Deep learning algorithm for  
mine detection on satellite images

Vladimir Khryashchev  
P.G. Demidov Yaroslavl State UniversityYaroslavl, Russia  
v.khryashchev@uniyar.ac.ru

Anna Ostrovskaya  
People’s Friendship University of RussiaMoscow, Russia  
ostrovskaya\_aa@rudn.university

Vladimir Pavlov  
P.G. Demidov Yaroslavl State UniversityYaroslavl, Russia  
i@yahon.ru

Alexander Semenov  
People’s Friendship University of RussiaMoscow, Russia  
semenov.venture@mail.ru

Abstract — The aim of our research was to create a deep learning algorithm for automated mine detection on high-resolution aerial images. Dice coefficient, which compares results of algorithms with real masks, was used to measure the quality of developed model. On the whole this paper demonstrates how convolutional neural network implemented on modern GPUs can be applied for the detection of mines on satellite images.

Keywords — computer vision, image segmentation, high-resolution photos, mines.

# Introduction

Currently, object detection task is in the focus of modern researchers. Automatic image segmentation allows to detect areas of interest, such as buildings, woods, angicultural fields, etc. Most approaches of solving this problem propose the usage of deep learning algorithms [1].

The progress in the development of high-performance computers with graphics processing units (GPU) allowed researchers to create and launch convolutional neural networks (CNNs) that have millions of parameters. To solve solving modern problems of computer vision, such deep learning algorithms exceed traditional methods and experts in some cases. CNNs demonstrated their advantage in tasks of object detection and image segmentation [2].

The task of image segmentation is usually reformulated as classification at the pixel level. The slowest approach to solve this problem is manual image segmentation. Nonetheless, this process is time-consuming and prone to mistakes. Automatic segmentation allows to process an image after it has been received [3].

This paper presents developed convolutional neural network for mine detection on satellite images. The main advantage of such networks is that image descriptors are formed during the training process. CNNs have become ubiquitous in computer vision since AlexNet [4] won the ImageNet Challenge: ILSVRC 2012 [5].

One of the most successful algorithms for image segmentation is fully convolutional networ (FCN). The basic idea of FCN is the usage of fully connected layer with a convolution layer in the end of model, while other previous layers extract necessary features from input images [6].

The authors of paper [7] present U-Net architecture - a special type of FCNs, which was developed to detect areas of interest in biomedical images. Later this model was applied to satellite images [8]. The U-Net architecture has skipped connections to combine features obtained at different stages of the network. In article [9] U-Net allows to get a high value of Dice coefficient for the task of building detection: 0.75.

This article consists of six parts. The first part is devoted to the usage of convolutional neural networks. It also provides an overview of articles related to the task of object detection on high-resolution aerial images. The second part pays attention on some available databasets of satellite images. The third part describes the architecture of developed convolutional neural network and some features of learning process. The fourth part presents the results of numerical experiments of deep learning algorithm. In the conclusion there is summarized the research. And finally, the last sections represents references.

# Overview of aerial image datasets

A satellite is a mechanical object that periodically rotates on the Earth orbit to performs certain functions and tasks for television, military intelligence, satellite communications, meteorological observations, etc. At present, there are many satellites which are capable of making high-resolution images of the Earth surface. Table 1 presents some popular satellites.

TABLE I. Some popular satellites

|  |  |  |
| --- | --- | --- |
| **Model of satellite** | **Launch year** | **Price of images** |
| Landsat 8 | 2013 | Free |
| Sentinel | 2014 | Free |
| MODIS | 2002 | Free |
| WorldView | 2016 | Paid |
| QuickBird | 2001 | Paid |
| GeoEye-1 | 2008 | Paid |
| IKONOS | 1999 | Paid |
| Jilin-1 | 2015 | Paid |
| SPOT-6 | 2012 | Paid |
| Gaofen-2 | 2014 | Paid |
| TripleSat | 2015 | Paid |

The features of obtaining satellite images of the Earth include the following peculiarities:

* Satellites simultaneously take photos of the same territory. The first image is black and white, the most detailed. The second one is color, with a resolution below average. It is impossible to take high-resolution color photos, because the light is refracted in the Earth's atmosphere. In addition, it is needed to stretch the color snapshot, because in digital form it turns out less black and white image. Finally both images are combined.
* Space cameras identify colors differently, so the original satellite images do not look like natural photos due to diffraction and scattering in the Earth’s atmosphere. In order to make colors normal for human perception, color correction is essential to be performed.
* Shooting conditions and camera type cause a shifting effect. In this case it is needed to implement eliminating distortions and transforming the original image into its orthogonal projection.
* As a rule, information from the satellite comes in a form of matrix of radiation intensity distribution.

An image dataset is the important part of training and efficiency evaluation of different machine learning algorithms. Nowadays, there were collected some datasets of satellite images.

The Jilin-1 database [10] contains high-resolution panchromatic and multichannel color images from the satellite of the same name, which was launched on October 7, 2015 from Jiuquan Satellite Launch Base in China. Every panchromatic and color image of Jilin-1 dataset has a spatial resolution of 0.72 m/pixel and 2.88 m/pixel respectively. High-resolution aerial photos from Jilin-1 satellite apply for various mapping applications such as environmental monitoring, forestry, energy, land development and agriculture. Examples of images from the Jilin-1 dataset are shown in Fig. 1.

The SPOT-6 database [11] includes aerial panchromatic and four-channel (blue, green, red, near-IR) images. SPOT-6 optical satellite was built by AIRBUS and successfully launched on September 9, 2012 from the Satish Dhawan Space Center in India. This satellite is capable of imaging the Earth with a resolution of 1.5 m panchromatic and 6 m multispectral. It offers imaging products to customers in defense, agriculture, deforestation, environmental monitoring, coastal surveillance, engineering, oil, gas and mining industries. Examples of images from the SPOT-6 dataset are shown in Fig. 2.

In our research there were used 23 four-channel (blue, green, red, near-IR) satellite images from the Planet database. Every aerial photo of this dataset has a resolution of 8000 × 8000 pixels with a spatial resolution from 1 to 2 m / pixel. Mine sizes in the Planet database are varied from 50x50 pixels to 1200x1200 pixels.

# C:\Users\User\Downloads\hurricane-harvey-houston-texas.jpg

1. Examples of images from the Jilin-1 database

# C:\Users\User\Downloads\hurricane-harvey-houston-texas.jpg

1. Examples of images from the SPOT-6 database

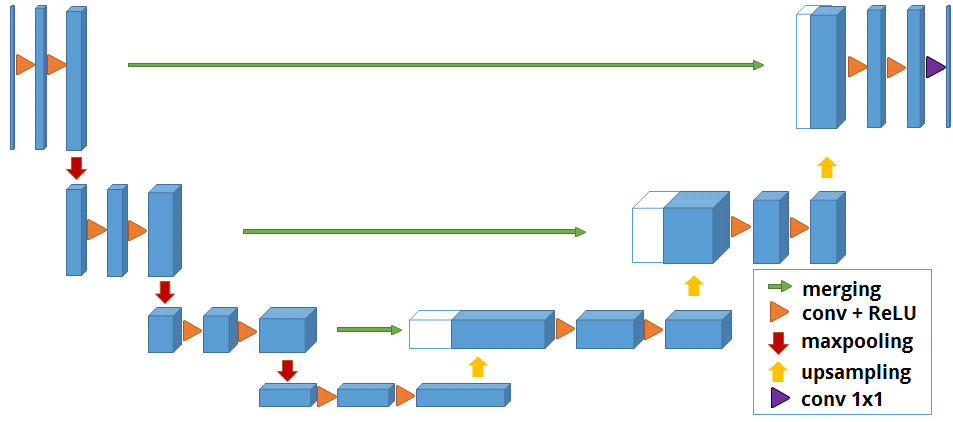
# Neural Network Architecture

This article presents developed deep learning algorithm based on CNN. According to brand new results of scientific researches, this type of neural networks aims at quick and high-quality image segmentation [12].

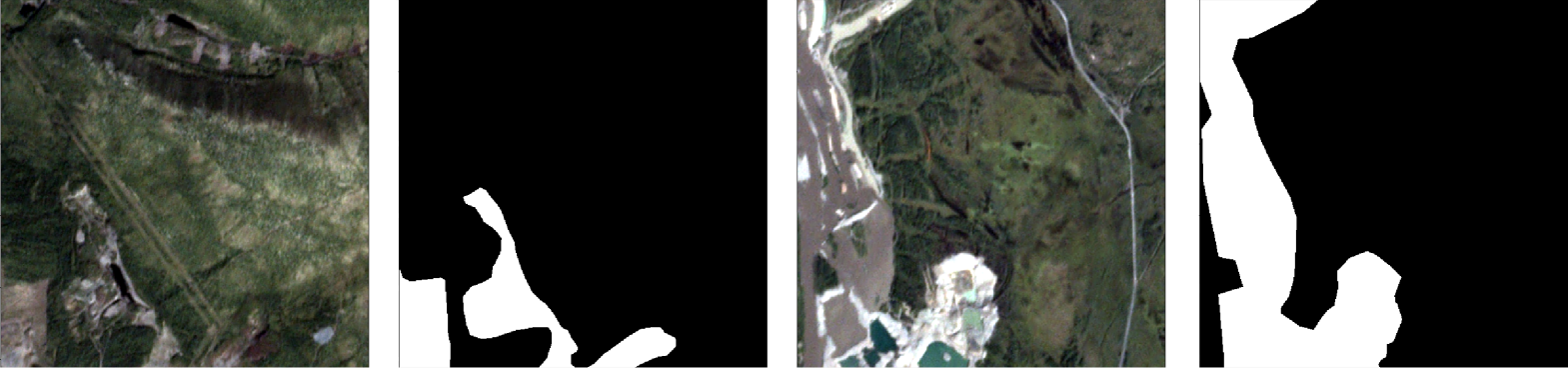
In our research there was used a modification of the well-known CNN architecture: U‑Net [8]. Original U-Net consists of 2 parts: encoder and decoder. The architecture of original U-Net is shown in Fig. 3.

The encoder is a typical CNN which includes four blocks. Every encoder block consists of two convolutional layers with 3 × 3 filter, ReLU activation function applied to each of them and a maxpooling operation with 2 × 2 filter and stride 2. Unlike original U-Net presented in article [9], in modified U-Net there were used two encoder for RGB and NIR channels of input satellite image respectively. Feature maps from such encoders were united before decoder part. The decoder contains three blocks. Each decoder block contains an upsampling operation with 2 × 2 filter combining with a map of two corresponding encoder blocks for RGB and NIR channels, two convolutional layers with 3 × 3 filter and ReLU activation function applied to each of them. The last layer of the network is a convolution layer with 1 × 1 filter, which implements classification at pixel level.

The modified version of U-Net was developed by means of Tensoflow library. It is an open-source software library for high performance numerical computation which is also used for machine learning applications. Tensorflow contains numerous implementations of commonly used neural network building blocks and ready tools to preprocess images. Moreover, this library allows to train networks on GPU [13].



1. Original U-Net



1. Examples of sliced images and masks

# Numerical Results

Numerical experiments for developed deep learning algorithm were carried out on satellite images of the Planet database. Information about the location of mines was extracted from json files and generated as black-and-white masks, where white pixel belongs to mines.

Image processing is an integral part of Earth remote sensing, but, unfortunately, at the moment there are no fully automatic processing methods without human intervention.

The main shortage of satellite images of the Planet database was the deregulation of color and contrast histograms: some images looked very dark whereas other photos were light-colored. Therefore for each image of the Planet dataset there was implemented equalization of histogram for every color channel.

Modified U-Net requires images of 512 × 512 pixels, so before the training and validation of model, image and masks of dataset have been cut on parts of appropriate size by data windowing. Examples of sliced images and masks are shown in Fig. 4.

The quality of developed image segmentaion algorithm was evaluated by Dice coefficient (). This metric takes its values from the interval [0, 1] and calculates by the following formula:

 

where is a power of intersection and is a sum of powers for real mask and predictions [14]. In our task, numerator and denominator can be calculated by following formulae

 

 

where are normalized values of pixels for expert markup and predicted masks respectively.

Developed CNN was launched on NVIDIA DGX-1 supercomputer of AI-centre of P.G. Demidov Yaroslavl State University [15]. As the algorithm of numerical optimization Adam optimizer was choosen. This optimizer combines best approaches from gradient descent and momentum optimizers and shows optimal and fast convergence for most tasks of machine learning [16].

The deep learning algorithm was trained using modified binary cross-entropy () with Dice loss, which calculating by the following formulae:

 

 

On each training iteration, the model updated its weights after running through the network of a batch of 64 samples. The training finishes after completing 100 epochs.

According to numerical, modified U-Net showed an ability to learn on training dataset: loss value was equal to 0.09 and Dice coefficient achieved the value of 0.9. However, test results were not so well: in spite of the fact the validation loss has the same value as for training dataset, Dice coefficient did not achive the minimal threshold of 0.5 for acceptable image segmentation. Such problems may have been occurred because of some reasons:

* Objects of interest look different: their sizes vary from 50 to thousands of pixels.
* The database contains areas were marked as target objects by mistake.

# Conclusions

This article presents research results of modified U-Net for mine detection on satellite images of Planet database. Before training of developed model masks, which were generated from json files, were sliced on smaller parts together with high-resolution aerial photos. The training of modified U-Net was carried out on supercomputer NVIDIA DGX-1 of AI-centre of P.G Demidov Yaroslavl State University. The quality of deep learning algorithm was evaluated by Dice coefficient. The developed model was learned using modified binary cross-entropy with Dice loss. Despite high values of metrics on training dataset, the value of Dice coefficient on validation dataset did not show acceptable results. The further research wil be directed to solve problems occurred during numerical experiments, in particular create different classifiers for mines of different sizes or use other deep learning algorithms for object detection together with developed model.

##### Acknowledgment

The article was prepared with the financial support of the Ministry of Education of the Russian Federation as part of the research project No. 075-15-2019-249 connected with the implementation of applied scientific research on the following topic: «Development of applied solutions for processing and integration of large volumes of diverse operational, retrospective and the thematic data of Earth's remote sensing in the unified geospace using smart digital technologies and artificial intelligence» (identifier RFMEFI57517X0167). The authors are also grateful to the AI-centre of P.G. Demidov Yaroslavl State University for providing an access to NVIDIA DGX-1 supercomputer.

##### References

1. Y. Goodfellow, Y. Bengio, A. Courville. *Deep Learning*. The MIT Press, 2016, 800 p.
2. S. Raschka, *Python Machine Learning*. Packt Publishing Ltd., 2015, 454 p.
3. S. Ghosh, N. Das, I. Das, U. Maulik. Understanding Deep Learning Techniques for Image Segmentation. Web: https://arxiv.org/pdf/1907.06119.pdf.
4. A. Krizhevsky, I. Sutskever, G. E. Hinton, “Imagenet classification with deep convolutional neural networks”, Advances in neural information processing systems, 2012, pp. 1097–1105.
5. ImageNet Large Scale Visual Recognition Challenge 2012 (ILSVRC2012), Web: http://image-net.org/challenges/LSVRC/2012/.
6. E. Shelhamer, J. Long, T. Darrell. Fully Convolutional Networks for Semantic Segmentation. Web: https://arxiv.org/pdf/1605.06211.pdf.
7. 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. 234–241.
8. V. Khryashchev, L. Ivanovsky, V. Pavlov, A. Ostrovskaya and A. Rubtsov, “Comparison of Different Convolutional Neural Network Architectures for Satellite Image Segmentation”, 2018 23rd Conference of Open Innovations Association (FRUCT), Bologna, 2018, pp. 172-179.
9. L. Ivanovsky, V. Khryashchev, V. Pavlov, A. Ostrovskaya, “Building Detection on Aerial Images Using U-NET Neural Networks”, 2019 24th Conference of Open Innovations Association (FRUCT), Moscow, Russia, 2019, pp. 116-122.
10. Jilin-1 Optical Satellite. Web: https://www.satimagingcorp.com/satellite-sensors/jilin-1-satellite-sensor-1m/jilin-1-optical-satellite.
11. SPOT-6 Satellite Images. Web: https://www.satimagingcorp.com/gallery/spot-6.
12. A. Muller, S. Guido. *Introduction to Machine Learning with Python: A Guide for Data Scientists*. O’Reilly, 2016, 400 p.
13. N. Shukla. *Machine Learning with Tensorflow*. Manning Publications, 2018, 272 p.
14. P. Rychly, “A lexicographer-friendly association score”, Proceedings of the Second Workshop on Recent Advances in Slavonic Natural Language Processing, RASLAN 2008, pp. 6–9.
15. AI-centre of P.G Demidov Yaroslavl State University. Web: http://demid.ai.
16. D. P. Kingma, J. Ba. Adam: A Method for Stochastic Optimization. Web: https://arxiv.org/abs/1412.6980.