

Agricultural parcel localization on satellite images using U-Net-based neural network

Maria Pavlova
Vision systems laboratory
IITP RAS (Kharkevich Institute)
Moscow, Russia
m.pavlova@visillect.com

Alexey Savchik
Vision systems laboratory
IITP RAS (Kharkevich Institute)
Moscow, Russia
savsmail@gmail.com

Lev Teplyakov
Vision systems laboratory
IITP RAS (Kharkevich Institute)
Moscow, Russia
teplyakov.l@ya.ru

Mikhail Zagarev
Digital Agro LLC
Moscow, Russia
mzagarev@gmail.com

Igor Kukoev
Digital Agro LLC
Moscow, Russia
kukoevigor72@gmail.com

Anton Grigoryev
Vision systems laboratory
IITP RAS (Kharkevich Institute)
Moscow, Russia
grigoryev@visillect.com

Abstract—This work considers the problem of automatic delineation of agricultural parcels on satellite images based on true-color images and NDVI vegetation index maps from Sentinel-2 satellites (10 m ground sampling distance). The problem is solved using a U-Net-based convolutional neural network. We consider problem formulation either as parcel mask or boundary detection; the multiclass (simultaneous) training did not prove to be effective. The approach looks promising and applicable for automated land mapping for agricultural monitoring systems.

Keywords—U-Net, convolutional neural network, precision agriculture, automated mapping

I. INTRODUCTION

In this paper, we consider the problem of automating the mapping of agricultural fields using satellite data of 10-meter spatial resolution. This task is relevant both in cadastral accounting and in agricultural monitoring. High-precision manual mapping of parcels is a labor-intensive process, and knowledge about the boundaries of parcels is an essential element for solving other tasks of agricultural monitoring, in particular, evaluating various indicators of productivity and land condition when using precision farming approaches.

There are many works on the related problem of crop classification [1]–[3]. The problem of field mapping automation is not so studied, although there are some works [4], [5].

Some works, including [4], [5], use high-resolution data, which makes the problem easier. For example, high-resolution 80-cm allow to detect single trees [6]. However, lower-resolution data is more widely available thanks to research programs such as Sentinel-2 [7]. It provides regularly updated (about two times per week) multispectral satellite imagery in resolutions from 10 to 60 meters per pixel, depending on the spectral channel.

Determining the most suitable spectral ranges for mapping is not a trivial task. This paper uses the fact that the so-called vegetation indices-images calculated from images in different wavelength ranges [8], [9] - have long been known and widely used to solve problems of agricultural monitoring.

In this paper, in addition to the image in the visible range, we investigate the use of the NDVI (Normalized Difference Vegetation Index) as input data for automated mapping. It is a normalized relative vegetation index, which is useful in crop monitoring problems [10]. NDVI is calculated from the ratio of the observed intensity of red (RED) and near-infrared (NIR) channels:

$$NDVI = \frac{NIR - RED}{NIR + RED}.$$

Currently, algorithms based on training in artificial neural networks, in particular using convolutional layers, are widely used in various problems of image analysis [4], [5], [11], [12]. A particular case of full-convolution architecture, showing good quality in segmentation problems, is the U-Net family of neural networks [13], which implements a multiscale approach to image analysis. We can also note one of the universal segmentation algorithms MaskRCNN [14], in which object detection is accompanied by further pixel-by-pixel segmentation using convolutional layers. This algorithm allows one to detect individual objects and their exact boundaries, including when their images intersect. However, this more complicated method is not required for defining masks or parcel boundaries when objects do not intersect, and there is no problem with partial obscuration.

II. METHOD

We used a neural network approach to automatically detect parcels. In this approach we model detection with a function $f_w : X \rightarrow Y$. The f_w function maps multi-channel images $x \in X = [0, 1]^{h \times w \times c}$ to single-channel images $y \in Y = [0, 1]^{h \times w}$. The pixels $y_{i,j} \in [0, 1]$ of such an image contain confidence estimates that the i, j pixel in the source image refers to a parcel (or a parcel boundary in case of boundary detection). The type of f_w function depends on the task to be solved and defines the architecture of an artificial neural network.

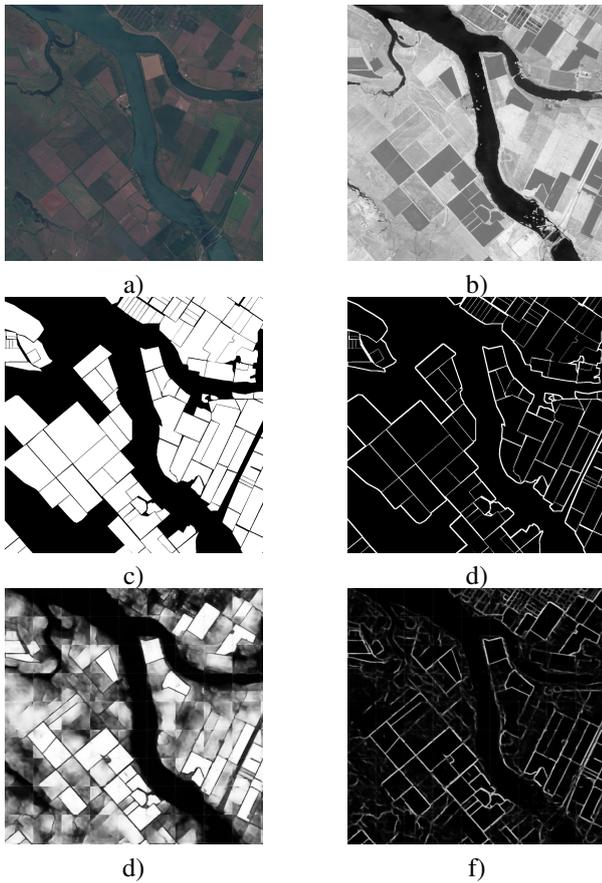


Fig. 1. Image example and the resulting predictions: a) color data (TCI), b) NDVI vegetation index, c) ground truth binary mask, d) ground truth boundary mask, e) parcels mask prediction, f) boundary mask prediction.

A sequential network is a simple example of an artificial neural network. Such a network consists of inputs $x = h_0$ and a few trainable functions (layers) $h_i = f_i(h_{i-1})$, $i = \{1, \dots, n\}$, applied sequentially. The last layer output $y = h_n$ is treated as the neural network output. For the image processing tasks, the convolutional layers are usually used. In this case, the input and the output of the layer are images, i.e., 3d tables (tensors).

$$h_i^{x,y,ct} = \text{ReLU} \left(\sum_{dx,dy,cf,ct} w_{dx,dy,cf,ct}^i h_{i-1}^{x+dx,y+dy,cf} \right),$$

where $\text{ReLU}(x) = \max(x, 0)$. The idea of such filters is to locally process different image parts in the same way. This also allows to simultaneously optimize weights w on different parts of the image. Convolutions have typically small kernel size, usually 3×3 , (bigger ones are reducible to several such ones). To make the output size to be equal with the input size, an input image is usually zeros-padded.

Function f_w parameters (or weights) w are automatically tuned on the collected dataset $\{x_i, y_i\}$ with the expertly marked position of the parcels (and their borders). The w is changed so to minimize the loss function (or empirical

risk) $L(w) = \sum_i l(f_w(x_i), y_i)$, where l is the two-class cross entropy.

$$l(y, \hat{y}) = \begin{cases} -\log(y) & , \hat{y} = 1 \\ -\log(1 - y) & , \hat{y} = 0. \end{cases}$$

U-Net-like architectures [13] are commonly used in problems of image segmentation and pixel-wise classification. Neural networks of this type have an encoder-decoder architecture with several layers of different resolutions. Also, such architecture is not sequential and has shortcut connections. U-Net-like networks have several advantages, including:

- a sufficiently large receptive field, which allows a neural network to make each pixel decision based on a relatively wide spatial neighborhood;
- a computational efficiency due to the multiscale approach.

The network of this architecture works with a small number of weights in the original resolution image, then with a large number of weights in the image of a much lower resolution and, in the end, combines the low-resolution results with high-resolution data from the original image for more accurate pixel prediction. In this paper, we used a neural network with 32 filters on the first layer and the smallest network scale of $1/8$ (see Fig. II).

A. Training

A dataset consisting of 122 4-channel images of 22 areas of the Earth surface for the period 04/05/2018 – 11/11/2018 from the Sentinel-2 satellite imagery archive was prepared. The size of each image is 1030×1030 pixels, imagery resolution is 10 m/pixel. The first 3 channels of each image are visible colors (TCI in Sentinel-2 nomenclature), and as the fourth channel, the NDVI vegetation index map is used (calculated by bands 4 and 8 of the original multispectral image), see Fig. 1ab.

Fields and similar structures were manually marked on each full-color image in the form of polygonal contours, resulting in 400 average objects being marked on all images. The field boundary mask was constructed as follows: the field mask was morphologically dilated with a square window of 10×10 , after which the points included in the dilated mask but not in the original ones were considered as the boundary points (see Fig. 1cd).

The dataset was divided into training and test parts, consisting of 17 regions with 94 images total and 5 sections with 28 images, respectively.

The network has been trained for a 25 epochs with 500 batches. Each batch contains 32 random 128×128 patches, cropped randomly from the original images. The used optimizer is Adam [15] — one of stochastic gradient descent methods. We used a loss function consisting of a cross entropy and l_2 — regularization with 0.0001 coefficient to prevent overfitting.

TABLE I. METHODS COMPARISONS

Input Data	Output Data	Results AUC-ROC/AUC-PR
TCI	boundaries	$0.80 \pm 0.00 / 0.39 \pm 0.00$
TCI	parcels	$0.82 \pm 0.01 / 0.53 \pm 0.02$
TCI	parcels & boundaries	(b) $0.78 \pm 0.00 / 0.36 \pm 0.00$
TCI	parcels & boundaries	(p) $0.77 \pm 0.02 / 0.47 \pm 0.06$
NDVI	boundaries	$0.79 \pm 0.00 / 0.39 \pm 0.01$
NDVI	parcels	$0.80 \pm 0.01 / 0.51 \pm 0.04$
NDVI	parcels & boundaries	(b) $0.77 \pm 0.01 / 0.34 \pm 0.02$
NDVI	parcels & boundaries	(p) $0.78 \pm 0.02 / 0.43 \pm 0.05$
TCI and NDVI	boundaries	$0.79 \pm 0.02 / 0.36 \pm 0.03$
TCI and NDVI	parcels	$0.81 \pm 0.01 / 0.45 \pm 0.08$
TCI and NDVI	parcels & boundaries	(b) $0.78 \pm 0.01 / 0.37 \pm 0.00$
TCI and NDVI	parcels & boundaries	(p) $0.81 \pm 0.02 / 0.53 \pm 0.01$
	random boundaries	$0.50 \pm 0.00 / 0.06 \pm 0.00$
	random parcels	$0.50 \pm 0.00 / 0.39 \pm 0.00$

III. RESULTS

We use pixel-wise classification metrics. The most common one is the AUC-ROC metric. To calculate it we consider all possible thresholds to obtain the all possible classifiers

$$c_{i,j}^T = [y_{i,j} > T], T \in \mathbb{R}.$$

These classifiers' parameters on the TPR and FPR coordinates plot the so-called ROC curve. The area under this curve is the desired value. However, this metric may be low informative if the classes are highly unbalanced. In our case, the number of background non-borders pixels is much bigger than the number of object pixels (borders). To reduce this effect, we also calculate the AUC-PR metric built similarly on the values of precision and recall instead of TPR and FPR.

We did a few computational experiments to study dependence by the input channels: TCI, NDVI, TCI+NDVI. Also, we did a few experiments to check the multiclass method, in which the neural network learns to predict parcels and its boundaries simultaneously. In each case, the experiment was repeated three times to estimate not only the value but its standard deviation also.

For comparison, the value of metrics for a random classifier is given. For boundaries & parcels prediction, different boundary (b) and parcels (p) results are presented. The results are listed in Tab. I.

The presented results show that a neural network is significantly higher than the random algorithm, which has $AUC - ROC$ and $AUC - PR$ equals to $0.5 / 0.4$ for the fields and to $0.5 / 0.06$ for the boundaries. The results of the multitask training are not significantly different from the usual ones. The quality of border detection by AUC-PR metrics is much lower than that of AUC-ROC, because of a much lower share of boundary pixels compared to the share of parcels pixels. The results of the network look strongly correlated with the correct answer (see Fig. 1de), which is probably the most important thing in this task: both the metrics themselves and the ground truth values are not very reliable, as it is difficult to check whether it is a field boundary at a given point or it is another visually similar structure. We see that

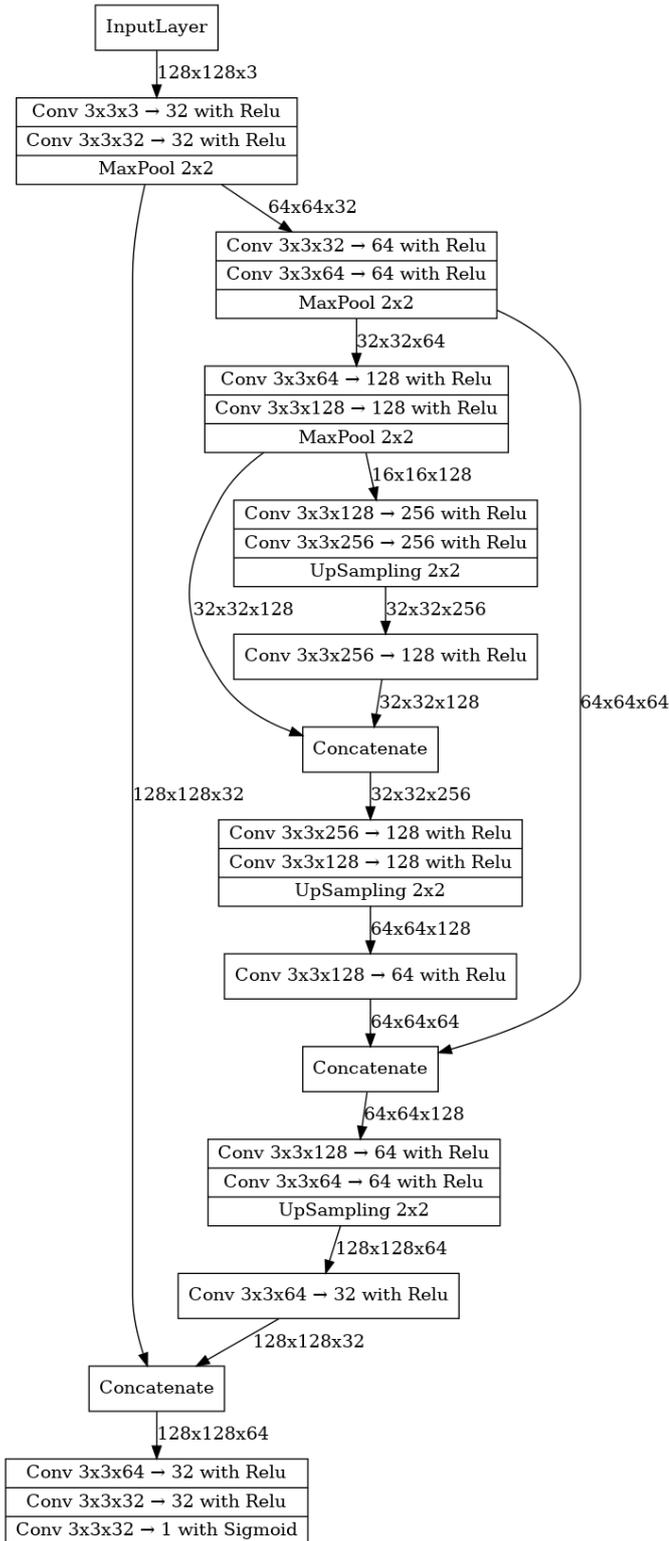


Fig. 2. The neural network architecture. It is a U-Net-based one with encoder, decoder, and shortcut connections.

this approach is applicable to the construction of agricultural monitoring systems and the appropriate optimization for a particular application scenario. In the case of a larger dataset, there are no obstacles to obtain good results.

IV. CONCLUSION

The paper considers the problem of determining the location of agricultural parcels using multispectral satellite images. The input images contain visible range imagery, NDVI vegetation index maps, or both. The output ground truth annotation contains parcels, boundaries, or both.

The results demonstrated the applicability of the U-Net network architecture for this task with at least 0.7 AUC-ROC metric value for the parcels and boundaries. The comparison of different variants of the training task setting showed that using vegetation index maps may be useful in this task. However, in the absence of infrared images using only the image in the visible range shows the comparable result. Multiclass training did not show any advantages.

In general, the obtained results show the perspective of U-Net architecture neural networks application for solving the tasks of large-scale automated agricultural monitoring using freely available satellite data of medium spatial resolution (10 m/px in the considered case of Sentinel-2). Further development of this work can be the construction of more relevant metrics (and, accordingly, loss functions) for the task, using more multispectral information and historical images of the same parcels to improve the accuracy and relevance of recognition results.

REFERENCES

- [1] S. M. Borzov, M. A. Guryanov, and O. I. Potaturkin, "Study of the classification efficiency of difficult-to-distinguish vegetation types using hyperspectral data," *Computer Optics*, vol. 43, no. 3, pp. 464–473, 2019.
- [2] S. A. Bibikov, N. L. Kazanskii, and V. A. Fursov, "Vegetation type recognition in hyperspectral images using a conjugacy indicator," *Computer Optics*, vol. 42, no. 5, pp. 846–854, 2018.
- [3] A. A. Varlamova, A. Y. Denisova, and V. V. Sergeev, "Earth remote sensing data processing for obtaining vegetation types maps," *Computer Optics*, vol. 42, no. 5, pp. 864–876, 2018.
- [4] G. M. Musyoka, "Automatic delineation of small holder agricultural field boundaries using fully convolutional networks," Master's thesis, University of Twente, Enschede, The Netherlands, 2018.
- [5] M. M. Alemu, "Automated farm field delineation and crop row detection from satellite images," Master's thesis, University of Twente, Enschede, The Netherlands, 2016.
- [6] A. Y. Denisova, A. A. Egorova, V. V. Sergeev, and L. M. Kavelenova, "Requirements for multispectral remote sensing data used for the detection of arable land colonization by tree and shrubbery vegetation," *Computer Optics*, vol. 43, no. 5, pp. 846–856, 2019.
- [7] M. Drusch, U. Del Bello, S. Carlier, O. Colin, V. Fernandez, F. Gascon, B. Hoersch, C. Isola, P. Laberinti, P. Martimort et al., "Sentinel-2: Esa's optical high-resolution mission for gmes operational services," *Remote sensing of Environment*, vol. 120, pp. 25–36, 2012.
- [8] N. Pettorelli, J. O. Vik, A. Mysterud, J.-M. Gaillard, C. J. Tucker, and N. C. Stenseth, "Using the satellite-derived ndvi to assess ecological responses to environmental change," *Trends in ecology & evolution*, vol. 20, no. 9, pp. 503–510, 2005.
- [9] X. Jinru and B. Su, "Significant remote sensing vegetation indices: A review of developments and applications," *Journal of Sensors*, vol. 2017, pp. 1–17, 2017.
- [10] M. Boori, K. Choudhary, and A. Kupriyanov, "Crop growth monitoring through sentinel and landsat data based ndvi time-series," *Computer Optics*, vol. 44, no. 3, pp. 409–419, 2020.
- [11] J. Davidse, "Semi-automatic detection of field boundaries from high-resolution satellite imagery," Master's thesis, Wageningen University, The Netherlands, 2015.
- [12] A. Garcia-Pedrero, C. Gonzalo-Martín, and M. Lillo-Saavedra, "A machine learning approach for agricultural parcel delineation through agglomerative segmentation," *International journal of remote sensing*, vol. 38, no. 7, pp. 1809–1819, 2017.
- [13] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," *International Conference on Medical image computing and computer-assisted intervention*, Springer, pp. 234–241, 2015.
- [14] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask r-cnn," *Proceedings of the IEEE international conference on computer vision*, pp. 2961–2969, 2017.
- [15] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.