UDC 004.02
DOI: 10.15587/1729-4061.2023.272951

RECOGNITION OF AERIAL PHOTOGRAPHY OBJECTS BASED ON DATA SETS WITH DIFFERENT AGGREGATION OF CLASSES

Pylyp Prystavka
Doctor of Technical Sciences, Professor*
Kseniia Dukhnovska
PhD**
Oksana Kovtun
Corresponding author
PhD, Associate Professor**
E-mail: kovok@ukr.net
Olga Leshchenko
PhD, Associate Professor
Department of Networking and Internet Technologies ***
Olha Cholyshkina
PhD, Associate Professor
Department of Computational Mathematics and Computer Modeling
Interregional Academy of Personnel Management
Frometivska str., 2, Kyiv, Ukraine, 03039
Vadym Semenov*
*Department of Applied Mathematics
National Aviation University
Lubomyra Huzara ave., 1, Kyiv, Ukraine, 03058
Department of Programming and Computer Equipment*
***Taras Shevchenko National University of Kyiv
Volodymyrska str., 60, Kyiv, Ukraine, 01601

The object of this work is the recognition algorithms of aerial photography objects, namely, the analysis of recognition accuracy based on data sets with different aggregation classes.

To solve this problem, an information system for object recognition based on aerial photography data has been developed. An architecture based on neural network architectures of the ConvNets group with structural modifications was chosen and used to create the information system. The use of a convolutional neural network of the ConvNets group in the architecture of the information system for the recognition of objects of aerial photography gives high accuracy rates when training the information system and validating its results. But the authors did not find any studies on the learning of the neural network of the ConvNets group. Therefore, it was decided to conduct an analysis in which case the ConvNets network will provide validation results with higher accuracy when the training takes place on datasets with or without class aggregation.

The authors performed an analysis of the accuracy of recognition of aerial photography objects based on data sets with different aggregation classes. The dataset used for neural network training consisted of 3-channel labeled images of 64x64 pixels size. Based on the analysis, the optimal number of epochs for training is selected, which makes it possible to recognize aerial photography objects with greater accuracy and speed. It was concluded that greater accuracy in image classification is achieved for sampling without crossing data from different classes (without aggregation of classes). The result of the work is recommended for use in the automation of dataset filling and information filtering of visual images

Keywords: recognition of aerial photography objects, classification of data sets, recognition accuracy, neural network of the ConvNets group

Received date 16.11.2022
Accepted date 19.01.2023
Published date 28.02.2023

***How to Cite**: Prystavka, P., Dukhnovska, K., Kovtun, O., Leshchenko, O., Cholyshkina, O., Semenov, V. (2023). Recognition of aerial photography objects based on data sets with different aggregation of classes. Eastern-European Journal of Enterprise Technologies, 1 (2 (121)), 6–13. doi: https://doi.org/10.15587/1729-4061.2023.272951*

1. Introduction

The rapid development of intelligent human activity in the XXI century contributes to the development of various technologies, primarily digital. Increasing the equipment and a variety of gadgets requires software improvement. Although machine learning dates back to the second half of the last century, the development of this industry was not possible in the absence of appropriate technologies. Over the past 10 years, the development of computer vision based on neural network pattern recognition has been unique.

The capabilities of modern information technologies make it possible to recognize not only static objects but also do it in real time. Even modern smartphones can already recognize fingerprints, the face of a person with a variety of emotions, and some even the retina. Object recognition is the basis of many different identification programs that are practically used in many industries such as technical diagnostics, medical diagnostics, letter recognition, speech, aerial photography recognition, military sphere, and others. Technologies for recognizing the results of aerial photography are actively used, namely during the analysis of the results of aerial reconnaissance of objects monitoring at strategic, operational, and tactical levels, monitoring of the earth's surface, air and water space, climate and environmental control, the agricultural sector, etc.

The development and research of algorithms that provide an opportunity to recognize objects of aerial photography is an urgent task in recent years.

2. Literature review and problem statement

Computer vision technologies are used in many branches of human activity. For example, work [1] reports the results of research on the information technology of computer vision for remote measurement of symptoms of schizophrenia using smartphones. Study [2] shows the use of computer vision technologies to measure the color of mozzarella cheese. Research using computer vision technologies in [3] makes it possible to obtain reliable data for the analysis of physiological and environmental indicators of the soybean crown. The variety of application of information technologies of computer vision is proved by a large number of scientific works, such as [1–3].

But it should be noted that there is no universal mathematical apparatus that would make it possible to form a general formalized approach to the construction of computer vision systems. Currently, a large number of scientific papers address the development of information technologies of computer vision. In [4], the results of experimental studies into the dependence of informativeness of images on the results of overlaying previous filters for processing digital images, depending on the values of the parameters of the methods, are given. In this paper, it is established that the use of analysis algorithms using a sliding window can significantly increase the resolution of analysis in the time domain while maintaining a sufficiently high ability in the frequency domain.

In works [5, 6], the use of filters used to suppress pulsed noise is considered. Among such filters, special attention is paid to the median filter. Its simplicity and ability to maintain edges have led to widespread use in the field of image processing and computer vision. However, issues such as moderate to high uptime of the standard median filter algorithm and relatively low performance when the image is strong damaged by impulse noise led to the development of several variants of the algorithm. One set of algorithm options focuses on creating quality results while the other set is aimed at reducing work time.

The development of computer vision algorithms through the use of neural networks attracts great attention of scientists. In works [7, 8], the use of an artificial neural network of counter-propagation (CP-ANN), which has such advantages as the ability to learn and classify, is considered. CP-ANN neural networks still have some limitations in pattern recognition tasks when they encounter ambiguity during the learning process, leading to incorrect classification of the self-organizing Kohonen map (K-SOM). This issue affects CP-ANN performance. Therefore, this paper proposes a new strategy for improving CP-ANN according to the Gram-Schmidt algorithm (GSHM) as a stage of preprocessing of the source data without changing its architecture. For experimental testing, three examples of datasets from different areas were used, such as correlation, yield, and fertilizer. To get the results, the work relies on two simulations. The first simulation uses CP-ANN, and datasets are entered into the network without prior processing. The second simulation uses a modified artificial counter-propagation neural network (MCP-ANN), and the datasets are pre-processed through the GSHM unit. The results of the experiment show that the proposed MCP-ANN recognize all models with a classification accuracy of 100 % versus 62.5 % for CP-ANN in the correlation data set. In addition, the MCP-ANN proposed in the paper reduce the lead time and learning parameter values in all datasets compared to CP-ANN. Thus, the proposed approach based on the GSHM algorithm significantly improves CP-ANN performance.

In [9, 10], the algorithms of deep convolutional neural networks (CNN) in various problems of computer vision are investigated, their potential for classifying images of multispectral remote sensing has not been thoroughly studied. In particular, the application of deep CNN using optical remote sensing data has focused on classifying air and satellite data at very high resolution due to the similarity of this data to large ones datasets in computer vision. Accordingly, this study presents a detailed study of state-of-the-art deep learning tools for classifying complex wetland classes using RapidEye multispectral optical images. In particular, the paper explores the capacity of seven well-known deep networks, namely DenseNet121, InceptionV3, VGG16, VGG19, Xception, ResNet50, and InceptionResNetV2, for wetland mapping in Canada. In addition, the classification results obtained from deep CNN are compared with results based on conventional machine learning tools, including Random Forest and Support Vector Machine, to further evaluate the effectiveness of the former for classifying wetlands. The results show that full convnet training using five spectral ranges is superior to other strategies for all networks. InceptionResNetV2, ResNet50 and Xception distinguish as the three largest convnets, providing the most modern classification accuracy of 96.17 %, 94.81 %, and 93.57 %, respectively. The classification accuracy obtained using the support vector machine (SVM) and random forest (RF) is 74.89 % and 76.08 %, respectively, which is significantly inferior to CNN.

As can be seen from the above, scientific research on computer vision algorithms is timely.

3. The aim and objectives of the study

The aim of this work is to analyze the accuracy of recognition of aerial photography objects based on data sets with different class aggregation. Based on the analysis, the optimal division into classes of input data is determined. This will make it possible to recognize aerial photography objects with greater accuracy and speed.

To achieve the set aim, the following tasks have been solved:

– to develop a mathematical model to determine the accuracy of the classification of aerial photography objects;

– on its basis to implement information technology for the classification of aerial photography objects using neural networks.

4. The study materials and methods

The object of study: processes of recognition of classes of digital images. The subject of research: neural network methods for recognizing classes of digital images according to aerial photography.

Software for information technology was implemented on a computer with a processor Core^{TM} i9 11900F and RAM of 32 GB. An image is fed to the input of the software, at the output a class is obtained to which the input image belongs.

The dataset used to train a neural network consists of 3-channel marked images of 64x64 pixels in size. In general, the dataset consists of the following classes of aerial photography data objects: houses and trees, houses and roads, trees and roads, forests, non-vegetation fields, vegetation fields, earthen fields, water. In total, the datset consisted of 10,000 images. Among them: 5000 is a training sample, 2500 is a test sample, 2500 is a control sample.

The data is distributed unevenly across classes, so in classes in which too many images some of the images were removed, and in classes in which there were very few images, the number of the latter was increased by augmentation of part of the images.

Fig. 1 shows examples of elements of a data set from the class “houses and trees”.

The complete set of data classes consists of 9 classes (Fig. 2).

To obtain aggregated classes, the class “houses and trees” and “houses and roads” were combined into the class “houses”. Classes “houses and trees”, “trees and roads” and “forests” were combined into the class “trees”. The class “non-vegetation fields”, “vegetation fields” and “earthen fields” were combined into the class “fields”. The classes “houses and roads”, “trees and roads” and “earthen roads” were merged into the class of “roads”.

An aggregated set of data classes consists of 5 classes (Fig. 3).

Fig. 4 shows examples of data set elements from the aggregated class “trees”.

The input set is not evenly distributed. Distribution of the input data of the initial sample by class is shown in Fig. 5.

To increase the average accuracy of recognition and speed up learning, 300 images from each of the classes were selected. For aggregated classes, 900 images were selected. Since for the aggregated class of water there were not enough images, some of the images were augmented.

Augmentation is an artificial increase in the number of images by modifying (threshold to an arbitrary angle, reflection (horizontal or vertical), scaling, or a combination of these transformations) of the original image.

The input data to the network is the image size (64,64,3), the first two values correspond to the size of the image (height and width), the third value corresponds to the number of channels in the image.

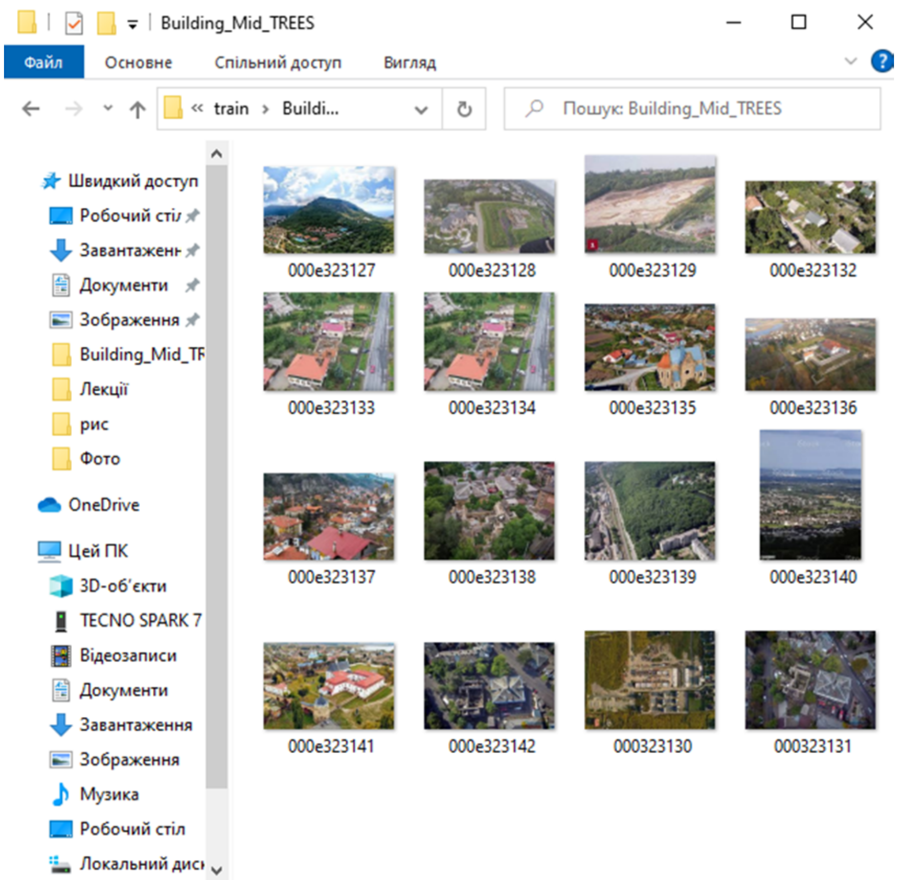

Fig. 1. Examples of dataset

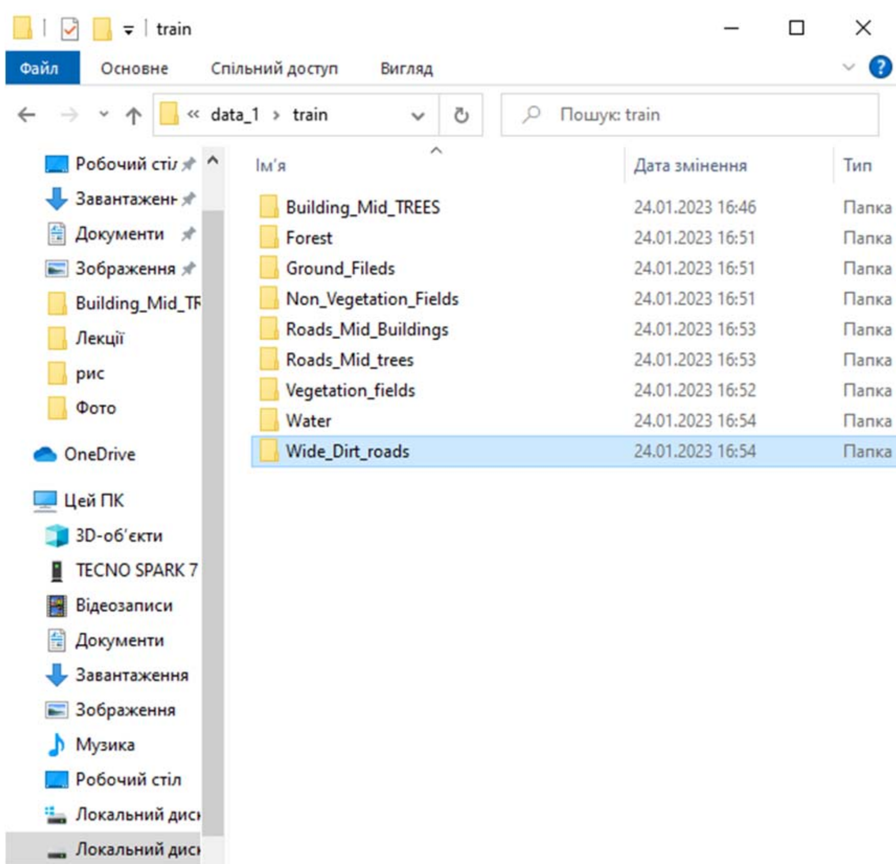

Fig. 2. A complete set of data classes

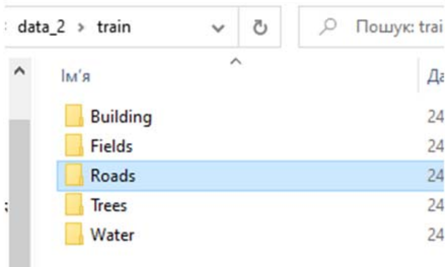

Fig. 3. Aggregated data classes

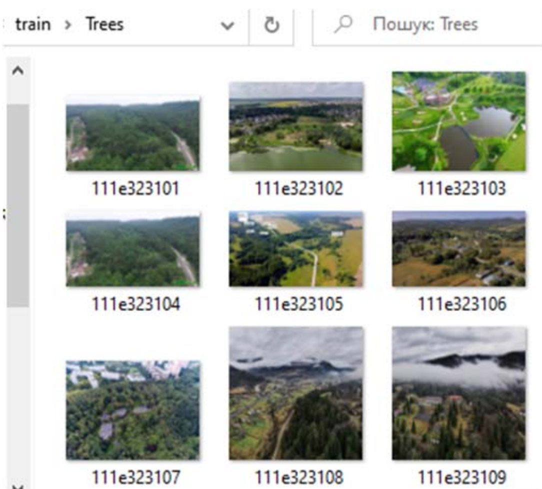

Fig. 4. Example of images from an aggregated data class "trees

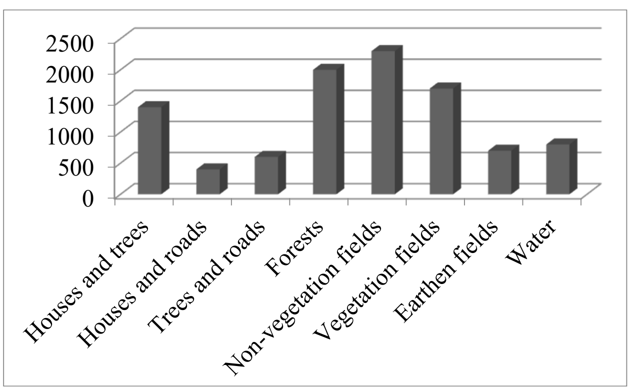

Fig. 5. Distribution of classes of the initial sample

5. Results of research using the information system for recognizing aerial objects

5. 1. Mathematical model for classification of aerial photography objects

Let a set of images be given

$$\Omega_{N,n}=\left\{X_l : l=\overline{1,N}\right\}=\left\{X_l : \left(x_{l,1},x_{l,2},\ldots,x_{l,n}\right)\right\},$$

where N is the number of images in the set $\Omega_{N,n}$;

n – the number of attributes by which the image is recognized (classified);

X_l – l-th image, $l=\overline{1,N}$;

$(x_{l,1}, x_{l,2}, \ldots, x_{l,n})$ – n-dimensional vector of signs of the l-th image, $l=\overline{1,N}$;

x_{li} – i-th attribute, l-th image, $l=\overline{1,n}$, $l=\overline{1,N}$;

Each image must be recognized, that is, attributed to a specific class. Therefore, you need to determine the classification function:

$$f\left(S_k,X_l\right)\rightarrow[0;1], \tag{1}$$

S_k is a class, $k=\overline{1,K}$; where K is the number of classes to which an image can belong.

The classification function $f(X_l, S_k)$ takes the value 1 when the image X_l belongs to the class S_k (recognized as an object of class S_k), in the case of another case it will be equal to 0.

The number of images in the test sample is N. After recognizing these images, we denote as N_p the number of images correctly attributed to the class, N_f – the number of images mistakenly attributed to classes. Then the recognition accuracy will be defined as:

$$T=\frac{N_p}{N}. \tag{2}$$

The loss V shall be called the value determined by the formula:

$$V=\frac{N_f}{N}. \tag{3}$$

The problem of image recognition is to find such a function $f(X_l,S_k)$ so that the accuracy T is maximum.

5. 2. Information technology for the classification of aerial photography objects using neural networks

It is difficult to analytically determine the function $f(X, S)$ from (1), so it was decided to build it by means of neural networks.

To create information technology, an architecture based on networks and the ConvNets group with modifications in the structure was selected and used [11] (Fig. 6). This network has proven effective for work with images and generally contains 184457 neurons.

As can be seen from Fig. 6, the network consists of an input layer, a full-linked output layer, three subdiscretizing layers, three convolutional layers, and two aggregation layers. In general, the recognition of aerial photography objects using the neural network of the ConvNets group consists of a set of similar stages, each stage has three layers in its composition – a convolution layer, an activation layer, and a merging function. Convolution to solve the problem of recognition of objects of aerial photography occurs using the core 3×3.

Convolutional layer outputs are inputs for subdiscretizing network layers. The task of these layers are the operations of combining and reducing the dimensionality of the feature vector by taking the maximum value.

After entering the input layer, the following transformations are performed: ReLU-activation of serial convolution with cores 3×3, max – pooling subdiscretization and, normalization of packets. Due to the normalization of packets, the following is achieved: faster convergence of the model; independence in the training of network layers; the pace of learning increases because packet normalization ensures that the outputs of the nodes of the neural network will not accept very large or very small values; there is a decrease in the dependence on the initial initialization of weights.

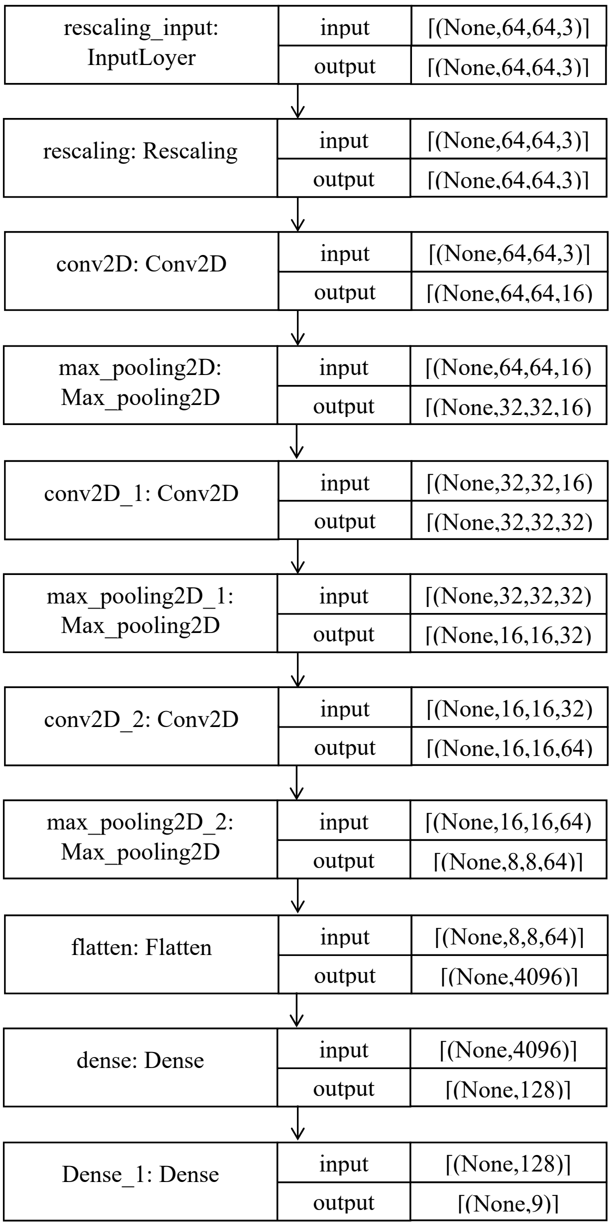

Fig. 6. Network architecture

This is followed by successive 3×3 convolutions to highlight image features, max – pooling subdiscretization to seal the resulting feature maps. ReLU was chosen as the activation function because it provides the highest results in terms of model accuracy. ReLU can be written as:

$$f(x)=\max(0,x),$$

where $f(x)$ is the output signal;

x – input signal.

A piece of code that implements this neural network architecture is shown in Fig. 7.

The scheme of the information system is shown in Fig. 8.

First you need to break the input data into classes. Then, for each class, you need to check if there are too many images in it. If there are too many images, you need to remove the extra images. If there are not enough images, you need to augment the image. Over time, it is necessary to build a neural network, to train it on the processed data.

Neural network training took place according to the algorithm of the fastest descent, this algorithm involves several steps. The first step is to set the weight of neurons with random values. The second is the calculation of the weights of the neural network by direct propagation. Initial data is fed to the network input and the output signals of neurons of all layers of the network are calculated, After that, the differences between the obtained and expected values at the output are calculated, thus error signals are formed. Next, you need to make a return move: the effect of the error signal on preceding neurons. At this step, an error surface gradient vector is formed, which will indicate the direction of the shortest descent along the error surface from a given point in order to reduce the error. This whole cycle of learning is called the epoch. At the beginning of each epoch, all images of the objects of aerial photography of the training sample are sequentially fed to the input of the neural network, the output results of the neural network are compared with the expected result. An error is calculated based on these results. A big error entails a larger gradient. Training occurs until the total error is less than the given value.

In the process of learning, it is necessary to collect metrics and save them to a file for further processing.

Then it is necessary to check the results on a test sample and save the metrics of the test sample for their further processing.

As a result, the information system makes it possible to visualize learning outcomes and validation. So, Fig. 9, 10 show the dependence of the mean quadratic deviation σ and the mathematical expectation M of the functions of accuracy of learning and validation on epochs for data without class aggregation (Fig. 9) and with class aggregation (Fig. 10).

The dependence of the standard deviation σ and the mathematical expectation M of the loss function for learning and validation on epochs, for non-aggregated classes, is shown in Fig. 11, and with class aggregation – in Fig. 12.

```
Model = Sequintial([
layers.experimental.preprocessing.Rascaling(1./255, input_shape=(img_height,
img_width,3)),
layers.Conv2D(16, 3, padding = 'same', activation = 'relu'),
layers.MaxPooling2D(),
layers.Conv2D(32, 3, padding = 'same', activation = 'relu'),
layers.MaxPooling2D(),
layers.Conv2D(64, 3, padding = 'same', activation = 'relu'),
layers.MaxPooling2D(),
layers.Flatten(),
layers.Dense(128, activation = 'relu),
layers.Dense(num_classes)
])
```

Fig. 7. Code implementing network architecture

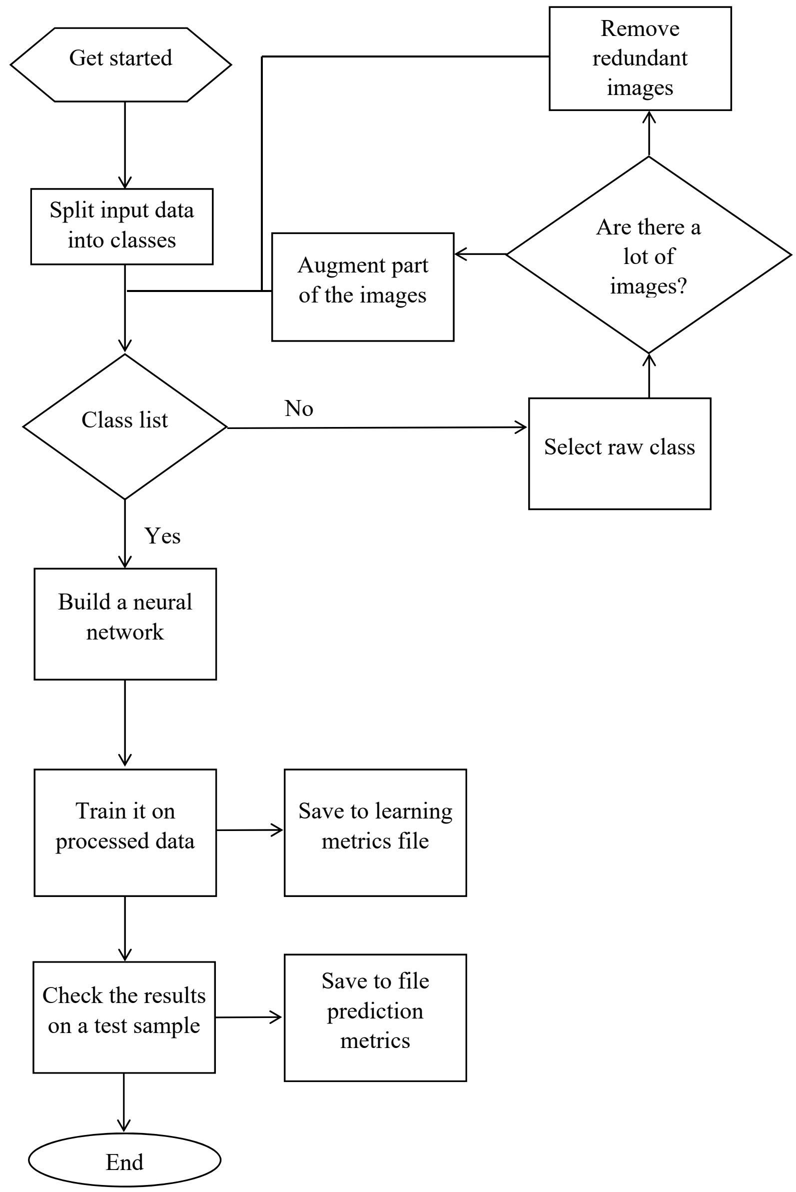

Fig. 8. Scheme of the software operation

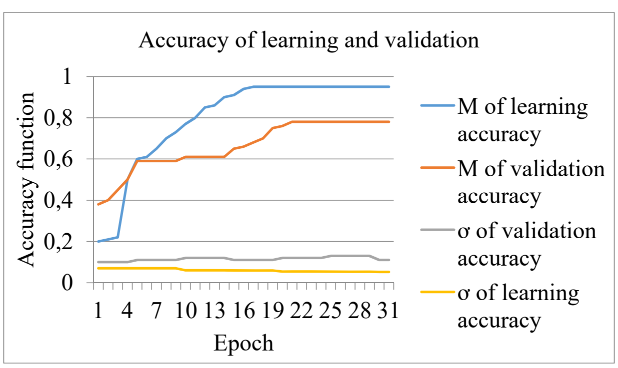

Fig. 9. Dependence of the values of the standard deviation of the accuracy function on the number of epochs in the sample without class aggregation

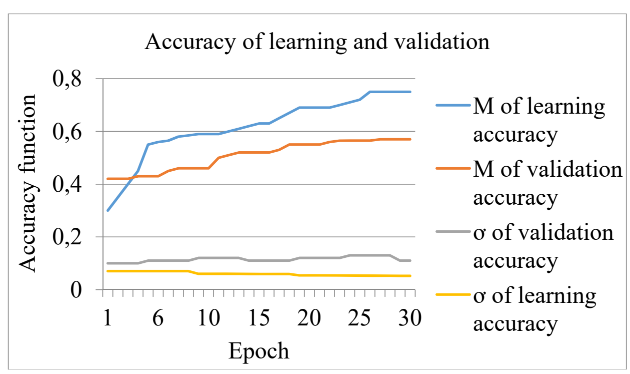

Fig. 10. Dependence of the standard deviation and accuracy function on the number of epochs in the sample with class aggregation

Also, the information system makes it possible to view the accuracy of recognition for each class separately (Fig. 13).

From Fig. 13 it can be concluded that for each class the recognition accuracy is different, the classes "vegetation fields" and "non-vegetation fields" are most accurately recognized.

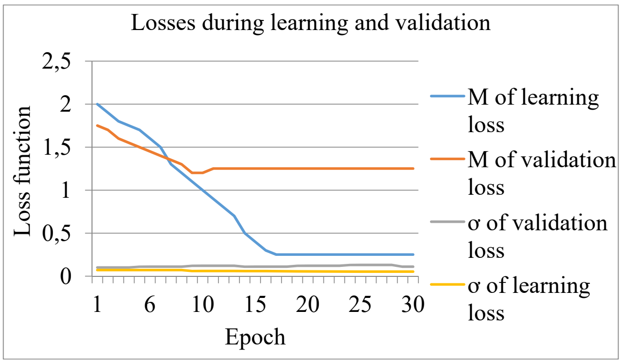

Fig. 11. Dependence of the standard deviation and loss function on the number of epochs in the sample without class aggregation

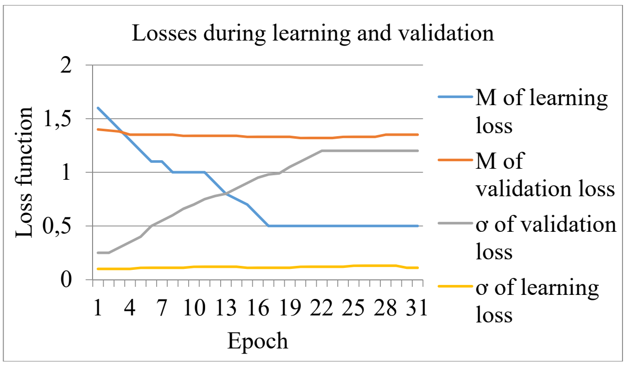

Fig. 12. Dependence of the standard deviation and the average loss function on the number of epochs in the sample with class aggregation

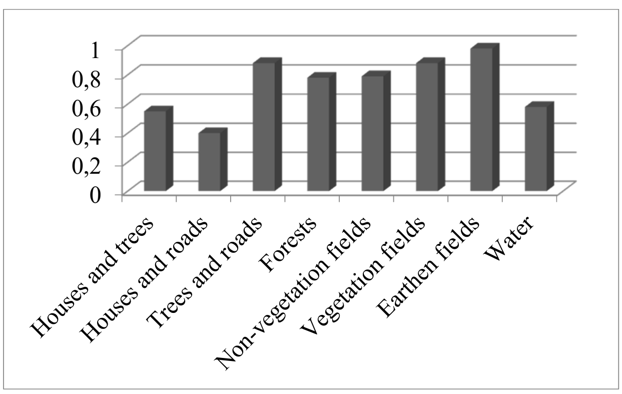

Fig. 13. Recognition accuracy for each class

6. Discussion of results of using information technology for recognition of aerial photography objects

The use of convolutional neural network of the ConvNets group in the architecture of the information system for recognizing aerial photography objects gave high accuracy indicators when training the information system and validating its results. This is confirmed by the works of other authors [12, 13]. As can be seen from Fig. 9, 10, the accuracy of recognition in final epochs ceases to increase, therefore, the neural network has reached its maximum accuracy. From Fig. 11, 12 it is concluded that the loss function at the last epochs ceases to subside for the test sample, therefore the neural network has reached the minimum number of losses.

But no studies have been found on the training of the ConvNets group's neural network. Therefore, it was decided to analyze in which case the network of the ConvNets group will provide validation results with higher accuracy when learning takes place on datasets with or without class aggregation. Comparing Fig. 9, 10, you can see that when data from different classes overlap, the recognition accuracy decreases, that is, the learning accuracy is higher for the sample without class aggregation. The losses when training on datsets is used without class aggregation are smaller, as can be seen from the Fig. 11, 12. This allows us to conclude that for information systems that include ConvNets groups in their network architecture, training should be carried out on datasets without class aggregation.

It is recommended to use our work's result in the automation of the filling of datasets, and information filtering of visual images.

A limitation on the use of the proposed technology is its dependence on the training data obtained during the warm season. For example, for classes where trees are present, further research should pay attention to the recognition of "trees in autumn", "trees in winter", "trees in summer". Similarly, it can be noted that Ukraine has, on a territorial basis, different types of landscape or different types of development (industrial, high-rise urban, rural buildings, etc.). For better recognition of different types of buildings, it is desirable to further increase the number of training classes, which is confirmed by the general conclusion of the study, namely: higher quality of recognition for non-aggregated data.

A possible direction for further research may be to analyze the transfer of learning results to remote sensing data obtained from satellites. In addition, it is relevant to use the results in solving the problem of structural description of the terrain for positioning the aircraft, under conditions of interference with supuputnikov navigation systems. The latter task relates to dual-use tasks and is relevant in aviation unmanned aerial surveillance and aerial reconnaissance systems.

7. Conclusions

1. In the development of information technology, a matha ematical model was used to classify objects with optimization by classification accuracy based on the ConvNets group wrapping neural network with modifications in the structure. Based on comparative analysis, it was established that the capacity of the model, based on the neural network, depends on the initial aggregation of classes, which reveals a wide prospect of studying the issue of regularization of complex non-parametric approximation models, which are neural networks of direct propagation, by detailing the markup of training sets. It is shown that this model provides better recognition for non-aggregated classes, which is consistent with the theoretical assumption about the quality of a posteriori assessment of data distribution, subject to a better assessment of the a posteriori distribution (i.e., more detailed markup of training data). On the other hand, this fact is not obvious and required practical proof for this particular subject area – the results of aerial photography.

2. The information system for recognizing image classes according to aerial photography, optimized for class aggregation, based on the network of the ConvNets group, was developed, programmatically implemented, and described;

training, testing, and practical verification were conducted. According to the test results, it was determined that for data without class aggregation, the probability of recognition is equal to 0.96, compared with the accuracy for data with class aggregation, which is 0.83. It is established that with the optimal number of epochs, to improve the accuracy of the classification of images, it is worth using partitioning into classes with a minimum intersection of data in classes.

Funding

The study was conducted without financial support.

Data availability

All data are available in the main text of the manuscript.

References

1. Abbas, A., Yadav, V., Smith, E., Ramjas, E., Rutter, S. B., Benavidez, C. et al. (2021). Computer Vision-Based Assessment of Motor Functioning in Schizophrenia: Use of Smartphones for Remote Measurement of Schizophrenia Symptomatology. Digital Biomarkers, 5 (1), 29–36. doi: https://doi.org/10.1159/000512383
2. Minz, P. S., Saini, C. S. (2021). Comparison of computer vision system and colour spectrophotometer for colour measurement of mozzarella cheese. Applied Food Research, 1 (2), 100020. doi: https://doi.org/10.1016/j.afres.2021.100020
3. Gao, S., Guan, H., Ma, X. (2022). A recognition method of multispectral images of soybean canopies based on neural network. Ecological Informatics, 68, 101538. doi: https://doi.org/10.1016/j.ecoinf.2021.101538
4. Prystavka, P., Dukhnovska, K., Kovtun, O., Leshchenko, O., Cholyshkina, O., Zhultynska, A. (2021). Devising information technology for determining the redundant information content of a digital image. Eastern-European Journal of Enterprise Technologies, 6 (2 (114)), 59–70. doi: https://doi.org/10.15587/1729-4061.2021.248698
5. Appiah, O., Asante, M., Hayfron-Acquah, J. B. (2022). Improved approximated median filter algorithm for real-time computer vision applications. Journal of King Saud University - Computer and Information Sciences, 34 (3), 782–792. doi: https://doi.org/10.1016/j.jksuci.2020.04.005
6. Liu, L., Chen, C. L. P., Zhou, Y., You, X. (2015). A new weighted mean filter with a two-phase detector for removing impulse noise. Information Sciences, 315, 1–16. doi: https://doi.org/10.1016/j.ins.2015.03.067
7. Belattar, S., Abdoun, O., Haimoudi, E. K. (2022). A Novel Strategy for Improving the Counter Propagation Artificial Neural Networks in Classification Tasks. Journal of Communications Software and Systems, 18 (1), 17–27. doi: https://doi.org/10.24138/jcomss-2021-0121
8. Kravchenko, Y., Leshchenko, O., Dakhno, N., Deinega, V., Shevchenko, H., Trush, O. (2020). Intellectual Fuzzy System Air Pollution Control. 2020 IEEE 2nd International Conference on Advanced Trends in Information Theory (ATIT). doi: https://doi.org/10.1109/atit50783.2020.9349334
9. Mahdianpari, M., Salehi, B., Rezaee, M., Mohammadimanesh, F., Zhang, Y. (2018). Very Deep Convolutional Neural Networks for Complex Land Cover Mapping Using Multispectral Remote Sensing Imagery. Remote Sensing, 10 (7), 1119. doi: https://doi.org/10.3390/rs10071119
10. Prystavka, P., Dolgikh, S., Kozachuk, O. (2022). Terrain Image Recognition with Unsupervised Generative Representations: the Effect of Anomalies. 2022 12th International Conference on Advanced Computer Information Technologies (ACIT). doi: https://doi.org/10.1109/acit54803.2022.9913178
11. Muhtasim, D. A., Pavel, M. I., Tan, S. Y. (2022). A Patch-Based CNN Built on the VGG-16 Architecture for Real-Time Facial Liveness Detection. Sustainability, 14 (16), 10024. doi: https://doi.org/10.3390/su141610024
12. Zhang, H., Luo, X. (2022). The Role of Knowledge Creation-Oriented Convolutional Neural Network in Learning Interaction. Computational Intelligence and Neuroscience, 2022, 1–12. doi: https://doi.org/10.1155/2022/6493311
13. Hong, M., Choe, Y. (2019). Wasserstein Generative Adversarial Network Based De-Blurring Using Perceptual Similarity. Applied Sciences, 9 (11), 2358. doi: https://doi.org/10.3390/app9112358