



HAL
open science

A RESIDUAL DENSE GENERATIVE ADVERSARIAL NETWORK FOR PANSHARPENING WITH GEOMETRICAL CONSTRAINTS

Anaïs Gastineau, Jean-François Aujol, Yannick Berthoumieu, Christian Germain

► **To cite this version:**

Anaïs Gastineau, Jean-François Aujol, Yannick Berthoumieu, Christian Germain. A RESIDUAL DENSE GENERATIVE ADVERSARIAL NETWORK FOR PANSHARPENING WITH GEOMETRICAL CONSTRAINTS. 27th IEEE international conference on image processing (ICIP 2020), Oct 2020, Abou Dabi, United Arab Emirates. hal-02859866

HAL Id: hal-02859866

<https://hal.science/hal-02859866v1>

Submitted on 8 Jun 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

A RESIDUAL DENSE GENERATIVE ADVERSARIAL NETWORK FOR PANSHARPENING WITH GEOMETRICAL CONSTRAINTS

Anais GASTINEAU^{*†}, Jean-François AUJOL^{*}, Yannick BERTHOUMIEU[†] and Christian GERMAIN[†]

^{*} Univ. Bordeaux, Bordeaux INP, CNRS, IMB, UMR 5251, F-33400 Talence, France

[†]Univ. Bordeaux, Bordeaux INP, CNRS, IMS, UMR 5218, F-33400 Talence, France

ABSTRACT

The pansharpening problem consists in fusing a high resolution panchromatic image with a low resolution multispectral image in order to obtain a high resolution multispectral image. In this paper, we adapt a Residual Dense architecture for the generator in a Generative Adversarial Network framework. Indeed, this type of architecture avoids the vanishing gradient problem faced when training a network by re-injecting previous information thanks to dense and residual connections. Moreover, an important point for the pansharpening problem is to preserve the geometry of the image. Hence, we propose to add a regularization term in the loss function of the generator: it preserves the geometry of the target image so that a better solution is obtained. In addition, we propose geometrical measures that illustrate the advantages of this new method.

Index Terms— Pansharpening, Generative Adversarial Network, residual dense network, regularization, remote sensing.

1. INTRODUCTION

Satellites are essential for Earth’s observation. They acquire images at various scales depending on the objective of the satellite. In remote sensing, the spatial resolution is expressed as the size of the ground captured by one pixel. The spectral resolution is given both by the number of bands of the image and by the bandwidth of the signal captured by the sensors producing the images.

Satellites usually offer multispectral and panchromatic images. Panchromatic images have a high spatial resolution and only one spectral band, in contrast with the multispectral images that have a low spatial resolution and several spectral bands. Pansharpening techniques are then used in order to fuse the panchromatic and the multispectral images to ob-

tain an image with the spatial resolution of the panchromatic image and the spectral resolution of the multispectral image.

This problem can be formulated as the reconstruction of an image u thanks to the observed panchromatic image P and the multispectral image y . This leads to consider the following model:

$$\begin{cases} y^k = SH^k u^k + B^k, \forall k \leq K \\ P = \sum_{k \leq K} \alpha_k u^k + B \end{cases}, \quad (1)$$

where S is a subsampling operator, H the blur operator, B a gaussian noise and K the number of bands of the multispectral images. The second equation of problem (1) supposes that the panchromatic image can be approached by a linear combination of the bands of the image we are looking for. This model was introduced by *Ballester et al.* [1] and it is commonly used [2, 3].

2. STATE OF THE ART AND CONTRIBUTIONS

Several types of approaches have been proposed to solve the pansharpening problem:

- i) Component substitution methods that use a linear transformation on the multispectral data to separate spatial details from spectral details in order to substitute the spatial details by the panchromatic image [4, 5].
- ii) Multi-resolution analysis approaches which combine methods that decompose panchromatic and multispectral images into a sequence of signals or pyramids with decreasing information content. This allows to add high frequencies of the panchromatic image to the solution [6, 7].
- iii) Variational or bayesian methods that solve an inverse problem by giving an a priori on the solution [1, 2, 3].
- iv) Learning methods that model the relation between variables, here the panchromatic image and the multispectral images by the composition of several levels (convolution, etc.), without the need of a model such as (1).

Many approaches using deep learning have been proposed and they give state of the art results. *Masi et al.* [8] adapt a CNN used for super-resolution for the pansharpening problem. This CNN mimics the behavior of a sparse represen-

The authors would like to thank the team of the thematic recipe users of the program ORFEO of accompaniment Pléiades for the Pléiades images (©CNES (2012,2013), distribution Airbus DS/Spot Image) used within the framework of this study. This study has been carried out with financial support from the French State, managed by the French National Research Agency (ANR) in the frame of the "Investments for the future" Programme IdEx Bordeaux - SysNum (ANR-10-IDEX-03-02). The authors acknowledge the support of the French Agence Nationale de la Recherche (ANR) under reference ANR-18-CE92-0050 SUPREMATIM.

tation. *Palsson et al.* [9] use a 3-D CNN because they consider multispectral images as 3D images: two spatial dimensions and one spectral dimension. This makes it possible to better model the inter bands spectral correlation. *Guo et al.* [10] propose a CNN robust to the inconsistencies across satellites by using a 4 layers network with dilated multilevel blocks. This dilated multilevel block allows to make full use of features extracted by the convolutional layers. Moreover, to avoid overfitting they propose to use a l_2 -regularization term on the weights in the loss function.

An important point for the pansharpening problem is to preserve the geometry of the panchromatic image. To do so, *Yang et al.* [11] propose to train a network PanNet in the high frequency (HF) domain to preserve the structure and the geometry of the panchromatic image. To preserve the spectral information, the multispectral image is propagated thanks to residual connection in a ResNet architecture.

More recently, there is the emergence of GANs for the generative problem. Generative Adversarial Networks (GANs) are a class of unsupervised learning algorithms introduced by *Goodfellow et al.* [12]. This type of network seeks to mimic any data distribution. In general, a GAN is a generative model where two networks are competing each other. The first network is the generator G_θ . It generates a sample, while its adversary, the discriminator network D_η tries to detect whether a sample is real or if it is the result of the generator. For example, *Liu et al.* [13] propose a GAN based method. They choose a simple architecture, i.e. a CNN as generator and discriminator.

In this paper, we propose a new GAN based framework by first considering a Residual Dense architecture for the generator. Motivated by the efficiency of the Residual dense architecture to train networks for many applications (super-resolution [14] or denoising [15] for instance), we propose to adapt this type of architecture. Indeed, dense and residual connections allow to re-inject the information across the network and so to avoid the vanishing gradient problem faced during training.

In addition, to preserve the geometry, we propose to add a regularization term in the loss function of the generator. This term is based on the assumption that the gradient of the fused image can be aligned at each point with the gradient of the target image by regarding the orientation of the vector given by the gradient. This new method is called RDGAN-Geom.

3. RESIDUAL DENSE BASED ARCHITECTURE USED FOR THE GENERATOR

In the literature, based on GANs, the authors, for example [13, 18], usually use to minimize the following loss functions:

$$L(G_\theta) = \sum_{i \leq N} \alpha \log(D_\eta(G_\theta(z))) + \delta \|u - G_\theta(z)\|_1 \quad (2)$$

the loss function for the generator G_θ and:

$$L(D_\eta) = \sum_{i \leq N} \log(1 - D_\eta(G_\theta(z))) + \log(D_\eta(u)) \quad (3)$$

the loss function for the discriminator D_η .

Here, N is the batch size, $z = [\uparrow y, P]$ is the input of the network G_θ with $\uparrow y$ corresponding to the low resolution multispectral y image upsampled with a bicubic interpolation to the size of the panchromatic image P , $G_\theta(z)$ is the output of the generator, θ and η the parameters of G_θ and D_η to optimize and u is the target image.

Since we work in a residual framework, the output of the generator $G_\theta(z)$ is a residual image. This residue contains information about spatial and spectral details. To obtain the final multispectral image we have to add the residue to the multispectral image $\uparrow y$. For the discriminator D_η , Equation (3) is used but formulated in a residual way.

We consider a Residual Dense architecture as in Figure 1 for the generator. This architecture allows us to keep advantages of residual [16] and dense [17] architectures. These types of architectures were introduced in order to solve the vanishing gradient problem, often faced during the training process when using a deep neural network. The Residual Dense architecture, in Figure 1, is composed of p residual dense blocks B_i , $i \leq p$ presented in Figure 2, and it takes in input the concatenation of the panchromatic and the multispectral images.

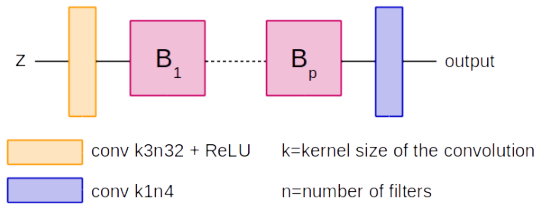


Fig. 1: General architecture used for the generator, where the input $z = [P, \uparrow y]$ is the concatenation $[\cdot]$ of the panchromatic image P and $\uparrow y$ the multispectral image $y = (y^1, \dots, y^N)$ resized to the size of P . Blocks B_i , $i \leq p$, are residual dense blocks described in Figure 2.

Each residual dense block, Figure 2, is composed of four convolutional layers. Dense connections [17] are represented by the concatenation of outputs of different layers. H_l represents the non-linear transformation (convolution, ReLU and pooling) of the l^{th} layer of one block and v_l represents the output of this layer. The l^{th} layer in input is the feature maps of all previous layers of the block. Therefore, it can be written as follows:

$$v_l = H_l([v_0, \dots, v_{l-1}]),$$

where $[\cdot]$ represents the concatenation. It means that each layer has access to all previous information. Furthermore, each blocks is composed of one residual connection, represented by the addition of the input of the block with the outputs of the last layers. In this way, the information in input

of the block is transmitted without any modification.

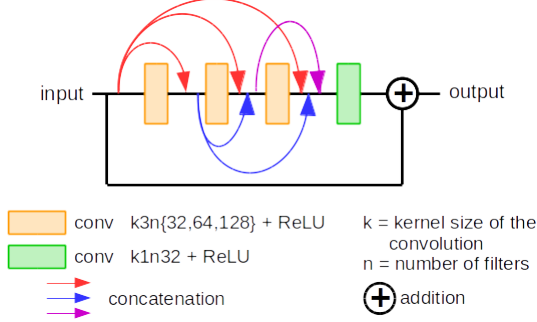


Fig. 2: Architecture based on residual dense blocks B_i for $i = 1, \dots, p$. Dense connections are represented by the concatenation of each output of previous layers and residual connections by the addition of the two different layer outputs.

This type of architecture trains a deep neural network by considering residual and dense connections. The aim is to re-inject the information obtained in output of some layers in order to avoid a gradient equal or close to zero.

4. GEOMETRICALLY CONSTRAINED LOSS FUNCTION

Thus, finding the spatial details and recovering the geometry in the fused image is very important for the pansharpening problem. In order to do this, *Yang et al.* use the high frequencies (HF) of the panchromatic and the multispectral image to train a ResNet architecture, PanNet [11]. Therefore, they minimize the loss function:

$$\mathcal{L} = \|f_\omega(P_{HF}, \uparrow y_{HF}) + \uparrow y - u\|_2^2, \quad (4)$$

where f_ω is the network PanNet, u the target image and P_{HF} and $\uparrow y_{HF}$ are high pass information of P and $\uparrow y$.

On the contrary, to take into account the geometry, we propose to consider a third regularization term in the equation (2), a geometrical term, initially proposed by *Ballester et al.* [1] in a variational framework for the pansharpening problem. We propose to minimize the following loss function:

$$\begin{aligned} L(G_\theta) = & \sum_{i \leq N} \alpha \log(D_\eta(G_\theta(z) + \uparrow y)) + \\ & \delta \|G_\theta(z) + \uparrow y - u\|_1 + \\ & \beta \sum_{x \in \Omega} |\nabla u(x)^\perp \cdot \nabla(G_\theta(z)(x) + \uparrow y(x))| \end{aligned} \quad (5)$$

for the generator G_θ , where $\nabla(\cdot)$ is the gradient operator, \perp is the orthogonal vector and Ω the image domain. This third term forces the alignment of the level lines of each band of the solution with each band of the target image. This amounts to transferring the geometry of the target image to the desired pansharpened image. This results in using the inner product between the orthogonal gradient vectors of the target image and the gradients of the output image. Indeed, this inner product is equal to zero when vectors are collinear and consequently the geometry of the target image is preserved.

5. EXPERIMENTS

5.1. Dataset and quality evaluation

For most satellites, the resolution factor between the panchromatic and the multispectral images is 4. We use images of the Pléiades satellite to train and test the network. The ground sampling distance of Pléiades is 0.7 m for the panchromatic image and 2.8 m ground sampling distance for the multispectral image. To train and test, the satellite images are cropped into samples of size 128×128 . Finally, 4408 samples are used for training and 412 for testing.

Several criteria are used to measure the spatial or spectral quality of the fused images. We use the Spectral Angle Mapper (SAM) measure. This criterion measures the spectral distortion between the target image and the fused image. The Cross Correlation (CC) coefficient evaluates the spatial distortion between both images by computing the inter-bands and intra-bands correlations. The Root Mean Square Error (RMSE) gives a global error corresponding to the l_2 norm and the Peak Signal to Noise Ratio (PSNR).

In order to compare the geometrical similarities of two images X and Y , we propose the geometrical metric M_A :

$$M_A(X, Y) = \frac{1}{|\Omega|} \sum_{\Omega} \|AX - AY\|_2, \quad (6)$$

where $|\Omega|$ is the number of pixel and $A = \nabla$ or $A = \Delta$ with Δ the Laplacian operator. The metric M_∇ allows to compare the edges of two images and M_Δ allows to compare the points detected on each image [19].

Since all the previous measures need a target image, the Wald's protocole [20] is used to train the network. It consists in reducing the spatial resolution of the panchromatic and the multispectral images. Then the degraded images are used to train the network. Hence it becomes possible to compare the fused image obtained by the different methods with the original multispectral image. This protocol offers therefore a way of checking the different properties of the considered methods.

5.2. Details of implementation

The proposed method was implemented with Tensorflow and with ADAM optimizer to minimize the loss function with an initial learning rate of 0.0002 and a momentum of 0.5. The batch size is set to 19.

5.3. Results

We compare the proposed method with some state-of-the-art pansharpening methods. The pansharpening methods in comparison are the coefficient injection methods GLP [21] and GSA [22], the variational method P+XS [1] and the PanNet [11] and PSGAN [13] networks. For a better comparison, we train both networks on the Pléiades dataset and the training time is about 24 hours for PanNet. The training time is about 15 hours for the PSGAN method and between 10 and 12 hours when considering the ResNet or the Residual Dense architecture.

Method	PSNR	CC	SAM	RMSE
ideal value	max	1	0	0
GSA [22]	19.73	0.877	0.2997	27.40
GLP [21]	19.47	0.877	0.2998	28.14
P+XS [1]	19.37	0.860	0.317	28.48
PanNet [11]	28.36	<u>0.950</u>	<u>0.157</u>	10.30
PSGAN [13]	<u>26.59</u>	0.952	0.155	<u>10.93</u>
PSGAN-Geom	27.18	0.955	0.145	10.14
GAN ResNet	28.38	0.960	0.153	10.07
GAN ResNet-Geom	28.42	0.961	0.152	10.02
GAN ResNet HF	28.24	0.9595	0.154	10.22
GAN ResNet HF-Geom	28.26	0.960	0.152	10.18
RDGAN	29.37	0.969	0.141	8.94
RDGAN-Geom	29.38	0.969	0.138	8.93
RDGAN HF	29.17	0.967	0.145	9.15
RDGAN-Geom HF	29.20	0.968	0.144	9.11

Table 1: Quantitative results obtained on test images, best results are in bold and worst underlined comparing learning approaches.

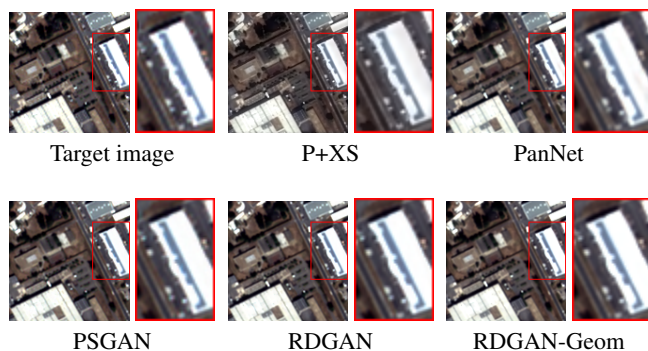


Fig. 3: Visual results obtained with the different methods on urban area. Only comparing RGB images does not allow to conclude about the major differences between results.

First, the coefficient injection methods GLP [21] and GSA [22] and the variational method P+XS [1] do not give satisfying quantitative results unlike learning approaches (Table 1).

Then, when the geometrical term is added to the PSGAN method while keeping the same architecture provided by Liu *et al.* [13], quantitative results are improved but the architecture of the PSGAN method is very simple. By changing the architecture of the generator for a ResNet architecture proposed in the PanNet method [11], quantitative results are similar to those obtained with the PanNet approach. However, when considering the geometrical term (5) with the ResNet architecture, results are improved.

Finally, motivated by the results obtained in many applications, the proposed method RDGAN-Geom, composed of several residual dense blocks, gives the best quantitative results and these results are better when adding the geometrical term (5) in the loss function.

Visually on Figure 3, it is more difficult to appreciate visual differences between all learning methods if we only consider RGB visual comparison. Nevertheless, on the dif-

ference images in Figure 4, we can see a better preservation of the geometry when considering the RDGAN-Geom network. In order to quantify the results obtained in Figure 4, we use the metrics M_{∇} and M_{Δ} introduced in Section 5.1. Quantitative results in Table 2 confirm visual results and we can conclude that our RDGAN-Geom gives best results while preserving the geometry.

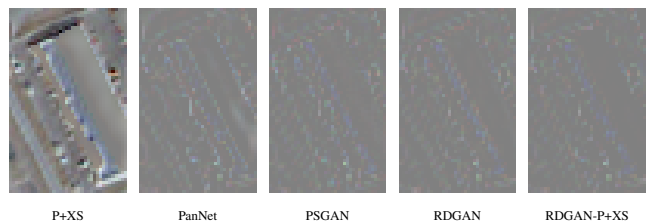


Fig. 4: Images of the differences between the target image zoom and the result obtained for each method on the zoomed part of figure 3. The method GAN Residual Dense-P+XS helps to better preserve the geometry of the target image.

Model	M_{∇}	M_{Δ}
Ideal value	0	0
P+XS	<u>17.22</u>	<u>23.96</u>
PanNet	9.07	13.52
PSGAN	8.66	13.18
RDGAN	8.20	12.55
RDGAN-Geom	8.10	11.87

Table 2: Quantitative results obtained on test images. These results show a better preservation of the geometry for the GAN Residual Dense architecture while considering the P+XS term in the loss function. Best results are in bold and worst underlined.

Yang *et al.* [11] use the high frequencies of the multispectral and the panchromatic images in input of their network to preserve edges and structure. To compare this network with ours, we also test when giving the high frequencies in input of our network when considering a ResNet or a Residual Dense architecture for the generator (named with HF in Table 1). Quantitative results show that the high frequencies have few influence in our GAN framework whatever the architecture considered.

6. CONCLUSION

In conclusion, we propose to use a Residual Dense architecture for the generator in a Generative Adversarial Network with a regularization term in the loss function for the pansharpening problem. The proposed term added to the loss function forces the alignment of the gradient of the output with the gradient of the target image at each pixel in order to preserve the geometry.

Furthermore, the combination of the architecture and the term in the loss function allows to get better quantitative and visual results, preserving geometry in images.

7. REFERENCES

- [1] C. Ballester, V. Caselles, L. Igual, J. Verdera, and B. Rougé, “A Variational Model for P+XS Image Fusion,” *IJCV*, vol. 69, no. 1, pp. 43-59, 2006.
- [2] J. Duran, A. Buades, B. Coll, and C. Sbert, “A non local variational model for pansharpening image fusion,” *SIAM*, vol. 7, no. 2, pp. 761-796, 2015.
- [3] F. Palsson, J. Sveinsson, M. Ulfarsson, and J. Benediktsson, “A New Pansharpening Method Using an Explicit Image Formation Model Regularized via Total Variation,” *IEEE International Geoscience and Remote Sensing Symposium*, 2012.
- [4] M. Gonzalez-Audicana, J.L. Saleta, R. Garcia Catalan, and R. Garcia, “Fusion of Multispectral and Panchromatic Images Using Improved IHS and PCA Mergers based on Wavelet Decomposition,” *IEEE TGRS*, vol. 42, No. 6, pp. 1291-1299, 2004.
- [5] W. Carper, T. Lillesand, and R. Kiefer, “The use of Intensity-Hue-Saturation transformations for merging SPOT panchromatic and multispectral image data,” *Photogramm Eng Remote Sensing*, vol. 56, No. 4, pp. 459-467, 1990.
- [6] M. Gonzalez-Audicana, X. Otazu, O. Fors, and A. Seco, “Comparison between Mallat’s and the ‘A trous’ discrete wavelet transform algorithms for the fusion of multispectral and panchromatic images,” *Int J Remote Sens*, vol. 26, no. 3, pp. 595-614, 2005.
- [7] P.J. Burt and E.H. Adelson, “The Laplacian pyramid as a compact image code,” *IEEE Transactions on communications*, vol. 31, No. 4, pp. 532-540, 1983.
- [8] G. Masi, D. Cozzolino, L. Verdolina, and G. Scarpa, “Pansharpening by Convolutional Neural Network,” *Remote Sensing*, vol. 8, no. 7, pp. 594-616, 2016.
- [9] F. Palsson, J. Sveinsson, and M. Ulfarsson, “Multispectral and Hyperspectral Image Fusion Using 3D Convolutional Neural Network,” *IEEE Geoscience and remote sensing*, 2017.
- [10] Y. Guo, F. Ye, and H. Gong, “Learning an Efficient Convolution Neural Network for Pansharpening,” *Algorithms*, vol. 12, pp. 16, 2019.
- [11] J. Yang, X. Fu, Y. Hu, Y. Huang, X. Ding, and J. Paisley, “PanNet: A Deep Network Architecture for Pan-Sharpener,” *IEEE International Conference on Computer Vision*, October 2017.
- [12] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, “Generative Adversarial Nets,” *Neural Information Processing Systems Proceedings*, 2014.
- [13] X. Liu, Y. Wang, and Q. Liu, “PSGAN: A Generative Adversarial Network for Remote Sensing Image Pansharpening,” *IEEE International Conference on Image Processing*, 2018.
- [14] Y. Zhang, Y. Tian, Y. Kong, B. Zhong, and Y. Fu, “Residual Dense Network for Image Super-Resolution,” *IEEE Conference on Computer Vision and Pattern Recognition*, 2018.
- [15] Y. Song, Y. Zhu, and X. Du, “Dynamix Residual Dense Network for Image Denoising,” *Sensors*, 2019.
- [16] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” *IEEE Conference on Computer Vision and Pattern Recognition*, 2016.
- [17] G. Huang, Z. Liu, L. van der Maaten, and K. Weinberger, “Densely connected convolutional networks,” *IEEE Conference on Computer Vision and Pattern Recognition*, 2017.
- [18] Y. Zhang, X. Li, and J. Zhou, “SFTGAN : a generative adversarial network for pan-sharpening equipped with spatial feature transform layers,” *Journal of Applied Remote Sensing*, vol. 13, no. 2, 2019.
- [19] G. Aubert, J.-F. Aujol, and L. Blanc-Feraud, “Detecting codimension-two objects in an image with Ginzburg-Landau models,” *International Journal of Computer Vision*, 2005.
- [20] L. Wald, T. Ranchin, and M. Mangolini, “Fusion of satellite images of different spatial resolution: Assessing the quality of resulting images,” *Photogramm. Eng. Remote Sens.*, vol. 63, no. 6, 1997.
- [21] G. Vivone, R. Restaino, and J. Chanussot, “Full Scale Regression-based Injection Coefficients for Panchromatic Sharpening,” *IEEE Transactions on Image Processing*, vol. 27, no. 7, 2018.
- [22] R. Restaino, M. Dalla Mura, G. Vivone, and J. Chanussot, “Context-adaptive Pansharpening Based on Image Segmentation,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 2, 2017.