COMPUTER SCIENCES AND INFORMATION TECHNOLOGIES

[1]O. I. Chumachenko,
[2]A. T. Kot

FORMATION OF A LEARNING SET FOR THE TASK OF IMAGE PROCESSING

[1,2]Technical Cybernetic Department, National Technical University of Ukraine "Ihor Sikorsky Kyiv Polytechnic Institute," Kyiv, Ukraine
E-mails: [1]chumachecko@tk.kpi.ua   ORCID 0000-0003-3006-7460, [2]anatoly.kot@gmail.com

Abstract—The problem of forming a training set for the task of image processing is considered. It is shown that this task is of great importance in the construction of intelligent medical diagnostic systems in which convolution neural networks are used for image processing (results of ultrasound, CT and MRI). Due to the lack of elements of the training sample, it is proposed, on the one hand, to use approaches of artificial data multiplication based on the initial training sample of a fixed volume, and on the other hand, to use methods that reduce the need for large training samples, both through the use of ensemble topology (hybrid neural networks), and by applying the transfer learning approach. An algorithm for the formation of a training set for image processing tasks is developed based on the modification of the initial input information with the calculation of the confidence measure of the obtained sample.

Index Terms—Transfer learning; learning set; convolution neural networks; ensemble topology; image processing.

I. INTRODUCTION

One of the problems arising in the construction of machine learning algorithms and, in particular, algorithms of classification of patterns is the formation of the initial training data. When solving complex practical problems of machine learning, the stage of preparation and organization of training data becomes the most important, given the general trend towards using data analysis algorithms that have a universal structure (neural networks, compositional algorithms, etc.).

At the same time, the search for objects, preliminary processing and the formation of their formalized descriptions (features) with their subsequent preparation for use for training certain algorithms requires significant costs of time and human resources. This is especially evident under images, sound, text processing.

In connection with the active development of deep neural networks in the last decade, the formation of a set of training data is especially important, since in many problems deep neural networks demonstrate a quality that is significantly superior to other machine learning algorithms, however, in order to obtain such a gain in quality, it is necessary to use training set of very large size (up to several million images, while training requires a large amount of computational resources.

As the complexity of the tasks being solved in the field of data analysis and the creation of artificial intelligence systems, the need for such actions will grow.

It should also be noted that in a number of problems, obtaining representative training samples is also difficult for objective reasons. Usually this is due to the need to product and conduct experimental research. In medicine, there are often objectively limitations associated with the need to obtain training data for the diagnosis of diseases, however, the preparation of such data can take months or years.

Unfortunately, insufficient attention is paid to the formation of the training set, often these issues are completely ignored and the theoretical base is insufficiently developed to explain the phenomena that arise in the process of training data set forming.

II. BASIC CONCEPTS

Learning set. Let there be a set of inputs X, a set of correspondent outputs Y, and there is an objective function $y* : X \to Y$, whose values $y_i = y*(x_i)$ are known only for a finite subset of inputs $\{x_1, ..., x_L\} \subset X$. The set of pairs $X^L = (x_i, y_i)_{i=1}^{L}$ is called a training set. The learning task is to restore the dependence $y*$ from the training set X^L, that is, to construct the decision function $a : X \to Y$, which would approximate the objective function $y*(x)$, and not only on the objects of the training set, but and on the whole set X [1].

The training method is a mapping $\mu : (X \times Y)^L \to A,$ which assigns a certain decision function $a : X \to Y$ to an arbitrary finite training set $X^L = (x_i, y_i)_{i=1}^L.$

It is also said that the training method builds the decision function a from the training set X^L [1].

III.   DATA SET MODELS

As a data set model, a probabilistic model and a case data model, a data model based on the application of the Transfer Learning approach, can be used. Transfer Learning allows you to use the experience gained while solving one problem to solve another, similar problem. The neural network is first trained on a large amount of data, then on the target set.

Modern machine learning is based on a probabilistic data model [1]. It is considered that training set $X^L = (x_i, y_i)_{i=1}^L$ is a sample from the general population of some objects, while the sample should reflect the basic properties of the general population. It is also assumed that the probability of the appearance of objects of a certain type in the training sample is equal to the probability of the appearance of these objects in the population.

The probabilistic data model has several disadvantages [1].

1) The results in the probabilistic model strongly depend on the ratio of the number of objects of different types in the sample.

2) A probabilistic data model is convenient to use only if the data is a homogeneous (unimodal) population, which is not true in most machine learning problems.

In the case data model, the entire space of possible objects X is divided into some types of data (cases). Each case is characterized by a membership function $f(x)$ (for each object it determines the degree of membership in a given case) and importance w. At the same time, the importance of the case may change during the development of the system, depending on the customer's requirements. In simpler terms, a case is a set of conditions that some kind of objects from a set of data satisfy. Arbitrary nesting of cases is allowed. Within one case, you can use the standard probabilistic data model. The choice of cases in the object space can be performed arbitrarily and serves solely for the convenience of system operation describing.

The case model makes it possible to achieve stability with respect to changes in the relative number of objects of different types in the set. It also simplifies the description and understanding of the phenomena when training the algorithm, since in different areas of the object space the training set can have completely different properties (for example, different data density), and the probabilistic model allows you to evaluate only the properties of the training set or the system as a whole.

IV.   MODEL OF THE OBSERVABLE FEATURES FORMATION

The training object x_i is determined by a set of observable features $f_1, \ldots, f_q$ which are calculated with help of the values of some variables $z_1, \ldots, z_L$ (we will call them generating variables). These variables can be divided into several types: *Target variables, variables of intraclass distinction, external variables, random variables* [28].

V.   CRITERIA FOR THE FORMATION OF A TRAINING SAMPLE

Based on the analysis of terms and various learning procedures for classification processes (pattern recognition), it can be concluded that the future quality of decision-making is influenced, on the one hand, by the qualitative and quantitative composition of the training sample and, on the other hand, by the qualitative and quantitative composition of the space of informative features. However, these metrics are not very informative. Firstly, there can be a lot of data, but they are all the same, secondly, even if all objects are different, some areas of the feature space may remain empty, and, thirdly, in the very procedure of forming the training set, they can mistakes be laid.

In turn, training samples are characterized by such indicators as representativeness (as belonging to the general population [2], [3], [4]), volume and expert confidence. The feature space can be characterized by statistical indicators of information content, expert confidence in the composition of features and dimension.

The specified qualitative and quantitative indicators characterizing the training samples and the space of features are mainly empirical in nature with a pronounced fuzzy definition. Proceeding from this, to describe the input indicators, taking into account the existing terminology in the field of fuzzy decision-making logic and the theory of confidence, to designate the integral characteristics of the training sample, we will introduce the concept of a measure of confidence in the learning abilities of the sample (MCS), and to indicate the classification possibility of the space of characteristic space – a measure of confidence in the feature space (MCF). We will give the MCS and

MCF indicators the property of a measure of confidence in the decisions made, defining the area of their change in the range from 0 to 1, in which zero corresponds to complete distrust to the training sample or the composition of informative features, and one corresponds to complete confidence to them.

Full confidence in the training set and feature composition means that there is a potential possibility of classification decision rules synthesis that are never "wrong".

Similarly, for a training sample, we will define the concept of a measure of confidence to the representativeness of a sample (MCR), a measure of confidence to the sample size (MCS), a measure of confidence of experts in a sample (MCES). For the space of features, we define the concept of a measure of confidence to the informative or informational value (MDI), measure of confidence of experts to the composition of features (MCEF), measures of confidence to the dimension (number) of informative features (MCN).

The task and calculation of the selected indicators can be carried out: by a group of highly qualified experts; by statistical criteria on various samples, including small ones; using mixed strategies (experts, statistical calculations, fuzzy structures and operations on them). In order to simplify the subsequent entries, we will introduce new designations for confidence measures: MCS $\leftrightarrow I_1$; MCF $\leftrightarrow I_2$; MCR $\leftrightarrow I_3$; MCS $\leftrightarrow I_4$; MCES $\leftrightarrow I_5$; MDI $\leftrightarrow I_6$; MCEF $\leftrightarrow I_7$; MCN $\leftrightarrow I_8$.

Taking into account the introduced definitions, it is proposed to evaluate the classification capabilities of the training data in accordance with the following approach.

1) Training and control samples are formed, and at the expert level the composition of indicators $I_1 = F_1 (I_3, I_4, I_5)$ and $I_2 = F_2 (I_6, I_7, I_8)$, is determined, where F_1 and F_2 are the aggregation functionals of the components I_1 and I_2.

2) At the expert level, the method of calculating each of the constituent indicators I_1 and I_2 from the following list is determined: expert opinion, statistical estimates, mixed strategy.

If the main work is performed by experts (numerical estimate of measure confidence (membership functions, etc.)), then, taking into account the complexity of the problem being solved, in accordance with the recommendations [5], [6], the quantitative composition of the expert group is determined and, based on the results of solving text problems, its coherence is determined work with the calculation of the coefficient of concordance *W*. If $W > 0.7$, then the expert group proceeds to the solution of the given tasks. Otherwise, the composition of the group is qualitatively corrected.

If elements of fuzzy decision-making logic modified for solving classification problems using methods of exploratory analysis are used to calculate measure confidence, then using the recommendations of [7], [3], it is provided a synthesis of combined decision rules for calculating selected components from the following general list:

$$\{I_3, I_4, I_5, I_6, I_7, I_8\}.$$

Folding criteria I_3, I_4, I_5, I_6, I_7, I_8 into one criterion I_9, we obtain

$$I_9 = f \sum_{i=3}^{8} \alpha_i \, I_i,$$

where α_i are weights that determine the contribution of particular indicators I_i $(i = 3,...,8)$, respectively.

The measure of confidence I_9 can be used for both training ($I_{9 \, tr}$) and testing ($I_{9 \, t}$) samples.

The obtained values of measures I_9 allow one to specify the degree of confidence in the synthesized decision rules, since take into account not only the work of the classification rules themselves, but also the features of the data that are involved in the learning processes and testing of the work of the automated classification system.

V. Problem Statement

A set of pairs $X^L = \left(x_i, y_i \right)_{i=1}^{L}$ is given, where $x_i, y_i, i = \overline{1, L}$ input and corresponding output reference signals that are obtained as a result of the research. It is necessary, on the basis of criterion I_9 (a measure of classification confidence in the data), to evaluate the quality of the sample and, if necessary, which is determined by the value of the generalized error of the neural network, obtained as a result of test sample use, to generate adding data and use it to expand the training set, or use other approaches for classification problem solution under short learning samples.

VI. Possible Problems in the Formation of the Training Set

Let's consider some possible problems and errors in the formation of the training set

Background patterns. In machine learning problems, an object can be specified by a set of feature values $f_1, ..., f_q$ and values of target variables $z_1, ..., z_L$. The task of machine learning is to find patterns between the values of the observed features and the target variables. At the same time, on the basis of each specific training object, without taking into account other objects, any dependence $y_k(x_i)$ characteristic of this object can always be considered

true. When considering a large number of various objects from all possible regularities, only a small number of really significant regularities will remain. Note that based on the small amount of data, there is no approach to differ the right pattern from the false one. We will call such false regularities, arising from lack of data, background regularities. In fact, some types of retraining involve memorizing background regularities. An example of a background pattern is the relationship between the image class and the color of one particular pixel.

Lack of training objects of a certain type. The simplest example of an error in the formation of a training set is that if there is no data of a certain type in it (some area of the object space is not covered, in the case data model there are no objects of some case), the algorithm will not be able to correctly learn to classify them. In this case we mean objects in the space of generating variables $z_1, \ldots, z_L$, and not in the space of features $f_1, \ldots, f_q$.

It would be logical to add here an insufficient number of training objects of a certain type, however, in different cases, a different number of objects is sufficient, with different learning algorithms this problem will manifest itself in completely different ways, therefore we will assume that this problem is included in the following two problems.

Lack of data of a certain type regarding the feature system. The feature system $f_1, \ldots, f_q$ generates some partitioning of the data set into cases, each case corresponds to a certain narrow set of feature values, and the more cases are the more diverse and complex the features are. If some of these cases are not covered by objects from the training set, or the probability distribution within the case will incorrectly reflect the properties of the general population, training may turn out to be incorrect. Note that with increasing complexity of the feature system, the requirements for the training set increase.

Some of the generating variables do not vary. An important special case of the problem of lack of data of a certain type. Very often, when forming a training set, some of the generating variables always have the same values or a very narrow range of values.

Imbalance. An unreasonable from a semantic point of view, violation of the ratios of the amount of data of different types in the set of data under consideration, leading to an unreasonable overestimation of the influence on the result of some and an underestimation of the influence or complete ignoring of other data, and, as a consequence, to the adoption of nonoptimal decisions. This problem arises in situations of uneven representation of different classes, when some classes of images are represented by significantly less data than others.

Imbalance is especially critical when using decision trees. Note that imbalance is a fairly general class of phenomena that can arise not only in the process of forming a training set. The imbalance problem is often solved by using various types of normalization.

External regularities. In training set $X^L = (x_i, y_i)_{i=1}^{L}$, there may be dependencies between the external and target variables, which the learning algorithm can learn as true due to imbalances, learning feature, or a feature system. For example, in training set $X^L = (x_i, y_i)_{i=1}^{L}$, all male faces are photographed during daylight hours, and all female faces are photographed during dark hours. In such conditions, correct learning is possible only if the feature system f_b f_q does not have features that depend on the image illumination, as soon as such features are added, the algorithm will not work correctly on a sample from another source. The presence of such regularities is an example of an error in the formation of a training set. A sign of a similar problem – if, after adding better features or features based on information of a different kind, the algorithm stops working correctly.

VII.   Ways to Add Data to the Training Set in Image Processing Tasks

Adding (artificial data reproduction (ADR)) data is one of the simplest and most effective ways to improve the quality of the training set. A number of sources use essentially similar terms: artificial expansion, augmentation, morphing – data multiplication [9] – [11]. A simple addition of arbitrary data is not always effective; it is often required to add data of a certain type to improve the recognition quality.

Data augmentation. Modification of existing images in order to expand the training set. It is actively used in training deep neural networks, as well as in conditions of a lack of labeled data. Compression / stretching, horizontal display, rotation, random shift in color space, random or regular change of some pixels are applied. It is considered that adding completely random noise is ineffective, you should add noise due to data (only potentially possible distortions in real data). A significant drawback of this method is that most of the background *regularities* are preserved.

Use of a generative adversarial neural network. There is another effective way to multiply images – this is the use of a generative adversarial neural network (GAN) [12], [13], which is an architecture that consists of a generator and a discriminator. The

architecture of this network consists of two different networks. One neural network, a generator, creates random new data instances, and the other, a discriminator, evaluates them for authenticity.

Hard samples mining [14]. A classic problem in the tasks of searching for objects in an image is the need to maintain a sufficient number of hard negative samples (training examples that are similar to an object of interest, but are not) in the training set. The difficulty arises due to the fact that in natural conditions such objects are rare, therefore, special methods are used to find them and add them to the training set (hard samples mining). The key assumption in these methods is that the objects of interest to us are very similar to each other. Typically, data augmentation, adaptive search, pattern search, machine learning-based methods are applied. It is of interest to apply thematic modeling methods to find complex negative examples.

Simulate adding data. When training deep neural networks, the use of the dropout method is considered mandatory [17]: random zeroing of the activations of some neurons in the network when the next training image is fed to it (usually, 20–50% of neurons are randomly selected in each layer). Without this technique, the neural network "learns" a large number of background patterns due to the fact that the complexity of the model exceeds the amount of available data. Essentially, dropout is a simulation of adding data to a training set. Inside the learning algorithm, we simulate the variability of the data - a randomly modified version of the real image arrives at the input of the deeper levels of the network (it means an image that would not cause the activation of zeroed neurons), although such data is not actually in the training set. The disadvantage of this method is that the addition of such data can be simulated, which, in principle, cannot be in reality, due to which the recognition accuracy may suffer. It is of interest to create modifications of this method that take into account the nature of the data.

VIII. APPROACHES FOR CLASSIFICATION PROBLEM SOLUTION UNDER SHORT LEARNING SAMPLES

Ensemble synthesis. One of the most common ways to improve the classification accuracy in conditions of a fixed size of training sample is the construction and training of ensembles of classifiers. This approach is based on the idea that combining independent classifiers into an ensemble makes it possible to compensate for their individual shortcomings through collective voting, which ensures a higher classification accuracy and greater stability to random outliers in the processed data. The

following works of the authors [16] are devoted to the problem of constructing ensembles of classifiers.

Transfer learning. Transfer learning is used to improve a learner from one domain by transferring information from a related domain.

The transfer learning purpose in our case is to accumulate the knowledge necessary to solve the image processing problem and use it to solve a target problem that is close in meaning.

Here domain is defined by a pair

$$D = \{X, P(x)\},$$

where X is the feature space, $P(x)$ is the probability distribution of a random variable

$$x = \{x_1, x_2, ..., x_n\} \in X.$$

The two domains are mismatched $D_1 \neq D_2$, if feature spaces differ $X_1 \neq X_2$, either distribution $P(x_1) \neq P(x_2)$.

The task, is determined by the pair $T = \{Y, f(\cdot)\}$, where Y is the label space, $f(\cdot)$ is the prediction function.

The task is not explicitly observed, it is trained on the training set of pairs of elements $(x_i, y_i), x_i \in X, y_i \in Y$.

The prediction function is the distribution of the conditional probability that a certain label is observed for a given attribute description $f(x) = P(y|x)$.

Two tasks are mismatched $T_1 \neq T_2$, if label spaces are different $Y_1 \neq Y_2$, or the probability distribution $P(Y_1|x_1) \neq P(Y_2|x_2)$.

Given the original T_S in the D_S domain and the T_T in the D_T domain, the goal of transfer learning is to improve the quality of the prediction function $f_T(\cdot)$ of the target problem in the D_T domain using the knowledge gained from training the T_S problem in the D_S domain, where $D_T \neq D_S$ or $T_T \neq T_S$ (domains or tasks do not match).

There are the next types of transfer learning:
- instance-based transfer learning;
- feature representation transfer;
- parameter transfer;
- transfer learning for relational domains.

The main results of transfer learning are given in [27].

IX. PROBLEM SOLUTION

In this paper it is realized the direct artificial data propagation based on the transformations of the

initial training data, especially often when generating images, such transformations are used as rotation by a certain random angle, compression and stretching vertically and horizontally, tilt, mirror reflection, cropping, displacement, and many others [18] – [20]. This group of methods also includes the noise of the initial data, as well as various morphing transformations, similar to those described in [21], where new data are generated by "crossing" the initial data with each other.

An increase in the volume of the training sample due to is most often used in image recognition problems, therefore, these methods are primarily focused on image processing. Especially often when generating images, such transformations are used as rotation by a certain random angle, compression and stretching vertically and horizontally, tilt, mirror reflection, cropping, displacement, and many others [18] – [20]. In article [22], algorithms were proposed for introducing artificial realistic deformation of face images, on the basis of which the initial training sample of original images was multiplied and used to train the Viola–Jones algorithm (an AdaBoost class algorithm). It is shown that in a similar way the size of the initial training sample can be reduced by a factor of 10 (1000 original face images instead of 10,000) with a decrease in the recognition probability by no more than 2–4%.

The problem of deep NN training for image processing used in the medical diagnostics system is the complexity of obtaining a training sample. The training sample is obtained on the basis of the available small number of examples by rotating them at small angles in different planes with subsequent image processing in a different way [16].

X.   RESULTS

In the training sample [reference to the sample], 305 images of MRI results were considered to analyze the stages of liver fibrosis.

The number of images of MRI results for each stage of fibrosis:

$$F0 - 78 \quad F1 - 38 \quad F2 - 67$$

$$F3 - 47 \quad F4 - 85.$$

Stages of fibrosis for each image were determined by biopsy. There were 229 pictures in the initial sample. There were 76 images in the control sample.

The neural network was trained on an NVIDIA Tesla K80 computing processor with 12 GB of dedicated video memory. To implement the neural network, the Python programming language was used using the Keras library (with the TensorFlow backend) as a high-level neural network library).

Resnet101 [26] was used as the neural network.

During the experiment, the network was first trained on the original training sample for 50 epochs. (Fig. 1).

The neural network was then tested on a test dataset. The result of the network on the test data set for each of the classes (Fig. 2).

Image modification in order to increase the training sample is actively used in the training of neural networks, as well as in conditions of shortage of tagged data. Used contraction (stretching), rotation, random shift in color space, blurring of some pixels, scaling.

To improve the network recognition results of defective classes, it was decided to use the methods described above to increase the data set.

The scaling and blurring method of some pixels was checked. To do this, the experiment took a sample with a defective class F2 and increased the number of images in it by the methods described above.

Then the neural network was trained in this sample for 100 epochs (Fig. 3).
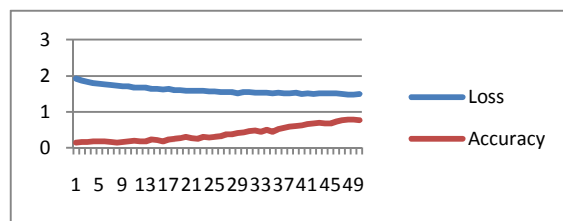

Fig. 1. Graph of losses and accuracy when learning the network Resnet101 for 50 epochs on the initial data set
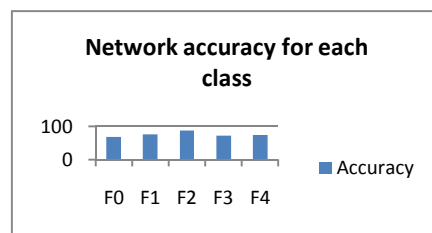

Fig. 2. The overall accuracy of the model was 76.417%
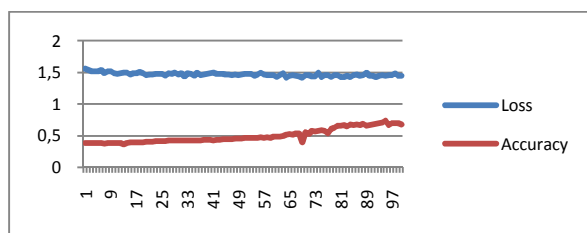

Fig. 3. Graph of losses and accuracy when learning the Resnet101 network for 100 epochs on a data set with a defective class F2, which was artificially expanded by scaling and blurring some pixels

The neural network was then tested on a test data set. The result of the network on the test data set for each of the classes (Fig. 4).
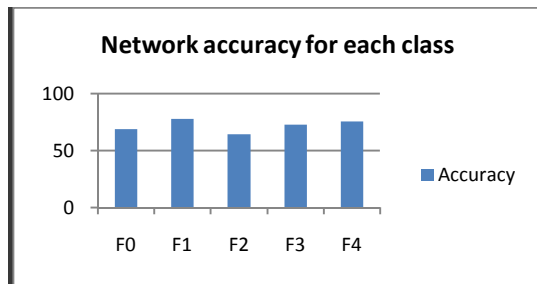
Network accuracy for each class

Fig. 4. The overall accuracy of the model was 71.4%

XI. CONCLUSIONS

It is considered the problem of intelligence diagnostic system construction for identification of liver fibrosis stages. As a principal technical mean it is used Magnetic Resonance Imaging. It is justified the need of Residual neural network for image processing of MRI results for identification of liver fibrosis stages due to its high accuracy of classification problem solution, taking into account the necessity of liver texture analysis. The absence of quality and enough learning sample is compensated by transfer learning approach use. The final diagnosis is formed with help of fuzzy inference system use.

Although liver biopsy is the standard of reference for diagnosis and staging of liver fibrosis, it is an invasive procedure with possible complications , so the proposal approach is less traumatic but more complicated.

REFERENCES

[1] K. Vorontsov, "Matematicheskie metody obucheniya po pretsedentam (teoriya obucheniya mashin)," *Mathematical Methods of Training on Precedents (the Theory of Machine Learning)*. Avaliable at: http://www.machinelearning.ru/wiki/images/6/6d/Voron-ML-1.pdf (accessed September 2015).

[2] V. P. Borovnikov, *Statistica for Students and Engineers*. Moscow: Computer Press, 2011, 301 p. (in Russian)

[3] Yu. B. Kotov, *New mathematical approaches to problems of medical diagnostics*. Moscow: Edinotorial URSS, 2004, 328 p. (in Russian)

[4] URL: http://www.science-education.ru/113-11527, (date of the application: 18.02.2014).

[5] M. V. Artemenko and A. S. Babkov, "Classification of methods for predicting the behavior of systems," *Modern problems of science and education*, no. 6, 2013. URL: www.science-education.ru/113-11527 (date of the application: 19.07.2014).

[6] N. A. Korenevsky and E. B. Ryabkova, "Method for the synthesis of fuzzy decision rules based on information about the geometric structure of multidimensional data," *Bulletin of the Voronezh State Technical University*, vol. 7, no. 8, pp. 128–137, 2011. (in Russian)

[7] V. P. Borovnikov, *Statistica for students and engineers*. Moscow: Computer Press, 2011, 301 p. (in Russian)

[8] E. Mangalova and I. Petrun'kina, "Prediction Capacity of Wind Power Plants Based on Non-Parametric Algorithm, K Nearest to Neighbors," *Doklady vserossiyskoy nauchnoy konferentsii AIST'2013 [Reports of the All-Russian Scientific Conference AIST'2013]*. Ekaterinburg, 2013, pp. 1–8. (in Russian)

[9] Labeled Faces in the Wild. http://vis-www.cs.umass.edu/lfw/.

[10] The Facial Recognition Technology (FERET) Database. http://www.itl.nist.gov/iad/humanid/feret/feret_master.html

[11] E. Mangalova and I. Petrunkina, "Forecasting the power of wind power plants on the basis of a nonparametric algorithm of k nearest neighbors," *Reports of the All-Russian scientific conference AIST'2013*. 2013, pp. 1–8. (in Russian)

[12] R. Latypova, Neural networks [Text]. Moscow: LAP Lambert Academic Publishing, 2012, 180 p. (in Russian)

[13] A. Radhakrishna, Frequency-tuned Salient Detection [Electronic resource]. Access mode: http://infoscience.epfl.ch/ (date of the application: 05.05.2019)

[14] O. Canavet and F. Fleuret, "Efficient Sample Mining for Object Detection," *Proceedings of the Asian Conference on Machine Learning (ACML)*, 2014, pp. 48–63.

[15] Amazon Mechanical Turk. Avaliable at: https://www.mturk.com/mturk/welcome (accessed September 2015)

[16] Michael Z. Zgurovsky, Victor M. Sineglazov, Olena I. Chumachenko, *Artificial Intelligence Systems Based on Hybrid Neural Networks*, Springer, 2020, 390 p. [Electronic resource]. Access mode: https://link.springer.com/book/10.1007/978-3-030-48453-8. Customer can order it via https://www.springer.com/gp/book/9783030484521

[17] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov, "Dropout: A Simple Way to Prevent Neural Networks From Overfitting," *The Journal of Machine Learning Research*, vol. 15, no. 1, pp. 1929–1958, 2014.

[18] L. Yaeger, R. Lyon, and B. Webb, "Effective Training of a Neural Network Character Classifier for Word Recognition," *NIPS*, 1996.

[19] D. C. Ciresan, U. Meier, L. M. Gambardella, and J. Schmidhuber, "Deep Big Simple Neural Nets Excel on Handwritten Digit Recognition," *Neural Computation.* vol. 22(12), 2010.

[20] P. Y. Simard, D. Steinkraus, and J. C. Platt, "Best Practices for Convolutional Neural Networks Applied to Visual Document Analysis*," Int'l Conf. Document Analysis and Recognition*, 2003.

[21] S. V. Kachalin, "Increasing the stability of learning large neural networks by adding small training samples of parent examples, synthesized biometric descendant examples," *Proceedings of the scientific and technical conference of the cluster of Penza enterprises that ensure the security of information technologies.* Penza-2014, vol. 9, pp. 32–35. (in Russian)

[22] A. V. Akimov and A. A. Sirota, "Models and algorithms for artificial data propagation for training face recognition algorithms using the Viola–Jones method," *Computer Optics*, vol. 40, no. 6, pp. 911–918, 2016. (in Russian)

[23] H. Guo and H. L. Viktor, "Learning from Imbalanced Data Sets with Boosting and Data Generation: The DataBoost IM Approach," *ACM SIGKDD Explorations Newsletter*, vol. 6(1), pp. 30–39, 2004.

[24] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic Minority Oversampling Technique," *J. Artificial Intelligence Research*, vol. 16, pp. 321–357, 2002.

[25] Nitesh V. Chawla, Aleksandar Lazarevic, Lawrence O. Hall and Kevin W. Bowyer, "SMOTEBoost: Improving Prediction of the Minority Class in Boosting," *in 7th European Conference on Principles and Practice of Knowledge Discovery in Databases– Cavtat-Dubrovnik*, Croatia, September 22–26, 2003, pp. 107–119.

[26] He K, Zhang X, Ren S, Sun J (2015a) Deep Residual Learning for Image Recognition. Multimed Tools Appl 77:10437–10453. doi:10.1007/s11042-017-4440-4

[27] Karl Weiss, Taghi M. Khoshgoftaar and DingDing Wang, "A survey of transfer learning," *J Big Data*, 3:9, pp. 1–40, 2016. doi:10.1186/s40537-016-0043-6

[28] Jing Zhang, Wanqing Li, and Philip Ogunbona, "Recent Advances in Transfer Learning for Cross-Dataset Visual Recognition: A Problem-Oriented Perspective," University of Wollongong, Australia DONG XU, University of Sydney, Australia. arXiv:1705.04396v3 [cs.CV] 20 May 2019.

Chumachenko Olena. orcid.org/0000-0003-3006-7460.
Doctor of Engineering Science. Associate Professor.
Technical Cybernetic Department, National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute," Kyiv, Ukraine.
Education: Georgian Polytechnic Institute, Tbilisi, Georgia, (1980).
Research area: system analysis, artificial neuron networks.
Publications: more than 80 papers.
E-mail: chumachecko@tk.kpi.ua

Kot Ananatoliy. Post-graduate student.
Technical Cybernetic Department, National Technical University of Ukraine "Ihor Sikorsky Kyiv Polytechnic Institute," Kyiv, Ukraine.
Education: National Technical University of Ukraine "Ihor Sikorsky Kyiv Polytechnic Institute," Kyiv, Ukraine, (2017).
Research area: artificial Intelligence.
Publications: 6.
E-mail: anatoly.kot@gmail.com

О. І. Чумаченко, А. Т. Кот. Формування навчального набору для задачі обробки зображень
Розглянуто проблему формування навчального набору для задачі обробки зображень. Показано, що ця задача має велике значення при побудові інтелектуальних медичних діагностичних систем, в яких згорткові нейронні мережі використовуються для обробки зображень (результати УЗД, КТ та МРТ). Через відсутність елементів навчальної вибірки пропонується, з одного боку, використовувати підходи штучного множення даних на основі початкової навчальної вибірки фіксованого обсягу, а з іншого боку, використовувати методи, що зменшують потреба у великих навчальних зразках як за допомогою використання ансамблевої топології (гібридні нейронні мережі), так і шляхом застосування підходу до трансферного навчання. Розроблений алгоритм формування навчального набору для завдань обробки зображень, заснований на модифікації вихідної вхідної інформації з розрахунком міри достовірності отриманої вибірки. Умови індексу – навчання за допомогою трансферу навчальний набір; згортка нейронних мереж; ансамблева топологія; обробка зображень.
Ключові слова: навчання за допомогою трансферу; навчальний набір; згорткові нейронні мережі; ансамблева топологія; обробка зображень.

Чумаченко Олена Іллівна. orcid.org/0000-0003-3006-7460.
Доктор технічних наук. Доцент.

Кафедра технічної кібернетики, Національний технічний університет України «Київський політехнічний інститут ім. Ігоря Сікорського», Київ, Україна.
Освіта: Грузинський політехнічний інститут, Тбілісі, Грузія, (1980).
Напрямок наукової діяльності: системний аналіз, штучні нейронні мережі.
Кількість публікацій: понад 80 наукових робіт.
E-mail: chumachecko@tk.kpi.ua

Кот Анатолій Тарасович. Аспірант.
Кафедра технічної кібернетики, Національний технічний університет України «Київський політехнічний інститут ім. Ігоря Сікорського», Київ, Україна.
Освіта: Національний технічний університет України «Київський політехнічний інститут ім. Ігоря Сікорського», Київ, Україна, (2017).
Напрям наукової діяльності: штучний інтелект.
Кількість публікацій: 6.
E-mail: anatoly.kot@gmail.com

Е. И. Чумаченко, А. Т. Кот. Формирование обучающей выборки для задачи обработки изображений

Рассмотрена задача формирования обучающей выборки для задачи обработки изображений. Показано, что эта задача имеет большое значение при построении интеллектуальных медицинских диагностических систем, в которых используются сверточные нейронные сети для обработки изображений (результатов УЗИ, КТ и МРТ). Из-за отсутствия элементов обучающей выборки предлагается, с одной стороны, использовать подходы искусственного умножения данных на основе начальной обучающей выборки фиксированного объема, а с другой стороны, использовать методы, уменьшающие потребность в больших обучающих выборках, как за счет использования ансамблевой топологии (гибридные нейронные сети), так и за счет применения подхода трансферного обучения. Разработан алгоритм формирования обучающей выборки для задач обработки изображений, основанный на модификации исходной входной информации с вычислением меры достоверности полученной выборки.
Ключевые слова: трансферное обучение; обучающий набор; сверточные нейронные сети; ансамблевая топология; обработка изображения.

Чумаченко Елена Ильинична. orcid.org/0000-0003-3006-7460.
Доктор технических наук. Доцент.
Кафедра технической кибернетики, Национальный технический университет Украины «Киевский политехнический институт им. Игоря Сикорского», Киев, Украина.
Образование: Грузинский политехнический институт, Тбилиси, Грузия, (1980).
Направление научной деятельности: системный анализ, искусственные нейронные сети.
Количество публикаций: более 80 научных работ.
E-mail: chumachecko@tk.kpi.ua

Кот Анатолий Тарасович. Аспирант.
Кафедра технической кибернетики, Национальный технический университет Украины «Киевский политехнический институт им. Игоря Сикорского», Киев, Украина.
Образование: Национальный технический университет Украины «Киевский политехнический институт им. Игоря Сикорского», Киев, Украина, (2017).
Направление научной деятельности: искусственный интеллект.
Количество публикаций: 6.
E-mail: anatoly.kot@gmail.com