Convolutional Neural Networks for the Extraction of Built-Up Areas from Sentinel-2 Images

Khelifa Djerriri
Centre des Techniques Spatiales
Département Observation de la Terre
Arzew, Algeria
kdjerriri@cts.asal.dz

Reda Adjouj
Djillali Liabes University
Departement of Computer Science
Sidi Bel Abbes, Algeria
adjoudj@univ-sba.dz

Dalila Attaf
Centre des Techniques Spatiales
Département Observation de la Terre
Arzew, Algeria
dattaf@cts.asal.dz

Abstract

Monitoring of the human-induced changes and the automatic mapping of urban areas were always a main concern to researchers in the field of remotely sensed image processing. Thus, several techniques have been proposed to saving technicians from interpreting and digitizing hundreds of areas by hand. In this work, we propose to exploit the benefit of Sentinel-2 images to extract urban areas. The approach relies on deep features extraction using a pre-trained convolutional neural network (CNN) and random forest classification. Experiments are performed on PoDelta area, in Italy. Results, validated with a Kappa index over 0.74, illustrate the great interest of Sentinel-2 in operational projects, such as Corine land-cover mapping, and since such an approach can be conducted on very large areas, such as the European or global scale.

Keywords: Sentinel-2; urban areas; convolutional neural network, random forest classification.

1 Introduction

Monitoring of the human-induced changes and especially of urban areas was always a key task that is also increasingly in demand for a number of applications (urban planning, health monitoring, ecology, etc.) [1].

Understanding the urban growth phenomenon is among the major issues that public services have to deal with. Today, more than half of the world population lives in urban areas, and it is estimated that this will reach up to two-thirds by 2025 [2].

The recent launch of the Sentinel-2A satellite in June 2015 makes available data with a minimum spatial resolution of 10 m, 13 spectral bands, wide acquisition coverage and short time revisits, which opens a large scale of new applications [3].

Recently, deep learning has become the new state-of-the-art solution for image processing. Given its success, deep learning based techniques have been intensively used in several distinct tasks of different domains, including remote sensing, where they have demonstrated excellent performance on different tasks, such as VHSR, hyper spectral and LIDAR data classification [4,5,6], buildings and roads extraction [7,8], or image pansharpening [9].

In practice, very few people train an entire Convolutional Network from scratch (with random initialization), because it is relatively rare to have a dataset of sufficient size. Instead, it is common to pretrain a CNN on a very large dataset (e.g. ImageNet, which contains 1.2 million images with 1000 categories), and then use it a fixed feature extractor, which is our case.

In this paper, a method based on exploiting Sentinel-2 images is proposed to extract urban areas. The approach relies on deep features extraction using a pre-trained CNN and random forest classification. Experiments are performed on data from the PoDelta area, in Italy.

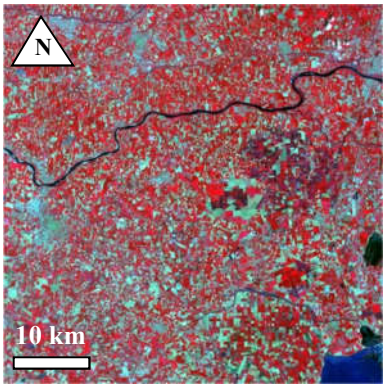
2 Data and methods

In this research, a patch-based analysis approach is adopted. It is performed in three main steps: superpixel based image segmentation and patches generation, CNN based feature extraction and supervised classification.

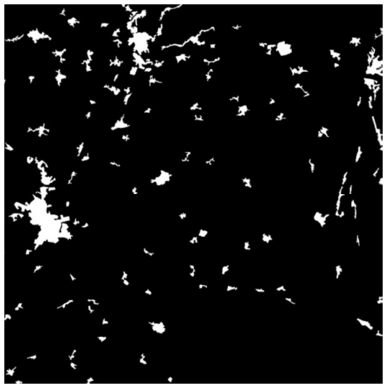
2.1 Data description

In this work, we evaluate our approach on the Sentinel-2 dataset that consists of 2 multispectral sub-images acquired over the area of PoDelta, Italy on 04 July 2015. The size of the two images is 5490x5490 pixels and they are composed of 3 channels (NIR-R-G). The spatial resolution of the images is 10 meters per pixel. The reference set used for the example was extracted from the vector version of Corine Land Cover (CLC) product of year 2006 (Copernicus Land Monitoring Services, http://land.copernicus.eu/pan-european/corine-land-cover/clc-2006/). The positive examples for the "built-up" class were extracted from the CLC classes of continuous urban fabric (code 1.1.1), discontinuous urban fabric (code 1.1.2), and industrial or commercial units (code 1.2.1). The CLC reference set can be considered as affected by thematic and spatiotemporal noise as regarding the target classes in the Sentinel-2 image. A time gap of nine years (2006 for the CLC and 2015 for the Sentinel-2 data) can be assumed between the image data and the CLC data [10]. The reference set used for the example is shown in Figure 1. The positive examples for the built-up class as extracted from the CLC source, are represented in white.

Figure 1: Presentation of the study area.



a. Training image



b. Training Ground truth

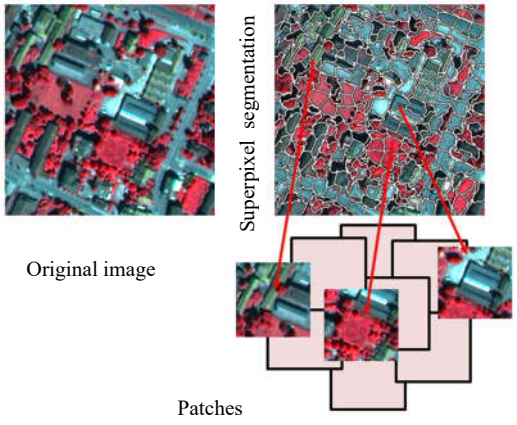
c. Localization of the study area

2.2 Superpixel-based labeling approach

During the feature extraction, training and testing phases, CNN needs input windows of a fixed size. Our analysis is thus performed based on patches derived from superpixel segmentation instead of the time consuming sliding windows strategy [11]. In order to take into account the limits between the different classes in the original image, we compute superpixels with the SLIC algorithm (Simple Linear Iterative Clustering) [11].

Superpixels typically cover the whole image; they are distributed regularly with respect to the nature of the input image. The desirable variation of superpixels size is preferably small and the boundary of superpixels has to match the natural boundary of the different objects present in the image. The parameters of the superpixel extraction are tuned in order to get around 100~200 pixels by superpixel, which is found as the best compromise between the appearance of the final classified image and the segmentation computation time. The patches are 32 by 32 pixels subimages representing the contextual neighborhood of the objects. One patch is generated for each superpixel centered on the pixel corresponding to its centroid. Figure 2 presents the process of patch creation.

Figure 2. Generation of patches from VHSR image



2.3 Convolutional Neural Networks (CNN)

CNNs have some characteristics that distinguish them from traditional feed-forward neural networks. Unlike traditional feed-forward layers, convolutional layers have neurons with limited receptive fields allowing the processing of local image region that affects a particular element in the output. Moreover, as their name reflects, the output of this layer is computed as a spatial convolution using a learned filter over its input. Two main contributions have been the proposal of the rectified linear unit (ReLu) [12], which allows a faster training, and the dropout strategy [13] to reduce overfitting.

CNN architecture typically comprises several layers of different types [14]:

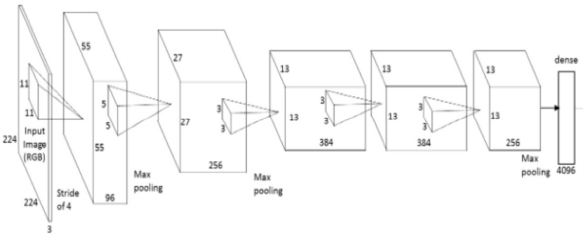
1) Convolutional layers. They compute the convolution of the input image with the weights of the network. These layers are characterized by few parameters: the size of filters, the filter spatial support, the step between different windows and an optional zero-padding which controls the size of the layer output. As the layers are deeper, the features extracted from the image are higher-level.

2) Pooling layers. The mission of these layers is to reduce the size of the input layer through some local non-linear operations. Their most important parameters are the support of the pooling window and the step between different windows.

3) Normalization layers. Their objective is to improve generalization of the CNN. Neurons typically used in these layers are sigmoid type Fully-connected layers. These layers have the capacity of abstracting the low-level information generated in previous layers for a final decision. Figure 3 shows the architecture of the CNN used as feature extractor.

.

Figure 3: pre-trained CNN architecture.



2.4 Random forest classification

Random forests (RF) are similar to tree classifiers, given that these do not use a method of "bootstrapping" that can be improved upon. In training, the random forest algorithm creates multiple trees, each trained on a bootstrapped sample of the original training data, and searches only across a randomly selected subset of the input variables to determine a split (for each node). For classification, each tree in the random forest casts a unit vote for the most popular class at input x. The output of the classifier is determined by a majority vote of the trees. The random forest algorithm can handle high dimensional data, which is the case of deep features and use a large number of trees in the ensemble. As each tree is only using a portion of the input variables in a random forest, the algorithm is considerably lighter than conventional bagging (bootstrap aggregating) [15] with a comparable tree-type classifier.

3 Experiments and results

After SLIC based segmentation of the training image (Figure 1 a), for each superpixel the deep feature was calculated using the pre-trained CNN for a patch of 32 by 32 pixels around its centroid. A supervised random forest classifier with 50 trees was trained, then tested on unseen test sub-image (figure 4 a). In this work, CNN is exploited only as a feature extractor and classification experiments have been made by using external classifier. The deep features, are obtained by removing the last classification layer and considering the output of previous layers. The individual pixels of the superpixels were labeled as the class designed by the random forest classifier.

We evaluated the classification of deep features by employing two (02) different criteria: the estimated Cohen's Kappa statistic (k), and the F1 score (F1). The first metric k is an overall accuracy metric which compensates for the chance agreement between classes. It is calculated by multiplying the total number of pixels in all the ground truth classes (N) by the sum of the confusion matrix diagonals (x_{kk}), subtracting the sum of the ground truth pixels in a class times the sum of the classified pixels in that class summed over all classes ($\sum_k x_{k\Sigma}x_{\Sigma k}$), and dividing by the total number of pixels squared minus the sum of the ground truth pixels in that class times the sum of the classified pixels in that class summed over all classes.

$$k = \frac{N\sum_k x_{kk} - \sum_k x_{k\Sigma}x_{\Sigma k}}{N^2 - \sum_k x_{k\Sigma}x_{\Sigma k}} \qquad (1)$$

The second score F1 is the average of the harmonic means between precision and recall for each class. This measure is sensitive to class accuracy, but additionally takes into account the number of correctly classified pixels over the number of predicted labels for each class (built-up and non-built-up classes) [16].

$$F1 = 2.\frac{precision.recall}{precision.recall} \qquad (2)$$

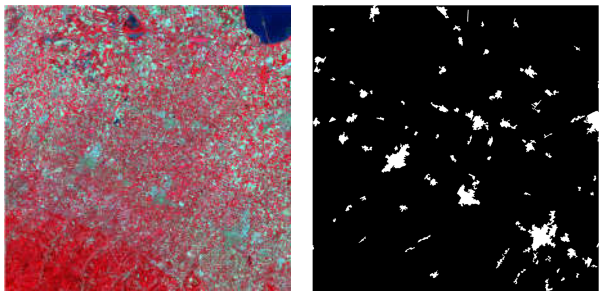
The results of the superpixel-based classification using the pre-trained CNN features have been compared to four (04) state of the art texture features: Integrative co-occurrence matrices (ICOO) and opponent Gabor (OGabor) features, nad wavelets features [17]. The comparison study was conducted using three bands image (nir, red and green spectral bands). Resulting classified images are given in Figure. 4. The evaluation of classification in terms of precision, recall, F1 score and Kappa index are summarized in Table 1.

Table 1: classification results of the test image.

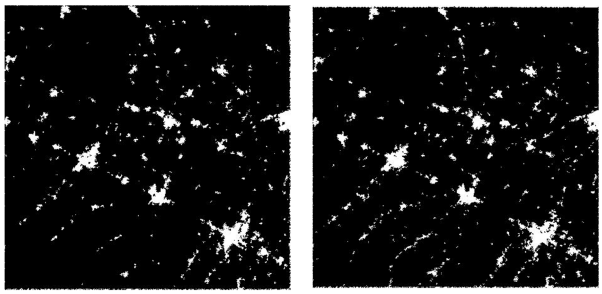
	Precision	Recall	F1 score	Kappa
CNN	**0.8228**	**0.7171**	**0.7663**	**0.7407**
ICOO	0.7603	0.6993	0.7285	0.7163
OGabor	0.6606	0.6901	0.6750	0.6613
Wavelet	0.7972	0.6767	0.7320	0.7195

Table 1 shows that the CNN based method, using deep features improves F score and kappa statistic compared to literature texture analysis methods. For example the CNN based classification recorded an improvement of 0.03 kappa index compared the integrative co-occurrence matrices based texture analysis. The improvement was 0.08 and 0.02 compared to the opponent Gabor, and wavelet methods respectively.

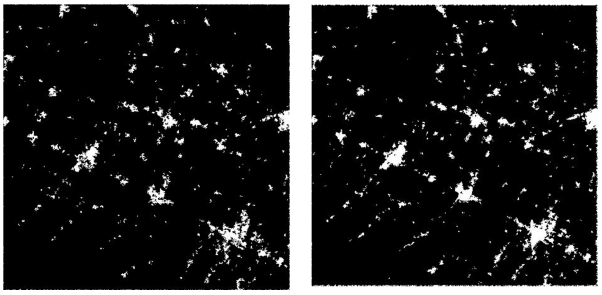
Figure 4: classification maps of the test image.



a. test image b. test ground truth



c. CNN based map d. ICOO map



e. Gabor map f. Wavelet map

Figure 4 shows that the CNN based map visual quality outperforms the remaining maps, which is the most similar to the test ground truth map. The previous results show the applicability of using pre-trained deep convolutional neural networks as feature extractor to classify Sentinel-2 images and extract urban areas.

4 Conclusion

We have applied pre-traind CNN deep features and random forest classifier for the detection of urban areas on Sentinel 2 images. The deep convolutional neural network consists of 23 layers. The proposed CNN based approach extracts deep features from 32-by-32 image patches that are extracted through using of superpixel segmentation based strategy. The obtained features are then classified using external supervised Random forest classifier. The experiments demonstrated high kappa statistic and F1 score of the CNN deep features classification. The experimental results were promising and showed that the proposed method outperforms state of the art texture analysis methods.

References

[1] A. Lefebvre, C. Sannier, C., and T. Corpetti, Monitoring urban areas with Sentinel-2A data: Application to the update of the Copernicus high resolution layer imperviousness degree. Remote Sensing, 8(7), 606. 2016

[2] United Nations. World Population Prospects: The 2015 Revision; Technical Report; United Nations: New York, NY, USA, 2015

[3] J. Radoux,; G. Chomé, D. Jacques, F. Waldner, N. Bellemans, N. Matton, C. Lamarche, R. D'Andrimont and P. Defourny, Sentinel-2's potential for sub-pixel landscape feature detection. Remote Sensing, 8, 488. 2016

[4] W. Zhao, S. Du. Learning multiscale and deep representations for classifying remotely sensed imagery. ISPRS Journal of Photogrammetry and Remote Sensing, 113, 155-165. 2016

[5] Y. Chen, Z. Lin, X Zhao, G. Wang, and Y. Gu. Deep learning-based classification of hyperspectral data. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 7(6), 2094-2107. 2014

[6] H. Guan, Y. Yu, Z. Ji, J. Li, and Q. Zhang. Deep learning-based tree classification using mobile LiDAR data. Remote Sensing Letters, 6(11), 864-873. 2015

[7] M. Vakalopoulou, K. Karantzalos, N. Komodakis, and N. Paragios. Building detection in very high resolution multispectral data with deep learning features. In 2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS) (pp. 1873-1876). 2015.

[8] V. Mnih, G. E Hinton. Learning to detect roads in high-resolution aerial images. In European Conference on Computer Vision (pp. 210-223). Springer Berlin Heidelberg. 2010

[9] W. Huang, L. Xiao, Z. Wei, H. Liu, and S. Tang. A new pan-sharpening method with deep neural networks. IEEE Geoscience and Remote Sensing Letters, 12(5), 1037-1041. 2015

[10] M. Pesaresi, C. Corbane, A. Julea, J. Florczyk, V. Syrris, and P. Soille, Assessment of the added-value of sentinel-2 for detecting built-up areas. Remote Sensing, 8(4), 299. 2016

[11] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua. and S. Susstrunk,. SLIC Superpixels Compared to State of-the-Art Superpixel Methods. Pattern Analysis and Machine Intelligence, IEEE Transactions 34(11), pp. 2274–2282. 2012

[12] K. Jarrett, K. Kavukcuoglu, Y. Lecun, et al. What is the best multi-stage architecture for object recognition? In: 2009 IEEE 12th International Conference on Computer Vision, IEEE, 2009, pp. 2146–2153.

[13] G. E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever. and R. Salakhutdinov,. Improving neural networks by preventing co-adaptation of feature detectors. arXiv preprint arXiv:1207.0580. 201

[14] M. Castelluccio, G. Poggi, C. Sansone and L. Verdoliva. Land use classification in remote sensing images by convolutional neural networks. arXiv preprint arXiv:1508.00092. 2015

[15] J. XIA, "Multiple classifier systems for the classification of hyperspectral data", 2014

[16] M. Volpi, V. Ferrari; Semantic segmentation of urban scenes by learning local class interactions, In IEEE CVPR 2015 Workshop Looking from above: when Earth observation meets vision (EARTHVISION), Boston, USA, 2015.

[17] F. Bianconi, R. Harvey, P. Southam, and A. Fernández. Theoretical and experimental comparison of different approaches for color texture classification. Journal of Electronic Imaging, 20(4), 043006-043006, 2011