Convolutional Neural Network for Satellite Imagery

Vladimir Khryashchev, Vladimir Pavlov, Andrey Priorov, Evgeniya Kazina
P.G. Demidov Yaroslavl State University
Yaroslavl, Russian Federation
v.khryashchev@uniyar.ac.ru, i@yajon.ru, andcat@yandex.ru, kazinaevgeniya@gmail.com

***Abstract*—Information extracted from aerial photographs has found applications in different areas including urban planning, crop and forest management, disaster relief, and climate modeling. In many cases information extraction is still performed by human experts, making the process slow, costly, and error prone. The goal of this investigation is to develop methods for automatically extracting the locations of objects such as water resource, forest and urban areas from aerial images. We analyze patterns in land using large-scale satellite imagery data which is available worldwide from third-party providers. For training, given the limited availability of standard benchmarks for remote-sensing data, we obtain ground truth land use class labels carefully sampled from open-source surveys, in particular the Urban Atlas land classification dataset of 20 land use classes across 300 European cities. The developed algorithms are based on the implementation of a relatively new approach in the field of deep machine learning - a convolutional neural network. We show how deep neural networks implemented on modern GPUs can be used to efficiently learn highly discriminative image features.**

I. Introduction

Examining aerial imagery allows identifying objects and determining various properties of the identified objects. Aerial image interpretation find applications in many diverse areas including urban planning, crop and forest management, disaster relief, and climate modeling. Much of the work, however, is still performed by human experts, and only a few semi-automated systems that work in limited domains are in use today and no fully automated systems currently exist [1], [2].

To date, databases containing high-resolution images (about 100 pixels per square meter) have been created. This fact underscores the need for automated aerial image interpretation methods. Recent applications of large-scale machine learning to such high-resolution imagery have produced object detectors with impressive levels of accuracy [3]-[5], suggesting that automated aerial image interpretation systems may be within reach.

In machine learning applications, aerial image interpretation is usually formulated as a pixel labeling task. Given an aerial image like the one shown in Fig. 1, the goal is to produce either a complete semantic segmentation of the image into classes such as building, road, tree, grass, and water [3], [4] or a binary classification of the image for a single object class [5]-[7].

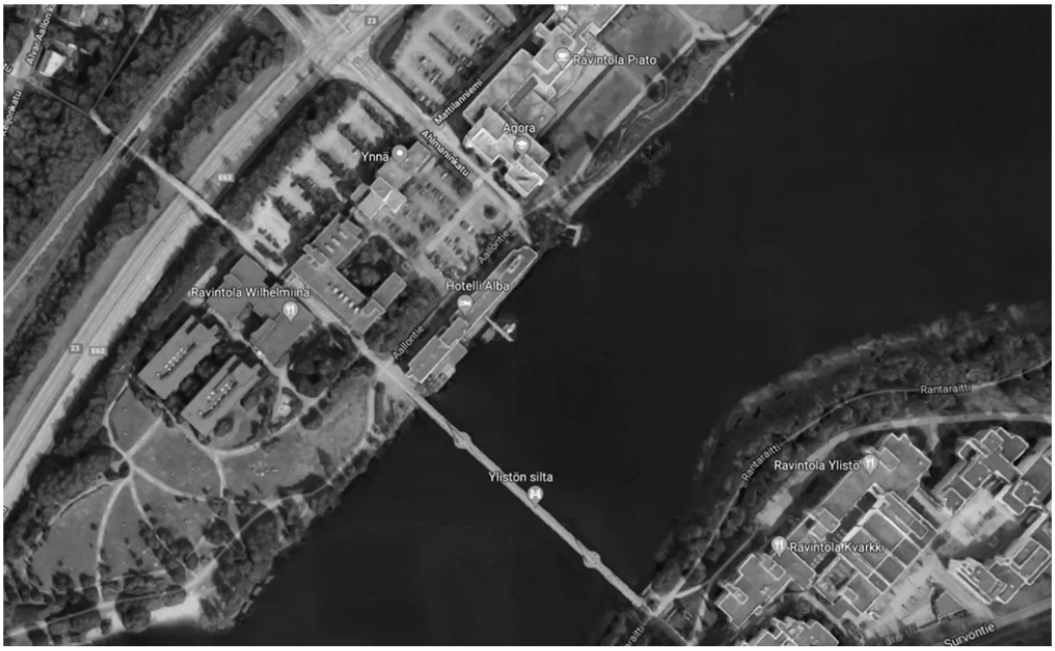

Fig 1. Marked aerial image of University of Jyvaskyla campus

Object detection is a common task in computer vision, and refers to the determination of the presence or absence of specific features in image data. Once features are detected, an object can be further classified as belonging to one of a pre-defined set of classes. This latter operation is known as object classification. Object detection and classification are fundamental building blocks of artificial intelligence. A major challenge with the integration of artificial intelligence and machine learning in aerial image analysis is that these tasks are not executable in real-time or near-real-time due to the complexities of these tasks and their computational costs. One of the proposed solutions is the implementation of a deep learning-based software which uses a convolutional neural network algorithm to track, detect, and classify objects from raw data in real time. In the last few years, deep convolutional neural networks have shown to be a reliable approach for image object detection and classification due to their relatively high accuracy and speed [8]–[11].

Convolutional neural networks have become ubiquitous in computer vision ever since AlexNet [12] popularized deep convolutional neural networks by winning the ImageNet Challenge: ILSVRC 2012 [13]. However, these advances to improve accuracy are not necessarily making networks more efficient with respect to size and speed. In many real world applications such as robotics, self-driving car and augmented reality, the recognition tasks need to be carried out in a timely fashion on a computationally limited platform.Furthermore, a CNN algorithm enables to convert object information from the immediate environment into abstract information that can be interpreted by machines without human interference.

II. Learning of convolutional neural network

Our work is devoted to the analysis of the use of convolutional neural networks for detecting earth surface types using remote sensing data. For deep learning of convolutional neural networks we used the marked image database UrbanAtlas [14]. Urban Atlas contains images of 21 classes. Images obtained from the Landsat-8 satellites [15] are used for estimation of automatic object detection quality. Examples of images from Landsat-8 satellites are shown on Fig. 2.

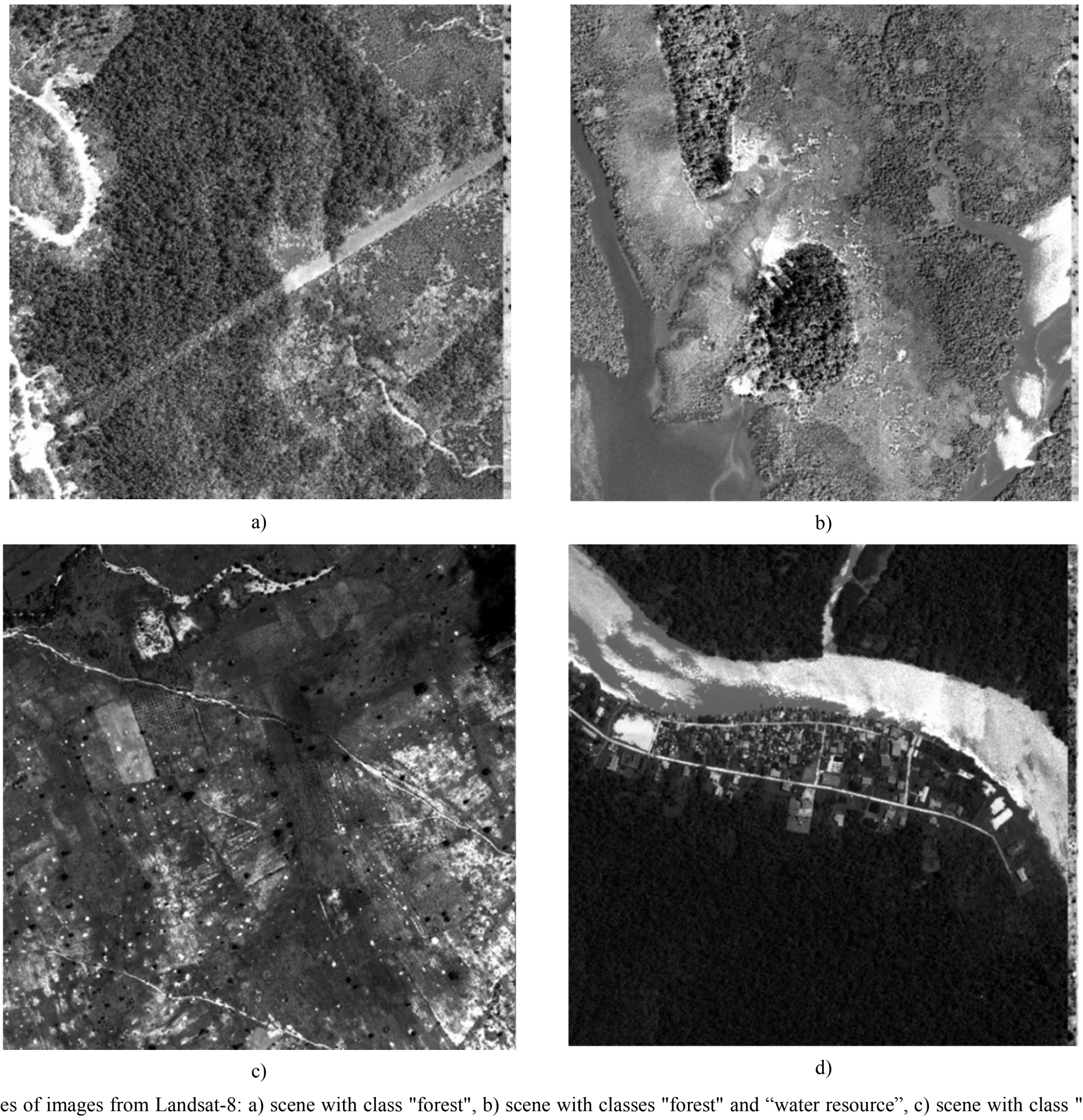

Fig. 2. Examples of images from Landsat-8: a) scene with class "forest", b) scene with classes "forest" and “water resource”, c) scene with class "agriculture", d) scene with classes "forest", “urban area” and “water resource”

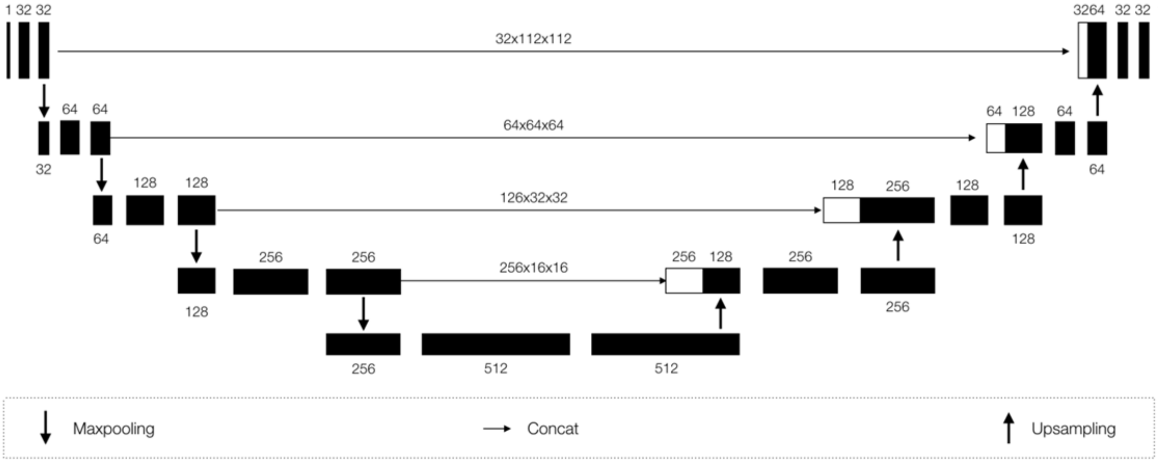

Fig. 3. U-NET network architecture

In this research we used the U-NET architecture of a convolutional neural network presented on Fig. 3. The network consists of two parts: a scraping conveyor (on the left) and an expanding network (on the right), which are represented by 23 convolutional layers.

The convolution conveyor operation is a sequential execution of a convolution layer (3x3) followed by a 2x2 linear rectification unit (ReLU) to clamp the negative part of the scalar value, followed by the operation of maxpooling 2x2 layers in step 2 for downsampling. At each sampling step, we double the number of channels of the output function.

Expanding network at each step performs a two-fold up-convolution, followed by a 2x2 convolution layer that reduces the number of function channels, followed by 2 3x3 convolutions and a 2x2 linear rectification block. Cropping is necessary at each stage because of the loss of pixels on the border after each stage of convolution. 1x1 convolution layer is performed to match each 64-component output vector to the classification classes at the last stage.

This architecture is allowed to reduce the spatial resolution of the image at the initial stage and then increase it preliminarily combining it with the image data and passing through other convolution layers. The network serves as a kind of filter.

The testing was carried out on the supercomputer NVIDIA DGX-1 in the Center for Artificial Intelligence of the Yaroslavl University named by. P.G. Demidov. To accelerate network operation, the training and testing processes of the convolutional neural network were carried out in parallel, on a large number of independent streams of the graphic processor of the video card.

III. Experimental results

To analyze the accuracy of the object detection algorithm, we compared the contours of automatically detected areas with areas of expert markup. To analyze the accuracy of the object detection algorithm, the selected regions were compared with the areas by previously marked by experts. For the experiment we selected 100 space images where there is an interest class. Then the procedure of automatic detection of the object in the image was carried out. Next we made detection of objects by our detector on neural networks. The resulting multipoligons we compared to the percentage of intersection.

$$R = \frac{S_{pred} \cap S_{exp}}{S_{exp}},$$

where $S_{pres}-$ detected multipolygon, $S_{exp}-$ expert multipoligon.

Examples of images with detected and expert multipolygon is shown on Fig. 4. A white solid line identifies the contour, drawn by the detector, a black solid line is a multipoligon, drawn by an expert.

The results of the experiment are given at Table I. In this table **total objects** – is the total number of expert markings of multipoligons of this class and **detected objects** – is the number of multipolygons that were detected by the detector as a result of satellite image analysis.

TABLE I. Object detection results

	Total objects	**Detected objects**	**False detection**	**Average percentage of intersection**
Forest	200	168	21	92.3%
Water	40	145	47	81.7%
Agriculture	220	208	11	96.1%

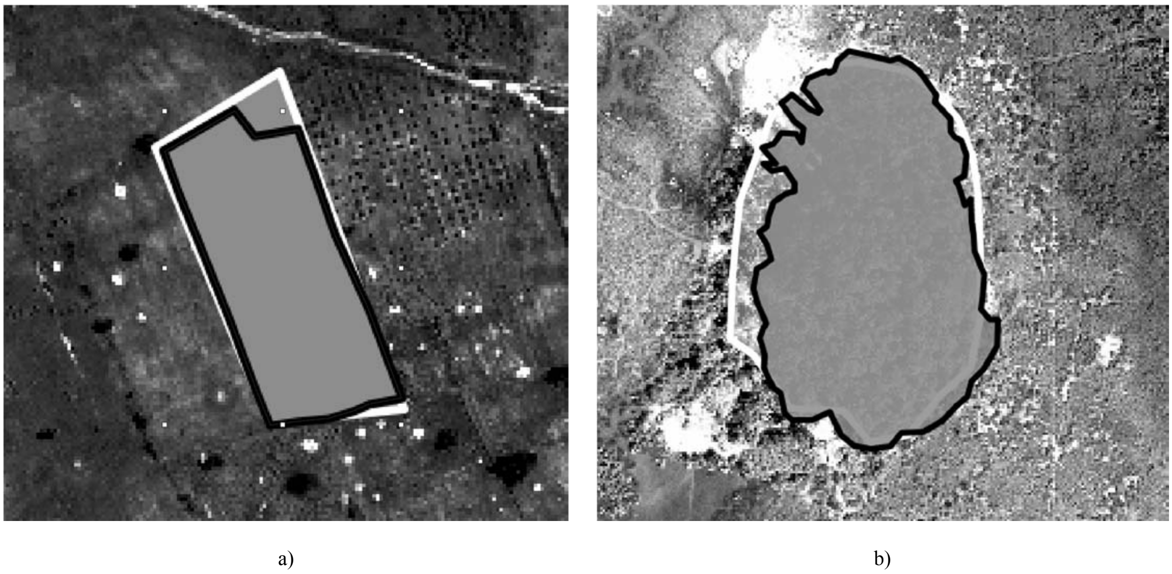

Fig. 4. Images from Landsat-8 with automatically (white curve) and manual (black curve) marked region: a) "agriculture", b) "forest"

As can be seen from Table I, the highest percentage of intersections of areas of detected objects of the "Agriculture" class. This is due to the clarity of the boundaries and the apparent visual separation of the surrounding objects. When we detect “Water”, we have a lot separate parts of the water resource are allocated. This is due to the presence of ice and other objects over rivers and lakes. Forest territory has an average detection value of 92.3% of the intersection of areas due to inaccurate allocation of boundaries of forest territory.

IV. Conclusions

In this paper was presented a research about using convolutional neural networks for detection geo-objects on the satellite images from Landsat-8. We use U-NET convolutional neural network architecture for implementing the algorithm of computer vision. The neural network we trained by the marked image base "Urban Atlas". The final classification accuracy is 81.7% for objects such as "water resource", 92.3% for objects of the "forest" class and 96.1% for objects of the "agriculture" class. The considered algorithm can be applied for the semantic analysis of images from the satellite: allocation of the territories of cities, control of construction and other.

References

[1] E. P. Baltsavias, “Object extraction and revision by image analysis using existing geodata and knowledge: current status and steps towards operational systems”, *ISPRS Journal of Photogrammetry and Remote Sensing*, 58(3-4): 129-151, January, 2004.

[2] H. Mayer, “Object extraction in photogrammetric computer vision”, *ISPRS Journal of Photogrammetry and Remote Sensing*, 63(2):213-222, March 2008.

[3] S. Kluckner, and H. Bischof, “Semantic classication by covariance descriptors within a randomized forest”, *In Computer Vision Workshops (ICCV)*, pages 665-672. IEEE, 2009.

[4] S. Kluckner, T. Mauthner, P. M. Roth, and H. Bischof, “Semantic classication in aerial imagery by integrating appearance and height information”, *In ACCV, volume 5995 of Lecture Notes in Computer Science*, pages 477-488. Springer, 2009.

[5] V. Mnih, and G. Hinton, “Learning to detect roads in high-resolution aerial images”, *In Proceedings of the 11th European Conference on Computer Vision (ECCV)*, September 2010.

[6] V. Mnih, and G. Hinton, “Learning to label aerial images from noisy data”, In Andrew McCallum and Sam Roweis, editors, Proceedings of the 29th Annual International Conference on Machine Learning (ICML 2012), June 2012.

[7] P. Dollar, Z. Tu, and S. Belongie, “Supervised learning of edges and object boundaries”, *In Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 1964{1971, 2006.

[8] J. Sherrah, “Fully Convolutional Networks for Dense Semantic Labelling of High-Resolution Aerial Imagery”, Web: https://arxiv.org/abs/1606.02585.

[9] T. Qu, Q. Zhang, and S. Sun, “Vehicle detection from high-resolution aerial images using spatial pyramid pooling-based deep convolutional neural networks”, *Multimed. Tools Appl.* October 2017, Volume 76, Issue 20, pp 21651–21663.

[10] K. He, X. Zhang, S. Ren, and J. Sun, “Spatial pyramid pooling in deep convolutional networks for visual recognition”, *IEEE Trans. Pattern Anal. Mach. Intell.* 2015, 37, 1904–1916.

[11] Q. Wang, C. Rasmussen, and C. Song, “Fast, Deep Detection and Tracking of Birds and Nests”. *In Proceedings of the International Symposium on Visual Computing*, Las Vegas, NV, USA, 12–14 December 2016; pp. 146–155.

[12] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks”, *In Advances in neural information processing systems*, pages 1097–1105, 2012.

[13] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, et al., “Imagenet large scale visual recognition challenge”, *International Journal of Computer Vision*, 115(3):211–252, 2015.

[14] “European Union. 2011. Urban Atlas”, Web: https://www.eea.europa.eu/data-and-maps/data/urban-atlas.

[15] “Landsat8”, Web: https://en.wikipedia.org/wiki/Landsat_8.