Applied Mathematics and Information Sciences
An International Journal

http://dx.doi.org/10.18576/amis/170220

Fast Facial Expression Recognition System: Selection of Models

L. Atymtayeva [1,*], *M. Kanatov*[2], *A. M. A Musleh*[3], *and G. Tulemissova*[1]

[1]Faculty of Engineering and Natural Sciences at Suleyman Demirel University, Kaskelen, Kazakhstan
[2]Faculty of Automatic and Information Technology at Satbayev University, Almaty, Kazakhstan
[3]College of Business and Finance, Ahlia University, Manama, Kingdom of Bahrain

Received: 22 Nov. 2022, Revised: 22 Dec. 2023, Accepted: 16 Feb. 2023
Published online: 1 Mar. 2023

Abstract: Facial Expression Recognition (FER) is rapidly developing field of Computer Vision and Pattern Recognition directions. FER can be helpful for various purposes: in security systems for aggression recognition, in education for students interests recognition, in marketing for customer satisfaction and in the many other fields. Usually we can distinguish seven common facial expressions for all persons. However, it is often important to know: whether a person is positive or negative. This paper describes the recognition system for facial expression in real time, which defines relatively fast and accurate the positive or negative emotion of the faces in the camera view and selection of the architecture of Deep CNN. This effect is a result of the combination of facial detection algorithms and classification algorithms based on the convolutional neural networks. We have compared different datasets (FER2013 and AffectNet) and provided the experiment results in different cases of the FER models and classes, convolutional layers, and filters. We have found that for 8 classes of FER expressions the architecture model M2 is the best model. It has the best accuracy and works about 2 times faster on GPU and 3 times faster on CPU than M1 model. It has been also found out that training on AffectNet dataset is significantly better than training on FER2013 dataset due to the differences in number of samples in the given datasets.

Keywords: deep convolutional neural networks (Deep CNN), deep learning, computer vision, facial expression recognition (FER), pattern recognition

1 Introduction

Emotions are very important in the human interactions. Emotions define the essence of the person. People reveal their emotional state through the signals to others and perceive others emotional state through the signal as well as during communication. Some emotions enrich the meaning of human communication, others affect to the favorable or adverse actions towards others, therefore emotions become a decision maker instruments. Correct recognition of human emotions allows you to predict what intentions a person has. Thus, emotion recognition systems are a crucial condition when implementing future information systems and information technologies, for the purpose of providing human-centered computer interfaces that have the ability to react to the users' emotional states.

Human-being recognize others emotions based on the context. For example, a smile depending on the context can be interpreted as a signal of politeness, irony, joy, or greeting. Humans express their emotions through various ways such as facial expressions, speech, gestures and textual emoticons.

Facial Expression Recognition is more clear tools for defining of emotions. Investigations of facial expressions have been launched many years ago [1], [2]. In 1971 Ekman and Friesen have proposed six main universal emotions [3] that can be common for all nations and cultures [4]. They have established the following facial expressions: disgust, fear, happiness, surprise, sadness and anger. Using a discrete concept of emotions, Plutchik proposed his own model [5], which consists of eight basic emotions: joy, trust, surprise, anticipation, sadness, fear, anger and disgust. However, he adopted a new approach, where emotions can be mixed. He proposed that eight basic emotions could be mixed into twenty four new emotions.

Later Russel [6] proposed a new dimensional model, where emotions can be mapped into a two-dimensional space. This model consists of two features. The first

* Corresponding author e-mail: lyazzat.atymtayeva@sdu.edu.kz

feature is an arousal which defines how sleepy or alert the emotion is. The second one is a valence which defines if the emotion is negative or positive.

Schematically the interrelationships between these two features can be shown in the Figure 1.

Fig. 1: Samples from CK+ dataset

Currently, we can distinguish seven types of facial expressions that are common for all nations and cultures and may be used in the facial recognition researches: disgust, fear, happiness, surprise, sadness, anger and neutral [4] , [5].

Starting from the 21st century, Computer Vision investigations are developing very fast. In general, there are a lot of different recognition systems. One of them is Automated Facial Expression Recognition (AFER). AFER as an area of computer vision has appeared only the last two decades. First researches have been concerned the analysis of facial expressions [7] later the first attempts to create AFER were conducted [8].

Fast developing of AFER couldn't be possible without the boost of artificial intelligence (AI) technologies. Among the best tools of AI in the area of Image Recognition is Deep Convolutional Neural Networks (Deep ConvNets or Deep CNN) which uses the main principles and algorithms of Deep Learning and Artificial Neural Networks. The algorithms and tools of Deep CNN was proposed by Yann Lecun [9] and J.Redmon et al. [10] (You Only Look Once - YOLO). First great results in the using of DeepCNN in Image Recognition was done by Alex Krizhevsky [11] which won ILSVRC-2012 competition using CNNs for image recognition task and applying ImageNet dataset [12] with a large margin of results.

Thus, any architecture of Artificial Neural Networks [13] can be created for special features, which are extracting from the image. Extraction is based on some methods and algorithms, or can consists of preprocessing procedures. Also, vectorized pixels of the image can be used as a features for CNNs. Features must be extracted the same for training and recognizing. CNNs represent a structure, that is similar to the visual perception of the human brain therefore the recognition procedure starts from feature extraction from the visual objects. Then, data follows to fully connected Artificial Neural Networks for classification. Using different layers (convolution layers, subsampling layers and neuron layers) CNNs can manipulate the data. The convolutional and subsampling layers make feature extraction tasks. The neuron layer is an output of all CNNs architecture and it works for classification task.

Actually we can produce the image (face) recognition in any time by using the standard tools for image making and recognition. However, it is crucial to make the process of facial recognition more quickly and accurate. Also, it is important to mean that the equipment for getting the images usually has the standard characteristics.

In this paper we propose the method that recognizes positive and negative facial expressions using Deep Convolutional Neural Networks. We study the algorithms and train the datasets for the acceleration of the processes of image (face) recognition with high enough accurate by using the principles of AFER and applying different architecture of Deep CNNs . Also, we use only standard video surveillance cameras for different purposes in the public places.

The main idea of this paper is revealing by the next sections which are devoted to the description of datasets, used models and methods, and demonstration of the results of experiments with datasets and their training.

2 Datasets

In our combined researches we used the different datasets. For example, we applied PASCAL-Visual Object Classes 2012 (PASCAL-VOC 2012) dataset [14] for the task of the person detection [15]. This dataset consists of 4 087 RGB images with person in the different situations, and 10 129 labels of person. The main principle of the person detection task implementation is the following. For each images authors of the dataset prepare .txt file with coordinates of persons in the images. Some txt files can have more than one labels. It depends how many persons in the image.

In comparing with Viola-Jones' Pedestrian detection method [16] we used Deep CNNs and could find the following advantages. Viola-Jones' method detects only full body of the person, while Deep CNNs could define person from the any angle, even if not the full part of the person is shown on the image.

In our facial expression recognition (FER) researches we distinguish the datasets that can be divided into several categories and groups. For instance, three groups of datasets such as [16]:

–RGB and GrayScaled (JAFFE, CK, CK+, Oulus-CASIA, etc.);
–3D(BU-3DFE) ;
–Thermal (Oulus-CASIA).

The first group (RGB and GrayScaled dataset) usually consists of the posed or natural samples. The posed samples are usually collected in the laboratory by using special camera position and distances for good images of facial emotions. Normally, in posed datasets images have a good illumination and resolution. In the datasets with natural samples, faces can be of any resolution, different illumination and at any angle relative to the camera. These samples are regularly collected from movies.

In our work [17] we used Extended Cohn-Kanade (CK+) and Karolinska Directed Emotional Faces (KDEF) datasets for the improving of learning process and implementing the preprocessing methods.

There are also sequential datasets that are made from the video frames, where each emotion is demonstrated as a short video of facial expression changing from neutral to some emotion. These datasets are widely used in Recurrent Convolutional Neural Networks (RCNN) [9]. However, using RCNN is computationally expensive and not always suitable for the fast implementation.

This paper contains analysis of two FER datasets: FER-2013 [18] and AffectNet [19]. FER-2013 was proposed in the ICML 2013 Challenges. It contains 35887 images, with 4953 images of "Anger" faces, 547 images of "Disgust" faces, 5121 images of "Fear" faces, 8989 images of "Happiness" faces, 6077 images of "Sadness" faces, 4002 images of "Surprise" faces and 6198 images of "Neutral" faces. All images are collected by using Google Search API and special keywords like "blissful", "enraged" etc. All samples are grayscale and have size 48x48 pixels. They are illustrated in the Figure 2.

Fig. 2: Samples of FER2013 dataset

Dataset AffectNet is the largest dataset for facial expression recognition systems. This dataset consist of one million images with facial landmarks and 450 000 manually annotated images. Images of AffectNet are collected from the internet by querying emotion related keywords from Google, Bing and Yahoo. Used keywords for searching facial emotion images were combined with special words that could getting the strings as *joyful girl*, *furious young lady* and etc. Then, these strings have been translated into other five languages: Farsi, German, Portuguese, Spanish and Arabic. Faces from all collected images can be detected by OpenCV Face Recognition system. Average Resolution of the images is 425x425 pixels.

AffectNet dataset has eight facial emotion categories. They are Neutral, Happy, Sad, Surprise, Fear, Anger, Disgust and Contempt. Also, it has None, Uncertain and Non-face categories of images, which could not be put into the above groups. In the given work, we consider only first eight groups with manually annotated normal facial expression images. As it was described before, AffectNet is very large dataset that allows us to provide more accurate analysis and dataset training. The total number is about million images. The number of manually annotated images by special labelers are 450 000, where number of neutral images is 80 276; happy ones is 146 198; sad ones is 29 487; surprise– 16 288; fear– 8 191; disgust– 5 264; anger– 28 130; contempt– 5 135; none – 35 322; uncertain – 13 163; none-face – 88 895. Figure 3 shows some samples of manually annotated images of AffectNet dataset.

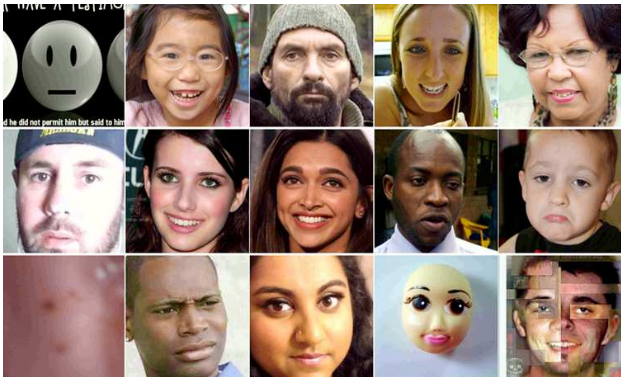

Fig. 3: Samples of AffectNet dataset

3 Convolutional Neural Network Model

The Architecture of Convolutional Neural Network (CNN) , that we use in our work (see Figure 4), consists of nine two-dimensional convolutional layers, three maxpooling layers, four fully connected layers and Softmax [19] activation layer. For convolution operation we applied 3x3 kernel with 'same' padding and 'ReLU'[17] activation after each convolutional layer. After every three convolutional layers we used two dimensional maxpooling layer with 2x2 kernel. It changes input image height and width and makes them two times less. For training the model we used FER2013 [18] and AffectNet [19] datasets with several different input images shapes. The sizes of them are 48x48,96x96 and 128x128.

In the following subsections we describe the results of experiments with the datasets containing the images of facial expressions after using CNN model.

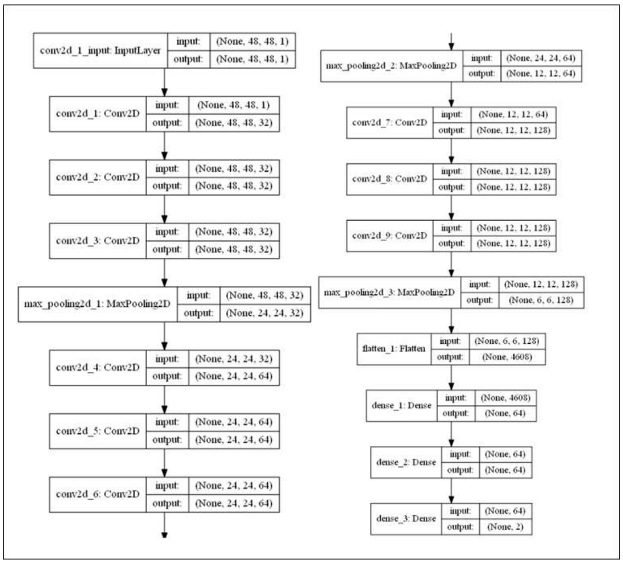

Fig. 4: Used Convolutional Neural Networks Architecture

3.1 Experiment 1

In the first experiment we trained our CNN model using FER2013 dataset. Based on one-vs-all [20] [21] classification method, FER dataset was divided into two classes – positive - “happy” and negative - “others”. For the first class we used all 8 989 Happy images from FER dataset. As negative, we randomly have chosen the same number of images from other classes. Totally, created from FER one-vs-all dataset consists of 17 978 images, which was divided into training set (14 382), validation set (1 797) and test set (1 797). Training process is shown on the Figure 5. Minimal validation error is 0.33 and the best validation accuracy is 87.70 %.

Some different researches have been conducted with FER dataset and the authors have got different level of accuracy. They claimed that with ConvNet can be achieved the validation accuracy of 91,01% [22]. Other good results have been received with KDEF and JAFFE facial image datasets [23]. With DenseNet-161 [23] the researchers achieved 96,51% (on KDEF) and 99,52% (on JAFFE) respectively.

The same experiment with mirroring data augmentation operation increased the best validation accuracy up to 88.32% and decreased validation error to 0.29. Confusion matrix of testing results is illustrated on the Figure 6a. Also, we tested this trained model using AffectNet Validation set (see Figure 6b).

We resized all images to resolution 48x48 pixels and made them GrayScaled. Then, we have chosen all 462 Happy images from validation set and randomly selected the same number of negative images from other classes. Totally, we have collected 924 images. Figure 6b demonstrates the testing results on AffectNet validation set.

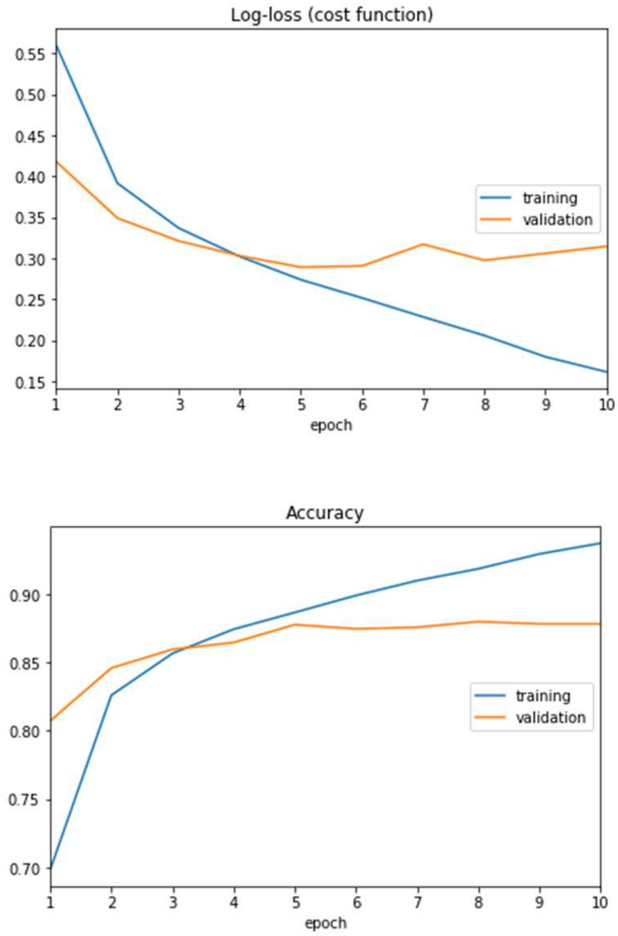

Fig. 5: Training process: dataset FER-2013, input shape – 48x48x1

3.2 Experiment 2

In the second experiment we trained our CNN model using AffectNet dataset. We used only manually annotated part of AffectNet. Since annotated images were filtered by special labelers (actually by humans), the discovered mistakes are much smaller in this part. Even in this case, training set is large enough and it has a sufficient number of images. Exactly, it has 414 799 images (neutral – 74 814, happy – 134 415, sad – 25 459, surprise – 14 090, fear – 6 378, anger – 3 803, disgust – 24 882, contempt – 3 750, none – 33 088, uncertain – 11 645, non-face – 82 415). Validation set has 5500 images (500 images for each class). Using DLIB machine learning toolkit [24] we filtered images from the first eight classes where faces were detected. Then, all filtered images were resized to 48x48 pixels and GrayScale. The same operation was done for validation set. As a result, we’ve got 268 987 grayscale images of training set and 3 716 images of validation set. The number of Happy images of training set is 126 384. That is almost half of

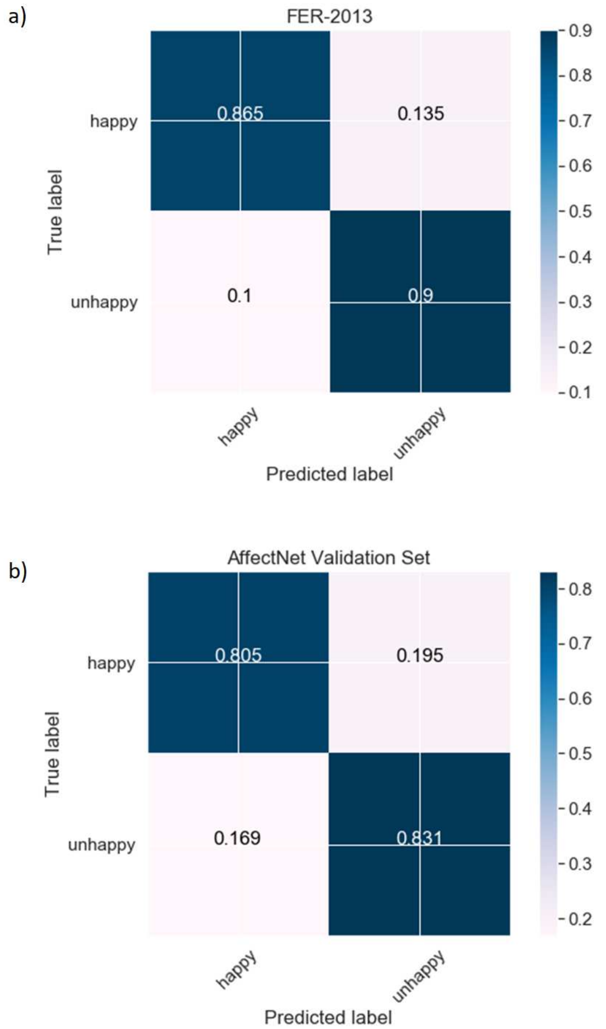

Fig. 6: First experiment confusion matrixes (a) - for FER-2013 test set, (b) – for AffectNet validation set

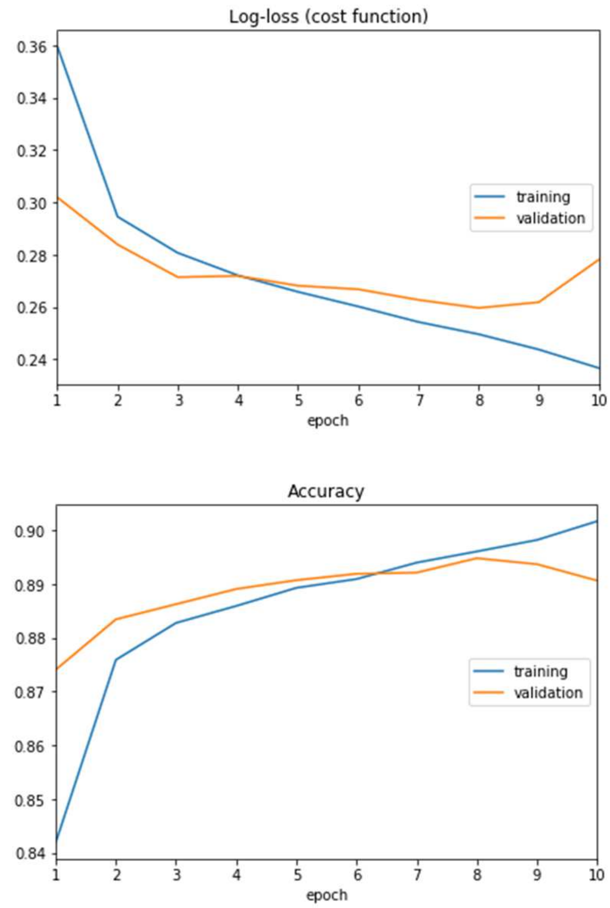

Fig. 7: Training process: dataset - AffectNet, input shape – 48x48x1

the total number. That is why, we divided training set into two classes, where positives are all happy images and negatives are all other images. The number of negatives is 142 603. This training set was used as total dataset and after shuffling was divided into three parts: training set – 90% (242 088 images), validation set – 5% (13 449 images) and test set – 5% (13 450 images).

Figure 7 demonstrates learning process on the training set and validation accuracy and error after every epoch. The best value of accuracy on validation set during the training is 89,5%, the minimum value of error is 0.260. Testing on test set, which created from the main training set, shows 88,8% of accuracy and 0.277 value of error. Testing on test set of FER-2013 shows 84,7% of accuracy and 0.386 value of error.

Results are illustrated on the Figure 8(a-b). Later, this model was tested on all FER-2013 dataset and main validation part of AffectNet dataset, which was used in the section 3.1. Both of results are shown on the Figure 9(a-b).

3.3 Experiment 3

The previous two experiments explain training on FER-2013 and AffectNet datasets with the same Deep CNN Architecture. In the third experiment we have tested DeepCNN architecture with different hidden layers and have analyzed how they influence to model accuracy and computation time on GPU and CPU. All models were trained and tested using AffectNet Training set as it was shown in the section 3.2 (Experiment 2). Table 1 shows the training results of fifteen different models. This table consists of 8 columns. First column is the DeepCNN Model Name with the information about the convolutional layers and fully connected layers model. The second column shows how many classes of FER expressions of the AffectNet datasets we have trained. In

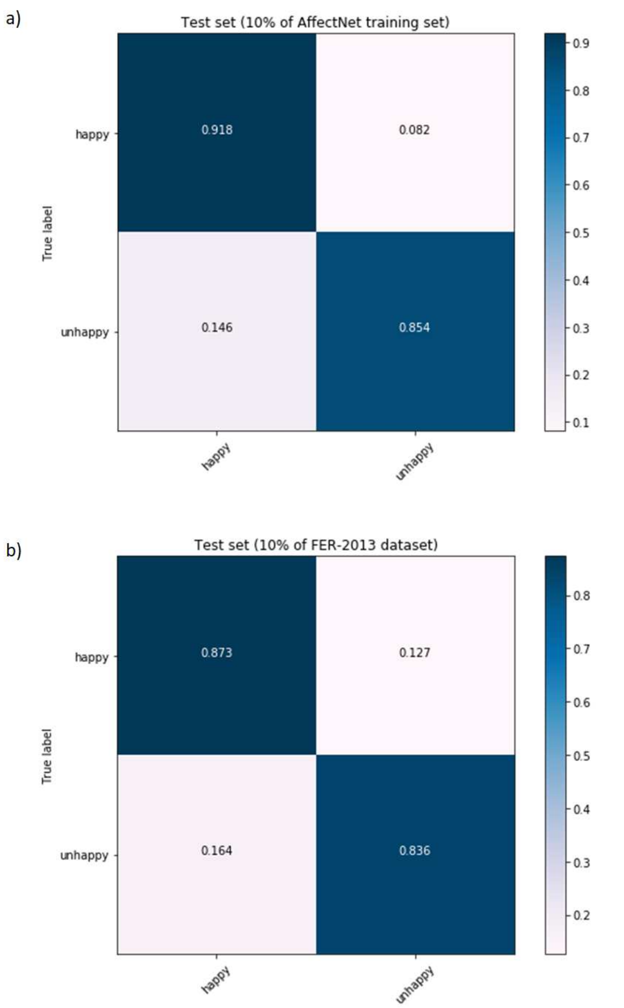

Fig. 8: Confusion matrixes for model trained on AffectNet dataset, (a) – testing results of using test set created from 10% of AffectNet training set; (b) – testing results of using test set of FER-2013 used in the section 3.1.

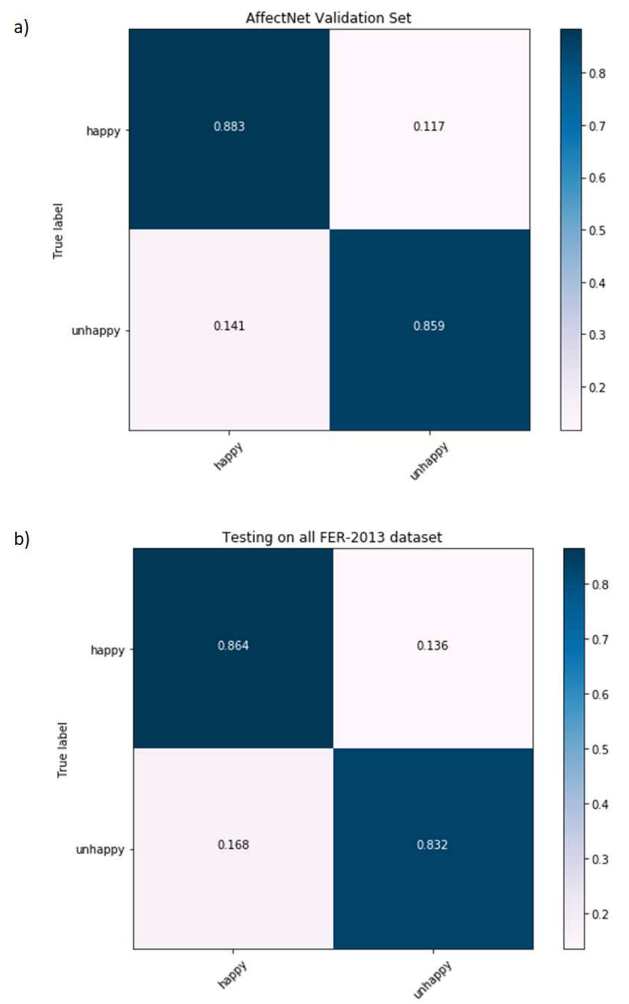

Fig. 9: Confusion matrixes for model trained on AffectNet dataset, (a) – testing results of using AffectNet validation set, used also in the section 3.1; (b) – testing results of using all FER-2013 dataset.

our experiments we used 2 and 8 classes of FER expressions. The third column contains the information how many parameters model has. The fourth and fifth columns are testing results , i.e. accuracy and error. Sixth and sevens columns have a computation time of 1000 samples on GPU and CPU. For GPU computation we used Nvidia Quadro M2000 which has a 768 cuda cores. For CPU computation we selected CPU Quad-Core Intel Core i7 of MacBook Pro Retina 15 (2013 early). Every computation was repeated 10 times and the best results you can see in the Table 1. Here the last (eight) column shows the size of model. Models from M1 to M5 have nine convolutional layers with different numbers of filters. Also, the models contain three MaxPooling layers with 2x2 pooling size that downscale input images. The largest model M1 has 2 513 026 parameters and shows the best accuracy (88,74%) in the table for two classes of FER expressions. However, we can also notice that it takes more time for computation.

We also can notice that difference in accuracy of M1 and M5 for the two classes of FER is 0.20 while the difference of errors is 0.008. In general, the difference between the models M1 and M5 is very small, however, M5 works about 2 times faster on GPU, and 11 times faster on CPU than M1. In contrast, training with eight classes shows the different values in compare with two classes of the model. Differences between M1 and M2 models for eight classes of FER in the Table 1 shows that accuracy is reduced from 72,18% to 63,48% while the error is increased from 0.818 to 1.085. The speed is about the same for both cases of FER classes. Thus, for two

Table 1: Training results of 15 models with different convolutional layers and filters for two and eight FER Class

Models-CLayers-FullCLayers	FERClass	Param-s	AccTest	TestError	GPURun	CPURun	ModelSizeMB
M1-9(3x64,3x128,3x256)/2x64	2	2513026	88.74	0.286	4.23	17.17	30.239
	8	2513416	72.18	0.818	4.24	17.32	30.243
M2-9(3x32,3x64,3x128)/2x64	2	779458	88.54	0.278	2.51	5.7	9.436
	8	779848	72.24	0.819	2.55	5.8	9.442
M3-9(3x16,3x32,3x64)/2x32	2	195170	87.9	0.291	2.39	2.59	2.376
	8	195368	70.85	0.84	2.5	2.75	2.427
M4-9(3x8,3x16,3x32)/2x16	2	48946	86.54	0.321	2.3	1.66	2x16
	8	49048	66.7	0.975	2.4	1.69	0.316
M5-9(3x4,3x8,3x16)/2x8	2	12314	87.57	0.294	2.59	1.5	0.168
	8	12368	63.48	1.085	2.41	1.5	0.168
M6-9(2x64,2x128,2x256)/2x64	2	1738434	88.42	0.322	3.07	10.33	20.92
	8	1738824	70.92	0.91	3.14	11	20.926
M7-9(3x32,2x64,2x128)/2x64	2	585698	87.81	0.31	2.22	3.63	2.452
	8	586088	70.24	0.883	2.48	3.7	2.455
M8-9(2x16,2x32,2x64)/2x32	2	146674	87.66	0.298	2.11	1.95	0.696
	8	146872	69.95	0.889	2.3	1.89	0.698
M9-9(2x8,2x16,2x32)/2x16	2	36794	86.67	0.317	2.17	1.36	0.257
	8	36896	65.88	1.005	2.05	1.43	0.259
M10-9(2x4,2x8,2x16)/2x8	2	9262	84.07	0.361	2.03	1.19	0.146
	8	9316	62.03	1.129	2.17	1.2	0.149
M11-9(1x64,1x128,1x256)/2x64	2	963842	87.87	0.316	2.15	3.78	11.612
	8	964232	68.71	0.913	2.44	3.91	11.617
M12-9(1x32,1x64,1x128)/2x64	2	391938	87.66	0.298	1.86	1.54	4.749
	8	392328	68.83	0.922	2.08	1.59	4.753
M13-9(1x16,1x32,1x64)/2x32	2	98178	85.88	0.336	1.85	1.08	0.494
	8	98376	66.91	0.966	1.93	1.07	0.496
M14-9(1x8,1x16,1x32)/2x16	2	24642	84.05	0.363	1.75	0.93	0.2
	8	24744	62.86	1.091	1.75	0.94	0.202
M15-9(1x4,1x8,1x16)/2x8	2	6210	78.69	0.475	1.7	0.9	0.126
	8	6264	61.23	1.175	1.84	0.89	0.126

classes of FER the reducing of the number of training parameters, the number of convolutional layers, filter or the number of fully connected layers may lead to little loosing of accuracy. However, for eight classes FER models we cannot reduce many parameters for speeding up. The above treatment can be used in different fields, such as quantum information processing [25]-[28].

4 Conclusion

This work is dedicated to find the best model for recognizing Happy Facial Expression that is computationally cheap and can work in the real-time systems. For these purposes we conducted the series of experiments with various datasets and numbers of models. In the Experiment 1 (3.1) and Experiment 2 (3.2) we found out that training on AffectNet dataset is better than training on FER2013 datasets. These two datasets differ by number of samples. AffectNet is much larger, therefore, even after filtering AffectNet training set has much more samples than FER2013. Figure 5 and Figure 7 illustrate that model trained on AffectNet dataset recognizes Happy expressions better.

In the Experiment 2 (3.2) and Experiment 3 (3.3) we trained 15 different DeepCNN architec-tures with different convolutional layers, filters and different number of parameters. We have analyzed all proposed models, and got results that have been cumulated in the Table 1. For small tasks such as “Happy vs All” we can reduce a large number of parameters and get more fast model with little loosing of accuracy. But, in other cases that use eight classes of Facial Expression Recognition unfortunately we cannot reduce a large number of parameters. Extended tasks need more complex architecture and more trainable parameters. As we have found, for eight classes architecture M2 is the best model. Because it has the best accuracy and works about twice times faster on GPU and three times faster on CPU than M1.

5 Acknowledgement

This research was supported by research community of Suleyman Demirel University (SDU). We thank our colleagues from SDU (Kazakhstan) who provided insight and expertise that greatly assisted the research. We express our hopes that they will agree with the conclusions and findings of this paper.

References

[1] G.-B. D. de Boulogne and R. A. Cuthbertson, *The Mechanism of Human Facial Expression,* Cambridge University Press, UK (1990).

[2] C. Darwin, *The expression of emotion in man and animals,* Oxford University Press, London UK (1872).

[3] P. Ekman and W. V. Friesen, *Constants across cultures in the face and emotion,* Pers. Soc. Psychol, vol. 17., no. 2, 124-129 (1971).

[4] P. Ekman, *An argument for basic emotions,* Cognition and Emotion, vol. 6., no. 3, 169-200 (1992).

[5] R. Plutchik, *Nature of emotions,* Am. Sci., vol. 89, no. 4, 344-350 (2002).

[6] J. A. Russell, *A circumplex model of affect,* J. Pers. Soc. Psychol. vol. 39, no. 6, 161-117 (1980).

[7] K. Mase, *Recognition of facial expression from optical flow,* IEICE transactions-E., vol. 74, no. 1, 3474-3483 (1991).

[8] Graves, Alex, et al., *Facial expression recognition with recurrent neural networks,* Proc. IWCTS’08 (2008).

[9] LeCun, Y., and Bengio, Y., *Convolutional networks for images, speech, and time series,* The handbook of brain theory and neural networks, vol. 10, no. 3361 (1995).

[10] Redmon, J., and Farhadi, A., *YOLO9000: better, faster, stronger,* arXiv preprint (2017).

[11] Krizhevsky, A., Sutskever, I., and Hinton, G. E., *Imagenet classification with deep convolutional neural networks,* In Advances in neural information processing systems, 1097-1105 (2012).

[12] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei- Fei., *ImageNet: A Large-Scale Hi-erarchical Image Database,* In CVPR09 (2009).

[13] Haykin, S., *Neural Networks and Learning Machines,* Prentice Hall, 3 edition (2008).

[14] Everingham, M., Eslami, S. A., Van Gool, L., Williams, C. K., Winn, J., and Zisserman, A., *The pascal visual object classes challenge: A retrospective,* International journal of computer vision, vol.1, no. 111, 98-136 (2015).

[15] Rifkin, Ryan, and Aldebaro Klautau, *In defense of one-vs-all classification,* Journal of machine learning research, 101-141 (2004).

[16] M.Kanatov, L.Atymtayeva, *Deep Convolutional Neural Network based Person Detection and People Counting System,* Int.J. AETA, vol.7, no.3, 21-25 (2018).

[17] M.Kanatov, L.Atymtayeva and M.Mendes, *Improved Facial Expression Recognition with Xception,* Int. J. Appl. Math. Inf. Sci., vol. 13, no. 5, 859-865 (2019).

[18] D. E. P. L. C. a. I. J. Goodfellow, *Challenges in representation learning: A report on three machine learning contests,* International Conference on Neural Information Processing, 117-124 (2013).

[19] Mollahosseini, A., Hasani, B., and Mahoor, M. H., *Affectnet: A database for facial expression, valence,and arousal computing in the wild,* arXiv preprint arXiv:1708.03985 (2017).

[20] Peter Burkert, Felix Trier, Muhammad Zeshan Afzal, Andreas Dengel and Marcus Liwicki, *DeXpression: Deep Convolutional Neural Network for Expression Recognition,* arXiv: 1509.05371v2, (2016).

[21] Rifkin, Ryan, and Aldebaro Klautau, *In defense of one-vs-all classification,* Journal of machine learning research, 101-141 (2004).

[22] Tanoy Debnath, Md. Mahfuz Reza, Anichur Rahman, Amin Beheshti, Shahab S. Band & HamidAlinejad Rokny, *Four layer ConvNet to facial emotion recognition with minimal epochs and the significance of data diversity,* Scientific Reports 12, 6991 (2022).

[23] M. A. H. Akhand , Shuvendu Roy, Nazmul Siddique, Md Abdus Samad Kamal, and Tetsuya Shimamura, *Facial Emotion Recognition Using Transfer Learning in the Deep CNN,* Electronics 2021, vol. 9, no. 10, 1036 (2021).

[24] King D. E. Dlib-ml, *A machine learning toolkit,* Journal of Machine Learning Research, vol. 10, no. 5, 1755-1758 (2009).

[25] A.-S.F. Obada, M. Abdel-Aty, *Influence of the stark shift and kerr-like medium on the evolution of field entropy and entanglement in two-photon processes*, Acta Physica Polonica B, 31(3), 589–599 (2000).

[26] A.-H. Abdel-Aty, H. Kadry, M. Zidan, E. A. Zanaty, M. Abdel-Aty, A quantum classification algorithm for classification incomplete patterns based on entanglement measure, Journal of Intelligent and Fuzzy Systems, 38(3), 2817–2822 (2020) DOI: 10.3233/JIFS-179566

[27] M. Zidan, A.-H. Abdel-Aty, A. Younes, I. El-khayat, M. Abdel-Aty, A novel algorithm based on entanglement measurement for improving speed of quantum algorithms,

Applied Mathematics and Information Sciences, 2018, 12(1), 265–269 (2018) DOI: 10.18576/amis/120127

[28] M. Abdel-Aty, S. Furuichi, A.-S.F. Obada, Entanglement degree of a nonlinear multiphoton Jaynes-Cummings model Journal of Optics B: Quantum and Semiclassical Optics, 4(1), 37–43 (2002) DOI: 10.1088/1464-4266/4/1/306

Lyazzat Atymtayeva received the Ph.D and Doctor of Science degree in Mechanics, Mathematics and Computer Science at Al-Farabi National University, Kazakhstan. Now she is working as associate professor in Information Systems at Suleyman Demirel University. Her research interests are in the areas of mechanics, applied mathematics and computer science including the numerical and rigorous mathematical methods and models for mechanical engineering and computer science, intelligent and expert systems in Information Security, Artificial Intelligence and Machine Learning, Project Management and Human-Computer Interaction. She has published research papers in reputed international journals of mathematical and computer sciences. She is a reviewer and an editor of international journals in mathematics and information sciences.

Maksat Kanatov master of Computer science, researcher at Satbayev University, Kazakhstan. His research interest includes Machine Learning, Computer Vision, Deep Convolutional and Recurrent Neural Networks, Automated Facial Expression Recognition systems using Deep CNNs.

Gulfarida Tulemissova received the Ph.D (Candidate of Technical Sciences) in Mathematical and Computer Modeling at the Research Institute of the Information Technology, Kyrgystan. Now she is working as associate professor in Computer Science at Suleyman Demirel University. Her research interests are in the areas of Internet of Things, FPGA, Embedded systems, Artificial Intelligence, Multiservice Networks, Software Defined Networks, Knowledge management, Intellectual capital, Information theory. She has published research papers in reputed international journals of mathematical and computer sciences. She is reviewer and editor of international journals in computer networks and information sciences.

Abdalmuttaleb M.A. Musleh Al-Sartawi (Ph.D.) is the Chairperson of the Accounting, Finance & Banking Department at Ahlia University, the Kingdom of Bahrain. He is ranked by the Stanford University as one of the World's Top 2% Scientists in 2022. Moreover, he is the Editor-in- Chief of the International Journal of Electronic Banking (IJEBank), Inderscience and the Associate Editor of the Journal of Sustainable Finance & Investment, Taylor & Francis. Al-Sartawi received his PhD in Accounting from the University of Banking and Financial Science (UBFS). He has presented and published many papers in regional and international conferences and journals. In addition, Al-Sartawi is the chair and the editor of two international conferences: the European, Asian, Middle Eastern, North African Conference on Management & Information Systems (EAMMIS) and the International Conference on Global Economic Revolutions (ICGER). He is a member of various editorial boards and technical committees in international refereed journals and conferences. Al-Sartawi is a member of several international organizations and associations such as the European Accountants Association (EAA), Bahrain Economists Society, the Middle East Economic Association (MEEA), the International Islamic Marketing Association (IIMA), the Arab Academy for Banking and Financial Sciences, and the Palestinian Accounting Association. Besides his experience in teaching, he is vastly experienced in local and international quality standards and accreditation requirements. He is responsible for developing and implementing the Bahrain Quality Assurance (BQA) and the National Qualifications Framework (NQF) systems at his department. In terms of international accreditation, he has experience in preparing reports as per AACSB standards, related to Assurance of Learning and intellectual contribution standards. Al-Sartawi has worked closely with a number of international professional bodies in mapping his department's programs with professional certifications, resulting in several exemptions and accreditations such as ACCA, CFA, and CIMA.