Application of Satellite Image Segmentation for Urban Planning Optimization

Vladimir Khryashchev [1+], Leonid Ivanovsky [2], Anna Ostrovskaya [1], Alexander Semenov [1]

[1] People's Friendship University of Russia (RUDN University)

[2] P.G. Demidov Yaroslavl State University

Abstract. This article presents research results of a convolution neural network for building detection on high-resolution aerial images of Planet database. Jaccard index was used for analysis of the quality of machine learning algorithm. This index of similarity compares results of algorithms with real masks. The masks were sliced on smaller parts together with images before training of developed model. The convolution neural network was launched on NVIDIA DGX-1 supercomputer, which was provided by AI-center of P.G Demidov Yaroslavl State University. The problem of building detection on satellite images can be put into practice for urban planning, building control, search of the best locations for outlets etc.

Keywords: machine learning, aerial image segmentation, building detection.

1. Introduction

Currently, the task of object detection on high-resolution satellite images is in the focus of researchers. Automatic image segmentation allows to extract areas of interest on aerial photos, such as buildings, roads or vehicles. It is suggested deep learning algorithms are able to solve this problem.

The task of image segmentation is usually reformulated as a classification on a pixel-wise level. The simplest and the slowest way to solve this problem is a manual segmentation of satellite images, but it is time-consuming and tedious process which is subjected to human errors [1]. Therefore, the great interest of researchers aims at automatic image segmentation. It makes possible to process satellite images immediately after receiving it. Satellite image segmentation finds its application in business, urban planning, building control, forest management, meteorology and land-use agriculture.

In this article we present a developed convolutional neural network (CNN). The main advantage of CNNs is that they can detect objects in real time while being computationally less expensive and superior in performance at the side of traditional machine learning algorithms. Features in these networks are formed automatically in the process of training. The structure of CNNs is parallel and fit to the architecture of graphics processing units (GPUs) which contains thousands of cores to perform many similar at the same time. Nowadays in tasks of computer vision CNNs excel even some experts in quality [2].

One of the most successful algorithms of image segmentation is based on fully convolutional networks (FCNs). The basic idea of FCNs is the usage of fully connected layer with a convolution layer at end, while other layers extract necessary features from input data [3]. FCN shows perfect results of object detection on the dataset of PASCAL VOC 2011 challenge [4].

Soon FCNs have been improved to U-Net, which had received a lot of interest for segmentation of biomedical images [5]. Later this model showed acceptable results of object detection on satellite images. U-

---

[+] Corresponding author.
E-mail address: vhr@yandex.ru.

Net uses skip-connections and combinations of low-level and higher-level feature maps. Using U-Net the authors of paper [6] get the value of Sorensen-Dice coefficient is equal to 0.75 for building detection. In paper [7] U-Net shows good results for detection of trees, crops and waterway: the value of Sorensen-Dice coefficient is equal to 0.75 too.

This article consists of six parts. The first part introduces readers to the problem of satellite image segmentation. It also contains an overview of certain papers for object detection on aerial photos. The second part is devoted to some available databases of satellite images. The third section describes the developed CNN and peculiarities of training of models. Moreover, in this part there is mentioned about Keras framework for the designing of deep learning algorithms. The forth part presents results of numerical experiment for the developed model. In the conclusion there is summarized the research. And finally, the last section represents references.

2. Databases of High-Resolution Aerial Images

The Pleiades-1B database [8] contains colour images from Pleiades-1B satellite. This satellite was successfully launched on the 2th of December in 2012 for mapping large scale areas including engineering and construction projects, monitoring of mining and industrial complexes, natural hazards and rescue operations. This dataset is notable for different angles of shooting. Each image of Pleiades-1B database has a spatial resolution of 0.5 m / pixel. Examples of images from the Pleiades-1B database are shown in Fig. 1.
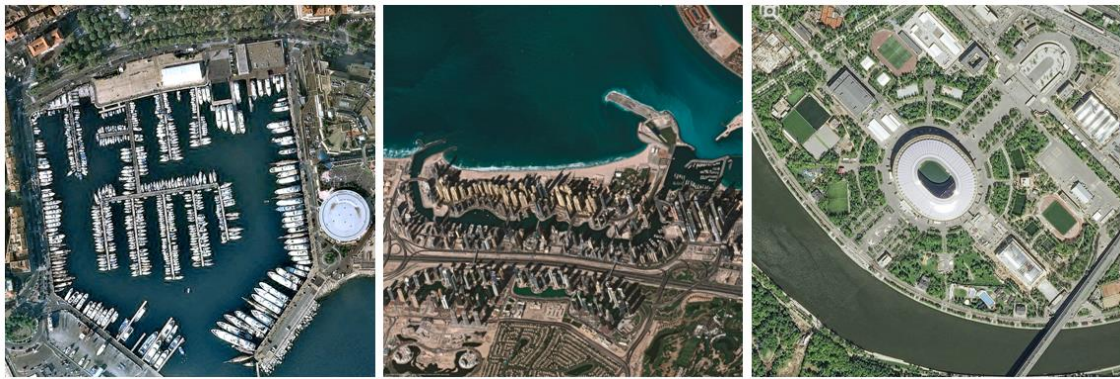

Fig. 1: Examples of images from the Pleiades-1B database.

In our research there were used 14 colour satellite images from the Planet database. Every image of this dataset has a resolution of 16384×16384 pixels with a spatial resolution of 0.5 m / pixel. Images of private database cover the area of 3 Russian cities: Yaroslavl, Rybinsk and Moscow. Examples of images from the Planet database are shown in Fig. 2.

Fig. 2: Examples of images from the Planet database.

3. Deep learning Algorithm

3.1. Convolutional neural network

This article presents a developed CNN - U-Net. This type of neural networks has a special architecture, aimed at fast and high-quality detection of various objects [1]. The research of work of developed models continues the research, which was provided in papers [7, 9].

Keras library with Tensorflow framework as a backend was used for development of CNN. Keras is an open-source library written in Python. It is built on Tensorflow framework and contains various implementations of commonly used neural network building blocks, such as layers, activation functions and optimizers, and ready tools to pre-process images and text data. In other words, Keras offers a higher-level, more intuitive set of abstractions to develop deep learning models [10]. In addition, this library allows to train developed models on GPU.

As it is said in [7], U-Net consists of 2 parts: an encoder (left) and a decoder (right). The encoder is a neural network which has a typical CNN architecture. The decoder of U-Net has a similar architecture as the encoder, except merging operations and a maxpooling operation, which was replaced with an upsampling layer. The last convolutional layer of U-Net with 1×1 convolution window classifies every pixel. In summary, U-Net has about 2 million trainable parameters.

3.2. Peculiarities of training

CNN require considerable computational resources, as a consequence training and test stages were implemented on a large number of independent streams of GPU by means of parallel computing technology NVIDIA CUDA. This cross-platform technology is supported by all modern NVIDIA graphics cards [11]. As an algorithm of numerical optimization Adam optimizer was chosen. This optimizer combines the best approaches from gradient descent and momentum optimizers and shows optimal and quick convergence for most tasks of machine learning [12]. Binary cross-entropy was chosen as a loss function. The training finishes after completing 100 epochs, while on each training step of epoch a batch of 36 samples was passed through the developed model.

Numerical experiments for developed algorithms were performed on satellite images of the Planet database. Information about the location of buildings was generated as black-and-white masks, where a pixel is coloured to white if it belongs to buildings. Satellite image segmentation concerns the usage of parts of aerial photos, which are fed to the input of CNN. In our case U-Net require images of 512×512 pixels as an input data, so before the training of CNN each high-resolution photo and mask of dataset have been sliced on parts of appropriate size by data windowing. Examples of sliced images and masks are shown in Fig. 3. Every little part of sliced images corresponded to the needed small part of big generated mask.

Fig. 3: Sliced samples: prepared images and corresponding masks.

As a result of such transformations, prepared dataset contained 3264 images of 512×512 size. For numerical experiments the prepared dataset was shuffled and split on training and test set in ratio 80/20. Thus, the training set contained 2611 photos and the test set contained 653 photos. Train and test samples did not have same pictures. In our task, only 2 classes were taken into account: "buildings" and "not-buildings".

4. Numerical results

Developed CNN was launched on NVIDIA DGX-1 supercomputer, which was provided by AI-centre of P.G Demidov Yaroslavl State University.

As a rule, the quality of algorithms for image segmentation is evaluated by special coefficients for comparing the similarity of predicted and true masks. In our research, to estimate developed models there was used Jaccard index (J). This index is binary measure of similarity, possesses the value from $[0, 1]$ and can be calculated by the following formula:

$$J = \frac{|X \cap Y|}{|X \cup Y|},$$

where X is a set of real masks and Y are predictions [13-15]. Dependencies of values of Jaccard index from the number of epochs (E) for U-Net on test set are shown in Fig. 4.
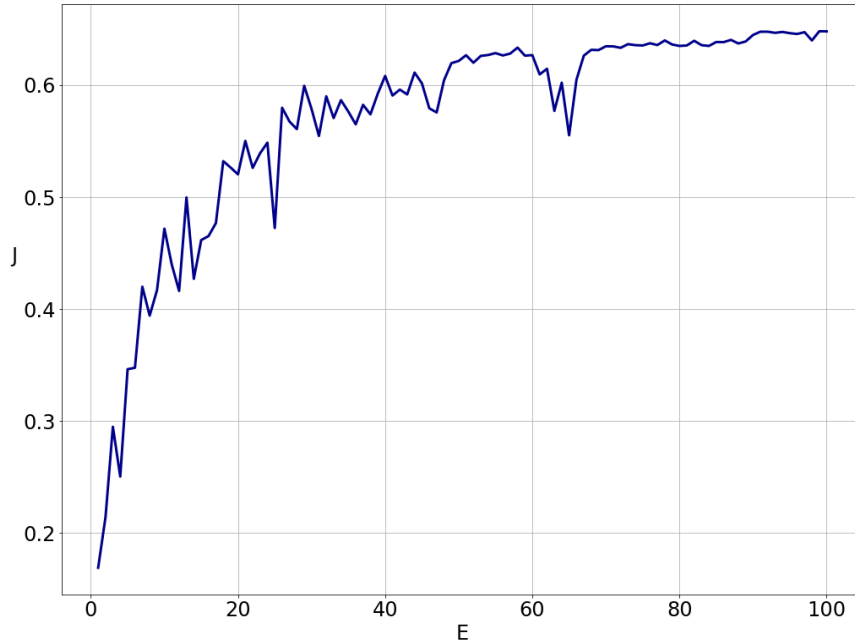
Fig. 4: Dependencies of Jaccard index on training epochs for developed convolutional neural networks.

According to the results presented in Fig. 4, U-Net shows acceptable results: the value of Jaccard index is equal to 0.65. The training of model lasted about 2 hours in spite of complexity of architecture of CNN.

One of applications of developed CNN might be a search of the best locations for outlets. No wonder U-Net can be used for building detection on high-resolution aerial photos. Therefore, it is possible to extract residential areas and estimate area population. Thus, we can calculate the approximate number of residents and, as a consequence, the number of possible customers when placing a store in a particular place. It allows to choose the best location for outlets.
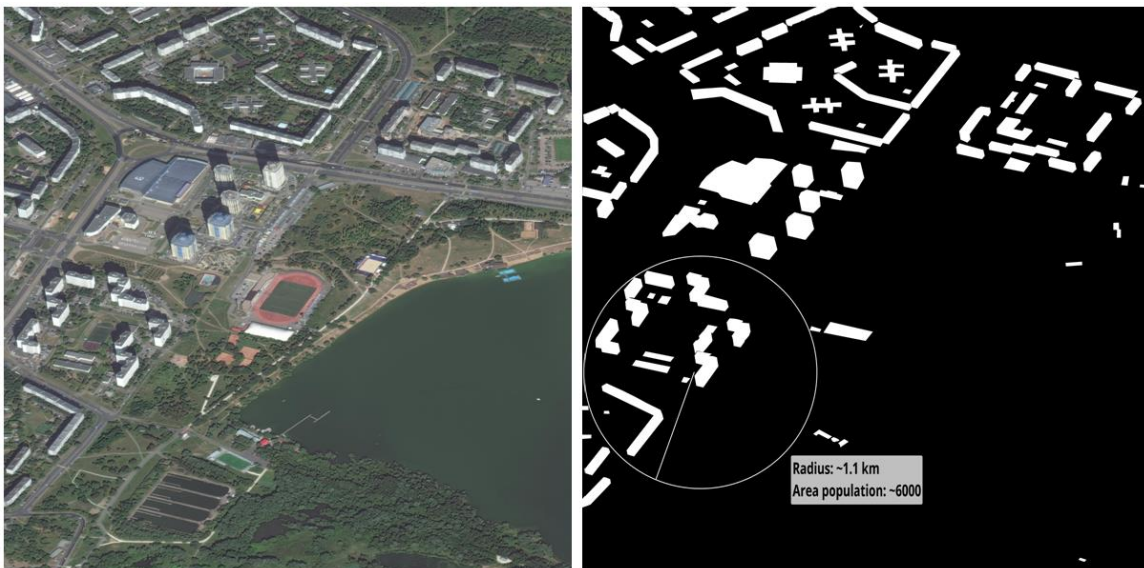

Fig. 5: Application of developed convolutional neural network for a search of the best locations for outlets.

5. Conclusion

This article presents research results of U-Net for building detection on high-resolution aerial images of Planet database. Before training of developed model masks were sliced on smaller parts together with images before training of developed model. The training of CNN lasted about 2 hours on GPU of supercomputer NVIDIA DGX-1 of AI-centre of P.G Demidov Yaroslavl State University. Jaccard index was used for analysis of the quality of machine learning algorithm. This index of similarity compares results of algorithms with real masks. For U-Net the value of Jaccard index is equal to 0.65. As a consequence, according to numerical results developed model can be used for building detection on aerial photos. The problem of building detection on satellite images can be put into practice for urban planning, building control, search of the best locations for outlets etc.

6. Acknowledgements

7. References

[1] Y. Goodfellow, Y. Bengio, A. Courville. *Deep Learning*. The MIT Press, 2016, 800 p.

[2] N. Shukla, *Machine Learning with Tensorflow*. Manning Publications, 2018, 272 p.

[3] E. Shelhamer, J. Long, T. *Darrellю Fully Convolutional Networks for Semantic Segmentationю* Web: https://arxiv.org/pdf/1605.06211.pdf.

[4] PASCAL VOC 2011 challenge. Web: http://host.robots.ox.ac.uk/pascal/VOC/voc2011/index.html.

[5] O. Ronneberger, P. Fischer, T. Brox. U-Net: Convolutional Networks for Biomedical Image Segmentation. *Medical Image Computing and Computer-Assisted Intervention (MICCAI), Springer, LNCS*, vol. 9351, 2015, pp. 234–341.

[6] G. Chhor, C. Bartolome Aramburu, I. Bougdal-Lambert. *Satellite Image Segmentation for Building Detection using U-net*. Web: http://cs229.stanford.edu/proj2017/final-reports/5243715.pdf.

[7] V. Khryashchev, L. Ivanovsky, V. Pavlov, A. Ostrovskaya and A. Rubtsov. Comparison of Different Convolutional Neural Network Architectures for Satellite Image Segmentation. *Proceedings of the 23rd Conference of Open Innovations Association FRUCT'23, Bologna, Italy*. 2018, pp. 172-179.

[8] *Pleiades-1B Satellite Sensor*. Web: https://www.satimagingcorp.com/satellite-sensors/pleiades-1b/.

[9] V. Khryashchev, A. Priorov, V. Pavlov, A. Ostrovskaya. Deep learning for region detection in high-resolution aerial images. *Proceedings of 16-th IEEE East-West Design & Test Symposium (EWDTS'2018), Kazan, Russia*, September 14 - 17, 2018. P. 792-796.

[10] A. Gulli, S. Pal. *Deep Learning with Keras*. Packt Publishing, 2017, 320 p.

[11] J. Sanders, E. Kandrot. *CUDA by Example: An Introduction to General-Purpose GPU Programming*. Addison-Wesley Professional, 2010, 320 p.

[12] D. P. Kingma, J. Ba. Adam. *A Method for Stochastic Optimization*. Web: https://arxiv.org/abs/1412.6980.

[13] J. Bell, H.M. Dee. *The subset-matched Jaccard index for evaluation of Segmentation for Plant Images*. Web: https://arxiv.org/pdf/1611.06880.pdf.

[14] J. VanderPlas. *Python Data Science Handbook: Essential Tools for Working with Data*. O'Reilly Media, 2016, 541 p.

[15] A. Muller, S. Guido. *Introduction to Machine Learning with Python: A Guide for Data Scientists.* O'Reilly, 2016, 400 p