How useful is region-based classification of remote sensing images in a deep learning framework ?

Nicolas Audebert[1,2], Bertrand Le Saux[1], Sébastien Lefèvre[2]
[1]: ONERA, The French Aerospace Lab
[2]: Université de Bretagne-Sud/IRISA
nicolas.audebert@onera.fr

IGARSS 2016, Beijing

Introduction

Deep learning for remote sensing

- The *deep features* + classifier baseline is becoming more and more popular for aerial images
- Excellent results in classification (Penatti et al., CVPR Workshop 2015)
- State-of-the-art for semantic mapping on the DFC using region-based deep features enhanced classifiers (Lagrange et al., IGARSS 2015)
- More and more (annotated) remote sensing data available that makes supervised learning a reality

Motivation and problem statement

Deep learning for remote sensing

Deep features as a baseline

Deep features

Deep networks as feature extractors (Razavian et al., CVPR Workshop, 2015)

- Get a pre-trained deep network (e.g. AlexNet) on ImageNet
  - ⇒ Optional: fine-tune the network on remote sensing data
- Extract the feature vector (e.g. from the last fc layer)
- Classification using a SVM, RF...

Why this baseline ?

- Deep nets trained on ImageNet (1M images, 1K classes)
  - ⇒ No RS data but huge variability (cars, dogs, cats...)
  - ⇒ Convolutional filters in the first layers are generic
- It works ! (Penatti et al., "Do deep features generalize from everyday objects to remote sensing and aerial scenes domains ?", CVPR Workshop, 2015)

Deep learning for remote sensing

Region-based classification

Why region-based classification ?

- Semantic mapping = giving a label to every pixel
- But labeling every pixel individually is time consuming
- Semantic mapping = segmentation + classification

Segmentation + Classification →

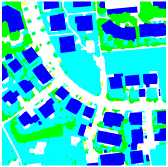


Figure: Region-based classification : segment then classify the regions

Deep learning for remote sensing

Semantic mapping baseline

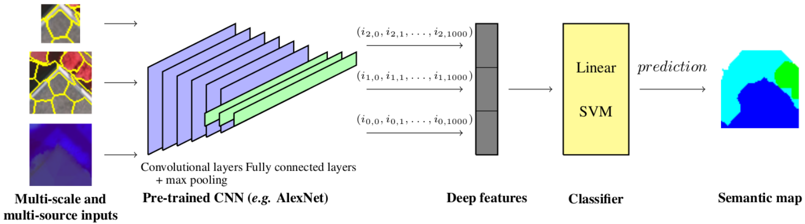
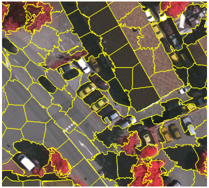


Figure: Deep learning based framework for semantic segmentation of remote sensing images

Problem statement

How to choose the regions to be classified ?



. . .

The ideal segmentation

- Homogeneous (w.r.t the ground truth) regions

$\Rightarrow$ Mixed classes in a region $\Rightarrow$ unavoidable classification error

- Regions of similar sizes ?

$\Rightarrow$ Better normalization for the machine learning pipeline

- But more importantly, results in a good final accuracy

Segmentation comparison

Comparing several segmentation algorithms

Experimental setup

Tested algorithms

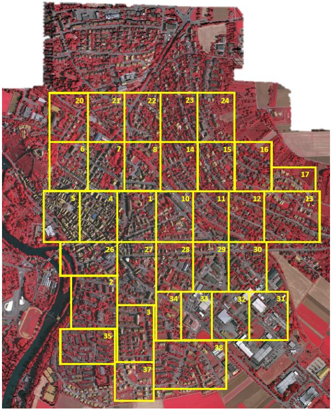
- Compare algorithms from both computer vision and remote sensing communities
  - Sliding window baseline
  - Superpixels (SLIC (Achanta et al., 2010), LSC (Li and Chen, 2015), Quickshift (Vedaldi and Soatto, 2008))
  - HSEG (Tilton et al., 2012)
  - Multiresolution Segmentation (Baatz and Schäpe, 2012)

Two approaches

- Image based segmentation: use generic pixel-related informations (colors, coordinates) ⇒ superpixels
- "Expert" segmentation: use a well-defined similarity criterion based on remote sensing data knowledge

Comparing several segmentation algorithms

Experimental setup



Dataset and metrics

- ISPRS Vaihingen Dataset (16 IRRG tiles, urban area)[1]
- Using our deep features based multi-scale baseline
- Metrics:
  - Segmentation (borders, region purity. . .)
  - Classification accuracy
- Algorithm parameters chosen by cross-validation

[1] http://www2.isprs.org/commissions/comm3/wg4/tests.html

Results

What is the segmentation impact on the classifier ?

Segmentation comparison

Algorithm	Regions	UE (%)	BR (%)	AP (%)	Oracle (%)
SLIC	**20 000**	**10.21**	84.07	75.10	89.91
LSC	22 800	11.37	91.13	71.54	85.83
Quickshift	21 000	11.66	88.34	72.90	83.61
MRS	23 500	13.12	**95.71**	**79.08**	**91.68**
HSEG	21 000	11.39	94.83	78.66	85.25

Table 1: Segmentation metrics on the ISPRS dataset

Algorithm	Regions	Acc. (%)	F1_car	κ
SLIC	**20 000**	82.20	0.54	0.76
LSC	22 800	**82.45**	**0.58**	0.76
Quickshift	21 000	82.05	0.52	0.75
MRS	23 500	80.53	0.56	0.73
HSEG	21 000	79.56	0.54	0.72
SW	23 800	81.22	0.53	0.74

Table 2: Classification metrics on the ISPRS dataset

Segmentation metrics

- Superpixels are generally less precise than expert-designed segmentation algorithms
- The oracle (perfect) classification suggests to use HSEG or MRS (“expert” segmentations)
- However, “good” segmentation $\not\Rightarrow$ good classification accuracy

What is the segmentation impact on the classifier ?

Segmentation comparison

Algorithm	Regions	UE (%)	BR (%)	AP (%)	Oracle (%)
SLIC	**20 000**	**10.21**	84.07	75.10	89.91
LSC	22 800	11.37	91.13	71.54	85.83
Quickshift	21 000	11.66	88.34	72.90	83.61
MRS	23 500	13.12	**95.71**	**79.08**	**91.68**
HSEG	21 000	11.39	94.83	78.66	85.25

Table 1: Segmentation metrics on the ISPRS dataset

Algorithm	Regions	Acc. (%)	F1_car	κ
SLIC	**20 000**	82.20	0.54	0.76
LSC	22 800	**82.45**	**0.58**	0.76
Quickshift	21 000	82.05	0.52	0.75
MRS	23 500	80.53	0.56	0.73
HSEG	21 000	79.56	0.54	0.72
SW	23 800	81.22	0.53	0.74

Table 2: Classification metrics on the ISPRS dataset

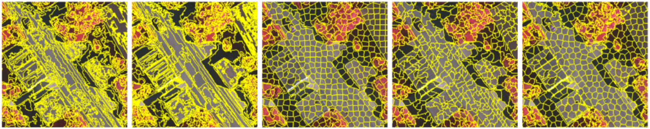
Analysis

Two complementary phenomena:

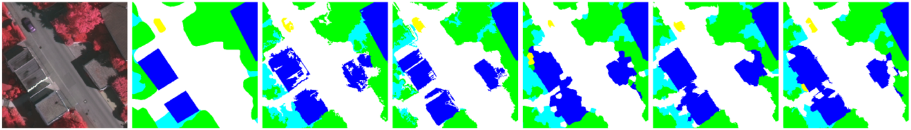
- Segmentation accuracy (adherence to boundaries, no mixed classes in one region . . .)
- Homogeneous samples ⇒ better classification as normalization decreases the knowledge to be inferred (e.g. the shape of the RoI in the processed patch)

What is the segmentation impact on the classifier ?

Qualitative comparison



(a) MRS (b) HSEG (c) SLIC (d) Quickshift (e) LSC



(a) Orthoimage (b) Ground truth (c) MRS (d) HSEG (e) SLIC (f) Quickshift (g) LSC

What is the segmentation impact on the classifier ?

Analysis

Expert segmentations

- \+ Good for object proposals
- \+ Very accurate segmentation
- \- Inhomogeneous regions in scale and shape

Superpixels

- \+ Bounded shape and scale
- \+ Accurate enough segmentation

Conclusion

Conclusion

Region-based classification pipeline with deep features

- Segmentation pre-processing by partitioning the image
- Deep features generation using a pre-trained CNN
- SVM-based classification

Choosing a segmentation algorithm

- Two criteria: segmentation accuracy and sample shape/size
- Superpixel algorithms perform similarly and outperform both sliding window baseline and expert segmentation
- Some segmentations can significantly outperform others on some specialized tasks (e.g. vehicle detection)

The end.

Thank you for your attention !

Feel free to ask questions !

Contact e-mail

nicolas.audebert@onera.fr

The authors would like to acknowledge the joint ONERA-TOTAL project NAOMI for providing the funding for this work.