IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 29 February 2024, accepted 17 May 2024, date of publication 22 May 2024, date of current version 31 May 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3404371

RESEARCH ARTICLE

Application a Committee of Kohonen Neural Networks to Training of Image Classifier Based on Description of Descriptors Set

VOLODYMYR GOROKHOVATSKYI[1], **IRYNA TVOROSHENKO**[1], **OLENA YAKOVLEVA**[1,2,3], **MONIKA HUDÁKOVÁ**[2], **AND OLEKSII GOROKHOVATSKYI**[4]

[1]Department of Informatics, Kharkiv National University of Radio Electronics, 61166 Kharkiv, Ukraine
[2]Department of Economics and Finance, Economics and Management Institute, Bratislava University of Economics and Management, 851 04 Bratislava, Slovakia
[3]Department of AI Research and Development, SYTOSS Ltd., 831 04 Bratislava, Slovakia
[4]Department of Informatics and Computer Engineering, Simon Kuznets Kharkiv National University of Economics, 61166 Kharkiv, Ukraine

Corresponding author: Iryna Tvoroshenko (iryna.tvoroshenko@nure.ua)

This work was supported by the European Union NextGenerationEU through the Recovery and Resilience Plan for Slovakia under Project 09I03-03-V01-00115.

ABSTRACT The research aims to improve structural methods of image classification based on a description as a set of keypoint descriptors. The focus is on implementing a classifier training using a committee of Kohonen networks separately for each description in an etalon database. The training result is a fixed set of data centroids, which ensures high classification speed. The improvement comprises implementing independent training for each etalon, which increases the accuracy of approximating descriptions with a set of centroids and, in general, guarantees that the classification efficiency remains at a decent level. Calculating the cluster centroids for each class prevents the influence of descriptors from other classes. In addition, independent training is effective in cases when the powers of etalon descriptions differ. The classification speed of the proposed method based on training a committee of Kohonen networks, compared to the traditional linear search method, increases proportionally to the ratio of the description power and the number of generated centroids. The presented results of experimental modeling of the proposed methods comprise a database of images of coins. The test sample is formed as a set of images from the etalon database by applying geometric transformations of shift, scale, and rotation. The testing stage has shown a high level of classification accuracy after a proposed training and revealed a practical opportunity to choose a network structure and parameters capable of providing the required level of accuracy and speed criteria for classification for an applied task.

INDEX TERMS Classification accuracy, classifier training, image classification, Kohonen network, set of descriptors.

I. INTRODUCTION

The problem of effective classification of visual object images under conditions of geometric transformations is one of the most difficult for modern computer vision systems [1], [2], [3], [4]. In recent years, its practical solution has been greatly facilitated by introducing neural networks [5], [6], [7]. A set of images of a particular class is fed to the network input after preliminary processing by humans to highlight the boundaries of the object in the background [8]. The network summarizes the data of each class on the basis of the analysis and processing of image fragments for the system and determines the characteristics of the class using the training procedure. The obtained characteristics are then used in the classification process as parameters, vectors, or weighting coefficients.

At present times, structural classification methods are also being intensively developed, where decision-making is based

The associate editor coordinating the review of this manuscript and approving it for publication was Zhenhua Guo.

on describing the image as a set of keypoint descriptors [2], [9], [10]. Keypoints and their descriptors are formed by special filters – detectors [11], [12].

Modern structural methods implement the classification process based on the "bag of words" model, where the class of an object is determined by the result of voting of the components – descriptors of the recognized object [2], [13]. In the traditional approach, the base set consists of the descriptions of etalons, and in the modifications, the same base is the centers or parameters of the descriptions [9], [14], [15]. Such methods are advantageous over neural networks, as they focus on alignment with the description or parameters of the etalons and do not require long-term training [10]. In addition, the application of such methods allows for the rapid change of the etalon base content and ensures invariance with geometric transformations.

Taking into account the mentioned advantages, it is of practical interest to combine the keypoint descriptors and neural network learning tools in classification methods. Such an approach positively simplifies the learning process by pre-selecting a set of descriptors for the etalon base and, at the same time, provides high performance in terms of both processing speed and classification accuracy [16]. In addition, the combined approach can be universally applied to visual data and also to any data that is given as a set of multidimensional vectors [17], [18].

The research aims to improve the performance of structural classification methods by implementing a training procedure using a committee of Kohonen networks. The primary focus is on increasing the level of classification accuracy while maintaining high performance after implementing the training results.

The objectives of the research are to study and compare various methods of network training, including ensembles of networks with separate training for etalon descriptions. The key task is to study the effectiveness of implemented learning models in classification methods, as well as to conduct an experimental evaluation of the proposed methods on the basis of image database analysis.

The article proposes:

1) A formal classification model using the results of separate multi-module training on a set of etalon descriptions. The ways of implementing the Kohonen network as a committee are analyzed. Implementing the presented ideas ensures high performance and speed of classifiers.
2) Ways to organize training on a set of etalon descriptions as a set of descriptors. The schemes of training on the set of descriptors of the entire database, as well as descriptions of the separate etalons, are discussed. A scheme for constructing a classifier based on the feature system formed after training is proposed.
3) Results of computer modeling. Software tools for implementing the discussed learning models and classification methods have been developed. Testing is performed on an image database. The experimental results confirm the effectiveness of the proposed approach.

II. LITERATURE REVIEW

Data quantization and neural networks are currently powerful tools for knowledge acquisition [5], [6], [7], [16], [19], [20], [21], [22]. Regarding object descriptions as a set of descriptors, the mentioned tools can significantly speed up the image classification process by pre-grouping large amounts of data [9], [13], [20]. This achievement results from high-speed data search when the input data is compared with the etalon data stored in the system's memory [9], [14].

One of the practically effective networks used for a wide range of data is the Kohonen network [19], [20]. Up to now, the use of the Kohonen network in performing a self-learning procedure for a set of descriptors and classifying images based on the obtained parameters has been studied [16], [23]. At the same time, various options for determining the centroids of quantized data both to form feature systems and to use in the learning process are discussed [21], [24].

However, a deeper study of network applications for data as a set of descriptors shows [2] that the vector values of data centroids obtained for different images can be very close to each other. The multidimensionality of the descriptors [13] that reflect the features of visual images is not the only explanation. In fact, the image signal is quite large in size and almost immensely diverse, so it actually contains a huge amount of information. It means that even very dissimilar images can often comprise almost identical or close fragments, which leads to similarity in their representation in the feature space. The closeness of the characteristics of the centroids calculated from the overall class training results reduces the classification accuracy in voting methods that implement the bag-of-words technology [2], [25].

To avoid or reduce this factor impact, it is necessary to increase the accuracy of the data representation for training and subsequent classification. One way is to obtain a concentrated description separately for each etalon as a result of training as the representation of not a single but several data centers [26], [27]. In this case, a network committee for training is used, where each component independently forms a fixed number of description centers for a separate etalon. The number of components of such an ensemble coincides with the number of etalons. At the same time, the Kohonen network apparatus acts as a means for the concentrated representation of a vast number of multidimensional vectors as a limited set of data centers. In addition, if a network is trained on a separate etalon, it helps to increase the consistency of the centers with the data in terms of the efficiency of further classification.

Note that along with committee usage, the classification scheme based on determining the relevance of the descriptions of the analyzed object and the etalon remains unchanged [2], [26]. Class vote counting can be organized independently for each etalon [9] or competitively on a set of calculated centroids for the entire database [26].

The discussed principle of improvement by quantitative expansion of data representation can be universally implemented for other methods of data concentration [14], [24], such as clustering [9], hashing [13], and estimation [28]. The parameter of a certain number of centers significantly affects the classification performance in general.

The author's monograph [19] contains the result of studying the variety of approaches to the construction and training of Kohonen networks, which makes it possible to adapt the network structure to the method of processing and the type of data being analyzed. In general, the use of the Kohonen network in computer vision systems is aimed at identifying the most important image features, which are subsequently used for recognition [20], [21], [22], [23], [29].

Researchers identify such positive properties of the Kohonen network as the ability to recognize stable clusters in the analyzed data and establish the proximity of classes [19]. Thus, it is possible to improve the understanding of the studied data structure in order to implement effective classification rules when classes are already defined by their representatives [16]. The Kohonen network is trained by the successive approximations method [21]. In this case, the training objective is not only error minimization but also adaptation of the internal network parameters to maximum consistency with the input data [22]. In [16], we proposed a network variant that trains on a set of descriptors for the full content of the etalon database. An analysis of metrics for classification based on integrated image features is presented in [30].

The article [31] discusses options for ensemble training of convolutional neural networks for image recognition. The effectiveness of various rules for voting the outputs of these networks is studied: hard (majority) voting, where the minority of votes obeys the majority; hard voting with a threshold for the minimum number of votes received; weighted and soft probabilistic voting. The same approaches are also relevant for classification based on voting of description descriptors [10], [26]. In general, it is currently believed that ensemble learning using several simple network models improves classification performance while maintaining or increasing the accuracy rate compared to a single but complex learning model [32], [33], [34], [35].

As it may be concluded from the literature analysis, introducing neural network learning into structural classification methods can not only improve classification speed but also provide a sufficient level of efficiency.

III. CLASSIFICATION BY DESCRIPTION AS DESCRIPTORS SET

The classification is performed within a database of N etalons (class representatives) as a set of E descriptions of etalon images: $E = \{E_1, E_2, \ldots, E_N\}$. E is a training sample and also the basis for classification by comparison with the etalon [9]. Each etalon description $E_k \subset E$ in the classifier formalism represents a separate class. The class k with the description E_k is defined as an infinite set of images obtained from the etalon k by applying to it a multi-parameter group of geometric transformations, including displacement, rotation, scaling, the effect of which does not remove the object of interest from the field of view [1], [3], [14]. The fact is that modern keypoint detectors, such as Oriented FAST and Rotated BRIEF (ORB), Binary Robust Invariant Scalable Keypoints (BRISK), accelerated-KAZE (AKAZE), ensure the invariance of the resulting description to such a group of transformations [11], [12].

The description of the etalon $E_k = \{e_v(k)\}_{v=1}^{s}$ is given as a finite set of S descriptors of keypoints in the space B^n of binary vectors, $e_v(k) \in B^n$, $s = card\ E_k$ is the number of descriptors in the set [2]. Each descriptor of the $e_v(k)$ database E has a parameter k of the class number, and the total number of features – descriptors in the base set E is $card\ E = sN$. The classifier R_1 is based on the element-by-element analysis of the input description $Z = \{z_v\}_{v=1}^{s}$ of the object, assigns the descriptor $z_v \in Z$ to one of the classes according to the rule

$$R_1 : z_v \to \{1, \ldots, N\}. \tag{1}$$

The class k of the analyzed descriptor z_v of the object according to rule (1) is defined as the argument of the minimum distance on the set E of descriptors of all classes

$$R_1 : k = \arg \min_{i=1,\ldots,N;\ d=1,\ldots,s} \rho(z_v, e_d(i)). \tag{2}$$

In the competitive model (2), it is advisable to use the Hamming distance, which calculates the number of distinct bits in two binary vectors [13], [26]. Equation (2) minimizes the value of the distance by the variable i (class number) on the set of descriptors of the database E. At the same time, equation (2) can organize the determination of the minimum both for the full list of $s{\cdot}N$ database descriptors and separately for each of the N etalons (with a limit on the value of the minimum), followed by the determination of the winning class by the highest number of votes [14].

Let's introduce a vector $\{h_i\}_{i=1}^{N}$ with integer values to accumulate the class votes obtained by applying model (2) to the entire set of object descriptors Z. Based on the implementation of R_1 for each $z_v \in Z$ in accordance with (2), we determine the class number k, and then increment the accumulator $h_k = h_k + 1$ for the corresponding class number. In fact, the vector $\{h_i\}_{i=1}^{N}$ is the distribution of the number of votes for the object descriptors in the class system.

As a result of processing the description Z of the object, we accumulate the vector $\{h_i\}_{i=1}^{N}$. The class of the object is determined by the rule R_2 as the maximum argument

$$R_2 : Z \to E_k \left| (k = \arg \max_{i=1,\ldots,N} h_i)\ \&(h_k \geq \delta_h), \right. \tag{3}$$

where δ_h is a threshold for the minimum number of votes, which is determined experimentally for a given database as a set of etalons. If the $h_k \geq \delta_h$ condition is not met, the object class is not set (rejection of classification).

The sequence of classification rules R_1, R_2 implements a classifier based on the decisions of local classifiers obtained for the object's components [2], [13]. It is resistant to distortions of individual components due to possible interference [36].

IV. CLASSIFICATION WITH THE IMPLEMENTATION OF TRAINING OUTCOMES

A. APPLICATION OF THE KOHONEN NETWORK FOR DATA CENTERS

The procedure of quantizing the data on M of non-intersecting groups by the Kohonen network (vector quantization) [16], [19] is applied to each etalon description E_i, $i = 1, \ldots, N$, independently of other etalon descriptions. Note that the result of quantization depends on the initial M values of the centroids $W = \{w_j\}_{j=1}^{M}$ of the network [37]. The training process is carried out according to the competition model on the set of centroids $W = \{w_j\}_{j=1}^{M}$, where for the data component z the number v of the nearest center is determined as

$$v = \arg \min_{j=1,\ldots,M} \rho(w_j, z). \tag{4}$$

Most often, a model (4) is used in an online training scheme where data is received one by one, and the system responds by changing the centroid system at each step by modifying the neuron with the number v:

$$w_v(t+1) = w_v(t) + \alpha(t)(z(t) - w_v(t)). \tag{5}$$

In the model (5), t is the training iteration step, $z(t)$ is the value of the learning component at the step t, and $\alpha(t)$ is the training rate coefficient. Expressions (4), (5) are one of the common variants of the winner-takes-it-all rule if only the center defined by (4) changes [19]. When training on the set E_i of descriptors, the value t means the number of the descriptor in the description: $t = \{1, 2, \ldots, s\}$.

Given the binary nature of the analyzed data, to ensure convergence and reduce the training time, data normalization is implemented by dividing by the Hamming norm $\chi[z]$ for the binary vector z (number of unit bits) [13], [30].

$$\chi[z] = \sum\nolimits_{\alpha=1}^{n} z^{\alpha},$$

where z^{α} is the bit with the number α in the binary vector $z \in B^n$. In traditional training schemes, the coefficient $\alpha(t)$ decreases from some initial value as t increases. The most common model is $\alpha(t) = \alpha_0/t$ with some a priori set initial value α_0 of the training rate.

The data quantization error is traditionally calculated using the formula [37]:

$$\beta = \sum\nolimits_{i=1}^{M} \sum\nolimits_{v=1}^{s(i)} \rho^2(z_v, w_i), \tag{6}$$

where $s(i)$ is the power of the i-th cluster created as a result of training, z_v is the cluster element with the number v. Note that the value (6) as a characteristic of training efficiency does not directly affect the accuracy of data classification, but is only an a posteriori assessment of the cohesion of data around the centers identified as a result of training.

An important aspect of the training process, the results of which will be used for classification, is the selection of initial centroid values [19]. To do this, it is proposed to select the descriptor with the highest value of the informativeness parameter for each etalon descriptions. We calculate the informativeness of an arbitrary descriptor as the difference between the smallest distances to the set of descriptors from other etalons and our own etalon [16], [25].

Formally, for the vector $z \in E$ as a component element of $z \in E_i$ in the class system E, the criterion of information content $V(z, E)$ is introduced

$$V(z, E) = \rho_m(z, \overline{E_i}) - \rho_m(z, E_i), \tag{7}$$

where $\rho_m(z, \overline{E_i}) = \min_{v,u \neq i} \rho(z, e_v(u))$ is the distance from z to the set of elements of the database that do not belong to the class E_i, $e_v(u) \in E \backslash E_i$, $\rho_m(z, E_i) = \min_{v,u=i} \rho(z, e_v(u))$ is the distance from z to the nearest element from the class E_i.

In models (4)-(7), the Hamming metric can be used if the data centers are reduced to a binary form. However, requiring data normalization, it is more practical to use Euclidean or Manhattan metrics when implementing training (4), (5) [30].

B. CLASSIFICATION BASED ON TRAINING

The use of the system of centroids obtained after training for the classification process transforms model (2) into the form

$$R_1 : k = \arg \min_{i=1,\ldots,N;\ d=1,\ldots,M} \rho(z_v, w_d(i)), \tag{8}$$

where the minimum search is now performed exclusively on the set of created data centers.

Note that the set of centers $W = \{w_j\}_{j=1}^{M}$ has a much smaller power M compared to the power s of the set of description descriptors, as $M \ll s$. This result dramatically gains in processing speed when switching from a linear search within the database to a search on the set of centers [9], [14]. For example, for specific values of $s = 500$, $M = 3$, when switching from the model (2) to the model (8), processing speed increased more than 150 times!

Note that the classification model (8) can use both a metric for vectors with real values (e.g., Manhattan [30]) and the Hamming metric, as in the model (2). In the second case, the components of the $W = \{w_j\}_{j=1}^{M}$ centroid system should be transformed into a binary form.

If Kohonen's training is applied to the complete database E [16] without dividing E into subsets E_i, a set of L centroids $W = \{w_j\}_{j=1}^{L}$ common to the database E is gained. Then the model (8) for the first stage R_1 is transformed into the definition of the nearest center with the number

$$u = \arg \min_{d=1,\ldots,L} \rho(z_v, w_d). \tag{9}$$

As a result of applying equation (9) $\forall z_v \in Z$, some integer vector of the cluster representation $g(Z) = [g_1, g_2, \ldots, g_L]$ through the system of centers [9] is obtained for the analyzed object Z. Here, g_d is the number of descriptors of the analyzed description that fell into the cluster under the number d.

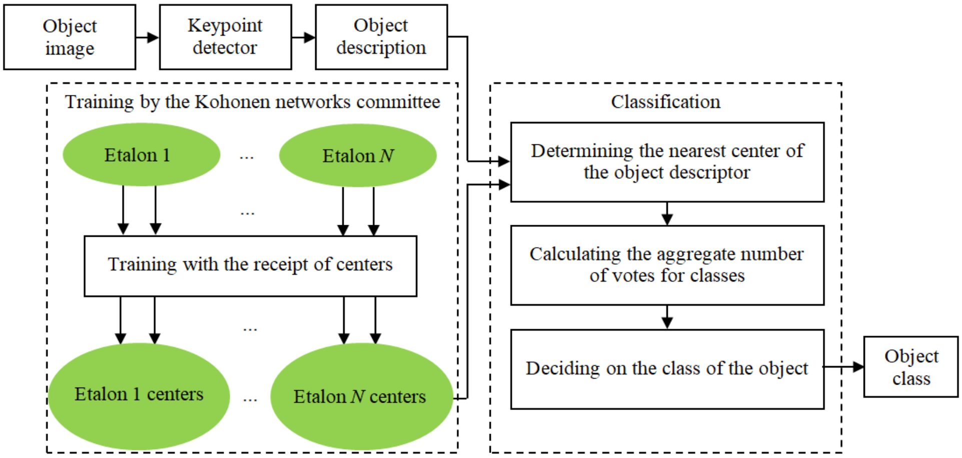


FIGURE 1. **Classification scheme using multi-module training.**

The vector $g(Z)$ is the distribution of the data of the analyzed description by a set of clusters, in contrast to $\{h_i\}_{i=1}^{N}$ as a distribution by a system of classes. In the same way, the cluster representation $g(E_i)$ can be obtained for each of the etalons E_i. The completed classification model (3) can now be implemented as

$$R_2 : Z \to E_k \left| k = \arg \min_{i=1,\ldots,N} \rho(g(Z), g(E_i)) \right., \tag{10}$$

where $\rho(g(Z), g(E_i))$ is the distance (e.g., Manhattan) between the vectors of the cluster representation of the object $g(Z)$ and the etalon $g(E_i)$. The constraint for classification according to the equation (10) is the condition $\rho(g(Z), g(E_k)) \leq \delta_\rho$, where δ_ρ is the largest acceptable value of the minimum distance between cluster views.

The proposed classification scheme using data centers obtained by the Kohonen network during training is shown in Figure 1.

The proposed multi-module training method separately for each of the etalons in the description database (Figure 1) improves the degree of approximation and data separation, which generally improves the classifier's performance. The classification performance will be evaluated traditionally by the value of the accuracy index pr, which is calculated by the ratio of the number of correctly classified objects q to their total number Q, used in the experiment [13], [30], [36].

$$pr = q/Q. \tag{11}$$

V. EXPERIMENTAL RESULTS AND DISCUSSION

A. METHODS AND CONDITIONS OF TRAINING

The simulation of the proposed methods was performed in the Google Colaboratory software environment in the Python programming language using the NumPy library for model building and fast computation, as well as the OpenCV computer vision library [38], [39].

Let's experimentally evaluate and compare the effectiveness of classification methods with these types of learning methods using software modeling:

1) Kohonen joint network using data from the complete database E of descriptions with the number of $L = N = 5$ centroids according to the model (9).
2) A committee of N Kohonen networks with one neuron ($M = 1$) for each etalon E_i, $i = 1, \ldots, N$ (each network is trained on a set of descriptors of its etalon).
3) A committee of N Kohonen networks (M centers for each etalon E_i).

The training of method 1 differs from the others as the training dataset uses the aggregate database of descriptors of all the etalons, which are then distributed among N class centers by a competitive quantization procedure. Thus, a traditional clustering method similar to k-means is implemented here, which is actually a batch variation of the Kohonen network training algorithm [16], [19], [21].

Methods 2 and 3 train separately for each etalon description. Separate training makes it possible to create a multi-module classifier (Figure 1), where each module is linked to a separate etalon. Training method 2 is a variant of more generalized method 3 for the case of $M = 1$.

At the beginning of training, for all methods, neurons with the values of the etalon descriptors that have the highest value of informativeness (8) are initiated, i.e., are the most unique representatives of their class. The data for training and classification are normalized by dividing by the Hamming norm $\chi[z]$ for a binary vector (number of single bits) [13]. To ensure the homogeneity of data processing in all the considered methods, the Manhattan metric is used for classification.

In method 3 of network training, the initial centers were randomly selected from the set of descriptors of each etalon. The flow of vectors from the description for training was also randomly generated.

Let's choose images of collectible coins with images of mythical gods as the etalons to be used for training the classifier (Figure 2) [40]. The size of the images is 770×770 pixels. The ORB detector is used to generate keypoints because its descriptor values are binary vectors

FIGURE 2. Images of the coins of the etalon base.

FIGURE 3. Examples of the set of coordinates ORB descriptors for etalons.

FIGURE 4. Examples of transformed images with keypoint coordinates.

of 256 bits [11], [12]. For the experiment, the value of the number of classes $N = 5$ was chosen, for method 3, and the number of centroids within the class $M = 3$. Figure 3 shows examples of etalon images with the generated keypoint coordinates.

To evaluate the classification accuracy, a test sample of images was formed and used, consisting of the etalons and images obtained by applying geometric transformations of rotation, scaling, and displacement to the images of the etalons. Examples of transformed images with keypoint coordinates are shown in Figure 4.

The total number of test images is 360. The classification accuracy was evaluated by the indicator (11). The class of the image was determined by the classifier R_2 by the maximum number of votes, and the truth of the solution was evaluated by the correspondence to the correct class.

B. ANALYSIS OF THE EXPERIMENTAL RESULTS

The testing was conducted taking into account the effect of the number of training epochs and the initial coefficient α_0 on the classifier performance using models (5), (9). During one training epoch, the entire set of descriptors is adjusted to the system of centers. The next epoch – training is repeated with a reduced coefficient α_0. The number of epochs varied from 5 to 25, and the α_0 coefficient varied between 0.1 and 0.9. An ensemble with a large number of epochs and a low initial learning coefficient is called a thorough configuration, and an ensemble with a small number of epochs and a large initial learning coefficient is called a fast configuration [7], [19], [31]. The following variants are considered in more detail: fast (5 epochs with an initial coefficient of $\alpha_0 = 0.75$, moderate (10 epochs, $\alpha_0 = 0.25$) and slow (50 epochs and coefficient $\alpha_0 = 0.1$).

The number of descriptors in the descriptions of the etalons and input (test) images is fixed at $s = 500$.

The Experiments Showed the Following Results: The classification on the test sample using training method 1 showed the following highest accuracy values: $pr = 0.4$ with $\alpha_0 = 0.25$ and the number of epochs 10; $pr = 0.39$ with $\alpha_0 = 0.75$ and the number of epochs 25.

The classification on the test set using training method 2 showed the following best accuracy values: $pr = 0.8$ with $\alpha_0 = 0.1$ and the number of epochs 5; $pr = 0.8$ with $\alpha_0 = 0.75$ and the number of epochs 25.

As you can see, training separately for each etalon yields a significantly higher classification accuracy of 0.8 compared to training on the full description base (accuracy of 0.4).

Now let's compare these indicators with the values of Table 1 obtained during training using the proposed method 3 for $M = 3$.

TABLE 1. Accuracy (11) of classification with training method 3.

Coefficient α_0	Number of epochs			
	5	10	25	50
0.1	0.769	0.800	**0.811**	0.788
0.25	0.636	0.658	0.802	0.755
0.5	0.580	0.627	**0.888**	**0.802**
0.75	0.619	0.660	0.788	0.750
0.9	0.630	0.602	0.683	0.722

The three most effective training configurations in terms of classification accuracy on the test set are highlighted in bold in Table 1. The results of Table 1 make it possible to determine, for practical purposes, the training characteristics that provide the required classification accuracy.

Comparing the obtained results, it can be concluded that the use of the Kohonen network for training the classifier leads to a significant increase in performance: the classification accuracy increases from the best results of 0.4 for the method without the committee to 0.8 for the etalon-independent training with one neuron and to an even higher level of 0.89 with three neurons. At the same time, the average value, taking into account all epochs and the learning rate, also increased significantly from 0.11 for method 1 to 0.713 for method 3.

The classification method based on the training method 1 showed significantly lower performance compared to the training method 2. Such low performance can be explained by the fact that in the training method 1, the training was performed on the aggregate descriptor base. In this case, the descriptors of "foreign" classes significantly affect the calculation of centers, as they can naturally be closer to the "foreign" center.

Table 2 shows the relative number of votes obtained for the transformed etalon images during the classification process when training with method 3 (500 descriptors, $\alpha_0 = 0.25$, 10 epochs). The row of Table 1 shows the distribution of the number of votes by the set of classes. Table 2 shows that, on average, test images from all classes are classified

TABLE 2. Relative number of votes of classes for training method 3.

Etalon number	Number of votes				
	1	2	3	4	5
Etalon 1	**0.383**	0.182	0.141	0.196	0.098
Etalon 2	0.134	**0.441**	0.117	0.187	0.121
Etalon 3	0.184	0.218	**0.269**	0.210	0.119
Etalon 4	0.157	0.201	0.154	**0.367**	0.121
Etalon 5	0.193	0.179	0.139	0.220	**0.269**

correctly since all the maxima of the number of votes of the "own" etalon are located on the diagonal. Unfortunately, mostly correct classification is not the case for learning methods 1 and 2. In addition, Table 2 shows that images of class 2 (the largest maximum) are classified best of all, and images of classes 3 and 5 are classified worst of all. The obtained indicators make it possible to evaluate and manage the effectiveness of the classification system by selecting the most effective training parameters and etalon images.

Separate experiments conducted for the proposed method 3 with a larger number of neurons per etalon confirmed a further increase in classification accuracy. With 5 neurons, the highest accuracy is $pr = 0.74$, with 10 neurons – $pr = 0.86$, with 25 neurons – $pr = 0.94$. At the same time, the average values, taking into account epochs and the training rate, also increased significantly. For example, the average values for 10 epochs are: 0.65 for 5 neurons, 0.80 for 10 neurons, and 0.91 for 25 neurons per etalon.

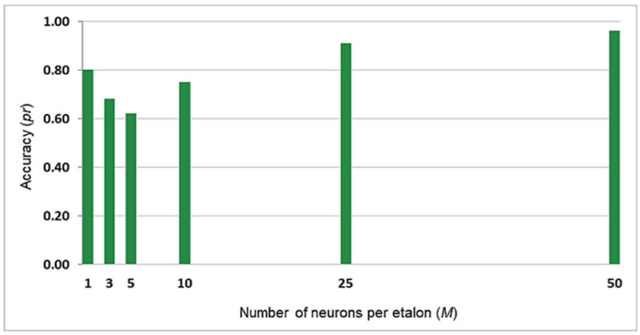


FIGURE 5. Typical dependence of classification accuracy on the change *M*.

Figure 5 shows a typical dependence of the classification accuracy rate (11) on the number of neurons per etalon for 10 epochs of training at $\alpha_0 = 0.25$. As it shown, there is a slight decrease in accuracy at the point $M = 5$, and then the growth of (11) continues.

As the number of M neurons grows, the time for training the model and performing classification also grows non-linearly. Figure 5 shows the dependence between the time of one training epoch (500 descriptors in the description) and the number of neurons. Table 3 shows that up to the value of $M = 5$, the time for testing slightly decreases, and the training time is minimal at 5 centroids simultaneously with the lowest accuracy (Figure 5). Table 3 also shows the classification time for one image depending on the number

TABLE 3. Dependence of training time and classification time on the parameter *M*.

Number of neurons per etalon	Time, s	
	training	classifications
1	1.68	0.061
3	1.82	0.043
5	0.40	0.010
10	1.70	0.037
25	1.86	0.052
50	2.28	0.086

of neurons used. After that, when the number of neurons exceeds 5, the training and classification time increases monotonically along with the accuracy. Repeated tests confirm the nonlinear nature of this relationship.

The obtained experimental data (Figure 5, Table 3) and the values of the tables similar to Tables 1 and 2 for the variety of training methods open up a practical opportunity to select specific parameters for an applied problem to implement the procedure for training a network of the desired quality. It is clear that training method 1 showed lower accuracy results mainly because the trained centers are obtained due to competition between classes. This competition resulted in some smoothing of the difference between the centers regarding training results, which affected the accuracy. It is known that the sets of descriptors of different classes have many common or close components [2], [13]. A more apparent difference between class centers for this method could be obtained by implementing supervised or teacher-assisted learning.

The classification speed for the proposed method based on training an ensemble of Kohonen networks (model (7)) compared to the traditional linear search method (model (2)) increases in proportion to the ratio s/M [9], [14]. For the values of $s = 500$, $M = 3$ used in the experiment, it resulted in an increasing processing speed of more than 150 times. As the number of M neurons increases, the classification accuracy naturally increases, but the gain in performance decreases. Based on our research, a network structure that will ensure the required level of classification accuracy and speed criteria in an applied task with specific descriptions of the etalons can be selected.

VI. CONCLUSION

This article develops the idea of combining the advantages of a neural network and a voting apparatus for descriptors of keypoints in an image into a single classification method. The committee of networks, through separate training, contributes to the qualitative formation of a limited set of integrated features of the etalon points, while voting based on consistency with the etalon descriptions simplifies the classification decision. The result is a high-speed classification method that provides a decent level of accuracy.

In this research, the network committee was used exclusively to train the classification system separately for each of the etalon images, which actually led to the synthesis of a multi-module classifier. The obtained data as a system of centroids is implemented in the classification process. Training of networks in an ensemble and matching with centroids can be organized in parallel. By separating the training sets and calculating the cluster centers for each class separately, we prevent the influence of descriptors from other classes. In addition, separate training can be implemented in situations where the descriptions of the etalons have different powers.

A significant contribution of this research is the introduction of the Kohonen network apparatus for class-separated etalon-based training of an image classifier on a description in the form of a set of keypoint descriptors. Based on the results of multi-module training, a system of centers for etalon descriptions was obtained, which significantly reduces the classification time. Experiments confirm the gain in performance with a guaranteed level of classification accuracy.

The discussed approaches to training and multi-module use of the Kohonen network can be universally applied to the analysis and classification of any objects whose descriptions are given by a set of multidimensional vectors.

Future research may be related to improving the classifier by using supervised learning tools for the network committee since the discussed approaches assume that class representatives – the etalons – are given. A supervised learning approach could also further improve the accuracy, especially for the first training method.

ACKNOWLEDGMENT

This work was supported by the European Union NextGenerationEU through the Recovery and Resilience Plan for Slovakia under Project 09I03-03-V01-00115. The authors acknowledge the support to the undergraduate students Oleksandr Kulyk and Viacheslav Klinov of the Department of Informatics for their participation in computer modeling and to the management of the Department of Informatics, Kharkiv National University of Radio Electronics for supporting the research.

CONFLICTS OF INTEREST

The authors declare that they have no conflicts of interest to report regarding the present research.

REFERENCES

[1] Q. Bai, S. Li, J. Yang, Q. Song, Z. Li, and X. Zhang, "Object detection recognition and robot grasping based on machine learning: A survey," *IEEE Access*, vol. 8, pp. 181855–181879, 2020, doi: 10.1109/ACCESS.2020.3028740.

[2] Y. Ibrahim Daradkeh, V. Gorokhovatskyi, I. Tvoroshenko, S. Gadetska, and M. Al-Dhaifallah, "Statistical data analysis models for determining the relevance of structural image descriptions," *IEEE Access*, vol. 11, pp. 126938–126949, 2023, doi: 10.1109/ACCESS.2023.3332291.

[3] Y. Putyatin, V. Gorohovatsky, A. Gorohovatsky, and E. Peredriy, "Projective methods of image recognition," in *Proc. XIV Int. Conf. Knowl. Dialogue Solution (KDS)*, Jul. 2008, pp. 37–42. [Online]. Available: https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=fca0c353f216d794477f1a89800746f0826d6056

[4] R. Szeliski, "Feature detection and matching," in *Computer Vision: Algorithms and Applications*. Cham, Switzerland: Springer, 2022, pp. 333–399.

[5] S. Zhang, J. Wang, X. Tao, Y. Gong, and N. Zheng, "Constructing deep sparse coding network for image classification," *Pattern Recognit.*, vol. 64, pp. 130–140, Apr. 2017, doi: 10.1016/j.patcog.2016.10.032.

[6] D. Padilla Carrasco, H. A. Rashwan, M. Á. García, and D. Puig, "T-YOLO: Tiny vehicle detection based on YOLO and multi-scale convolutional neural networks," *IEEE Access*, vol. 11, pp. 22430–22440, 2023, doi: 10.1109/ACCESS.2021.3137638.

[7] T. Yu, W. Chen, G. Junfeng, and H. Poxi, "Intelligent detection method of forgings defects detection based on improved EfficientNet and memetic algorithm," *IEEE Access*, vol. 10, pp. 79553–79563, 2022, doi: 10.1109/ACCESS.2022.3193676.

[8] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 2, pp. 318–327, Feb. 2020, doi: 10.1109/TPAMI.2018.2858826.

[9] Y. Ibrahim Daradkeh, V. Gorokhovatskyi, I. Tvoroshenko, and M. Zeghid, "Cluster representation of the structural description of images for effective classification," *Comput., Mater. Continua*, vol. 73, no. 3, pp. 6069–6084, Jul. 2022, doi: 10.32604/cmc.2022.030254.

[10] X. Zhang, F. X. Yu, S. Karaman, and S.-F. Chang, "Learning discriminative and transformation covariant local feature detectors," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 4923–4931, doi: 10.1109/CVPR.2017.523.

[11] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, "ORB: An efficient alternative to SIFT or SURF," in *Proc. Int. Conf. Comput. Vis.*, Nov. 2011, pp. 2564–2571, doi: 10.1109/ICCV.2011.6126544.

[12] *ORB Feature Detector and Binary Descriptor*. Accessed: Jan. 22, 2024. [Online]. Available: https://scikit-image.org/docs/dev/auto_examples/features_detection/plot_orb.html

[13] P. Flach, "Model ensembles," in *Machine Learning: The Art and Science of Algorithms That Make Sense of Data*. New York, NY, USA: Cambridge Univ. Press, 2012, pp. 330–342.

[14] V. A. Gorokhovatskiy, "Compression of descriptions in the structural image recognition," *Telecommun. Radio Eng.*, vol. 70, no. 15, pp. 1363–1371, Oct. 2011, doi: 10.1615/TelecomRadEng.v70.i15.60.

[15] S. Srivastava, P. Mukherjee, and B. Lall, "Characterizing objects with SIKA features for multiclass classification," *Appl. Soft Comput.*, vol. 46, pp. 1056–1066, Sep. 2016, doi: 10.1016/j.asoc.2015.12.014.

[16] V. Gorokhovatskyi and I. Tvoroshenko, "Image classification based on the Kohonen network and the data space modification," in *Proc. CEUR Workshop: Comput. Modeling Intell. Syst. (CMIS)*, vol. 2608, May 2020, pp. 1013–1026. [Online]. Available: http://ceur-ws.org/Vol-2608/

[17] H. Kuchuk, A. Podorozhniak, N. Liubchenko, and D. Onischenko, "System of license plate recognition considering large camera shooting angles," *Radioelectron. Comput. Syst.*, vol. 100, no. 4, pp. 82–91, Dec. 2021, doi: 10.32620/reks.2021.4.07.

[18] R. Scherer, "Computer vision methods for fast image classification and retrieval," in *Studies in Computational Intelligence*. Cham, Switzerland: Springer, 2020, pp. 107–118.

[19] T. Kohonen, "Learning vector quantization," in *Self-Organizing Maps*. Berlin, Germany: Springer, 2001, pp. 245–261.

[20] Z. Hu, Y. V. Bodyanskiy, and O. K. Tyshchenko, "Kohonen maps and their ensembles for fuzzy clustering tasks," in *Self-Learning and Adaptive Algorithms for Business Applications*. Leeds, U.K.: Emerald Publishing, 2019, pp. 51–77.

[21] I. N. da Silva, D. H. Spatti, R. A. Flauzino, L. H. B. Liboni, and S. F. dos Reis Alves, "Self-organizing Kohonen networks," in *Artificial Neural Networks*. Cham, Switzerland: Springer, 2017, pp. 157–172.

[22] Y. Pan, L. Zhang, and Z. Li, "Mining event logs for knowledge discovery based on adaptive efficient fuzzy Kohonen clustering network," *Knowl.-Based Syst.*, vol. 209, Dec. 2020, Art. no. 106482, doi: 10.1016/j.knosys.2020.106482.

[23] C. C. Aggarwal, "Machine learning with shallow neural networks," in *Neural Networks and Deep Learning*. Cham, Switzerland: Springer, 2023, pp. 73–117.

[24] P. Wang, E. Fan, and P. Wang, "Comparative analysis of image classification algorithms based on traditional machine learning and deep learning," *Pattern Recognit. Lett.*, vol. 141, pp. 61–67, Jan. 2021, doi: 10.1016/j.patrec.2020.07.042.

[25] A. Oliinyk, S. Subbotin, V. Lovkin, O. Blagodariov, and T. Zaiko, "The system of criteria for feature informativeness estimation in pattern recognition," *Radio Electron., Comput. Sci., Control*, vol. 43, no. 4, pp. 85–96, Mar. 2018, doi: 10.15588/1607-3274-2017-4-10.

[26] Y. I. Daradkeh, V. Gorokhovatskyi, I. Tvoroshenko, S. Gadetska, and M. Al-Dhaifallah, "Methods of classification of images on the basis of the values of statistical distributions for the composition of structural description components," *IEEE Access*, vol. 9, pp. 92964–92973, 2021, doi: 10.1109/ACCESS.2021.3093457.

[27] Y. Sun, Z. Li, X. Li, and J. Zhang, "Classifier selection and ensemble model for multi-class imbalance learning in education grants prediction," *Appl. Artif. Intell.*, vol. 35, no. 4, pp. 290–303, Feb. 2021, doi: 10.1080/08839514.2021.1877481.

[28] V. Gorokhovatskyi, I. Tvoroshenko, and O. Yakovleva, "Transforming image descriptions as a set of descriptors to construct classification features," *Indonesian J. Elect. Eng. Comput. Sci.*, vol. 33, no. 1, pp. 113–125, Jan. 2024, doi: 10.11591/ijeecs.v33.i1.pp113-125.

[29] Z. Hu, Y. V. Bodyanskiy, and O. K. Tyshchenko, "A deep cascade neural network based on extended neo-fuzzy neurons and its adaptive learning algorithm," in *Proc. IEEE 1st Ukraine Conf. Electr. Comput. Eng. (UKRCON)*, May 2017, pp. 801–805, doi: 10.1109/UKRCON.2017.8100357.

[30] F. Malik and B. Baharudin, "Analysis of distance metrics in content-based image retrieval using statistical quantized histogram texture features in the DCT domain," *J. King Saud Univ.-Comput. Inf. Sci.*, vol. 25, no. 2, pp. 207–218, Jul. 2013, doi: 10.1016/j.jksuci.2012.11.004.

[31] W. Mo, X. Luo, Y. Zhong, and W. Jiang, "Image recognition using convolutional neural network combined with ensemble learning algorithm," in *Proc. J. Phys., Conf.*, Jun. 2019, vol. 1237, no. 2, Art. no. 022026, doi: 10.1088/1742-6596/1237/2/022026.

[32] Z.-H. Zhou, "Ensemble learning," in *Machine Learning*. Singapore: Springer, 2021, pp. 181–210.

[33] J. Barraza, P. Melin, F. Valdez, and C. I. Gonzalez, "Modeling of fuzzy systems based on the competitive neural network," *Appl. Sci.*, vol. 13, no. 24, p. 13091, Dec. 2023, doi: 10.3390/app132413091.

[34] J. Mawane, A. Naji, and M. Ramdani, "A cluster validity for optimal configuration of Kohonen maps in e-learning recommendation," *Indonesian J. Elect. Eng. Comput. Sci.*, vol. 26, no. 1, pp. 482–492, Jan. 2022, doi: 10.11591/ijeecs.v26.i1.pp482-492.

[35] Y. Huang, Z. Cheng, Q. Zhou, Y. Xiang, and R. Zhao, "Data mining algorithm for cloud network information based on artificial intelligence decision mechanism," *IEEE Access*, vol. 8, pp. 53394–53407, 2020, doi: 10.1109/ACCESS.2020.2981632.

[36] V. Gorokhovatskyi, O. Peredrii, I. Tvoroshenko, and T. Markov, "Distance matrix for a set of structural description components as a tool for image classifier creating," *Adv. Inf. Syst.*, vol. 7, no. 1, pp. 5–13, Mar. 2023, doi: 10.20998/2522-9052.2023.1.01.

[37] C. C. Aggarwal and C. K. Reddy, "Concepts of visual and interactive clustering," in *Data Clustering: Algorithms and Application*. Boca Raton, FL, USA: CRC Press, 2014, pp. 483–500.

[38] *OpenCV Open Source Computer Vision*. Accessed: Jan. 12, 2024. [Online]. Available: https://docs.opencv.org/master/ index.html

[39] *OpenCV Compare Images*. Accessed: Jan. 14, 2024. [Online]. Available: https://www.delftstack.com/howto/python/opencv-compare-images/

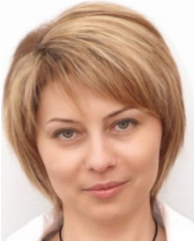
[40] *Images of Coins*. Accessed: Jan. 16, 2024. [Online]. Available: https://fama.ua/podarki/monety

VOLODYMYR GOROKHOVATSKYI received the M.Sc. degree in applied mathematics and engineering from Kharkiv National University of Radio Electronics (KNURE), the Ph.D. degree in management (technical systems), in 1984, and the Dr.Sc. degree in systems and means of artificial intelligence, in 2010.

He received an Internship with Dresden Technical University. He is currently a Professor with the Department of Informatics, KNURE. He has more than 200 scientific articles and seven monographs. His research interests include image and pattern recognition in computer vision systems, structural methods of image classification and recognition, multidimensional data analysis, and artificial intelligence.

IRYNA TVOROSHENKO received the Ph.D. degree in systems and means of artificial intelligence, in 2010.

She is currently an Associate Professor with the Department of Informatics, Kharkiv National University of Radio Electronics. She has published 173 scientific articles and educational and methodical articles include four study guides, seven monographs, 42 articles, 93 abstracts of reports, 14 lecture notes, and 13 methodological instructive regulations. She is fluent in modern programming languages and technologies, computer-aided mathematical modeling, and constantly expanding her range of scientific interests. Her research interests include image and pattern recognition in computer vision systems, structural methods of image classification and recognition, and fuzzy methods in artificial intelligence appliances.

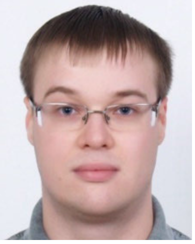


OLENA YAKOVLEVA received the Ph.D. degree in systems and means of artificial intelligence, in 2004.

In 2008, she obtained the academic title of an Associate Professor with the Department of Informatics. Since 2004, she has been teaching disciplines related to databases, big data, data visualization, development of information systems, artificial intelligence, and computer vision at Kharkiv University of Radio Electronics. Since 2017, she has been a Scientific Consultant in computer vision, machine learning, and intelligent systems at Sytoss Ltd. Since 2022, she has also been with Bratislava University of Economics and Management. She is the author of many publications on computer vision and intelligent systems. Her research interests include computer vision, machine learning, and intelligent systems.

MONIKA HUDÁKOVÁ received the Ph.D. degree in branch and cross-sectional economics from the Faculty of Operational Economics, Slovak University of Agriculture in Nitra, Slovak Republic, in 2000, and the Professor degree in economics and business management from Slovak University of Agriculture, in 2022.

Since 2022, she has been a University Professor with Bratislava University of Economics and Management and a Vice-Rector for Science and Research. She is currently a full-time Professor with the Department of Economics and Finance, where she constantly raises the visibility of issues in the field of economics, management, crisis management, and international management. She is also a full member of Czech Academy of Agricultural Sciences and Slovak Academy of Agricultural Sciences. She has received several awards for her long-standing international cooperation and development of science and research. Scientific research activity is documented by a rich, internationally accepted publication activity and participation in numerous research projects.



OLEKSII GOROKHOVATSKYI received the Ph.D. degree in systems and means of artificial intelligence from Kharkiv National University of Radio Electronics, in 2010.

He is currently an Associate Professor with the Department of Informatics and Computer Engineering, Simon Kuznets Kharkiv National University of Economics. His research interests include artificial intelligence, explainable artificial neural networks, computer vision, image processing, pattern and image classification, analysis, and recognition.