Per City-Block, Density Estimation at Build-Up Areas from Aerial RGB Imagery with Deep Learning

Theodosia Vardoulaki[a], Maria Vakalopoulou[a b], Konstantinos Karantzalos[a]
[a] Remote Sensing Laboratory, National Technical University of Athens, Greece
[b] Center for Visual Computing, CentraleSupélec, Inria, Université Paris-Saclay, France
E-mail: vtheodosia@gmail.com, mariavak@central.ntua.gr, karank@central.ntua.gr

***Abstract*—Estimating the density of the 'urban fabric' land cover classes is of major importance for various urban and regional planning activities. However, the generation of such maps is still challenging requiring significant time and labor costs for the per city-block analysis of very high resolution remote sensing data. In this paper, we propose a supervised classification approach based on deep learning towards the accurate density estimation of build-up areas. In particular, for the training procedure we exploit information both from maps (open street, google, etc) and from very high resolution RGB google image mosaics. A patch-based, deep learning model was trained against five land cover classes. During the prediction phase the per city-block classification procedure delivered the locations and percentages of impervious, soil and green regions. Experimental results and validation at two European cities *i.e.,* Athens and Bilbao, indicated overall accuracy rates of 95%. Results, also, highly match with the corresponding layers from the Copernicus Urban Atlas product.**

I. Introduction

Currently a 54% of the world's population lives in urban areas, while projections show that urbanization combined with the overall growth of the worlds population could add another 2.5 billion people to urban populations by 2050. Close to 90 percent of the increase is concentrated in Asia and Africa. In Europe, 75% of the population currently lives in cities and 80% is expected to do so by 2020.

Therefore, up-to-date and comparable information on land cover and land use are crucial to cope with emerging issues such as urban sprawl, the decrease in urban-green areas or the sustainable urban development in general [1], [2], [3]. An understanding of the implications of changes in land cover and land use is a fundamental part of planning for sustainable development. Urban and regional planners need accurate data in order to efficiently monitor and interpret land cover changes.

However, the generation of such maps including impervious surface or man-made object detection is not a trivial task and a significant amount of time and labour cost is required in order to acquire, process and analyse very high resolution ($\leq$2.5m) remote sensing data [4], [5], [6], [7]. An example is Urban Atlas [8] (from the Copernicus Land Monitoring Services) which consist of decent, pan-European comparable land use and land cover data for Large Urban Zones (>100.000 inhabitants). Urban Atlas provides reliable, inter-comparable, high-resolution land use maps for 305 large urban zones as well as their surroundings for the reference years of 2006 and 2012.

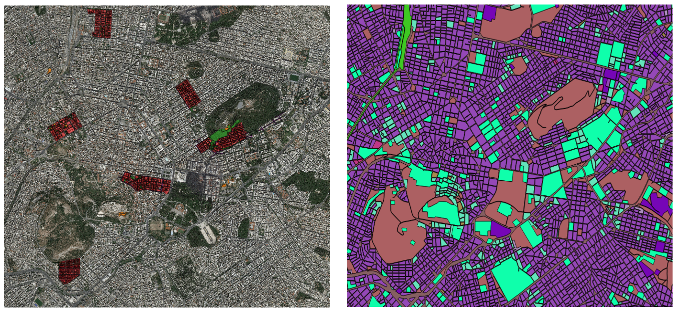


Fig. 1. Urban density at the city of Athens, Greece. On the left-hand side: the per-city block, ground truth data superimposed onto Google's image mosaic. On the right: the corresponding Urban Atlas Map

The motivation in this paper was to design a procedure for urban land cover mapping which is able to exploit all the available geoinformation including both very high resolution images as well as geodata from *e.g.,* open street or google maps. In particular, based on google image mosaics and the corresponding street maps all the city-blocks are detected and labelled. Then, a supervised classification approach was developed based on deep learning (DL) frameworks [9], [10] towards the accurate, per city-block, density estimation of build-up areas. The training procedure integrates information both from open street or google maps, as well as from very high resolution RGB image mosaics from google maps. The patch-based, deep-learning model was trained against five land cover classes of 'urban fabric'. During the prediction phase the per city-block classification procedure delivered the locations and percentages of impervious, soil and green regions.

The rest of the paper is organised as follows. In Section II we present the DL architecture used and we describe the training and testing procedure. Finally, in Sections III and IV we present and discuss our experimental results.

II. Methodology

During the first step of the developed methodology the city-blocks were detected and labeled with a unique number based on the different colors of the google map image which denote the roads (usually white, yellow), urban green (green color), built-up areas (grey color), *etc*. Then a supervised classification procedure based on the corresponding RGB image mosaics was developed towards the detection of the main 'urban fabric' land cover classes. These classes were then combined, during the last processing step, towards the per city-block estimation

978-1-5090-5808-2/17/$31.00

# of pixels	Reference Data Building	Building Shadow	Vegetation	Soil	Vegetation Shadow	Total	UA (%)
Classification							
Building	**225465**	2997	4967	3084	1226	237739	94.8
Building Shadow	1974	**8563**	1958	9	2344	14848	57.7
Vegetation	2592	1353	**23806**	230	1471	29452	80.8
Soil	632	6	319	**1393**	0	2350	59.3
Vegetation Shadow	243	693	543	27	**2436**	3942	61.8
Total	230906	13612	31593	4743	7477	**288331**	
PA (%)	97.6	62.9	75.4	29.4	32.6		
	Overall accuracy = 90.8% Kappa coefficient = 71.6%						

TABLE I. THE RESULTING CONFUSION MATRIX AFTER THE EMPLOYED CONVNET DEEP-LEARNING MODEL. THE RESULTING OVERALL ACCURACY WAS 90.8%, WHILE THE KAPPA COEFFICIENT WAS 71.6%.

of urban density.

A. *Training Procedure*

The goal was to design a robust classification framework that could effectively detect the main 'urban fabric' land cover classes and therefore the developed methodology is heavily based on a comprehensive training procedure. A deep learning, patch-based approach was employed for the classification of very high resolution images in five different classes. In particular, after extensive and laborious photo-interpretation, a significant amount of ground truth data was collected. The reference/ground truth data was collected for the classes *building, building shadow, vegetation, vegetation shadow* and *soil*. The selected classes represent the main per city-block land cover categories (*(i)* impervious surfaces: *building* and *building shadow*, *(ii)* urban green: *vegetation* and *vegetation shadow*, *(iii)* soil: *soil*) and they were adequate for the estimation of impervious surfaces, soil sealing and urban green.

For the training procedure a patch of 29x29 size has been selected and fed into a DL network. Using the available ground truth data randomly selected patches have been extracted and formed the training dataset. For the class *soil* a few samples were available therefore a data augmentation strategy with random samplings and random transformations has been used to add training patches. Using the google mosaic image for the training, we exploit only the RGB information for the classification.

Different CNN architectures exist in the literature with deep or shallow strategies. In our case, we selected a simple ConvNet network, with a shallow architecture both because of the relatively small size of the selected patch and the low computational complexity with high measured accuracies [9].

B. *Simple ConvNet Network*

A simple ConvNet Network consisting of 10 layers: 2 convolutional, 2 max pooling 3 transfer function and 3 fully connected has been used for the classification. More specifically, the raw input patch of size 29x29 is given as input to the first convolutional layer. Next comes a transfer function layer which applies the tanh function element-wise to the input tensor. It is followed by a max-pooling layer, which downsamples the training dataset. The next 3 layers follow the same pattern and result to the final fully connected layers that produce the final outputs.

Regarding the implementation, the model was trained with a learning rate of 1 for 36 epochs, while every 3 epochs the learning rate was reduced to half. The momentum was set to 0.9, the weight decay parameters to $5 \cdot 10^{-4}$ and the limit for the Threshold layer to 10^{-7}.

C. *Testing Procedure*

Regarding the testing procedure, a patch of 29x29 pixels was extracted for every pixel and was fed into the network. Then, each patch was classified into one of the five different classes.

III. EXPERIMENTAL RESULTS AND EVALUATION

For the validation of the developed methodology two urban test sites were selected *i.e.,* city of Athens and Bilbao. Both maps and very high resolution image mosaics ($\approx$ 1m ground resolution) have been acquired through google maps APIs and were used for the training and testing procedures. In particular, the employed RGB images cover an $7km^2$ area containing 254 city blocks. From these city blocks, 112 have been used for the training and 142 for testing the framework. Approximatively, 230.000 randomly selected patches have been used per class for the training.

Despite the important structural and spectral (due to the different dominating materials at each city) differences, a single training model was developed for both cities. In Table I the confusion matrix after the classification based on the simple ConvNet model is presented with an Overall Accuracy (OA) at 90.8% and with a Kappa coefficient at 71.6%. The relative lower accuracy rates reported for the classes *building shadow*, *vegetation shadow* and *soil* are mainly due to misclassification errors between the *building shadow* and *vegetation shadow* classes as well as between the *building* (mainly red, soil-like, building roofs) and the *soil* classes. However, the accuracy rates for the classes *building* and *vegetation* were higher than 80%.

In Figure 2 three different sub-regions from the Bilbao and Athens test sites are presented along with the different classification outputs and the corresponding ground truth. After a closer look, one can observe that the *building shadow* and *vegetation shadow* are the main classes that were mostly misclassified. This was most probably due to the absence of a near-infrared spectral band. The *building* and *vegetation* classes were detected with relative high accuracy rates.

(a) RGB image (b) Map (c) Ground truth (d) Proposed Approach

Fig. 2. The classification output (d) per city-block after the application of the developed method is presented. The corresponding google image mosaics (a) and maps (b) as well as the manually collected ground truth data (c) are presented for three different test sites in Athens and Bilbao.

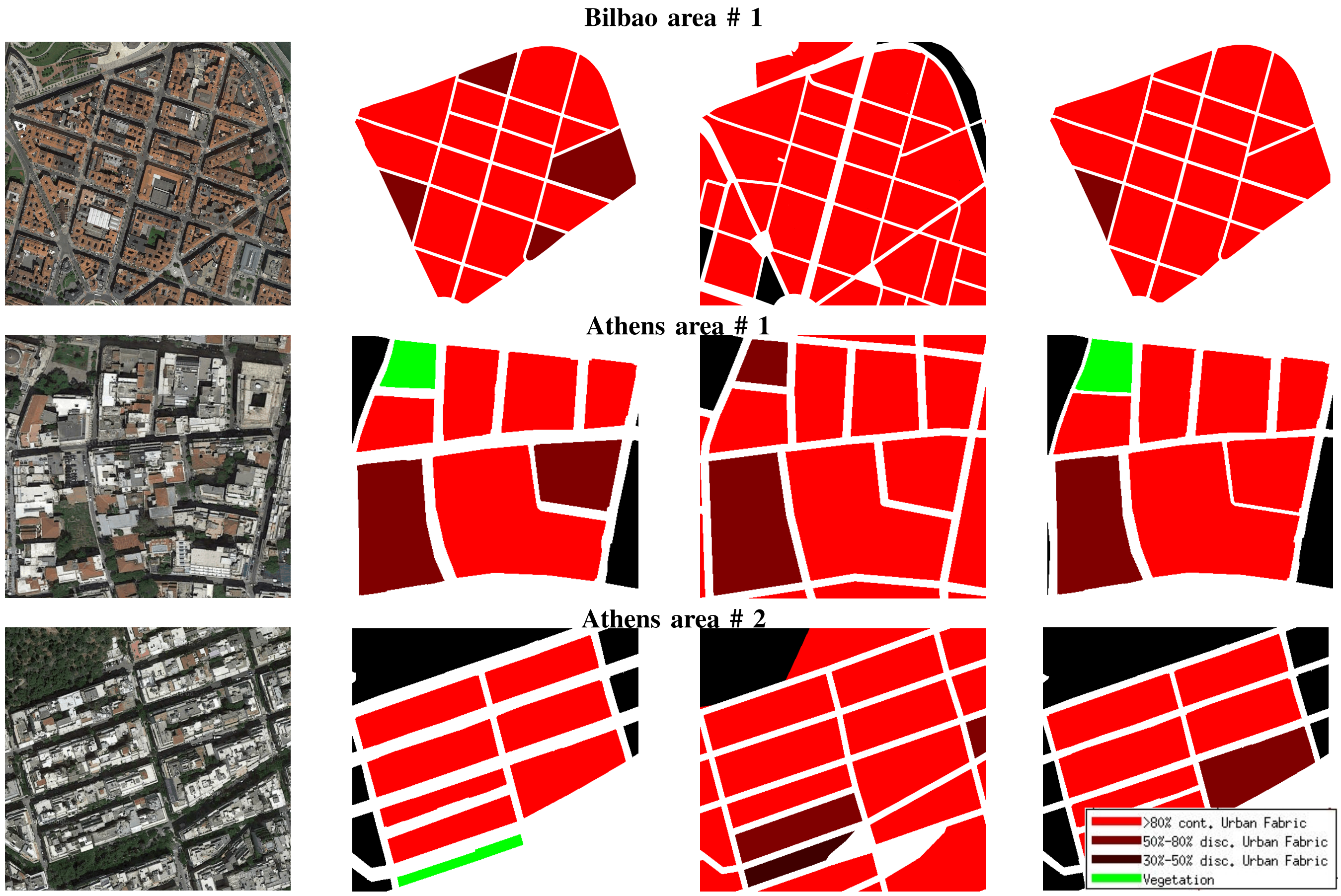
After the classification procedure, the classes were aggregated and based on the area that each one was covering at every city-block, its density was calculated. In particular, the density levels were calculated in accordance with the 'urban fabric' classes of Urban Atlas, namely the *(i)* 11100: *Continuous Urban Fabric (>80%)*, *(ii)* 11210: *Discontinuous Urban Fabric (50%-80%)* and *(iii) 11220: Discontinuous Urban Fabric (30%-50%)*. In Figure 3 the estimated per city-block densities are presented after the application of the proposed approach. The corresponding image, ground truth and Urban Atlas map are also presented for three test sites in Bilbao and Athens. There are certain differences between the ground truth, the Urban Atlas and the classification output per city-block. Even slightly, the developed method is more close with the densities derived from the ground truth data. Of course, one should take into account the different acquisition dates of the datasets employed for the Urban Atlas generation.

The quantitative evaluation between the ground truth and the developed methodology (Table II) indicated that the majority of city-blocks belonged to the *Continuous Urban Fabric (>80%)* class at both cities. The confusion matrix also indicated that the lowest accuracy rates were reported for the second class *Discontinuous Urban Fabric (50%-80%)*. In particular, 5 city blocks from the first class have been wrongly classified, mainly due to the fact that most free spaces were also covered from different artificial materials. Finally, it's worth mentioning that in both tested areas only three building blocks were allocated in the vegetation class and all was corrected classified.

Moreover, in Table III a quantitative evaluation between the ground truth and the Urban Atlas is presented. Urban Atlas was in accordance with the ground truth (>93%) as well as the proposed approach with the Urban Atlas (>94%) which implies that all products agree in most city-blocks.

	Reference Data				
# of blocks	> 80%	50%-80%	Vegetation	Total	UA (%)
Classification					
> 80%	**127**	5	0	132	96.2
50%-80%	2	**5**	0	7	71.4
Vegetation	0	0	**3**	3	100
Total	129	10	3	**142**	
PA (%)	98.5	50.0	100		
	Overall accuracy = 95.07% Kappa coefficient = 67.5%				

TABLE II. THE RESULTING CONFUSION MATRIX BETWEEN THE GROUND TRUTH AND THE PROPOSED APPROACH FOR THE ESTIMATED DENSITY PER CITY-BLOCK. THE OVERALL ACCURACY WAS 95.1%, WHILE THE KAPPA COEFFICIENT WAS 67.5%.



(a) RGB Image (b) Ground truth (c) Urban Atlas (d) Proposed Approach

Fig. 3. The estimated density per city-block (d) is presented after the application of the developed method. The corresponding map from Urban Atlas (c) as well as the manually collected ground truth data (b) and the RGB image mosaic (a) are also presented for three different test sites in Athens and Bilbao.

	Ground Truth vers. Urban Atlas			Proposed Appr. vers. Urban Atlas		
	Agree	Disagree	Total	Agree	Disagree	Total
Athens	54	9	63	55	8	63
Bilbao	78	1	79	79	0	79
Percentage	93.0%	7.0%	142	94.4%	5.6%	142

TABLE III. COMPARING THE DENSITIES FROM URBAN ATLAS AGAINST THE GROUND TRUTH AND THE PROPOSED APPROACH.

IV. CONCLUSION

In this paper, we proposed a supervised classification approach based on deep learning towards accurate density estimation at build up areas. In particular, for the training procedure we exploit information both from maps (*e.g.,* open street or google maps) as well as very high resolution RGB image mosaics. The employed patch-based, deep-learning model was trained against five land cover classes. During the prediction phase the per city-block classification procedure delivered the locations and percentages of impervious, soil and green regions. Experimental results and validation from two European cities *i.e.,* Athens and Bilbao, indicated overall accuracy rates at 95%. Results, also, highly match the corresponding ones from the Copernicus Urban Atlas product.

REFERENCES

[1] P. Gamba, "Human settlements: A global challenge for eo data processing and interpretation," *Proceedings of the IEEE*, vol. 101, no. 3, pp. 570–581, March 2013.

[2] M. Pesaresi, G. Huadong, X. Blaes, D. Ehrlich, S. Ferri, L. Gueguen, M. Halkia, M. Kauffmann, T. Kemper, L. Lu, M. A. Marin-Herrera, G. K. Ouzounis, M. Scavazzon, P. Soille, V. Syrris, and L. Zanchetta, "A global human settlement layer from optical hr/vhr rs data: Concept and first results," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 6, no. 5, pp. 2102–2131, 2013.

[3] P. Gamba, "Image and data fusion in remote sensing of urban areas: status issues and research trends," *International Journal of Image and Data Fusion*, vol. 5, no. 1, pp. 2–12, 2014.

[4] K. Karantzalos and D. Argialas, "A Region-Based Level Set Segmentation for Automatic Detection of Man-Made Objects from Aerial and Satellite Images," *Photogrammetric Engineering and Remote Sensing*, vol. 75, no. 6, pp. 667–678, 2009.

[5] K. Karantzalos and N. Paragios, "Large-Scale Building Reconstruction Through Information Fusion and 3-D Priors," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 48, no. 5, pp. 2283–2296, 2010.

[6] Q. Weng, "Remote sensing of impervious surfaces in the urban areas: Requirements, methods, and trends," *Remote Sensing of Environment*, vol. 117, pp. 34 – 49, 2012.

[7] I. Fernndez, F. J. Aguilar, M. A. Aguilar, and M. F. lvarez, "Influence of data source and training size on impervious surface areas classification using vhr satellite and aerial imagery through an object-based approach," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 7, no. 12, pp. 4681–4691, 2014.

[8] "Urban atlas," http://www.eea.europa.eu/data-and-maps/data/urban-atlas, accessed: 2016-09-15.

[9] M. Papadomanolaki, M. Vakalopoulou, S. Zagoruyko, and K. Karantzalos, "Benchmarking deep learning frameworks for the classification of very high resolution satellite multispectral data," *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. III-7, pp. 83–88, 2016.

[10] 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)*, July 2015, pp. 1873–1876.