Training set AERIAL SURVEY for Data Recognition Systems From Aerial Surveillance Cameras

Valery Zivakin [1], Oleksandr Kozachuk [1], Pylyp Prystavka [1] and Olha Cholyshkina [2]

[1] National Aviation University, Liubomyra Huzara ave. 1, Kyiv, 03058, Ukraine
[2] Interregional Academy of Personnel Management, Frometivska St. 2, Kyiv, 03039, Ukraine

Abstract

Recognition of terrain elements and objects from aerial image data wich obtained from aircraft air surveillance cameras, UAVs and high-resolution satellite images is a complex task and requires numerical factors to be taken into account. The created training set for intelligent systems AERIAL SURVEY allows solving a wide range of scientific and practical problems in the field of machine learning and neural network recognition of terrain elements. The purpose of this paper is to show the dynamics of the conducted research on the creation of the training set AERIAL SURVEY, identify problem areas and identify promising vectors for further development of this area. The training set already now allows solving a number of video analytics tasks, including dual-use ones: various aerial monitoring of roads, agricultural fields, forests, water bodies, aerial reconnaissance with the search for targets in the combat zone, use in tracking and target designation systems. In the course of further research, the aerial survey data set will allow solving such problems as information filtering of excessive information content of terrain images, automated formalization and description of new unlabeled classes, the transition from recognition of individual terrain elements to a structural description of the aerial survey area in order to automate its georeferencing and many others.

Keywords

Neural network element recognition, video analytics, monitoring, filtering

1. Introduction

The success of the widespread introduction of artificial intelligence technologies over the past decade is largely due to the development of machine learning methods, in particular, such a component as deep learning. That is why the value of data sets, on the basis of which the creation and training of intelligent systems for various purposes, is increasing. In recent years, issues related to data collection and their preparation for learning have already formed into an integral part of a separate area - machine learning engineering. The result of its deployment and implementation depends on how adequate these goals are for the sake of which a machine learning system or model is created. At any time, with machine learning, the researcher is not limited to a linear data processing algorithm. The learning model can be changed, improved depending on the purpose and circumstances of the applied task, and therefore requires the introduction of additional hyperparameters or the involvement of additional data. Of course, we are talking about complex applied problems, the solution of which is faced with the need to solve scientific and technical problems.

One of these tasks is the recognition of terrain elements and objects from aerial photography data obtained from aircraft surveillance cameras, UAVs and high-resolution satellite images as a special case of Earth remote sensing data. The complexity of this task is due to a large number of factors, so here are the most important ones (the order of mention is not critical):

Information Technology and Implementation (IT&I-2022), November 30 - December 02, 2022, Kyiv, Ukraine
EMAIL: Zivakin1993@gmai.com (A. 1); oleksandrkozachukk@gmail.com (A. 2); chindakor37@gmail.com (A. 3); greenhelga5@gmail.com (A. 4)
ORCID: 0000-0002-0420-0558 (A. 1); 0000-0003-3361-6197 (A. 2); 0000-0002-0360-2459 (A. 3); 0000-0002-0681-0413 (A. 4)

• Various shooting heights lead to the necessity to recognize objects at different scales and with different levels of detail;

• Recognition of objects during different seasons, under meteorological conditions and period of the day;

• Different shooting angle of objects creates perspective distortion of recognition targets;

• Interference that occurs when shooting: from aberration or micro-movement of the fixation camera on board the aircraft to distortion from ground-based means, such as electronic warfare;

• The necessity to identify not only objects in the image, but also often the solution of related tasks, such as setting the location of the shooting frame, the speed of processing a large amount of information per unit of time, preferably in near real time, etc.

In the complex solution of the problem of recognition of aerial data, special attention is required to the formation of a high-quality training set, on the basis of which it is already possible to form a problem statement for the creation of various machine learning systems, both surface and deep.

The relevance of the work on creating a training set of aerial survey data is especially important for Ukraine, as an aviation State that actively develops and uses various unmanned arial vehicles. With the outbreak of war with the Russian Federation, the use of recognition systems in the processing of aerial reconnaissance and air surveillance data in the combat zone acquired particular importance. According to the order [1] of the Cabinet of Ministery of Ukraine dated 08.30.2017 No. 600-r "Some issues of the development of critical technologies in the field of production and military equipment" and its Adition 1 as amended on 23.02.2022, in the list of critical technologies in the field of production of weapons and military techniques include but are not limited to:

• Technologies for figurative interpretation, selection and classification of targets for homing systems of high-precision weapons;

• Technologies for controlling robotic platforms, identifying, recognizing and tracking targets;

• Technologies of machine learning, artificial intelligence, neural networks for the design, production and operation of military and special equipment;

• Technologies for coding, transmission and receipt (automatic recognition, processing, analysis, generation, visualization) of information.

As you can see, it is difficult to overestimate the importance of creating, filling and using a training set for recognition of data from aerial surveillance cameras (aerial photography data).

That is why, starting from 2018, the Department of Applied Mathematics of the National Aviation University [2] began work in this area, in particular, the creation and filling of the Aerial Survey training set was started [3, 4]. First of all, the specified data set is considered by the authors as a launching pad for the deployment, research and continuous improvement of machine learning systems. During the creation of the ARIEL SURVEY set, the department's scientists, students and graduate students solved various problems, improving both the training set and the systems and information technologies based on it. Let set the purpose of this paper to show the dynamics of the conducted research, to outline problem areas and to identify promising directions for further development of this area.

2. Analysis of sources and studies

Consider, according to open sources, the existing developments in the field of training sets of aerial observation data (or may be suitable for such processing) and pay attention to deep learning technologies that have the prospect of being used when working with the mentioned sets.

Conventionally, the formation of sets can be divided into two options:

1. Formation of the most universal set for a wider coverage of possible tasks.
2. Formation of a dataset for a specific task.

Of course, the first option is a priority in case of an excess of time, and the second - in case of its absence. So in [5], the Inria dataset is presented, which was used for automatic pixel marking of aerial images. The coverage of 810 km is marked into two classes: structure and non-structure. At the same time, the team "captain Whu" focused on presenting the most representative datasets for solving segmentation problems [6, 7], object detection [8], and ect. It is logical to think that over time, any set needs to be supplemented and verified. Such networks are the most valuable.

In general, a large amount of data is currently freely available. For example, since 2006, the Environment Agency has been publishing sets collected on a plane from a few square kilometers to hundreds. Images are provided as a raster dataset in ECW (extended compressed wavelet), as true color (RGB), near infrared (NIR), or 4-band (RGBN) dataset. If the image was taken under incident response conditions and/or lighting conditions not be optimal, it is identified by the IR prefix. This set is constantly updated (last updated March 28, 2022) [9]. In addition, on the kaggle resource, you can find high-quality data sets for solving various problems, for example, a set for semantic segmentation [10] or Semantic Drone Dataset [11].

To work with a data set, those types of neural networks that are typical in image processing are suitable: convolutional (CNN), with an autoencoder (autoencoder), generative competitive networks (GAN). Thus, it is relevant to process the set using the VGG-16 network, which is an improved version of AlexNet [12], ResNet50 is a powerful base model [13], research with a convolutional autoencoder (Convolutional Autoencoder) [14], which gives various types of representations, as well as processing with complex models like MASK R-CNN [15] and DCNN [16].

The YOLO network [17] is also excellent for real-time object detection. Unlike previous object identification methods repurposed by classifiers for detection, YOLO proposes to use a pass-through method. The neural network simultaneously provides bounding boxes and probabilities of classes. YOLO delivers state-of-the-art results using a groundbreaking approach to object recognition, surpassing previous real-time object detection methods. There are various versions of the neural network, although YOLO does not add a new model architecture to the YOLO family, it does provide a new PyTorch training and deployment system that improves on the state of the art for object detectors.

3. Presenting main material

At the moment, the images in the “Aerial survey” dataset, which are the subject of this publication, are divided into 16 classes and have the structure shown in Fig. 1.

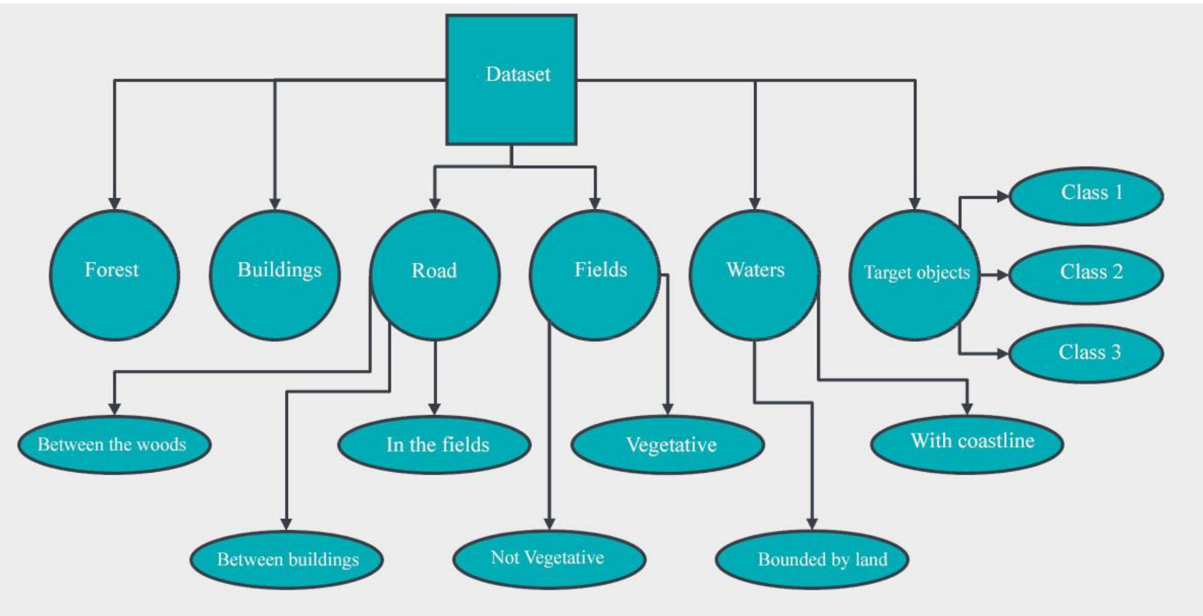


Figure 2: The structure of the “Aerial survey” dataset

In total, the dataset consists of almost eighty thousand 64x64 pixel images. In 2018, the first iteration of the formation of the dataset and work on image processing on it was carried out. At that time, the set consisted of about two thousand images with a 64x64 pixel extension, divided into four classes: water, road, building, target class [18, CNN-technology for recognition of objects for aerial photography data]. The study used a convolutional neural network consisting of two convolutional blocks and a multilayer perceptron. Three experiments were carried out in which the dataset was divided into training and test parts of different sizes, and the neural network was trained for a different number of epochs (50, 60 and 80). In each of them, the training and test data were shuffled randomly five times. The summarized results are presented below in Tab. 1.

Table 1

Results of the first training iteration

		The accuracy of recognition					
		1	2	3	4	5	Average accuracy by class
Classes	Water	0.7241	0.7381	0.7292	0.8723	0.7692	0.76658
	Road	0.5556	0.5556	0.6207	0.7442	0.7857	0.65236
	Building	0.746	0.8814	0.7049	0.7959	0.7759	0.78082
	Targeted	0.9107	0.7632	0.7609	0.8	0.7778	0.80252
Average accuracy		0.7341	0.7345	0.7039	0.8031	0.77715	0.75057

As of 2019, the dataset has been updated to 16,000 images divided into eleven classes: buildings, forests, fields, water bodies, roads, vehicle traces, civil engineering, poles, and three target classes. In addition, the study was also carried out on a new type of networks - a convolutional autoencoder with a classifier [18, Automated object recognition system based on aerial survey]. The results of the classifier, in the most general form, are presented in the diagram (Fig. 3) The results of the classifier are presented in more detail in the diagram (Fig. 2).

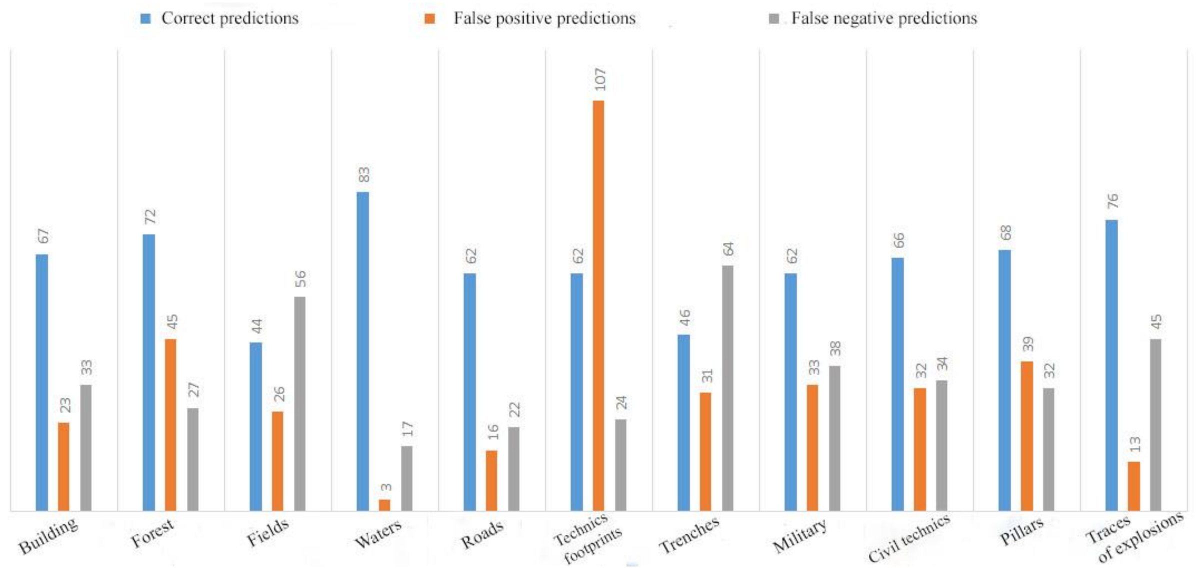


Figure 2: Recognition results by class [19]

Also, work continued with the convolutional network - the classifier. The number of images was additionally added and augmented (up to 16 thousand), and the classes were redistributed as follows: buildings, buildings between trees, forest, roads, roads between buildings, roads between trees, vegetation field, non-vegetation field, waters between trees, technics footprints, trenches, military and civil technics[18, Neural network recognition of image classes from aerial survey]. The work done, in general, contributed to the quality of recognition (Fig. 3).

In 2020, the number of images in the studies was increased to 22,000. In addition, as the analysis of the results of previous recognition works (Fig. 2, 3) shows, the “technics footprints” class, which is closely related to the recognition of roads, in particular dirt roads, required special attention. and different paths. Therefore, in [18, Information technologies for recognizing road classes based on aerial survey data], new elements of the dataset with different types of roads were created, about 40 thousand of the total number of images. Subsequently, a more detailed analysis of representations was carried out for architectures based on the autoencoder model. In the same work, for the representations obtained from the encoder, the principal component method was applied, according to which it turned out that the first three components contain more than 50% of the variability, and the first of them accounted for 28.9%, the second 12.3%, the third 8.9%.

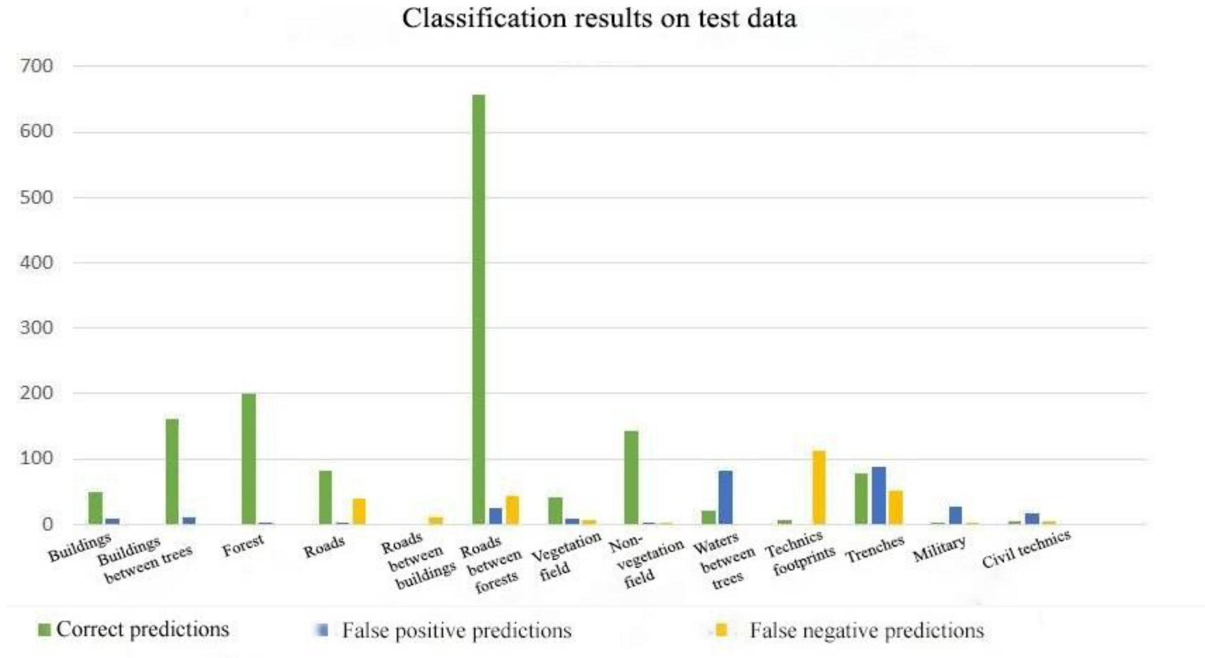


Figure 3: Classification results of augmented dataset 2019 [20]

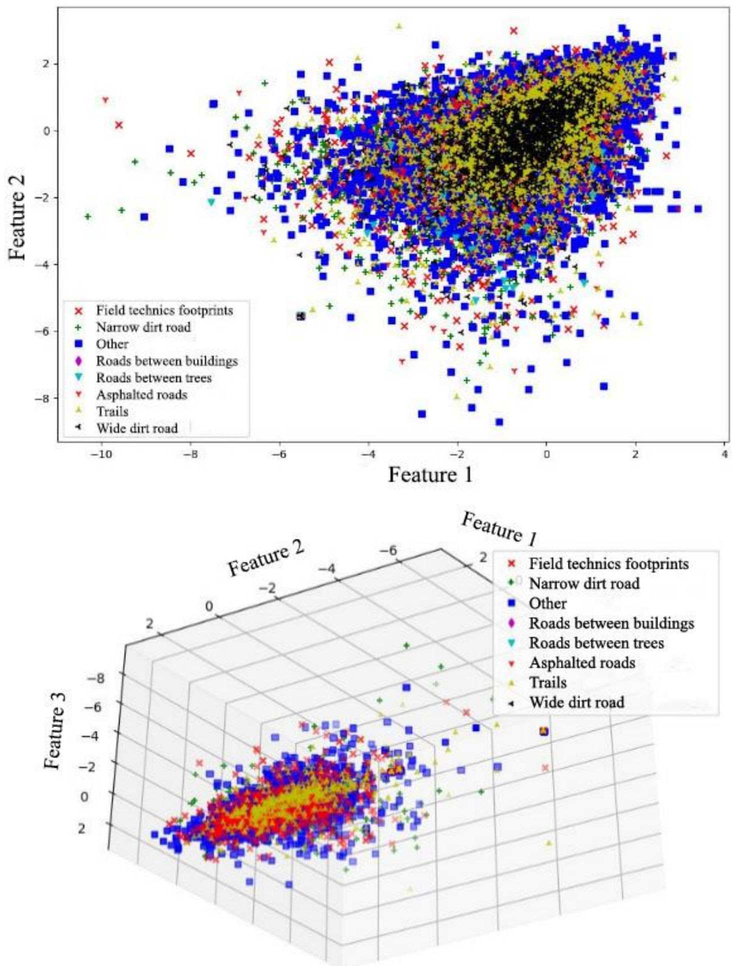
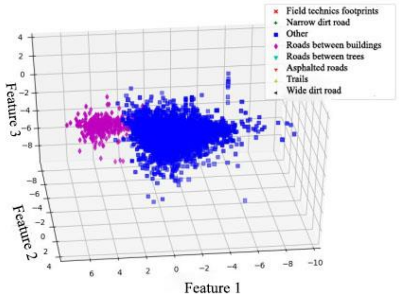
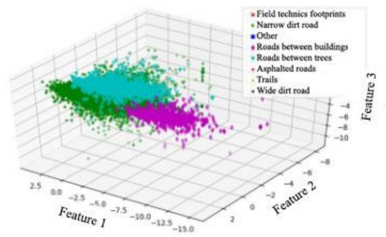
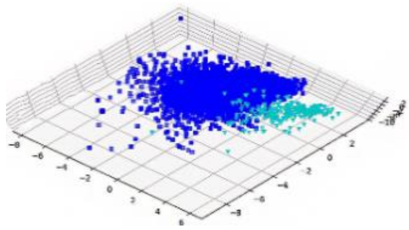
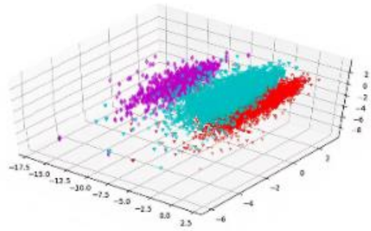
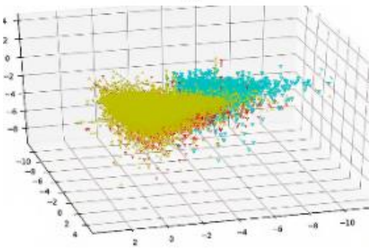


Figure 4: 2D and 3D Data Distribution Model in Principal Component Space

Through the application of the principal components method, a two-dimensional and three-dimensional model of data variability was built (Fig. 4). In the course of testing the classifier on different types of roads, the histogram of the top and bottom transitions was taken away (Fig. 5). In addition, the possibility of presenting road classes in the representation space to the autoencoder in the form of actually linearly separated sets was demonstrated (Tab. 2)

Table 2

Separation of classes among themselves in the space of representations

Class name	Picture
“Other” - blue color; "Roads between buildings" - pink color.	
"Roads between trees" - a shade of green; "Narrow dirt road" - green color; "Roads between buildings " - purple color.	
"Other" - blue color; "Roads between trees" - a shade of green.	
"Roads between trees" - a shade of green; "Asphalt roads" - red color; "Roads between buildings " - purple color.	
"Roads between the trees" - a shade of green; "Trails" - yellow color; "Asphalt roads" - red color.	

An interesting experience was the study of the generative network model (GAN). About twenty thousand images were processed by the network, after which a new sample was generated [18, Technology of neural network generation of a training set of digital images]. Both samples during the research served as both training and test for the convolutional network of the classifier. As can be seen from Table 3, the classifier showed the best recognition accuracy when trained on data generated by the GAN.

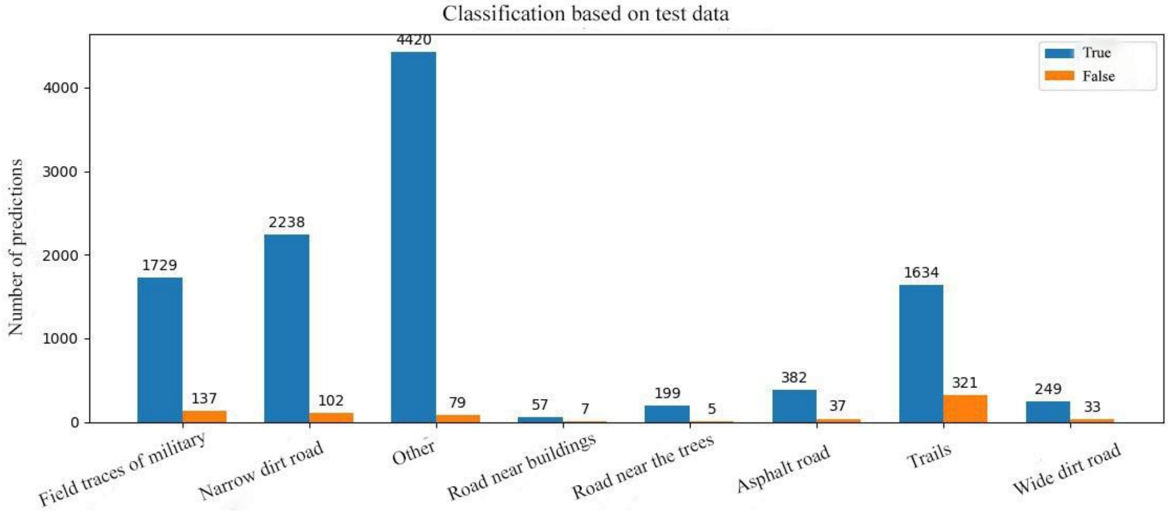


Figure 5: Histogram of true and false transfers

In addition, the possibility of presenting road classes in the representation space to the autoencoder in the form of actually linearly separated sets was demonstrated (Tab. 2).

Table 3
Accuracy of recognition, when using different samples

		Dataset for verification (test)	
		Real data	Generated data
Dataset for training	Real data	95%	80%
	Generated data	98%	83%

In addition, studies of individual classes were carried out. For example, the analysis of the recognition of specific forest plantations and individual trees [18, Automation of filling the training sample according to aerial observation data]. This work is interesting because the qualitative recognition of just such a class makes it possible to automate the calculation of forest plantations, as well as more reliable recognition of "mixed" classes, for example: buildings among trees or trees (forest plantations) along roads. In 2021, the data from the set was used to train modern architecture - MASK R-CNN. The network was used to detect buildings on live video. Positive results were obtained, even with the degree of Jaccard (IoU) 0.8, accuracy and completeness of recognition reached approximately 0.6.

Also at this stage, studies were carried out to identify data that were not formalized in the training set. That is, images with objects that were not present in the training set were attached to the test sets [18, Identification of aerial survey data classes not formalized in the training set]. The result of 36 experiments was demonstrated to add a fourth (unknown) to the neural network classifier trained in three classes. On average, for the case of adding one non-formalized class to three, its identification was carried out with an accuracy of ~68%. And in the case of adding one formalized class to four classes, approximately 63%.

Research continued on the representations obtained by using the encryption part of the autoencoder. In particular, the representations were subject to analysis, in which they were clustered by the k-means algorithm into a different number (from 10 to 17) of newly formed classes, compared with the initial eleven [18, Information technology for modeling the initial datasets of aerospace surveys based on neuromeasurement architectures with an autoencoder]. As an encoder architecture that solves the problem of data dimensionality reduction, the following network was built (Fig. 6).

The refore, using the PCA, the following representations were obtained (see Fig. 7). After applying the PCA, it was found that the first 10 principal components account for 74% of the data variability. The use of PCA made it possible to carry out cluster analysis in the feature space. An analysis of the dependence of the choice of the optimal clustering parameter and repeated retraining of the network on the resulting clusters was carried out. It is shown that using this approach, we obtain a comparable level of recognition - up to 97% on the test sample, and this opens up the possibility of researching approaches in the field of building self-learning systems.

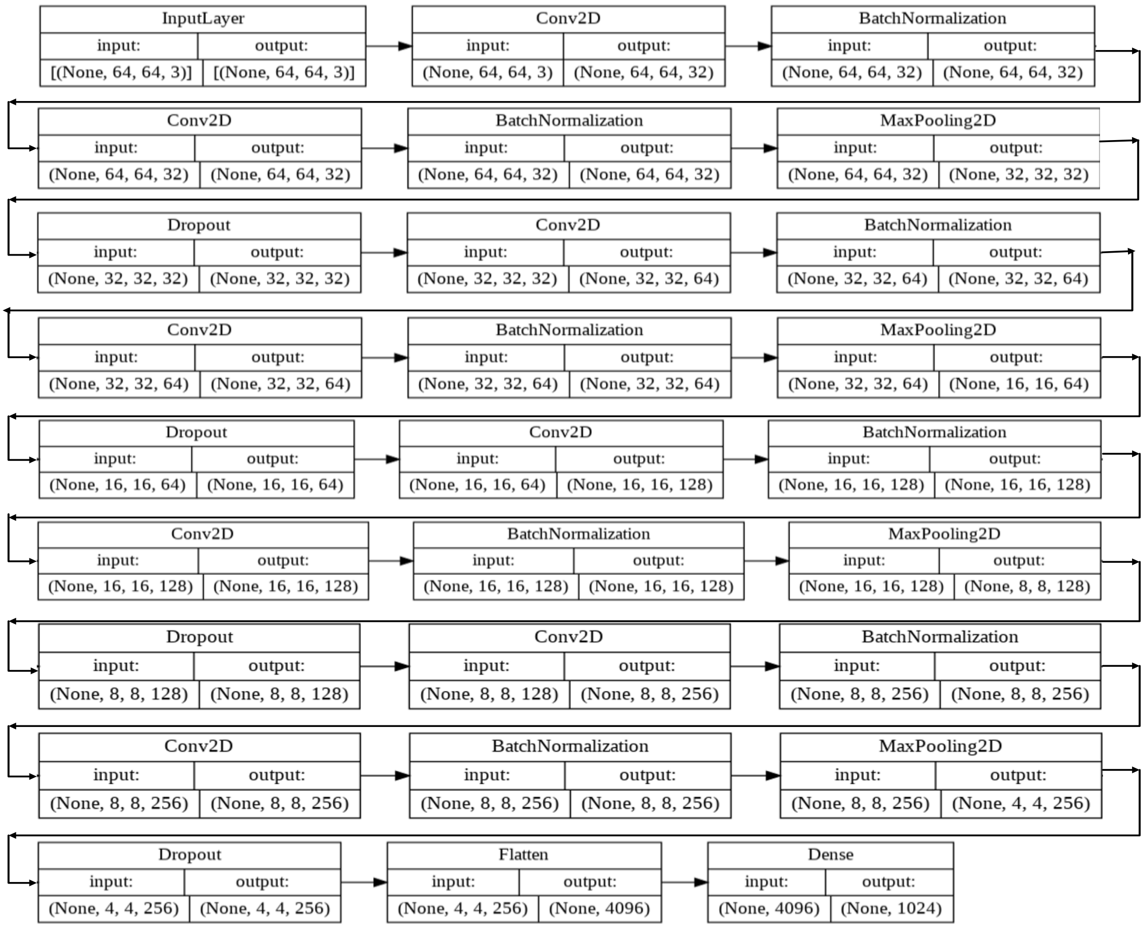


Figure 6: Used encoder layers (Decoder layers are respectively inversely proportional to encoder layers)

In several works, attention was paid to automating the filling of the training set of images. In the master's work [18, Information technology for classification of terrain types and object search based on aerial survey data], the idea was to carry out the initial segmentation of images, after which the selected clusters were used to automatically generate new classes for the training set. The disadvantage of this approach is that the training data obtained in this way must be reviewed by a human specialist and assigned a name to the class. However, as the practice of applying this approach has shown, in many cases this approach has significantly increased the number of images for such types of images: forests, fields, buildings, large water bodies. A more adequate automation of data set filling was demonstrated in a thesis [18, Information technology for automating the filling of the training set for neural network recognition]. In this case, an automated application was created that used the results of neural network training to process aerial data. Recognized image fragments were automatically marked up and could be added to an existing set of training images. Testing has shown that this approach reduces the cost of data labeling and improves the quality of recognition after training on the updated training set.

4. Conclusions and directions for further research

As even a brief review of the work performed shows, the availability of the Aerial Survey training set allows a wide range of scientific and practical research in the field of machine learning and neural network recognition of terrain elements. On the applied side, the presence of such a set already now allows solving a number of video analytics tasks, including dual-use ones: various aerial monitoring of roads, agricultural fields, forest plantations, water bodies, conducting aerial reconnaissance with the search for targets in the combat zone actions, use in tracking and target designation systems.

In addition to the indicated practical, the authors of this work also have a number of purely scientific interests and plans for working with the set. Let us outline some of them and a general perspective for further research.

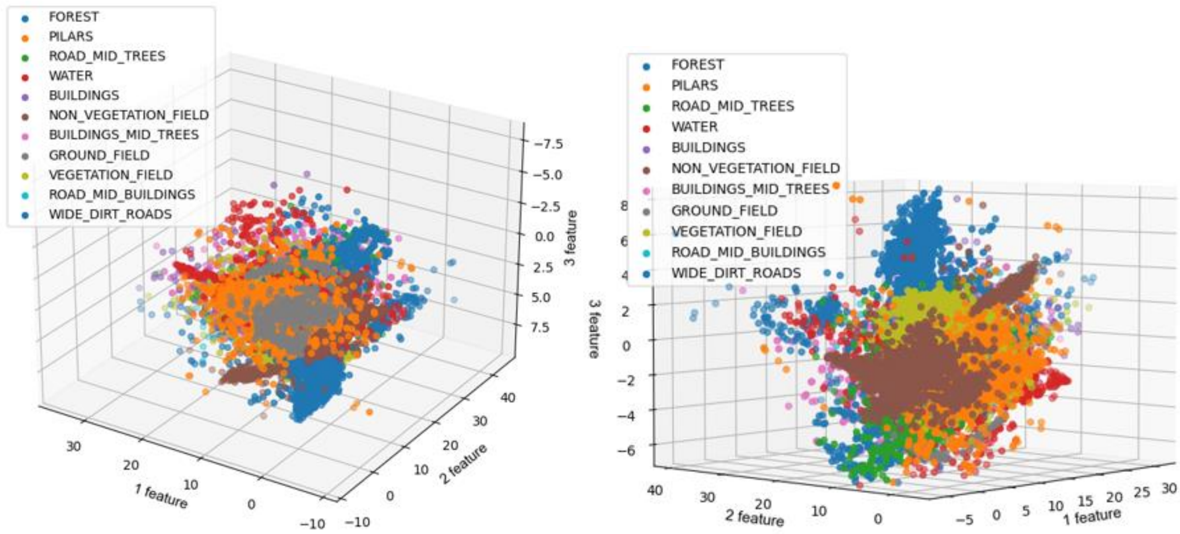


Figure 7: Space representation of the first three principal components on the initial set

If we strive for the universality of this training set, then we should take into account the need to label data for different seasons, which will lead to significant data heterogeneity within some classes, for example, trees, roads, fields, etc. In addition, even within Ukraine, the structure and elements of the landscape, vegetation, and even buildings are noticeably different. Increasing the class options presented in the set requires research into the impact on the quality of training and recognition. Perhaps a separate study will be the selection of some "core" of base classes within the overall architecture of the training set, using for "additional learning" classes of images characteristic of a particular season or type of terrain. In this context, the experience of automatic replenishment of the dataset, which can be performed on the basis of a stable training “core”, with adjustment to a specific monitoring task, is useful. Another area of research may be work on the generation of artificial data, on the principle of generative networks, or on the basis of research and use of representations in the autoencoder space. Working with low-dimensional representations also allows the study and primary analysis of the distributions of data presented in individual classes. The influence of heterogeneity or the presence of anomalies on the learning process is quite an interesting and promising area of research.

Information filtering of excessive information content of terrain images, automated formalization and description of new, unlabeled classes, the transition from the recognition of individual terrain elements to the structural description of the aerial survey area in order to automate its georeferencing - this is not a complete list of tasks that can be performed with the “Aerial survey” dataset in further studies.

5. Acknowledgements

The authors express their gratitude to the Aerorozvidka public association [4] for supporting the development of modern information technologies in Ukraine.

6. References

[1] Order of the Cabinet of Ministers of Ukraine, 2022. URL: https://zakon.rada.gov.ua/laws/show/600-2017-%D1%80#Text.

[2] National Aviation University, 2022.URL:http://applmaths.nau.edu.ua/index.php.
[3] Training data set "Aerial Survey". URL: https://drive.google.com/file/d/1BAmSRbYUyCnrPYn-jpHI7l_6qsNmc9o6/view.
[4] Aerial reconnaissance. Unmanned aerial vehicles, situational awareness, cybersecurity, 2022. URL: https://aerorozvidka.xyz/uk/
[5] E. Maggiori, Y.Tarabalka, G.Charpiat, and P. Alliez, Can Semantic Labeling Methods Generalize to Any City? The Inria Aerial Image Labeling Benchmark, IEEE International Symposium on Geoscience and Remote Sensing (IGARSS), Fort Worth, USA, 2017. URL: https://hal.inria.fr/hal-01468452/document.
[6] Xin-Yi Tong, Gui-Song Xia, Qikai Lu, Huangfeng Shen, Shengyang Li, Shucheng You, and Liangpei Zhang, Land-Cover Classification with High-Resolution Remote Sensing Images Using Transferable Deep Models, Remote Sensing of Environment, China, 2020. URL: https://arxiv.org/pdf/1807.05713.pdf.
[7] Syed Waqas Zamir, Aditya Arora, Akshita Gupta and others, iSAID: A Large-scale Dataset for Instance Segmentation in Aerial Images, Computer Vision and Pattern Recognition Conference (CVPR), Long Beach, USA, 2019. URL: https://openaccess.thecvf.com/content_CVPRW_2019/papers/DOAI/Zamir_iSAID_A_Large-scale_Dataset_for_Instance_Segmentation_in_Aerial_Images_CVPRW_2019_paper.pdf.
[8] Gui-Song Xia, Xiang Bai, Jian Ding, Zhen Zhu, Serge Belongie and others, DOTA: A Large-scale Dataset for Object Detection in Aerial Images, 2019. URL: https://arxiv.org/pdf/1711.10398.pdf.
[9] Vertical Aerial Photography, Environment Agency, 2022. URL: https://data.gov.uk/dataset/4921f8a1-d47e-458b-873b-2a489b1c8165/vertical-aerial-photography
[10] Semantic segmentation of aerial imagery, 2020. URL: https://www.kaggle.com/datasets/humansintheloop/semantic-segmentation-of-aerial-imagery.
[11] The Semantic Drone Dataset, Institute of Computer Graphics and Vision, 2022. URL: https://www.tugraz.at/index.php?id=22387
[12] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, International Conference on Learning Representations (ICLR), San Diego, USA, 2015. URL: https://arxiv.org/pdf/1409.1556.pdf.
[13] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep Residual Learning for Image Recognition, Computer Vision and Pattern Recognition Conference (CVPR), Las Vegas, USA, 2016, 770 – 778. URL:https://www.cvfoundation.org/openaccess/content_cvpr_2016/papers/He_Deep_Residual_Learning_CVPR_2016_paper.pdf.
[14] Yifei Zhang, A Better Autoencoder for Image: Convolutional Autoencoder, 1st ANU Bio-inspired Computing conference, Canberra, Australia, 2018. URL: http:// users.cecs.anu.edu.au/~Tom.Gedeon/conf/ABCs2018/paper/ABCs2018_paper_58.pdf.
[15] K. He, G. Gkioxari, P. Dollar, and R. Girshick, Mask R-CNN, Computer Vision and Pattern Recognition Conference (CVPR), Honolulu, Hawaii, 2017, 2961-2969. URL: https://openaccess.thecvf.com/content_ICCV_2017/papers/He_Mask_R-CNN_ICCV_2017_paper.pdf.
[16] Lars Lien Ankile, Morgan Feet Heggland, and Kjartan Krange, Deep Convolutional Neural Networks: A survey of the foundations, selected improvements, and some current applications, 2020. URL: https://deepai.org/publication/deep-convolutional-neural-networks-a-survey-of-the-foundations-selected-improvements-and-some-current-applications.
[17] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, You Only Look Once: Unified, Real-Time Object Detection, 2016. URL: https://doi.org/10.48550/arXiv.1506.02640
[18] Research of the Department of applied mathematic for young seintists. National Aviation University (NAU), Kyiv, Ukraine, 2018-2022. URL: http://applmaths.nau.edu.ua/category-view.php?cat=diplomas, http://applmaths.nau.edu.ua/show.php?id=281
[19] Gnatyuk, V. A. (2001). Mechanism of laser damage of transparent semiconductors. Physica B: Condensed Matter, 308-310, 935-938. doi:10.1016/S0921-4526(01)00865-1.
[20] Dudnik, A., Presnall, B., Tyshchenko, M., & Trush, O. (2021). Methods of determining the influence of physical obstructions on the parameters of the signal of wireless networks. Paper presented at the CEUR Workshop Proceedings, , 3179 227-240.
[21] Latsyshyn, A., Iatsyshyn, A., Kovach, V., Zinovieva, I., Artemchuk, V., Popov, O. Turevych, A. (2020). Application of open and specialized geoinformation systems for computer modelling studying by students and PhD students. Paper presented at the CEUR Workshop Proceedings, , 2732 893-908