PCA-CNN Hybrid Approach for Hyperspectral Pansharpening

Giuseppe Guarino, *Graduate Student Member, IEEE*, Matteo Ciotola, *Graduate Student Member, IEEE*, Gemine Vivone, *Senior Member, IEEE*, Giovanni Poggi, *Member, IEEE*, and Giuseppe Scarpa, *Senior Member, IEEE*

***Abstract*— This work proposes a simple yet effective method to adapt unsupervised convolutional neural networks (CNNs) from multispectral (MS) to hyperspectral (HS) pansharpening. Thus, it focuses on the fusion of a single high-resolution panchromatic (PAN) band with a low-resolution HS data cube. This is achieved by means of a decorrelation transform, following the principal component analysis (PCA) approach, which enables the compression of a significant portion of the HS image energy into a few bands. Afterward, a suitably adapted pansharpening network designed for four spectral bands is used to super-resolve only the principal components (PCs). Experiments demonstrate high performance in both quantitative and qualitative evaluations, favorably comparing against state-of-the-art methods.**

***Index Terms*— Convolutional neural network (CNN), hyperspectral (HS) image, image fusion, pansharpening, principal component analysis (PCA).**

I. Introduction

HYPERSPECTRAL (HS) images are widely employed to address diverse applications such as unmixing [1], change detection [2], object detection [3], semantic segmentation [4], classification [5]. However, due to technological constraints, high spectral resolution imposes a relatively low spatial resolution compared to other imaging systems, for example, multispectral (MS) sensors. Therefore, it is very useful to exploit efficient and effective pansharpening tools, capable of increasing the spatial resolution of the HS images while preserving their spectral quality, thanks to the fusion with a simultaneously acquired high-resolution panchromatic (PAN) image [6].

A survey about the topic has been presented in [6]. Moreover, a recent contest has been organized at IEEE WHISPERS 2022 and its outcomes have been reported in [7]. Most of the current solutions are actually generalizations of methods originally conceived for the classical MS pansharpening problem [8], where usually 4–8 bands are fused. Among these, component substitution (CS) [8], [9], [10] and multiresolution analysis (MRA) [11], [12], [13] are the most popular ones. In addition, Bayesian [14] or matrix factorization [15] solutions have also been employed.

In the last years, many deep learning-based solutions have been devised for the MS pansharpening problem [16], [17], [18], [19] as well as for the HS pansharpening problem [20], [21], [22], [23]. In [20], a new HS pansharpening framework via spectrally predictive convolutional neural networks (CNNs) has been proposed. Following the same research line, a new network aimed to keep under control the spectral distortion while progressively refining the spatial details has been presented in [24]. On other hand, a spatial-domain constraint between the fused image and the PAN has been imposed to the network in [23]. A popular research line leverages on the use of attention mechanisms declined in different manners. Channel and spatial attentions have been introduced in [21] for the problem at hand. Instead, in [22], an arbitrary scale attention upsampling module has been explored, while, in [25], a residual attention network has been considered. Finally, a dual-attention guided fusion block has been presented in [26]. Another approach that is worth to mention has been proposed in [27]. It is based on the use of a deconvolution long short-term memory network enclosed in a bi-directional architecture.

Despite the research efforts registered in the last years, mostly concerned on architectural aspects, some fundamental questions remain unanswered. One is about the training context. All the above-mentioned solutions, except [25], refer to supervised learning schemes, hence, because of the lack of ground-truth (GT), they are forced to rely on synthesized reduced-resolution (RR) datasets with unavoidable implications on generalization. This problem has been deeply analyzed in [18] for MS pansharpening, registering a paradigm shift toward the unsupervised approach, which gives the possibility to train networks on full-resolution (FR) real data [28], [29]. Another issue, peculiar for the HS case and rarely taken into account [30], is the variable number of bands to manage. This variability arises from acquisition errors, which often require the removal of subsets of bands. Last but not least, in

Manuscript received 7 June 2023; revised 22 August 2023; accepted 17 October 2023. Date of publication 20 October 2023; date of current version 2 November 2023. This work was supported by the European Union—NextGenerationEU within the National Biodiversity Future Center (NBFC) Project under Grant CN00000033 and Grant CUP B83C22002930006. *(Corresponding author: Giuseppe Scarpa.)*

Giuseppe Guarino, Matteo Ciotola, and Giovanni Poggi are with the Department of Electrical Engineering and Information Technology, University of Naples Federico II, 80125 Naples, Italy (e-mail: giuseppe.guarino2@unina.it; matteo.ciotola@unina.it; giovanni.poggi@unina.it).

Gemine Vivone is with the National Research Council–Institute of Methodologies for Environmental Analysis (CNR-IMAA), 85050 Tito, Italy, and also with the National Biodiversity Future Center (NBFC), 90133 Palermo, Italy (e-mail: gemine.vivone@imaa.cnr.it).

Giuseppe Scarpa is with the Department of Engineering, Parthenope University of Naples, 80143 Naples, Italy (e-mail: giuseppe.scarpa@uniparthenope.it).

Digital Object Identifier 10.1109/LGRS.2023.3326204

the HS case, the lack of data for training is more critical than for the simpler MS case where, however, adaptive strategies, such as, [17], have been considered.

Motivated by the above considerations, here we propose an HS pansharpening technique based on the recently proposed zoom pansharpening neural network (Z-PNN) [18], conceived for PAN-MS fusion, suitably adapted to the HS case leveraging on spectral decorrelation transforms, that is, a principal component analysis (PCA). The method is fully unsupervised (allowing training at FR), is adaptive, and can operate with an arbitrary number of bands.

II. Proposed Solution

In this work, we propose a plain and efficient solution to the problem of HS pansharpening, which leverages on the recently proposed method called Z-PNN [18], originally designed for MS pansharpening. In the following, we briefly recall Z-PNN, then we describe how it has been generalized to the case of HS data.

A. Z-PNN

The Z-PNN method [18] inherits the target-adaptive paradigm introduced in [17], but it is based on a new loss conceived to use FR images without GTs, hence allowing a fully unsupervised training and tuning. The target-adaptive modality involves two stages: adaptation and prediction. The former is an iterative tuning phase, where the whole PAN-MS input pair feeds the pretrained network. At each tuning iteration, the spatial and spectral consistencies of the output are quantified to provide updated parameters, thanks to a suitable composite loss that makes use of the input PAN and MS as references, respectively. After a predefined number of iterations (i.e., 100), the network parameters are eventually frozen and used for the last run of the network to provide the final pansharpened image.

More in detail, said M and P the MS and PAN input components, respectively, $\widehat{M}$ is the pansharpened image, and $\mathcal{D}$ is a suitable downscaling operator, the overall loss is given by

$$\mathcal{L}\left(M, P; \widehat{M}\right) = \mathcal{L}_\lambda\left(M; \mathcal{D}\left(\widehat{M}\right)\right) + \beta \mathcal{L}_S\left(P; \widehat{M}\right) \quad (1)$$

where the two loss terms are weighted by β to account for both spectral and spatial consistency. The spectral loss, $\mathcal{L}_\lambda$, is given by

$$\mathcal{L}_\lambda\left(M; \mathcal{D}\left(\widehat{M}\right)\right) = \| \mathcal{D}\left(\widehat{M}\right) - M \|_1 \quad (2)$$

where $\| \cdot \|_1$ indicates the ℓ_1-norm, and the low-resolution projection operator, $\mathcal{D}$, consists of a band-wise low-pass filtering followed by spatial decimation with a factor of R (resolution ratio):

$$\mathcal{D}(M(\cdot, \cdot, b)) = [M(\cdot, \cdot, b) * h_b] \downarrow_R \quad (3)$$

being h_b the point spread function for band b, while $\downarrow_R$ and $*$ are the decimation and convolution operators, respectively.

Besides, said $\rho^\sigma_{i,j}(\widehat{M}_b; P)$ the local (to a $\sigma \times \sigma$ window) correlation coefficient between P and band $\widehat{M}_b$ at location (i, j), the spatial loss term, $\mathcal{L}_S$, is given by

$$\mathcal{L}_S\left(P; \widehat{M}\right) = 1 - \langle \min\left(\rho^{\max}_{i,j,b}, \rho^\sigma_{i,j}\left(\widehat{M}_b; P\right)\right)\rangle \quad (4)$$

being $\langle\cdot\rangle$ the average operator along i, j and b. $\rho^{\max}_{i,j,b}$ is a correlation upper bound, estimated through a smoothed version of P and the 23-tap polynomial interpolation of M. By doing so, image locations, where the correlation coefficient have reached the corresponding bounding level, do not contribute to the gradient computation. Experimental evidence suggested to set $\sigma = R$.

B. HS Pansharpening by Z-PNN

In principle, the same methodology to pansharp MS images through CNN models could be applied to HS images by increasing the number of input bands from less than ten to hundreds and, in case, redefining the resolution ratio. In practice, this generalization to a much larger number of bands is a nontrivial task for several reasons.

1) It grows the network "weight" (number of parameters) and, with it, the GPU memory requirements.
2) It grows the volume of data for an effective training.
3) The availability of HS-PAN datasets for training is much limited compared to MS-PAN datasets.
4) It grows the statistical mismatch between synthetic RR data for supervised training and real FR test data, giving raise to generalization issues.
5) The number of spectral bands can vary from one dataset to another, even when they come from the same sensor (it often occurs that subsets of bands undergo sensing issues, necessitating to get rid of them).
6) Only a fraction of the spectral bands falls within the PAN bandwidth, so that can reliably be considered well correlated with it.

On the basis of the considerations made above, we focused our attention on Z-PNN for several reasons.

1) It is unsupervised, allowing to avoid the resolution downgrade protocol for generating labeled data.
2) It makes use of an online target-adaptive phase for parameter tuning, mitigating the scarcity of training data.
3) It allows to balance the spectral and spatial constraints through a simple hyperparameter.
4) It is relatively light.

To cope with the high and variable number of bands of the HS images, we apply Z-PNN on subsets of PCA transformed bands by means of the processing chain shown in Fig. 1 (hereinafter, we will refer to the proposed method as PCA-Z-PNN). The input low-resolution HS image is split first into two sets of bands, one comprising those wavelengths that overlap with the PAN bandwidth, the other gathering the remaining ones. To be clear, we split the input bands according to their block-wise autocorrelation matrix, allowing a differentiated (adaptive) setting of the involved hyperparameters. Then, inspired by [31], each set undergoes a decorrelating transform aimed to extract four principal components (PCs) to be forwarded to the Z-PNN pansharpening engine. Once the PCs are pansharpened, they are concatenated with the other (lower energy) interpolated

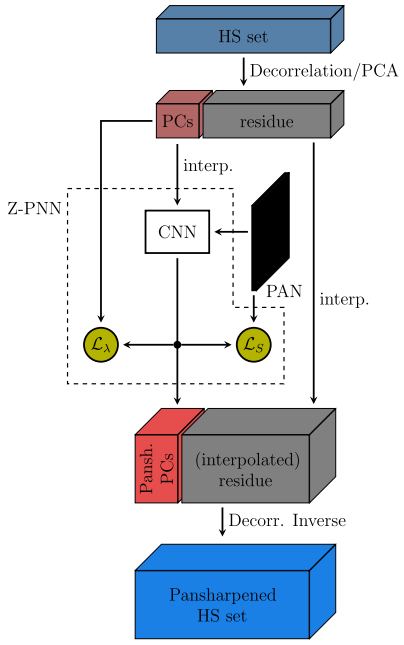

Fig. 1. Top-level flowchart of the proposed solution. The same scheme, with a different hyperparameter setting, is applied to each of the two subsets of HS bands. $\mathcal{L}_\lambda$ and $\mathcal{L}_S$ are the loss terms involved in (1)–(4).

components to be inversely transformed, coming back to the original spectral domain.

It is worth to observe that, thanks to its unsupervised and adaptive formulation, Z-PNN can work equally well in the transformed domain. The PCA transform, in fact, operates exclusively in the spectral dimension of the HS datacube, leaving unaltered the spatial information and the PAN, which is not involved. Hence, while the spatial consistency loss forces the transfer of the spatial information from the PAN to the HS bands through their compact PCs, the spectral loss term ensures the consistency between high- and low-resolution PCA bands and, indirectly, between high- and low-resolution HS bands. On other hand, the Z-PNN model is run in target-adaptive modality directly from scratch with random weights. This allows to prevent possible biases due to pretraining on nonPCA datasets or on PCA datasets built on different statistics (hence, different decorrelation transforms). On the downside, a computational overhead has to be counted at inference time because of the tuning operations. The separate processing of the two sets of bands allows us to find an optimal balance between the spectral and the spatial loss terms, that is, a different value for the hyperparameter, β, within the loss [see (1)] differing for bands that are more or less correlated with the PAN image.

III. Experimental Results

To assess the proposed approach, we used several PRISMA datasets (see Table I), comprising both RR datasets with GT and FR datasets without GT, released by the organizers of the 2022 WHISPERS HS pansharpening challenge [7]. One more FR image (Prato) is also enclosed just for validation purposes. The RR Milan image is also reserved to validation while the three remaining PRISMA images are used for test.

TABLE I

Datasets. The Resolution Ratio Is 6. The Last Three Datasets Are Available at RR, Hence With GT

Zone	Sensor	GSD	PAN size	#bands	Use	GT (RR)
Prato	PRISMA	5m	2400×2400	73	valid.	no
Bologna	PRISMA	5m	2400×2400	69	test	no
Florence	PRISMA	5m	2400×2400	63	test	no
Milan	PRISMA	30m	900×900	73	valid.	yes
Barcelona	PRISMA	30m	900×900	59	test	yes
Pavia/Univ.	ROSIS-3	1.3m	610×340	103	test	yes

TABLE II

Results on the RR Test Images (Barcelona and Pavia)

	Barcelona (PRISMA)			Pavia/Univ. (ROSIS-3)		
	$Q2^n$ (↑)	SAM (↓)	ERGAS (↓)	$Q2^n$ (↑)	SAM (↓)	ERGAS (↓)
GS [9]	0.5080	14.104	7.9535	0.7727	6.6799	4.4807
GSA [10]	0.6912	**3.3297**	2.8750	0.8731	**6.1354**	3.0682
AWLP [11]	0.6166	4.8929	4.1890	0.8415	7.6729	3.6439
MTF-GLP [12]	0.6207	3.6398	3.9223	0.8820	6.5759	3.1684
MF [13]	0.6189	5.5063	7.2229	0.8558	6.5684	3.9074
PCA [32]	0.4908	17.452	9.1831	0.7977	6.9850	3.9900
HyperPNN1 [20]	0.6491	5.2565	3.2416	0.7769	7.7578	4.2776
HSpeNet1 [24]	0.6857	4.3354	2.8756	0.8032	7.6207	4.0907
Proposed w/out split	06888	3.8133	3.1428	0.7181	6.9523	5.3625
PCA-Z-PNN	**0.6961**	3.6991	**2.8719**	**0.8862**	6.1988	**3.0578**

Beyond PRISMA images, we also considered a ROSIS-3 HS image (Pavia/University) test image for a more comprehensive experimental analysis. However, differently from PRISMA, ROSIS-3 does not provide a PAN image. Therefore, following a common practice [20], [21], an RR HS-PAN set has been synthesized by spatially downscaling the HS image, on one side, reaching a spatial resolution of 7.8 m (R equal to 6, again), and, on the other side, averaging bands sampled from the visible spectrum to get a synthetic PAN. As benchmark, in addition to the baseline methods given in [7], we also consider two recently proposed CNN approaches [20], [24]. Moreover, to prove the effectiveness of the band split strategy, we also consider the proposed method without band splitting.

Let us start with the discussion of the numerical assessment for RR datasets summarized in Table II. Here, the availability of the GT allows to quantify the synthesis error, which we measure through the most popular indexes for pansharpening: spectral angle mapper (SAM) [33], Erreur Relative Globale Adimensionelle de Synthèse (ERGAS) [34], and the generalized version [35] of the Universal Image Quality Index [36]. The proposed solution seems to work fairly (not extremely) well, in line with the state-of-the-art, with top scores on ERGAS and $Q2^n$. This should not surprise as our method is designed to provide its best on real data. In fact, learning exclusively on FR data, it can fit more or less well on synthetic data depending on the resolution downgrade process. Among the competitors, GSA seems to provide the best performance.

Besides, with the help of Fig. 2, we can have a look at some (cropped) pansharpening results, for which we show an RGB-like subset from the visible spectrum (top row) and a false color subset from the near-infrared-short-wave infrared (NIR-SWIR) range (bottom row).

The reading of these results is a bit controversial since most of the methods, including the proposed, provide outputs that

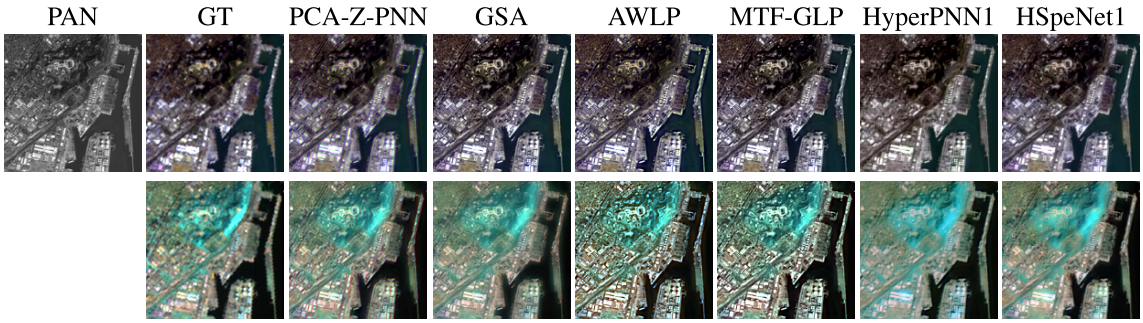

Fig. 2. Close-up on pansharpening results on the RR Barcelona image. (Top) Bands from the visible spectrum (wavelengths: 660.3, 587.8, 441.7 [nm]). (Bottom) Bands from the NIR-SWIR range (wavelengths: 2052.8, 1229, 770.5 [nm]).

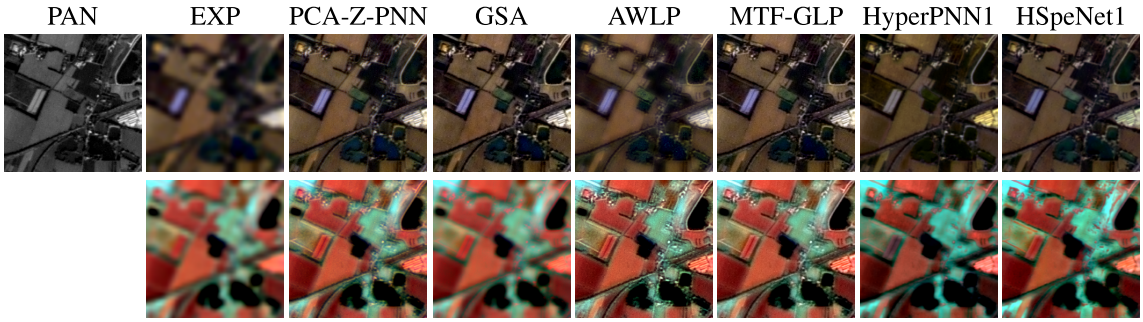

Fig. 3. Pansharpening results on the FR Bologna image. (Top) Bands from the visible spectrum (wavelengths: 660.3, 587.8, 441.7 [nm]). (Bottom) Bands from the NIR-SWIR range (wavelengths: 1585.6, 1037.8, 760.1 [nm]). EXP (interpolated HS) replaces the missing GT to provide a visual "spectral" reference.

TABLE III
NUMERICAL RESULTS AT FR (BOLOGNA AND FLORENCE)

	Bologna (PRISMA)			Florence (PRISMA)		
	D_λ (↓)	D_S^* (↓)	Q^* (↑)	D_λ (↓)	D_S^* (↓)	Q^* (↑)
GS [9]	0.2648	**0.0113**	0.7269	0.2073	0.0344	0.7654
GSA [10]	0.0512	0.0121	0.9373	0.0555	**0.0033**	0.9414
AWLP [11]	0.0274	0.0490	0.9249	0.0257	0.0367	0.9385
MTF-GLP [12]	0.0396	0.0341	0.9277	0.0331	0.0254	0.9423
MF [13]	0.0939	0.0663	0.8460	0.1044	0.0447	0.8556
PCA [32]	0.2949	0.0125	0.6963	0.1826	0.0094	0.8097
HyperPNN1 [20]	0.1203	0.0117	0.8695	0.1198	0.0152	0.8668
HSpeNet1 [24]	0.0748	0.0401	0.8881	0.0714	0.0149	0.9148
Proposed w/out split	0.0313	0.0637	0.9070	0.0502	0.0254	0.9256
PCA-Z-PNN	**0.0264**	0.0238	**0.9504**	**0.0170**	0.0096	**0.9735**

are sharper and more detailed than the GT. This is due to the synthesized PAN, which seems to be much more detailed than the GT. This also partially explains why the overall numerical results are not satisfying in this case. However, the PCA-Z-PNN solution seems to be closer to the GT than others.

Let us move to the experimental results for the FR datasets summarized in Table III and Fig. 3. The quality is assessed (refer to [7] for details) in terms of spectral (D_λ) and spatial (D_S^*) consistency or, synthetically, through their combination $Q^* = (1 - D_\lambda)(1 - D_S^*)$. The proposed solution consistently outperforms the competitors especially in terms of spectral consistency. On the other hand, the achieved spatial consistency, which is well known to be quite difficult to quantify [8], [37], and here measured through D_S^*, seems to be aligned with that of the best competitors. Overall, the proposal shows large gains over the best competitors in terms of Q^*. Indeed, if we move to the visual inspection of some sample results on Bologna with the help of Fig. 3, we can notice that the top method according to D_S^*, that is, GSA (neglecting methods whose spectral distortion is too high, e.g., above 0.1), provides too blurred results, at least compared to PCA-Z-PNN. The proposed solution, instead, works pretty well in both visible and NIR-SWIR ranges. Compared to the others, there are negligible spectral aberrations and the spatial details of the PAN are well transferred to the final fused image, also in the challenging NIR-SWIR range (bottom row). Among the competitors, HSpeNet1 seems to work fine on the spatial detail injection, although at a price of a larger spectral distortion, particularly noticeable in the visible spectral range and quantified through D_λ equal to 0.07, whereas PCA-Z-PNN achieves about 0.02.

As a final remark, we notice that in all the cases the band splitting strategy provides a remarkable performance gain.

A. Implementation Details and Complexity

Compared to the original formulation of Z-PNN [18], which runs about 100 tuning iterations starting from pretrained parameters, here, for the reasons discussed above, we run a much longer tuning (1000 iterations) but starting from random initial weights. However, the computational burden remains acceptable, thanks to a relatively light network architecture, which comprises just three convolutional layers complemented by a residual skip connection. In particular, we use the same model proposed in [17] for the 4-band MS case (refer to [17] for additional details). Under these conditions, for the given HS test images, we registered execution times of about 26 and 4 min for the FR and RR PRISMA images, respectively, and about 90 s for Pavia/Univ. ROSIS-3 image on an RTX 2080 Ti

NVIDIA GPU with 12 GB memory, which can be acceptable for many practical applications based on HS images.

Concerning the balance hyperparameter β, after several experiments carried out on the validation datasets (see Table I), we found that the optimal values for the two HS subsets are 0.5 and 0.25 for the set "covered" by the PAN and its complement, respectively, using the Adam optimizer with default parameters and a learning rate of $5e{-}5$.

IV. Conclusion

In this work, we proposed a simple yet effective CNN-based solution for HS pansharpening, which inherits the Z-PNN framework, conceived for the 4-band MS case, leveraging on the PCA decorrelation transform applied to subsets of HS bands. By doing so, we ended up with a hybrid solution having several properties: 1) it can be applied to any HS dataset, no matter how many spectral bands are considered; 2) it does not require data for pretraining; 3) it is adaptive to the target image with good generalization properties; and 4) it exclusively learns from FR real data.

References

[1] J. M. Bioucas-Dias et al., "Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 5, no. 2, pp. 354–379, Apr. 2012.

[2] C. Souza, "Mapping forest degradation in the eastern Amazon from SPOT 4 through spectral mixture models," *Remote Sens. Environ.*, vol. 87, no. 4, pp. 494–506, Nov. 2003.

[3] L. Yan, M. Zhao, X. Wang, Y. Zhang, and J. Chen, "Object detection in hyperspectral images," *IEEE Signal Process. Lett.*, vol. 28, pp. 508–512, 2021.

[4] J. L. García, M. E. Paoletti, L. I. Jiménez, J. M. Haut, and A. Plaza, "Efficient semantic segmentation of hyperspectral images using adaptable rectangular convolution," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2022.

[5] F. Luo, Z. Zou, J. Liu, and Z. Lin, "Dimensionality reduction and classification of hyperspectral image via multistructure unified discriminative embedding," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5517916.

[6] L. Loncan et al., "Hyperspectral pansharpening: A review," *IEEE Geosci. Remote Sens. Mag.*, vol. 3, no. 3, pp. 27–46, Sep. 2015.

[7] G. Vivone, A. Garzelli, Y. Xu, W. Liao, and J. Chanussot, "Panchromatic and hyperspectral image fusion: Outcome of the 2022 WHISPERS hyperspectral pansharpening challenge," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 16, pp. 166–179, 2023.

[8] G. Vivone et al., "A new benchmark based on recent advances in multispectral pansharpening: Revisiting pansharpening with classical and emerging pansharpening methods," *IEEE Geosci. Remote Sens. Mag.*, vol. 9, no. 1, pp. 53–81, Mar. 2021.

[9] C. Laben. and B. Brower, "Process for enhancing the spatial resolution of multispectral imagery using pan-sharpening," U.S. Patent 6 011 875, Jan. 4, 2000.

[10] B. Aiazzi, S. Baronti, and M. Selva, "Improving component substitution pansharpening through multivariate regression of MS+Pan data," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 10, pp. 3230–3239, Oct. 2007.

[11] G. Vivone et al., "A critical comparison among pansharpening algorithms," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 5, pp. 2565–2586, May 2015.

[12] B. Aiazzi, L. Alparone, S. Baronti, A. Garzelli, and M. Selva, "MTF-tailored multiscale fusion of high-resolution MS and pan imagery," *Photogramm. Eng. Remote Sens.*, vol. 72, no. 5, pp. 591–596, May 2006.

[13] R. Restaino, G. Vivone, M. D. Mura, and J. Chanussot, "Fusion of multispectral and panchromatic images based on morphological operators," *IEEE Trans. Image Process.*, vol. 25, no. 6, pp. 2882–2895, Jun. 2016.

[14] M. Simões, J. Bioucas-Dias, L. B. Almeida, and J. Chanussot, "A convex formulation for hyperspectral image superresolution via subspace-based regularization," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 6, pp. 3373–3388, Jun. 2015.

[15] N. Yokoya, T. Yairi, and A. Iwasaki, "Coupled nonnegative matrix factorization unmixing for hyperspectral and multispectral data fusion," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 2, pp. 528–537, Feb. 2012.

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

[17] 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.

[18] M. Ciotola, S. Vitale, A. Mazza, G. Poggi, and G. Scarpa, "Pansharpening by convolutional neural networks in the full resolution framework," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5408717.

[19] L.-J. Deng et al., "Machine learning in pansharpening: A benchmark, from shallow to deep networks," *IEEE Geosci. Remote Sens. Mag.*, vol. 10, no. 3, pp. 279–315, Sep. 2022.

[20] L. He, J. Zhu, J. Li, A. Plaza, J. Chanussot, and B. Li, "HyperPNN: Hyperspectral pansharpening via spectrally predictive convolutional neural networks," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 8, pp. 3092–3100, Aug. 2019.

[21] Y. Zheng, J. Li, Y. Li, J. Guo, X. Wu, and J. Chanussot, "Hyperspectral pansharpening using deep prior and dual attention residual network," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 11, pp. 8059–8076, Nov. 2020.

[22] L. He, J. Xie, J. Li, A. Plaza, J. Chanussot, and J. Zhu, "Variable subpixel convolution based arbitrary-resolution hyperspectral pansharpening," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5534519.

[23] W. G. C. Bandara, J. M. J. Valanarasu, and V. M. Patel, "Hyperspectral pansharpening based on improved deep image prior and residual reconstruction," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5520816.

[24] L. He, J. Zhu, J. Li, D. Meng, J. Chanussot, and A. Plaza, "Spectral-fidelity convolutional neural networks for hyperspectral pansharpening," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 5898–5914, 2020.

[25] J. Nie, Q. Xu, and J. Pan, "Unsupervised hyperspectral pansharpening by ratio estimation and residual attention network," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2022.

[26] P. Guan and E. Y. Lam, "Multistage dual-attention guided fusion network for hyperspectral pansharpening," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5515214.

[27] X. Wu, J. Feng, R. Shang, X. Zhang, and L. Jiao, "Multiobjective guided divide-and-conquer network for hyperspectral pansharpening," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5525317.

[28] S. Luo, S. Zhou, Y. Feng, and J. Xie, "Pansharpening via unsupervised convolutional neural networks," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 4295–4310, 2020.

[29] M. Ciotola, G. Poggi, and G. Scarpa, "Unsupervised deep learning-based pansharpening with jointly enhanced spectral and spatial fidelity," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5405417.

[30] J. Qu, S. Hou, W. Dong, S. Xiao, Q. Du, and Y. Li, "A dual-branch detail extraction network for hyperspectral pansharpening," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5518413.

[31] F. Palsson, J. R. Sveinsson, and M. O. Ulfarsson, "Multispectral and hyperspectral image fusion using a 3-D-Convolutional neural network," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 5, pp. 639–643, May 2017.

[32] P. S. Chavez and A. Y. Kwarteng, "Extracting spectral contrast in Landsat thematic mapper image data using selective principal component analysis," *Photogramm. Eng. Remote Sens.*, vol. 55, no. 3, pp. 339–348, 1989.

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

[34] L. Wald, *Data Fusion: Definitions and Architectures—Fusion of Images of Different Spatial Resolutions*. Paris, France: Les Presses de l'École des Mines, 2002.

[35] A. Garzelli and F. Nencini, "Hypercomplex quality assessment of multi/hyperspectral images," *IEEE Geosci. Remote Sens. Lett.*, vol. 6, no. 4, pp. 662–665, Oct. 2009.

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

[37] G. Scarpa and M. Ciotola, "Full-resolution quality assessment for pansharpening," *Remote Sens.*, vol. 14, no. 8, p. 1808, Apr. 2022.