LR-MS) and PAN images to produce high resolution multi spectral (HR-MS) images. The task becomes challenging to transfer maximum information from two sensors simultaneously. In this regard, various state of the art schemes are discussed below. Existing methods can be specified into four main categories namely: Component substitution (CS) methods, multi resolution analysis (MRA) based on image decomposition, variational optimization (VO), and deep learning (DL) based on convolutional neural network (CNN) and residual blocks.

Wu et al. [1], proposed pansharpening algorithm based on detail injection with overflow minimization and adaptive spectral contribution of MS bands to reduce spectral distortion. However, its real time application is limited due to use of guided filter (with less detailed information). In [2], bidirectional data properties are used to carry out blind model estimation for observation and hyperspectral image fusion. The scheme produces high quality results with sharp edges.

In [3], conditional random fields are used to model state and transition functions, related to blur and spatial relationships of HR-MS (corresponding to upsampled MS, and PAN images). The scheme efficiently produces impressive results while avoiding spectral distortions. However, convergence of state function requires more time as compared to modulation transfer function (MTF). In [4], variance and regression are used to define superpixel segments for texture descriptors of PAN image. Injection coefficients are then computed from superpixels alongwith MTF-matched generalized Laplacian pyramid method. Xu et al. [5], have used feed forward CNN with progressive loss function to boost network training and maintain ground truth consistency. The scheme produces favorable results while avoiding gradient disappearance. In [6], multi linear regression is used for polynomial
estimation of relationship between PAN image details (that are missing in the MS image). Injection coefficients, based on generalized linear approach, along with MTF filter estimation produces robust results; however, sometimes the scheme produces slightly poor spectral quality.

Palubinskas [7], used LR-and PAN images to estimate energy disbalance across different bands. The image adjustment is then used to improve results. In [8], adaptive high dimensional components of MS image along with scaling factors are used to minimize spectral distortion in sharpened image. The model is quite successful, however, it requires optimal values for better results. In [9], high pass filtering is used for contrast enhancement of PAN image. Regression coefficients between MS bands and PAN image are then used to generate better clarity fused image. Hybrid pansharpening method, based on IHS and à-trous wavelet decomposition, injects spatial information in MS bands while preserving spectral details [10]. However, its computational complexity is higher as compared to other CS methods.

Imani [11], proposed multi resolution pansharpening framework using morphological operators (with different structuring elements) to extract structural details from MS and PAN images. Max-absolute fusion rule is then used to generate fused results. Jiao and Wu [12] proposed restoration framework based on blind deblurring and back projection while avoiding ringing artifacts. The scheme uses Tikhonov regularization to perform blurring filter estimation while avoiding ringing artifacts. The scheme uses kernels of different sizes and channel wise rescaled multi-scale detail injection method. In [28], proposed scheme injects spatial information in MS bands while preserving spectral distortion. In [18], CNN based proximal details along with spatial dependent weights are used as priori information regularizer. Fu et al. [19] formulated HR-MS image as variational optimization problem based on regression and gradient difference of HR-PAN and LR-MS image (for spatial preservation).

Shallow structure architecture [20], uses three convolutional layers (with different sizes to improve results of HR-MS image. In [21], target-adaptive deep fusion network is proposed for pansharpening. The residual learning and \( L_1 \) loss function along with fine tuning pass show robust performance for pansharpening as compared to PNN [20]. Wei et al. [22] proposed two staged deep residual leaning based PNN (DRPNN), using multiple convolutional layers. The scheme shows favorable results for pansharpening based on residual estimation with the provision of skip connections. Pan-GAN [23], uses generator-discriminator setting in an unsupervised deep learning framework. The scheme uses two discriminators to preserve spectral and spatial details of MS and PAN images in resultant HR-MS image. Likely, Zhou et al. [24] used QNR (quality with no reference) based loss function discriminators along with for pansharpening. These methods produce good quality results without ground truth information. However, down-sampling methods, affect the performance evaluation of down-sampled HR-MS images, and may be further explored to avoid distortions.

Dian et. al [25], formulated reconstruction of high-resolution hyperspectral image (HR-HSI) as optimization solving problem through deep learning. Residual learning priors are returned to construct the target HR-HSI. However, the performance can be enhanced using nonlocal and low rank prior information. In [26], dilated and concatenated feature maps in a multi-scale setting are used to improve spatial content in target HR-MS image. In addition, the proposed scheme removes the batch normalization thus improves the time complexity of the network. Hu et. al [27], incorporates super resolution method in a deep learning framework to generate HR-MS image from LR-MS image. PAN sharpened image is then obtained from HR-MS image and details (extracted using super resolution and total loss methods). However, the scheme can be further improved by exploring multi-scale detail injection method. In [28], proposed scheme uses kernels of different sizes and channel wise rescaled maps to improve feature extraction. The scheme uses channel attention mechanism along with feature extraction and residual learning in a multi-scale fashion to improve network convergence time.

Recently, Shamsolmoali et al. [29] proposed a multi scale architecture with multi patch learning and gradient norms for robust object detection. The cross-scale connections enable adaptive channel selection and sizes (of feature maps) thus improves computational efficiency. Wang et al. [30], used wavelet decomposition, three dimensional CNN, and dimension amplification to improve super-resolution. The scheme can further be improved by following Heisenberg uncertainty principle (to overcome shortcomings of wavelet transform).
Similarly, rotation feature pyramid framework [31], uses multibox detection and region proposal network to improve region of interest and object localization. Scale adaptation can further enhance detection capability (focusing small objects) in aerial scenes. Yang et al. [32], uses upsampled MS image for spectral correction of high pass components. PanNet addresses the shortcoming of resNet while preserving spectral and spatial information. VO and DL based method [33], uses convolutional sparse encoding to extract shift invariant features in a variational model. Pansharpened image synthesis step is then merged with the training phase along with traditional optimization methods to boost fusion performance.

To overcome the limitations of the existing schemes, proposed scheme synchronously transfers spatial and spectral details in resultant image. The proposed siamese fusion network bank is initialized with UP-MS and down-sampled PAN images. The siamese networks bank then learns inter-band similarity among different MS and PAN bands to maximize obtainable details for fusion. These maps are passed to global feature synthesis step and then merged with the training phase along with traditional optimization methods to boost fusion performance.

The remaining part of the paper is structured as follows: Proposed system model and mathematical formulations are discussed in section 2. Demonstration of simulation results and findings of the experiments are discussed in section 3. Finally, section-4 concludes the paper along with some future directions.

II. PROPOSED SCHEME

Proposed model is divided into a two-stage framework. In first stage, a siamese network learns a common feature space between panchromatic and multispectral bands. The second stage is followed by fusing the output feature maps of the siamese network. The parameters of these two stages are shared across scales to add spatial information consistently across scales. The spectral information is preserved by adding appropriate skip connections from input multispectral image.

A. SIAMESE NETWORK BLOCK

Let the LR-MS image be represented by $I^M$ where $m \in \{1...M\}$ designates each of multi-spectral bands. The captured MS images have quarterly resolution as of PAN images. To make originally captured images workable, the PAN image is downsampled with a factor of 2 given as,

$$X^P = I^P \downarrow 2 \quad (2)$$

where $\downarrow$ shows convolution operator while $g$ represents linear interpolation [34] filter bearing symmetric and separable properties. Let $I^m$ and $X^m$ represent $m^{th}$ band in $I^M$ and $X^M$ respectively and the PAN image be represented by $I^P$. The captured MS images have quarterly resolution as of PAN images. To make originally captured images workable, the PAN image is downsampled with a factor of 2 given as,

$$X^P = I^P \downarrow 2 \quad (2)$$

Siamese network imitates deep fusion network to obtain corresponding stacked scale components $S_o$. To minimize the spectral and spatial distortions, global and local features, are jointly processed. The two-stage proposed framework uses
FIGURE 2. Representation of proposed siamese network based deep fusion framework.
siamese networks in a multi-scale setting to generate the fused target. For each scale \( \omega = \{ 1, 2, \ldots, \Omega \} \) the \( X^M_\omega \) and \( X^P_\omega \) (compatible in size) are cascaded and passed to the local feature fusion block to obtain corresponding stacked scale components \( S_\omega \). For each scale stacked input \( S_\omega \) (passed to the deep leaning framework) is defined as,

\[
S_\omega = [X^F_\omega, X^M_\omega]
\]  

(3)

The scale-space approximation of the target HR-MS image is given as,

\[
X^m_{\omega+1} = (\text{FuseNet}(S_\omega) + X^M_\omega) \dagger_2
\]  

(4)

where \( \text{FuseNet}(S_\omega) \) provides a deep convolutional mapping as shown in fig 2. The same network parameters are shared accross the different scales. The framework uses multi-spectral connections \( X^M_\omega \) for interpolated estimation of associated input image. Cascaded siamese network and \( \text{FuseNet}() \) are then recursively implemented using \( X^m_{\omega+1} \) and \( X^P \). The final reconstructed HR-MS image can be represented as,

\[
Y = \text{FuseNet}(S_\Omega) + X^M_\omega
\]  

(5)

where, \( \Omega \) indicates the desired scale for the fused target.

B. DEEP FUSION NETWORK

In order to minimize the spectral and spatial distortions, proposed residual learning framework simultaneously processes the global and local features. The proposed fully convolutional \( \text{FuseNet} \) is shown in figure 3. The output \( Y_j \) at \( j^{th} \) layer is calculated as,

\[
Y(j) = \alpha \left( W(j) * Y(j-1) + \delta(j) \right)
\]  

(6)

Here \( \alpha \) represents activation function while \( W(j) \) and \( \delta(j) \) indicate \( j^{th} \) layer associated filter weights and bias respectively. The first layer formulates shallow extraction of \( \omega^{th} \) input stack \( S_\omega \), i.e., \( Y(0) = S_\omega \). The extracted output \( Y(1) \) is then passed onto the stack of \( R \) residual learning blocks (each of which learns local features for fusion).

C. LOCAL FEATURES MAP ESTIMATION

Each residual learning block estimates local feature fusion map by linear weighting of the cascaded channels. The input to the residual learning block is sequentially passed through convolutional layers with filter size, \( f_s = 5 \times 5 \). The concatenated outputs from subsequent layers, along with block input, are then passed through final convolution layer using \( f_s = 1 \times 1 \) to estimate linearly weighted feature map.

Since the filter weights and biases are shared across \( R \) blocks, the output of \( j^{th} \) layer at \( i^{th} \) local fusion block be determined as \( Y_r(j) \), i.e.,

\[
Y_r(j) = \alpha \left( W_r(j) * Y_r(j-1) + \delta_r(j) \right)
\]  

(7)

The third layer in arbitrary local feature block can be represented as,

\[
Y_r(3) = [Y_r(0), Y_r(1), Y_r(2)]
\]  

(8)

D. GLOBAL FEATURES MAP ESTIMATION

The shallow feature map \( (Y_1) \) provides skip connection and passed as an additional input to each local fusion block, i.e.,

\[
Y_r(j) = Y_{r-1}(j) + Y(1), \quad \text{for} \ j > 1
\]  

(9)

Estimated local feature maps are then used for global feature map estimation. The concatenated outputs from \( R \) residual blocks are expressed as follows,

\[
Y_R = [Y_1, \ldots, Y_R]
\]  

(10)

Feature maps, \( Y_R \), are convolved using \( f_k = 1 \times 1 \) and passed from another convolutional layer using \( f_k = 5 \times 5 \) to generate \( M \) channel output. The provision of skip connections is useful for model convergence. Deep architectures with skip
Ground truth image HR-MS is downsampled to compute loss where, \( G_{\omega} \) for global fusion network, ensure better transfer of spectral information. Similarly, across scales parameter sharing block parameters are shared across scales to better learn spatial information. The difference at next stage is that the eight MS bands are again upsampled to make them workable with the original full scale HR PAN image. Loss function uses the ground truth HR-MS image in a multi-scale setting to preserve spectral and spatial information while avoiding distortions in the fused output. In this regard, comparisons are being made with latest deep learning models including CNN based Pansharpening (PNN) [20], enhanced PNN named as PNN+ [21], Deep Residual PNN (DRPNN) [22] and PanNet [32]. In addition to deep learning models, recently proposed Variational Pansharpening with Local Gradient constraints (VPLG) [19] is also included in comparison. The pre-trained models of all of these schemes are already available online.

### A. 8-BAND PANSHARPENING

The presented work considers pansharpening using PAN images and LR-MS images having 8 bands. Over 430 image pairs from WV-2 [37] and WV-3 [38] datasets are used in this evaluation. Satellite images provided by these sensors WV-2 and WV-3 have different spatial resolutions having size 512 × 512. Nearly 250 samples belong to WV-2 while remaining are captured by WV-3. Nearly one-third samples (by making random selection) of the acquired dataset are used for training phase, and remaining samples are used during the testing phase. Despite these sensors provide different-type of imagery, proposed model learns network parameters for both sensors simultaneously.

### B. REDUCED SCALE FUSION EVALUATION

The proposed method is compared for both reduced-scale and full-scale pansharpening assessment. Image quality at high and low resolutions are closely related [43]. As groundtruth HR-MS images are not available, fusion quality assessment is carried out while considering the low resolution MS image as ground truth \( G_{\omega} \). Similar to [32] PAN and MS images are downsampled and fused output is then compared at the resolution of available MS image.

### III. EXPERIMENTS & EVALUATIONS

To fully assess the robustness of the proposed scheme a large number of experiments are carried out using Worldview-2 (WV-2) [37] and Worldview-3 (WV-3) [38] datasets. Fusion quality of pansharpened image essentially depends on inclusion of spectral and spatial information while avoiding distortions in the fused output. In this regard, comparisons are being made with latest deep learning models including CNN based Pansharpening (PNN) [20], enhanced PNN named as PNN+ [21], Deep Residual PNN (DRPNN) [22] and PanNet [32]. In addition to deep learning models, recently proposed Variational Pansharpening with Local Gradient constraints (VPLG) [19] is also included in comparison. The pre-trained models of all of these schemes are already available online.

#### TABLE 1. Network parameters and average computational time.

| Scheme     | No. of parameters | Time(min:sec) |
|------------|-------------------|---------------|
| Proposed   | 135 \( K \)       | 0:21          |
| VPLG [19]  | /                 | 4:00          |
| PNN+ [21]  | 90 \( K \)        | 0:28          |
| DRPNN [22] | 1.64 \( M \)      | 2:23          |
| PanNet [32]| 83 \( K \)        | 1:31          |
| PNN [20]   | 125 \( K \)       | 0:18          |

Connections omit some layers in the neural network and feed the output of one layer as input (to next layers). Network parameters of \( \text{FuseNet} \) are optimized by minimizing the \( L_2 \) norm loss function defined as,

\[
L_2 = \sum_{\omega} \| G_{\omega}^M - Y_{\omega}^M \|^2 \tag{11}
\]

where, \( G_{\omega}^M \) is the ground truth HR-MS image and \( Y_{\omega}^M \) is the approximation of target output at corresponding scale \( \omega \). The ground truth image HR-MS is downsampling to compute loss function

\[
G_{\omega-1}^M = G_{\omega}^M \downarrow_2 \tag{12}
\]
Figure 4 shows visual illustration of results at different scales generated by the proposed fusion framework. Visual bands of low resolution UP-MS bands are shown in figures 4 (a) & 4(d). Similarly, down-sampled PAN and full scale HR-PAN images are shown in figures 4 (b) & 4(e). Corresponding intermediate and final HR-MS images generated by the proposed network are shown in figure 4(c) & 4(f). Figure 5 presents an example of reduced scale pansharpening for visual illustration using WV-3 satellite image pairs. Figure 5(a) shows reference MS image while figures 5(b)-(g) present results for the compared schemes and the proposed model. Corresponding differences (for each scheme and reference image) are shown in figure 6. Among the proposed schemes, DRPNN [22], (figure 5d) and VPLG [19], (figure 5f), produce visually better quality results. However, blurring effect can be seen in some areas of produced results. Further DRPNN [22], suffers from spectral distortion. PANnet [32], (figure 5e) renders pleasant output, even so, the results are slightly over-sharpened. PNN+ [21], (figure 5c) result shows some lack of clarity and suffers from blurring artifacts with some distortions. Similarly, PNN [20], (figure 5b) suffers from color distortion. In comparison to

and $Q_{SCC}$ higher values correspond to good quality fusion while lower values indicate poor quality fusion. The metric values range between $[-1, 1]$ for $Q_{UIQI}$ and $Q_{SCC}$ where 1 tends to be the best value. Likely, metric values range between $[0, \infty]$ for $Q_{SAM}$ and $Q_{ERGAS}$ where 0 being the best value.

### TABLE 2. Fusion quality assessment at reduced scale. Finest metric values corresponding to best fusion are bold faced.

| Scheme          | Mean values for WV-3 sensor data | Mean values for WV-2 sensor data |
|-----------------|----------------------------------|----------------------------------|
|                 | $Q_{SAM}$ [39] | $Q_{UIQI}$ [40] | $Q_{SCC}$ [41] | $Q_{ERGAS}$ [42] | $Q_{SAM}$ [39] | $Q_{UIQI}$ [40] | $Q_{SCC}$ [41] | $Q_{ERGAS}$ [42] |
| Proposed        | 3.7488           | 0.9272            | 0.9415            | 2.6190            | 3.5614           | 0.8114            | 0.9096            | 2.7605            |
| VPLG [19]       | 4.3710           | 0.9117            | 0.8944            | 3.2945            | 3.9255           | 0.7479            | 0.8771            | 3.1048            |
| PNN+ [21]       | 9.4245           | 0.6385            | 0.4968            | 7.7605            | 5.7245           | 0.7193            | 0.8167            | 4.2630            |
| DRPNN [22]      | 7.0739           | 0.8269            | 0.8331            | 5.1512            | 5.1577           | 0.7569            | 0.8539            | 4.0459            |
| PanNet [32]     | 4.8948           | 0.8607            | 0.8674            | 4.2456            | 4.7030           | 0.7458            | 0.8394            | 3.9146            |
| PNN [20]        | 8.3305           | 0.7711            | 0.7888            | 7.3133            | 5.5836           | 0.7300            | 0.8440            | 4.0532            |
existing schemes, the proposed framework (figure 5g) outputs good quality fused image that is close in appearance to the reference image.

Quantitative comparison of existing and proposed schemes is provided in Table 2. Table 2 reports mean values for pansharpened test images using WV-3 and WV-2 satellite data. The obtained metric values illustrate that proposed method outperforms latest pansharpening methods while minimizing spatial and spectral distortions simultaneously.

**C. FULL SCALE FUSION EVALUATION**

The learnt model is also tested for full scale pansharpening. In the absence of ground truth, the input LR-MS and PAN images are used as spectral and spatial references respectively. In case of full reference pansharpening evaluation, no-reference quality metrics are used for assessment based on degraded PAN image. In this regard, spectral and spatial distortions denoted by $D_{\lambda}$ [44] and $D_{s}$ [44] are computed. Additionally, quality with no-reference $QNR$ [45], is also used to estimate global distortion. Lower values of $D_{\lambda}$ and $D_{s}$ indicate better quality results (with minimum distortions). However, higher value of $QNR$ indicates better fusion for target HR-MS image.

Figure 7 illustrates full scale pansharpening case using WV-2 sensor images. Input PAN and interpolated MS image are shown in figure 7a and 7b respectively. Output generated

![Figure 5](image5.png)

**FIGURE 5.** An illustration of reduced scale pansharpening using WV-3 sensor images.

![Figure 6](image6.png)

**FIGURE 6.** Difference images corresponding to existing schemes and proposed model using WV-3 sensor images shown in figure 5.

**TABLE 3.** Fusion quality assessment at full Scale. Finest metric values showing minimum distortions are bold faced.

| Scheme       | Mean values for WV-3 sensor data | Mean values for WV-2 sensor data |
|--------------|----------------------------------|----------------------------------|
|              | $Q_{\lambda}$ | $Q_{s}$ | $QNR$ | $Q_{\lambda}$ | $Q_{s}$ | $QNR$ |
| Proposed     | 0.0423        | 0.0309  | 0.9282 | 0.0732        | 0.0342  | 0.8915 |
| VPLG [19]    | 0.0462        | 0.0132  | 0.9412 | 0.1661        | 0.0335  | 0.8100 |
| PNN [21]     | 0.1598        | 0.0558  | 0.8410 | 0.1624        | 0.0839  | 0.7334 |
| PanNet [32]  | 0.0576        | 0.0402  | 0.9058 | 0.1719        | 0.0799  | 0.7726 |
| DRPNN [22]   | 0.0662        | 0.0661  | 0.8742 | 0.2207        | 0.1192  | 0.7043 |
| PNN [20]     | 0.1444        | 0.0870  | 0.8221 | 0.2189        | 0.1318  | 0.6908 |

Table 3
FIGURE 7. An illustration of full scale pansharpening using WV-2 sensor images.

by PNN [20], (figure 7(c)), generates reasonable output, however, the scheme suffers from color distortion. PNN+ [21] (figure 7d) has better colors, however, suffers from artifacts, which are quite evident in the roof top areas. DRPNN [22] (figure 7e) produces good spatial features (edges are preserved) however spectral quality is suffered and several objects on the right side of the image are not easily distinguishable. PANnet [32] (figure 7f) renders good quality easily visible to distinguish along with colors are preserved. Still some less spectral details are included along with a slightly over-sharpened image. In resultant image of VPLG [19], (figure 7g), color details are preserved to a great extent; but suffers from some artifacts. In comparison to existing schemes, the proposed model (figure 7h) renders good quality fused image in all aspects.

Table 3 presents measures for spectral, spatial and global distortions. In comparison to other schemes, VPLG [19], and proposed scheme have better values. Corresponding table data for these schemes are similar. However, careful observation of metric values indicate that VPLG [19], tends to minimize spectral distortion in terms of $Q_s$, while the proposed model better minimizes spatial distortion, $Q_s$.

IV. CONCLUSION

This work presents a pansharpening framework employing siamese network in a multi-scale setting named (DSN-PAN). Siamese network extracts feature maps based on inter-band similarities among adjacent multispectral and enriched panchromatic bands. A local global fusion network generates pansharpened image at each level in the Gaussian pyramid with shareable learning parameters. Fusion performance of the proposed method validates its superiority for reduced and full scale pansharpening. In our future work, we may explore different backbones for the proposed framework.

REFERENCES

[1] L. Wu, Y. Yin, X. Jiang, and T. C. E. Cheng, “Pan-sharpening based on multi-objective decision for multi-band remote sensing images,” Pattern Recognit., vol. 118, Oct. 2021, Art. no. 108022.
[2] W. Wang, X. Fu, W. Zeng, L. Sun, R. Zhan, Y. Huang, and X. Ding, “Enhanced deep blind hyperspectral image fusion,” IEEE Trans. Neural Netw. Learn. Syst., early access, Aug. 30, 2021, doi: 10.1109/TNNLS.2021.305543.
[3] Y. Yang, H. Lu, S. Huang, Y. Fang, and W. Tu, “An efficient and high-quality pansharpening model based on conditional random fields,” Inf. Sci., vol. 553, pp. 1–18, Apr. 2021.
[4] H. Hallabia, H. Hamam, and A. B. Hamida, “A context-driven pansharpening method using superpixel based texture analysis,” Int. J. Image Data Fusion, vol. 12, no. 1, pp. 1–22, Jan. 2021.
[5] H. Xu, Z. Le, J. Huang, and J. Ma, “A cross-direction and progressive network for pan-sharpening,” Remote Sens., vol. 13, no. 15, p. 3045, Aug. 2021.
[6] P. Addesso, R. Restaino, and G. Vivone, “An improved version of the generalized Laplacian pyramid algorithm for pansharpening,” Remote Sens., vol. 13, no. 17, p. 3386, Aug. 2021.
[7] G. Palubinskas, “Model-based image adjustment for a successful pan-sharpening,” 2021, arXiv:2103.03062.
[8] E. K. Özay and B. Tunga, “A novel method for multispectral image pansharpening based on high dimensional model representation,” Expert Syst. Appl., vol. 170, May 2021, Art. no. 114512.
[9] G. Rohith and L. S. Kumar, “Pansharpening for better spectral and spatial clarity,” in Proc. 5th Int. Conf. Comput., Commun. Secur. (ICCCS), Oct. 2020, pp. 1–6.
[10] S. M. A. Wady, Y. Bentoutou, A. Bengerman, A. Bounoua, and N. Taleb, “A new IHS and wavelet based pansharpening algorithm for high spatial resolution satellite imagery,” Adv. Space Res., vol. 66, no. 7, pp. 1507–1521, Oct. 2020.
[11] M. Imani, “A multi-scale transform method based on morphological operators for pansharpening,” *AUT J. Model. Simul.*, vol. 52, no. 1, p. 10, 2020.

[12] J. Jiao and L. Wu, “Image restoration for the MRA-based pansharpening method,” *IEEE Access*, vol. 8, pp. 13694–13709, 2020.

[13] Y. Xie, W. Hu, H. Yang, N. Wu, and Y. Shen, “Detail information prior net for remote sensing image pansharpening,” *Remote Sens.*, vol. 13, no. 14, p. 2800, Jul. 2021.

[14] Z. Lai, L. Chen, G. Jeon, Z. Liu, R. Zhong, and X. Yang, “Real-time and effective pan-sharpening for remote sensing using multi-scale fusion network,” *J. Real-Time Image Process.*, vol. 18, pp. 1–17, Feb. 2021.

[15] Y. Chi, J. Li, and H. Fan, “Pyramid-attention based multi-scale feature fusion network for multispectral pan-sharpening,” *Appl. Intell.*, vol. 51, pp. 1–3, Aug. 2021. [Online]. Available: https://link.springer.com/journal/10498/volumes-and-issues

[16] P. Liu, “A new total generalized variation induced spatial difference prior model for variational pansharpening,” *Remote Sens. Lett.*, vol. 10, no. 7, pp. 659–668, 2019.

[17] G. Khadem and H. Ghassian, “A primal-dual method for total-variation-based pansharpening,” *Int. J. Remote Sens.*, vol. 42, no. 6, pp. 2072–2104, Mar. 2021.

[18] Z.-C. Wu, T.-Z. Huang, L.-J. Deng, J.-F. Hu, and G. Vivone, “VO-i-Net: An adaptive approach using variational optimization and deep learning for panchromatic sharpening,” *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–16, 2022.

[19] X. Fu, Z. Lin, Y. Huang, and X. Ding, “A variational pan-sharpening with local gradient constraints,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.* (CVPR), Jun. 2019, pp. 10265–10274.

[20] G. Masi, D. Cozzolino, L. Verdoliva, and G. Scarpa, “Pansharpening by convolutional neural networks,” *Remote Sens.*, vol. 8, no. 7, p. 594, Jul. 2016.

[21] G. Scarpa, S. Vitale, and D. Cozzolino, “Target-adaptive CNN-based pansharpening,” *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 9, pp. 5443–5457, Sep. 2018.

[22] Y. Wei, Q. Yuan, H. Shen, and L. Zhang, “Boosting the accuracy of multi-spectral image pansharpening by learning a deep residual network,” *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 10, pp. 1795–1799, Oct. 2017.

[23] J. Ma, W. Yu, C. Chen, P. Liang, X. Guo, and J. Jiang, “Pan-GAN: An unsupervised pan-sharpening method for remote sensing image fusion,” *Inf. Fusion*, vol. 62, pp. 110–120, Oct. 2020.

[24] H. Zhou, Q. Liu, and Y. Wang, “PGMAN: An unsupervised generative multitadversarial network for pansharpening,” *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, 2021, pp. 6316–6327.

[25] R. Dian, S. Li, A. Guo, and L. Fang, “Deep hyperspectral image sharpening,” *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 11, pp. 5345–5355, Nov. 2018.

[26] X. Fu, W. Wang, Y. Huang, X. Ding, and J. Paisley, “Deep multiscale detail networks for multiband spectral image sharpening,” *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 5, pp. 2090–2104, May 2021.

[27] J. Hu, C. Du, and S. Fan, “Two-stage pansharpening based on multi-level detail injection network,” *IEEE Access*, vol. 8, pp. 156442–156453, 2020.

[28] X. Li, F. Xu, X. Lyu, Y. Tong, Z. Chen, S. Li, and D. Liu, “A remote-sensing image pan-sharpening method based on multi-scale channel attention residual network,” *IEEE Access*, vol. 8, pp. 27163–27177, 2020.

[29] P. Shamsolmoali, J. Chanussot, M. Zareapoor, H. Zhou, and J. Yang, “Multichannel feature pyramid network for weakly supervised object detection in optical remote sensing images,” *IEEE Trans. Geosci. Remote Sens.*, early access, Aug. 30, 2021, doi: 10.1109/TGRS.2021.3106442.

[30] 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.

[31] P. Shamsolmoali, M. Zareapoor, J. Chanussot, H. Zhou, and J. Yang, “Rotation equivariant feature image pyramid network for object detection in optical remote sensing imagery,” 2021, arXiv:2106.00880.

[32] J. Yang, X. Fu, Y. Hu, Y. Huang, X. Ding, and J. Paisley, “PanNet: A deep network architecture for pan-sharpening,” in *Proc. IEEE Int. Conf. Comput. Vis.* (ICCV), Oct. 2017, pp. 5449–5457.

[33] H. Yin, “PSCSC-Net: A deep coupled convolutional sparse coding network for pansharpening,” *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–16, 2021.

[34] K. T. Gribbon and D. G. Bailey, “A novel approach to real-time bilinear interpolation,” in *Proc. 2nd IEEE Int. Workshop Electron. Design, Test Appl. (DELTA)*, Jan. 2004, pp. 126–131.

[35] X. Giorot and Y. Bengio, “Understanding the difficulty of training deep feedforward neural networks,” in *Proc. 12th Int. Conf. Artif. Intell. Statist.*, 2010, pp. 249–256.

[36] D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” 2014, arXiv:1412.6980.

[37] European Space Agency. (Oct. 1, 2021). WorldView-3: Siamese Networks Based Deep Fusion Framework for Multi-Source Satellite Imagery. [Online]. Available: https://earth.esa.int/eogateway/missions/worldview-3

[38] European Space Agency. (Oct. 1, 2021). WorldView-4: Siamese Networks Based Deep Fusion Framework for Multi-Source Satellite Imagery. [Online]. Available: https://earth.esa.int/eogateway/missions/worldview-4

[39] R. H. Yuhas, A. F. Goetz, and J. W. Boardman, “Discrimination among semi-arid landscape endmembers using the spectral angle mapper (SAM) algorithm,” in *Proc. Summaries 3rd Ann. JPL Airborne Geosci. Workshop*, 1992, pp. 147–149.

[40] Z. Wang and A. C. Bovik, “A universal image quality index,” *IEEE Signal Process. Lett.*, vol. 9, no. 3, pp. 81–84, Mar. 2002.

[41] J. Zhou, D. L. Civco, and J. A. Silander, “A wavelet transform method to merge Landsat TM and SPOT panchromatic data,” *Photogramm. Eng. Remote Sens.*, vol. 60, no. 9, pp. 1377–1387, 1994.

[42] L. Wald, T. Ranchin, and M. Mangolini, “Fusion of satellite images of different spatial resolutions,” Paris, France: Presses de l’Ecole, Ecole des Mines de Paris, 2002, p. 200. [Online]. Available: https://hal-mines-paristech.archives-ouvertes.fr/hal-00464703/document

[43] L. Wald, T. Ranchin, and M. Mangolini, “Fusion of satellite images of different spatial resolutions: Assessing the quality of resulting images,” *Photogramm. Eng. Remote Sens.*, vol. 63, no. 6, pp. 691–699, Nov. 1997.

[44] F. Palsson, J. R. Sveinsson, and M. O. Ulfarsson, “A new pansharpening algorithm based on total variation,” *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 1, pp. 318–322, Jan. 2014.

[45] L. Alparone, B. Aiazzi, S. Baronti, A. Garzelli, F. Nencini, and M. Selva, “Multispectral and panchromatic data fusion assessment without reference,” *Photogramm. Eng. Remote Sens.*, vol. 74, no. 2, pp. 193–200, Feb. 2008.