DOI: https://doi.org/10.15276/hait.06.2023.13
UDC 004.8

Pseudo-labeling of transfer learning convolutional neural network data for human facial emotion recognition

Olena O. Arsirii[1)]
ORCID: https://orcid.org/0000-0001-8130-9613; e.arsiriy@gmail.com. Scopus Author ID: 54419480900
Denys V. Petrosiuk[1)]
ORCID: https://orcid.org/0000-0003-4644-3678; d.petrosyuk1994@gmail.com. Scopus Author ID: 54419479400
[1)] Odessa Polytechnic National University, 1, Shevchenko Ave. Odessa, 65044, Ukraine

ABSTRACT

The relevance of solving the problem of facial emotion recognition on human images in the creation of modern intelligent systems of computer vision and human-machine interaction, online learning and emotional marketing, health care and forensics, machine graphics and game intelligence is shown. Successful examples of technological solutions to the problem of facial emotion recognition using transfer learning of deep convolutional neural networks are shown. But the use of such popular datasets as DISFA, CelebA, AffectNet, for deep learning of convolutional neural networks does not give good results in terms of the accuracy of emotion recognition, because almost all training sets have fundamental flaws related to errors in their creation, such as the lack of data of a certain class, imbalance of classes, subjectivity and ambiguity of labeling, insufficient amount of data for deep learning, etc. It is proposed to overcome the noted shortcomings of popular datasets for emotion recognition by adding to the training sample additional pseudo-labeled images with human emotions, on which recognition occurs with high accuracy. The aim of the research is to increase the accuracy of facial emotion recognition on the image of a human by developing a pseudo-labeling method for transfer learning of a deep neural network. To achieve the aim, the following tasks were solved: a convolutional neural network model, previously trained on the ImageNet set using the transfer learning method, was adjusted on the RAF-DB data set to solve emotion recognition tasks; a pseudo-labeling method of the RAF−DB set data was developed for semi-supervised learning of a convolutional neural network model for the task of facial emotion recognition; the accuracy of facial emotion recognition was analyzed based on the developed convolutional neural network model and the method of pseudo-labeling of RAF-DB set data for its correction. It is shown that the use of the developed method of pseudo-labeling data and transfer learning of the MobileNet V1 convolutional neural network model allowed to increase the accuracy of facial emotion recognition on the images of the RAF-DB dataset by 2 percent (from 76 to 78 %) according to the F1 estimate. At the same time, taking into account the significant imbalance of the classes, for the 7 main emotions in the training set, we have a significant increase in the accuracy of recognizing a few representatives of such emotions as surprise (from 71 to 77 %), fearful (from 64 to 69%), sad (from 72 to 76 %), angry with (from 64 to 74 %), neutral (from 66 to 71 %). The accuracy of recognizing the emotion of happy, which is the most common, decreased (from 91 to 86 %) Thus, it can be concluded that the use of the developed pseudo-labeling method gives good results in overcoming such shortcomings of datasets for deep learning of convolutional neural networks such as lack of data of a certain type, imbalance of classes, insufficient amount of data for deep learning, etc.

Keywords: pseudo-labeling data; semi-supervised learning; transfer learning; convolution neural networks; facial emotion recognition

***For citation*: Arsirii O. O., Petrosiuk D. V. “Pseudo-labeling of transfer learning convolutional neural network data for human facial emotion recognition“. *Herald of Advanced Information Technology.* 2023; Vol. 6 No.3: 203–214. DOI: https://doi.org/10.15276/hait.06.2023.13**

INTRODUCTION

Modern intelligent systems of computer vision and human-machine interaction, online learning and emotional marketing, health care and forensics, machine graphics and game intelligence have a basic foundation in the form of a model of social interaction.

The development of such a model requires information about the emotional the condition of a person, the acquisition of which is connected with the solution of the problem of Facial Emotion Recognition (FER) on images of human faces [1, 2]. The works show successful examples of technological solutions to the FER problem using transfer learning of deep Convolutional Neural Networks (CNN) [3, 4], [5, 6]. But the use of such popular datasets as DISFA, CelebA, AffectNet [7] for deep learning of CNN does not give good results in terms of FER accuracy because these training samples have fundamental disadvantages such as lack of data of a certain class, class imbalance, subjectivity and ambiguity of labeling, insufficient volume of data for deep learning, etc. Therefore, the topic related to increasing the accuracy of FER due to the use of transfer learning, which allows using an already trained CNN of the MobileNet V1 type on a large and more advanced dataset, is relevant ImageNet with further post-training already on a specialized and significantly smaller and unbalanced Real-world set Affective Faces Database (RAF-DB). At the same time, it is proposed to overcome the indicated shortcomings of RAF-DB by adding to the training sample additional pseudo-labeled images with human emotions, on which FER occurs with high accuracy. Let's consider the CNN transfer learning method and the most common ways of adding data to the training sample in more detail.

ANALYSIS OF EXISTING RESEARCH AND PUBLICATIONS

The approach of Transfer Learning, consists in the transfer of feature description functions obtained by the CNN model with multiple layers in the process of solving the initial recognition task to the target recognition task [3, 4], [5, 6], [8, 9]. The MobileNet family of networks is widely used for FER tasks on mobile platforms due to its ease (4-3M parameters) [10, 11], [12]. The MobileNet model is based on the structure of a deep separable convolution, which can transform a standard convolution into a deep convolution and a point convolution with a convolution kernel 1×1.

We briefly summarize the stages of transfer learning:

Stage 1. Convolutional layers are extracted from the previously trained model (*pre-train*).

Stage 2. The convolutional layers are “frozen” to avoid destroying any information they contain during further training (*train*).

Stage 3. Several new training layers are added on top of the frozen layers, which will learn to turn the old feature maps into predictions for the new data set.

Stage 4. New layers are trained on the *target* data set.

Stage 5. Fine-tuning of the entire CNN model. Namely, unblocking the “frozen” part and retraining the entire CNN model on the target data set with a very low training speed. As shown by numerous studies, the implementation of this stage allows the pre-trained CNN model to gradually adapt to new data.

Thus, for the implementation of transfer learning, in addition to the pre-train CNN model, is needed a target training data set, which, as a rule, requires expansion.

To solve the problem of expansion, methods of adding such data to the training sample are used, which do not create additional imbalances or false patterns within the sample [13, 14]. Such methods include: *Data augmentation*, *Hard Samples Mining*, *Generative Adversarial Networks*, *Dropout* as an imitation of adding data and *Pseudo-labeling*

Data augmentation consists in modifying the available images of the training sample using the operations of adding noise, rotation, scaling, mirroring, manipulation of color, contrast, as well as multiplicity according to a certain rule in order to expand the training sample and increase its diversity. An effective type of augmentation is CutOut [15] and Random Erasing [16] (painting random rectangles in the picture so that the CNN could not learn to recognize an object by one specific detail of its appearance, for example, to recognize a car by its wheel). As a rule, in the process of CNN training, the intensity of augmentations gradually decreases, which allows the neural network to better adapt to the initial distribution, at the same time improves its convergence and stability due to the increase in data diversity. [17, 18], [19].

Hard Samples Mining. In order for the neural network to distinguish the object from similar objects in the background, is added to the dataset *hard negative examples* – fragments of images that look like an object. At the same time, as hard negative examples of images used are images that released false positives from the network, which was trained in a small number of epochs [20, 21].

Generative Adversarial Networks (GAN) – a combination of two neural networks, in which two algorithms “generator” and “discriminator” work simultaneously. The task of the generator is to generate images of a given category. The task of the discriminator is to try to recognize the created image. In fact, GANs create their own training data. However, modern GANs often generate incorrect images, and there are also big problems with their convergence. However, you can still try to use them to generate facial images [22] (Fig. 1), taking into account positive examples of background generation or for adaptation of existing images to other conditions (for example, generation from night images [23, 24], [25]).

Dropout as an imitation of adding data. The CNN Dropout regularization method [26] is considered as an imitation of adding data. In this case, the reasoning is correct that for the layer *i* of the CNN there is no difference between whether the input data changes or the values of the outputs of the layer *i*-1 of the CNN change. The method showed good results for CIFAR-10 [27] – a dataset containing a small number of images (about 10,000), which is why the simulation of data addition for this dataset seems expedient.

Fig. 1. **An example of expanding the learning curve using generated images**
***Source:* compiled by the [22]**

Pseudo-labeling. Recently, for the data of the ImageNet set [28], one of the best results in competitive practice showed the direction of pseudo-labeling of the method self-supervised learning [29, 30], [31]. According to the scheme (Fig. 2), the neural network model is first trained on a collection of labeled data, then the answers (Predict) of the trained model are used to label a set of Unlabeled data, and then Pseudo-labeled data are used to train the model, while pseudo-labels are added only to those data whose predictions performed with a high degree of reliability.

The aim of the research is to develop a pseudo-labeling method Data for advanced transfer learning of deep convolutional neural networks increasing the accuracy of facial emotion recognition in a human face image.

THE AIM AND OBJECTIVES OF THE RESEARCH

To achieve the aim, it is necessary to solve the following tasks:

1) to solve FER problems, adjust the CNN model pre-trained on the ImageNet data set using the transfer learning method on the RAF-DB data set;

2) develop a method for pseudo-labeling the data of the RAF-DB data set for the semi-supervised learning of the CNN model for the FER problem;

3) analyze the accuracy of facial emotion recognition in a human face image based on the developed CNN model and the method of pseudo-labeling the RAF-DB data set for its correction.

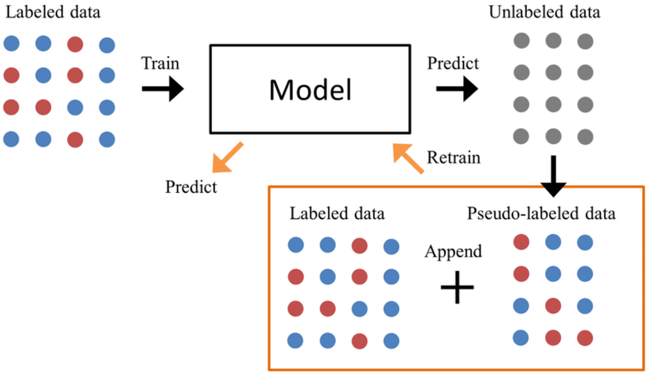


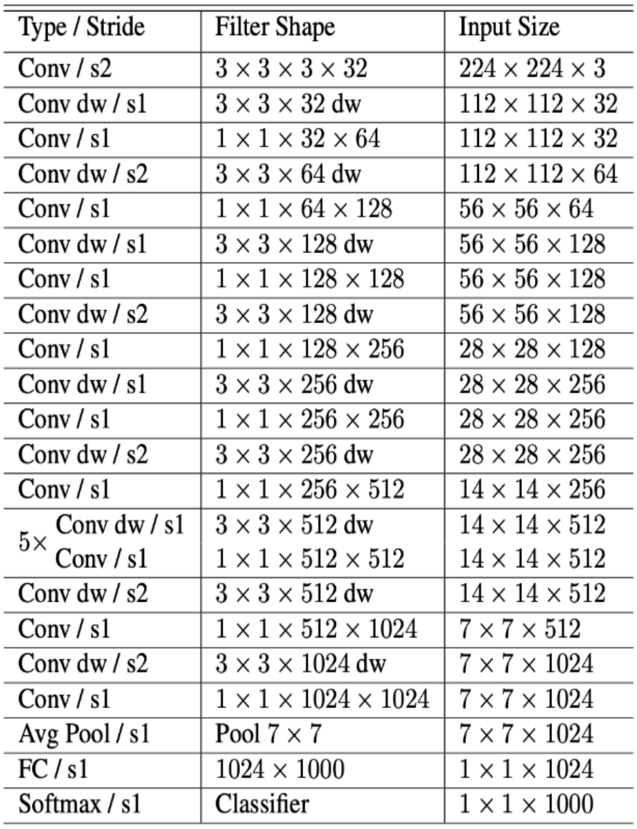
***Fig 2.* Scheme Pseudo-labeling**
***Source:* compiled by the [31]**

PRESENTATION OF THE MAIN RESEARCH MATERIAL

Adjusting the pre-trained CNN model to solve the FER problem

In the research carried out in the previous works of the authors [32, 33], when developing a CNN model for the FER problem, taking into account the requirements for resource intensity and learning speed, it was proposed to use CNN MobileNet V1 pre-trained on the dataset ImageNet (Fig. 3).

This version is designed for use in mobile applications and is the first mobile computer vision model obtained using framework TensorFlow. MobileNet uses depth-separated convolutions. This significantly reduces the number of parameters compared to a neural network with regular convolutions with the same depth of network. A depth-separated convolution consists of two operations: depth and pointwise convolution.



Type / Stride	Filter Shape	Input Size
Conv / s2	3 × 3 × 3 × 32	224 × 224 × 3
Conv dw / s1	3 × 3 × 32 dw	112 × 112 × 32
Conv / s1	1 × 1 × 32 × 64	112 × 112 × 32
Conv dw / s2	3 × 3 × 64 dw	112 × 112 × 64
Conv / s1	1 × 1 × 64 × 128	56 × 56 × 64
Conv dw / s1	3 × 3 × 128 dw	56 × 56 × 128
Conv / s1	1 × 1 × 128 × 128	56 × 56 × 128
Conv dw / s2	3 × 3 × 128 dw	56 × 56 × 128
Conv / s1	1 × 1 × 128 × 256	28 × 28 × 128
Conv dw / s1	3 × 3 × 256 dw	28 × 28 × 256
Conv / s1	1 × 1 × 256 × 256	28 × 28 × 256
Conv dw / s2	3 × 3 × 256 dw	28 × 28 × 256
Conv / s1	1 × 1 × 256 × 512	14 × 14 × 256
5× Conv dw / s1 Conv / s1	3 × 3 × 512 dw 1 × 1 × 512 × 512	14 × 14 × 512 14 × 14 × 512
Conv dw / s2	3 × 3 × 512 dw	14 × 14 × 512
Conv / s1	1 × 1 × 512 × 1024	7 × 7 × 512
Conv dw / s2	3 × 3 × 1024 dw	7 × 7 × 1024
Conv / s1	1 × 1 × 1024 × 1024	7 × 7 × 1024
Avg Pool / s1	Pool 7 × 7	7 × 7 × 1024
FC / s1	1024 × 1000	1 × 1 × 1024
Softmax / s1	Classifier	1 × 1 × 1000

Fig 3. **Architecture MobileNet V1**

Source: **compiled by the [12]**

After loading the CNN model MobileNet V1 is pre-trained on the dataset ImageNet add data augmentation layers to the input of the neural network to combat overtraining and improve training data. In this case, random mirroring of the image horizontally (*RandomFlip*), multiplicity of the image (*RandomZoom*), random rotation of the image (*RandomRotation*) and random transformation (*RandomTranslation*) are used as data augmentation.

In the TensorFlow API specifications Keras let's create a *ModelCheckpoint* callback to save the training history and the model itself. Callback is used in conjunction with assisted training *model.fit* to save the model or weights (in the checkpoint file) at some interval so that the model or weights can be loaded later to continue training from the saved state.

The Real-world Affective Faces (RAF-DB) dataset represents 29672 images of human facial emotion, labeled with seven main emotion classes (0 – *surprised;* 1 – *fearful;* 2 – *disgusted;* 3 –*happy;* 4 – *sad;* 5 – *angry;* 6 – *neutral*) or eleven complex (multiple-labeled) emotions, for example, happy-surprised, sad-angry, surprised-frightened. However, the faces in the images vary greatly in the age, gender and ethnicity of the subjects, head position, lighting conditions, occlusion (for example, glasses, facial hair or self-occlusion), post-processing operations (for example, various filters and special effects), etc. [34].

For transfer training, download the RAF-DB archive and unpack it, specifying such parameters as the total number of objects, image size, etc. It is also necessary to initialize the variable that corresponds to the main emotions in the dataset (Fig. 4a) and divide the loaded data into training and test parts (Fig. 4b). In total, we have 12.271 images in the training set and 3.068 in the test set. The relative distribution of uploaded images by classes of basic emotions indicates the imbalance of classes in the dataset.

Next, we begin the learning process. The callback parameter *save_best_* saves the model only if the accuracy of the model increases depending on the number of training epochs.

Using this parameter allows you to identify the best model, even if it was not obtained at the last training epoch. In our case, 10 epochs were used to train the CNN model. The structure of the CNN model after transfer learning for the FER problem is shown in Fig. 4c.

Taking into account the structure of the MobileNet V1 model (Fig. 3), the following can be noted. Conv2D-layers in Keras were chosen as intermediate layers of the obtained CNN model for the FER problem. This layer is similar to the Dense-layer, and contains weights and biases that are subject to optimization (matching). The Conv2D-layer also contains filters (“kernels”), the values of which are also optimized. This layer creates a convolution kernel, which together with the input of the layer creates a tensor (2-dimensional arrays) of the output data. Empirically, 3 Conv2D-layers were chosen with the number of filters 64, 128 and 256, with the *relu* activation function and with kernel sizes (5.5) (5.5) and (3.3), each of which is followed by a MaxPooling2D-layer of size kernels (2.2), also Dense-layers, one of which is the size of the filter (128) and the last one, with the number of outputs equal to the number of emotions, i.e. 7.

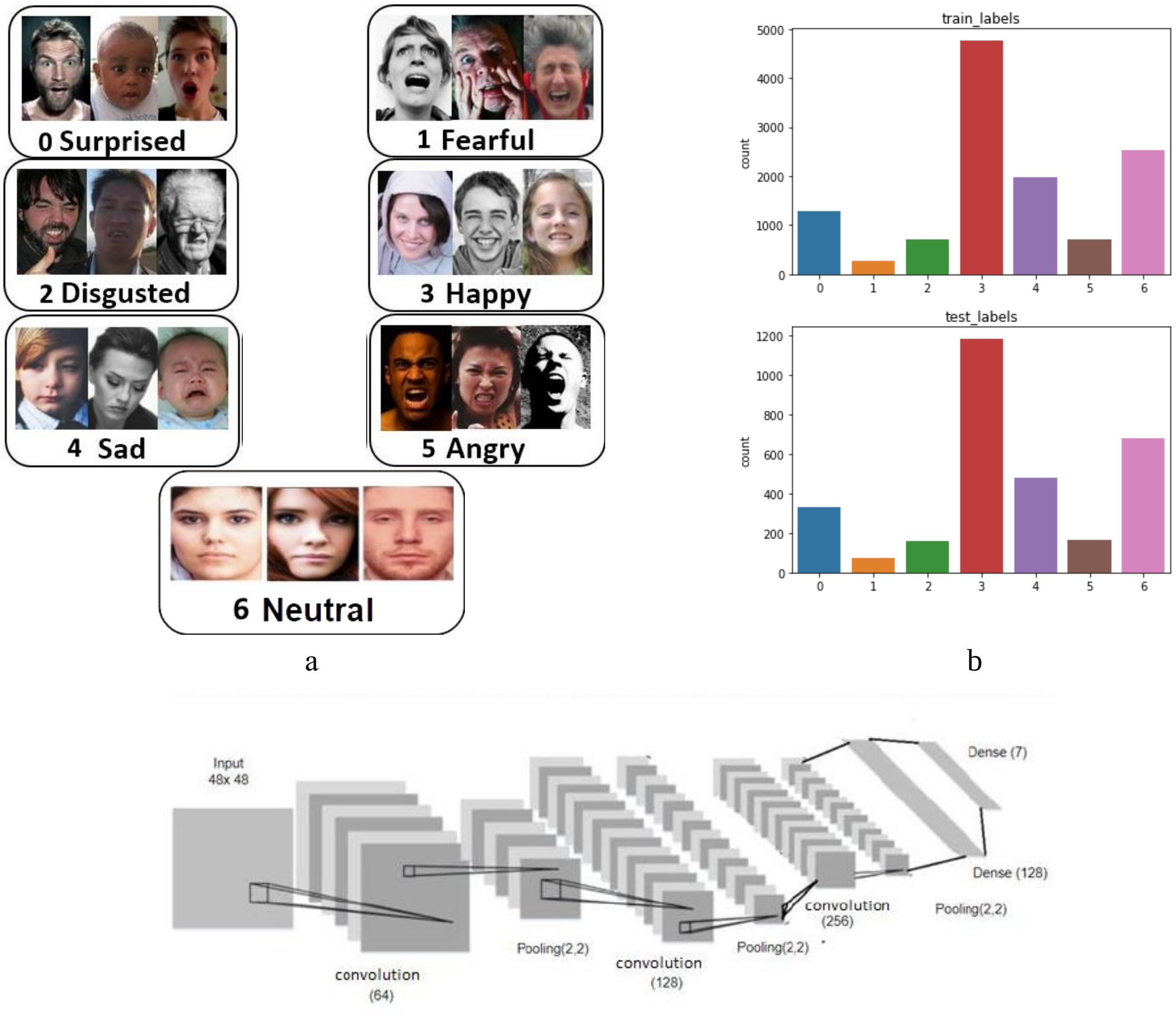


Fig. 4. **Adjustment of the pre-trained CNN model for solving the FER problem: a – labeled images of the RAF-DB dataset; b – distribution of images by classes in the training set Train_labels and test set Test_labels; c – the structure of the adjusted model CNN for the FER problem – CNN-FER)**

Source: **compiled by the authors**

A method of pseudo-labeling the RAF-DB data set for semi-supervised training of the CNN model for the FER task

When developing the pseudo-labeling method the training set *Train_labels* (Fig. 4, b) with the size of 12271 (*Train_labels* = 12271) it is necessary to divide into the training set (Y_{train}=8834), validation set (Y_{val}=2209) and “unlabeled set” ($Y_{unlabeled}$ =1228). That is, during the pseudo-labeling of the data, the labels of the class of images falling into $Y_{unlabeled}$ are ignored. Such a division is carried out using the *train_test_split* module libraries *Scikit-learn* with *stratify* and *random_state* parameters. The result of the data distribution of the dataset RAF-DB for further pseudo-labeling of data is shown in Fig 5. Note that the test part of the sample *Test_labels* remains unchanged.

Pseudo-labeling method these data were implemented using the teacher (CNN-FER) and student (CNN(Retrain)-FER) models.

It consists of four main stages:

Stage 1 train the CNN-FER model on labeled images (Y_{train} and Y_{val});

Stage 2 use the trained CNN-FER model for creating pseudo-labels on unlabeled images ($Y_{unlabeled}$);

Stage 3 add to training data (Y_{train} and Y_{val}) all confident predictions $Y_{unlabeled}$ with probability predicted (P) for (y ={0,1,2,3,4,5,6}) classes above a certain threshold (t)

P(y={0,1 ,2,3,4,5,6}|x) > t ;

Stage 4 train the CNN(Retrain)-FER model on the combination (Y_{train}+Y_{val}+$Y_{unlabeled}$) sample.

Stages 2, 3, 4 can be repeated as many times as the given value tallows.

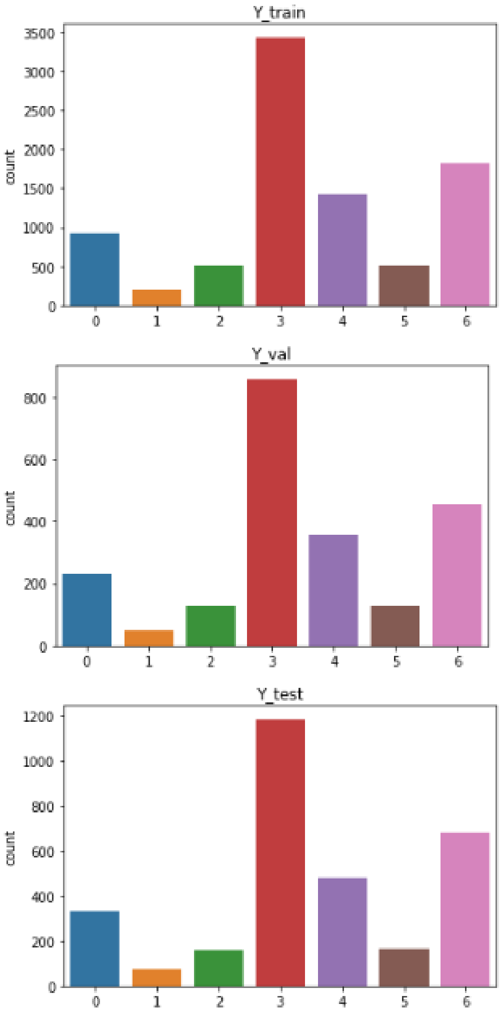


Fig. 5. **R distribution images by class in the elections $\mathbf{Y_{train}}$, $\mathbf{Y_{val}}$ and $\mathbf{Y_{unlabeled}}$**
Source: **compiled by the authors**

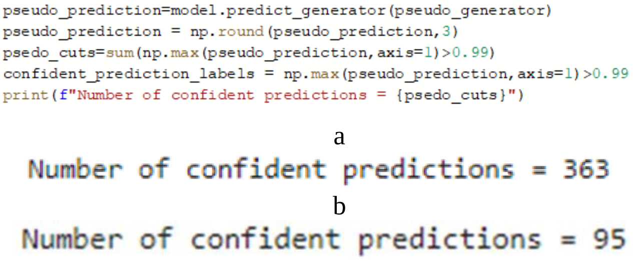
We briefly present the features of the implementation of the pseudo-labeling method. The $Y_{unlabeled}$ data will be presented as a generator using the *ImageDataGenerator* class in the API TensorFlow Keras specification.

In Fig. 6, according to steps 2, 3 and 4 of the pseudo-labeling method show the program code for facial recognition emotions in images $Y_{unlabeled}$ by the CNN-FER model (Fig. 6a) and the results of emotion recognition of the 7 specified classes in with probability predicted

P(y ={0,1,2,3,4,5,6}|x) > 0.99.

The result of the execution of the program code after the 1st and 2nd iterations is shown in Fig. 6b and Fig. 6c. As we can see, as a result of the first execution of step 2 on $Y_{unlabeled}$ images, it was possible to add 363 images ($Y_{pseudo\_labeled}$) to Y_{train} , the correct recognition of which is determined by the CNN-FER model with a probability of more than 0.99.

According to stage 4, the CNN(Retrain)-FER model is already trained using $Y_{train} + Y_{pseudo\_labeled}$.



```
pseudo_prediction=model.predict_generator(pseudo_generator)
pseudo_prediction = np.round(pseudo_prediction,3)
psedo_cuts=sum(np.max(pseudo_prediction,axis=1)>0.99)
confident_prediction_labels = np.max(pseudo_prediction,axis=1)>0.99
print(f"Number of confident predictions = {psedo_cuts}")
```

a

```
Number of confident predictions = 363
```

b

```
Number of confident predictions = 95
```

c

Fig. 6. **Results of using the method pseudo-labeling of data**
Source: **compiled by the authors**

After performing the second iteration of the pseudo-labeling method, it was possible to add 95 images ($Y_{pseudo\_labeled}$) to Y_{train} .

Analysis of facial emotion recognition accuracy based on the developed CNN-FER model and data pseudo-labeling method

Due to the large inconsistency of labeled data in the RAF-DB set (Fig. 4b and 5), the accuracy of facial emotion recognition in a human face image was assessed by the F1-measure value (the harmonic mean of the *Precision* and *Recall* indicators):

$$F_1 = 2\frac{Precision \times Recall}{Precision + Recall}, \quad (1)$$

$$Precision = \frac{TP}{TP+FP}, \quad (2)$$

$$Recall = \frac{TP}{TP+FN}, \quad (3)$$

where *TP* are true positive examples $TP_i = T_i$; *FP* are false positive examples $FP_i = \sum_{c \in Classe} F_{i,c;}$, *FN* are false negative examples $FN_i = \sum_{c \in Classe} F_{c,i}$.

In this case, the *TP*, *FP* and *FN* indicators were calculated using the rows and columns of the confusion matrix (Fig. 7a, Fig. 7d and Fig. 7e). And in Fig. 7b, Fig. 7c and Fig. 7f the characteristics of the accuracy of emotion recognition for the *F*1 and *Precision* and *Recall* metrics are shown.

The characteristics of the recognition accuracy are shown depending on the iteration number by the method of pseudo-labeling data (b – the first iteration; c – the second iteration; f – the third iteration).

As can be seen after the first iteration Fig. 7a and Fig. 7b), the CNN-FER model is the most accurate recognizes the emotions of *happy*, *angry*, *neutral* and *surprised*. And the overall recognition accuracy is 76 %. In the second iteration (Fig.7c and Fig.7d), already for the CNN(Retrain)-FER model, the following results of increasing the accuracy of recognition of the emotion of *surprised* (from 71 to 77 %), *fearful* (from 64 to 69 %), *sad* (from 72 to 76%), *angry* with (from 64 to 74%), *neutral* (from 66 to 71 %). The accuracy of recognizing the emotion of *disgusted* did not change, but the accuracy of recognizing the emotion of *happy*, which is the most common in the training sample, decreased (from 91 to 86 %). Thus, the overall accuracy of the CNN(Retrain)-FER model became 2 percent better compared to the CNN-FER model and in the end we have an overall accuracy of 78 %. Comparing the average recognition accuracy on the third iteration of the method (Fig. 7e and Fig. 7f) of pseudo-marking with the one obtained on the second, we can see that the result has deteriorated. Only the emotions of *happy* and *neutral* received an increase in accuracy (from 86 to 91 % and from 71 to 73 %), respectively.

Thus, the general conclusion is as follows: the use of the developed data pseudo-labeling method gives good results in overcoming such shortcomings of datasets for deep learning of convolutional neural networks as lack of data of a certain class, imbalance of classes, insufficient amount of data for deep learning, etc.

CONCLUSION

In general, the following conclusions can be drawn based on the results of the research conducted in the paper:

– The relevance of solving the problem of facial emotions recognizing on human images in the creation of modern intelligent systems of computer vision and human-machine interaction, online learning and emotional marketing, health care and forensics, machine graphics and game intelligence has been studied. Successful examples of technological solutions to the problem of emotion recognition using transfer learning of deep convolutional neural networks are shown.

– It has been studied that the use of such popular datasets as DISFA, CelebA , AffectNet , for deep learning of convolutional neural networks does not give good results in terms of the accuracy of emotion recognition because almost all training samples have fundamental flaws related to errors in their creation , such as the absence data of a certain type, imbalance of classes, subjectivity and ambiguity of labeling, insufficient amount of data for deep learning, etc.

– It is proposed to overcome the noted shortcomings of popular datasets for emotion recognition by adding to the training sample additional pseudo-labeled images with human emotions, on which recognition occurs with high accuracy.

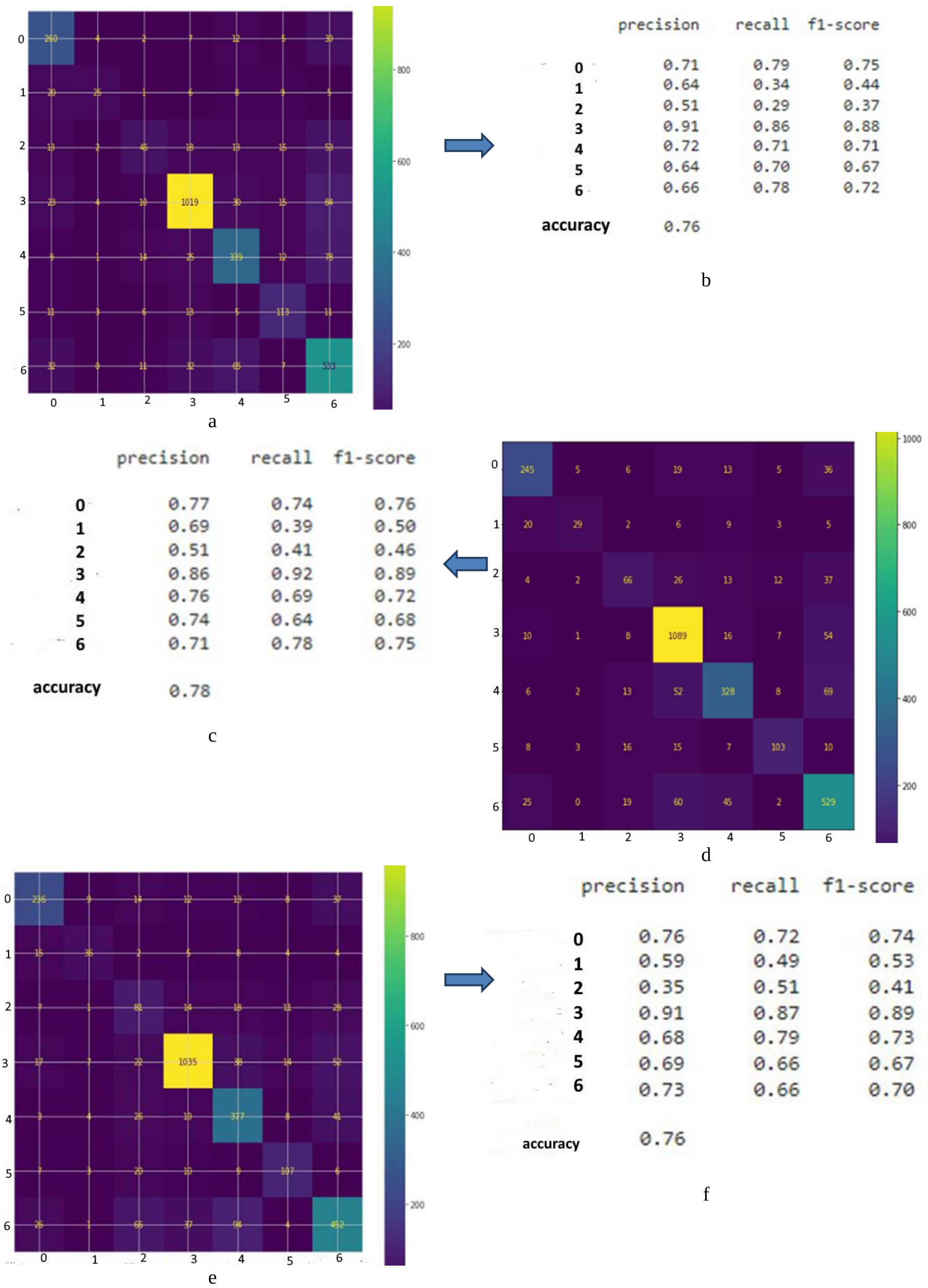
– It is shown that to solve the problem of expanding training datasets, methods are used to add such data to the training set that do not create additional imbalance or false intra-sample patterns. These methods include: Data augmentation, Hard Sample Mining, Generative Adversarial Networks, Dropout as an imitation of adding data, Pseudo-labeling. The advantages of Pseudo-labeling are shown.

– The aim of the research is determined, which consists in increasing the accuracy of facial emotion recognition on the image of a person's face by developing a method of pseudo-labeling data for transfer learning of a deep neural network.

– Pseudo-labeling method for solving facial emotion recognition tasks on the RAF-DB data set, the convolutional neural network model previously trained on the ImageNet set was improved.

– Pseudo-labeling method of the RAF-DB set data was developed for semi-supervised learning of a convolutional neural network model for the task of facial emotions recognizing on human images.

– The accuracy of emotion recognition in human images was analyzed based on the developed convolutional neural network model and the pseudo-labeling method of the RAF-DB data set for its correction.



	0	1	2	3	4	5	6
0	260	4	2	7	12	5	39
1	20	25	1	6	8	9	5
2	13	2	46	18	13	15	58
3	23	4	10	1019	30	15	84
4	9	1	14	25	339	12	78
5	11	3	6	13	5	113	11
6	32	0	11	32	65	7	533

a

	precision	recall	f1-score
0	0.71	0.79	0.75
1	0.64	0.34	0.44
2	0.51	0.29	0.37
3	0.91	0.86	0.88
4	0.72	0.71	0.71
5	0.64	0.70	0.67
6	0.66	0.78	0.72
accuracy	0.76		

b

	precision	recall	f1-score
0	0.77	0.74	0.76
1	0.69	0.39	0.50
2	0.51	0.41	0.46
3	0.86	0.92	0.89
4	0.76	0.69	0.72
5	0.74	0.64	0.68
6	0.71	0.78	0.75
accuracy	0.78		

c

	0	1	2	3	4	5	6
0	245	5	6	19	13	5	36
1	20	29	2	6	9	3	5
2	4	2	66	26	13	12	37
3	10	1	8	1089	16	7	54
4	6	2	13	52	328	8	69
5	8	3	16	15	7	103	10
6	25	0	19	60	45	2	529

d

	0	1	2	3	4	5	6
0	236	9	14	12	13	8	37
1	15	36	2	5	8	4	4
2	7	1	81	14	18	11	28
3	17	7	22	1035	38	14	52
4	3	4	26	19	377	8	41
5	7	3	20	10	9	107	6
6	26	1	66	37	94	4	452

e

	precision	recall	f1-score
0	0.76	0.72	0.74
1	0.59	0.49	0.53
2	0.35	0.51	0.41
3	0.91	0.87	0.89
4	0.68	0.79	0.73
5	0.69	0.66	0.67
6	0.73	0.66	0.70
accuracy	0.76		

f

Fig. 7. **Characteristics of the accuracy of facial emotions recognition in images of people using the method of pseudo-labeling data on the first (a, b), second (c, d) and third iteration of the semi-supervised learning (e, f)**

***Source:* compiled by the authors**

– It is shown that the use of the developed pseudo-labeling method data transfer training of the MobileNet V1 convolutional neural network model allowed to increase the accuracy of recognizing human emotions on images of the RAF-DB dataset by 2 percent (from 76% to 78%) according to the *F*1 score. At the same time, taking into account the significant imbalance of the classes of 7 basic emotions in the training sample, we have a significant increase in the accuracy of recognizing individual representatives of such emotions as *surprised* (from 71 to 77%), *fearful* (from 64 to 69%), *sad* (from 72 to 76%), *angry* from (from 64 to 74%), *neutral* (from 66 to 71%), the accuracy of recognizing the emotion of *happy*, which is the most common, decreased (from 91 to 86%).

–Pseudo-labeling method can be recommended for expanding the volume of training sets deep learning of convolutional neural networks, in order to overcome such shortcomings of datasets as lack of data of a certain class, imbalance of classes, insufficient amount of data for deep learning, etc.

REFERENCES

1. Cowie, R., Douglas-Cowie, E., Tsapatsoulis, N., Votsis, G., Kollias, S., Fellenz, W. & Taylor, J. G. “Emotion Recognition in Human-Computer Interaction.” *IEEE Signal Processing Magazine*. 2001; 18 (1): 32–80. DOI: https://doi.org/10.1109/79.911197.

2. Andalibi, Nazanin & Justin Buss. “The human in emotion recognition on social media: attitudes, outcomes, risks”. *In Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems.* Honolulu, USA. 2020. p. 1–16. DOI: https://doi.org/10.1145/3313831.3376680.

3. Li, W., Abtahi, F., Zhu, Z. & Yin, L. “Eac-net: Deep nets with enhancing and cropping for facial action unit detection”. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2018; 40 (11): 2583–2596. DOI: https://doi.org/10.1109/TPAMI.2018.2791608.

4. Lim, Y. K., Liao, Z., Petridis, S. & Pantic, M. “Transfer learning for action unit recognition”. *Humanoids 2017 IEEE RAS International Conference workshop Cooperative Autonomous Robot Experience (Presentation)*. 2018. DOI: https://doi.org/10.48550/arXiv.1807.07556.

5. Almaev, T., Martinez, B. & Valstar, M. “Learning to transfer: transferring latent task structures and its application to person-specific facial action unit detection”. *2015 IEEE International Conference on Computer Vision (ICCV)*. Santiago, Chile. 2015. p. 3774–3782. DOI: https://doi.org/10.1109/ICCV.2015.430.

6. Shao, Z., Liu, Z., Cai, J. & Ma, L. “JAA-Net: Joint facial action unit detection and face alignment via adaptive attention”. *International Journal of Computer Vision.* 2021; 129 (2): 321–340. DOI: https://doi.org/10.1007/s11263-020-01378-z.

7. “28 dataset results for Facial Expression Recognition (FER)”. – Available from: https://paperswithcode.com/datasets?task=facial-expression-recognition&page=1. – [Accessed: Jul. 2022].

8. Li, S. & Deng, W. “Deep facial expression recognition: A Survey”. *IEEE Transactions on Affective Computing*. 2018; 13 (3): 1195–1215. DOI: https://doi.org/10.1109/TAFFC.2020.2981446.

9. Samadiani, N., Huang, G., Cai, B., Luo, W., Chi, C., Xiang, Y. & He, J. “A review on automatic faci-al expression recognition systems assisted by multimodal sensor data”. *Sensors*. 2019; 19 (8): 1863. DOI: https://doi.org/10.3390/s19081863.

10. Howard, A., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M. & Adam, H. “Mobile nets: Efficient convolution neural networks for mobile vision applications”. *ArXiv*. 2017. DOI: https://doi.org/10.48550/arXiv.1704.04861.

11. Sandler, M., Howard, A., Zhu, M., Zhmoginov, A. & Chen, L.C. “Mobilenetv2: Inverted residuals and linear bottlenecks”. *IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Salt Lake City: USA. 2018. p. 4510–4520. DOI: https://doi.org/10.1109/CVPR.2018.00474.

12. Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M. & Hartwig, A. “MobileNets: Efficient convolutional neural networks for mobile vision applications”. *arXiv*. 2017. DOI: https://doi.org/10.48550/arXiv.1704.04861.

13. Gonzalez-Diaz, R., Gutiérrez-Naranjo, A. & Paluzo-Hidalgo, E. "Representative datasets: the perceptron case". 2019. – Available from: https://idus.us.es/bitstream/handle/11441/97961/Representative%20Datasets.pdf?sequence=1&isAllowed=y. – [Accessed: Jul. 2022].

14. Roh, Y., Heo, G. & Whang, S. E. "A survey on data collection for machine learning: a big data-ai integration perspective". *IEEE Transactions on Knowledge and Data Engineering*. 2021; 33 (4): 1328–1347. DOI: https://doi.org/10.1109/TKDE.2019.2946162.

15. Zhong, Z., Zheng, L., Kang, G., Li, S. & Yang, Y. "Random erasing data augmentation". *Proceedings of the AAAI Conference on Artificial Intelligence*. 2020; 34 (07): 13001–13008. DOI: https://doi.org/10.48550/arXiv.1708.04896.

16. Cubuk, E. D., Zoph, B., Mane, D., Vasudevan, V. & Le, Q. V. "Autoaugment: Learning augmentation strategies from data". *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2019. p. 113–123. DOI: https://doi.org/10.48550/arXiv.1805.09501.

17. Perez, L. & Wang, J. "The effectiveness of data augmentation in image classification using deep learning". *arXiv preprint arXiv:1712.04621*. 2017. DOI: https://doi.org/10.48550/arXiv.1712.04621.

18. Shorten, C. & Khoshgoftaar, T. M. "A survey on image data augmentation for deep learning". *Journal of Big Data*. 2019; 6 (1): 60. DOI: https://doi.org/10.1186/s40537-019-0197-0.

19. O'Gara, S. & McGuinness, K. "Comparing data augmentation strategies for deep image classification". *Irish Machine Vision and Image Processing Conference (IMVIP)*. 2019. DOI: https://doi.org/10.21427/148B-AR75.

20. Canavet, O. & Fleuret, F. "Efficient sample mining for object detection". *Proceedings of the Asian Conference on Machine Learning (ACML)*. 2014. p. 48–63.

21. Jin, S. Y., RoyChowdhury, A., Jiang, H., Singh, A., Prasad, A., Chakraborty, D. & Learned-Miller, E. "Unsupervised hard example mining from videos for improved object detection". *Proceedings of the European Conference on Computer Vision (ECCV)*. 2018. p. 316–333. DOI: https://doi.org/10.48550/arXiv.1808.04285.

22. Hao, H., Güera, D., Reibman, A. & Delp, E. "A Utility-Preserving GAN for face obscuration". *ICML 2019 Workshop on Synthetic Realities: Deep Learning for Detecting Audiovisual Fakes*. 2019. DOI: https://doi.org/10.48550/arXiv.1906.11979.

23. Meng, Y., Kong, D., Zhu, Z. & Zhao, Y. "From night to day: GANs based low quality image enhancement". *Neural Processing Letters*. 2019; 50 (1): 799–814. DOI: https://doi.org/10.1007/s11063-018-09968-2.

24. Ardiyanto, I., Soesanti, I. & Cahya Qairawan, D. "Night-today road scene translation using generative adversarial network with structural similarity loss for night driving safety". *In: Ahmed, K.R., Hassanien, A.E. (eds) Deep Learning and Big Data for Intelligent Transportation. Studies in Computational Intelligence*. 2021; 945: 119–133. DOI: https://doi.org/10.1007/978-3-030-65661-4_6.

25. Xie, H., Xiao, J., Lei, J., Xie, W. & Klette, R. "Image Scene Conversion Algorithm Based on Generative Adversarial Networks*". In: Cree, M., Huang, F., Yuan, J., Yan, W. (eds) Pattern Recognition. ACPR 2019. Communications in Computer and Information Science*. Springer, Singapore. 2019; 1180: 29–36. DOI: https://doi.org/10.1007/978-981-15-3651-9_4.

26.Srivastava, N., Hinton, G. E., Krizhevsky, A., Sutskever, I. & Salakhutdinov, R. R. "Dropout: A simple way to prevent neural networks from overfitting". *The Journal of Machine Learning Research*. 2014; 15 (1): 1929–1958.

27."CIFAR-10 Dataset". – Available from: https://www.cs.toronto.edu/~kriz/cifar.html. – [Accessed: Jul. 2022].

28."ImageNet Dataset". – Available from: http://www.image-net.org. – [Accessed: Jul. 2022].

29. Lee, D. H. "Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks". *Workshop on Challenges in Representation Learning, ICML*. 2013; 3 (2): 896.

30. Hendrycks, D., Mazeika, M., Kadavath, S. & Song, D. “Using self-supervised learning can improve model robustness and uncertainty”. *33rd Conference on Neural Information Processing Systems.* Vancouver, Canada. 2019. DOI: https://doi.org/10.48550/arXiv.1906.12340.

31. Kim, S. W., Lee, Y. G., Tama, B. A. & Lee, S. “Reliability-enhanced camera lens module classification using semi-supervised regression method”. *Applied Sciences*. 2020; 10: 3832. DOI: https://doi.org/10.3832.10.3390/app10113832.

32. Arsirii, O., Petrosiuk, D., Babilunha, O. & Nikolenko, A. “Method of transfer deap learning convolutional neural networks for automated recognition facial expression systems”. *Lecture Notes in Computational Intelligence and Decision Making. ISDMCI 2021. Lecture Notes on Data Engineering and Communications Technologies.* Springer, Cham. 2022; 77: 744–761. DOI: https://doi.org/10.1007/978-3-030-82014-5_51.

33. Petrosiuk, D., Arsirii, O., Babilunha, O. & Nikolenko, A. “Deep learning technology of convolutional neural networks for facial expression recognition”. *Applied Aspects of Information Technology*. 2021; 4 (2): 192–201. DOI: https://doi.org/10.15276/aait.02.2021.6.

34.“RAF-DB (Real-world Affective Faces)”. – Available from: https://paperswithcode.com/ dataset/raf-db. – [Accessed: Jul. 2022].

Conflicts of Interest: The authors declare no conflict of interest

Received 22.06.2023
Received after revision 05.09.2023
Accepted 15.09.2023

DOI: https://doi.org/10.15276/hait.06.2023.13
УДК 004.8

Псевдомаркування даних трансферного навчання згорткової нейронної мережі для розпізнавання емоцій на обличчі людини

Арсірій Олена Олександрівна[1)]
ORCID: https://orcid.org/0000-0001-8130-9613. e.arsiriy@gmail.com Scopus Author ID: 54419480900
Петросюк Денис Валерійович[1)]
ORCID: https://orcid.org/0000-0003-4644-3678; d.petrosyuk1994@gmail.com. Scopus Author ID: 54419479400
[1)] Національний університет «Одеська політехніка», пр. Шевченка, 1. Одеса, 65044, Україна

АНОТАЦІЯ

Показано актуальність розв'язання задачі розпізнавання емоцій на зображенні людини при створенні сучасних інтелектуальних систем комп'ютерного зору та людино-машинної взаємодії, онлайн-навчання та емоційного маркетингу, охорони здоров’я та криміналістики, машинної графіки та ігрового інтелекту. Показано вдалі приклади технологічних рішень задачі розпізнавання емоцій з використанням трансферного навчання глибоких згорткових нейронних мереж. Але використання таких популярних датасетів як DISFA, CelebA, AffectNet для глибокого навчання згорткових нейронних мереж не дає хороших результатів по точності розпізнавання емоцій тому, що майже всі навчальні вибірки мають принципові недоліки, пов’язані з похибками при їх створенні такими, як відсутність даних певного виду, незбалансованість класів, суб’єктивність та багатозначність маркування, недостатній для глибинного навчання об’єм даних, тощо. Запропоновано зазначені недоліки популярних датасетів для розпізнавання емоцій долати за рахунок додавання у навчальну вибірку додаткових псевдо-маркованих зображень з емоціями людини, на яких розпізнавання відбувається з високою точністю. Метою роботи є підвищення точності розпізнавання емоцій на зображенні обличчя людини за рахунок розробки методу псевдо-маркування для трансферного навчання глибокої нейронної мережі. Для досягнення мети вирішено такі завдання: скориговано на наборі даних RAF−DB для вирішення завдань розпізнавання емоцій модель згорткової нейронної мережі, попередньо навчену на наборі ImageNet за допомогою методу трансферного навчання; розроблено метод псевдо-маркування даних набору RAF−DB для полу-контрольованого навчання моделі згорткової нейронної мережі для задачі розпізнавання емоцій на зображенні людини; проаналізовано точність розпізнавання емоцій на зображенні людини на основі розроблених моделі згорткової нейронної мережі та методу псевдо-маркування даних набору RAF-DB для її коригування. Показано, що використання розробленого метода псевдо-маркування даних трансферного навчання моделі згорткової нейронної мережі MobileNet V1 дозволило підвищити точність розпізнавання емоцій людини на зображеннях набору даних RAF-DB на 2 відсотка (з 76% до 78%) за оцінкою F1. При цьому, враховуючи суттєву незбалансованість класів 7 основних емоцій в тренувальній виборці маємо суттєве збільшення точності розпізнавання нечисленних представників таких емоцій як *здивованості* (з 71 до 77 %), *страху* (з 64 до 69 %), *суму*(з 72 до 76 %), *злості* з (з 64 до 74 %), *нейтральності* (з 66 до 71%), точність розпізнавання емоції *щастя*, що є найбільш поширеною знизилась (з 91 до 86 %) Таким чином, можна зробити висновок, що використання розробленого метода псевдо-маркування дає гарні результати в подоланні таких недоліків датасетів для глибинного навчання згорткових нейронних мереж як відсутність даних певного виду, незбалансованість класів, недостатній для глибинного навчання об'єм даних, тощо.

Ключові слова: псевдо-маркування даних; полу-контрольоване навчання; трансферне навчання; згорткові нейроні мережі; розпізнавання емоцій на обличчі людини

ABOUT THE AUTHORS

Olena O. Arsirii - Doctor of Engineering Sciences, Professor, Head of the Department of Information Systems, Odessa Polytechnic National University, 1, Shevchenko Ave. Odessa, 65044, Ukraine

ORCID:https://orcid.org/0000-0001-8130-9613; e.arsiriy@gmail.com. Scopus Author ID: 54419480900
Research field: Information technology; artificial intelligence; decision support systems; machine learning; neural networks

Арсірій Олена Олександрівна - доктор технічних наук, професор, завідувач кафедри Інформаційних систем. Національний університет «Одеська політехніка», пр. Шевченка, 1. Одеса, 65044, Україна

Denys V. Petrosiuk - PhD Student of the Department of Information Systems. Odessa Polytechnic National University, 1, Shevchenko Ave. Odessa, 65044, Ukraine
ORCID: https://orcid.org/0000-0003-4644-3678; d.petrosyuk1994@gmail.com. Scopus Author ID: 54419479400
***Research field*:** Convolution Neural Networks; Facial Expression Recognition

Петросюк Денис Валерійович - аспірант кафедри Інформаційних систем. Національний університет «Одеська політехніка», пр. Шевченка, 1. Одеса, 65044, Україна