Università di Genova

Inria

Semantic Segmentation of Remote Sensing Images Combining Hierarchical Probabilistic Graphical Models and Deep 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
[2] Inria, UCA, Sophia-Antipolis Méditerranée Center, France

November 26, 2020

Outline

Introduction

Introduction

Semantic segmentation of remote sensing images with deep learning

Deep learning techniques achieve state-of-the-art results in semantic segmentation tasks

- Very high per-pixel accuracies
- Efficient reproduction of the shapes of the objects segmented
- Among the most successful architectures are the FCNs (*fully convolutional networks*)
- However, to attain high performances they need big datasets of spatially exhaustive ground truths
  - Only available in benchmark datasets, not in real applications
  - Require the involvement of expensive human experts for labeling
- Often computationally demanding

Introduction

Potential of probabilistic graphical models in semantic segmentation

Probabilistic graphical models (PGMs) have the ability to produce structured predictions

- Exploitation of contextual (spatial) information
- Markov models postulated on planar or multilayer graphs (*quadtrees*) are known as flexible and powerful stochastic models for spatial information
- For MRF, Markovianity is formulated with respect to a neighborhood of each node of the related graph
  - Hierarchical MRFs captures multiresolution relations (multiscale spatial information) but does not model the spatial context within each pixel grid
  - Markov mesh random fields (MMRFs) describes spatial interactions among the pixels (single resolution)

Objectives

Development of a novel semantic segmentation method for VHR remote sensing images combining the advantages of deep learning techniques and PGMs

- Exploit the information contained at different image scales in the network activations
- Integrate deep learning solutions with probabilistic graphical models
- To achieve accurate performances with lower requirements in terms of quality and quantity of ground truth maps

Proposed method

Overview of the proposed method

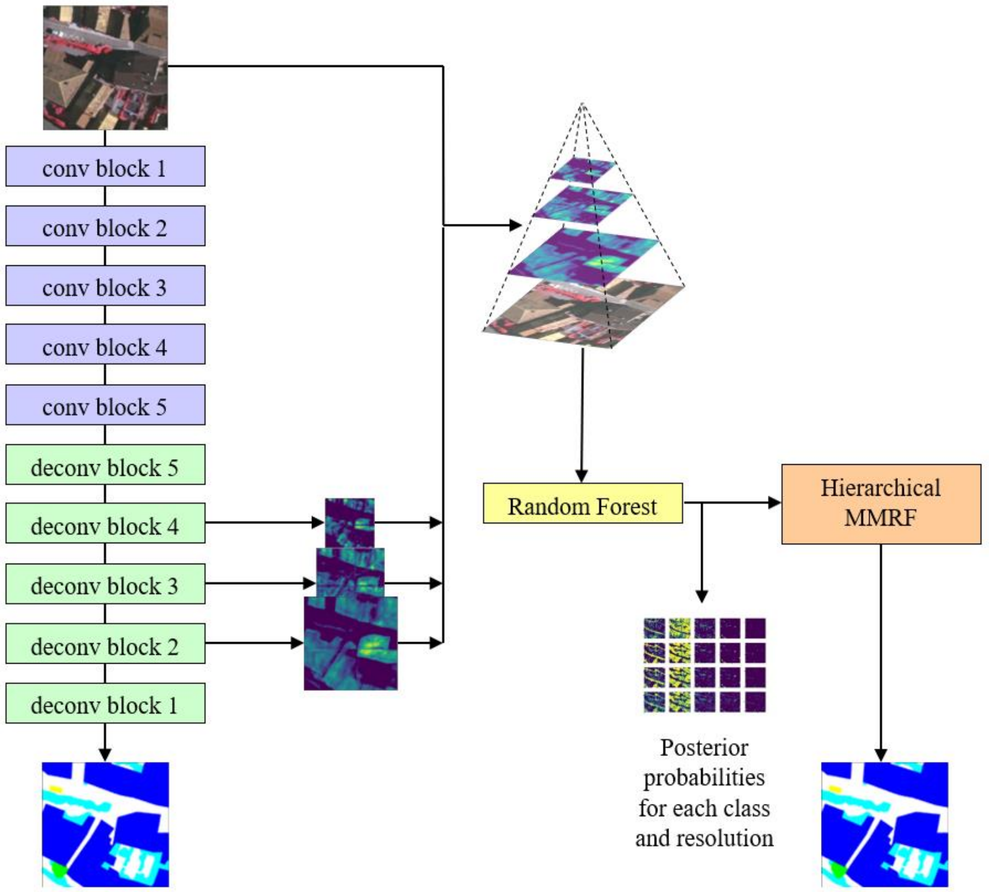
The proposed method for semantic segmentation involves the use of

- FCNs (U-Net or SegNet)
- Hierarchical causal Markov model
- Random forest (RF) ensemble

Objective: exploit the multiscale behavior of FCNs

The FCN is trained with a dataset of VHR images

- its activations at L different blocks (i.e., different spatial resolutions) are inserted in a *quadtree (level 1 to L-1)* with the channels of the original image in level 0



Overall architecture

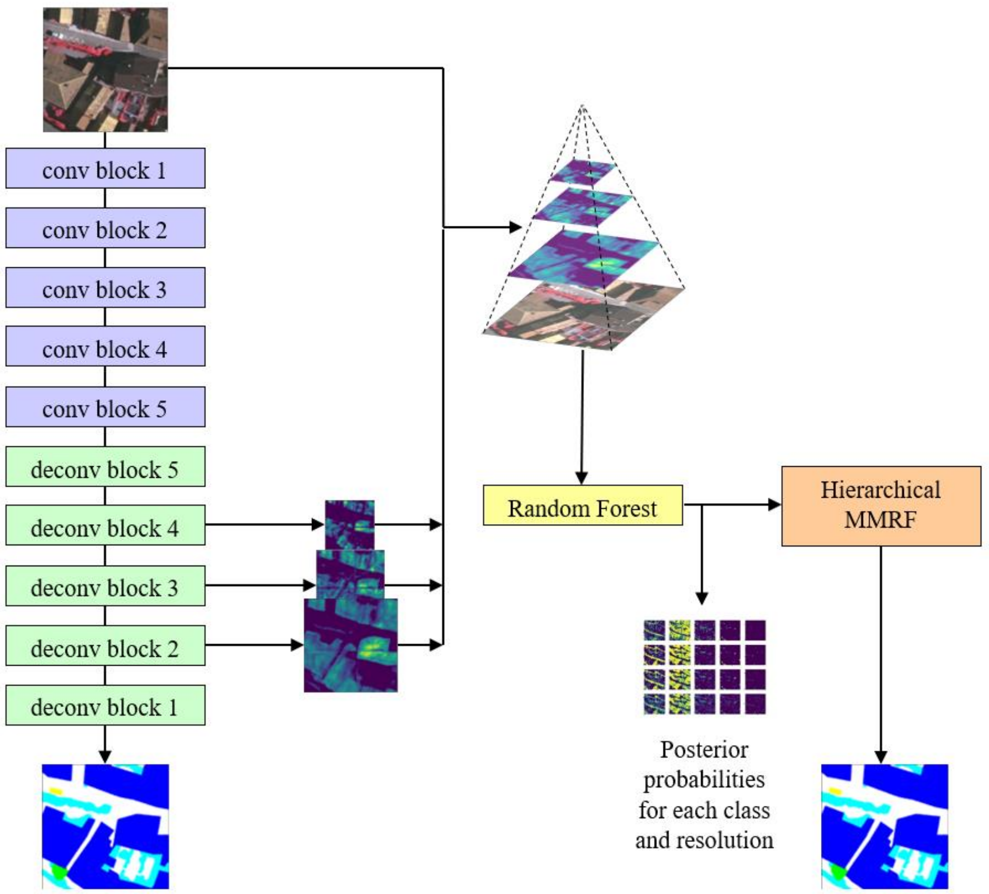
Overview of the proposed method

Feature representation extracted through the network activations is fed to the PGM

- Hierarchical causal Markov random field on the quadtree and spatial Markov chain (jointly)
- A pixel scan that combines both a zig-zag trajectory and a Hilbert space-filling curve is used to account for the dependencies within pixels, both inter-scale and intra-scale

Sample-wise posteriors are necessary to incorporate network activations into the PGM

- The quadtree is used to train the RF classifier
- To obtain the pixelwise posterior probabilities used for the inference equations of the model



Overall architecture

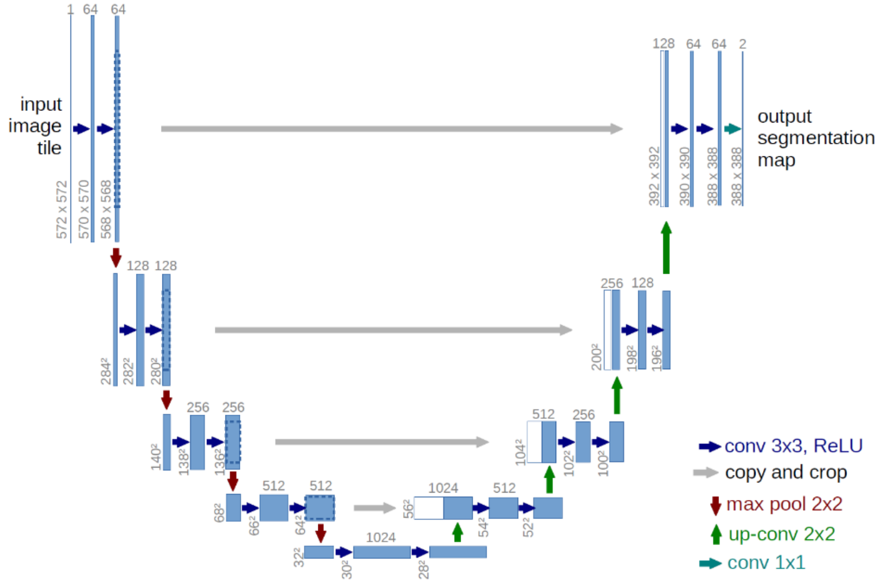
Deep learning architecture

Two different FCNs were adopted

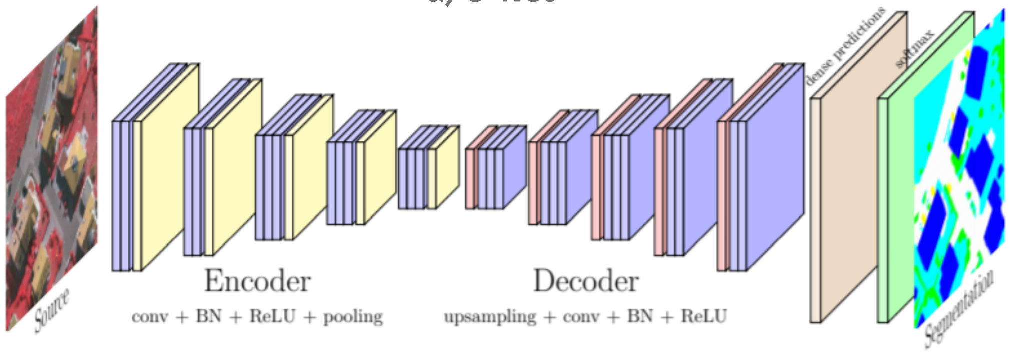
- U-Net and SegNet

These networks do not contain any dense layer

- Encoder-decoder architecture
  - The encoder performs the downsampling
  - The decoder addresses the upsampling and the classification
- Semantic segmentation that can yield outputs with the same size of the input
- 3 skip connections collect the activations of the network at three different resolutions
  - 128×128, 64×64, 32×32 pixels



a) U-Net

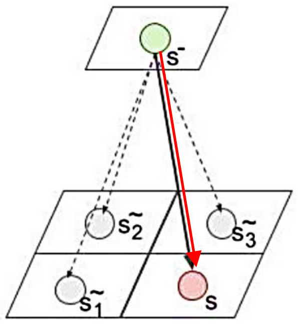


b) SegNet

Probabilistic graphical model

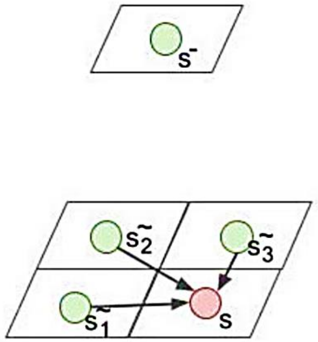
Hierarchical MRF on quadtrees

- Causal
- Efficient non-iterative inference
- Does not model spatial information within each scale



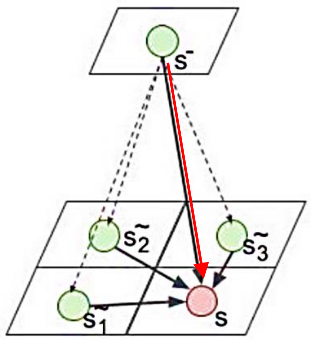
Planar MRF

- Models spatial information
- Generally non-causal



In the proposed method

- Markovianity between scales and within each layer
- Multiresolution fusion through *quadtree* topology



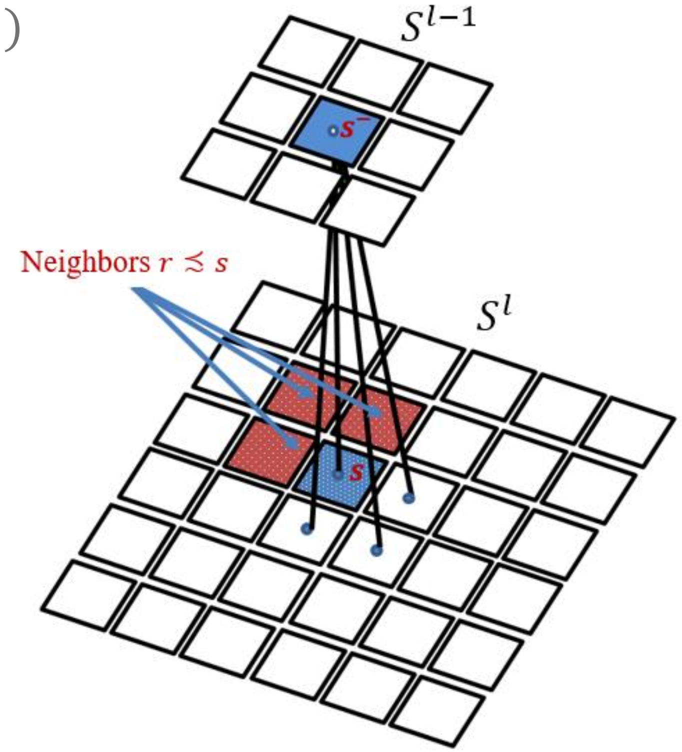
Hierarchical Markov PGM: properties

Markovianity of labels across scales and in each layer

$$P(\mathcal{X}^l|\mathcal{X}^{l-1},\mathcal{X}^{l-2},\dots,\mathcal{X}^0) = P(\mathcal{X}^l|\mathcal{X}^{l-1}) \propto \prod_{s\in S^l} P(x_s|x_r, r \precsim s)P(x_s|x_{s^-})$$

$$P(\mathcal{X}^0) = \prod_{s\in S^0} P(x_s|x_r, r \precsim s)$$

- A neighborhood relation is assumed in the pixel grid: $r \precsim s$ indicates that r is a causal neighbor of s
- The relation $\precsim$ is defined by a 1D scan of each layer of the quadtree → Markov chain (combination of zig-zag and Hilbert curve scans)
- $\mathcal{X}^0$ is a causal MRF on the root lattice S^0
- Conditional independence of feature vectors given the labels



Hierarchical Markov PGM: MPM formulation

The whole hierarchical PGM is causal → Marginal posterior mode (MPM) for inference.

Under some assumptions

$$P(x_s) = \sum_{x_{s^-}} P(x_s|x_{s^-})P(x_{s^-})$$

$$P\left(x_s|y_s^d\right) \propto P(x_s|y_s) \prod_{t \in s^+} \sum_{x_t} \frac{P(x_t|y_t^d)P(x_t|x_s)}{P(x_t)}$$

$$P\left(x_s \middle| x_s^c, y_s^d\right) \propto \frac{P\left(x_s|y_s^d\right)P(x_s|x_{s^-})P(x_{s^-})}{P(x_s)^{n_s}} \prod_{r \precsim s} P(x_s|x_r)P(x_r)$$

$$P(x_s|\mathcal{Y}) = \sum_{x_s^c} P(x_s|x_s^c, y_s^d)P(x_{s^-}|\mathcal{Y}) \prod_{r \precsim s} P(x_r|\mathcal{Y})$$

UniGe Inria

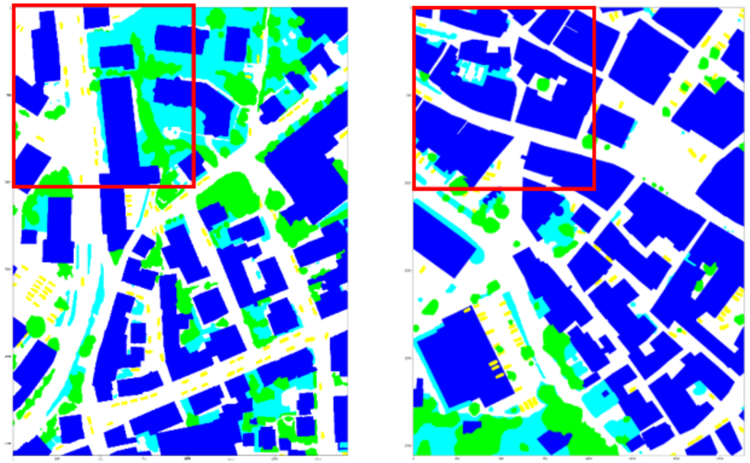
Experimental Results

Dataset

ISPRS 2D Semantic Labelling Challenge Vaihingen Dataset

- VHR aerial images with a resolution of 9 cm/pixel
- Ideal dataset, with dense, spatially exhaustive, pixel-level ground truths
- Six classes: buildings, impervious surfaces (e.g., roads), low vegetation, trees, cars, and clutter
- Red, green, and near-infrared channels and digital surface model (DSM)
- 33 image tiles of approximately 2100 × 2100 pixels
- 16 ground truth images: 12 used for training and 4 for testing

a) True orthophoto



b) Ground truth

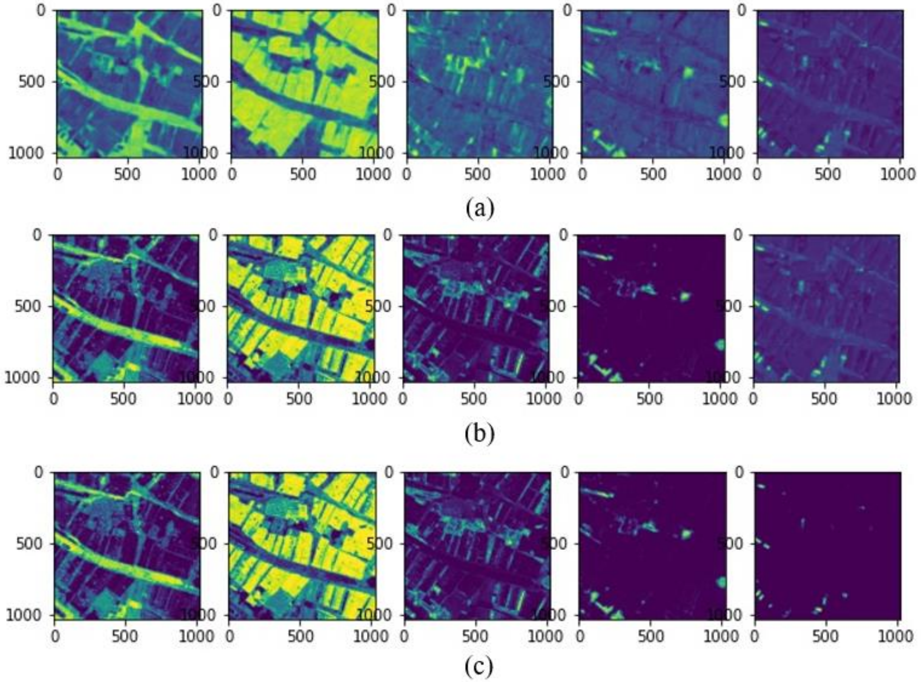
Experimental setup

Several training conditions were considered

- Full ground truth
- “Deteriorated” ground truth with a percentage of unlabeled pixels (either randomly or in blocks)
- Ground truth modified by morphological operators
  - Erosion and dilation
- These degradations are aimed at approaching real-world cases of limited and non-exhaustive ground truths

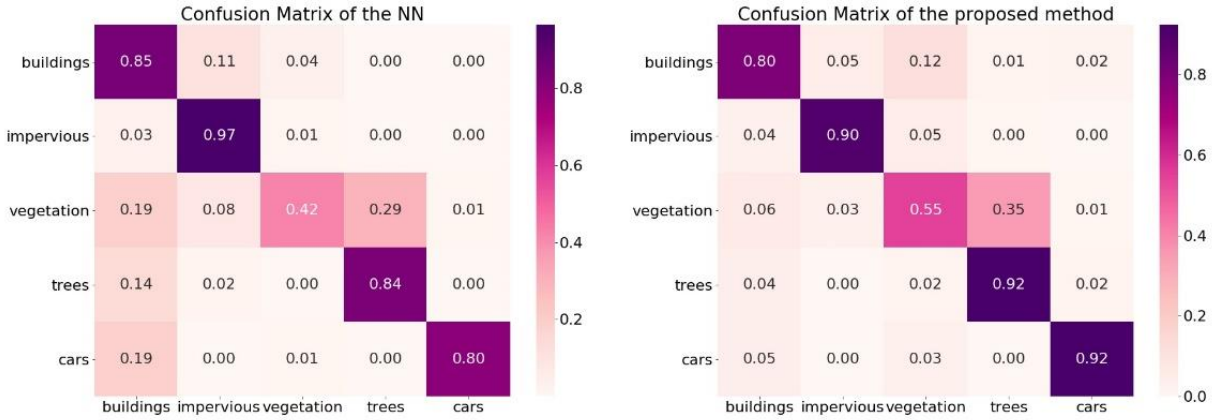
And several formulations

- Focusing on the posterior probabilities of the base of the quadtree

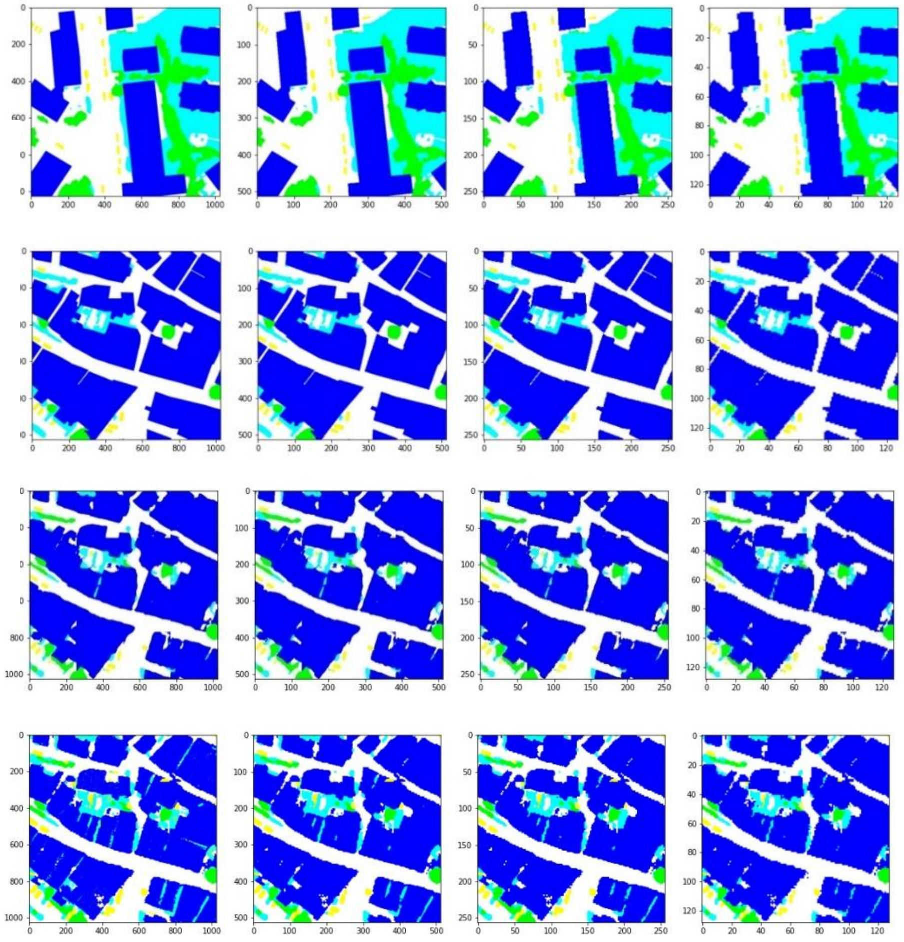


Posterior probabilities with the different techniques

Results with a standard U-Net



Confusion matrices



Results obtained with the full dataset

UniGe Inria

Results with a standard U-Net

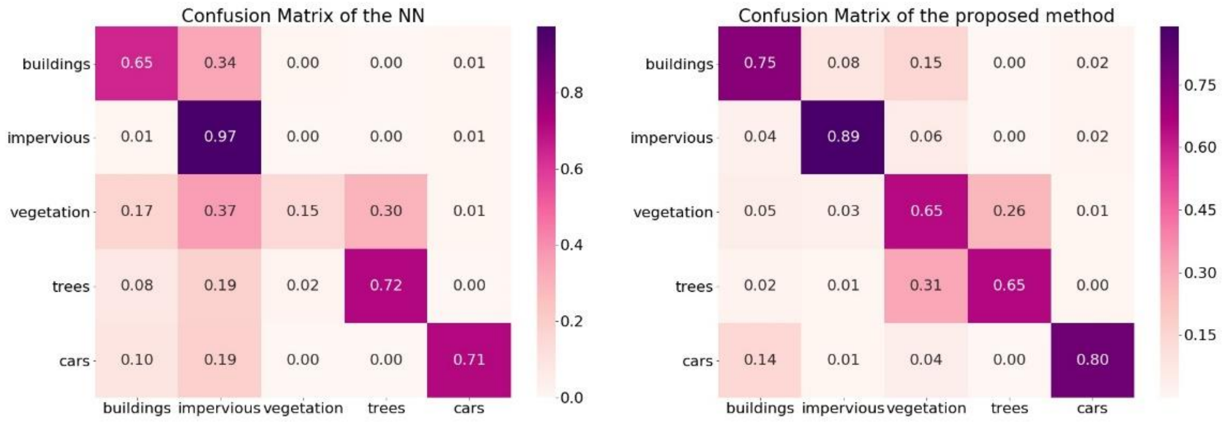
l	Proposed method				RF	Network
	RGB	PGM+NET	NET for cars	Resize		
0	0.86, 0.68, 0.64	0.86, **0.82**, 0.60	0.86, 0.81, 0.58	0.87, 0.80, 0.63	0.75, 0.50, 0.47	**0.90**, 0.78, **0.72**
1	0.88, 0.71, 0.68	0.86, **0.82**, 0.61	0.88, **0.82**, 0.62	0.88, 0.80, 0.65	**0.90**, 0.74, **0.73**	**0.90**, 0.78, 0.72
2	0.88, 0.70, 0.69	0.86, **0.81**, 0.61	0.88, **0.81**, 0.62	0.88, 0.80, 0.66	0.88, 0.71, 0.71	**0.90**, 0.78, **0.72**
3	0.88, 0.68, 0.70	0.87, 0.77, 0.62	0.88, **0.78**, 0.64	0.88, 0.76, 0.67	0.87, 0.65, 0.70	**0.90**, **0.78**, **0.72**

Table of overall accuracy, precision, and recall

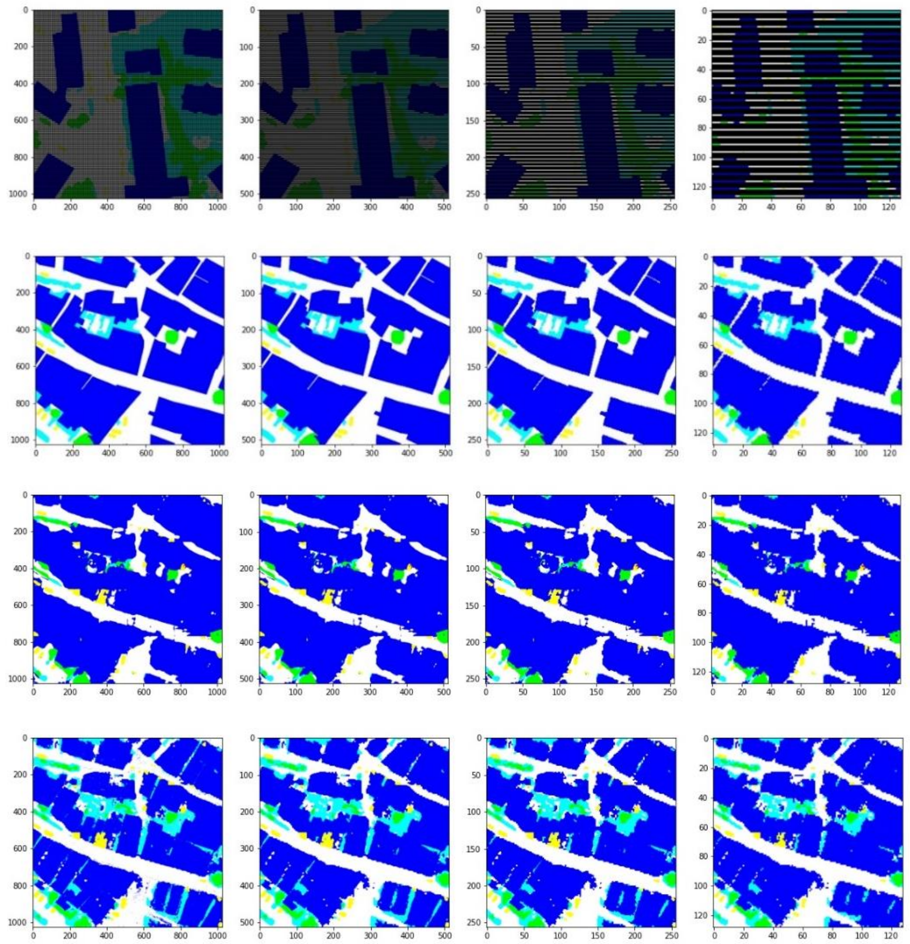
Table of Cohen's kappa coefficient and F1 score

l	Proposed method				RF	Network
	RGB	PGM+NET	NET for cars	Resize		
0	0.66, 0.74	0.69, 0.73	0.68, 0.74	0.70, 0.75	0.48, 0.51	**0.75**, **0.81**
1	0.70, 0.77	0.70, 0.74	0.71, 0.76	0.72, 0.77	0.74, 0.79	**0.75**, **0.81**
2	0.69, 0.77	0.70, 0.75	0.70, 0.77	0.73, 0.77	0.71, 0.77	**0.75**, **0.81**
3	0.69, 0.77	0.69, 0.76	0.70, 0.77	0.71, 0.77	0.67, 0.74	**0.75**, **0.81**

Results with 70% of unlabeled pixels in blocks



Confusion matrices



Results obtained with the 70% of unlabeled pixels

UniGe Inria

Results with 70% of unlabeled pixels in blocks

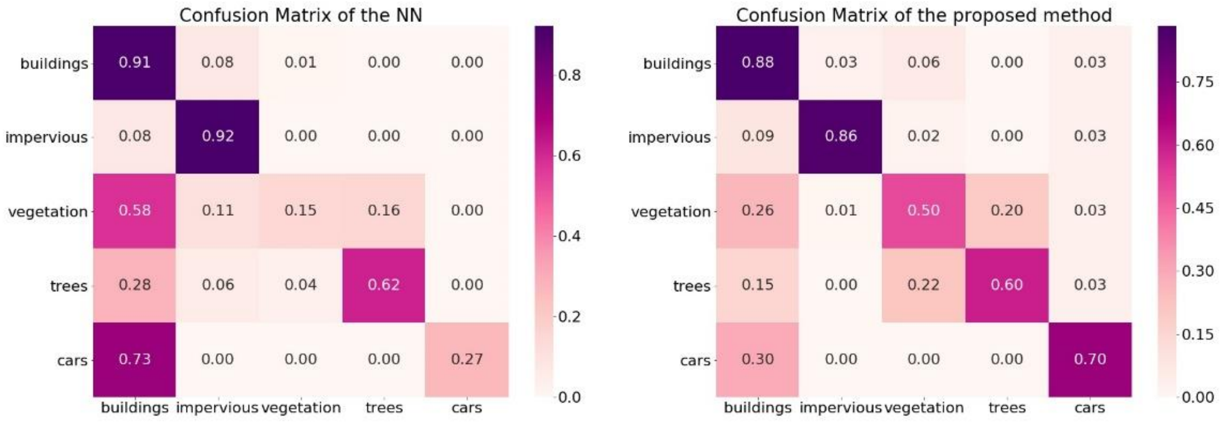
l	Proposed method				RF	Network
	RGB	PGM+NET	NET for cars	Resize		
0	**0.84**, 0.64, **0.67**	0.71, 0.74, 0.46	0.79, **0.75**, 0.54	0.83, **0.75**, 0.57	0.75, 0.50, 0.48	0.83, 0.64, **0.67**
1	0.85, 0.66, **0.67**	0.73, 0.75, 0.47	0.82, **0.76**, 0.54	0.84, 0.75, 0.58	**0.86**, 0.66, 0.64	0.83, 0.64, **0.67**
2	**0.85**, 0.65, 0.67	0.75, 0.74, 0.48	0.83, **0.75**, 0.55	**0.85**, 0.74, 0.59	0.84, 0.62, **0.69**	0.83, 0.64, 0.67
3	**0.86**, 0.64, 0.69	0.77, 0.71, 0.49	0.84, **0.72**, 0.56	**0.86**, 0.71, 0.60	0.81, 0.55, **0.72**	0.83, 0.64, 0.67

*Table of overall accuracy, **precision**, and **recall***

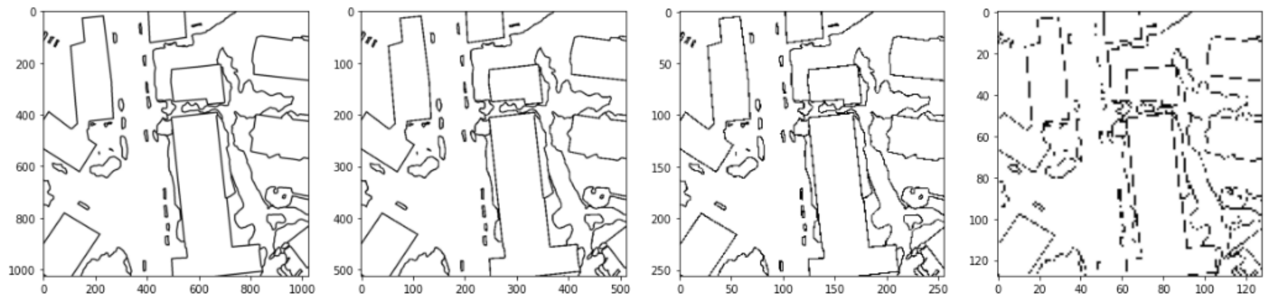
*Table of **Cohen's kappa coefficient** and F1 score*

l	Proposed method				RF	Network
	RGB	PGM+NET	NET for cars	Resize		
0	**0.65**, **0.70**	0.57, 0.53	0.63, 0.63	**0.65**, 0.69	0.49, 0.51	**0.65**, 0.64
1	**0.66**, **0.71**	0.58, 0.55	0.63, 0.67	0.65, **0.71**	0.65, 0.71	0.65, 0.64
2	**0.66**, **0.72**	0.58, 0.57	0.63, 0.68	**0.66**, **0.72**	0.65, 0.69	0.65, 0.64
3	**0.66**, **0.73**	0.58, 0.59	0.63, 0.70	0.65, **0.73**	0.62, 0.64	0.65, 0.65

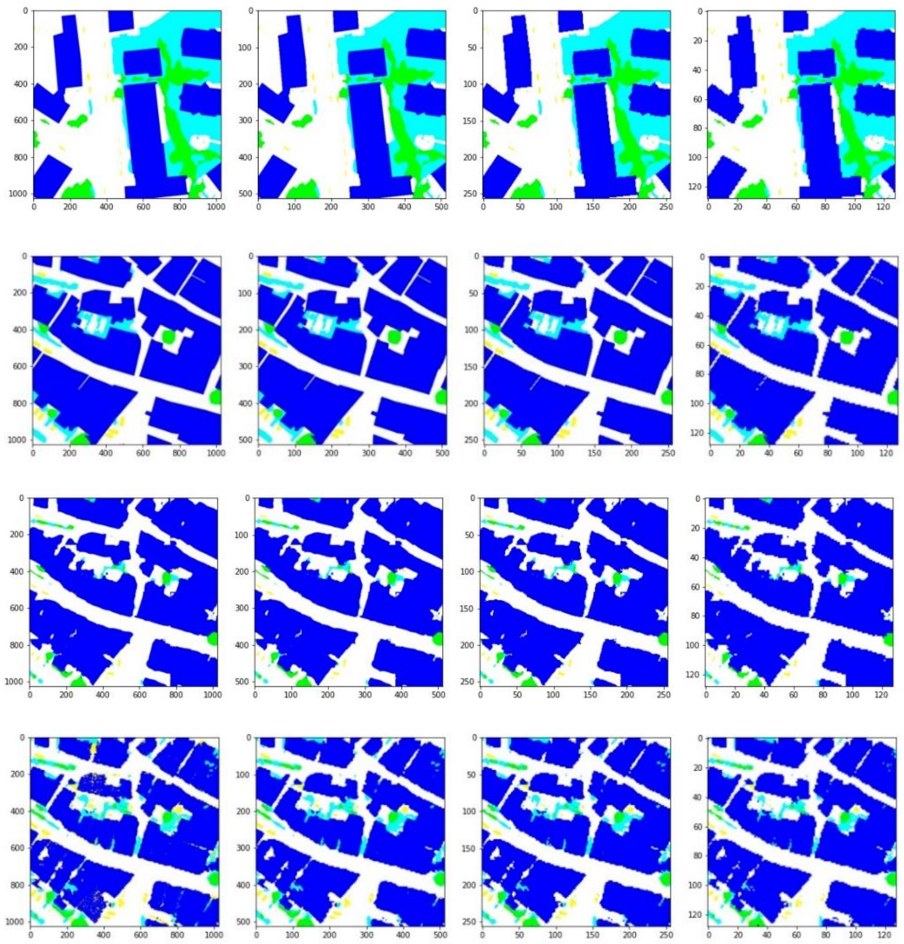
Results with the erosion morphological operator



Confusion matrices



Highlight of the eroded parts



Results obtained with the eroded dataset

Results with the erosion morphological operator

l	Proposed method				RF	Network
	RGB	PGM+NET	NET for cars	Resize		
0	0.86, 0.58, 0.55	0.76, 0.68, 0.55	0.85, **0.71**, 0.57	0.86, 0.66, 0.68	0.74, 0.47, 0.49	**0.87**, 0.57, **0.77**
1	0.86, 0.58, 0.54	0.78, 0.67, 0.56	0.86, **0.70**, 0.64	**0.87**, 0.66, 0.72	0.85, 0.52, 0.75	**0.87**, 0.57, **0.77**
2	0.86, 0.57, 0.54	0.78, 0.66, 0.57	**0.87**, **0.69**, 0.64	**0.87**, 0.64, 0.72	0.83, 0.48, 0.65	**0.87**, 0.57, **0.77**
3	0.86, 0.56, nan	0.80, 0.62, 0.58	**0.87**, **0.64**, 0.65	**0.87**, 0.60, 0.72	0.83, 0.50, 0.47	**0.87**, 0.57, **0.77**

*Table of overall accuracy, **precision**, and **recall***

*Table of **Cohen's kappa coefficient** and F1 score*

l	Proposed method				RF	Network
	RGB	PGM+NET	NET for cars	Resize		
0	0.56, 0.73	0.61, 0.60	0.63, 0.72	**0.67**, 0.74	0.48, 0.50	0.66, **0.75**
1	0.56, 0.74	0.61, 0.61	0.67, 0.74	**0.69**, **0.75**	0.61, 0.71	0.66, **0.75**
2	0.55, 0.74	0.61, 0.62	0.66, 0.74	**0.68**, **0.75**	0.55, 0.68	0.66, **0.75**
3	nan, 0.74	0.60, 0.64	0.64, 0.74	0.65, 0.74	nan, 0.68	**0.66**, **0.75**

Conclusion and future work

- **Novel method for semantic segmentation of remote sensing images mixing FCNs and hierarchical PGMs**
  - Surpasses the accuracy as per the *recall* of the standard FCNs studied
  - Outperforms the state-of-the-art in the classification of minority classes, while maintaining adequate classification results for all classes
  - Advantages are progressively more relevant as the training set is farther from the ideal densely-labeled case
- **Perspectives for future work**
  - Addition of feedforward neural networks to compute the pixelwise posterior probabilites to replace RF
    - Mix directly deep learning and PGMs without the addition of another classifier
  - Test the proposed method with another dataset
    - Same encoding of the classes but different complexity and features
  - Test with data associated with other applications
    - Natural disasters management (e.g., earthquakes, landslides, floods, etc.)

Thank you for your attention!

Università di Genova

Inria