IMAGE SEGMENTATION USING DEEP LEARNING METHODS

I. V. Saetchnikov

Belarusian State University, Minsk;
saetchnikovivan@gmail.com;
scientific adviser – Dr. Skakun Victor V., assistant professor

This paper discusses the image segmentation methods based on deep learning methods. The segmentation object dataset was formed by two sets of World View 3 satellite images: RGB images and a 16-channel multispectral images. For dataset, I develop image pre-processing algorithms based on CNN. As segmentation methods Convolutional Neural Network are used due to its possibility to process not only by their spectral differences, but also by their spatial attributes. Based on U-Net, DeepLab and FullConv architecture networks were developed for satellite image segmentation. Finally, Jacard indexes of 3 networks were compared. These results are primarily due to the classes unevenness. To increase accuracy, it is necessary to train separately the classes sets. Among the applications of CNN in satellite images segmentation, we can distinguish urban infrastructure localization for Smart City technology, the segmentation of agricultural fields for the precision agriculture etc.

Key words: image segmentation, convolutional neural network, autoencoder, satellite, multispectral image.

Success in deep neural networks implementation allowed to segment images of different nature and structure with an accuracy exceeding standard segmentation methods. Among practical applications of images segmentation, it is possible to select the following: medicine [1, 2], transport, video surveillance, biometric authentication, space etc.

Now satellites can photo the Earth's surface with the resolution up to 31 cm in the different spectral ranges. That allows to carry out the post-processing of images of the interest Earth located objects. One of such satellites – the World View 3 (Digital Globe company) – has three camera shooting modes which characteristics are presented in table 1.

Table 1

Characteristics of the WorldView3 satellite cameras.

Parameters	VNIR	CAVIS (RGB + P)	SWIR
Wavelength range	400 nm-1040 nm	400-800 nm	1195-2365 nm
Dynamic range	11 bit/pixel	11 bit/pixel	14 bit/pixel
Resolution	1.24 m	0.31 m	7.5 m
Number of channels	8 channels: 6 – visible, 2 – near-infrared.	4 channels	8 channels

Finally, dataset was formed by two sets of World View 3 satellite images: RGB images and 16-channel multispectral images. The multispectral image consisted of a multispectral part and a middle infrared component (1195–

2365 nm). Also 25 tagged tiff images from the Dstl Imagery Feature Detection dataset were used. The set consists of the following 10 classes: buildings, fences, roads, dirty roads, trees, fields, rivers, lakes, trucks, cars. The marking was presented in the Excel csv file in the form of polygons, according to which it is possible to paint over areas belonging to a certain class. At the same time, the classes on the images were unevenly distributed, the largest area was occupied by fields, the following were rivers, followed by trees and dirty roads. Due to the uneven classes and the proximity of the object's spectrum to different classes, the segmentation algorithm must be based not only on the pixels values, but also on their relative position [3]. Convolutional neural networks, which can separate objects not only by their spectral differences, but also by their spatial attributes, are well suited to the described requirements.

For the neural network input were selected 16 channels, covering the spectral range from 400 nm to 2365 nm, which includes the range of RGB-provided images. Due to different spatial resolution for the first 8 channels and the last 8 channels of 16 channel images, they are reduced to a single size of 1024×1024, which for both is larger than their original size. To increase the images spatial sizes, a convolutional neural network consisting of 19 inner blocks of the convolutional layer + Relu was used. The result of resolution increasing on the training images SSIM = 0.9215. But due to small number of labeled images (25 Images) and their large input dimensions, the input dimension 128×128 was chosen for the neural network.

For image segmentation three architectures of convolutional neural networks were chosen: U-Net, FullConv, DeepLab. Each of the architectures was processed for the required image input dimension.

Architecture of U-Net network was the following: 2 ConvLayer (32 feature maps) + 3 sets of 3 ConvLayer with 64, 128, 256 feature maps. Between an encoder and decoder there was 4 ConvLayer.

Architecture of FullConv network was the following: set of 2 ConvLayer (32 feature maps) + MaxPool, 4 sets of 3 ConvLayer (64, 128, 256 feature maps) + MaxPool, set of 2 ConvLayer (1024 feature map). The last layer was dropout (0.15 Accuracy).

DeepLab network architecture was the following: set of two ConvLayer + MaxPool, 3 sets of three ConvLayer (64, 128, 256 feature maps) + MaxPool, a set of three ConvLayer (512 feature map). Batch Norm + Dropout (0.25 Accuracy) were used to prevent retraining. Adam gradient descent optimization method with parameters: learning speed: 1e-3, $\beta 1 = 0.9$, $\beta 2 = 0.999$ and decay 1e-5 was used. The training process includes 30 epochs.

For all architectures, focal loss entropy was used as a cost function, which showed good convergence in training and considered the classes unevenness. Jacard index was used as a quality metric. For training and testing random

samples of 128×128 size from original images were taken. Thus, the shear augmentation principle was used for the dataset, which made it possible to improve the result and expand the data set. The results of training and validation over 30 epochs are demonstrated in table 2.

Table 2

Image segmentation results

Architecture	Jacard Index (testing)	Jacard Index (validation)
U-Net	0.652	0.682
FullConv	0.562	0.554
DeepLab	0.512	0.694

On the training and test sample, the best result in quality metrics shows U-Net, but on the validation sample DeepLab has the best result, which indicates a better generalizing ability of DeepLab compare to U-Net. Architecture FullConv showed worse result than U-net and DeepLab.

Image segmentation using deep learning methods is necessary to determine the relations between objects, as well as the context of objects in an image. Applications include face recognition, video surveillance, medicine and satellite image analysis. Especially, deep neural networks are suitable for analyzing satellite images, since they do not require real-time processing and allow to obtain better segmentation accuracy of even the smallest features compared to standard segmentation methods, such as thresholding methods, k-mean clustering method etc.

Reference

1. W-Net for Whole-Body Bone Lesion Detection on Ga 68-Pentixafor PET/CT Imaging of Multiple Myeloma Patients / L. Xu [et al.] // Molecular Imaging, Reconstruction and Analysis of Moving Body Organs, and Stroke Imaging and Treatment. 2017. P. 23–30. DOI: 10.1007/978-3-319-67564-0.
2. *Xu S., Dang H.* Deep residual learning enabled metal artifact reduction in CT // Medical Imaging 2018: Physics of Medical Imaging. 2018. P. 10–15. DOI 10.1117/1.JMI.6.1.011001.
3. *Badrinarayanan V., Kendall A., Cipolla R.*. SegNet: A Deep Convolutional Encoder-Decoder Spatio-temporal Road Detection from Aerial Imagery using CNNs 499 Architecture for Image Segmentation. 2015. DOI: 10.1109/TPAMI.2016.2644615.