



HAL
open science

Multimission, Multifrequency, and Multiresolution SAR Image Classification through Hierarchical Markov Models and Convolutional Networks

Martina Pastorino, Gabriele Moser, Sebastiano B Serpico, Josiane Zerubia

► **To cite this version:**

Martina Pastorino, Gabriele Moser, Sebastiano B Serpico, Josiane Zerubia. Multimission, Multifrequency, and Multiresolution SAR Image Classification through Hierarchical Markov Models and Convolutional Networks. IEEE Geoscience and Remote Sensing Letters, 2024, 21, pp.5. 10.1109/lgrs.2024.3414284 . hal-04613394

HAL Id: hal-04613394

<https://inria.hal.science/hal-04613394>

Submitted on 16 Jun 2024

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.



Distributed under a Creative Commons Attribution 4.0 International License

Multimission, Multifrequency, and Multiresolution SAR Image Classification through Hierarchical Markov Models and Convolutional Networks

Martina Pastorino, *Student Member, IEEE*, Gabriele Moser, *Fellow, IEEE*, Sebastiano B. Serpico, *Fellow, IEEE*, and Josiane Zerubia, *Fellow, IEEE*

Abstract—The availability of multimodal remotely sensed images calls for the development of methods capable to jointly exploit the information deriving from images acquired at different spatial resolutions, frequencies, and bands, taking advantage from their possible complementary features. This paper proposes to address this task in the case of multimission synthetic aperture radar (SAR) images, through a combination of fully convolutional networks (FCNs), hierarchical probabilistic graphical models (PGMs), and decision tree ensembles. The objective is to model the multimodal information collected at multiple spatial resolutions by distinct space missions with SAR payloads through the non-parametric formulation of FCNs and decision trees, and the spatial and multiresolution modeling capabilities of FCNs and hierarchical PGMs. The experimental validation is conducted with multimission SAR imagery acquired at X-, L-, and C-band respectively by COSMO-SkyMed, SAOCOM, and Sentinel-1 over Northern Italy. The results suggest the advantages of incorporating multifrequency radar acquisitions to reach accurate classification maps, and the multimodal fusion capabilities of the proposed methodology.

Index Terms—SAR imagery, multimission, multifrequency, multiresolution, classification, CNN, FCN, PGM, MRF

I. INTRODUCTION

THANKS to the current space missions for Earth observation, there is a significant availability of synthetic aperture radar (SAR) data, characterized by heterogeneous features such as frequencies, polarizations, and spatial resolutions [1]. This is especially relevant for land cover mapping. Multispectral images have been used extensively for this purpose, since the observations they collect in multiple spectral bands favor land-cover discrimination. However, their use is limited by cloud cover and illumination conditions. SAR does not suffer from these limitations, yet SAR data from each individual space mission are typically single-frequency with 1-to-4 channels (depending on polarimetry) [2], a scenario that limits their capability to discriminate land-covers. From this perspective, fusing data obtained from distinct SAR missions is particularly promising for land cover mapping, since it combines all-weather and day-and-night acquisition with the

possibility to jointly use multiple observations associated with distinct carrier frequencies and spatial resolutions [2]. In this framework, the challenge is to jointly exploit the heterogeneous multimodal information conveyed by multimission SAR to produce accurate semantic segmentation maps [3].

Extensive research has been conducted on semantic segmentation of multimodal remote sensing data [3]. However, in most cases, the combination of optical and SAR images has been addressed [4], [5], [6], whereas land cover mapping from multimission satellite SAR data has been scarcely investigated in the literature. For example, the paper in [7] focuses on the effect of integrating C- and X-band SAR data for the improvement of classification accuracy with a support vector machine. A few recent papers have addressed semantic segmentation of multifrequency SAR images, although collected from airborne platforms. The work in [8] presents an airborne SAR system, based on frequency-modulated continuous-wave radar and operating in X- and L-band, and shows that the merged X/L-band image, classified via maximum likelihood using the complex Wishart distribution, provides higher classification accuracy compared with the single X- and L-band images. In [9] the application of a level set method for the segmentation of multifrequency and polarimetric SAR images is investigated. The results show that the segmentation performances improve when multifrequency data are used.

This paper addresses the problem of the joint semantic segmentation of multifrequency, multipolarization, and multiresolution data collected by distinct SAR missions. For this purpose, the methodology in [10], which was previously developed for the classification of single-sensor single-resolution aerial optical imagery of urban areas, is extended here to this challenging SAR multimission case and is integrated with an information-theoretic registration algorithm to correct the spatial misalignments amongst the input multimission data. In particular, the potential of incorporating X-, L-, and C-band SAR satellite images is explored within the experimental validation of the developed method.

Specifically, according to the methodology in [10], the proposed approach integrates methodological contributions drawn from the most current families of image processing and machine learning algorithms — structured output learning, deep learning, and ensemble learning. The method combines a fully convolutional network (FCN) with a hierarchical probabilistic graphical model (PGM) on a quadtree and with decision tree ensembles. The main methodological difference

M. Pastorino, G. Moser, and S. B. Serpico are with the DITEN Department, University of Genoa, 16145, Genoa, Italy (e-mail: martina.pastorino@edu.unige.it).

M. Pastorino and J. Zerubia are with Inria, Université Côte d'Azur, 06902, Sophia-Antipolis, France.

The activity of the first three authors was partially supported by ASI in the framework of the project MultiBigSARData - ASI no. 2021-7-U.0; the support is gratefully acknowledged.

with respect to the technique in [10] is that multiresolution data fusion is necessary in the proposed approach and is addressed by generalizing the related FCN and PGM components, as described in the following of the paper.

On the one hand, the flexibility of PGMs allows the modeling of data characterized by mutual dependence relationships, which can be spatial, multiresolution, and semantic, depending on the considered context and application. Within the proposed approach, hierarchical PGM modeling, together with the aforementioned registration algorithm based on mutual information, allows fusing multimission data associated with distinct spatial resolutions.

On the other hand, the ensemble and deep learning techniques integrated in the developed methodology, thanks to their non-parametric formulation, make it possible to incorporate data with arbitrary probability distributions, thus effectively supporting the application to SAR data associated with both textured image areas, typically following K [2], Fisher [11], generalized Gaussian Rayleigh [12], and generalized Gamma distributions [13], and with areas without texture, usually well modeled by Gamma, Weibull or log-normal distributions [2].

II. METHODOLOGY

The objective of the proposed methodology is to integrate satellite SAR data associated with different resolutions, frequencies, and polarizations to obtain a land-cover map at the finest observed spatial resolution. Ideally, the input set of multimission SAR images should be acquired on the same date. In practice, a short time series of SAR images, collected by the various missions on different dates with rather small temporal intervals, can be used, under the assumption that no relevant land-cover changes have occurred during the overall acquisition timeframe. Thanks to the adoption of a quadtree topology, the proposed method also allows to obtain a set of classification maps jointly generated at all spatial resolutions of the input multimission dataset.

The architecture of the proposed method is shown in Fig. 1. First, we note that multimission images may generally be affected by spatial misalignments. Therefore, the proposed method is integrated with the algorithm in [14], with the aim of automatically registering the input SAR images acquired by the different missions. This algorithm was originally developed for large-scale optical-SAR multisensor image registration and is applied here to the case of multimission SAR data. In particular, the method combines an information-theoretic approach to area-based registration [15], based on the maximization of a mutual information functional through the Powell numerical algorithm, with a tile-based strategy for large-scale imagery. The input image is divided into a set of partially overlapping tiles, and a tile-wise processing approach defines a piecewise rotation-scale-translation mapping that approximates a globally non-homogeneous transform across the whole imaged scene, favoring spatial consistency in the overlapping areas. Details can be found in [14]. Here, it is worth noting that the image at the finest spatial resolution is used as the reference (master) image in this registration process.

After the registration module, the remaining part of the proposed architecture is based on a generalization of the

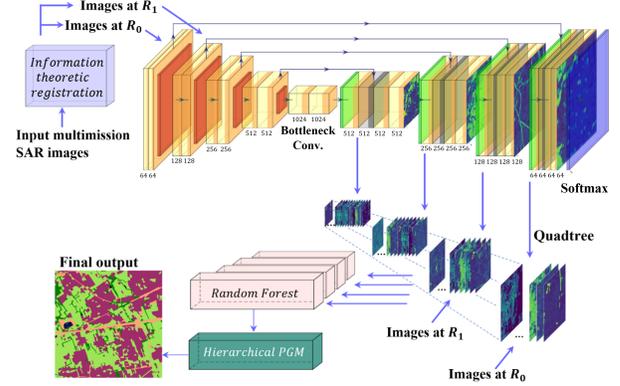


Fig. 1. Overall architecture of the proposed method in the case of input multimission images at two spatial resolutions R_0 and R_1 .

technique in [10] and comprises three main methodological components (see Fig. 1): an FCN, a hierarchical PGM on a quadtree, and L random forest (RF) classifiers [16], where L is the number of levels of the quadtree. The method is FCN-agnostic, meaning that any kind of FCN can be employed in the proposed framework, provided its encoder-decoder architecture yields a classification output with the same size of the input image. The activation function of the output layer is a softmax, to ensure that the output predictions can be formally interpreted as probabilities. The nonlinear activation function of the intermediate layers can be arbitrarily chosen (e.g., ReLU). The convolutional blocks include batch normalization, convolutional layers, and pooling layers of size 2×2 . Therefore, the inputs of two consecutive blocks have a power-of-2 relation in size. Differently from the method in [10], in which single-resolution optical imagery was the input of the FCN, here in the proposed approach, the input multiresolution SAR images resulting from the registration are fed to the neural network. Since the ratio between the different resolutions of the input images is assumed to be equal to a power of 2, depending on the original resolutions of the multimission images, resampling and antialiasing filtering may be necessary (in the example shown in Fig. 1, two resolutions R_0 and R_1 are considered). The feature maps produced by the FCN at different scales in the convolutional blocks of the decoder are extracted and inserted in a quadtree structure, matching the power-of-2 relation between resolutions. A ground truth for the training of the FCN is assumed available on the spatial finest resolution, only.

Multiresolution information is further modeled by the proposed hierarchical PGM, which consists of a combination of a hierarchical Markov random field (MRF) on a quadtree [17], [18] and a planar MRF on each level of the quadtree [19], to access both multiscale and spatial-contextual information between neighboring pixels. The levels of the quadtree topology represent different resolutions and contain both the feature maps of the hidden layers of the decoder of the FCN and the registered multimission SAR images at the same resolution.

Given the set of $l = 1, \dots, L$ pixel grids I^l at the different resolutions in the quadtree (with $l = L$ and $l = 0$ indicating the grids at the finest and the coarsest spatial resolutions, respectively), let x_i and y_i be the resulting feature vector associated

with a generic pixel i in the l -th level I^l and the corresponding class label, respectively ($i \in I^l$). If $\mathcal{Y}^l = \{y_i\}_{i \in I^l}$ is the random field of the labels on the l -th level, the hierarchical PGM is defined as [17]:

$$P(\mathcal{Y}^l | \mathcal{Y}^{l-1}) \propto \prod_{i \in I^l} P(y_i | y_h, h \lesssim i) P(y_i | y_{i^-}), \quad (1)$$

where y_{i^-} represents the label of the parent pixel $i^- \in I^{l-1}$ of $i \in I^l$ and $h \lesssim i$ indicates that pixel h is a causal neighbor of pixel i ($h, i \in I^l$). Details on the definition of causal neighbor and on the related pixel visiting scanning issues can be found in [10], [18]. Here, we only recall that the resulting model is completely causal, therefore, the inference can be performed with an efficient non-iterative algorithm, based on the marginal posterior mode (MPM) [17], [20] criterion, which requires input pixelwise posterior probabilities $P(y_i | x_i)$.

An RF classifier is used to estimate these pixelwise posterior probabilities for each level of the quadtree, in order to link the feature representation extracted by the neural network and the Bayesian inference structure of the hierarchical PGM. Here, RF is adopted as a flexible and powerful decision tree ensemble to model the pixelwise posteriors $P(y_i | x_i)$ conditioned on the heterogeneous feature vectors x_i , which generally includes feature maps from the FCN and multifrequency multipolarization SAR data. In principle, any non-parametric estimator of posterior probabilities could be used in this framework. RF was adopted thanks to its limited computational burden, robustness to overfitting, and simple model selection [16], [21]. A separate RF classifier is trained on each l -th level. The ground truth to train this RF is obtained by subsampling, to the grid I^l , the input training map available at the finest resolution ($l = 0, 1, \dots, L - 1$). Algorithmic details can be found in [10]. Here, we only emphasize that, thanks to the non-parametric formulation of neural networks and decision tree ensembles, and to the multiscale architecture of FCNs and hierarchical PGMs, the proposed technique is able to fuse multisensor multiresolution data.

III. EXPERIMENTAL VALIDATION

The experimental validation was carried out with a dataset of multimission SAR images acquired by COSMO-SkyMed (CSK), SAOCOM (SA), and Sentinel-1 (S1) in 2021, over Lombardy, Northern Italy. The CSK and SA acquisitions are Stripmap GTC (geocoded terrain corrected) with pixel spacing of 2.5 m and around 9 m, respectively. The S1 images are TOPSAR [22], with pixel spacing of 10 m. The SA and S1 images are dual-pol, cross-polarized VH and VV, while the CSK acquisitions have polarization HH. The DUSAF¹ archive (“Land use and land cover of the Lombardy Region” map) which contains land-cover information for the year 2018, was used as a guideline to define a ground truth (GT).

In the resulting dataset, the CSK acquisitions were multilooked at a spatial resolution of 5 m to reduce the impact of speckle. The dualpol SA images were resampled at 10 m, both to respect the power-of-2 relation of the quadtree and to match the pixel spacing of the S1 data. The S1 and the

resampled SA images were stacked in a 4-channel tile. The three acquisitions were registered using the method in [14] and using the CSK image as master image. Input data at 5 m represent the finest resolution in these experiments, hence it is fed to the FCN in its first layer, whereas the images at 10 m enter the FCN in the second convolutional block. Analogously, 5 m and 10 m images are inserted in the leaf level and in the second level of the quadtree, respectively, together with the FCN feature maps on the same two pixel grids. The images were acquired by the three satellite sensors over summer 2021, thus not presenting any seasonal variations (e.g., volume of water bodies, vegetation) among one another. These images, properly split into spatially disjoint training and test sets, were used to assess the performances of the proposed architecture.

The GT consisted of five semantic land-cover classes: built-up areas, low vegetation, tall vegetation, clutter, and water. Clutter is not only a minority class – representing a negligible percentage of the pixels in the images – but is also of limited interest as a target land cover as it only includes mixed surfaces (e.g., beaches, quarries, dumps, degraded areas, detrital accumulations). Given the mismatch in time between the GT (of 2018) and the radar acquisitions (of 2021), and the possible changes in the extension of the water bodies during the years, the labels for the class “water” were manually corrected to match their actual extension in the SAR images.

The quantitative results of the proposed method on the test set are reported in Table I. Details of the related classification maps are shown in Fig. 2. Specifically, the classification maps generated by applying the proposed method to the complete three-mission dataset are obtained at the resolution of 5 m, the same of the X-band data and the finest in the input multiresolution dataset. However, for comparison purposes, we ran the proposed system also considering as inputs the images provided by each single mission, alone, and the images deriving from two combinations of two missions (namely, CSK-SA, and CSK-S1). The results related to the combination SA-S1 have not been included in Table I as they are much less accurate than the other ones.

These experiments suggest that the combination of X- (CSK), L- (SA), and C-band (S1) data allows for more effective discrimination of the classes, with more accurate results than when using single-mission data, in terms of overall accuracy (OA), producer accuracy (PA, or recall), and user accuracy (UA, or precision).

The quantitative accuracies for the minority classes “tall vegetation” and “water” are generally higher for the experiments including the CSK data than when only the SA or S1 images were used. This is justified by the finer resolution of the CSK acquisitions, allowing to capture more details of these classes characterized by fewer samples in the dataset. The results for the majority classes appear to be more stable across all the various experiments, confirming the previous interpretation: the coarser spatial resolution does not strongly affect the majority classes “built-up” and “low vegetation”. The results of the proposed approach, when applied to images collected by the sensors of two missions, provided that CSK is one of the two, have higher UA than the results obtained from single-sensor data, and are comparable in terms of OA and PA

¹<https://www.geoportale.regione.lombardia.it/>

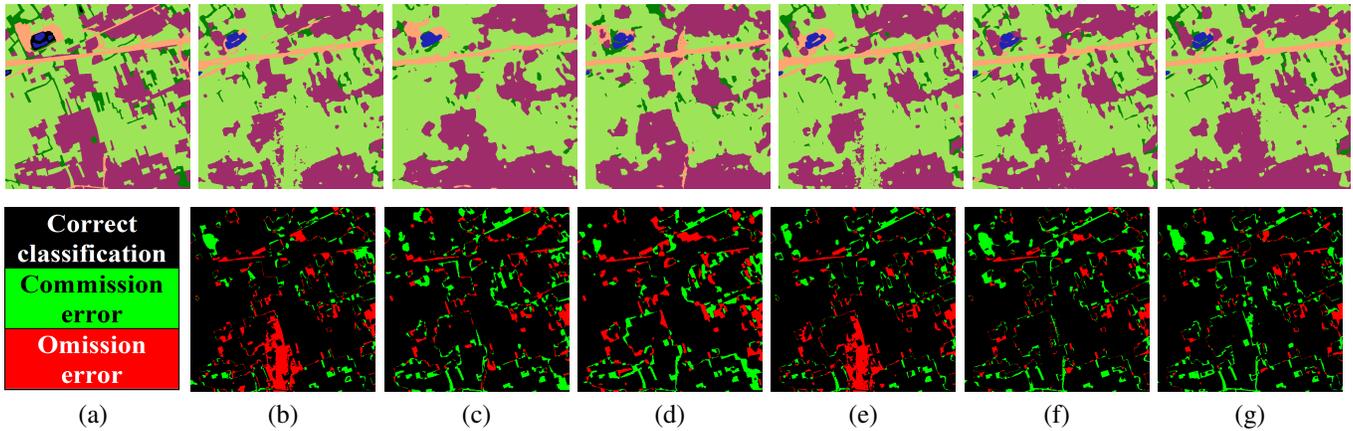


Fig. 2. Upper row: GT (a) and classification maps obtained by the proposed method when applied to (b) CSK, (c) SA, (d) S1, (e) CSK and SA, (f) CSK and S1, and (g) CSK, SA, and S1 data. Classes: built-up (pink), low vegetation (light green), tall vegetation (dark green), clutter (orange), and water (blue). (b) to (g): Product processed under a license of the Italian Space Agency (ASI) and the Argentinian Space Agency (CONAE); Original COSMO-SkyMed Product - ©ASI - (2021); Original SAOCOM Product - ©CONAE - (2021). Lower row: correctly classified pixels and commission and omission errors for the built-up class are displayed in different colors, for each classification map.

TABLE I

TEST-SET RESULTS. PER-CLASS VALUES ARE F1 SCORES. PA AND UA ARE AVERAGED OVER THE CLASSES. Z_1 , Z_2 STATISTICS OF THE MCNEMAR TEST TO VALIDATE WHETHER THE DIFFERENCES IN ACCURACY ARE STATISTICALLY SIGNIFICANT: Z_1 IS THE MCNEMAR'S STATISTICS OF THE PROPOSED METHOD AS COMPARED TO HRNET. Z_2 IS THE MCNEMAR'S STATISTICS OF THE PROPOSED METHOD, WHEN APPLIED TO ALL THREE SENSORS, AS COMPARED TO THE OTHER CASES (SINGLE SENSOR OR COUPLE OF SENSORS).

Data (architecture)	built-up	low veg.	tall veg.	clutter	water	OA	PA	UA	κ	Z_1	Z_2
X-band (proposed)	86.42	83.79	39.57	42.56	97.69	80.84	68.54	72.22	67.81	–	-84.87
X-band (HRNet)	85.48	81.57	32.94	40.02	98.18	79.06	65.56	72.28	64.44	-113.70	–
L-band (proposed)	81.57	77.81	20.02	42.85	89.53	74.90	58.55	70.38	56.98	–	-173.39
L-band (HRNet)	83.80	79.28	14.32	36.31	37.71	76.27	58.41	58.61	60.21	-5.72	–
C-band (proposed)	84.41	79.73	29.23	42.64	88.13	77.68	64.69	67.92	62.14	–	-128.25
C-band (HRNet)	82.18	76.37	21.34	42.93	92.66	75.56	62.01	70.98	57.90	-119.55	–
X-, L-bands (proposed)	85.93	81.96	36.76	49.71	99.11	79.96	68.03	76.09	66.09	–	-114.43
X-, L-bands (HRNet)	83.90	79.29	22.29	51.51	99.75	77.62	66.93	74.54	62.21	-51.31	–
X-, C-bands (proposed)	85.77	83.32	40.92	36.24	96.12	80.34	66.14	73.40	66.86	–	-77.34
X-, C-bands (HRNet)	83.50	77.11	31.46	38.17	95.81	75.63	64.13	69.95	59.42	-248.63	–
X-, L-, and C-bands (proposed)	86.93	85.09	40.76	48.85	99.63	82.15	70.56	76.67	69.82	–	–
X-, L-, and C-bands (HRNet)	85.24	81.94	37.48	47.66	95.45	79.69	66.41	74.74	65.59	-109.39	–

to the maps generated using single-sensor CSK imagery (see Table I). On the contrary, in the experiments involving all three radar bands, all the overall metrics are higher than in any other considered data configuration, thus confirming the advantages of modeling the complementary information contained in such multimodal dataset through the proposed approach.

The results obtained by the proposed technique were compared with those of HRNet [23], a neural architecture dealing with multiscale information through multiresolution subnetworks connected in parallel and trained with the same data configurations. As it is possible to note in Table I, the proposed method, leveraging both multiresolution and spatial contextual information, achieves higher accuracies than HRNet on both the overall metrics (OA, κ , and the average PA and UA) and on most of the class-wise accuracies. Considering only the experiments on single-sensor acquisitions, HRNet reaches higher UA in the case of single-sensor CSK and S1, thanks to the more accurate discrimination of the classes “clutter” and “water”, and a higher OA for single-sensor SA data. However, in the case of the fusion of all three input images, the proposed approach, as compared to HRNet, gains 2.46%, 4.15%, 1.93%, and 4.23% in OA, PA, UA, and κ , respectively. This suggests its effectiveness at taking benefit from the input

X-, L- and C-band imagery, and confirms the potential of the integration of deep learning and probabilistic graphical modeling for multimodal remote sensing image analysis.

The classification maps shown in Fig. 2 confirm the effectiveness of the proposed method at modeling spatial and multiresolution information to perform land-cover mapping from input multimission SAR data. In particular, the maps generated by the proposed method visually well differentiate the land covers in the considered scene. Concerning the vegetated zones, properly discriminated into “low vegetation” and “tall vegetation”, a majority and a minority class, respectively, the classification of the latter is improved in the experiments containing the CSK image. This may be interpreted in relation to the finer spatial resolution of this image in the considered dataset. The visualization of the omission and commission errors, shown in Fig. 2 for the class “built-up”, indicates the improvements obtained by the multiresolution and multifrequency combination of multiple sensor acquisitions through the proposed approach.

This experimental validation, conducted with images from the CSK, SA, and S1 satellite missions, points out the effectiveness of the integration of FCNs and PGMs for the semantic segmentation not only of aerial optical imagery (as

considered in [10]) but also of multimission radar images. The performances shown in Table I suggests that, in terms of OA, PA, UA, and κ , on the considered dataset, the proposed method achieves more accurate results than HRNet in almost all of the configurations, from single-mission to multimission input data. The statistical significance of the difference between the aforementioned results was confirmed by the outcome of McNemar's test [24] (see Table I) between the proposed method and HRNet (Z_1) and between the proposed method in the 3-sensor case and the proposed method itself in the single-sensor or 2-sensor cases (Z_2). We recall that a negative value of Z indicates that the proposed method is more accurate than the one used for comparison and $|Z| > 1.96$ indicates that the difference is statistically significant with respect to the common 5% level of significance [24].

IV. CONCLUSION

In this paper, a methodology for land-cover mapping from input multifrequency, multiresolution, and multimission SAR data has been presented. For this purpose, the proposed methodology integrates FCNs, hierarchical PGMs, and decision tree ensembles. The objective is to exploit the spatial and multiresolution modeling capabilities of hierarchical Markov models and deep convolutional networks to perform multimodal satellite image fusion and produce classification maps at the different spatial resolutions of the input dataset.

The proposed method generalizes a previous technique developed for aerial optical imagery of urban areas to the case of the joint classification of multimission SAR images at distinct spatial resolutions. The experimental validation, conducted with X/L/C-band images from COSMO-SkyMed, SAOCOM, and Sentinel-1, suggests the potential of the proposed approach in this challenging case of semantic segmentation of multimission radar images. The reported results show how the integration of multisensor information, characterized by different frequencies, spatial resolutions and bands, allows obtaining accurate ground mapping results. In particular, the classification accuracy improves when the three considered SAR bands are jointly used, as compared to the results achieved using data from only one or two missions. The proposed approach achieves more accurate fusion performances than HRNet, a recent state-of-the-art multiresolution deep learning technique, in terms of OA, PA, UA, and κ . The classification maps generated by the developed method appear visually smooth and their comparison with the maps obtained using input data from only one or two sensors confirm the multimission fusion capabilities of the technique.

Future developments may involve the extension to longer temporal series, characterized by non-negligible land-cover changes. This would require the integration of the proposed method with change detection and multitemporal modeling, for example, through the combination with recurrent neural networks [25] and/or latent Markov models [17].

ACKNOWLEDGMENT

Project carried out using COSMO-SkyMed Products, © of the Italian Space Agency (ASI), and SAOCOM Products, © of the Argentinian Space Agency (CONAE) delivered under a license to use by ASI and CONAE.

REFERENCES

- [1] J. A. Richards, *Remote sensing digital image analysis*. Springer, 2022.
- [2] C. Oliver and S. Quegan, *Understanding synthetic aperture radar images*. SciTech Publishing, 2004.
- [3] L. Gómez-Chova, D. Tuia, G. Moser, and G. Camps-Valls, "Multimodal classification of remote sensing images: a review and future directions," *Proc. of the IEEE*, vol. 103, no. 9, pp. 1560–1584, 2015.
- [4] V. H. R. Prudente, S. Skakun, L. V. Oldoni, H. A. M. Xaud, M. R. Xaud, M. Adami, and I. D. Sanches, "Multisensor approach to land use and land cover mapping in Brazilian Amazon," *ISPRS J. Photogramm. Remote Sens.*, vol. 189, pp. 95–109, 2022.
- [5] G. Joshi, R. Natsuaki, and A. Hirose, "Neural network fusion processing and inverse mapping to combine multisensor satellite data and analyze the prominent features," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 16, pp. 2819–2840, 2023.
- [6] Y. J. E. Gbodjo, O. Montet, D. Ienco, R. Gaetano, and S. Dupuy, "Multisensor land cover classification with sparsely annotated data based on convolutional neural networks and self-distillation," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 14, pp. 11 485–11 499, 2021.
- [7] J. Kun, W. Bingfang, L. Qiangzi, and T. Yichen, "Improvement of classification accuracy integrating C- and X-band synthetic aperture radar data," in *IEEE MAPE*, 2009, pp. 340–345.
- [8] H. Zhang and C. Ling, "Multichannel simultaneous dual-band fully polarimetric airborne synthetic aperture radar: system features and experimental results," *J. Appl. Remote Sens.*, vol. 12, no. 3, 2018.
- [9] J. Yin and J. Yang, "A modified level set approach for segmentation of multiband polarimetric SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 11, pp. 7222–7232, 2014.
- [10] M. Pastorino, G. Moser, S. B. Serpico, and J. Zerubia, "Semantic segmentation of remote-sensing images through fully convolutional neural networks and hierarchical probabilistic graphical models," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, no. 5407116, pp. 1–16, 2022.
- [11] H. Sportouche, J.-M. Nicolas, and F. Tupin, "Mimic capacity of fisher and generalized Gamma distributions for high-resolution SAR image statistical modeling," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 10, no. 12, pp. 5695–5711, 2017.
- [12] G. Moser, J. Zerubia, and S. Serpico, "SAR amplitude probability density function estimation based on a generalized gaussian model," *IEEE Trans. Image Process.*, vol. 15, no. 6, pp. 1429–1442, 2006.
- [13] E. W. Stacy, "A generalization of the Gamma distribution," *Ann. Math. Stat.*, vol. 33, no. 3, pp. 1187–1192, 1962.
- [14] D. Solarna, L. Maggiolo, G. Moser, and S. B. Serpico, "A tiling-based strategy for large-scale multisensor optical-SAR image registration," in *IEEE IGARSS*, 2022, pp. 127–130.
- [15] J. Le Moigne, N. S. Netanyahu, and R. D. Eastman, *Image registration for remote sensing*, 1st ed. Cambridge University Press, 2011.
- [16] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.
- [17] Z. Kato and J. Zerubia, "Markov random fields in image segmentation," *Found. Trends Signal Process.*, vol. 5, no. 1-2, pp. 1–155, 2012.
- [18] M. Pastorino, A. Montaldo, L. Fronda, I. Hedhli, G. Moser, S. B. Serpico, and J. Zerubia, "Multisensor and multiresolution remote sensing image classification through a causal hierarchical markov framework and decision tree ensembles," *Remote Sens.*, vol. 13, no. 5, 2021.
- [19] M. Pastorino, F. Gallo, A. Di Febraro, G. Moser, N. Sacco, and S. B. Serpico, "Multimodal fusion of mobility demand data and remote sensing imagery for urban land-use and land-cover mapping," *Remote Sens.*, vol. 14, no. 14, 2022.
- [20] J. Laferté, P. Pérez, and F. Heitz, "Discrete Markov image modeling and inference on the quadtree," *IEEE Trans. Image Process.*, vol. 9, no. 3, pp. 390–404, 2000.
- [21] N. Altman and M. Krzywinski, "Ensemble methods: bagging and random forests," *Nat. Methods*, vol. 14, pp. 933–934, 2017.
- [22] F. De Zan and A. Monti Guarnieri, "TOPSAR: Terrain observation by progressive scans," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 9, pp. 2352–2360, 2006.
- [23] K. Sun, B. Xiao, D. Liu, and J. Wang, "Deep high-resolution representation learning for human pose estimation," in *IEEE/CVF CVPR*, 2019, pp. 5686–5696.
- [24] Q. McNemar, "Note on the sampling error of the difference between correlated proportions or percentages," *Psychometrika*, vol. 12, no. 2, pp. 153–157, 1947.
- [25] M. Alameh, Y. Abbass, A. Ibrahim, G. Moser, and M. Valle, "Touch modality classification using recurrent neural networks," *IEEE Sens. J.*, vol. 21, no. 8, pp. 9983–9993, 2021.