CFC-MCRF: Multiresolution Fusion and Segmentation of Hyperspectral and Panchromatic Remote Sensing Images with deep learning and CRFs

1st Martina Pastorino
DITEN, University of Genoa
16145 Genoa, Italy
martina.pastorino@edu.unige.it

2nd Gabriele Moser
DITEN, University of Genoa
16145 Genoa, Italy
gabriele.moser@unige.it

3rd Sebastiano B. Serpico
DITEN, University of Genoa
16145 Genoa, Italy
sebastiano.serpico@unige.it

4th Josiane Zerubia
INRIA, Université Côte d'Azur
06902 Sophia Antipolis, France
josiane.zerubia@inria.fr

***Abstract*—This paper proposes a multiresolution supervised approach for the joint classification and fusion of panchromatic and hyperspectral data. The approach combines a fully convolutional network (FCN) and a hierarchical conditional random field (CRF) for multiresolution feature extraction and fusion. The goal of the proposed methodology is to produce a land cover map at the fine spatial resolution of the panchromatic channel, while integrating the rich spectral information of the hyperspectral channels. The deep learning architecture consists of a two-branch encoder, separately mining the spatial-spectral information provided by the different input data, and a decoder which performs the first multiresolution fusion. The posterior probability estimates computed by the FCN at the panchromatic and hyperspectral levels are further fused and modeled by a multiresolution CRF model. This CRF approximates the behavior of the ideal fully connected CRF in a computationally tractable manner through a clustering step, with the aim of integrating not only local but also long-distance spatio-spectral relationships. The proposed technique has been applied to PRISMA data from the Italian Space Agency for the land cover mapping of an area in Emilia Romagna, Italy.**

***Index Terms*—hyperspectral imagery, multiresolution fusion, semantic segmentation, FCN, CRF**

I. Introduction

Hyperspectral image classification is a highly investigated area of remote sensing, for which several techniques providing very accurate results have been developed over the course of the years. In particular, several approaches for supervised classification of hyperspectral data based on the analysis of spatial and spectral information have been proposed [1–11] employing different image processing techniques, such as Markov random fields (MRFs), conditional random fields (CRFs), dictionary learning, support vector machines, and deep learning.

Unlike the classification of single-resolution hyperspectral data, panchromatic-hyperspectral multiresolution classification is a very new problem, not largely addressed in the literature so far. Panchromatic-multispectral classification techniques have been developed through MRF-based approaches [12, 13], deep learning [14], and pyramid representations [15], but with reference only to sensors characterized by a limited number of spectral channels. RGB-hyperspectral classification has also been addressed through pyramidal models [16].

This paper proposes a method for the multiresolution supervised classification of hyperspectral-panchromatic data, with the goal of generating a classification map at the fine spatial resolution of the panchromatic channel, while taking advantage of the rich spectral information contained in the hyperspectral channels. The resulting multiresolution fusion problem is very promising from the point of view of the extraction of thematic information, thanks to the opportunity to benefit from both spatial and spectral information captured by the two different sensors.

The proposed multiresolution fusion technique for classification is based on deep learning and probabilistic graphical models. Deep neural networks have proven to be greatly effective in satellite image classification. Convolutional neural network (CNN) and fully convolutional network (FCN) architectures are characterized by a multiresolution structure, with convolutional and pooling layers [17]. This structure typical of two-dimensional CNNs naturally extends to the three-dimensional datacube associated with a hyperspectral image. Such extensions make use of 3D (3D-CNN) or 1D convolutions along the channel set (equivalent to the wavelength axis; 1D-CNN) [18].

The proposed method includes an FCN devoted to the integration of the hyperspectral and panchromatic information at the corresponding native spatial resolution, and to the estimation of the posterior probabilities of the thematic classes, given both the hyperspectral and panchromatic observations. The network accounts for the first step of multiresolution information fusion, which is then completed by the integration with the proposed multiresolution CRF. The proposed method also aims to address, in the context of multiresolution panchromatic-hyperspectral classification, the challenge

This work was partially supported by the Italian Space Agency (ASI) within the Framework of the Project under Grant PRISMA-Learn-ASI no. 2022-12-U.0. PRISMA Products, ©ASI, delivered under a license to use by ASI.

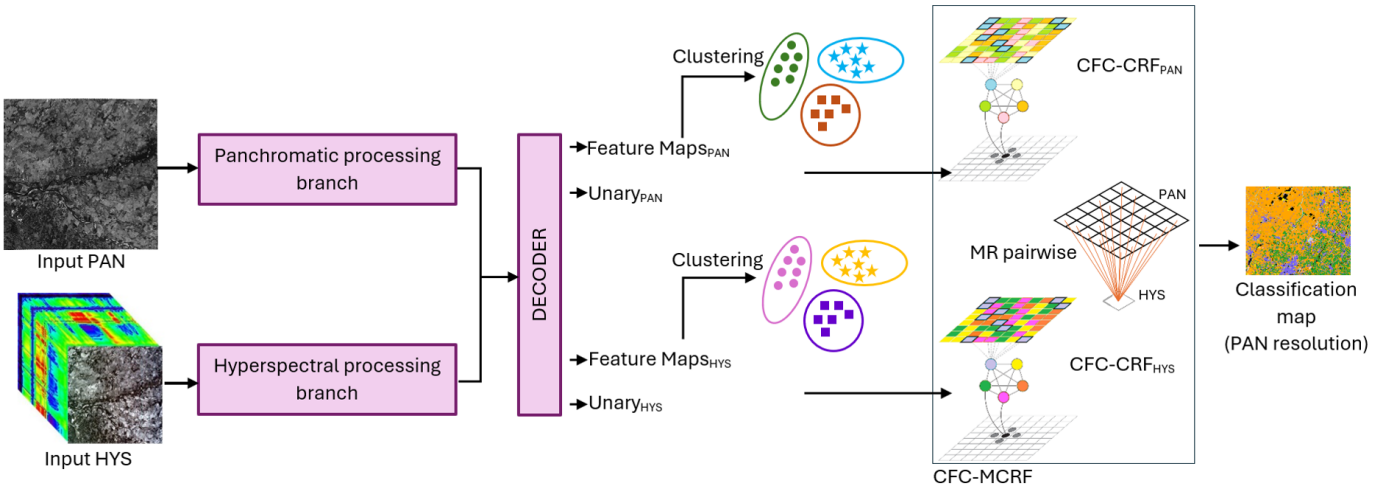

Fig. 1. Overall architecture of the proposed methodology. CFC-MCRF connects each pixel (black circle) with its neighbors (gray circles) in the single resolution lattice and with the h most similar clusters (colored circles), which are fully connected, and then connects the pixels at multiple resolutions.

of scarce training sets typical of realistic remote sensing applications, by incorporating not only local but also long-range spatial-spectral relations with the CRF. The developed CRF model aims to approximate the behavior of an ideal fully connected CRF, in a computational tractable way. A set of clusters is computed on the two sets of posterior probabilities estimated by the network at the multiple resolutions, each cluster corresponds to a virtual node. The CRF characterizes the interactions between (i) neighboring pixels at the same resolution, (ii) virtual nodes in a fully connected mesh at the same resolution, (iii) pixels and virtual nodes in a dedicated neighborhood system at the same resolution, and (iv) parent-child pixels at multiple resolution. The application of state-of-the-art graph-based minimum energy algorithms, involving graph cut [19], ensures convergence to solutions characterized by strong optimality conditions.

The technique extends the previous method in [20] for the classification of single-resolution RGB optical aerial images with a spatial resolution of a few centimeters, to the classification of multiresolution panchromatic-hyperspectral satellite images, and the method in [21] through the hierarchical CRF.

II. Methodology

The proposed methodology performs multiresolution and multispectral image data fusion for semantic segmentation through a novel spatial-spectral neural network architecture and a cluster fully connected multiresolution CRF (CFC-MCRF, see Fig. 1). The focus is on panchromatic, $\mathcal{X}^p \in \mathbb{R}^{1\times H\times W}$ and hyperspectral, $\mathcal{X}^h \in \mathbb{R}^{C_h\times \frac{H}{r}\times \frac{W}{r}}$ acquisitions (r is the spatial resolution ratio, and the ratios H/r and W/r are assumed to be integer), with the goal of leveraging the complementary strengths of high spatial resolution (panchromatic) data and rich spectral resolution (hyperspectral) data.

Overall, the methodology combines multiresolution data fusion through a deep learning-based feature extraction, clustering, and CRF-based long-range spatial dependencies modeling, resulting in an effective framework for semantic segmentation of panchromatic and hyperspectral data. A further description is provided in the following subsections.

A. Spatial-spectral FCN architecture

The deep learning module is composed of an FCN [22] characterized by a two-branch encoder and a decoder. The encoder processes the panchromatic and hyperspectral information separately: one branch focuses on the high spatial resolution panchromatic data, X_p, through a series of convolutions and downsampling maxpooling layers, while the other handles the high spectral resolution hyperspectral data, X_h, through 3D pointwise convolutions followed by max pooling along the spectral dimension. The 3D convolutions aim at extracting a meaningful representation from the hyperspectral channels, to benefit from its rich spectral content.

The features extracted from both branches are fused in a bottleneck layer, which learns the combined representations of spatial and spectral data. To ensure a match between the two original spatial resolutions, the sequence of pooling layers have an overall size which is a multiple of the original resolution ratio. The decoder reconstructs the segmentation map by progressively upscaling the features.

The network produces feature maps ($\mathcal{F}^p$ and $\mathcal{F}^h$) and posterior probabilities (P^p and P^h) at the native resolutions of the panchromatic and hyperspectral channels. The set derived from hyperspectral data processing is the output of the hyperspectral branch of the encoder, whereas the set at the fine resolution of the panchromatic image is the output of the decoder of the network.

B. Cluster fully connected multiresolution CRF (CFC-MCRF)

In order to compute the virtual nodes at multiple resolutions, a clustering through k-means is applied, separately, to the multiresolution feature maps $\mathcal{F}^p$ and $\mathcal{F}^h$. This step effectively joins similar pixels all over the feature maps in the same cluster, thus allowing connections through points at any distance

on the image itself. The result is two sets of clusters $\mathcal{C}^p$ and $\mathcal{C}^h$. For each cluster c in each of the two sets, the cluster feature vector $x_c \in \mathbb{R}^n$ is defined as the centroid of the cluster, and the label $y_c \in \Omega$ (Ω being the set of thematic classes) is given by the maximum of the posterior probability computed by the neural network averaged over cluster c.

Let $\mathcal{I}^p$ and $\mathcal{I}^h$ be pixel lattices associated with the panchromatic and the hyperspectral images, respectively, and let $\mathcal{S}(i')$ be the patch of pixels at the fine resolution corresponding to one pixel $i' \in \mathcal{I}^h$ at coarser resolution. The proposed multiresolution CRF model is a multiresolution extension of the methodology proposed in [20], and the total energy is defined as follows:

$$\mathcal{U}(\mathcal{Y}^p, \mathcal{Y}^h | \mathcal{X}^p, \mathcal{X}^h) = \mathcal{E}^p(\mathcal{Y}^p|\mathcal{X}^p) + \mathcal{E}^h(\mathcal{Y}^h|\mathcal{X}^h) + \xi \sum_{\substack{i' \in \mathcal{I}^h \\ i \in \mathcal{S}(i') \subset \mathcal{I}^p}} V_{ii'}(y_i^p, y_{i'}^h | f_i^p, f_{i'}^h) \quad (1)$$

where the first two terms are single resolution energy functions corresponding to the cluster-based approximated fully connected CRF (CFC-CRF [20]) at the panchromatic and the hyperspectral resolutions, respectively, and the last term is a further pairwise potential between the labels at the finer spatial resolution (y_i, $i \in \mathcal{S}(i') \subset \mathcal{I}^p$) and the corresponding label at coarser resolution ($y_{i'}$, $i' \in \mathcal{I}^h$). This further potential is defined as a standard Potts model [23].

The random fields of observations and labels are $\mathcal{X}^\star = \{f_i^\star, x_c^\star\}_{i \in \mathcal{I}^\star, c \in \mathcal{C}^\star}$ and $\mathcal{Y}^\star = \{y_i^\star, y_c^\star\}_{i \in \mathcal{I}^\star, c \in \mathcal{C}^\star}$, respectively. The $\star$ represents the two possible different input lattices ($\star = p$ or h) and $f_i^\star$ is the value of the corresponding feature map extracted by the FCN on pixel i. It is important to note that, while the pixelwise posterior probability may be different on the two image lattices, the set Ω of thematic classes is the same at both resolutions.

The two single resolution energy functions (the first two terms of (1)) describe a CRF model with up to pairwise nonzero clique potentials, approximating a fully connected behavior through a clustering partition. The $\star$ has been omitted in for brevity:

$$\mathcal{E}(\mathcal{Y}|\mathcal{X}) = \sum_{i \in \mathcal{I}} D_i(y_i|f_i) + \lambda_{\mathcal{I}\mathcal{I}} \sum_{\substack{j \in \partial i \\ i \in \mathcal{I}}} V_{ij}(y_i, y_j | f_i, f_j) + \gamma \sum_{c \in \mathcal{C}} D_c(y_c|x_c) + \lambda_{\mathcal{C}\mathcal{C}} \sum_{\substack{c,d \in \mathcal{C} \\ c \neq d}} V_{cd}(y_c, y_d | x_c, x_d) + \lambda_{\mathcal{I}\mathcal{C}} \sum_{\substack{i \in \mathcal{I} \\ c \in \bar{\partial} i}} V_{ic}(y_i, y_c | f_i, x_c). \quad (2)$$

D_i and D_c are the unary potentials for the pixel and cluster layers, and are computed as the pixelwise log-posterior probability predicted by the FCN (the softmax). V_{ij} is the pixel pairwise potential and enforces spatial smoothness between neighboring pixels, V_{cd} prompts similar clusters to be assigned to the same class, and V_{ic} is the pixel-cluster pairwise potential. The pairwise potentials in equation (2) are defined by a contrast-sensitive Potts potential [23] to favor consistency in the labeling while simultaneously weighting on the similarity among the corresponding features. γ, ξ, and the λ-terms are weights to balance the contributions of each term.

The energy function is minimized through the $\alpha - \beta$ swap graph cut method [24] which decomposes a multiclass inference problem in a sequence of binary ones. It converges to a local minimum with strong optimality properties [24–26].

III. Experimental Results

A. Dataset and experimental setup

The experimental validation was conducted with a dataset acquired over Emilia Romagna, Italy, in mainly urban areas, consisting of three PRISMA images collected in summer 2022. The PRISMA acquisitions include a 5-m spatial resolution panchromatic image and a 30-m spatial resolution hyperspectral image with 234 spectral bands. Hence, the resolution ratio is equal to 6. The ground truth was derived with a regional land use archive[1]. The land cover classes correspond to different vegetated areas (cultivated fields, low and high vegetation), water bodies and built-up areas (buildings and roads). The dataset was split into two disjoint subsets for training and testing.

The parameters of the k-means clustering and the CRF are experimentally set through trial and error and are reported in Table I. The standard deviation σ of the Gaussian kernel in the contrast sensitive Potts pairwise potential is the median Euclidean distance between all considered pairs of feature vectors. The number h of nearest centroids (see Fig. 1) is set to 4 to capture long-range information while keeping a rather low computational cost. More details on the meaning and setting of these parameters are presented in [20].

The proposed spatial-spectral FCN computes the unary potentials – the posterior probabilities – at the two different resolutions. Two different architectures were considered, both deriving from a U-Net [27] architecture, with a double branch encoder including channelwise spectral convolutions. One presents a lower number of convolutional blocks and filters (see Table II for the number of trainable parameters), in the following, they will be referred to as FCN_λ and (lightweight) L-FCN_λ. The networks are trained on an RTX2080Ti GPU.

[1] https://geoportale.regione.emilia-romagna.it

TABLE I
Parameters of the proposed CFC-MCRF method.

Parameter	Panchromatic	Hyperspectral
$\lambda_{\mathcal{I}\mathcal{I}}$ (pixel pairwise)	2	2
$\lambda_{\mathcal{C}\mathcal{C}}$ (cluster pairwise)	1	1
$\lambda_{\mathcal{I}\mathcal{C}}$ (pixel-cluster pairwise)	1	1
γ (cluster unary)	$\lvert\mathcal{I}^p\rvert/k^p = 5625$	$\lvert\mathcal{I}^h\rvert/k^h = 2500$
$\mathcal{I}$ (image patch)	600×600	100×100
k (number of clusters)	64	4
h (nearest centroids)	4	4

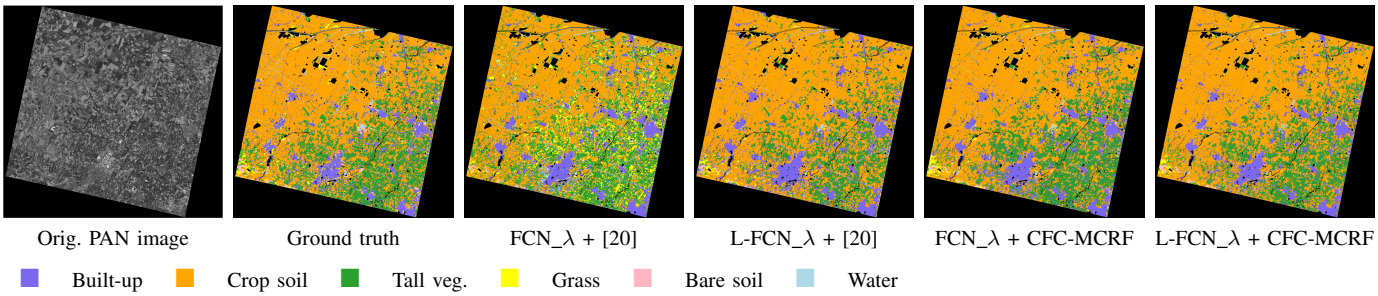

Fig. 2. Test ground truths and classification maps for the test tiles in the PRISMA Emilia Romagna dataset. Original PRISMA Product and processed under a license of ASI ©ASI (2022).

TABLE II
TEST-SET RESULTS FOR THE EMILIA ROMAGNA DATASET. PER-CLASS VALUES ARE RECALLS. OVERALL ACCURACY (OA), RECALL, PRECISION, AND F1 SCORE ARE AVERAGED OVER THE CLASSES.

Architecture	built-up	crop soil	tall veg.	grass	bare soil	water	OA	recall	precision	F1 score	# trainable param.
FCN_λ	81.30	81.97	78.07	8.37	26.51	27.53	77.54	50.63	67.91	58.00	$17{\cdot}10^6$
L-FCN_λ	86.44	91.82	62.34	15.17	20.62	30.05	81.16	51.08	68.03	58.34	$38{\cdot}10^4$
FCN_λ + CFC-CRF [20]	81.57	82.12	**90.48**	18.95	26.52	**49.40**	80.97	58.18	65.52	61.63	$17{\cdot}10^6$
L-FCN_λ + CFC-CRF [20]	**87.03**	94.89	65.07	22.57	22.46	30.05	83.86	53.68	70.43	60.92	$38{\cdot}10^4$
FCN_λ + CFC-MCRF	85.73	92.73	83.15	**27.99**	29.80	41.86	86.57	**60.21**	79.51	**68.53**	$17{\cdot}10^6$
L-FCN_λ + CFC-MCRF	83.98	**96.91**	81.15	14.80	**31.95**	42.40	**88.19**	58.53	**80.74**	67.87	$38{\cdot}10^4$

B. Discussion

The results obtained on the PRISMA dataset are shown in Fig. 2 and in Table II, in terms of recall for each class, overall accuracy, and class-averaged recall, precision, and F1 score.

As can be seen from the results in Table II, the proposed method, when combined with both neural architectures, effectively distinguishes the three primary land cover classes: built-up areas ("built-up") and vegetated areas ("crop soil" and "trees"). The results are generally less accurate for the minority classes, "bare soil", "grass" , and "water." This is primarily because the ground truth for this region is derived from a land-use archive with a lower spatial resolution compared to the original pachromatic resolution. Consequently, the minority classes identified have few available training samples, leading to less accurate classification results. Furthermore, the regional archive dates back to 2020, causing a misalignment between the nominal and effective extent of the land cover classes, in particular for the ones subject to temporal changes (e.g., the extent of low vegetation and water bodies). In general, the results of L-FCN_λ are slightly more accurate.

The two networks were integrated with the proposed CFC-MCRF and with the CFC-CRF proposed in [20]. The latter was employed as a comparison technique. As it was formulated for single resolution optical imagery, it was applied to a multiscale single-resolution tensor constructed with the feature maps of the panchromatic and the hyperspectral channels resampled at a common spatial resolution (5 m) and undergoing feature reduction. In both cases, the integration of the CRF, with its ability to model both local and long-range spatial-spectral relationships through pixel neighborhoods and clusters, resulted in improvements in the accuracy of all considered classes and in the average accuracies. In particular, higher overall classification accuracies can be noted with the proposed multiresolution CFC-MCRF, leading to an overall accuracy of 88%, precision of 80% and F1 score of about 70%, despite the aforementioned issues remarked in the dataset.

Fig. 2 displays the classification maps obtained using the two different neural networks with the addition of the CFC-CRF in [20] and the proposed CFC-MCRF. The results obtained by the method closely reflect the ground truth, particularly for the majority classes, with no significant over-smoothing or evident false alarms (except for the class "grass" with the FCN_λ + [20]). Notably, the proposed method demonstrates its ability to generate classification maps where the spatial boundaries between classes are well-defined.

IV. CONCLUSION

This paper presented a supervised method for the joint fusion and classification – or semantic segmentation – of panchromatic and hyperspectral images. The proposed method combines FCNs and a multiresolution cluster level fully connected CRF to model the spatial and spectral information provided by the multiresolution and hyperspectral input data and to generate a classification map at the finer of the two observed spatial resolutions.

The experimental validation was conducted with PRISMA images for the land cover mapping of a mainly urban zone. The results suggest the effectiveness of the proposed technique in fusing and classifying hyperspectral-panchromatic imagery at the finest spatial resolution. In particular, the integration of the proposed CRF improves classification accuracy, as compared to using a purely neural model.

While the proposed architecture achieves accurate overall results, the discrimination of minority classes still remains a challenge in datasets with limited training samples or outdated ground truth data. Future work will focus on improving the

robustness of the framework against such issues, integrating more advanced contextual models, and exploring semi-supervised techniques to address sample scarcity.

REFERENCES

[1] E. Kordi Ghasrodashti, M. S. Helfroush, and H. Danyali, "Sparse-based classification of hyperspectral images using extended hidden Markov random fields," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 11, no. 11, pp. 4101–4112, 2018.

[2] X. Cao, X. Wang, D. Wang, J. Zhao, and L. Jiao, "Spectral–spatial hyperspectral image classification using cascaded Markov random fields," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 12, no. 12, pp. 4861–4872, 2019.

[3] O. Okwuashi and C. E. Ndehedehe, "Deep support vector machine for hyperspectral image classification," *Pattern Recognition*, vol. 103, p. 107298, 2020.

[4] B. Pan, X. Xu, Z. Shi, N. Zhang, H. Luo, and X. Lan, "Dssnet: A simple dilated semantic segmentation network for hyperspectral imagery classification," *IEEE Geoscience and Remote Sensing Letters*, vol. 17, no. 11, pp. 1968–1972, 2020.

[5] H. Sun, X. Zheng, and X. Lu, "A supervised segmentation network for hyperspectral image classification," *IEEE Transactions on Image Processing*, vol. 30, pp. 2810–2825, 2021.

[6] Y. Liang, X. Zhao, A. J. X. Guo, and F. Zhu, "Hyperspectral image classification with deep metric learning and conditional random field," *IEEE Geoscience and Remote Sensing Letters*, vol. 17, no. 6, pp. 1042–1046, 2020.

[7] Z. Zhong, J. Li, D. A. Clausi, and A. Wong, "Generative adversarial networks and conditional random fields for hyperspectral image classification," *IEEE Transactions on Cybernetics*, vol. 50, no. 7, pp. 3318–3329, 2020.

[8] Z. Gong, L. Tong, J. Zhou, B. Qian, L. Duan, and C. Xiao, "Superpixel spectral–spatial feature fusion graph convolution network for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–16, 2022.

[9] W. Zhu, C. Zhao, S. Feng, and B. Qin, "Multiscale short and long range graph convolutional network for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–15, 2022.

[10] X. Zhao, J. Niu, C. Liu, Y. Ding, and D. Hong, "Hyperspectral image classification based on graph transformer network and graph attention mechanism," *IEEE Geoscience and Remote Sensing Letters*, vol. 19, pp. 1–5, 2022.

[11] M. Ahmad, S. Shabbir, S. K. Roy, D. Hong, X. Wu, J. Yao, A. M. Khan, M. Mazzara, S. Distefano, and J. Chanussot, "Hyperspectral image classification—traditional to deep models: A survey for future prospects," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 15, pp. 968–999, 2022.

[12] G. Storvik, R. Fjortoft, and A. H. S. Solberg, "A Bayesian approach to classification of multiresolution remote sensing data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, no. 3, pp. 539–547, 2005.

[13] G. Moser, A. De Giorgi, and S. B. Serpico, "Multiresolution supervised classification of panchromatic and multispectral images by Markov random fields and graph cuts," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 9, pp. 5054–5070, 2016.

[14] S. Liu, H. Zhao, Q. Du, L. Bruzzone, A. Samat, and X. Tong, "Novel cross-resolution feature-level fusion for joint classification of multispectral and panchromatic remote sensing images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–14, 2022.

[15] W. Ma, J. Shen, H. Zhu, J. Zhang, J. Zhao, B. Hou, and L. Jiao, "A novel adaptive hybrid fusion network for multiresolution remote sensing images classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–17, 2022.

[16] Q. Xu, X. Yuan, C. Ouyang, and Y. Zeng, "Attention-based pyramid network for segmentation and classification of high-resolution and hyperspectral remote sensing images," *Remote Sensing*, vol. 12, no. 21, 2020.

[17] I. Goodfellow, Y. Bengio, and A. Courville, *Deep learning*. Boston, Massachussetts: USA: MIT Press, 2016.

[18] C. Pelletier, G. I. Webb, and F. Petitjean, "Temporal convolutional neural network for the classification of satellite image time series," *Remote Sensing*, vol. 11, no. 5, 2019.

[19] V. Kolmogorov and C. Rother, "Minimizing nonsubmodular functions with graph cuts-a review," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 7, pp. 1274–1279, 2007.

[20] L. Maggiolo, D. Marcos, G. Moser, S. B. Serpico, and D. Tuia, "A semisupervised CRF model for CNN-based semantic segmentation with sparse ground truth," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–15, 2022.

[21] M. Pastorino, G. Moser, S. B. Serpico, and J. Zerubia, "Multiresolution fusion and classification of hyperspectral and panchromatic remote sensing images," in *IEEE/CVF Winter Conference on Applications of Computer Vision Workshops (WACVW 2025)*, 2025.

[22] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, pp. 3431–3440, 2015.

[23] Z. Kato and J. Zerubia, "Markov random fields in image segmentation," *Found. Trends Signal Process.*, vol. 5, no. 1-2, pp. 1–155, 2012.

[24] Y. Boykov, O. Veksler, and R. Zabih, "Fast approximate energy minimization via graph cuts," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 11, pp. 1222–1239, 2001.

[25] V. Kolmogorov and R. Zabin, "What energy functions can be minimized via graph cuts?" *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 2, pp. 147–159, 2004.

[26] Y. Boykov and V. Kolmogorov, "An experimental comparison of min-cut/max- flow algorithms for energy minimization in vision," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 9, pp. 1124–1137, 2004.

[27] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Med. Image Comput. Comput. Ass. Interv.*, ser. LNCS, vol. 9351. Springer, pp. 234–241, 2015.