CNN-BASED PANSHARPENING OF MULTI-RESOLUTION REMOTE-SENSING IMAGES

Giuseppe Masi, Davide Cozzolino, Luisa Verdoliva and Giuseppe Scarpa

DIETI, University Federico II of Naples, Italy
{firstname.lastname}@unina.it

ABSTRACT

We propose a convolutional neural network for the pansharpening of remote-sensing imagery. A very compact architecture is designed, which enables accurate training even with small-size datasets. Prior knowledge on the remote sensing domain is taken into account by augmenting the input with several maps of radiometric indices. Extensive experiments on images from various multiresolution sensors show the proposed CNN to outperform the current state of the art in terms of both full-reference and no-reference measures.

Index Terms— pansharpening; segmentation; super-resolution; machine learning; convolutional neural networks.

I. INTRODUCTION

In order to deliver data with high resolution in both the spatial and spectral domains modern remote-sensing systems rely on multiresolution images. A high (spatial) resolution panchromatic band (PAN) is complemented by a low resolution multispectral (MS) stack. Signal processing methods are then used to process these sources jointly, or to pansharpen them, thus generating a datacube with the highest spectral and spatial resolution.

In [1] traditional methods for pansharpening are clustered in two main categories, component substitution and detail injection. With the first approach, the MS is transformed in some suitable domain where one of the components is replaced by the high-resolution PAN image before up-sampling and back-transforming the whole stack. Performance depends closely on the type of transform used. Good results have been observed with Brovey transform (BT) [2] and Gram-Schmidt (GS) spectral sharpening [3]. To reduce spectral distortion, an adaptive version of GS is proposed in [4] while [5] relies on the partial replacement of components (PRACS).

Detail injection, instead, consists in extracting high-frequency content from the PAN and injecting it into the up-sampled version of the MS. The extraction can be carried out through suitable multiresolution analysis tools, like the á Trous Wavelet transform (ATWT) or Laplacian Pyramids (LP) [6]. To avoid distortion and artifacts, a model is necessary [7] which relates the scales at which data are available, based on the modulation transfer functions (MTFs) of the sensors. Some of the most promising injection-based methods are SFIM [8], Indusion [9], and algorithms based on ATWT [7], [10], and LP [11], [12], [13]. Lately, methods based on sparse representations have also gained some popularity.On the contrary, little attention has been devoted to deep learning, despite the impressive results observed in computer vision, image processing, and also remote sensing [14], [15].

In [16] we proposed a convolutional network for remote-sensing image pansharpening, building upon the architecture proposed in [17] for the closely related super-resolution problem. After adapting the architecture to the pansharpening case, we exploit domain-specific knowledge by augmenting the input with a number of radiometric indices. The resulting architecture is very compact, allowing easy and effective training, and providing promising performance. Here we carry out a deeper experimental analysis, comparing results with a large number od state-of-the-art methods on three datasets comprising images acquired by the Ikonos, GeoEye-1 and WorldView-2 sensors. A sample result of the proposed pansharpening method is visible in Fig.1. We show on the left the input PAN and MS images at their original resolutions, then the interpolated MS, and finally the pansharpened output (all MS images are projected on a suitable RGB space for visualization). The spatial resolution of the PAN is fully preserved and a high spectral fidelity is ensured. Results on all the datasets confirm the excellent performance of the proposed CNN.

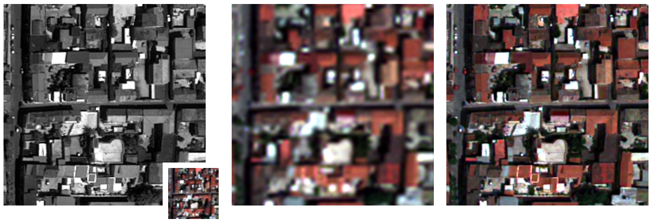

Fig. 1. Sample result of the proposed pansharpening method. From left to right: input multiresolution image acuired by the GeoEye-1 sensor, interpolation of MS, pansharpened image.

II. CNN-BASED PANSHARPENING

In [17] it was shown that some state-of-the-art super-resolution methods, based on sparse coding and dictionary learning, can be implemented by means of a simple convolutional network. Besides proving the versatility of CNNs, this result has significant practical implications. The theoretical equivalence with a conventional method provides precious guidelines for the design of an effective and compact CNN architecture. From this good starting point, training can be used to optimize jointly all layers of the network. Moreover, architectural changes can be easily explored starting for the basic solution to further improve performance.

Following this approach, in [16] we proposed a three-layer CNN for the pansharpening of remote sensing images. Preliminary experiments made clear that several features extracted from the first layer of the net were remarkably close to well-known radiometric indexes, such as the normalized vegetation index (NDVI) and the normalized water index (NDWI). Therefore, we decided to precompute maps accounting for these indexes and provide them as input to the net together with the PAN and upsampled MS components. After some more structural developments, we obtained the three-layer architecture shown pictorially in Fig.2.

The input image y_0 comprises the B upsampled MS bands, the PAN, and the $B_{\rm rad}$ upsampled radiometric index maps. Because of the nature of the problem, with the output image having the same spatial resolution as the PAN, only convolutional layers are needed. Therefore, all layers implement convolution followed by non-linear activation

$$y_{i+1} = f_i(y_i * w_i + b_i),$$

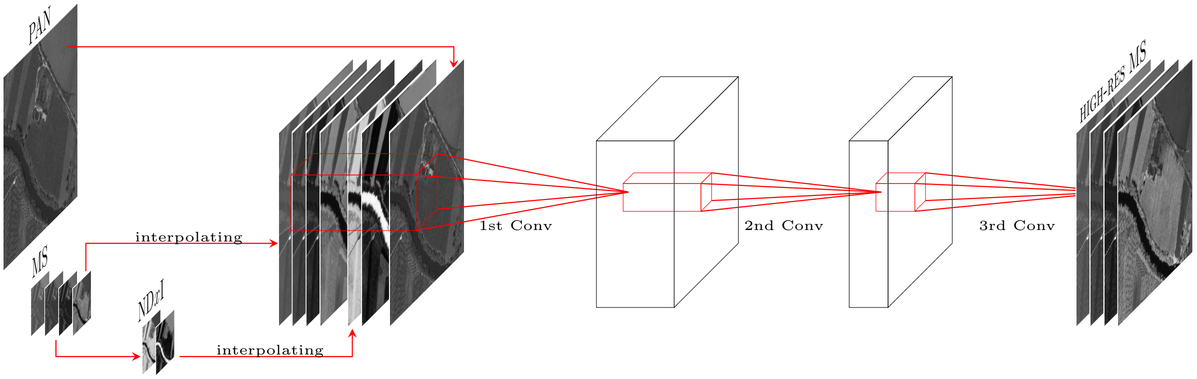

Fig. 2. Proposed CNN architecture for pansharpening. MS components and index maps are upsampled and joined with the PAN at the input.

using n_{i+1} filters with kernels w_i of size $n_i \times (k_i \times k_i)$ and bias vectors b_i. Some hyper-parameters are fixed, $n_0 = B + B_{\rm rad} + 1$ and $n_3 = B$, while the others are chosen based on preliminary experiments and are summarized below for the three datasets used in the expriments. A ReLU non-linearity is used in the first two layers, while an identity is used in the last one.

The first convolutional layer has resulted to be the most sensitive to the choice of the hyper-parameters. In particular experimental tests suggested to use $n_1 = 48$ filters in the first layer for all three sensors, but with different receptive fields ($k_1 \times k_1$): 9×9 for GeoEye-1 and WorldView-2, and 5×5 for Ikonos. The second and third layer, instead, have the same hyper-parameters for all sensors. In particular $n_2 = 32$ filters compose the second layer and 5×5 receptive fields are uses in both the second and the third layers: $k_2 = k_3 = 5$.

It is worth underlining that this is quite a simple CNN architecture, relatively shallow, and hence easy to train even with a small dataset. This is extremely important for the remote sensing field, where training data are often scarce as opposed to the millions images typically available for computer vision applications.

III. PERFORMANCE ASSESSMENT

To test the performance of the proposed method we designed three datasets comprising multi-resolution images acquired by some of the most popular sensors in the field, that is, GeoEye-1, IKONOS, and WorldView-2. Details on such datasets are reported in [16]. We assessed performance for the proposed method and a large number of reference techniques, listed in Tab.1, using the experimental protocol proposed in [1]. Results are reported in Tables 2, 3, and 4. Both full-reference (Q4, Q, SAM, ERGAS, SCC) and no-reference ($\mathbf{D}_\lambda$, $\mathbf{D}_S$, QNR) measures are used. The former are computed, assuming scale-invariance, by working on subsampled datacubes.

In the tables, methods are listed in temporal sequence (oldest first) except for tightly related methods that are grouped together. This organization allows one to catch at a glance the progress of the state of the art. We showed in blue the best performance and in red the second best. For all three datasets and all measures, our CNN solution provides always the best result, with the only exception of two no-reference results, $\mathbf{D}_\lambda$ on IKONOS and $\mathbf{D}_S$ on WorldView-2, where the CNN ranks third. The comparison with the second best result is especially meaningful, as it makes clear that deep learning guarantees in most cases an impressive performance gain. This holds in particular for full reference measures. For example, the CNN method reduces SAM distortion by 33%, 20%, and 25% on the three datasets with respect to the second best. Likewise, for the Q measure, CNN reduces significantly the gap towards the ideal optimum of 1, going even from 0.883 to 0.940 for GeoEye-1. Results are less striking on no-reference measures, ringing a bell on possible distortions on the output images. So we turn to visual inspection, which is of paramount importance for remote-sensing applications.

We inspected and compared carefully the pansharpened images provided by all techniques. For some selected details, one for each sensor, we show in Fig.3 the output images (to be observed on a computer screen with suitable zoom) generated by the methods that, according to numerical results, appear to be the most promising. In extreme summary, these images confirm the good pansharpening quality of the proposed method. Together with C-BDSD it provides the sharpest results, but the latter introduces visible spectral distortions. On the down side, we noticed some artificial patterns in flat areas of the IKONOS images. These may be due to the training on low-resolution patches, relying therefore on a scale-invariance property which is not guaranteed to hold. Future work will address this problem and focus on possible architectural changes.

A	BT : Brovey Transform [2]
B	GS : Gram Schmidt [3]
C	GSA : Gram Schmidt Adaptive [4]
D	HPF : High-Pass filtering based Fusion [18]
E	SFIM : Smoothing Filter-based Intensity Modulation[8]
F	ATWT-M2 : A Trous Wavelet Transform with Model 2 [7]
G	ATWT-M3 : A Trous Wavelet Transform with Model 3 [7]
H	ATWT-UIM : ATWT w/ Unitary Injection Model [10]
I	MTF-GLP : Gen. Lapl. Pyramid with MTF-matched filter [6]
J	MTF-GLP-CBD : (with regression based injection model) [12]
K	MTF-GLP-HPM : (with multiplicative injection model) [11]
L	MTF-GLP-HPM-PP : (with Post-Processing) [13]
M	AWLP : Additive Wavelet Luminance Proportional [19]
N	Indusion : wavelet transform with additive injection model [9]
O	BDSD : Band-Dependent Spatial-Detail [20]
P	C-BDSD : BDSD with nonlocal extension [21]
Q	PRACS : Partial Replacement Adaptive Comp. Substitution [5]
X	PNN : proposed CNN-based Pansharpening

Table I. Techniques under comparison: heading letters are used to index the following tables.

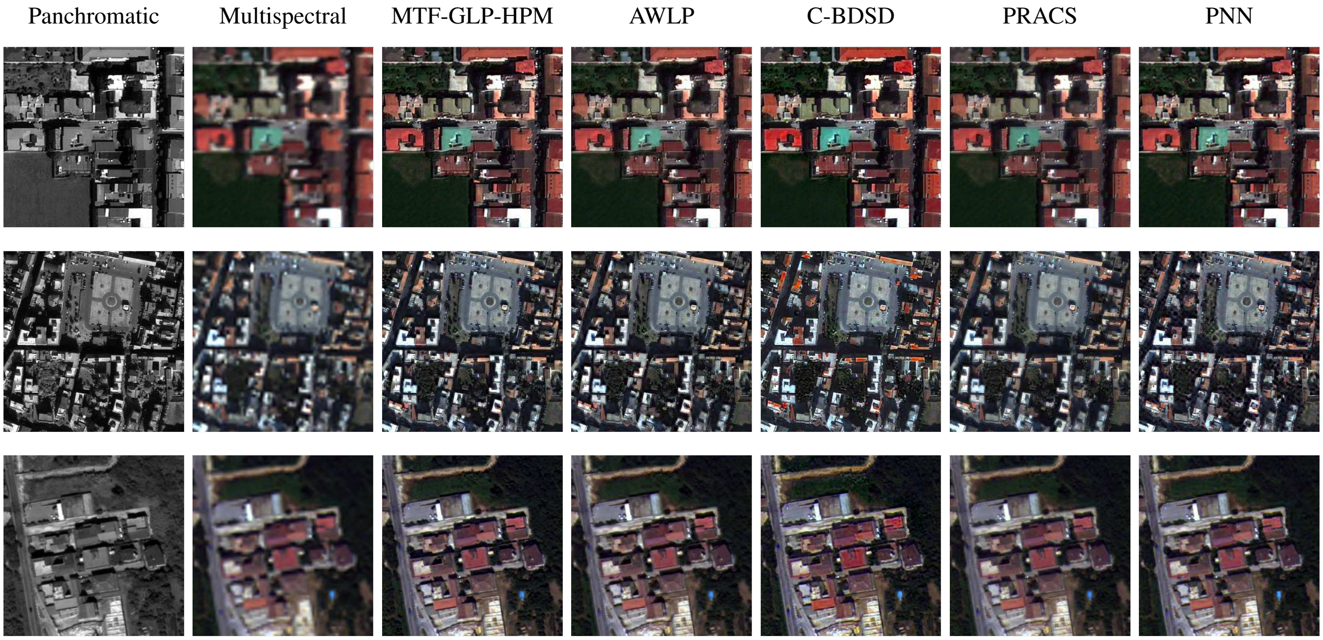

Fig. 3. Comparison of pansharpened images from selected techniques. Top to bottom: GeoEye-1, IKONOS, WorldView-2.

	Q4	**Q**	**SAM**	ERGAS	**SCC**	**D**$_\lambda$	**D**$_S$	**QNR**
	$\to 1$	$\to 1$	$\to 0$	$\to 0$	$\to 1$	$\to 0$	$\to 0$	$\to 1$
A	.649	.832	3.256	2.861	.846	.090	.125	.797
B	.643	.824	3.674	2.947	.829	.082	.121	.807
C	.724	.854	3.637	2.480	.809	.132	.193	.701
D	.691	.850	3.277	2.597	.810	.138	.163	.723
E	.697	.852	3.240	11.332	.727	.133	.157	.732
F	.582	.779	3.520	3.151	.798	.073	.089	.844
G	.600	.790	3.554	3.072	.794	.071	**.071**	.863
H	.719	.865	3.255	2.413	.825	.148	.182	.699
I	.728	.867	3.268	2.365	.826	.155	.187	.688
J	.731	.854	3.651	2.486	.807	.130	.182	.712
K	.735	.871	**3.220**	5.034	.788	.152	.181	.695
L	.706	.847	3.566	8.620	.792	.182	.196	.659
M	.717	.861	3.629	2.613	.787	.124	.152	.743
N	.574	.777	3.536	3.548	.760	.127	.126	.765
O	**.739**	**.883**	3.338	**2.234**	.852	.049	.099	.857
P	**.739**	.878	3.481	2.437	**.859**	.083	.134	.795
Q	.699	.856	3.236	2.429	.811	**.047**	.087	**.869**
X	**.809**	**.940**	**2.131**	**1.566**	**.915**	**.032**	**.061**	**.908**

Table II. Performance comparison on the GeoEye-1 dataset.

	Q4	**Q**	**SAM**	ERGAS	**SCC**	**D**$_\lambda$	**D**$_S$	**QNR**
	$\to 1$	$\to 1$	$\to 0$	$\to 0$	$\to 1$	$\to 0$	$\to 0$	$\to 1$
A	.606	.750	3.459	2.727	.871	.113	.218	.695
B	.616	.767	3.358	2.669	.884	.087	.194	.737
C	.716	.833	2.927	2.064	.898	.120	.194	.712
D	.688	.823	3.017	2.255	.893	.139	.204	.686
E	.693	.829	2.949	2.186	.901	.138	.199	.691
F	.517	.702	3.489	3.119	.809	.113	.152	.752
G	.557	.724	3.580	3.032	.818	.124	.145	.749
H	.705	.832	2.969	2.154	.901	.149	.218	.666
I	.712	.834	2.942	2.091	.904	.154	.225	.657
J	.718	.834	2.922	2.055	.898	.124	.187	.715
K	.717	.842	2.882	2.055	.907	.152	.218	.664
L	.686	.808	3.124	2.224	.902	.189	.247	.612
M	.714	.838	**2.842**	2.112	.906	.138	.195	.695
N	.592	.766	3.280	2.796	.850	.126	.161	.734
O	.719	**.857**	2.914	**1.985**	.908	**.039**	**.088**	**.876**
P	**.720**	.856	2.910	2.055	**.916**	.071	.121	.817
Q	.659	.802	2.993	2.359	.873	**.049**	.114	.842
X	**.760**	**.900**	**2.283**	**1.663**	**.941**	.051	**.073**	**.879**

Table III. Performance comparison on the Ikonos dataset.

IV. CONCLUSIONS

In this paper we have proposed to use CNNs to address the pansharpening task. To improve performance we augmented the input by including several maps of nonlinear radiometric indices. We tested the proposed method against a number of state-of-the-art references obtaining a very good performance under all metrics, both full-reference and no-reference, and also in terms of subjective quality. Future research on this topic will exploit the full potential of deep learning. In particular, we will test the use of further external inputs, such as textural features [22] or information derived from external segmenters [23].

V. REFERENCES

[1] G. Vivone, L. Alparone, J. Chanussot, M. Dalla Mura, A. Garzelli, G. A. Licciardi, R. Restaino, and L. Wald, "A critical comparison among pansharpening algorithms," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 5, pp. 2565–2586, May 2015.

	Q4	**Q**	**SAM**	ERGAS	**SCC**	**D$_\lambda$**	**D$_S$**	**QNR**
	→ 1	→ 1	→ 0	→ 0	→ 1	→ 0	→ 0	→ 1
A	.694	.828	3.778	2.991	.884	.052	.111	.842
B	.715	.848	3.826	2.903	.884	.037	.109	.857
C	.821	.895	3.737	2.203	.865	.062	.096	.847
D	.789	.891	3.503	2.410	.882	.060	.082	.862
E	.791	.893	3.500	2.383	.887	.059	.080	.865
F	.636	.789	3.975	3.380	.831	.053	.091	.860
G	.703	.818	4.065	3.160	.839	.067	.074	.862
H	.811	.903	3.452	2.226	.891	.077	.095	.834
I	.821	.905	3.431	2.130	.897	.077	.097	.833
J	.822	.896	3.689	2.175	.867	.064	.090	.851
K	**.824**	**.908**	3.449	**2.091**	**.901**	.075	.095	.837
L	.795	.881	3.840	2.439	.885	.115	.117	.782
M	.812	.904	**3.418**	2.256	.897	.066	.084	.854
N	.692	.837	3.726	3.202	.840	.055	.064	.883
O	.811	.905	3.744	2.264	.891	.048	**.038**	.915
P	.800	.894	3.989	2.636	.894	.025	**.045**	**.930**
Q	.790	.878	3.699	2.410	.852	**.023**	.073	.905
X	**.851**	**.944**	**2.576**	**1.602**	**.939**	**.020**	.048	**.932**

Table IV. Performance comparison on the WorldView-2 dataset.

[2] Alan R Gillespie, Anne B Kahle, and Richard E Walker, "Color enhancement of highly correlated images. II. Channel ratio and "chromaticity" transformation techniques," *Remote Sensing of Environment*, vol. 22, no. 3, pp. 343 – 365, 1987.

[3] C.A. Laben. and B.V. Brower, "Process for enhancing the spatial resolution of multispectral imagery using pan-sharpening," *U.S. Patent 6011875, 2000.*, 2000.

[4] 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.

[5] J. Choi, K. Yu, and Y. Kim, "A new adaptive component-substitution-based satellite image fusion by using partial replacement," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 1, pp. 295–309, Jan 2011.

[6] B. Aiazzi, L. Alparone, S. Baronti, and A. Garzelli, "Context-driven fusion of high spatial and spectral resolution images based on oversampled multiresolution analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 40, no. 10, pp. 2300–2312, Oct 2002.

[7] T. Ranchin and L. Wald, "Fusion of high spatial and spectral resolution images: the ARSIS concept and its implementation," *Photogrammetric engineering and remote sensing*, vol. 66, no. 1, pp. 49–61, 2000.

[8] J.G. Liu, "Smoothing filter based intensity modulation: A spectral preserve image fusion technique for improving spatial details," *International Journal of Remote Sensing*, vol. 21, no. 18, pp. 3461–3472, 2000.

[9] M.M. Khan, J. Chanussot, L. Condat, and A. Montanvert, "Indusion: Fusion of multispectral and panchromatic images using the induction scaling technique," *IEEE Geoscience and Remote Sensing Letters*, vol. 5, no. 1, pp. 98–102, Jan 2008.

[10] G. Vivone, R. Restaino, M. Dalla Mura, G. Licciardi, and J. Chanussot, "Contrast and error-based fusion schemes for multispectral image pansharpening," *IEEE Geoscience and Remote Sensing Letters*, vol. 11, no. 5, pp. 930–934, May 2014.

[11] B. Aiazzi, L. Alparone, S. Baronti, A. Garzelli, and M. Selva, "An MTF-based spectral distortion minimizing model for pan-sharpening of very high resolution multispectral images of urban areas," in *GRSS/ISPRS Joint Workshop on Remote Sensing and Data Fusion over Urban Areas*, May 2003, pp. 90–94.

[12] L. Alparone, L. Wald, J. Chanussot, C. Thomas, P. Gamba, and L.M. Bruce, "Comparison of pansharpening algorithms: Outcome of the 2006 GRS-S Data-Fusion Contest," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 10, pp. 3012–3021, Oct 2007.

[13] J. Lee and C. Lee, "Fast and efficient panchromatic sharpening," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 1, pp. 155–163, Jan 2010.

[14] H. Wang, S. Chen, F. Xu, and Y.-Q. Jin, "Application of deep learning algorithms to MSTAR data," in *IEEE IGARSS*, 2015, pp. 3743–3745.

[15] M. Castelluccio, G. Poggi, C. Sansone, and L. Verdoliva, "Land use classification in remote sensing images by convolutional neural networks," in *arXiv:1508.00092*, 2015.

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

[17] C. Dong, C.C. Loy, K. He, and X. Tang, "Image super-resolution using deep convolutional networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, no. 2, pp. 295–307, Feb 2016.

[18] P.S. Chavez and J.A. Anderson, "Comparison of three different methods to merge multiresolution and multispectral data: Landsat TM and SPOT panchromatic," *Photogrammetric Engineering and Remote Sensing*, vol. 57, no. 3, pp. 295 – 303, 1991.

[19] X. Otazu, M. Gonzalez-Audicana, O. Fors, and J. Nunez, "Introduction of sensor spectral response into image fusion methods. application to wavelet-based methods," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 10, pp. 2376–2385, Oct 2005.

[20] A. Garzelli, F. Nencini, and L. Capobianco, "Optimal MMSE pan sharpening of very high resolution multispectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 1, pp. 228–236, Jan 2008.

[21] A. Garzelli, "Pansharpening of multispectral images based on nonlocal parameter optimization," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 4, pp. 2096–2107, April 2015.

[22] R. Gaetano, G. Scarpa, and G. Poggi, "Recursive texture fragmentation and reconstruction segmentation algorithm applied to VHR images," in *IEEE IGARSS*, 2009, pp. IV101–IV104.

[23] R. Gaetano, G. Masi, G. Scarpa, and G. Poggi, "A marker-controlled watershed segmentation: Edge, mark and fill," in *IEEE IGARSS*, 2012, pp. 4315–4318.