DOI: https://doi.org/10.15276/hait.05.2022.7
UDC 004.93

Comparative analysis of classifiers for face recognition on image fragments identified by the FaceNet neural network

Marina V. Polyakova[1)]
ORCID: https://orcid.org/0000-0001-7229-7657; marina_polyakova1@rambler.ru. Scopus Author ID: 57017879200
Dmitry Yu. Kozak[1)]
ORCID: https://orcid.org/0000-0002-3549-9222; dmkozakofficial@gmail.com
Natalia A. Huliaieva[1)]
ORCID: https://orcid.org/0000-0002-8827-6612; gulyaeva.nata72@gmail.com
[1)] Odessa National Polytechnic University, 1, Shevchenko Ave. Odessa, 65044, Ukraine

ABSTRACT

As a result of the analysis of the literature, the based methods of face recognition on fragments of color images were identified. These are flexible comparison in graphs, hidden Markov models, principal component analysis, and neural network methods. The analyzed methods of face recognition are mainly characterized by significant computational costs and low recognition performance. An exception is the neural network methods of face recognition, which, after completing the training, make it possible to obtain a high recognition performance at low computational costs. However, when changing the prototype images of faces, it often becomes necessary to redefine the network architecture and retrain the network. The specificity of neural network methods is also the complexity of selecting the network architecture and its training. Such papers are devoted to the use of neural networks only for extraction of feature vectors of face images. The classification of the obtained feature vectors is then performed by known methods, namely, thresholding, a linear support vector machine, nearest neighbors, random forest. It has been observed that the lighting conditions in which the images were obtained and the turning of the head affect the shape of the separating surface and can decrease the feature vector classification performance for face images. Therefore, to improve the classification performance, it was decided to use correlation for prototype matching, a non-linear support vector machine and logistic regression. The performed experiment showed that correlation for prototype matching of low-light face images is characterized by higher classification performance compared to the thresholding. Moreover, the use of the Pearson and Spearman correlation coefficients showed similar results, and when using the Kendall correlation coefficient, the worst classification performance was obtained compared to the Pearson and Spearman coefficients. The research of the classification performance of images of faces that differ in head turn using correlation for prototype matching, a non-linear support vector machine and logistic regression showed the following. Correlation for prototype matching is more appropriate to use with small amounts of data due to the high classification performance and low computational complexity, since a small amount of data does not require a significant number of comparisons. However, on large amounts of data, the non-linear support vector machine requires less computation and shows similar classification performance. Using the results of the experiment, the researcher can select classification methods for a specific set of face images, preliminarily representing them with feature vectors using the network FaceNet.

Keywords: Face recognition; FaceNet; convolutional neural network; correlation for prototype matching; support vector machine; logistic regression; deep learning

For citation*: Polyakova M. V., Kozak D. Yu., Huliaieva N. A. Comparative analysis of classifiers for face recognition on image fragments identified by the FaceNet neural network.** ***Herald of Advanced Information Technology**. 2022; Vol. 5 No. 2: 91–101. DOI: https://doi.org/10.15276/hait.05.2022.7**

INTRODUCTION

An increasing number of companies are beginning to use in security systems access control and management subsystems with face recognition from a video stream in real time [1, 2]. Thanks to such subsystems, it is possible to significantly increase the security of the enterprise and its employees [3].

In addition, face recognition is used in systems for processing frames of video streams from surveillance cameras on city streets and in public places, for example, to search for wanted persons. Face recognition is also used in Apple and Samsung smartphones to unlock the screen. A number of social networks use face recognition to automatically detect user's friends in photos.

Recently, the requirements for the effectiveness of systems that use face recognition are increasing, since the insufficient performance of face recognition leads to illegal entry of intruders into the territory of a protected facility, allows to unlock a smartphone and gain access to the user's personal data, commit theft and fraud in trade and banking, contribute to the emergence of dangerous situations in crowded places.

The face image processing in the listed applications involves the solution of two problems.

The first one is the face detection on the image. It is usually solved by scanning video stream frames with a window smaller than the frame size [4, 5]. In this case, computational difficulties arise due to the using of windows of different locations and sizes [6, 7]. In this paper, we consider the second problem of face image processing, i.e., face recognition on a fragment selected in the original image [8, 9].

To solve this problem, features of face images are extracted, and then the estimated feature vectors are classified [10]. Therefore, the performance of face recognition is affected by both the selecting of face image features, which is automatic or depending on the experience of the researcher, and the selecting of the classifier of the obtained feature vectors. In addition, it is known that the face recognition performance decreases for low-light images, as well as if the turning of the head or the angle of shooting is varied [11].

1. ANALYSIS OF RECENT RESEARCH AND PUBLICATIONS

To solve the problem of face recognition on a fragment selected in the original image, the flexible comparison on graphs [4, 10], hidden Markov models [12], principal component analysis [11, 12], neural network methods [8, 9] are used.

The paper devoted to neural network methods for face recognition on image fragments [5, 13]. A specificity of these methods is the automatic extraction of image features and the mapping of inputs to network outputs as a result of its training. To train the network, a selection of pre-prepared examples of image feature vectors and their corresponding outputs is used. After training, the network is capable of recognizing previously unused face images. The advantage of these methods is the high performance of face image classification with the correct settings of the network parameters. The disadvantages of neural network methods are the informal nature of the process of selecting the network architecture, the complexity of network training. When the network architecture or initial data varies, network retraining is required.

The use of a deep learning convolutional neural network involves the automatic extraction of features of face images in convolutional layers, followed by classification of the resulting feature vectors by fully connected layers [8, 13]. The disadvantage of this approach is that it assumes the ability of fully connected layers to generalize the obtained results to face images that are not included in the training set. In addition, the feature vectors obtained with convolutional layers usually have a large dimension (about a thousand). Then additional fully connected layers significantly increase the number of learnable network parameters. This requires more training samples and increases training time, which is already a problem for deep learning.

Therefore, in [9, 14], it was proposed to use the FaceNet deep learning convolutional neural network only for extracting features of face images. The coefficients of this network were adjusted in such a way that the distance between the feature vectors of images of similar faces was less than the distance between the feature vectors of images of different faces. This was achieved by optimizing the triplet loss function, which based on the distances from the feature vector of the analyzed image to the feature vectors of similar and different face images. A thresholding was used to classify the feature vectors of face images obtained using FaceNet.

The FaceNet network achieves high recognition performance for face image of people of different races, with different face shapes and different facial expressions. So, after training on the Labeled Faces in the Wild dataset [15], which contains approximately 3 million images, the percentage of correct recognition for the FaceNet network on the same dataset was 99.42 %, and on the YouTube Faces Database it was 95.12 % [9, 14]. In [5], to improve the face recognition performance it was proposed to apply a k-nearest neighbor classifier, a linear support vector machine (SVM), and a random forest to image feature vectors obtained using FaceNet. Compared to the thresholding, higher recognition performance was obtained on the Labeled Faces in the Wild image database [15]; especially for the first two classifiers [5].

2. FORMULATION OF THE PROBLEM

The results of face recognition with feature extraction by the FaceNet neural network presented in the literature differ depending on the classifier used. The selection of the classifier is determined by the shape of clusters in the multidimensional feature space and the presence of data outliers. So, illumination, contrast, turning of the head determines the shape of the separating surface during classification and effects on the face recognition performance.

In addition, the classification performance of image feature vectors depends on the reseached data set. In this paper images of the Face Place database [16] are researched, which differ in illumination and turning of the head. To improve the face recognition performance compared to a thresholding, it is proposed to apply correlation for prototype matching, nonlinear SVM and logistic regression to the feature vectors obtained using the FaceNet network.

Correlation for prototype matching shows the best results in comparison with other classifiers in the processing of deterministic signals in Gaussian white noise. Logistic regression assumes linear separability of image classes in the feature space. Nonlinear SVM allows one to construct a smooth separating surface between two classes of images. This surface has a more complex shape compared to a hyperplane. However, it is quite difficult to test these assumptions about the shape of clusters and the illumination for the researched set of face images in the multidimensional feature space obtained using the FaceNet network. In such cases, experimental research is carried out on the expediency of using the selected classifiers for face image recognition.

The aim of the paper is a comparative analysis of classifiers for an grounded selection when recognizing faces on fragments of color images are identified using feature vectors extracted by the FaceNet convolutional neural network.

3. FACE RECOGNITION ON FRAGMENTS OF COLOR IMAGES WITH FEATURE VECTORS EXTRACTED BY THE FACENET

The problem of face recognition on fragments of color images is formulated as follows. There is a database containing, for each unique identifier i, photographs I_{ij} with the image of the face of person $i \in N_I$, $j=1,\ldots, M$; where N_I is the set of identifiers of persons in the database under consideration, M is the number of photo for one identifier value containing the face of the same person. The number of unique face identifiers is limited. For the original color image I, it is necessary to determine the unique identifier i of the person present in the image, or, if the person is not in the database, send an appropriate message.

To recognize faces on fragments of color images the following steps were used: obtaining input images; face detection in the image; geometric transforms of the face image; identification and classification of this image; output of recognition result [9, 14, 17].

The input images are obtained by capturing frames of the webcam video stream in real time.

To localize the face in the image, the histogram of oriented gradients algorithm was used, which assumes that the shape of an object on the image can be described by the distribution of edge directions [18].

With the help of geometric transforms of the face image, namely, affine transforms of rotation and scaling that preserve parallel lines, the eyes were first centered, and then the mouth [9, 14], [17].

The identification of a person's face on a selected image fragment was performed using the FaceNet deep learning neural network. FaceNet belongs to the class of networks used to mapping from a two-dimensional input face image to a one-dimensional feature vector (Table 1). These feature vectors are extracted in such a way that for images of the same person they are located closer to each other in Euclidean space than for images of different people [9].

Table 1. **FaceNet network architecture used for face image feature extraction**

Layer type	Layer output size	Depth
Convolutional (7×7×3, 2)	112×112×64	1
Max pooling + normalization	56×56×64	0
Inception module (2)	56×56×192	2
Normalization + max pooling	28×28×192	0
Inception module (3a)	28×28×256	2
Inception module (3b)	28×28×320	2
Inception module (3c)	14×14×640	2
Inception module (4a)	14×14×640	2
Inception module (4b)	14×14×640	2
Inception module (4c)	14×14×640	2
Inception module (4d)	14×14×640	2
Inception module (4e)	7×7×1024	2
Inception module (5a)	7×7×1024	2
Inception module (5b)	7×7×1024	2
Average pooling	1×1×1024	0
Fully Connected	1×1×128	1
L_2 normalization	1×1×128	0

Source: **compiled by the [9]**

After the identification stage, 128-dimensional feature vectors were obtained for the input face image and prototype face images, for comparison of which a thresholding was used in [9, 14], [17]. A distance threshold was set in advance in order to make a decision that the face from the input image is the same as in the prototype face image. The lower the threshold, the greater the similarity with the prototype. The distances from the feature vector of the input image to the feature vectors of the prototypes were calculated. If the distance did not exceed the threshold, a conclusion was made about the similarity of the face in the input image with the prototype face.

The thresholding is fast and easy to implement, but the threshold value varies depending on the illumination of the recognized face in the image. The use of a single threshold value for different illumination conditions of face images leads to a decreasing of the recognition performance. Therefore, in this paper, to feature vectors of face images obtained using the FaceNet network, correlation for prototype matching, a nonlinear SVM, and logistic regression were used as classifiers.

Correlation for prototype matching

According to the correlation for prototype matching, the correlation coefficients between the feature vector of the classified image and the feature vectors of the prototype images are calculated [19]. The similarity with the prototype is determined by the highest value of the correlation coefficient. The theoretical base of this method is its optimality in detecting a deterministic signal in Gaussian white noise.

In this work, correlation coefficients of Pearson, Spearman, and Kendall were used and researched for correlation with the standard of face images.

The Pearson correlation coefficient characterizes the existence of a linear relationship between two quantities. Let two vectors of features identifying a face image be given as $x=(x_1, \ldots, x_n)$, $y=(y_1, \ldots, y_n)$, where n is the number of face image features. The Pearson correlation coefficient is calculated by the formula

$$r_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})\,(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}},$$

where $\bar{x}, \bar{y}$ are the sample means of x and y, $r_{xy} \in [-1, 1]$. If $|r_{xy}|=1$, then x, y are linearly dependent. If r_{xy} is close to 0, then x, y are linearly independent [19].

Spearman and Kendall correlation coefficients are measures of linear relationship between random variables. These coefficients are rank coefficients, that is, not numerical values are used to estimate the strength of the dependence, but the corresponding ranks. The coefficients are invariant with respect to any monotonic transform of the measurement scale. Spearman correlation coefficient for two feature vectors $x=(x_1, \ldots, x_n)$, $y=(y_1, \ldots, y_n)$, identifying face images is calculated by the formula

$$\rho = 1 - \frac{6}{n(n-1)(n+1)} \sum_{i=1}^{n}(R_i - S_i),$$

where R_i is the rank of element x_i in feature vector x; S_i is the rank of element y_i in the feature vector y.

The Kendall correlation coefficient is calculated as

$$\tau = 2T/(n(n-1)),$$

where $T = \sum_{i<j} \operatorname{sign}(x_j - x_i)\operatorname{sign}(y_j - y_i)$.

Coefficients $\rho \in [-1, 1]$ and $\tau \in [-1, 1]$. Equality $\rho=1$ or $\tau=1$ indicates a strict direct linear dependence, and $\rho=-1$ or $\tau=-1$ indicates inverse dependence.

Correlation for prototype matching differs from other classifiers its high classification performance. However when processing with big data, this classifier is characterized by significant computational costs, since it is necessary to compare the input feature vector with all prototype feature vectors each time. When processing a significant number of images, it is more expedient to use SVM.

Support Vector Machine

SVM was originally intended to classify objects into two classes. It allows you to choose the optimal location of the separating surface so that it is located at the maximum distance from the elements of each of the classes, i.e. in the middle of some strip that separates these elements. The separating surface is constructed as a result of learning the SVM on the training set of examples $\{x_k, y_k, k=1, \ldots, K\}$, where x_k is the feature vector of the kth example, $x_k \in R^n$, n is dimension of the feature space (in our case $n=128$), $y_k \in \{-1, 1\}$ is the corresponding output of the kth example, K is the number of the training set examples [20]. When learning the SVM, the Lagrange multipliers α_k, k=1, …, K; are determined by solving the dual problem of minimization of the functional

$$L(w) = \sum_{i=1}^{K} \alpha_i - \frac{1}{2}\sum_{i=1}^{K}\sum_{j=1}^{K} \alpha_i\,\alpha_j\,y_i y_j \varphi(x_i, x_j),$$

with constraints $\sum_{i=1}^{K} \alpha_i\,y_i = 0$, $0 \le \alpha_k \le C$, $k=1,\ldots, K$. C is a regularization parameter that can be selected based on the experience of the researcher or obtained together with Lagrange multipliers when solving the optimization problem; $\varphi(x, x_k)$ is the kernel function [20]. In this paper the radial basis function kernel $\varphi(x, x_k)=\exp(-\gamma\|x - x_k\|^2)$ is used. γ is the parameter of the radial basis function, $\|\cdot\|$ is the norm in the space L_2 [21]. After obtaining the values of the Lagrange multipliers α_k, $k=1,\ldots, K$; the separating surface $f(x)$ is defined by the formula $f(x) = \sum_{k=1}^{K} \mathrm{y_k}\, w_k \varphi(x, x_k)$.

The use of support vectors reduces the computational costs of training and classification. Learning the SVM leads to solving a quadratic optimization problem, which has a unique solution. This, as well as the obtaining of a separating strip of maximum width between classes, improves the classification performance. The disadvantages of the SVM are the complexity of selecting of the kernel function and the regularization parameter. In addition, it is essential the low robustness to noise.

The classification result is strongly influenced by outliers and noise in the original data.

Logistic regression

Logistic regression is used to build a linear classifier that describes the a posteriori probabilities that objects belong to classes [21]. The linear function $f(x)=w^{T}x+b$ is used as a separating surface, where $w=(w_1,\ldots, w_n)$ is the vector of weight coefficients, $x\in R^n$ is the feature vector of the classified object, b is the bias parameter. The parameters of the logistic regression are estimated from the training set $\{x_k, y_k, k=1, \ldots, K\}$ as a result of minimizing the loss function $E(w,b)$ in w and b:

$$E(w,b)=\frac{1}{K}\sum_{i=1}^{K}\ln(1+\exp(-y_i f(x_i))+C\sum_{j=1}^{n}w_j^2.$$

The first term of the $E(w,b)$ corresponds to the logarithm of the logistic loss function, which evaluates the approximation error on training set examples. The second term is regularizing. It provides the ability of the classifier to generalize, and $C>0$ is the regularization parameter [22].

Stochastic gradient descent was used to estimate the vector w of logistic regression weights and parameter b. The application of this method for optimizing the functional $E(w,b)$ implies that only one sample of the training set or some subsample of the training set is used to calculate the gradient of the loss function to calculate the approximation of the parameters w, b at the current iteration. This is difference of stochastic gradient descent from classical gradient descent, where the gradient of the cost function is calculated as the sum of the gradients from each sample of the training set [23].

The advantages of logistic regression are low computational costs when processing a large amount of data, the interpretability of regression parameters, and the ability to obtain the probabilities of classifying an object to different classes. However, if the separating surface has a complex shape, then the classification performance using logistic regression is reduced.

4. EXPERIMENTAL RESEARCH OF CLASSIFIERS FOR AN INFORMED SELECTION IN FACE RECOGNITION

For research, the Face Place image dataset which is open for use in research and publications was selected [16]. This dataset contains images of more than 200 people of different races with constant lighting, different facial expressions, real emotions and masking. Images are in jpeg format with a resolution of 250x250 72 dpi and 24-bit color.

The results of the classification of feature vectors for face images using the researched methods and the thresholding [9, 14], [17] were evaluated by calculating the probability of correct classification *Accuracy*, precision *Pr*, recall *Rc*, and F-measure *F* [4, 24], [25]:

$$F=\frac{2PrRc}{Pr+Rc},\qquad Pr=\frac{TP}{TP+FP},$$

$$Rc=\frac{TP}{FN+TP},\qquad Accuracy=\frac{TP+TN}{2},$$

where *TP* is the percentage of images from the class labeled “Accept” that are correctly assigned to the class labeled “Accept”; *FP* is the percentage of images from the class labeled “Reject” that are incorrectly assigned to the class labeled “Accept”; *FN* is the percentage of images from the class labeled “Accept” that are incorrectly assigned to the class labeled “Reject”; *TN* is the percentage of images from the class labeled “Reject” that are correctly assigned to the class labeled “Reject”.

First, the face classification performance was evaluated using correlation for prototype matching. 20 prototype face images were selected (Fig. 1) and 27 images to be recognized (Fig. 2). Images F1-F9 belonged to a person on the M1 prototype, images F10-F18 belonged to a person on the M2 prototype, images F19-F27 belonged to a person on the M3 and M4 prototypes (the same person is on the M3 and M4 prototypes), and other images had to be rejected. Each of the 27 images differs in illumination and head rotation.

Correlation for prototype matching was researched using Pearson, Spearman, and Kendall correlation coefficients. It is compared with a thresholding with a coefficient of 0.6 selected in [9, 14]. Preliminary research have shown that in order to match the input images to prototypes, it is sufficient to set thresholds of 0.95; 0.95; 0.85 for Pearson, Spearman and Kendall correlation coefficients respectively.

Analyzing the obtained results (Table 2) note that correlation for prototype matching showed the classification performance higher than the thresholding. The Pearson correlation coefficient was chosen for use in the further experiment. Its values for images with the same head rotation, but with different illumination, when compared with the prototype, differed less than the values of the Spearman and Kendall correlation coefficients.

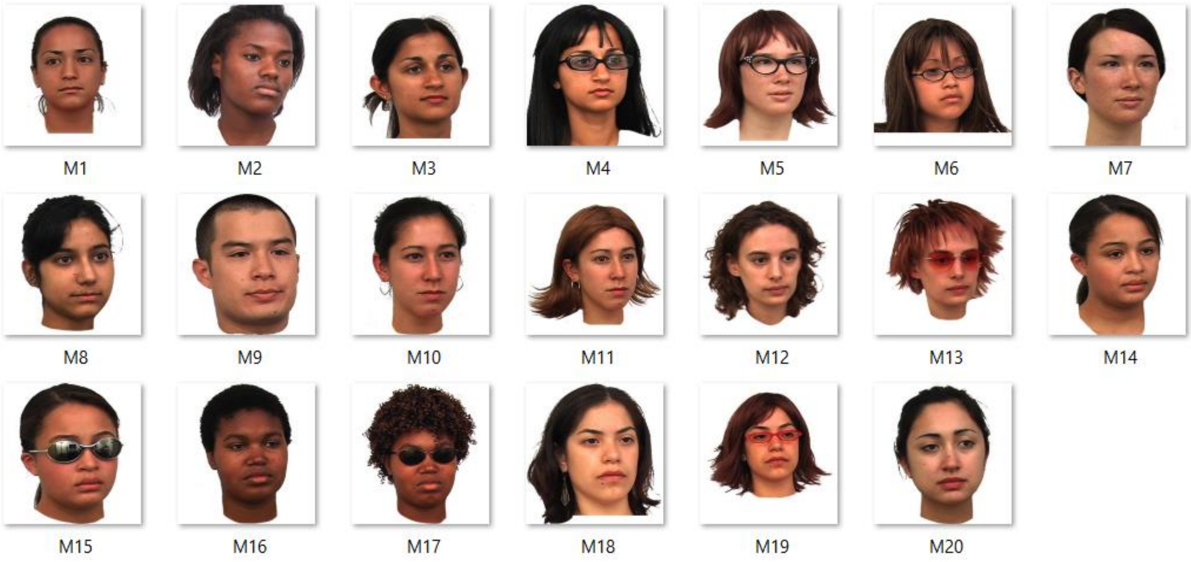

Fig. 1. **Prototype images to research of the classification performance using correlation for prototype matching**
***Source:* compiled by the [16]**

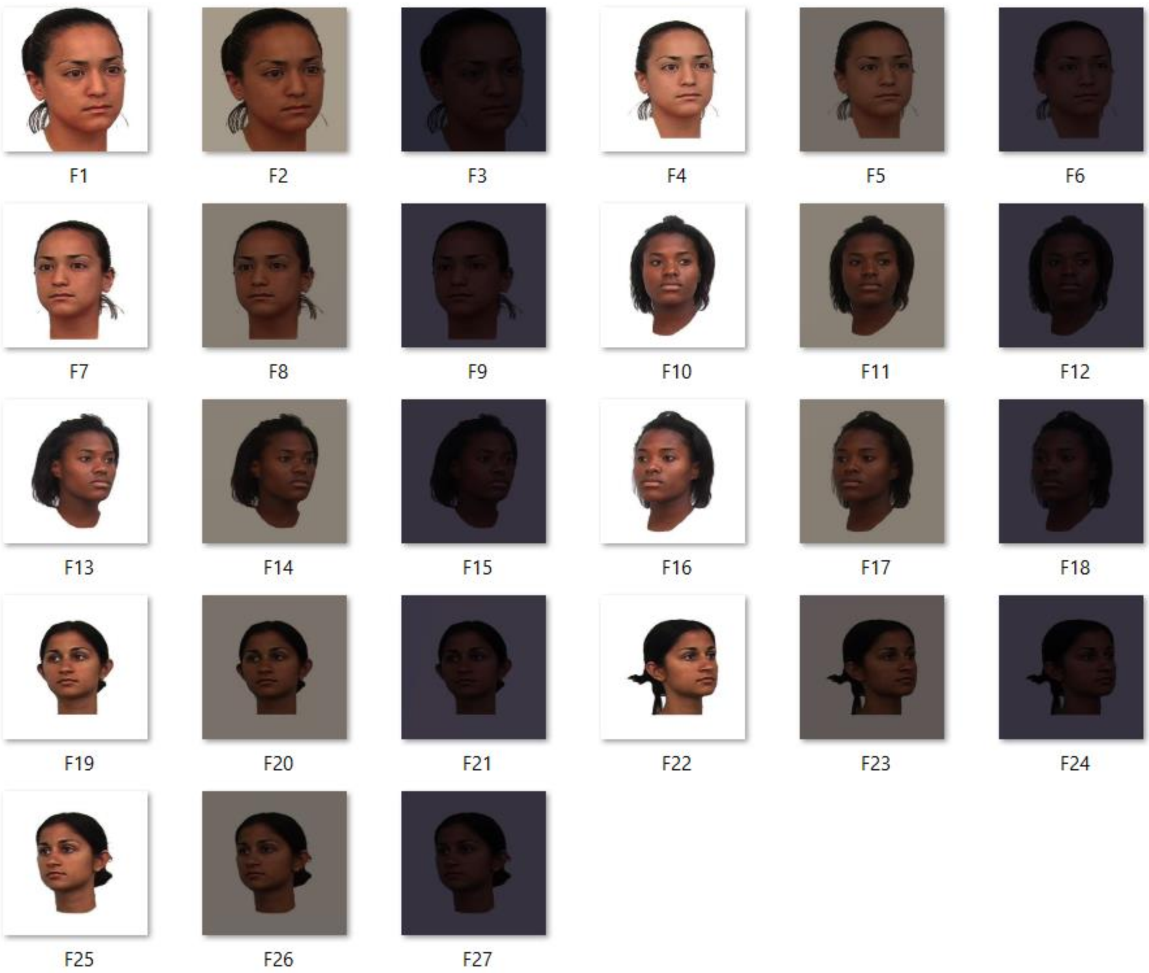

Fig. 2. **Input images that are to be recognized to research of the classification performance using correlation for prototype matching**
***Source:* compiled by the [16]**

ISSN 2663-0176 (Print)
ISSN 2663-7731 (Online)

Table 2. **The results of the classification of face images with correlation for prototype matching**

Classifier	*Accuracy*	*Pr*	*Rc*	*F*
Thresholding	0.920	0.456	1.000	0.626
Correlation for prototype matching (Pearson's coefficient)	1.000	1.000	1.000	1.000
Correlation for prototype matching (Spearman coefficient)	1.000	1.000	1.000	1.000
Correlation for prototype matching (Kendall coefficient)	0.989	1.000	0.833	0.908

Source: **compiled by the authors**

Also the face image classification performance is researched using correlation for prototype matching, non-linear SVM, and logistic regression. Ten prototype face images were selected (Fig. 3) and 45 images to be recognized (Fig. 4). Images of the same person differed in the head rotation. The input image U21-U25 corresponded to the prototype G1, images U26-U30 corresponded to the prototype G2, images U31-U35 corresponded to the prototype G3, images U36-U40 corresponded to the prototype G4, images U41-U45 corresponded to the prototype G5, images U1-U20 had to be rejected.

450 epochs were used to train the logistic regression with regularization in the space of L_2 functions, and 1000 epochs were used to train the SVM.

Analyzing the obtained results (Table 3) note that correlation for prototype matching showed the classification accuracy 3.5 % higher than the thresholding, precision 6.5 % higher than the thresholding, F-measure 3 % higher than the thresholding. In recall similar results were obtained. The non-linear SVM showed the precision 6.5% higher than the thresholding, classification accuracy and F-measure up to 1.5 % higher than the thresholding. However, the non-linear SVM is characterized by the recall 4.5 % lower than the thresholding.

It is known that as a result of classification to achieve an increasing both precision and recall simultaneously is in principle impossible. An increase in recall, which is typical for more “optimistic” classifiers, leads to an increase in *FP*, the number of false positive recognitions (the number of images from the class labeled “Reject” incorrectly assigned to the class labeled “Accept”) and a decrease in precision. Correlation for prototype matching, non-linear SVM, and logistic regression used in this paper are more “pessimistic” compared to the thresholding. It means that their use leads to an increase in *FN* which is the number of false negative recognitions (the number of images from the class with the label “Accept” incorrectly assigned to the class with the label “Reject”) and a decrease in recall.

Fig.3. **Prototype images to research of the classification performance using correlation for prototype matching, non-linear SVM, and logistic regression**

Source: **compiled by the [16]**

Table 3. **The results of face image classification with correlation for prototype matching, support vector machine, and logistic regression**

Classifier	*Accuracy*	*Pr*	*Rc*	*F*
Threshold	0.964	0.940	1,000	0.969
Logistic regression	0.822	1.000	0.680	0.810
Support Vector Machine	0.978	1.000	0.961	0.980
Comparison with the standard (Pearson's coefficient)	0.998	1.000	0.996	0.998

Source: **compiled by the authors**

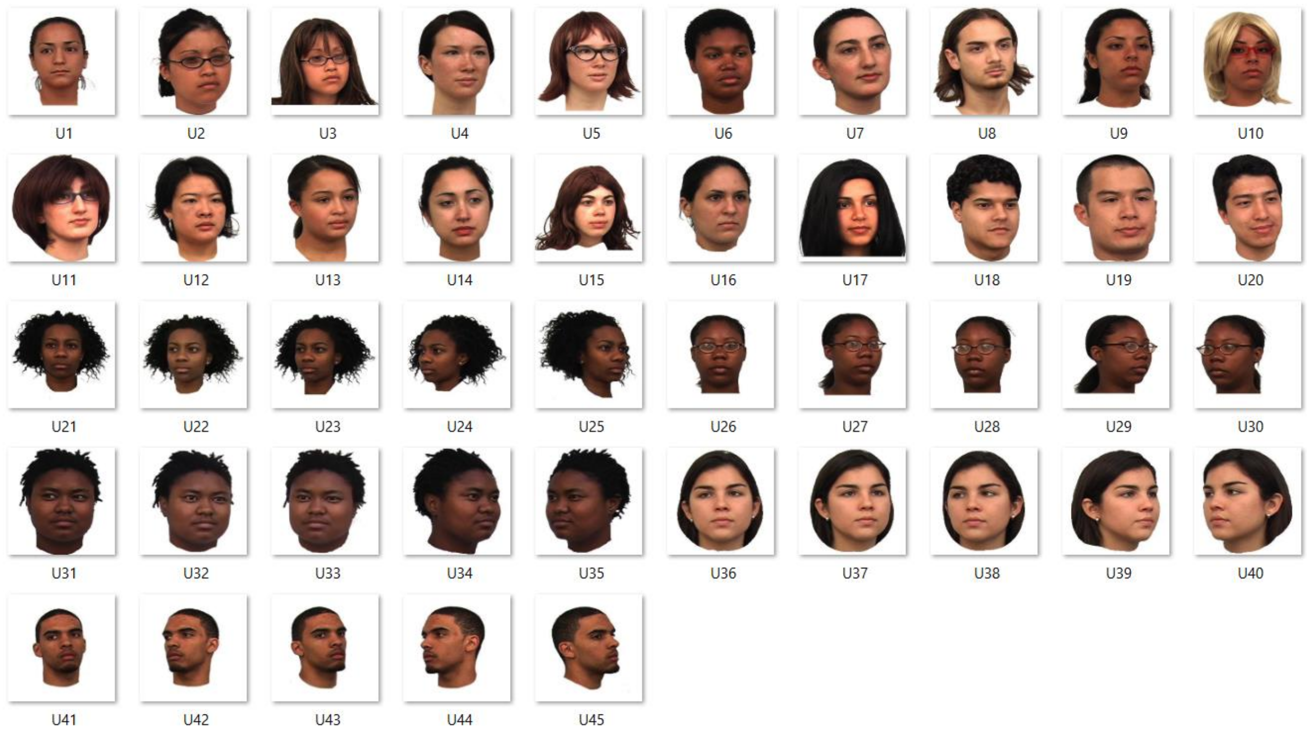

Fig.4. **Input images that are to be recognized to research of the classification performance using correlation for prototype matching, non-linear SVM, and logistic regression**
Source: **compiled by the [16]**

CONCLUSIONS

As a result of the analysis of the literature, a number of papers are devoted to the use of neural networks only for extraction of feature vectors of face images. The classification of the obtained feature vectors is then performed by known methods, namely, thresholding, a linear support vector machine, nearest neighbors, random forest. It has been observed that the lighting conditions in which the images were obtaned and the turning of the head affect the shape of the separating surface and can decrease the feature vector classification performance for face images. Therefore, to improve the classification performance, it was decided to use correlation for prototype matching, a non-linear support vector machine and logistic regression.

The performed experiment showed that correlation for prototype matching of low-light face images is characterized by higher classification performance compared to the thresholding. Moreover, the use of the Pearson and Spearman correlation coefficients showed similar results, and when using the Kendall correlation coefficient, the worst classification performance was obtained compared to the Pearson and Spearman coefficients. The research of the classification performance of images of faces that differ in turning of the head showed the following. Correlation for prototype matching is more appropriate to use with small amounts of data due to the high classification performance and low computational complexity, since a small amount of data does not require a significant number of comparisons. However, on large amounts of data, the non-linear support vector machine requires less computation and shows similar classification performance. Using the results of the experiment, the researcher can select classification methods for a specific set of face images, preliminarily representing them with feature vectors using the network FaceNet.

Further research can be devoted to determining the relevancy of features obtained as a result of the application of the network FaceNet. It is expedient to use the calculated values of the feature relevancy rate to reduce the dimension of the feature space.

REFERENCES

1. Manjula, V. S. & Baboo, L. D. “Face detection, identification and tracking by PRDIT algorithm using image database for Crime investigation”. *Int. J Comput. Appl.* 2012; Vol. 38(10): 40–46. DOI: https://doi.org/10.5120/4741-6649.

2. Lander, K., Bruce, V. & Bindemann, M. “Use-inspired basic research on individual differences in face identification: implications for criminal investigation and security”. *Cognit. Res., Princ. Implications*. 2018; Vol. 3(1): 1–13. DOI: https://doi.org/0.1186/s41235-018-0115-6.

3. Li, L., Mu, X., Li, S. & Haipeng, P. A review of face recognition technology. *IEEE Access*. 2020; Vol. 8: 139110–139120. DOI: https://doi.org/10.1109/ACCESS.2020.3011028.

4. Levchuk, S. A., & Yakimenko, A. A. "Study of the characteristics of face recognition algorithms" (in Russian). *Collection of scientific papers of NSTU*. 2018; Vol. 3–4 (93): 40–58. DOI: https://doi.org/10.17212/2307-6879-2018-3-4-40-58.

5. Sanchez-Moreno, A. S., Olivares-Mercado, J., Hernandez-Suarez, A., Toscano-Medina, K., Sanchez-Perez, G. & Benitez-Garcia, G. “Efficient face recognition system for operating in unconstrained environments”. J. Imaging. 2021; Vol. 7(9): 161–172. DOI: https://doi.org/10.3390/jimaging7090161.

6. Ruvinskaya, V. M. & Timkov, Y. Y. “Deep learning technology for videoframe processing in face segmentation on mobile devices”. *Herald of Advanced Information Technology.* 2021; Vol.4. No.2: 185–194. DOI: https://doi.org/10.15276/hait.02.2021.7.

7. Sheremet, O. I., Korobov, O. Ye., Sadovoi, O. V. & Sokhina, Yu. V. “Intelligent system based on a convolutional neural network for identifying people without breathing masks”. *Applied Aspects of Information Technology*. 2020; Vol. 3. No. 3: 133–134. DOI: https://doi.org/10.15276/aait.03.2020.2.

8. Petrosiuk, D. V., Arsirii, O. O., Babilunha, O. Ju. & Nikolenko, A. O. “Deep learning technology of convolutional neural networks for facial expression recognition”. *Applied Aspects of Information Technology*. 2021; Vol.4. No.2: 192–201. DOI: https://doi.org/10.15276/aait.02.2021.6.

9. Schroff, F., Kalenichenko, D. & Philbin, J. “FaceNet: A unified embedding for face recognition and clustering”. *In: Proc. of the IEEE Computer Society Conference on Computer Vision and Pattern*. 2015. p. 815–823. DOI: https://doi.org/10.1109/CVPR.2015.7298682.

10. Zahorodnia, D., Pigovsky, Yu., Bykovyy, P., Krylov, V., Paliy, I. & Dorotvor, I. “Structural statistic method identifying facial images by contour characteristic points”. *In: Proc. of the 8th IEEE Int. Conf. on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications*. 2015. p. 293–297. DOI: https://doi.org/10.1109/CVPR.2015.7298682.

11. Singh, S. & Graceline J. S. “Face recognition system”. *IJERT*. 2019; Vol. 8 (5): 263–266.

12. Samonenko, I. Yu. & Volchenkov, M. P. “About automatic face recognition” (in Russian). *Intelligent systems. Theory and Applications*. 2005; Vol. 9: 153–157.

13. He, K., Zhang, X., Ren, S. & Sun, J. “Deep residual learning for image recognition”. *In: Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVRP).* 2016. p. 770–778. DOI: https://doi.org/10.1109/CVPR.2016.90.

14. ”Machine learning is fun! Part 4: Modern face recognition with deep learning”, 2020. – Available from: https://medium.com/@ageitgey/machine-learning-is-fun-part-4-modern-face-recognition-with-deep-learning-c3cffc121d78. – [Accessed Dec 2020].

15. ”Labeled faces in the wild home", 2020. – Available from:. – [Accessed Dec 2020].

16. “OpenFace”. 2020. – Available from: https://cmusatyalab.github.io/openface. – [Accessed Dec 2020].

17. Kozak, D. Yu. & Polyakova, M. V. “Face recognition on low contrast images with the convolution neural network FaceNet”. *In: Proceedings of the 1st International Scientific and Practical Conf. “Modern research in world science”*. 2022. p. 435–442. – Available from: https://sci-conf.com.ua/i-mezhdunarodnaya-nauchno-prakticheskaya-konferentsiya-modern-research-in-world-science-17-19-aprelya-2022-goda-lvov-ukraina- archive/. – [Accessed Dec 2020].

18. Cerna, L. R., Camara-Chavez, G. & Menotti, D. “Face detection: histogram of oriented gradients and bag of feature method”. *In: Proceedings of the 2013 International Conference on Image Processing, Computer Vision, and Pattern Recognition*. 2013. p. 1–5.

19. Gonzalez, R. C. & Woods, R. E. “Digital image processing (3rd Edition)”. *Publ. Prentice Hall.* New York: USA. 2008.

20. Haikin, S. “Neural networks: a complete course (2nd Edition)” (in Russian). *Publ. Williams*. Moscow: Russian Federation. 2006.

21. Muller, A. & Guido, S. “An introduction to machine learning with Python. A guide for data scientists” (in Russian). *Publ. Alpha-book LLC.* St.-Petersburg: Russian Federation. 2017.

22. Ak, M. F. "A comparative analysis of breast cancer detection and diagnosis using data visualization and machine learning applications". *Healthcare*. 2020; Vol. 8 (2): 111–134. DOI: https://doi.org/10.3390/healthcare8020111.

23. Kashirina, I. L. & Demchenko, M. V. "Research and comparative analysis of optimization methods used in training neural networks" (in Russian). *Bulletin of the Voronezh State University. Series: System Analysis and Information Technology*. 2018; No 4: 123–132. DOI: https://doi.org/10.17308/sait.2018.4/1262.

24. Polyakova, M. V. & Krylov, V. N. "Data normalization methods to improve the quality of classification in the breast cancer diagnostic system". *Applied Aspects of Information Technology*. 2022; Vol. 5. No. 1: 55–63. DOI: https://doi.org/10.15276/aait.04.2022.5.

25. Polyakova, M. V. & Nesteryuk, A. G. "Improvement of the color text image binarization method using the minimum-distance classifier". *Applied Aspects of Information Technology*. 2021; Vol. 4. No. 1: 57–70. DOI: https://doi.org/10.15276/aait.01.2021.5.

Conflicts of Interest:the authors declare no conflict of interest

Received 14.01.2021
Received after revision 12.03.2021
Accepted 17.03.2021

DOI: https://doi.org/10.15276/hait.05.2022.7
УДК 004.93

Порівняльний аналіз класифікаторів для розпізнавання облич на фрагментах зображень, які ідентифіковано нейронною мережею FaceNet

Марина Вячеславівна Полякова[1)]
ORCID: https://orcid.org/0000-0001-7229-7657; marina_polyakova1@rambler.ru. Scopus Author ID: 57017879200
Дмитро Юрійович Козак[1)]
ORCID: https://orcid.org/0000-0002-3549-9222; dmkozakofficial@gmail.com
Наталя Анатоліївна Гуляєва[1)]
ORCID: https://orcid.org/0000-0002-8827-6612; gulyaeva.nata72@gmail.com
[1)] Одеський національний політехнічний університет, пр. Шевченка, 1. Одеса, 65044, Україна

АНОТАЦІЯ

В результаті аналізу літератури були виділені основні методи розпізнавання облич на фрагментах кольорових зображень: гнучке порівняння на графах, приховані марківські моделі, аналіз головних компонентів, нейромережеві методи. Проаналізовані методи розпізнавання облич, відомі з літератури, в основному характеризуються значними обчислювальними витратами та невисокою якістю розпізнавання. Винятком є нейромережеві методи розпізнавання облич, які після завершення навчання дозволяють отримати високу якість розпізнавання при малих обчислювальних витратах. Однак при зміні еталонних зображень облич часто виникає необхідність довизначення архітектури мережі та перенавчання мережі. Особливостями нейромережевих методів є складність вибору архітектури мережі та її навчання. Ряд робіт присвячено використанню нейронних мереж лише для побудови векторів ознак зображень облич. Класифікація отриманих векторів ознак виконується відомими методами: порівнянням з порогом, лінійною машиною опорних векторів, найближчих сусідів, випадковим лісом. Було помічено, що умови освітлення, в яких отримані зображення, і поворот голови впливають на форму поверхні, що розділяє, і можуть погіршити якість класифікації векторів ознак для зображень облич. Тому для підвищення якості класифікації вирішено використовувати кореляційне зіставлення з еталоном, нелінійну машину опорних векторів і логістичну регресію. Проведений експеримент показав, що кореляційне зіставлення з еталоном в умовах поганого освітлення зображень осіб відрізняється вищими значеннями показників якості класифікації порівняно з пороговим класифікатором. Причому застосування коефіцієнтів кореляції Пірсона та Спірмена показало подібні результати, а при використанні коефіцієнта кореляції Кенделла було отримано гірші значення показників якості класифікації порівняно з коефіцієнтами Пірсона та Спірмена. Дослідження якості класифікації зображень облич, що відрізняються поворотом голови, із застосуванням кореляційного зіставлення з еталоном, нелінійної машини опорних векторів та логістичної регресії показало наступне. Кореляційне зіставлення з еталоном більш доцільно використовувати при малих обсягах даних завдяки високій якості класифікації та невеликій обчислювальній складності, оскільки малий обсяг даних вимагає великої кількості

порівнянь. Однак на великих обсягах даних нелінійна машина опорних векторів вимагає меншої кількості обчислень і показує схожу якість класифікації. Використовуючи результати проведеного експерименту, дослідник може вибрати методи класифікації для конкретного набору зображень облич, попередньо представивши їх векторами ознак за допомогою мережі FaceNet.

Ключові слова: розпізнавання облич; FaceNet; згорткова нейронна мережа; кореляційне зіставлення з еталоном; машина опорних векторів; логістична регресія; глибоке навчання

ABOUT THE AUTHORS

Marina V. Polyakova - D.Sc (Eng), Associated Professor, Professor of Department of Applied Mathematics and Information Technologies. Odessa National Polytecnic University, 1, Shevchenko Ave. Odessa, 65044, Ukraine
ORCID: https://orcid.org/0000-0001-7229-7657; marina_polyakova1@rambler.ru. Scopus Author ID: 57017879200
***Research field*:** intelligent data analysis; machine learning; digital image processing

Марина Вячеславівна Полякова - доктор технічних наук, доцент, професор кафедри Прикладної математики та інформаційних технологій. Одеський національний політехнічний ун-т, пр. Шевченка, 1. Одеса, 65044, Україна

Dmitry Yu. Kozak - master, Department of Applied Mathematics and Information Technologies.
Odessa National Polytechnic University, 1, Shevchenko Ave. Odessa, 65044, Ukraine
ORCID: https://orcid.org/0000-0002-3549-9222; dmkozakofficial@gmail.com.
***Research field*:** Intelligent data analysis; machine learning; digital image processing

Дмитро Юрійович Козак - магістр кафедри Прикладної математики та інформаційних технологій. Одеський національний політехнічний ун-т, пр. Шевченка, 1. Одеса, 65044, Україна

Natalia A. Huliaieva - senior lector, Department of Applied Mathematics and Information Technologies. Odessa National Polytecnic University, 1, Shevchenko Ave. Odessa, 65044, Ukraine.
ORCID: https://orcid.org/0000-0001-2345-6789; gulyaeva.nata72@gmail.com.
***Research field*:** Intelligent data analysis; mathematical economics

Наталя Анатоліївна Гуляєва - старший викладач кафедри Прикладної математики та інформаційних технологій. Одеський національний політехнічний ун-т, пр. Шевченка, 1. Одеса, 65044, Україна