UDK 004.93:004.032.26

RECOGNITION OF UNDERLYING SURFACE USING A CONVOLUTIONAL NEURAL NETWORK ON A SINGLE-BOARD COMPUTER

D.A. Paulenka
Software Engineer,
UIIP NASB,
dmitri.pavlenko@gmail.com

V.A. Kovalev
Head of Biomedical Image Analysis Laboratory, Ph.D.,
UIIP NASB,
vassili.kovalev@gmail.com

E.V. Snezhko
Leading Researcher, Ph.D.,
UIIP NASB,
eduard.snezhko@gmail.com

V.A. Liauchuk
Researcher, Ph.D.,
UIIP NASB,
vitali.liauchuk@gmail.com

E.I. Pechkovsky
Leading Software Engineer,
UIIP NASB,
zver927@gmail.com

The United Institute of Informatics Problems of the National Academy of Sciences of Belarus, (UIIP NASB, 6 Surganova Str., 220012, Minsk, Belarus)

Abstract. The article presents development results for hardware and software system (micromodule), which detects and classifies underlying surface images of Earth. Given device has size 5.2×7.4×3.1 cm, mass 52 g and uses convolutional neural network based on MobileNetV2 architecture for image classification. The micromodule can be installed on board of a small spacecraft or a light unmanned aerial vehicle (drone). The information provided in this paper could be useful for engineers and researchers who are developing compact budget mobile systems for processing, analyzing and recognition of images.

Keywords: image recognition, convolutional neural network, deep learning, single-board computer, edge device, mobile system

Introduction. Automatic classification of underlying surface images, as well as search and recognition of objects in video and photo images are important tasks, the solution of which ensures the appropriateness of their application. Automatic terrain recognition using images of the underlying surface eliminates the long and tedious viewing of video materials by ground control operators, reduces time required for data preparation, increases reliability of choice, decreases power consumption from on-board sources due to the automatic regions of interest selection and diminish delay time between received signal and response to it.

Modern single-board microcomputers and smartphones have sufficient computing power to install software for automatic search and classification of images. This allows data preprocessing directly on the receiving device (edge device) without transferring a large amount of information to the server. Thus, the data preprocessing can be done aboard an unmanned aerial vehicle (UAV) or a small spacecraft. The authors developed a micromodule which is a hardware and software system and consists of a single-board microcomputer with necessary auxiliary equipment and a convolutional neural network (CNN) based on MobileNetV2 architecture for data processing.

Main micromodule functions

1. The micromodule is an autonomous system for automatic images recognition of underlying surface of various types: different kinds of forest, industrial structures, reservoirs, roads, fields, shrubs, country houses, farmland of various types with vegetation at different stages of growth, etc.
2. Target classes are set in the flight task by means of CNN which is prepared using Deep Learning technology.
3. Through the built-in video camera, the micromodule receives surface images and then classifies them in accordance with the flight task. If necessary, frames of the video sequence can be stored on the micromodule's external media for subsequent analysis.
4. The interaction between micromodule and UAV is minimal. And reduced to real-time transferring from micromodule to UAV the identified class number of a current underlying surface and the probability of its reliable determination. Class "0" means that the current image does not belong to any of the target classes specified in the flight task. The decision to perform high-quality shooting, recording and image transmission to the ground control is made outside of the micromodule.
5. The micromodule is a computing system with all necessary software development tools, operating system and interface equipment. Implementation with circuit solutions from separate specialized components such as microcontrollers, digital signal processors (DSP), chipsets, systems on a chip, etc. is not considered by the authors in this article.

Micromodule's hardware

Micromodule's hardware (Figure 1) consists of protective case, single-board Raspberry Pi Zero Wireless microcomputer, microSDHC memory card, Raspberry Pi Zero V1.3 Mini Camera, 720 mAh ROBITON Li-Po battery, battery charging microchip and connecting wires.

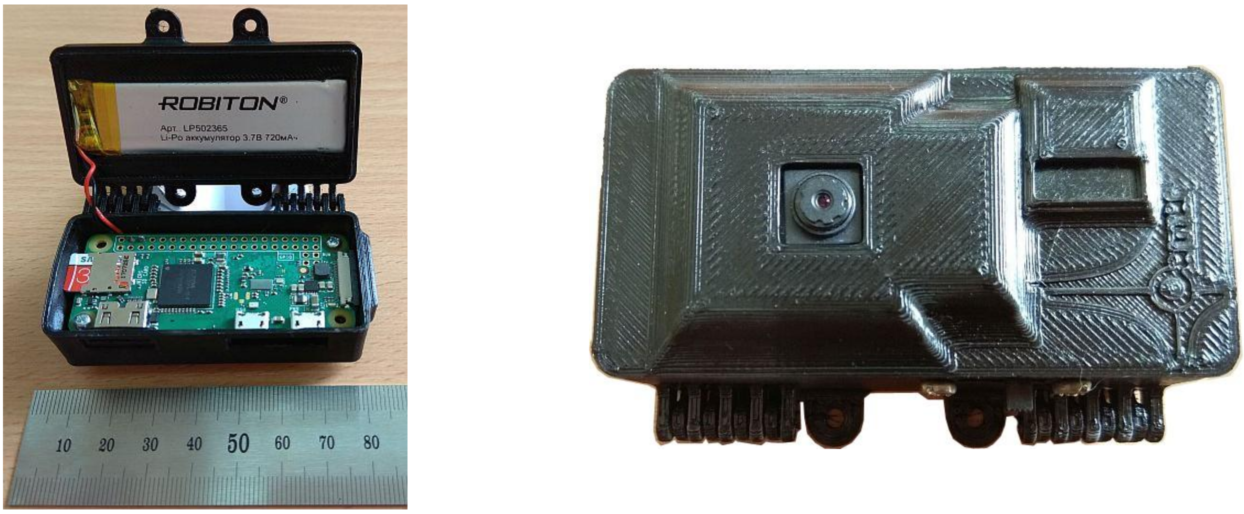

Figure *1*. – Micromodule's hardware

To reduce production time, decrease manufacturing complexity and product cost the protective case was made on a 3D printer by the layer-by-layer extrusion of molten yarn from polycarbonate material.

The advisability of equipment selection, comparative analysis of computing microplatforms and technical requirements to hardware and software system are carefully described in [1, 2].

Micromodule size is 5.2×7.4×3.1 см. Micromodule weight is 52 г. Maximum power consumption at payload is 1,75 ± 0,25 Wt.

Micromodule's software

Micromodule's software consists of flying and ground parts (Figure 2).

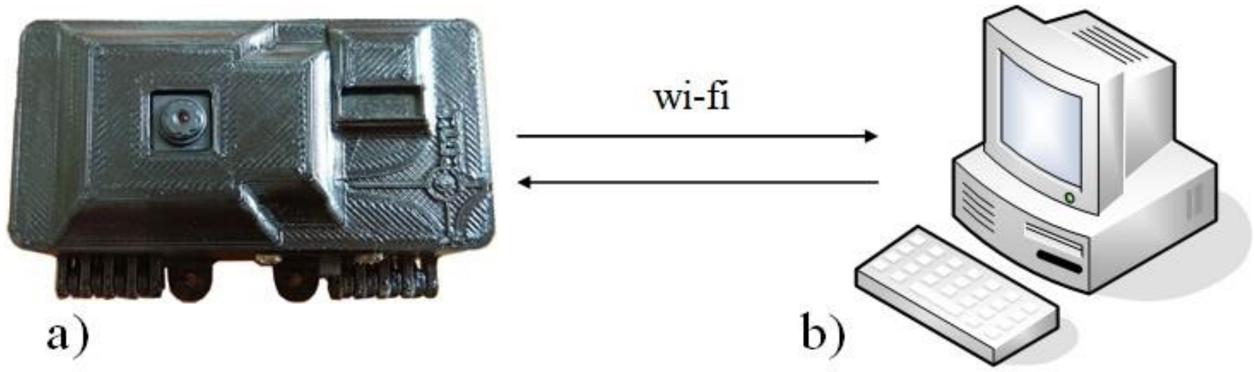

Figure *2*. – Flying (a) and ground (b) parts of the micromodule

The flying part processes images and video stream frames, picks and compresses them. The ground part creates image datasets and constructs flight tasks. A wireless Wi-Fi connection is established between ground and flying parts to launch flight task, exchange data and monitor software functioning. In its turn, the flying part operates in two modes: the image processing mode and the video stream processing mode.

In the image processing mode, the flying part of the micromodule gets input *image files*, performs real-time data recognition, selection and compression and returns images with required classes of underlying surface according to the flight task.

In the video stream processing mode, the flying part of the micromodule gets input *video stream frames*, performs real-time data recognition, selection and compression and returns frames with required classes of underlying surface according to the flight task.

Micromodule's software is written in the Python and C/C++ programming languages. Operating system of the flying part is Raspbian Jessie. Operating system of the ground part is Ubuntu Linux.

The human operator of the ground part forms a flight task and copies it to the flying part. The flying part is mounted on a quadcopter or a UAV. Micromodule is attached to the aircraft but is not interlinked with its software or hardware. During the flight, micromodule in real-time mode processes video stream frames from the camera or scans input satellite images received via Wi-Fi. At the output, it has indexes and images of the found target classes in accordance with the flight task. For image processing, micromodule uses CNN which is based on the MobileNetV2 architecture [3].

Micromodule's flying part processes frames at a speed of more than three frames per second (see the next section).

The architecture of developed convolutional neural network

The main function of the micromodule is the real-time images recognition of the underlying surface, that is, the analysis of a current video sequence frame or an image from the scanning window and assigning it to one of the classes indicated in the flight task.

Micromodule has low computational speed of a one-core processor with ARMv6 architecture (1 GHz) and small size of RAM (512 MB). This does not allow the use of any well-known neural networks (AlexNet, VGG16, GoogLeNet, etc.). Therefore, a specialized CNN architecture based on MobileNetV2 was implemented [3].

Developed CNN has eleven blocks (106 layers in total), each block is the same as in MobileNetV2, as well as an output block adapted for image classification of Earth's underlying surface. In addition to reducing the number of main blocks from thirteen to eleven, a series of optimization techniques to speed up calculations were used. These techniques are taken into account the micromodule's hardware architecture. In general, optimization techniques reduced the time to get the output CNN vector by about five times on the same data compared with the original MobileNetV2 architecture.

The process of solving an image recognition task has two stages:

- training the neural network on sample images of underlying surface target classes, which is performed on the ground part of the micromodule's software;
- real-time image recognition on the flying part of the micromodule's software.

Software for flight task construction on the micromodule's ground part uses Keras library and trains a neural network on a powerful video card (Graphical Processing Unit, GPU). Flight task construction result is a trained neural network. This trained neural network, or rather a file with weights of the trained neural network is then transferred to the micromodule's flying part.

For real-time image classification on the micromodule's flying part, a special functions library was developed using C programming language. This function library imports CNN's weights to the neural network model and then calculates output vector values from the input images.

With this implementation, the classification time for a single three-channel (RGB) image 96×96 pixels in size on a single-core processor with ARMv6 architecture and a frequency of 1 GHz is on average 280 ms per image.

Image recognition quality assessment

Two image datasets with various underlying surface classes were prepared (Figure 3).

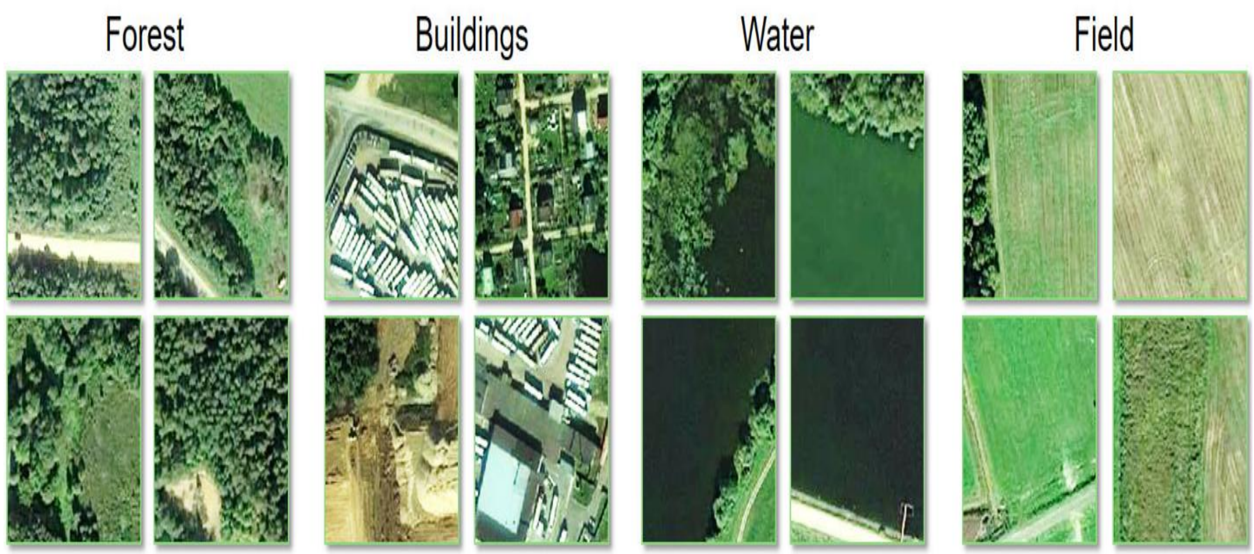

Figure 3. – Examples of underlying surface images

Maps with underlying surface were downloaded using free of charge SASPlanet software (URL: http://www.sasgis.org). Map projection is "Mercator / Google Maps". Map scale is Z18.

Maps markup and creation of image dataset with underlying surface classes were made by "Manual image annotation with polygons" software tool (URL:

https://github.com/foobar167/junkyard/tree/master/manual_image_annotation1). This tool was developed by the authors of the article and is a component of the micromodule's ground part.

The first image dataset consists of eight classes: lake, city, field, forest, highway, road, river, village. To create the flight task, 40 000 images, each with size 256×256 pixels were used. To avoid overfitting effect, satellite maps for training were taken from Minsk city neighborhood, and for testing from Brest city and Lake Naroch neighborhood. Resulting confusion matrix for eight classes of developed CNN is in the Table 1, which gives the relative number of images that are assigned to one of the classes.

Table 1. – Confusion matrix for eight classes of developed CNN

class	lake	city	field	forest	highway	road	river	village
lake	**0.81**	0.00	0.05	0.13	0.00	0,00	0,01	0,00
city	0.03	**0.60**	0.04	0.01	0.07	0,01	0,01	***0,23***
field	0.04	0.00	**0.95**	0.00	0.00	0,01	0,00	0,00
forest	***0.60***	0.00	0.02	**0.32**	0.00	0,00	0,05	0,00
highway	0.00	0.01	0.04	0.01	**0.73**	0,11	0,10	0,01
road	0.00	0.00	0.01	0.00	0.05	**0,91**	0,04	0,00
river	0.03	0.00	0.09	0.02	0.05	***0,29***	**0,47**	0,04
village	0.00	0.00	0.01	0.00	0.00	0,06	0,00	**0,92**

Image percentage with minimum confidence coefficient 0.9 is equal to 0.86. Average classification accuracy is equal to 0.71.

The table shows that the class "city" is often confused with the class "village", which is obvious, because "city" and "village" are similar by the texture. Class "forest" frequently falsely defined as class "lake", possibly because the "lake" stands out along with the forested shores. Class "river" is confused with the "road", possibly due to the fact that country roads often go along the rivers, so the rivers were marked along with the roads in the images. I.e. there are several classes on the image: lake and forest, river and road, etc.

For comparison, Table 2 shows the confusion matrix for the popular neural network GoogLeNet, from which it can be seen that the recognition quality is only slightly better than in Table 1.

Table 2. – Confusion matrix for eight classes of GoogLeNet

class	lake	city	field	forest	highway	road	river	village
lake	**0,91**	0,00	0,01	0,05	0,01	0,00	0,02	0,00
city	0,10	**0,60**	0,00	0,00	0,15	0,01	0,01	***0,13***
field	0,04	0,00	**0,94**	0,00	0,00	0,01	0,01	0,00
forest	***0,55***	0,00	0,00	**0,44**	0,00	0,00	0,01	0,00
highway	0,01	0,00	0,05	0,00	**0,73**	0,07	0,13	0,00
road	0,01	0,00	0,00	0,00	0,01	**0,95**	0,03	0,00
river	0,01	0,00	0,00	0,00	0,01	***0,11***	**0,87**	0,00
village	0,00	0,00	0,00	0,00	0,00	0,08	0,00	**0,91**

Image percentage with minimum confidence coefficient 0.9 is equal to 0.88. Average classification accuracy is equal to 0.79.

The second image dataset consists of eleven classes: canyon, city, desert1, desert2 (with a different texture), field, forest, lake, mountains, savannah, sea, swamp. In total 282 700 images, each with size 256×256 pixels were used.

Table 3. – Confusion matrix for eleven classes of developed CNN

class	canyon	city	desert1	desert2	field	forest	lake	mountains	savannah	sea	swamp
canyon	**9991**	0	0	0	0	0	8	1	0	0	0
city	0	**8000**	0	0	0	0	0	0	0	0	0
desert1	0	0	**5967**	33	0	0	0	0	0	0	0
desert2	0	0	0	**5000**	0	0	0	0	0	0	0
field	0	0	0	0	**7500**	0	0	0	0	0	0
forest	0	0	0	0	0	**6000**	0	0	0	0	0
lake	0	0	0	0	0	0	**6000**	0	0	0	0
mountains	16	0	0	0	17	0	0	**6959**	0	0	4
savannah	3	1	0	0	0	0	0	0	**7984**	0	0
sea	0	0	0	0	2	0	0	0	0	**6998**	0
swamp	0	0	0	0	0	0	0	0	0	0	**5500**

High recognition quality for the second image dataset is explained by the fact that it was *intentionally* made up of very similar textures of Earth, which, unlike the first image dataset, are completely different from each other.

A comparative analysis of the recognition quality between the developed CNN and the popular neural network architectures: LeNet, AlexNet and GoogLeNet was carried out. The results of the comparative analysis are given in [4] and show that the CNN architecture developed at UIIP NASB is comparable in quality to the popular architectures of deep convolution networks.

Conclusion

A hardware and software system (micromodule) for image classification of the underlying surface of the Earth was developed at UIIP NASB.

Micromodule's hardware has low power consumption, small size and weight. Cost of micromodule components is less than hundred Belarusian rubbles.

Developed convolutional neural network is reliable and is comparable by work quality to the popular deep convolutional neural network architectures. Small image processing time allows to install this software on mobile platforms with a small computing power. However, recognition accuracy depends on the image dataset, which is used to train developed neural network.

In a further development of the hardware and software system authors plan to move from the *image classification* task to the *object detection* task.

Gratitude. This work was carried out with the financial support by the project No. 3.2.4.1 of the Union State programme "Technology-SG".

References

[1] Kovalev V.A., Paulenka D.A., Snezhko E.V., Liauchuk V.A. Comparative analysis of budget computing platforms for a portable micromodule of on-board image classification // BIG DATA and Advanced Analytics: collection of materials of the fourth international scientific and practical conference. (Minsk, Belarus, May 3 – 4, 2018) / editorial board: M. Batura [etc.]. – Minsk, BSUIR, 2018. – pp. 31–42.

[2] Сравнительный анализ вычислительных платформ для бортового микромодуля предварительного распознавания изображений / В.А. Ковалев [и др.] // Информатика. – ISSN 1816-0301 (print). ISSN 2617-6963 (online). 2018. − Т. 15, № 3. – С. 7–21.

[3] MobileNetV2: Inverted Residuals and Linear Bottlenecks, Sandler M, Howard A, Zhu M, Zhmoginov A, Chen L.C. arXiv preprint. arXiv:1801.04381, 2018.

[4] Интеллектуальная технология распознавания подстилающей поверхности Земли / С.В. Кругликов, В.А. Ковалев, Д.А. Павленко, Э.В. Снежко, В.А. Левчук // Радиоэлектронные технологии. – 2019. – № 1. – С. 90–94.

РАСПОЗНАВАНИЕ ПОДСТИЛАЮЩЕЙ ПОВЕРХНОСТИ ЗЕМЛИ С ПОМОЩЬЮ СВЕРТОЧНОЙ НЕЙРОННОЙ СЕТИ НА ОДНОПЛАТНОМ МИКРОКОМПЬЮТЕРЕ

Д.А. Павленко
Инженер-программист 1 кат. ОИПИ НАН РБ

В.А. Ковалев
Кандидат технических наук, зав. лаб. анализа биомедицинских изображений ОИПИ НАН РБ

Э.В. Снежко
Кандидат технических наук, ведущий научный сотрудник ОИПИ НАН РБ

В.А. Левчук
Кандидат технических наук, научный сотрудник

Е.И. Печковский
Ведущий инженер-программист

Объединенный институт проблем информатики (ОИПИ) Национальной академии наук Беларуси (ОИПИ НАН РБ, ул. Сурганова 6, 220012, Минск, Республика Беларусь)

Аннотация. Приводятся результаты разработки программно-аппаратного комплекса (микромодуля) по обнаружению и классификации изображений подстилающей поверхности Земли. Полученное устройство имеет размеры 5,2×7,4×3,1 см, массу 52 г и использует сверточную нейронную сеть на основе архитектуры MobileNetV2 для классификации изображений. Микромодуль может использоваться на борту малых космических аппаратов либо легких беспилотных летательных аппаратов (дронов). Приведенные в статье сведения могут быть полезны инженерам и научным работникам, разрабатывающим компактные бюджетные мобильные системы обработки, анализа и распознавания изображений.

Ключевые слова: распознавание изображений, сверточная нейронная сеть, глубокое обучение, одноплатный микрокомпьютер, edge device, мобильная система