Теоретические основы информатики

Theoretical foundations of computer science

УДК 004:932

ИДЕНТИФИКАЦИЯ ОБЪЕКТОВ ЗЕМНОЙ ПОВЕРХНОСТИ НА ОСНОВЕ АНСАМБЛЕЙ СВЕРТОЧНЫХ НЕЙРОННЫХ СЕТЕЙ

Е. Е. МАРУШКО[1)]***, А. А. ДУДКИН***[1)]***, С. ЧЕН***[2)]

[1)]*Объединенный институт проблем информатики НАН Беларуси, ул. Сурганова, 6, 220012, г. Минск, Беларусь*
[2)]*Сианьский институт оптики и точной механики Китайской академии наук, Шэньси, 710119, г. Сиань, Китай*

В работе предлагается методика идентификации объектов на изображениях поверхности Земли, основанная на сочетании методов машинного обучения. В качестве исходных моделей рассматриваются различные варианты многослойных сверточных нейронных сетей и машин опорных векторов. Предлагается также гибридная сверточная нейронная сеть, которая комбинирует признаки, выделенные нейронной сетью и экспертами. Оптимальные значения гиперпараметров моделей вычисляются методами сеточного поиска с использованием *k*-кратной перекрестной проверки. Показана возможность повышения точности идентификации на основе ансамблей указанных моделей нейронных сетей. Эффективность предложенного подхода демонстрируется на примере изображений, полученных радаром с синтезированной апертурой.

Образец цитирования:
Марушко ЕЕ, Дудкин АА, Чен С. Идентификация объектов земной поверхности на основе ансамблей сверточных нейронных сетей. *Журнал Белорусского государственного университета. Математика. Информатика.* 2021;2:114–123 (на англ.).
https://doi.org/10.33581/2520-6508-2021-2-114-123

For citation:
Marushko EE, Doudkin AA, Zheng X. Identification of Earth's surface objects using ensembles of convolutional neural networks. *Journal of the Belarusian State University. Mathematics and Informatics.* 2021;2:114–123.
https://doi.org/10.33581/2520-6508-2021-2-114-123

Авторы:
Евгений Евгеньевич Марушко – научный сотрудник лаборатории идентификации систем.
Александр Арсентьевич Дудкин – доктор технических наук, профессор; заведующий лабораторией идентификации систем.
Сиантао Чен – кандидат наук (обработка сигналов и информации); доцент базовой лаборатории технологий спектральной обработки изображений.

Authors:
Evgenii E. Marushko, researcher at the laboratory of identification systems.
marushkoee@gmail.com
Alexander A. Doudkin, doctor of science (engineering), full professor; head of the laboratory of identification systems.
doudkin@newman.bas-net.by
Xiangtao Zheng, PhD (signal and information processing); associate professor at the key laboratory of spectral imaging technology.
xiangtaoz@gmail.com

Ключевые слова: сверточная нейронная сеть; машина опорных векторов; ансамбль нейронных сетей; изображение поверхности Земли; дистанционное зондирование; радар с синтезированной апертурой.

Благодарность. Работа частично поддержана Белорусским фондом фундаментальных исследований и национальным фондом естественных наук Китая (проект № Ф20-017).

IDENTIFICATION OF EARTH'S SURFACE OBJECTS USING ENSEMBLES OF CONVOLUTIONAL NEURAL NETWORKS

E. E. MARUSHKO[a]***, A. A. DOUDKIN***[a]***, X. ZHENG***[b]

[a]*United Institute of Informatics Problems, National Academy of Sciences of Belarus, 6 Surhanava Street, Minsk 220012, Belarus*
[b]*Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences, Shaanxi, Xi'an 710119, China*
Corresponding author: E. E. Marushko (marushkoee@gmail.com)

The paper proposes an identification technique of objects on the Earth's surface images based on combination of machine learning methods. Different variants of multi-layer convolutional neural networks and support vector machines are considered as original models. A hybrid convolutional neural network that combines features extracted by the neural network and experts is proposed. Optimal values of hyperparameters of the models are calculated by grid search methods using *k*-fold cross-validation. The possibility of improving the accuracy of identification based on the ensembles of these models is shown. Effectiveness of the proposed technique is demonstrated by the example of images obtained by synthetic aperture radar.

Keywords: convolutional neural network; support vector machine; neural network ensemble; Earth's surface image; remote sensing; identification; synthetic aperture radar.

Acknowledgements. This work was partially supported by the Belarusian Republican Foundation for Fundamental Research and the National Foundation of Natural Sciences of China (project No. F20-017).

Introduction

Remote sensing of the Earth is an observation of the Earth's surface by ground, aviation and space imaging means. Wavelengths received by the imaging equipment ranges from visible optical to radio waves. A multi-channel equipment of the passive and active types is used, that registers an electromagnetic radiation. Passive sensing methods use the natural reflected or secondary thermal radiation of Earth's objects due to solar activity. Active methods use stimulated emission of the objects, initiated by an artificial source. Remote sensing data is characterised by a large degree of dependence on the transparency of the atmosphere.

Digital data are represented as a two-dimensional image for each spectral range in the form of a matrix (two-dimensional array) of numbers $I(i, j)$ of the intensity of radiation received by a sensor from elements of the Earth's surface, which correspond to pixels of the image, where (i, j) are coordinates of the pixels. If the image is obtained in several ranges of the electromagnetic spectrum, it is represented by a three-dimensional matrix consisting of the numbers $I(i, j, k)$, where k is the number of the spectral channel. Thus, the information obtained during remote sensing is data with spatial relationships between the features $I(i, j)$.

Artificial neural networks (NNs) are successfully applied for solving image processing problems including object identification. Researches in the field of increasing the efficiency of identification based on NN theory are carried out in the following two main directions.

1. Development of the unique most appropriate multi-layer hybrid NN model for object identification on images which combining some popular NN models to solve the realistic identification process efficiently. This model is constructed from at least two different types of NNs. The first part of the architecture is aimed to image pre-processing and feature extraction, the second – directly to object detection. There are known different NN combinations for this goal: multi-layer perceptron, convolutional neural network (CNN), self-organising map, long short-term memory, NN realisation of principle component analysis, several support vector machines (SVMs), recurrent NN, etc. [1–4].

2. Development of ensembles of neural networks (ENN). They are sets of NNs making decision by averaging the results of individual NNs improving the identification quality [3; 5; 6].

In recent years, deep NNs have been most successfully used for processing on images obtained by Earth remote sensing [7]. Our work combines the above-mentioned approaches by using an ensemble of hybrid CNNs and SVMs to solve an image identification problem: to discern the nature of image components (objects). The main contribution of the paper is to distinguish the objects of the known class from the objects of alien classes (one-class classification). Effectiveness of the proposed technique is demonstrated on the task to identify objects of two classes on images obtained by synthetic aperture radar.

Methods and algorithms

The architecture of CNNs was proposed by LeCun [8] and it is aimed at effective image recognition. The NN architecture got its name because of convolution operations, where each image fragment is multiplied by the convolution matrix (kernel) element by element, and the result is summed up and written to the same position [9]. CNN is usually an alternation of convolutional layers, pooling layers, dense layers and an output layer. Additionally, a dropout layers can be used.

The dense (fully connected) layer connects each neuron with all neurons at the previous level. Each connection has its own weight. In the convolution layer (in contrast to the dense layer), a neuron is connected only with a limited number of neurons of the previous level. The convolutional layer is similar to the use of the convolution operation, which uses only a weight matrix of a small size (convolution kernel). The pooling layer performs dimension reduction. This can be done in various ways, but the method of selecting the maximum element is often used – the previous layer output is divided into cells, the maximum value is selected in each cell that is transferred to the next layer. The dropout layer is a matrix of coefficients and can be used together with all mentioned layers for a weight regularisation. The regularisation (dropout) consists in changing the NN structure: each neuron turns off with a certain probability at a stage of a training using stochastic gradient descent. The training is performed on a thinned NN, and a gradient step is made for the remaining weights. The output layer performs a class of identified objects.

The accuracy of identifying objects can be improved using ENNs [10–12]. It is necessary to realise the variability of NNs in the ensemble. The following approaches or their combinations can be applied for this purpose:

- using different parts of the training set;
- random initialisation of NN weights;
- variation of NN architectures in the ensemble (adding hidden layers, adding or deleting neurons of the hidden layer).

The output value of the ensemble is formed as a weighted sum of the outputs of the individual NNs. This approach is illustrated in fig. 1. For the case with one output neuron, the result is calculated by the equation

$$y = \sum_{i=1}^{n} y_i w_i,$$

where n is the number of the NNs; y_i is the output of the i NN; w_i is the weight of the i NN, which is calculated by the formula

$$w_i = \frac{A_i}{\sum_{j=1}^{n} A_j},$$

where A_i is a chosen measure of error calculated for the i NN; n is the number of NNs.

SVM is one of the most popular supervised learning methods proposed by Vapnik [13]. It creates a hyperplane or a set of hyperplanes in a multi-dimensional space that can be used to solve problems of classification, regression and other close problems. If the data is linearly inseparable, a non-linear kernel is applied, which allows to map the source data to a space of higher dimension, where an optimal separating hyperplane can exist.

The method is recommended for processing a small set of features. Therefore, it can be chosen as the main method for forming a model from manually selected features of objects in the image. Thus, CNNs that receive input data directly in the form of images and SVMs that make decisions on selected features of objects can be combined into an ensemble (fig. 2). This scheme can be modified by submitting additional features, formed without using images, directly to the input of the SVM classifiers. CNN can be modified similarly. The network can be divided into several branches for data processing (fig. 3).

One branch performs automatic feature extraction on the image using standard CNN layers, the weighting coefficients of which are determined by gradient methods during training. The other branch may include a set of predetermined processing procedures to form an additional set of features for each input image. Also, the sets of external features can be submitted to the hybrid model. This model involves two stages of training.

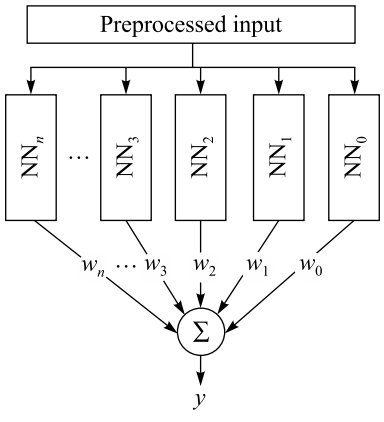

Fig. 1. Weighted ensemble

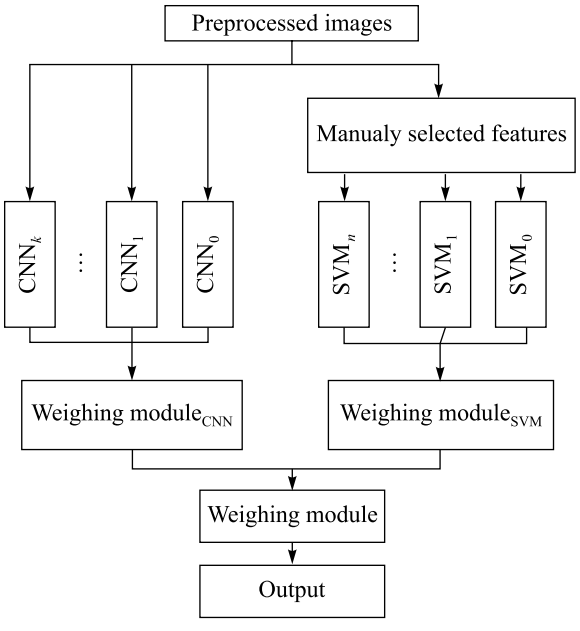

Fig. 2. Ensemble of CNN and SVM models

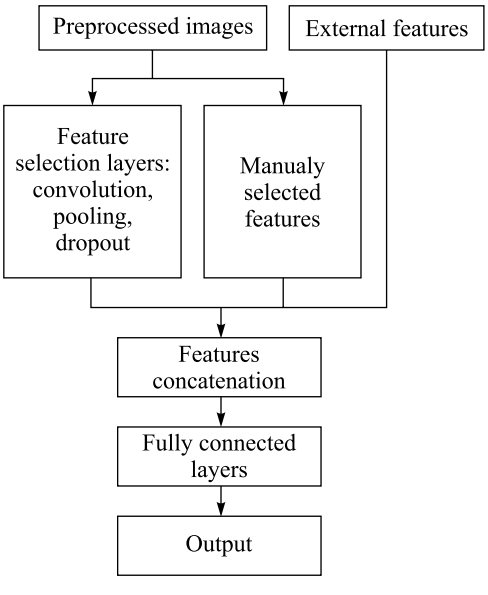

Fig. 3. Hybrid convolutional neural network

At the first stage, the first branch of the network is trained until sufficient accuracy is achieved, or before stopping by early stopping methods. At the second stage, the weights of the convolutional layers of the network are fixed and the training is carried out only for fully connected layers, where the features from convolutional layers, the manual set of features and the external features come together.

The proposed technique based on an ensemble of models for identifying objects of remote sensing of the Earth consists of the following steps.

Step 1: description of source data, the objects for identification and model quality measures (problem statement).

Step 2: formation of a training set: data collection, preprocessing, marking of the output set.

Step 3: searching additional features to solve the problem.

Step 4: expansion of the training set with additional features.

Step 5: splitting the training set into training and test sets.

Step 6: determination of the model's architecture based on the source data.

Step 7: determination of the hyperparameters range of the selected architecture.

Step 8: determination of the optimal model hyperparameters by grid or random search [14] using the *k*-fold cross-validation on the training set.

Step 9: forming an ensemble based on cross-validated models.

Step 10: if the model satisfies the quality measure on the test set, then the problem is solved, otherwise, it is necessary to expand the data set and go to step 4.

Experiments

The experimental data are images obtained using synthetic aperture radar (SAR), which allows taking radar images of the Earth's surface and objects on it, regardless of meteorological conditions and the level of the natural light of the area under observation. They include [15]:

• the images in two polarisation modes: horizontal – horizontal (HH), horizontal – vertical (HV); each image contains one object: a ship or an iceberg (fig. 4);

• incidence angle;

• data set: 1604 images with the size 75 × 75.

Data set was divided into 80 % training part and 20 % test part, so we use 1283 samples for training ENN.

Additionally, experiments were carried out on an extended data set. A simple augmentation technique was used for this: horizontal flip, vertical flip, 90-degree clockwise rotation, horizontal flip and vertical flip for the rotated image. In this case, we have 7698 samples for training.

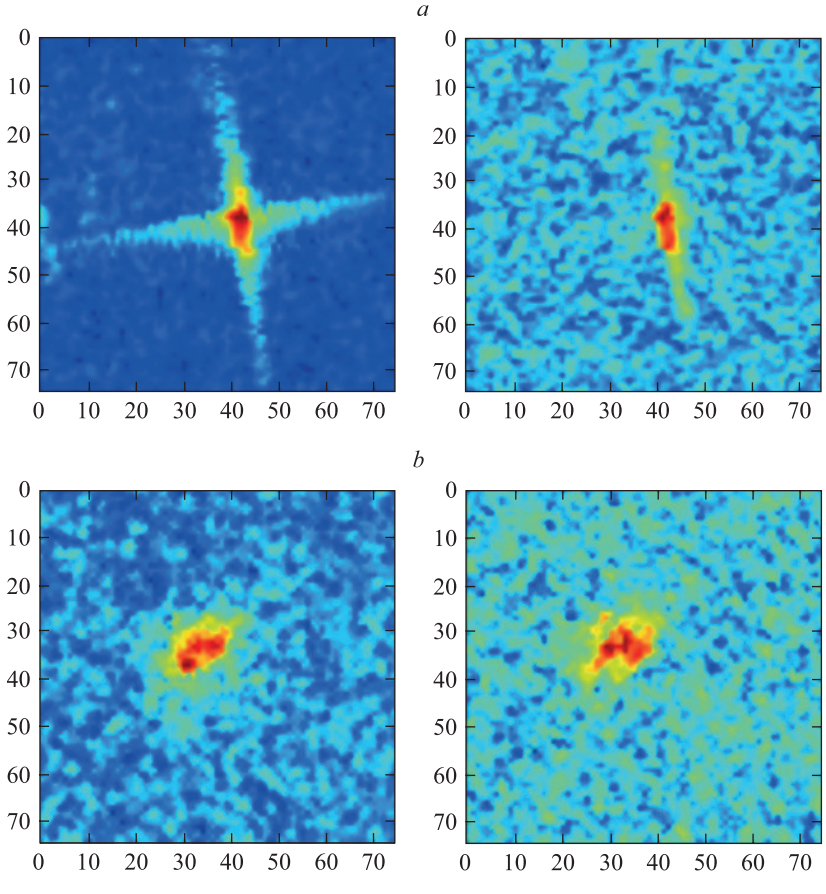

Fig. 4. Sample images: *a* – ship; *b* – iceberg

Training part was used for cross validated grid search, and test part was used for evaluation.

The task is to identify objects of two classes: an iceberg or a ship, which is essentially a binary classification task.

Efficiency in the classification problem can be assessed using accuracy – this is a basic measure that shows the proportion of correct model responses. For the binary classification problem, when the model derives the class probabilities, the logarithmic loss (logloss) function is used:

$$L = -\frac{1}{l}\sum_{i=1}^{l}\left(y_i \log\left(\hat{y}_i\right) + \left(1 - y_i\right)\log\left(1 - \hat{y}_i\right)\right),$$

where $\hat{y}_i$ is a model response on the i object; y is a true class label on the i sample; l is the number of samples.

The minimisation of L can be represented as the task of maximising accuracy by a penalty for incorrect predictions. However, it should be noted that L is heavily penalised for the classifier's confidence in the wrong answer. Therefore, an error on one object can give a significant increase in the total error. Such samples are often outliers, which must be filtered or treated separately.

Based on the analysis of the initial data, the following baseline CNN is proposed (network parameters were chosen empirically):

- size of the input layer: 75 × 75 × 3;
- 2D convolutional layer 1 consists of 64 kernels with size 3 × 3, the activation function is rectified linear unit (ReLU);
- pooling layer with 2D max pooling and pool size of 2 × 2;
- dropout with probability 0.3;
- 2D convolutional layer 2: 128 kernels with size 3 × 3, RELU;
- 2D max pooling 2 × 2;
- dropout with probability 0.3;
- 2D convolutional layer 3: 128 kernels with size 3 × 3, RELU;
- 2D max pooling 2 × 2;
- dropout with probability 0.3;
- 2D convolutional layer 4: 64 kernels with size 3 × 3, RELU;
- 2D max pooling 2 × 2;
- dropout with probability 0.3;
- fully connected layer of 1024 neurons, RELU;
- dropout with probability 0.5;
- fully connected layer of 512 neurons, RELU;
- dropout with probability 0.5;
- the output fully connected layer of two outputs with a softmax activation function.

The CNN accepts a pseudo image in which the first image channel is represented by the HH channel of the original data, the second channel is represented by the HV image channel, and the third image channel is represented by their composition in form of a normalised sum.

The CNN training is performed using the Adam stochastic gradient algorithm [16]. Cross-validation is performed for $k = 5$. Each model was trained for 60 epochs. The batch size is 32. Early stopping procedure [23] was used with patience equal to 10. Starting learning rate parameter was 1e-4. It was reduced by factor equal to 0.5 based on «reduce LR on Plateau» [24] algorithm with the patience equal to 7 (the patience means the number of epochs with no improvement after which the learning rate begins to reduce). Finally, the ensemble of five CNN models was formed.

In addition to the proposed ENN, single models and ensembles based on them were considered using the following widely used architectures: VGG16 [17], ResNet50 [19], EfficientNet-B0 [20], Xception [18], MobileNet-v2 [22], DenseNet-121 [21]. To build the ensemble base model convolutional layers were taken from each model and three fully connected layers were added.

During the experiments, architectures with a small number of weights were selected, since the input data has a low dimension and a low detail. The ensembles of the best models from all trained models were also analysed.

For feature extraction, all input images were binarised using manually selected threshold. After this operation there was a mask for target objects on each image that presented any pseudo image. Also 61 features extracted from the pseudo images:

- 10 moments of the 1st and 2nd order calculated for both HH and HV images that describe an object shape;
- global statistics (mean, maximum, minimum, variance for both HH and HV images);
- differences in global statistics (3 features for both HH and HV images);
- global statistics in the masked area (mean, maximum, minimum, variance, a sum for HH and HV images);
- local statistics (maximum local standard deviation, 6 maximum values differences, variance for both HH and HV images);
- incidence angle.

Using this set of features, the ensemble of SVM models with a non-linear Gaussian kernel is trained, for which the optimal parameters C (SVM hyperparameter) and γ (Gaussian kernel parameter) of the model were determined by random search. Also, the weighted ensemble of CNN and SVM models is formed. Additionally, a hybrid CNN model is formed (see fig. 3) that combines the features extracted by convolutional layers and the set of 61 manual features.

The result of the evaluation of these models is presented in table 1. The result of the evaluation of model ensembles on augmented data is presented in table 2.

Table 1

Model evaluation on Statoil/C-CORE Iceberg Classifier without augmentation

Model	Logloss	Accuracy	Parameters
Five-fold cross validated			
EfficientNet-B0	0.482	82.118	5 886 621
ResNet50	0.392	86.044	26 211 201
Xception	0.471	86.667	23 484 969
VGG16	0.599	61.246	15 765 313
MobileNet-v2	0.695	50.156	4 095 041
Baseline CNN	0.321	86.791	888 897
DenseNet-121	0.336	87.601	8 612 417
Five-fold ensemble			
EfficientNet-B0	0.306	87.539	29 433 105
ResNet50	0.267	87.850	131 056 005
Xception	0.286	90.031	117 424 845
VGG16	0.567	79.439	78 826 565
MobileNet-v2	0.695	50.156	20 475 205
Baseline CNN	0.279	86.916	4 444 485
DenseNet-121	0.240	89.408	43 062 085
Top-5 model ensemble	0.227	91.277	75 533 421
Top-10 model ensemble	0.226	90.966	160 489 110
Weighted ensemble by loss			
EfficientNet-B0	0.293	87.227	29 433 105
ResNet50	0.272	87.227	131 056 005
Xception	0.279	90.343	117 424 845
VGG16	0.454	81.308	78 826 565
MobileNet-v2	0.695	50.156	20 475 205
Baseline CNN	0.279	87.227	4 444 485
DenseNet-121	0.252	89.408	43 062 085
Top-5 model ensemble	0.228	91.900	75 533 421
Top-10 model ensemble	0.221	92.211	160 489 110
SVM ensemble	0.303	86.292	7241
Top-5 CNN + SVM ensemble	0.227	92.523	75 540 662
Hybrid CNN	0.248	90.654	905 025

Table 2

Model evaluation on Statoil/C-CORE Iceberg Classifier on augmented data

Model	Logloss	Accuracy	Parameters
Five-fold cross validated			
EfficientNet-B0	0.434	85.234	5 886 621
ResNet50	0.332	87.664	26 211 201

Ending table 2

Model	Logloss	Accuracy	Parameters
Xception	0.381	87.850	23 484 969
VGG16	0.672	54.330	15 765 313
MobileNet-v2	0.542	75.514	4 095 041
Baseline CNN	0.302	86.854	888 897
DenseNet-121	0.326	87.539	8 612 417
Five-fold ensemble			
EfficientNet-B0	0.279	89.408	29 433 105
ResNet50	0.249	89.097	131 056 005
Xception	0.256	90.966	117 424 845
VGG16	0.661	50.156	78 826 565
MobileNet-v2	0.366	87.227	20 475 205
Baseline CNN	0.277	88.162	4 444 485
DenseNet-121	0.243	90.654	43 062 085
Top-5 model ensemble	0.243	90.966	90 405 973
Top-10 model ensemble	0.241	90.031	155 490 078
Weighted ensemble by loss			
EfficientNet-B0	0.277	89.408	29 433 105
ResNet50	0.261	89.097	131 056 005
Xception	0.255	90.654	117 424 845
VGG16	0.596	70.405	78 826 565
MobileNet-v2	0.313	87.539	20 475 205
Baseline CNN	0.269	87.539	4 444 485
DenseNet-121	0.244	90.343	43 062 085
Top-5 model ensemble	0.252	89.719	90 405 973
Top-10 model ensemble	0.245	90.343	155 490 078

First of all, it should be noted that with the selected training parameters in the base data set, some models could not find a solution (see VGG16 and MobileNet-v2 in table 1). For MobileNet-v2, the situation is improved with a data set augmentation. The data set augmentation technique used for this task has increased the accuracy of single models and an ensemble of models of the same architecture.

As can be seen, the complication of the model architecture gives a slight increase in accuracy. At the same time, the number of weighting parameters of the model greatly is increased when using a more complex architecture.

The weighted ensemble makes the possibility to improve the accuracy of some models on the base data set. At the same time, on the extended data set, the accuracy of the ensemble either remained unchanged or slightly is decreased. It can be said that weighting improves the overall accuracy of the ensemble in the case when it can contain both too weak and strong models. Otherwise, when all models are about the same level, weighting does not improve the classification.

The SVM ensemble, as a classical approach, has shown low accuracy for this task in comparison with the ENNs. However, the ensemble of the CNN and the SVM models shows the highest accuracy on the test data set.

So, the combination of models of different architectures and training methods can significantly increase the efficiency of classification. At the same time, the amount of consumed resources also increases accordingly. For an ensemble of five models, the memory consumption is increased fivefold. And since the models learn independently, there is no way to use shared weights. Also, when each model is applied sequentially to the input data, the inference time is increased fivefold. It makes sense to use the models with low memory consumption in this case. Also, independent model training gives a possibility to produce parallel inference on the models without increasing time.

Conclusion

The technique based on an ensemble of models for identifying objects of Earth remote sensing images was proposed. It includes the following steps: preparing data, object feature extraction, creating base identification models, optimising the model's hyperparameters, construction of the ensemble, processing the data by the ensemble.

The technique was applied to binary clastering of images obtained by synthetic aperture radar. Evaluation of the proposed models on experimental data has showed that one of the effective ways to increase accuracy in machine learning tasks is to form an ensemble of heterogeneous models trained on different sets of input features.

References

1. Kim M, Choi W, Jeon Y, Liu L. A hybrid neural network model for power demand forecasting. *Energies*. 2019;12(5):931. DOI: 10.3390/en12050931.

2. Frankel A, Tachida K, Jones R. Prediction of the evolution of the stress field of polycrystals undergoing elastic-plastic deformation with a hybrid neural network model. *Machine Learning: Science and Technology*. 2020;1(3):035005. DOI: 10.1088/2632-2153/ab9299.

3. Liu H, Yang R, Wang T, Zhang L. A hybrid neural network model for short-term wind speed forecasting based on decomposition, multi-learner ensemble, and adaptive multiple error corrections. *Renewable Energy.* 2021;165:573–594. DOI: 10.1016/j.renene.2020.11.002.

4. Ma C, Du X, Cao L. Analysis of multi-types of flow features based on hybrid neural network for improving network anomaly detection. *IEEE Access.* 2019;7:148363–148380. DOI: 10.1109/ACCESS.2019.2946708.

5. Berkhahn S, Fuchs L, Neuweiler I. An ensemble neural network model for real-time prediction of urban floods. *Journal of hydrology*. 2019;575:743–754. DOI: 10.1016/j.jhydrol.2019.05.066.

6. Cheng B, Wu W, Tao D, Mei S, Mao T, Cheng J. Random cropping ensemble neural network for image classification in a robotic arm grasping system. *IEEE Transactions on Instrumentation and Measurement*. 2020;69(9):6795–6806. DOI: 10.1109/TIM.2020.2976420.

7. Large scale visual recognition challenge [Internet; cited 29.01.2021]. Available from: http://image-net.org/challenges/LSVRC/2016/results.

8. LeCun Y, Boser B, Denker JS, Henderson D, Howard RE, Hubbard W, et al. Backpropagation applied to handwritten zip code recognition. *Neural computation*. 1989;1(4):541–551.

9. Goodfellow I, Bengio Y, Courville A. *Deep Learning*. Cambridge: MIT Press; 2016. 781 p.

10. Parikh D, Polikar R. An ensemble-based incremental learning approach to data fusion. *IEEE Transactions on Systems, Man and Cybernetics*. *Part B: Cybernetics*. 2007;37(2):437450. DOI: 10.1109/TSMCB.2006.883873.

11. Marushko EE, Doudkin AA. Ensembles of neural networks for forecasting of time series of spacecraft telemetry. *Optical Memory and Neural Networks*. 2017;26(1):47–54. DOI: 10.3103/S1060992X17010064.

12. Kourentzes N, Barrow D, Crone S. Neural network ensemble operators for time series forecasting. *Expert Systems with Applications.* 2014;41(9):4235–4244. DOI: 10.1016/j.eswa.2013.12.011.

13. Vapnik V. *The nature of statistical learning theory.* 2nd edition. New York: Springer; 1999. 314 p.

14. Bergstra J, Bengio Y. Random search for hyper-parameter optimization. *Machine Learning Research*. 2012;13:281305.

15. Statoil/C-CORE iceberg classifier challenge. Data [Internet; cited 29.01.2021]. Available from: https://www.kaggle.com/c/statoil-iceberg-classifier-challenge/data.

16. Kingma DP, Ba J. Adam: a method for stochastic optimization. arXiv:1412.6980. 2017 [cited 29.01.2021]: [15 p.]. Available from: https://arxiv.org/abs/1412.6980.

17. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. arXiv:1409.1556 [Preprint]. 2015 [cited 29.01.2021]: [14 p.]. Available from: https://arxiv.org/abs/1409.1556.

18. Chollet F. Xception: deep learning with depthwise separable convolutions. In: IEEE Computer Society. *2017 IEEE Conference on computer vision and pattern recognition (CVPR); 2017 July 21–26; Honolulu, USA.* Los Alamitos: IEEE; 2017. p. 1251–1258. DOI: 10.1109/CVPR.2017.195.

19. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: IEEE Computer Society. *Proceedings of the IEEE conference on computer vision and pattern recognition; 2016 June 27–30; Las Vegas, Nevada*. Los Alamitos: IEEE; 2016. p. 770–778. DOI: 10.1109/CVPR.2016.90.

20. Tan M, Le QV. Efficient net: rethinking model scaling for convolutional neural networks. arXiv:1905.11946 [Preprint]. 2020 [cited 29.01.2021]: [11 p.]. Available from: https://arxiv.org/abs/1905.11946.

21. Huang G, Liu Z, Van Der Maaten L, Weinberger KQ. Densely connected convolutional networks. In: IEEE Computer Society. *2017 IEEE Conference on computer vision and pattern recognition (CVPR); 2017 July 21–26; Honolulu, USA.* Los Alamitos: IEEE; 2017. p. 2261–2269. DOI: 10.1109/CVPR.2017.243.

22. Sandler M, Howard A, Zhu M, Zhmoginov A, Chen LC. MobilenetV2: inverted residuals and linear bottlenecks. In: IEEE Computer Society. *2018 IEEE/CVF Conference on computer vision and pattern recognition; 2018 June 18–23; Salt Lake City, USA.* Los Alamitos: IEEE; 2018. p. 4510–4520. DOI: 10.1109/CVPR.2018.00474.

23. Prechelt L. Early stopping – but when? In: Orr GB, Müller K-R, editors. *Neural Networks: tricks of the trade*. Berlin: Springer; 1998. p. 55–69.

24. Goyal P, Dollar P, Girshick R, Noordhuis P, Wesolowski L, Kyrola A, et al. Accurate, large minibatch SGD: training imagenet in 1 hour. arXiv:1706.02677 [Preprint]. 2018 [cited 29.01.2021]: [12 p.]. Available from: https://arxiv.org/abs/1706.02677.

Received 03.02.2021 / revised 29.06.2021 / accepted 29.06.2021.