HAL open science

Classification of multimission SAR images based on Probabilistic Graphical Models and Convolutional Neural Networks

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

▶ **To cite this version:**

Martina Pastorino, Gabriele Moser, Sebastiano B Serpico, Josiane Zerubia. Classification of multimission SAR images based on Probabilistic Graphical Models and Convolutional Neural Networks. IEEE IGARSS 2023- International Geoscience and Remote Sensing Symposium, Jul 2023, Pasadena (CA), United States. pp.1420-1423, 10.1109/IGARSS52108.2023.10283356 . hal-04109475

HAL Id: hal-04109475

https://inria.hal.science/hal-04109475v1

Submitted on 30 May 2023

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

CLASSIFICATION OF MULTIMISSION SAR IMAGES BASED ON PROBABILISTIC GRAPHICAL MODELS AND CONVOLUTIONAL NEURAL NETWORKS

Martina Pastorino[1,2], *Gabriele Moser*[1], *Sebastiano B. Serpico*[1], *and Josiane Zerubia*[2]

[1] University of Genoa, DITEN dept., Genoa, Italy, martina.pastorino@edu.unige.it.
[2] Inria, Université Côte d'Azur, Sophia-Antipolis, France.

ABSTRACT

The problem of the semantic segmentation of multimodal images is characterized by the challenge of jointly exploiting information deriving from images possibly acquired at different spatial resolutions, frequencies, and bands. 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 spatial and multiresolution information contained in multimodal remote sensing images collected by distinct space missions with SAR payloads. The experimental validation is conducted with COSMO-SkyMed and SAOCOM images over Northern Italy. The results show that the proposed methodology is capable to reach accurate classification maps from input multimission SAR imagery.

***Index Terms*—** multimodal SAR imagery, multisensor, semantic segmentation, CNN, FCN, PGM, MRF

1. INTRODUCTION

Given the current advances in space missions for Earth observation, it is possible to have access to multimodal satellite imagery, characterized by different synthetic aperture radar (SAR) modalities, and therefore frequencies, bands, polarizations, and spatial resolutions [1]. The combination of data obtained from different sensors allows trade-offs between different acquisitions, possibly containing complementary information. Methods for the analysis of multimodal data are getting more and more popular to fully exploit all the available information in the field of remote sensing applications [2].

In the framework of semantic segmentation, deep learning (DL) models are currently major state-of-the-art techniques. Several models have been developed for remote sensing image classification [3]. Amongst them, fully convolutional networks (FCNs) have proven capable to obtain accurate classification results over complex remote sensing scenes [4, 5]. Various DL-based methods have been proposed to deal with multisource data [2, 6, 7, 8]. However, as a drawback, in order to reach high performances, DL methods generally require large training datasets, difficult to retrieve [9] and rarely available in remote sensing applications. Sometimes they may also poorly model spatial consistency [10].

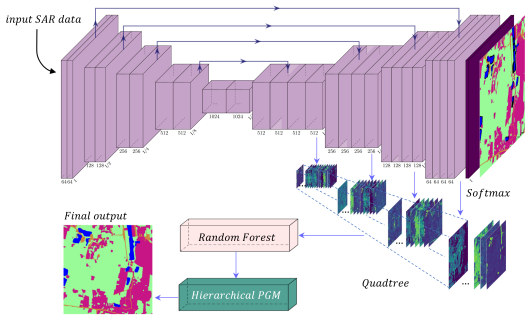


Fig. 1. Overall architecture of the method.

At the same time, stochastic models such as probabilistic graphical models (PGMs) are powerful tools for image processing tasks, as they can be employed for structured prediction problems within Bayesian inference schemes. Particularly, for 2D image analysis, random fields, such as Markov random fields (MRFs) [11] are capable to model spatial and multiresolution information [11, 12].

The approach in [13], in particular, adopts a quadtree topology by defining, in each of the associated lattices, a Markov chain-based model for the spatial dependencies between pixels. The result is a set of classification maps jointly generated at all spatial resolutions of the input multimission dataset. A hierarchical PGM is defined over the quadtree, in order to exploit the intrinsic multiscale behavior of FCNs [14], allowing to incorporate multiresolution information through its hidden layers.

This paper extends the methodology previously developed in [13] for the classification of optical aerial imagery to the classification of data obtained from different SAR missions on the same date, exploiting the statistical, multifrequency and multiresolution information provided by multiple sensors and their intrinsically complementary nature.

University of Genoa and Université Côte d'Azur (UCA) are part of the Ulysseus Alliance (European University). https://ulysseus.eu/

2. METHODOLOGY

The model (shown in Fig. 1) comprises three main methodological components: an FCN, a hierarchical PGM, and a decision tree ensemble (e.g., random forest) [15]. The latter aims to link the feature representation of the FCN and the Bayesian inference of the PGM through a suitable set of pixelwise posterior probabilities.

Starting with the DL component, any FCN can be employed in the proposed framework: given its encoder-decoder architecture, an FCN is capable to yield classification outputs with the same size of the input, making it a convenient tool for semantic segmentation tasks.

The multimodal information is further modeled through the hierarchical PGM which consists of a combination of a hierarchical and a planar MRF to access both multiscale and spatial-contextual information between neighboring pixels. Hierarchical MRFs are known to be causal [11], while this feature is not guaranteed for planar MRFs. Hence, a neighborhood relation $\precsim$ is introduced in the pixel grid S to ensure the causality of the whole PGM: $r \precsim s$ indicates that pixel r is a causal neighbor of pixel s ($r, s \in S$, $r \prec s$). Given $\mathcal{X} = \{x_s\}_{s \in S}$, the random field of the discrete class labels x_s of all pixels $s \in S$, and a set of $l = 1, \ldots, L$ pixel grids at different resolutions, the hierarchical PGM is defined as:

$$P(\mathcal{X}^l|\mathcal{X}^{l-1}) \propto \prod_{s \in \mathcal{X}^l} P(x_s|x_r, r \precsim s) P(x_s|x_s^-) \qquad (1)$$

where the first part of the equation derives from the hierarchical Markovian property, $P(\mathcal{X}^l|\mathcal{X}^{l-1}, \ldots, \mathcal{X}^0) = P(\mathcal{X}^l|\mathcal{X}^{l-1})$, and x_{s^-} represents the class label of the parent site $s^- \in S^{l-1}$ of pixel $s \in S^l (l = 1, \ldots, L; L$ indicating the grid at the finest spatial resolution). The hierarchical Markov model is defined over a quadtree. More details on the methodology can be found in [13, 16].

This approach was originally developed for the semantic segmentation from aerial optical imagery of urban areas. Here, it is extended to the challenging case of semantic segmentation from input very high resolution (VHR) satellite SAR multimission imagery. Ideally, a collection of SAR images taken on the same date is used. Operatively, a time series of SAR images, collected by distinct missions on different dates with rather small mutual time distances, are used, under the assumption that no relevant changes have occurred in the land cover of the monitored scene during the overall acquisition timeframe. In the proposed approach, the resulting multimission SAR data are incorporated into the aforementioned quadtree topology, whose levels represent different resolutions and contain, therefore, both the feature maps of the hidden layers of the decoder of the FCN and the original bands of the multimodal input SAR images, inserted respecting the original relationship between resolutions. Furthermore, the network included in the proposed method is trained with a multimission dataset of SAR images, set in input to the encoder respecting the relationship between their native resolution to take advantage of the multiscale topology of FCNs.

In view of the causality of the overall hierarchical PGM, an efficient non-iterative inference algorithm can be used for the final inference. The marginal posterior mode (MPM) [12, 17] criterion is employed as it is especially advantageous for applications involving multiscale information, because it penalizes errors according to their scale and it can be defined on quadtrees with three efficient recursive steps [16]. The proposed method is summarized in Algorithm 1.

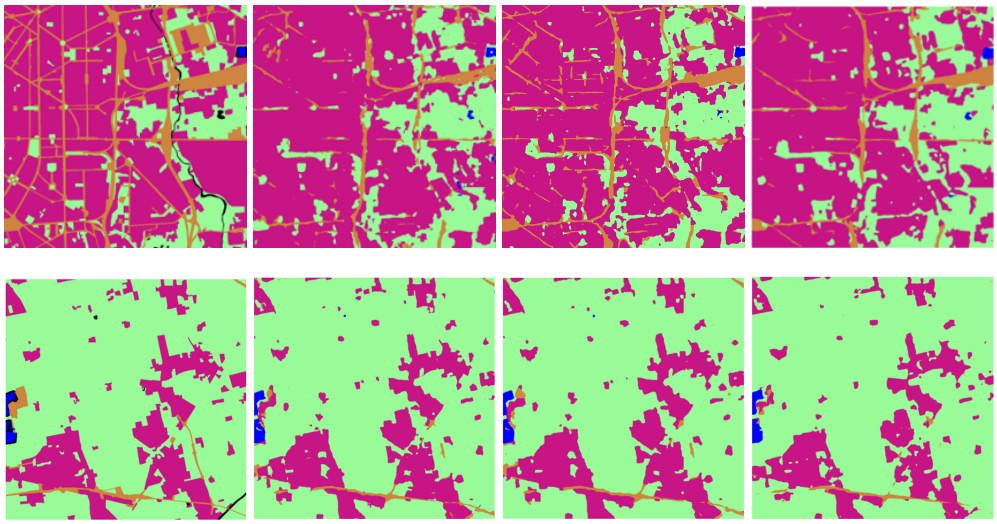
Algorithm 1 FCN, MPM on the hierarchical PGM, and RF

1: **Training of the FCN** with the input multimission SAR dataset at different convolutional blocks
2: **Creation of the $L-$levels quadtree** containing, in the random field $\mathcal{Y} = \{y_s\}_s \in S$ of the observations, the network feature maps and the original channels of the image to classify the corresponding resolution
3: **Beginning of the MPM:**
4: **First top-down pass:** estimation of the priors $P(x_s)$
5: **RF classifier** estimation of the posteriors $P(x_s|y_s)$ through RF
6: **Bottom-up pass:** estimation of $P(x_s|y_s^d)$ and $P(x_s|x_s^c, y_s^d)$, where y_s^d collects the observations of all descendants of s in the tree (including s), x_s^c collects the labels of all sites connected to s (x_{s^-} and $\{x_r\}_{r \precsim s}$)
7: **Second top-down pass:** estimation of $P(x_s|\mathcal{Y})$ at each level of the quadtree
8: **Output:** maximization of $P(x_s|\mathcal{Y})$

3. EXPERIMENTAL VALIDATION

The method was applied to a dataset of multimission SAR images acquired by COSMO-SkyMed and SAOCOM in 2021 over Lombardy, Northern Italy. It consists of Stripmap GTC (Geocoded terrain corrected) images with pixel spacing of 2.5 m and around 9 m, respectively. The polarization of the SAOCOM images is VH and VV, while COSMO-SkyMed acquisitions are HH-polarized. The DUSAF[1] ("Land use and land cover of the Lombardy Region") data archive, containing land cover information for the year 2018 was used as guideline to define a ground truth (GT). Starting from this archive, four semantic macro-classes of interest were selected for the experiments: urban areas, vegetation, clutter, and water. "Clutter" is relatively meaningless as a land cover class, since it comprises all the surfaces that do not explicitly belong to the other semantic classes (in this dataset: beaches, quarries, dumps, degraded areas, detrital accumulations), hence it is highly mixed. Moreover, it represents a negligible percentage of pixels in the images.

[1] https://www.regione.lombardia.it/wps/portal/istituzionale/HP/DettaglioServizio/servizi-e-informazioni/Enti-e-Operatori/Territorio/sistema-informativo-territoriale-sit/uso-suolo-dusaf/uso-suolo-dusaf



(a) (b) (c) (d)

Fig. 2. **GT and classification maps:** (a) GT and classification maps obtained with (b) FCN (U-Net [18]), (c) the proposed method, and (d) HRNet [19]. Classes: urban (pink), vegetation (pale green), clutter (orange), and water (blue). (b) to (d): 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).

Table 1. Test-set results. Per-class scores are recalls. OA, PA, and UA stand for overall accuracy, producer accuracy, and user accuracy, respectively.

Architecture	urban	vegetation	clutter	water	OA	PA	UA	F1 score
U-Net [18]	**0.89**	0.78	0.29	0.92	0.77	0.72	0.70	0.71
HRNet [19]	0.85	**0.83**	0.34	**1.0**	0.75	0.75	0.77	0.76
Proposed ($\vartheta = 0.2, \psi = 0.8$)	**0.89**	0.78	0.44	0.97	**0.79**	0.77	**0.79**	**0.78**
Proposed ($\vartheta = 0.5, \psi = 0.8$)	0.87	0.80	0.44	0.93	0.78	0.76	**0.79**	0.77
Proposed ($\vartheta = 0.8, \psi = 0.8$)	0.83	**0.83**	**0.48**	0.96	0.77	**0.78**	0.77	0.77

The resulting dataset consists of tiles of size 1024×1024 pixels, with the COSMO-SkyMed images multilooked at a resolution of 5 m to reduce the impact of the speckle, and the dualpol SAOCOM images resampled at 10 m to respect the power-of-two relation typical of a quadtree. These images were acquired by the two satellites over the summer 2021, thus not presenting any land cover discrepancy due to seasonal changes (e.g., volume of water bodies, seasonal vegetation). These images, properly split into the training and test sets, were used to assess the performances of the proposed architecture. Given the mismatch in time between the GT and the radar acquisitions, 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 on the SAR images.

The proposed technique depends on three main hyperparameters: L, the number of levels in the quadtree (set to $L = 4$ for the experiments), and two real-valued parameters ϑ and ψ, representing the transition probabilities in the hierarchical PGM; ϑ and ψ were optimized empirically by trial-and-error.

The proposed method for the classification of multimission images was compared to HRNet [19], another technique dealing with multiscale information through multiresolution subnetworks connected in parallel.

The quantitative results are reported in Table 1 and suggest that incorporating multiresolution and spatial contextual information through the proposed approach guarantees a better discrimination of the land cover classes, with gains in terms of producer accuracy (PA), or recall, user accuracy (UA), or precision, and overall accuracy (OA). From the presented results, it is possible to notice that lower values of the transition probability across the scales, ϑ, correspond to higher OA and UA, but slightly lower PA, hence suggesting that, in this case, information at coarser resolution does not contribute significantly to the final prediction. Experiments with lower values of ψ, not reported for brevity, were characterized by lower values in terms of both PA and UA.

The classification maps obtained by the proposed method and the techniques used for comparison are shown in Fig. 2 and suggest the effectiveness of the proposed method in com-

bining SAR images coming from different sensors, at different frequencies, spatial resolutions, and polarizations, to perform land cover mapping from input multimission SAR data. In particular, the map generated by the proposed approach visually well discriminates the land covers in the considered scene. This experimental validation, conducted with images from the COSMO-SkyMed and SAOCOM satellite missions, points out the effectiveness of the integration of FCN and PGM methodological components for the semantic segmentation not only of aerial optical imagery but also of multimission radar images.

4. DISCUSSION AND CONCLUSION

This paper presents a method for the joint classification of multimission SAR images through FCNs, hierarchical PGMs and decision tree ensembles. The idea is to make use of the multiresolution modeling structure of the aforementioned techniques to fully exploit multimodal, possibly complementary, information.

The experimental validation, conducted with images from COSMO-SkyMed and SAOCOM, shows the potentials of the approach previously developed for both the semantic segmentation of aerial optical imagery and single sensor radar data in the case of multimission radar images.

The results suggest the effectiveness of the proposed method, which achieves higher performances than standard FCNs and other multiresolution techniques in terms of OA, precision, and recall. The classification maps, as well, appear to be smooth and prove the capabilities of the proposed method in discriminating the land cover classes considered.

5. 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. 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.

6. REFERENCES

[1] J. A. Richards, *Remote sensing digital image analysis: an introduction*, 5th ed. Springer, 2013.

[2] 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.

[3] F. Lateef and Y. Ruichek, "Survey on semantic segmentation using deep learning techniques," *Neurocomputing*, vol. 338, pp. 321–348, 2019.

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

[5] S. Šćepanović, O. Antropov, P. Laurila, Y. Rauste, V. Ignatenko, and J. Praks, "Wide-area land cover mapping with sentinel-1 imagery using deep learning semantic segmentation models," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 14, pp. 10 357–10 374, 2021.

[6] Q. Weng, H. Chen, H. Chen, W. Guo, and Z. Mao, "A multisensor data fusion model for semantic segmentation in aerial images," *IEEE Geoscience and Remote Sensing Letters*, vol. 19, pp. 1–5, 2022.

[7] A. Farooq, X. Jia, J. Hu, and J. Zhou, "Transferable convolutional neural network for weed mapping with multisensor imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–16, 2022.

[8] X. Zhao, R. Tao, W. Li, W. Philips, and W. Liao, "Fractional Gabor convolutional network for multisource remote sensing data classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–18, 2022.

[9] W. Wu, H. Li, X. Li, H. Guo, and L. Zhang, "PolSAR image semantic segmentation based on deep transfer learning—realizing smooth classification with small training sets," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 6, pp. 977–981, 2019.

[10] F. Ma, F. Gao, J. Sun, H. Zhou, and A. Hussain, "Weakly supervised segmentation of SAR imagery using superpixel and hierarchically adversarial CRF," *Remote Sens.*, vol. 11, no. 5, 2019.

[11] S. Li, *Markov random field modeling in image analysis*, 3rd ed. Springer, 2009.

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

[13] 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 Transactions on Geoscience and Remote Sensing*, vol. 60, no. 5407116, pp. 1–16, 2022.

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

[15] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.

[16] 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.

[17] 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.

[18] 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.

[19] K. Sun, B. Xiao, D. Liu, and J. Wang, "Deep high-resolution representation learning for human pose estimation," in *2019 IEEE CVPR*, pp. 5686–5696, 2019.