Modification of U-Net neural network in the task of multichannel satellite images segmentation

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

Roman Larionov  
P.G. Demidov Yaroslavl State UniversityYaroslavl, Russia  
r.larionov@uniyar.ac.ru

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

Abstract — Results of training of convolutional neural network for satellite four-channel image segmentation are performed. Input images contain blue, green, red and near‑infrared channels. The algorithm was trained to detect buildings and other urban areas. Modification of the U‑Net neural network with two encoders was used. The values of Sorensen coefficient and Jaccard index were calculated for 16 different urban regions.

Keywords — image segmentation, satellite images, convolutional neural network, deep learning

# Introduction

A satellite is a mechanical object that periodically rotates on the Earth orbit to performs certain functions and tasks for global media, 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.

Satellites receive hundreds terabytes of photos every day. Therefore, the development of methods for their automatic processing is very relevant. Modern satellites are capable to make photos with spatial resolution of 3 m/pixel and less. It’s possible to detect such small objects as buildings, landfills, etc. The development of technologies allowed to use methods of deep learning for satellite image segmentation [1].

This paper presents developed image processing method based on convolutional neural network (CNN). Such networks are capable for real‑time detecting and classifying objects. CNN have millions of parametres that are automatically matched through the training. CNN has shown its effectiveness in different computer vision tasks [2].

Automatic segmentation is an important part of aerial image pre-processing. Today, convolutional neural networks are widely used for this task. In particular, U-Net neural network architecture has shown its effectiveness in medical image segmentation [3]. Also U-Net has shown good results is satellite image segmentation [4]. The main advantage of this architecture is that algorithm can show good results even with a small training datasets. The mathematical structure of CNN is parallel, so graphic processors units (GPU) are ideal instrument to work with CNN [5].

There are some specific requirements for satellite images segmentation, containing different urban areas[6]:

* Size and type of buildings and urban structures may significantly vary from cottages to huge city buildings. The algorithm should detect objects of any



1. Sample images from Spacenet database

size very well. The usage of multiple encoders in neural network structure could solve this problem.

* The separation of objects with a high density of location. Algorithm should be penalized for bad separation of objects during training to improve output mask quality. This is achieved by careful selection of the loss function.
* Trained model should be invariant to rotations. This problem can solved by data augmentation.
* Aerial images have different spatial resolution. It’s important to create algorithm which has an ability to generalize perfectly.
* Algorithms should be noise robust. Images are shooted in different weather situations. It is possible that there is a noise in some photos, for example haze and glare from reflective surfaces.

This article presents the results of training of developed neural network for satellite multispectral image segmentation in order to detect buildings and urban areas. Modification of the U-Net architecture with the second encoder is proposed. The research continues previous investigations [7,8].

The rest of the paper organized as follows. The second part describes image datasets. Architecture of neural network is shown in the third part. The fourth part presents results of numerical experiments with big aerial image database containing different urban areas. Current research is summarized in the conclusion part.

# Satellite Images Database

Neural network model was pre-trained on images of Spacenet database. WorldView-2 and WorldView-3 satellites were used to make eleven-bit photos, all images are eight-channel. The database contains subsets with marked buildings for training deep learning algorithms [9]. We used subset of Khartoum (region of Spacenet database) to pre‑train developed CNN. Examples of images from Spacenet database are shown at Fig. 1.

Our model was trained on images of 16 different regions of Russian Federation. Every image has an approriate mask marked by experts. The dataset covers about 30 square kilometres. The images of this database contain blue, green, red and near infrared (NIR) channels with a spatial resolution of 3 metres per pixel.

The developed CNN requires input images of 256×256 pixels. Because of this the launch of CNN training on each image and approriate mask of dataset have been cut on two non-intersecting stripes. Each stripe was divided on patches of 256×256 pixels with step of 128 pixels, so pathes intersected by half.



1. U-Net neural network architecture with 2 encoders

To increase the training set, there were applied three types of data augmentation:

* Rotations on 90, 180, 270 degrees and reflections. The training set has increased 8 times after this procedure.
* Chromatic distortions. Images were translated from RGB color space to HSV color space. Random values were added to HSV coordinates of images. For the NIR channel, instead of chromatic distortions, random values from [-0.06, +0.06] interval were added. (NIR values were normalized in [0, 1] interval).
* Random shifts, scales and small degree rotations of patches.

As a result, the training image dataset contains 9784 batches each of them consist 16 images of 256×256 pixels.

# Neural Network Architecture

In the current investigation the modification of well known U‑Net architecture was used. Its classical structure is described in [3].

Original U-Net structure was modified. We have used two different encoders for RGB and NIR channels (fig. 2). The outputs of encoders were concatenated before being linked with approriate layers. As a result, the neural network has 38 convolutional layers, 37 ReLU activation functions, 37 operations of batch normalization, 1 sigmoid activation function, 10 maxpooling operations, 5 upsampling operations, 11 merge operations.

# Numerical Results

The quality of the segmentaion algorithms was evaluated by Sorensen-Dice coefficient (dice) and Jaccard index (IoU).

The coefficients can be calculated by following formulas (1-4), where x and y are values of pixels, X is marked by experts mask, Y is algorithm prediction:

 

 

 

 

Developed CNN was launched on NVIDIA DGX-1 supercomputer. Adam optimizer with learning rate 0.001 was used in the learning procedure. Lovász-Softmax loss was chosen as the loss function. The training finishes after completing 100 epochs.

The comparison of training results for original U-Net with four‑channel images as inputs and modification with two encoders for RGB and NIR channels is shown at fig. 3.

During preliminary training on the Spacenet dataset, the value of the Sorensen coefficient reached a value of 0.84, a Jaccard index of 0.77. The values of the Sørensen coefficient and the Jaccard index for all 16 regions for the algorithm with pre-training and without pre-training are shown at Fig. 4.

Example of input image and the result of segmentation is shown on Fig. 5. Sorensen coefficient and Jaccard index are equal 0.78 and 0.67 respectively.



1. Values of Sorensen coefficients for original U-Net and modified U-Net for images of 16 regions



1. Values of Sorensen coefficient and Jaccard index of pre-treined model and the algorithm without pre-training



1. Example of input image and result of segmentaion for developed algorithm

# Conclusions

This paper describes the training process of developed convolutional neural network intended for the segmentation of satellite images. U-Net architecture with two encoders was proposed to work with four-channel images. The algorithm was pre-trained on Spacenet image dataset. The value of Sorensen coefficient is equal 0.78, Jaccard index is 0.67.

Future directions of research may include:

* Categorization of buildings based on area size and encoders for separate classes.
* Usage of object boundaries as the third class for detection.

Developed algorithm can used for assessing the level of urbanization of various regions and tracking the construction of large objects.

##### Acknowledgment

This article was prepared with the financial support of the Ministry of science and education of the Russian Federation under the agreement No. 075-15-2019-249 from 04.06.2019 (identifier works RFMEFI57517X0167).

##### References

1. Zhang, Liangpei, Lefei Zhang, and Bo Du. “Deep learning for remote sensing data: A technical tutorial on the state of the art.” IEEE Geoscience and Remote Sensing Magazine (2016): 22-40.
2. S. Seferbekov, V. Iglovikov, A. Buslaev, A. Shvets. Feature Pyramid Network for Multi-Class Land Segmentation. Web: https://arxiv.org/pdf/1806.03510.pdf.
3. 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.
4. Zhang, Z., Liu, Q., & Wang, Y. (2018). “Road extraction by deep residual u-net”. IEEE Geoscience and Remote Sensing Letters, 15(5), pp. 749-753.
5. A. Gulli, S. Pal., *Deep Learning with Keras*, Packt Publishing, 2017, 320 p.
6. X. X. Zhu et al., “Deep Learning in Remote Sensing: A Comprehensive Review and List of Resources,” in IEEE Geoscience and Remote Sensing Magazine, vol. 5, no. 4, pp. 8-36, Dec. 2017.
7. V. V. Khryashchev, V. A. Pavlov, A. Priorov and A. A. Ostrovskaya, "Deep Learning for Region Detection in High-Resolution Aerial Images," 2018 IEEE East-West Design & Test Symposium (EWDTS), Kazan, 2018, pp. 1-5.
8. L. Ivanovsky, V. Khryashchev, V. Pavlov and 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.
9. SpaceNet Database, Web: http://explore.digitalglobe.com/spacenet