K. YUVCHENKO, V. YESILEVSKYI, O. SEREDA

HUMAN EMOTION RECOGNITION SYSTEM USING DEEP LEARNING ALGORITHMS

The **subject** of research in the article is the software implementation of a neural image classifier. The work examines emotions as a special type of mental processes that express a person's experience of his attitude to the surrounding world and himself. They can be expressed in different ways: facial expressions, posture, motor reactions, voice. However, the human face has the greatest expressiveness. Technologies for recognizing companies to improve customer service use human emotions make decisions about interviewing candidates and optimize the emotional impact of advertising. Therefore, the purpose of the work is to find and optimize the most satisfactory in terms of accuracy algorithm for classifying human emotions based on facial images. The following tasks are **solved**: review and analysis of the current state of the problem of "recognition of emotions"; consideration of classification methods; choosing the best method for the given task; development of a software implementation for the classification of emotions; conducting an analysis of the work of the classifier, formulating conclusions about the work performed, based on the received data. An image classification **method** based on a densely connected convolutional neural network is also used. **Results**: the results of this work showed that the method of image classification, based on a densely connected convolutional neural network, is well suited for solving the problems of emotion recognition, because it has a fairly high accuracy. The quality of the classifier was evaluated according to the following metrics: accuracy; confusion matrix; precision, recall, f1-score; ROC curve and AUC values. The accuracy value is relatively high – 63%, provided that the data set has unbalanced classes. AUC is also high at 89%. **Conclusions**. It can be concluded that the obtained model with weights has high indicators of recognition of human emotions, and can be successfully used for its purpose in the future.

Keywords: object detection; classification of objects; supervised learning; recognition of emotions.

Introduction

Emotion is a special type of mental processes that expresses a person's feelings and attitude to the world around him or her. Emotions can be transmitted in different ways: facial expressions, posture, motor reactions, voice. However, the most expressive is the human face. Technologies for recognizing companies to improve customer service use human emotions make decisions about interviewing candidates and optimize the emotional impact of advertising.

Using facial emotion recognition, businesses can process images and videos in real time to monitor video streams or automate video analytics, thereby reducing costs and making the lives of their users better.

Recognizing human emotions in an image with complex background structure is the first and main challenge in solving the emotion recognition task.

The task of emotion recognition is very relevant today in various fields, including medicine (helps people with autism to recognize the emotions of others and establish normal communication), sociology, robotics, marketing, psychology, etc.

Face recognition is still a challenging problem due to the variety of head poses and background settings.

People always try to hide or even falsify their true facial expressions. The huge challenges in facial expression recognition are finding and understanding facial expressions. Over the past few years, science has made sufficient progress, and tasks such as landmark detection, object modeling and much more can be performed using computer vision.

However, researchers usually pretended that their systems should work in a controlled environment. In other words, the face and background should not be difficult to identify. Meanwhile, other scientists are recognizing facial expressions in the wild using deep neural networks so that the machine can react like a human to various factors without limitations, such as dynamic lighting, head position, etc.

In the nineteenth century, one of the most important works on the analysis of facial emotions, which is directly related to the modern field of automatic facial expression recognition, was the work of Charles Darwin [1]. In 1872, the scientist wrote a dissertation on this problem. This laid the foundation for general provisions for determining the emotions of humans and animals. Darwin grouped the names of emotions. The classification looks like this:

– bad mood, anxiety, grief, despair and despair;
– joy, elation, love, tender feelings and devotion;

– reflection, meditation, irritability and gloom;

– hatred and anger;

– contempt, scorn, disgust, guilt and pride;

– surprise, amazement, fear and terror;

– attention to oneself, shame, bashfulness and modesty.

Darwin also classified the facial changes that occur for each of the above classes of expressions. For example, "clenching of the muscles around the eyes in case of sadness", "tight closing of the mouth during contemplation", etc.

Another important milestone in the study of facial expressions and human emotions was the work of Paul Ekman and his colleagues from the 1970s. Their work had a huge impact on the development of modern automatic facial expression recognizers.

Suwa et al conducted the earliest study of automatic facial expression recognition in 1978. As a result, a model was created to learn facial expressions from a sequence of images using 20 tracking arguments. Research has been conducted since the late 1980s and early 1990s, when mobile computing power of the economy became available. This helped to develop face detection and tracking algorithms in the early 1990s. At the same time, research on human–computer interaction (HCI) and affective computing (AC) began [2].

The aim of this work is to find and optimize the most satisfactory in terms of accuracy algorithm for classifying human emotions from face images.

Analysis of recent research and publications

FaceReader

Many researchers have turned to using automated facial expression analysis software to more objectively assess emotions. FaceReader is used worldwide in over 1000 universities, research institutes and companies in most industries. The consumer and psychological benefits of FaceReader and its usability are being researched.

FaceReader software is fast, flexible, objective, accurate and easy to use. It immediately analyzes your data (live, video or still images), saving valuable time. The ability to record audio and video allows you to hear what people say, for example, during human-computer interaction or while viewing stimuli. FaceReader uses the following emotion categories: joy, surprise, sadness, anger, disgust, contempt, fear. Ekman [3] described these emotional categories as basic or universal emotions. Obviously, facial expressions vary in intensity and are

often a combination of several emotions. In addition, there are quite a few interpersonal varieties.

FaceReader works in three stages:

1) Face detection. FaceReader uses the popular Viola-Jones algorithm [4] to detect the presence of a face in an image.

2) Accurate face modeling using an algorithmic approach based on the Active Appearance method [5] described by Cootes and Taylor [6]. At this stage, the model is trained using a database of annotated images. This method describes over 500 key points on the face and the facial texture, e.g.:

a) points that cover the face (the part of the face analyzed by FaceReader);

b) points on the face that are easy to recognize (lips, eyebrows, nose and eyes).

3) The actual classification of facial expressions is performed by training an artificial neural network [7]. More than 10000 manually annotated images were used as training material.

Texture is also important as it gives additional information about the state of the face. The key points only describe the general position and shape of the face, but do not provide any information, for example, about the presence of wrinkles and the shape of eyebrows (such features, although they may seem less important, play a significant role in the classification of facial expressions). FaceReader works in three stages:

1) Face detection. FaceReader uses the popular Viola–Jones algorithm [4] to detect the presence of a face in an image.

2) Accurate face modeling using an algorithmic approach based on the Active Appearance method [5] described by Cootes and Taylor [6]. At this stage, the model is trained using a database of annotated images. This method describes over 500 key points on the face and the facial texture, e.g.:

a) points that cover the face (the part of the face analyzed by FaceReader);

b) points on the face that are easy to recognize (lips, eyebrows, nose and eyes).

3) The actual classification of facial expressions is performed by training an artificial neural network [7]. More than 10,000 manually annotated images were used as training material.

Texture is also important as it gives additional information about the state of the face. The key points only describe the general position and shape of the face, but do not provide any information; for example, about the presence of wrinkles and the shape of

eyebrows (such features, although they may seem less important, play a significant role in the classification of facial expressions).

EmoDetect

This is the name of the program that helps to determine the psycho-emotional state of a person by a sample of images by identifying six emotions: happiness, surprise, sadness, anger, fear, disgust. EmoDetect program contains the principles of coding facial expressions FACS Action Units, developed by Paul Ekman.

During the study, a person is shown various stimuli on a monitor, and a camera is located in front of the test participant, which shoots video.

The system processes the video according to the following algorithm: the image of a person whose facial expression is to be classified is captured. Then the face is extracted from the image using the Jones method to determine the elements. Next, the algorithms extract the corresponding elements from the cropped image of the face. Trained classifiers match these features with the corresponding emotions.

The code is implemented using OpenCV, an open source computer vision library.

The final stage is the fixation of emotions with reference to the time of demonstration and visualization of the results in the form of graphs and tables.

Viola–Jones method

The Viola–Jones method is used to detect objects in images. This method can recognize different classes of images, but the main task in its creation is face recognition. The algorithm of this method is able to detect objects very reliably and quickly enough to work in real time.

The Viola–Jones detector is a cascade of classifiers.

It has the following features:

– Haar features are used;

– images are presented in a holistic form;

– boosting is used.

Haar feature is a rectangular primitive. From the very beginning, the authors proposed the main four primitives, but the OpenCV library uses additional primitives that improve the quality of recognition with additional tests from non-standard points of view.

Each feature is paired with a threshold, and the decision of the feature is determined by comparing its value with the threshold. The Viola–Jones method implements a fast way to calculate attribute values that uses the integral representation of the image. The integral representation of an image is a matrix that has the same dimensions as the original image. The value of its elements is defined as the sum of ridge intensity peaks located on the left and top. In such a matrix, the sum of pixel values in an arbitrary rectangle is calculated for a constant time.

The detection process is performed by sliding the detection window over the entire image. A cascade solution is calculated for each window. In the case of a positive answer, it is assumed that the desired object is located inside the window. After one pass of the image is completed, the window size is increased. The window size is increased until the predefined size is reached. A smaller percentage of magnification improves the detection rate but increases the overall processing time [8].

Materials and methods

Nowadays, with the spread of artificial intelligence, such fields as machine learning and its branches, deep learning and neural networks have gained immense popularity. Learning requires software and tools such as classifiers that feed huge amounts of data, analyze it and perform useful functions. The goal of the classification process is to categorize all pixels of a digital image into one of several classes. Usually multispectral data is used to perform the classification. Indeed, the spectral pattern present in the data for each pixel is used as a numerical basis for categorization.

Image classification is probably the most important part of digital image analysis. Classification between objects is a complex process, so it is an important task for the field of computer vision. The essence of image classification is that an image is labeled and put into a number of predefined classes. There are potentially classes to which an image can belong. Manually checking and classifying images can be a tedious task, especially if there are a large number of them. Therefore, it is very useful if it is possible to automate this whole process using computer vision.

The algorithm for performing image classification consists of four stages.

– Image preprocessing, which aims to improve the image data (features). To do this, unwanted aspects are removed and some important features of the image are enhanced so that computer vision models can use this enhanced data to work. The image preprocessing steps include reading the image, resizing the image, and

enhancing the data (gray-scaling, reflection, Gaussian blur, histogram, alignment, and rotation).

– Object detection. This stage localizes the object, that is, the image is segmented and the position of the object is determined.

– Feature extraction and training. This is a crucial stage where statistical or deep learning techniques are used to identify the most interesting patterns in the image, features that may be unique to a particular class and that will later help the model to discriminate between different classes. (The process where a model learns features from a dataset is called model training.)

– Object classification. In this step, the detected objects are grouped using an appropriate classification technique that compares the image patterns with the target patterns.

In recent years, the application of machine learning techniques in everyday life has become commonplace. Many modern websites and devices use machine learning algorithms, ranging from automatic recommendations for watching movies, ordering food or buying groceries, to personalized online radio broadcasts and recognizing human emotions in an image.

Machine learning can be broadly defined as computational techniques that use experience to improve performance or to make accurate predictions. Since the success of a learning algorithm depends on the data used, machine learning is closely related to data analysis and statistics. In general, machine learning methods are data mining techniques that combine fundamental concepts of computer science with statistics, probability, and optimization [9]. It consists of the development of efficient and accurate prediction algorithms, and if more time is devoted and more input samples of documents are specified for training, then the predictions of such models will be better. As in other areas of computer science, some critical indicators of the quality of algorithms are their temporal and spatial complexity.

Neural networks are widely used to classify emotions from facial expressions. They employ both supervised (training with a teacher) and unsupervised (training without a teacher) neural architectures for emotion classification.

In this paper, a supervised learning algorithm is used to perform the image classification task.

Supervised (learning with a teacher) classification is based on the idea that the user can select samples of pixels in the image that are representative of certain classes, and then the software processes the image using training sites to classify all other pixels in the image.

The training sites (also known as test sets or input classes) are selected based on the user's knowledge. The user also sets boundaries on how similar other pixels must be to be grouped together. These bounds are often set based on the spectral characteristics of the training region. The user also defines the number of classes to which the image belongs.

The main goal of emotion recognition is to identify human emotional states based on face images. Deep convolutional neural network (DCNN) architectures have been successfully used for image classification [10, 11]. As the network deepens, the input information (or back propagation gradient) passes through many layers and may disappear by the time it reaches the end of the network. Several DCNN architectures have been proposed to address this problem. Highway Networks [12], ResNet (Residual Neural Network) [13] and DenseNet 169 traverse the signal from one layer to another by identifying intermediate connections between layers. FractalNets [14] repeatedly combines multiple parallel sequences with different numbers of convolutional units to obtain a large nominal depth while maintaining many short paths in the network. Although these methods differ in terms of network architecture and learning strategy, they all share a key characteristic: creating short paths from earlier layers to later layers. The DenseNet 169 model has slightly higher overall accuracy than the ResNet and FractalNets models. It will be used in this paper.

In 2017, DenseNets [15] was proposed at the CVPR 2017 conference. They started by trying to build a deeper network based on the idea that if a convolutional network has shorter connections between its layers close to the inputs and outputs, then this deep convolutional network can be more accurate and more efficient for training. Unlike ResNet, which adds a skip connection that bypasses the nonlinear transformation, DenseNet adds a direct connection from any layer to any subsequent layer. Hence, a l layer receives the feature maps of all former layers from x_0 to x_l

$$x_l = H_l\left(\left[x_0, x_1, \ldots, x_{l-1}\right]\right),$$

where $\left[x_0, x_1, \ldots, x_{l-1}\right]$ belongs to the spectrum of the feature map taken from the layers $0, 1, 2, \ldots, l-1$.

Fig. 1 shows the architecture of a five-layer tightly coupled block.

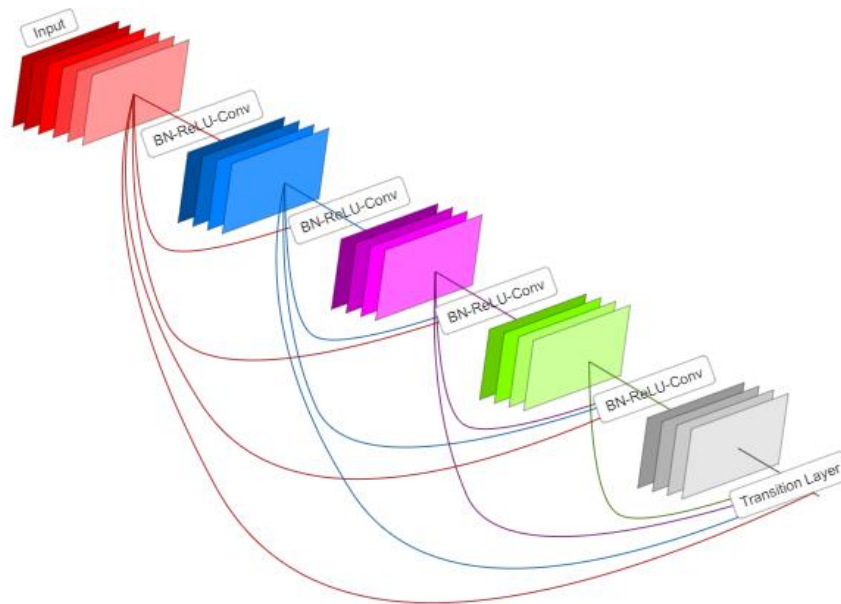Table 1 shows the proposed architecture of the DenseNet 169 network.

Fig. 1. Architecture of a five-layer densely connected unit of the neural network

Table 1. *Architecture of DenseNet 169 network*

Layers	Output size	DenseNet 169
Convolution	112×112	7×7 max. value, step 2
Association	56×56	3×3 convolution
Dense Block (1)	56×56	$\begin{bmatrix} 1\times1 \\ 3\times3 \end{bmatrix} \times 6$
Transition layer (1)	56×56	1×1 convolution
	28×28	2×2 mean value, step 2
Dense Block (2)	28×28	$\begin{bmatrix} 1\times1 \\ 3\times3 \end{bmatrix} \times 12$
Transition layer (2)	28×28	1×1 convolution
	14×14	2×2 mean value, step 2
Dense Block (3)	14×14	$\begin{bmatrix} 1\times1 \\ 3\times3 \end{bmatrix} \times 32$
Transition layer (3)	14×14	1×1 convolution
	7×7	2×2 mean value, step 2
Dense Block (4)	7×7	$\begin{bmatrix} 1\times1 \\ 3\times3 \end{bmatrix} \times 32$
Classification layer	1×1	7×7 total mean value, step 2
	1000	1000D full-coupled convolution

Let us build an algorithm for solving the problem of recognizing human emotions from a face image.

At the first stage, a set of data is collected for training and testing the classifier. In our case, we use FER2013 (Facial Expression Recognition 2013 Dataset) [16]. The dataset contains 35685 examples of 48×48 pixel grayscale face images divided into training and testing sets. The images are classified based on the emotions manifested in facial expressions (happiness, neutrality, sadness, anger, surprise, disgust, fear).

The second stage involves data processing, i.e. splitting the dataset into training, validation and testing samples, as well as preprocessing the

images themselves: bringing them to the same size (48×48 pixels) and color model (RGB).

At the third stage, we build a neural network model. First, we use the previously trained on ImageNet dataset DenseNet169 for feature extraction. Next, we create a classifier, the set of layers of which is shown in fig. 2. We train the model on a previously prepared training set with the following parameters: optimization – stochastic gradient descent, loss function – cross entropy and

metric – accuracy (to assess the quality of the classifier). Also, in order to improve the accuracy of the model, we use one of the transfer learning methods – Fine Tuning ("pre-training", "fine tuning"). After all training processes, we evaluate the quality of the model using appropriate metrics.

As a result of this algorithm, a model will be built that will be able to recognize human emotions with high accuracy.

```python
def classifier(inputs):
    x = tf.keras.layers.GlobalAveragePooling2D()(inputs)
    x = tf.keras.layers.Dense(256, activation="relu", kernel_regularizer = tf.keras.regularizers.l2(0.01))(x)
    x = tf.keras.layers.Dropout(0.3)(x)
    x = tf.keras.layers.Dense(1024, activation="relu", kernel_regularizer = tf.keras.regularizers.l2(0.01))(x)
    x = tf.keras.layers.Dropout(0.5)(x)
    x = tf.keras.layers.Dense(512, activation="relu", kernel_regularizer = tf.keras.regularizers.l2(0.01))(x)
    x = tf.keras.layers.Dropout(0.5) (x)
    x = tf.keras.layers.Dense(NUM_CLASSES, activation="softmax", name="classification")(x)

    return x
```

Fig. 2. Software implementation of the image classifier

Research results and their discussion

The program for solving problems of human emotion recognition is implemented in Python programming language. The code is completely written in one interactive web notebook with a runtime based on Google tensor processors on the Kaggle platform.

This application classifies human face images into seven different emotional states: happiness, neutrality, sadness, anger, surprise, disgust, fear. That is, the purpose of this program is to build and train a classifier that

would presumably tell us what emotion a person is depicted in the photo.

To train the classifier, we used the FER2013 dataset. We loaded it into the runtime environment, pre-processed the images, and generated training and test sets. We selected the optimal parameters, such as the number of training epochs – 30, the number of fine-tuning epochs – 20, the standard image size – 48×48 pixels and the battle size (number of objects in one iteration) – 64. We analyzed the training sample for distribution by classes, the visualization of the graph is shown in fig. 3.

Fig. 3. Distribution of images of the training sample by classes

66

ISSN 2522-9818 (print)
ISSN 2524-2296 (online)   *Innovative technologies and scientific solutions for industries. 2022. No. 3 (21)*

Further in the program, based on the pre-trained neural network DenseNet169, a model for feature extraction was obtained. A classifier with tightly connected layers was built. To improve the accuracy of the model and reduce the probability of overfitting, the Fine Tuning method was used. After training the model, we evaluated the quality of classification on a test sample.

We used Pandas libraries for data processing, NumPy – for matrix computations, Matplotlib and Plotly – for visualization, scikit-learn and TensorFlow (Keras) – for deep learning model development.

To classify emotions, depending on the input data and compare the results, a deep learning model was built to which the faces of people in the image were applied. The model built for this study is based on a densely connected convolutional neural network (DCCN).

The quality of the classifier is evaluated by the following metrics: accuracy (fig. 4); confusion matrix (fig. 5); precision, recall, f1-score (fig. 6); ROC curve and AUC value (fig. 7).

The accuracy value is relatively high – 63%, provided that the dataset has unbalanced classes. AUC also has a high value – 89%.
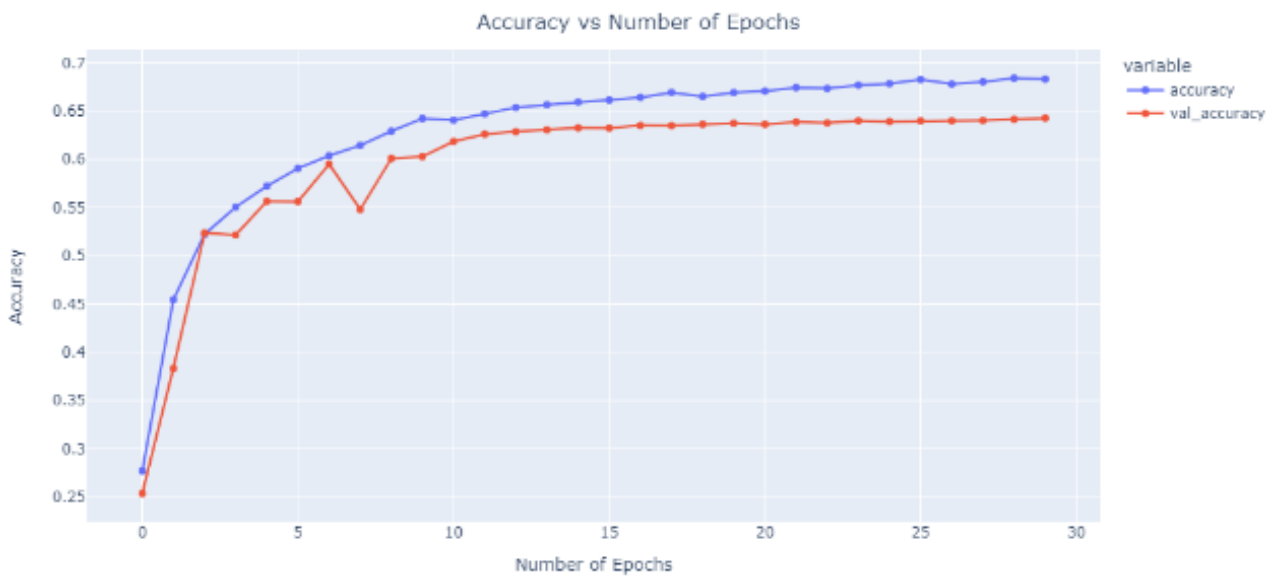

Fig. 4. Distribution of accuracy value by epochs

Fig. 5. Image of the matrix of inconsistencies

```
              precision    recall  f1-score   support

           0       0.52      0.61      0.56       958
           1       0.00      0.00      0.00       111
           2       0.47      0.34      0.39      1024
           3       0.87      0.86      0.87      1774
           4       0.55      0.69      0.61      1233
           5       0.51      0.48      0.49      1247
           6       0.75      0.75      0.75       831

    accuracy                           0.63      7178
   macro avg       0.52      0.53      0.52      7178
weighted avg       0.62      0.63      0.62      7178
```
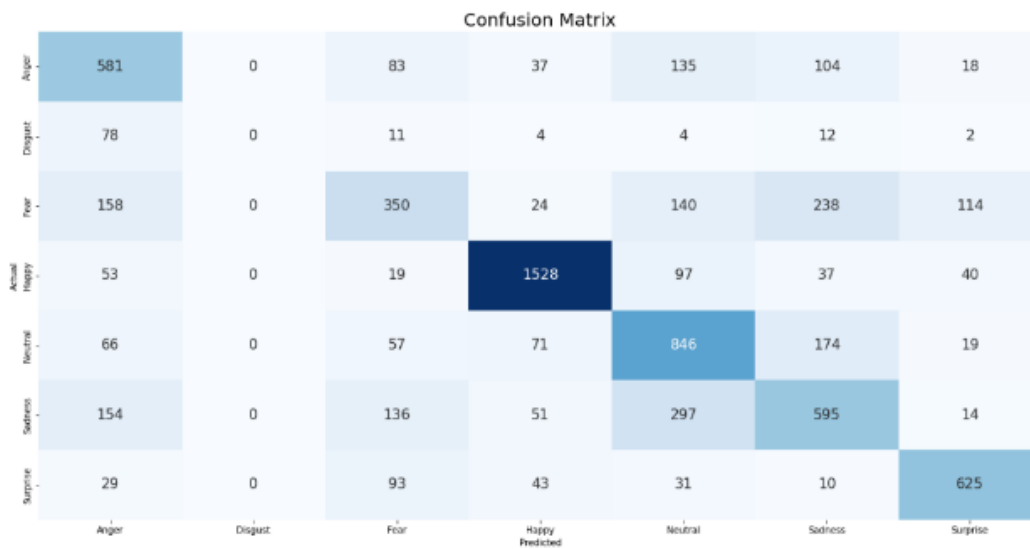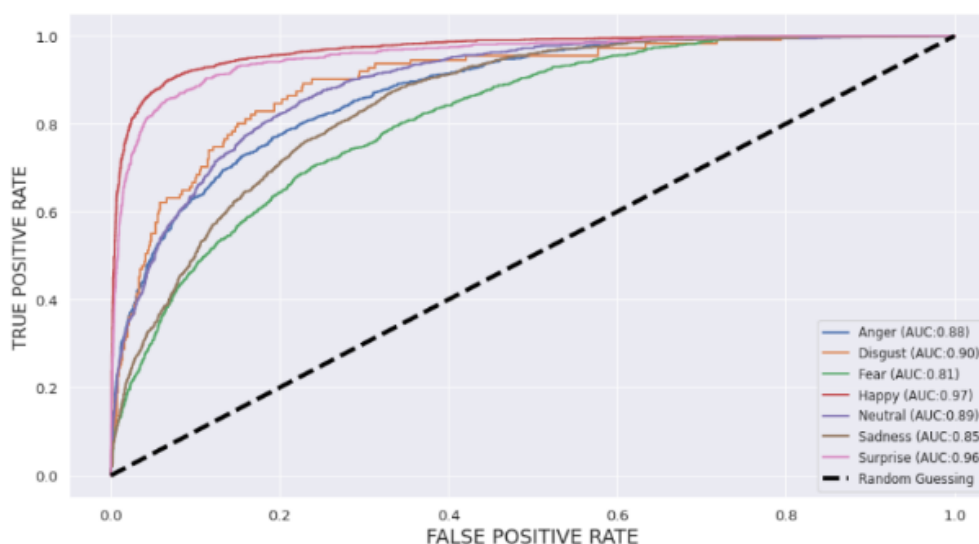
Fig. 6. Values of precision, recall, f1-score metrics

Fig. 7. Graph of the ROC curve

Conclusions and prospects for further development

In this paper, the methods of emotion classification are considered in detail and the best method for the purpose is chosen. Using the FER2013 dataset, the software implementation of the selected neural network was successfully trained and tested. The quality of the classifier was tested on a large number of different metrics: accuracy, precision, recall, f1-score, confusion matrix, ROC-AUC. Given the indicators, we can say that this work corresponds to the current level of scientific and technical knowledge in the field of deep learning.

The relevance of the topic is growing every year and the solution of such problems is becoming important in many fields.

The results obtained can be used to improve the work of robots and bots to make their conversation with a person look more emotional. Emotion recognition can also be used to collect user feedback on any project or service in business. Thanks to the created system, it becomes possible to analyze in detail what emotions a particular product and service evokes in consumers and make more accurate forecasts for future sales and changes in the popularity of products and services.

References

1. Hess, U., Thibault, P. (2009), "Darwin and Emotion Expression", *American Psychologist*, Vol. 64, No. 2, P. 120–129. DOI: https://doi.org/10.1037/a0013386

68

ISSN 2522-9818 (print)
ISSN 2524-2296 (online)   *Innovative technologies and scientific solutions for industries. 2022. No. 3 (21)*

2. Turabzadeh, S. (2018), "Facial expression emotion detection for real-time embedded systems", *Technologies*, No. 6 (1), 17 p. DOI: https://doi.org/10.3390/technologies6010017

3. Ekman, P. (1970), "Universal facial expressions of emotion", *California Mental Health Research Digest*, Vol. 8, No. 4, P. 151–158.

4. Viola, P., Jones, M. (2001), "Rapid object detection using a boosted cascade of simple features", *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*, Kauai, HI, USA, 8–14 December, P. 1–9.

5. Coots, T., Edwards, G., Taylor, Ch. (2001), "Active Appearance Model", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 23, No. 6, P. 681–685.

6. Coots, T., Taylor, K. (2004), "Statistical appearance models for computer vision", *Technical Report, University of Manchester, Wolfson Image Analysis Group, Imaging Science and Biomedical Engineering*, Manchester M13 9PT, United Kingdom, 125 p.

7. Van Dyke, Parunak H. (1995), *Neural Networks for Pattern Recognition*. New York: Christopher M. Bishop, Oxford University Press, 482 p.

8. Yashina, E., Artiukh, R., Pan, N., Zelensky, A. (2019), "Information technology for recognition of road signs using a neural network", *Innovative technologies and scientific solutions for industries,* No. 2 (8). DOI: https://doi.org/10.30837/2522-9818.2019.8.130

9. Mehryar, M., Rostamizadeh, A., Talwalkar, A. (2018), *Foundations of Machine Learning*, Cambridge: The MIT Press, 505 p.

10. Krizhevsky, A., Sutskever, I., Hinton, G. E. (2017), "Image Netclassification with deep convolutional neural networks", *Communications of the ACM*, Vol. 60, No. 6, P. 84–90.

11. Huang, G., Sun, Y., Liu, Z. (2016), "Deep networks with stochastic depth", *European Conf. on Computer Vision (ECCV)*, Amsterdam, Netherlands, October, P. 646–661.

12. Srivastava, R. K., Greff, K., Schmidhuber, J. (2015), "Training very deepnetworks", *Advances in Neural Information Processing Systems 28 (NIPS 2015)*, Montreal, Canada, December, P. 2377–2385.

13. He, K., Zhang, X., Ren, S. (2016), "Deep residual learning for image recognition", *The IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, 26 June – 1 July, P. 1–9.

14. Larsson, G., Maire, M., Shakhnarovich, G. (2017), "Fractalnet: ultra-deepneural networks without residuals", *5th International Conference on Learning Representa-tions*, Toulon, France, April, P. 1–11.

15. Gao, H., Zhuang, L., Laurens van der Maaten, Kilian, W. (2017), "Weinberger Densely Connected Convolutional Networks", *The IEEE on Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, USA, 21–26 July, P. 1–9.

16. FER2013 (Facial Expression Recognition 2013 Dataset), available at: https://paperswithcode.com/dataset/fer2013 (last accessed 12.05.2022).

Відомості про авторів / Сведения об авторах / About the Authors

Ювченко Катерина Сергіївна – бакалавр зі спеціальності "Прикладна математика", Харківський національний університет радіоелектроніки, студентка кафедри прикладної математики, Харків, Україна, e-mail: kateryna.iuvchenko@nure.ua; ORCID ID: https://orcid.org/my-orcid?orcid=0000-0002-5921-3049

Ювченко Екатерина Сергеевна – бакалавр по специальности "Прикладная математика", Харьковский национальный университет радиоэлектроники, студентка кафедры прикладной математики, Харьков, Украина.

Yuvchenko Kateryna – Bachelor of Applied Mathematics, Kharkiv National University of Radioelectronics, Student of Department of Applied Mathematics, Kharkiv, Ukraine.

Есілевський Валентин Семенович – кандидат технічних наук, доцент, Харківський національний університет радіоелектроніки, доцент кафедри прикладної математики, Харків, Україна; e-mail: valentyn.yesilevskyi@nure.ua; ORCID ID: https://orcid.org/ 0000-0002-5935-1505

Есилевский Валентин Семенович – кандидат технических наук, доцент, Харьковский национальный университет радиоэлектроники, доцент кафедры прикладной математики, Харьков, Украина.

Yesilevskyi Valentyn – PhD (Engineering Sciences), Associate Professor, Kharkiv National University of Radio Electronics, Associate Professor at the Department of Applied Mathematics, Kharkiv, Ukraine.

Середа Олена Григорівна – Харківський національний університет радіоелектроніки, старший викладач кафедри прикладної математики, Харків, Україна; e-mail: olena.sereda@nure.ua; ORCID ID: https://orcid.org/0000-0003-3579-0092

Середа Елена Григорьевна – Харьковский национальный университет радиоэлектроники, старший преподаватель кафедры прикладной математики, Харьков, Украина.

Sereda Olena – Kharkiv National University of Radio Electronics, Senior Lecturer at the Department of Applied Mathematics, Kharkiv, Ukraine.

СИСТЕМА РОЗПІЗНАВАННЯ ЕМОЦІЙ ЛЮДИНИ З ВИКОРИСТАННЯМ АЛГОРИТМІВ ГЛИБОКОГО НАВЧАННЯ

Предметом дослідження є програмна реалізація нейронного класифікатора зображень. У статті розглянуто емоції як особливий вид психічних процесів, що виражають переживання людини, її ставлення до навколишнього світу та до себе. Емоційна сфера може виражатися різними способами: мімікою, позою, руховими реакціями, голосом. Проте найбільшу виразність має обличчя людини. Технології для розпізнавання людських емоцій використовуються для поліпшення обслуговування клієнтів, прийняття рішення після співбесіди з кандидатами, для збільшення емоційного впливу реклами тощо. Тому **метою** роботи є знаходження та оптимізація найбільш задовільного, з погляду точності, алгоритму класифікації емоцій людини із зображенням обличчя. У статті вирішуються такі **завдання:** огляд і аналіз сучасного стану "розпізнавання емоцій"; розгляд методів класифікації; вибір найкращого методу для поставленого завдання; розроблення програмної реалізації для класифікації емоцій; проведення аналізу роботи класифікатора, формулювання висновків про проведену роботу на основі отриманих даних. У статті також використовується метод класифікації зображень, оснований на щільно зв'язаній згортковій нейронній мережі. Отримані **результати** довели, що для вирішення завдань розпізнавання емоцій добре підходить метод класифікації зображень, оснований на щільно зв'язаній згортковій нейронній мережі, оскільки зазначений метод має досить високу точність. Оцінена якість класифікатора за такими метриками: accuracy; confusion matrix; precision, recall, f1-score; ROC-крива та значення AUC. Значення accuracy відносно високе – 63%, за умови, що набір даних має незбалансовані класи. AUC також має високе значення – 89%. **Висновки.** Отримана модель має високі показники розпізнавання емоцій людини та надалі може успішно використовуватися за призначенням.

Ключові слова: виявлення об'єктів; класифікація об'єктів; навчання з учителем; розпізнавання емоцій.

СИСТЕМА РАСПОЗНАВАНИЯ ЭМОЦИЙ ЧЕЛОВЕКА С ИСПОЛЬЗОВАНИЕМ АЛГОРИТМОВ ГЛУБОКОГО ОБУЧЕНИЯ

Предметом исследования является программная реализация нейронного классификатора изображений. В работе рассмотрены эмоции как особый вид психических процессов, раскрывающих переживания человека, его отношения к окружающему миру и самому себе. Эмоции могут быть выражены разными способами: мимикой, позой, двигательными реакциями, голосом. Однако наибольшую выразительность имеет лицо человека. Технологии для распознавания человеческих эмоций используются фирмами для улучшения обслуживания клиентов, принятия решений о собеседовании с кандидатами и для оптимизации эмоционального воздействия рекламы. Поэтому **целью** работы является нахождение и оптимизация наиболее удовлетворительного, с точки зрения точности, алгоритма классификации эмоций человека с изображением лица. В статье решаются следующие **задачи**: обзор и анализ современного состояния задачи распознавания эмоций; рассмотрение методов классификации; выбор наилучшего способа для поставленной задачи; разработка программной реализации для классификации эмоций; проведение анализа работы классификатора, формулирование выводов о проделанной работе на основе полученных данных. В статье также используется метод классификации изображений, основанный на плотно связанной сверточной нейронной сети. **Результаты** показали, что для решения задач распознавания эмоций хорошо подходит метод классификации изображений, основанный на плотно связанной сверточной нейронной сети, поскольку он имеет достаточно высокую точность. Произведена оценка качества классификатора по следующим метрикам: accuracy; confusion matrix; precision, recall, f1-score; ROC-кривая и AUC. Значение accuracy относительно высокое – 63%, при условии, что набор данных имеет несбалансированные классы. AUC также имеет высокое значение – 89%. **Выводы**. Полученная модель имеет высокие показатели распознавания эмоций человека и в дальнейшем может успешно использоваться по назначению.

Ключевые слова: обнаружение объектов; классификация объектов; обучение с учителем; распознавание эмоций.

Бібліографічні описи / Bibliographic descriptions

Ювченко К. С., Есілевський В. С., Середа О. Г. Система розпізнавання емоцій людини з використанням алгоритмів глибокого навчання. *Сучасний стан наукових досліджень та технологій в промисловості*. 2022. № 3 (21). С. 60–69. DOI: https://doi.org/10.30837/ITSSI.2022.21.060

Yuvchenko, K., Yesilevskyi, V., Sereda, O. (2022), "Human emotion recognition system using deep learning algorithms", *Innovative Technologies and Scientific Solutions for Industries*, No. 3 (21), P. 60–69. DOI: https://doi.org/10.30837/ITSSI.2022.21.060