DETECTION OF URBAN TREES IN MULTIPLE-SOURCE AERIAL DATA (OPTICAL, INFRARED, DSM)

Lionel Pibre, Advisors: Marc Chaumont, Gérard Subsol, Dino Ienco, and Mustapha Derras

ABSTRACT

Standard Remote Sensing analysis uses machine learning methods such as SVMs with HOG or SIFT descriptors, but in recent years neural networks are emerging as a key tool regarding the detection of objects. Due to the heterogeneity of remote sensing information (optical, infrared, DSM) the combination of multi-source data is still an open issue. In this paper, we focused on localization of urban trees, and we evaluate the performances of CNNs compared to standard classification methods that employ descriptor-based representation.

Index Terms— Deep Learning, Machine Learning, Detection, Localization, Multi-source data

1. INTRODUCTION

In remote sensing, many object detection methods use machine learning and combines the extraction of descriptors such as HOG and efficient classifiers such as SVM [1]. In recent years, Convolutional Neural Networks (CNNs) [2] appeared, integrating in a single optimization scheme these two steps. In the case of multi-source data (optical, infrared, LiDAR), it is not easy to combine different types of information since they provide measures that can be very different considering dimensionality, range values and/or scales. It is therefore necessary to standardize the method and can have a great influence on the results. CNNs can deal with these issues by normalizing the input values and, at the same time, learning a discriminative model. In this abstract, we propose to assess the performance of CNNs compared to methods using image descriptor and a classifier in processing multi-source data. We will compare two well known CNNs, AlexNet and GoogleNet and two machine learning methods based on the same HOG image descriptor [3] but with two different powerful classifiers. As application example, we will take the detection and the localization of urban trees in aerial data composed of optical, near infrared and Digital Surface Model (DSM) measurements.

L. Pibre is with the LIRMM laboratory, University of Montpellier, and with Berger-Levrault company, Montpellier, France (email: lionel.pibre@lirmm.fr).

M. Chaumont is with LIRMM laboratory, University of Montpellier and University of Nîmes, France (email: marc.chaumont@lirmm.fr).

G. Subsol is with LIRMM laboratory, CNRS, Montpellier, France (email: gerard.subsol@lirmm.fr).

D. Ienco is with UMR-TETIS laboratory, IRSTEA, France (email: dino.ienco@irstea.fr).

M. Derras is with Berger-Levrault company, Montpellier, France (email: mustapha.derras-levrault.com).

2. RELATED WORK

Object detection constitutes an important task in the field of image analysis [4, 5].

In [6], the authors propose a taxonomy of the different object detection strategies organized in three families: template-based, knowledge-based and machine learning-based methods. The conclusion is that most of the approaches are still dominated by handcrafted features such as Histogram of Oriented Gradients (HOG) or Bag-of-Words (BoW).

Considering the task to detect trees in urban areas, in [7], the authors propose to combine spectral, hyperspectral and LiDAR data to classify different species of trees. The works heavily relies on the construction and selection of handcrafted features for each type of source (i.e. NDVI). It also investigates which source of information needs to be retained in order to increase the classification performances. The final classification is accomplished by Linear Discriminant Analysis.

In the last decade, Deep Learning [8] methods start to show interesting results in general image analysis tasks [9]. Such techniques have the ability to jointly learn i) new features and ii) the associated classifier.

Considering the Remote Sensing field, despite the increasing popularity of Deep Learning approaches, currently, most of the object detection methods are based on handcrafted features that are successively employed as input for machine learning classifiers. Recently, some works [10, 11] start to exploit deep learning for object classification and detection but, unfortunately, none of them leverage such techniques in the context of multi-source data (i.e. spectral and LiDAR) with the purpose of directly learn new data representation avoiding handcrafted features.

3. METHOD

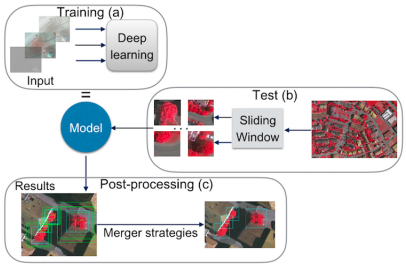

Fig. 1. Overview of the proposed method.

A general outline of our method is presented in Figure 1. We

train a CNN classifier to discriminate between the class "Tree" and the class "Other". The training set is composed of images having all the same size (Figure 1.a). For the test phase, a multi-scale sliding window is applied on the new image. Each sliding window is then sent to the CNN in order to get a probability of belonging to the class "Tree" or "Other" (Figure 1.b). Since the sliding window is applied at different scales, several predictions on the same image area will be output. We successively merge all these outputs [12] in order to get a final and accurate bounding box result of trees in the images (Figure 1.c). In our experiments we used two types of fusions based mechanism: i) on relative areas and ii) overlapping areas [12].

To assess the results, we compute the overlap ratio between the detected bounding box and the ground truth. The ground truth is obtained by which manual segmentation as in the Pascal Voc challenge[1].

All experiments were realized on the Vaihingen database with a 5-fold cross validation. This data set was captured over Vaihingen in Germany[2]. It consists of three areas, inner city, high riser and residential area. The first area is situated in the centre of the city of Vaihingen, it is characterized by dense development consisting of historic buildings having rather complex shapes, but also has some trees. The second area is characterized by a few high-rising residential buildings that are surrounded by trees. The third area is a purely residential area with small detached houses.

4. RESULTS

	AlexNet	GoogleNet	HOG+SVM	HOG+RF
Area				
Recall	56.38%	65.24%	26.66%	38.67%
Precision	43.36%	46.14%	0.95%	7.77%
F-Measure	0.47	0.53	0.01	0.1
Overlap				
Recall	59.62%	49%	21%	33.47%
Precision	31.79%	28.47%	1.54%	10.47%
F-Measure	0.4	0.34	0.03	0.13

Table 1. Results given by the two CNNs and the two machine learning methods.

Table 1 shows the results we obtained with the different methods: two CNNs, AlexNet and GoogleNet and two machine learning methods, Random Forest and SVM both with the HOG descriptor and the two fusion methods "Area" and "Overlap".

The results under the "Area" line are the results we have obtained using an area fusion and the results under the line "Overlap" are the results obtained using a overlap fusion.

As we can note, best results are obtained with CNNs when area fusion is considered. In fact, the area fusion appears more restrictive than the fusion overlap. Since CNNs create characteristic vectors with a high level of abstraction, this fusion allows them to greatly reduce the number of false positive and therefore have better accuracy.

[1]http://host.robots.ox.ac.uk/pascal/VOC/voc2012/

[2]The Vaihingen data set was provided by the German Society for Photogrammetry, Remote Sensing and Geoinformation (DGPF) [13]: http://www.ifp.uni-stuttgart.de/dgpf/DKEP-Allg.html.

The Random Forest and the SVM achieve performance well below those of CNNs. Contrary to CNNs, with both methods, the overlap fusion gives better performance than the area fusion. The performances obtained with these methods are extremely low. This may due to the fact that trees are often very close from each other making them difficult to differentiate on the basis of their contours (see Figure 2) which are emphasized by the HOG descriptor.

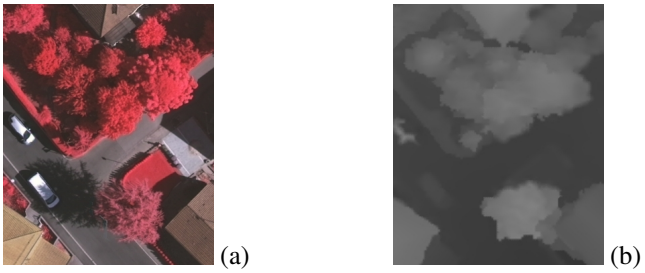

Fig. 2. Example of test image: (a) Red, Green and Near Infrared (b) Digital Surface Model.

5. REFERENCES

[1] Vladimir N Vapnik, "An overview of statistical learning theory," *IEEE Transactions on Neural Networks*, vol. 10, no. 5, pp. 988–999, 1999.

[2] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[3] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2005*, June 2005, vol. 1, pp. 886–893.

[4] W. Diao, X. Sun, X. Zheng, F. Dou, H. Wang, and K. Fu, "Efficient saliency-based object detection in remote sensing images using deep belief networks," *IEEE Geosci. Remote Sensing Lett.*, vol. 13, no. 2, pp. 137–141, 2016.

[5] J. Han, D. Zhang, G. Cheng, L. Guo, and J. Ren, "Object detection in optical remote sensing images based on weakly supervised learning and high-level feature learning," *IEEE Trans. Geoscience and Remote Sensing*, vol. 53, no. 6, pp. 3325–3337, 2015.

[6] G. Cheng and J. Han, "A survey on object detection in optical remote sensing images," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 117, pp. 11–28, 2016.

[7] M. Alonzo, B. Bookhagen, and Dar A Roberts, "Urban tree species mapping using hyperspectral and lidar data fusion," *Remote Sensing of Environment*, vol. 148, pp. 70–83, 2014.

[8] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 52, no. 8, pp. 436–444, 2015.

[9] A. Krizhevský, I. Sutskever, and G. Hinton, "Imagenet classification with deep convolutional neural networks," in *Neural Information Processing Systems*, 2012, pp. 1097–1105.

[10] D. Marmanis, M. Datcu, T. Esch, and U. Stilla, "Deep learning earth observation classification using imagenet pretrained networks," *IEEE Geosci. Remote Sensing Lett.*, vol. 13, no. 1, pp. 105–109, 2016.

[11] L. Zhang, G.-S. Xia, T. Wu, L. Lin, and X.-C. Tai, "Deep learning for remote sensing image understanding," *J. Sensors*, vol. 2016, pp. 7954154:1–7954154:2, 2016.

[12] Pierre Sermanet, David Eigen, Xiang Zhang, Michaël Mathieu, Rob Fergus, and Yann LeCun, "Overfeat: Integrated recognition, localization and detection using convolutional networks," *International Conference on Learning Representations*, 2014.

[13] Michael Cramer, "The dgpf-test on digital airborne camera evaluation–overview and test design," *Photogrammetrie-Fernerkundung-Geoinformation*, vol. 2010, no. 2, pp. 73–82, 2010.