The Usage of Grayscale or Color Images for Facial Expression Recognition with Deep Neural Networks

Dmitry A. Yudin1[0000-0002-1407-2633]\*, [Alexandr V. Dolzhenko](https://www.researchgate.net/scientific-contributions/2136080001_Aleksandr_Dolzhenko)2 [0000-0002-4426-5209] and Ekaterina O. Kapustina2[0000-0003-3831-1683]

1 Moscow Institute of Physics and Technology (National Research University),

Moscow Region, Dolgoprudny, Institutsky Per. 9, 141700, Russia

2 Belgorod State Technological University named after V.G. Shukhov, Belgorod, Kostukova Str. 46, 308012, Russia

\*yudin.da@mipt.ru

**Abstract.** The paper describes usage of modern deep neural network architectures such as ResNet, DenseNet and Xception for the classification of facial expressions on color and grayscale images. Each image may contain one of eight facial expression categories: “Neutral”, “Happiness”, “Sadness”, “Surprise”, “Fear”, “Disgust”, “Anger”, “Contempt”. As the dataset was used AffectNet. The most accurate architecture is Xception. It gave classification accuracy on training sample 97.65%, on cleaned testing sample 57.48% and top-2 accuracy on cleaned testing sample 76.70%. The category “Contempt” is worst recognized by all the types of neural networks considered, which indicates its ambiguity and similarity with other types of facial expressions. Experimental results show that for the considered task it does not matter, the color or grayscale image is fed to the input of the algorithm. This fact can save a significant amount of memory when storing data sets and training neural networks. The computing experiments was performed using graphics processor using NVidia CUDA technology with Keras and Tensorflow deep learning frameworks. It showed that the average processing time of one image varies from 4 ms to 30 ms for different architectures. Obtained results can be used in software for neural network training for face recognition systems.

**Keywords:** image recognition, classification, facial expression, emotion, face, deep learning, convolutional neural network.

1. Introduction

Currently, significant progress has been made in creating efficient image recognition algorithms based on the use of deep neural networks [1, 2, 3]. As a rule, such algorithms require the presence of a large number of images that are obtained in different lighting and noise conditions. They need huge amounts of memory for storage as well as for training. There are subject areas for which it is advisable to study the possibility of using grayscale images instead of color during training of recognition algorithms. This can reduce by three times the need for RAM or hard disk space.

One of these areas is the task of person’s facial expression recognition. The standard for determining the type of facial expressions is the Emotional facial action coding system (EMFACS-7), proposed by W. Friesen and P. Ekman in 1983 [4]. This generally accepted standard identifies seven basic types of emotions: 1) anger, 2) contempt, 3) disgust, 4) fear, 5) happiness, 6) sadness, 7) surprise. Additionally, it is considered a neutral facial expression.

At the initial stage, methods for emotions recognition on human face images were associated with manual selection of features: Gabor wavelets [5], local binary patterns [6], geometric deformation features on image sequences [7], 3D Surface Features [8], etc. Modern approaches are based on the automatic generation of image features based on deep convolutional neural networks, some of them use the prior alignment technique [9], some recognize facial expressions on images as they are [1, 10], including tuning of pre-trained networks on the task of face identification [11]. Also deep spatial-temporal networks were proposed for emotion recognition on video sequences [12]. Deep learning is also actively used to analyze facial expressions from face three-dimensional model [13].

There are a number of commercial services that implement closed-ended emotion recognition methods: Face API from Microsoft Azure [14], Amazon Emotion API [15], Affectiva Emotion SDK [16], etc.

However, the recognition of facial expressions on images in complex conditions of variable light, noise, uncomfortable perspective is still an important topic for further research.

For the study of approaches based on neural networks, there are many different data sets that differ in shooting conditions, the variety of people photographed, and the number of images per class. Some popular datasets and their features are listed in Table 1:

* Cohn-Kanade AU-Coded Expression Database [17] (CK), it is shown the statistics for the case with the first two and last two frames of image sequences from database,
* The Japanese Female Facial Expression Database [5] (JAFFE),
* Facial Expression Recognition Challenge [18] (FER2013),
* Facial expressions Repository [19] (FE),
* SoF dataset [20] (SoF),
* AffectNet [21].

The largest of them is AffectNet dataset (a total of more than 1 million images). In addition to manually labeled data, it contains automatically annotated images that researchers or developers can label and check on their own if necessary.

This paper discusses the issue of facial expression recognition on static images using modern deep learning methods, as well as choosing the format of the input data. On the one hand, color images provide additional information about a person’s face, on the other hand, using grayscale images reduces the effect of shooting conditions: light level, type of light source, etc. To choose one of these forms of image representation, it is necessary to conduct experiments with various architectures of neural networks with different sizes of input images.

**Table 1.** Datasets for facial expression recognition

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Database details | CK | JAFFE | FER2013 | FE | SoF | AffectNet |
| Image size | 640x490 -  720x480 | 256x256 | 48x48 | 23x29 -  355x536 | 640x480 | 129x129 -  4706x4706 |
| Image style | portrait | portrait | cropped face | cropped face | portrait | cropped face |
| Image type | grayscale,  color | grayscale | grayscale | grayscale,  color | color | color |
| Facial expression categories: | |  |  |  |  |  |
| Neutral | 324 | 30 | 6194 | 6172 | 667 | 75374 |
| Happy | 138 | 31 | 8989 | 5693 | 1042 | 134915 |
| Sad | 56 | 31 | 6077 | 220 | 237 (sad/ anger/disgust) | 25959 |
| Surprise | 166 | 30 | 4002 | 364 | 145 (surprise/ fear) | 14590 |
| Fear | 50 | 32 | 5121 | 21 | 0 | 6878 |
| Disgust | 118 | 29 | 547 | 208 | 0 | 4303 |
| Anger | 90 | 30 | 4953 | 240 | 0 | 25382 |
| Contempt | 36 | 0 | 0 | 9 | 0 | 4250 |
| Total: | 978 | 213 | 35883 | 12927 | 2091 | 291651 |

1. Task Formulation

In this paper we will solve the task of determining one of the eight facial expression categories (“Neutral”, “Happiness”, “Sadness”, “Surprise”, “Fear”, “Disgust”, “Anger”, “Contempt”) on grayscale or color images with cropped faces, see the Fig. 1.

We had taken modern and the biggest open-source dataset – AffectNet which contains 287651 images as training sample and 4000 images (500 images per class) as testing sample [21]. Samples include images of different sizes from 129×129 to 4706×4706 pixels that are obtained from different cameras in different shooting conditions.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | | | | | | | |
| 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 |

**Fig. 1.** Examples of labeled images with facial expressions from AffectNet Dataset: 0 – Neutral, 1 – Happiness, 2– Sadness, 3 – Surprise, 4 – Fear, 5 – Disgust, 6 – Anger, 7 – Contempt

To solve the task it is necessary to develop various variants of deep neural network architectures and to test them on the available data set with 1-channel (grayscale) and 3-channel (color) image representation. We must determine which image representation is best used for the task of facial expression recognition. Also, we need to select the best architecture that will provide best performance and the highest quality measures of image classification: accuracy, precision and recall [22].

1. Dataset preparation

AffectNet [21] was chosen as the main dataset, which is one of the largest modern datasets for facial expression recognition. However, it contains relatively few images for the “Fear”, “Disgust” and “Contempt” categories compared to other categories. To conduct experiments for learning neural networks, augmentation of images was carried out, and a balanced training sample was formed with 10,000 images per class.

For image augmentation we have used 5 sequential steps:

1. Coarse Dropout – setting rectangular areas within images to zero. We have generated a dropout mask at 2 to 25 percent of image's size. In that mask, 0 to 2 percent of all pixels were dropped (random per image).
2. Affine transformation – image rotation on random degrees from -15 to 15.
3. Flipping of image along vertical axis with 0.9 probability.
4. Addition Gaussian noise to image with standard deviation of the normal distribution from 0 to 15.
5. Cropping away (cut off) random value of pixels on each side of the image from 0 to 10% of the image height/width.

Results of this augmentation procedure are shown on Fig. 2.



**Fig. 2.** Examples of augmented images for Training sample 2 (balanced)

As the most of open-source datasets AffectNet contains wrong ground truth labels for cropped faces (Fig. 2). We had cleaned the testing sample for more correct evaluation of classifiers. As a result we create Testing sample 2 which have 3210 images.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| C:\Users\yuddim\YandexDisk\MIPT\Publications\2019 - Нейроинформатика-2019\To paper - wrong.jpg | | | | | | | |
| Errors in “Neutral” category | Errors in “Happiness” category | Errors in “Sadness” category | Errors in “Surprise” category | Errors in “Fear” category | Errors in “Disgust” category | Errors in “Anger” category | Errors in “Contempt” category |

**Fig. 3.** Examples of wrong ground truth labels in testing sample of AffectNet Dataset

Details of the datasets used in this research are given in Table 2.

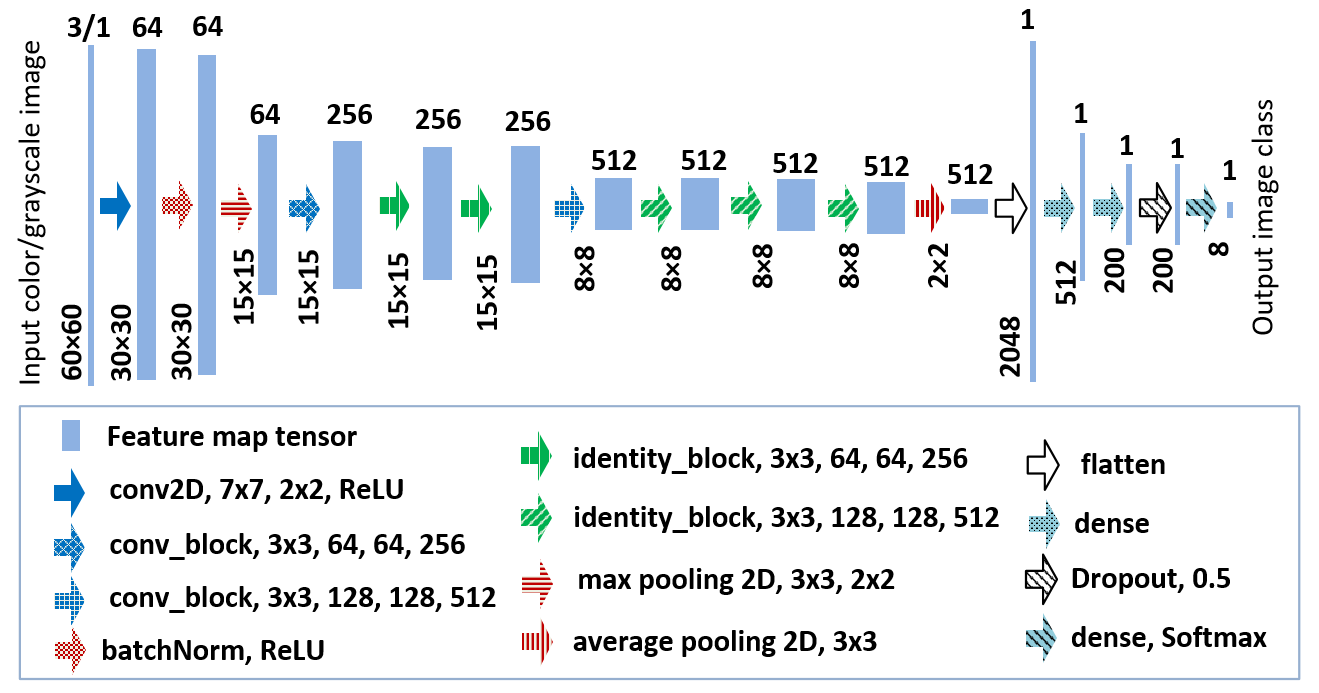
**Table 2.** Training and testing samples of used dataset

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Facial expression category | Training sample 1 | Training sample 2 (balanced) | Testing sample 1 | Testing sample 2 (cleaned) |
| 0 - Neutral | 74874 | 10000 | 500 | 490 |
| 1 - Happiness | 134415 | 10000 | 500 | 451 |
| 2 - Sadness | 25459 | 10000 | 500 | 473 |
| 3 - Surprise | 14090 | 10000 | 500 | 453 |
| 4 - Fear | 6348 | 10000 | 500 | 477 |
| 5 - Disgust | 3803 | 10000 | 500 | 359 |
| 6 - Anger | 24882 | 10000 | 500 | 351 |
| 7 - Contempt | 3749 | 10000 | 500 | 156 |
| Total: | 287621 | 80000 | 4000 | 3210 |

1. Classification of Emotion Categories using Deep Convolutional Neural Networks

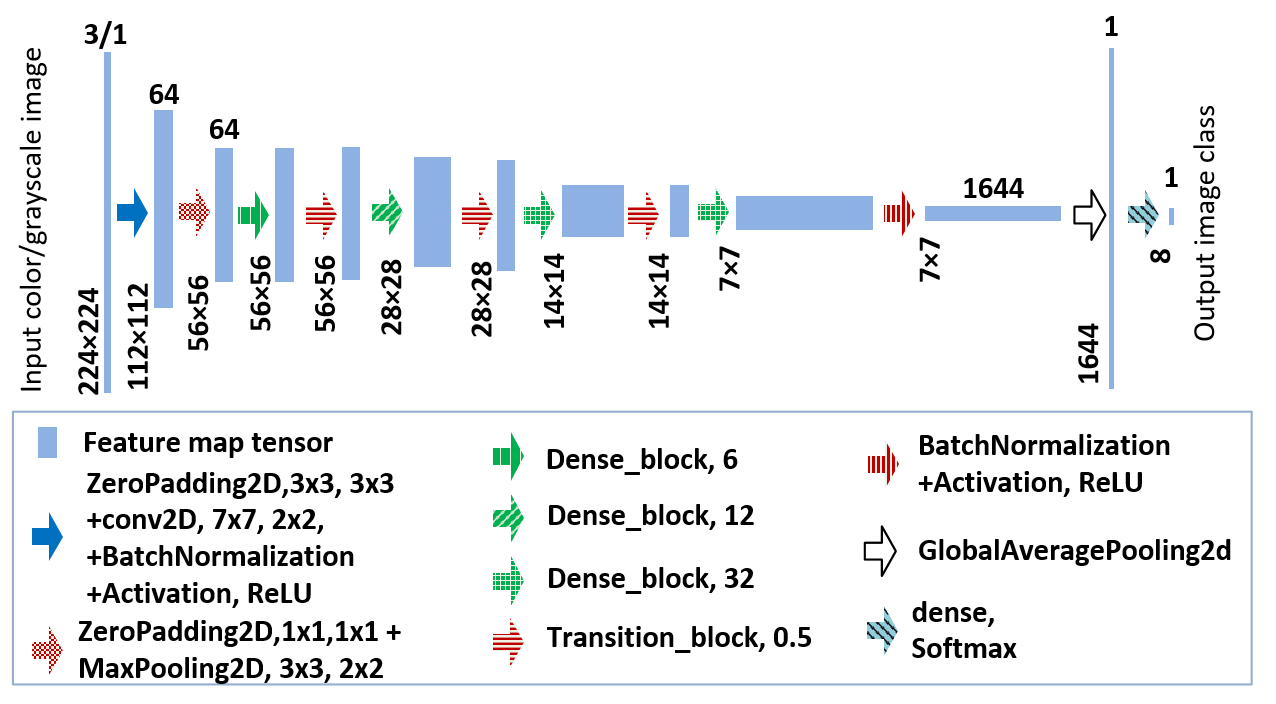
In this paper to solve formulated task we investigate the application of a deep convolutional neural networks of three architectures:

* ResNetM architecture inspired from ResNet [23] and implemented by authors in previous works [24]. It has input tensor 120x120x3 for color images and 120x120x1 for grayscale images. Its structure is shown in Fig. 4 and contains 3 convolutional blocks, 5 identity blocks, 2 max pooling layers, 1 average pooling layer and one output dense layer. First 11 layers and blocks provide automatic feature extraction and the last one fully connected layer allows us to find one of five image classes corresponding to input image. ResNetM net was trained on full Training sample 1.



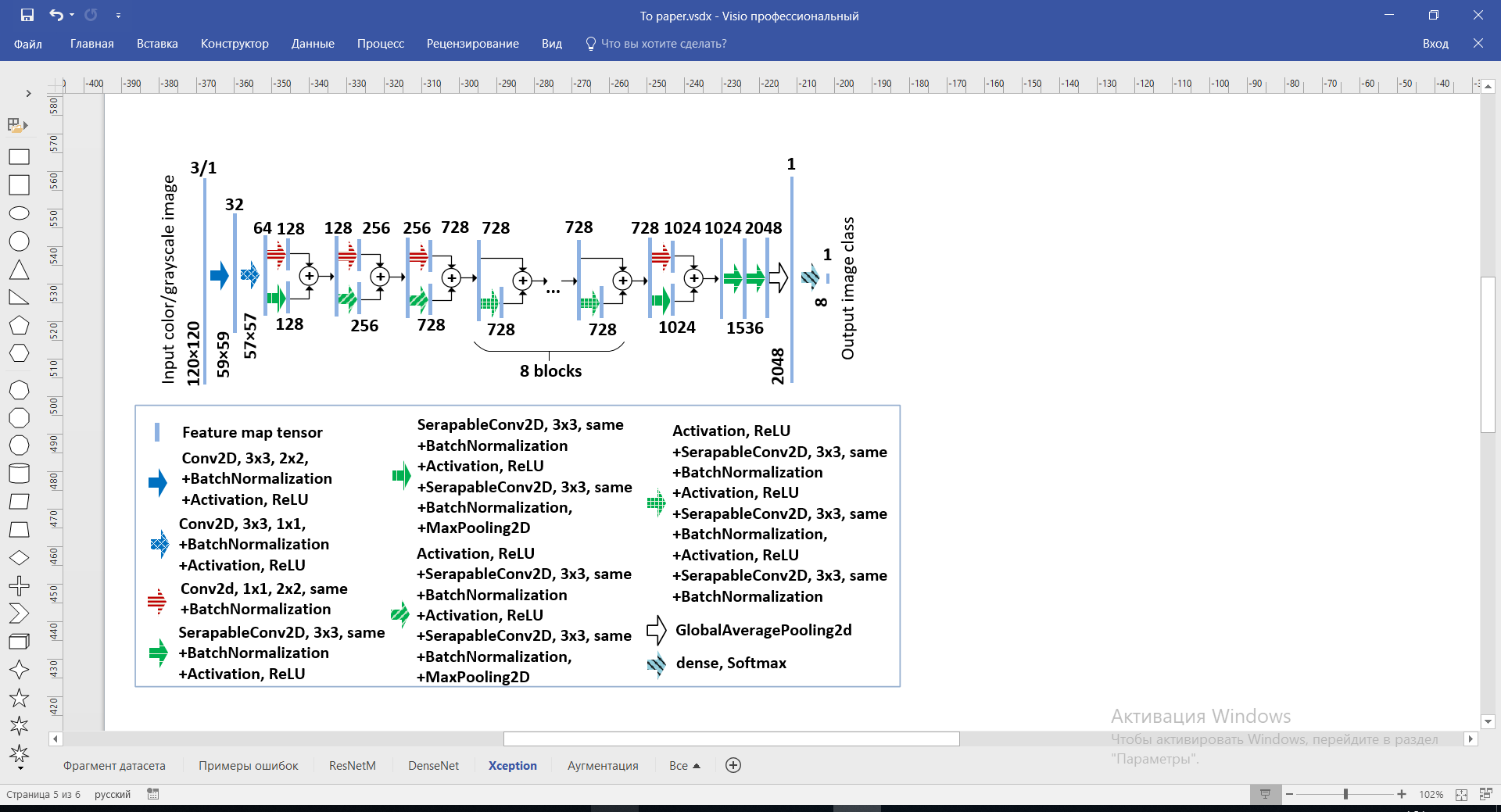
**Fig. 4.** ResNetM architecture.

* DenseNet architecture is based on DenseNet169 model [25] with input tensor 224x224x3 for color images and 224x224x1 for grayscale images. Its structure uses alternating Dense and Transition blocs (Fig. 5). The dataset from Training sample 2 containing 4000 images per class was prepared for DenseNet training.



**Fig. 5.** DenseNet architecture.

* Xception architecture [26] with changed input tensor to 120x120x3 for color images and 120x120x1 for grayscale images. This structure is a development of the Inception [27] and is based on prospective Separable convolutional blocks architectures (see Fig. 6). Xception net was trained on balanced Training sample 2.

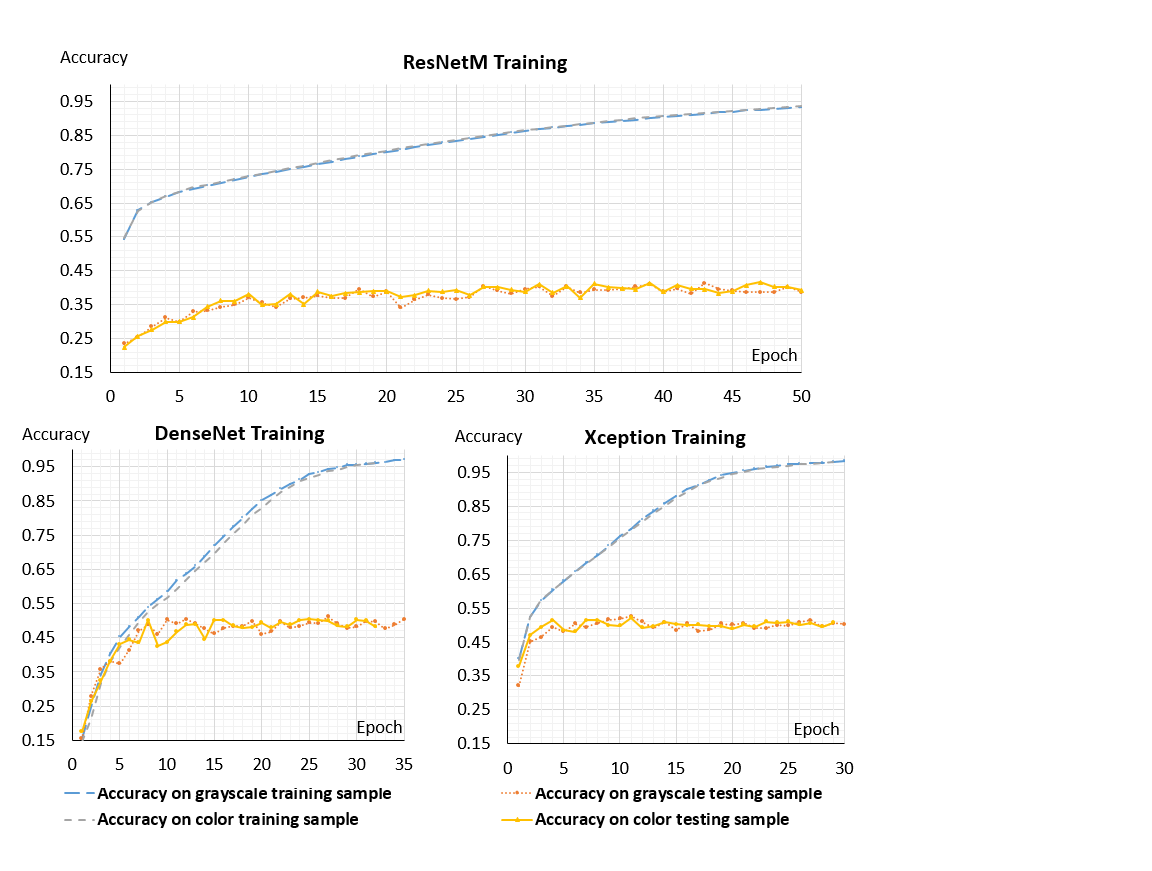


**Fig. 6.** Xception architecture.

Output layer in all architecture has 8 neurons with “Softmax” activation function. All input images are pre-scaled to a size of 60 × 60 pixels for ResNetM architecture, 120x120 pixels for Xception architecture and 224x224 pixels for DenseNet169. Neural networks works with color (three-channel) and grayscale (one-channel) images.

To train the neural networks we have used “categorical crossentropy” loss function, Stochastic Gradient Descent (SGD) as training method with 0.001 learning rate. Accuracy is used as classification quality metric during training. The batch is consisted of 5 images.

The training process of deep neural networks is shown in Fig. 7. The training experiment was carried out for 50 learning epochs using our developed software tool implemented on Python 3.5 programming language with Keras + Tensorflow frameworks [28]. We can see that DenseNet and Xception networks have similar speed and accuracy, while ResNetM achieves much lower accuracy rates on test samples compared to them.



**Fig. 7.** Training of deep neural networks with ResNetM, DenseNet and Xception architectures.

The calculations had performed using the NVidia CUDA technology on the graphics processor of the GeForce GTX 1060 graphics card with 6.00GB, central processor Intel Core i-5-8300H, 4 Core with 2.3GHz and 24 GB RAM.

Table 3 shows the results of the facial expression recognition on training and test samples with color or grayscale images using ResNetM, DenseNet169 and Xception architectures.

Analysis of the obtained results shows the highest accuracy and on all samples Xception architecture with grayscale input images: 97.65% on training sample, 57.48% on testing sample 2 and top-2 accuracy 76.70%. It also has the greatest and more balanced values of precision and recall for almost all categories (classes) of facial expression except for “Anger” и “Contempt”.

ResNetm is significantly faster than all other architectures: about 4 ms for processing a single image against 12 ms for Xception and 30 ms for DenseNet. Also, this architecture has the highest recognition recall for the “Happiness” category.

**Table 3.** Quality of facial expression recognition on AffectNet Dataset

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Metric | ResNetM | | DenseNet | | Xception | |
| Color | Grayscale | Color | Grayscale | Color | Grayscale |
| Accuracy on train sample | 0.9283 | 0.9139 | 0.9168 | 0.9428 | 0.9686 | **0.9765** |
| Accuracy on test sample 2 | 0.4844 | 0.4781 | 0.5520 | 0.5427 | 0.5654 | **0.****5748** |
| Top-2 acc. on test sample 2 | 0.6748 | 0.6766 | 0.7467 | 0.7371 | 0.7355 | **0.****7670** |
| Classif. time per image, s | **0,0042** | 0.0047 | 0.0305 | 0.0299 | 0.0120 | 0.0123 |
| Weights number | 2613392 | **2607120** | 12656200 | 12649928 | 20877872 | 20877296 |
| Size of model, Mb | 10.654 | **10.629** | 51.933 | 51.908 | 83.826 | 83.823 |
| Size of train sample on HDD, Mb | 26373.7 | 9520.9 | 3432.7 | 1210.6 | 7624.1 | 2670.2 |
| Size of train sample  in operative memory, Mb | 12425.2 | 4141.7 | 19267.6 | 6422.5 | 13824.0 | 4608.0 |
| Quality metrics on test sample 2 (cleaned) | | | | | | |
| Neutral (0): Precision | 0.375 | 0.4083 | 0.4838 | **0.5644** | 0.5422 | 0.5223 |
| Neutral (0): Recall | 0.6061 | 0.5000 | 0.5490 | 0.3755 | 0.4592 | **0.5735** |
| Happiness (1): Precision | 0.5214 | 0.5325 | 0.7363 | 0.7701 | 0.7363 | **0.7973** |
| Happiness (1): Recall | **0.9468** | 0.9268 | 0.7428 | 0.7428 | 0.8049 | 0.7849 |
| Sadness (2): Precision | 0.5184 | 0.4103 | 0.6070 | 0.5617 | 0.5221 | **0.6099** |
| Sadness (2): Recall | 0.4165 | 0.5370 | 0.4735 | 0.4715 | **0.6490** | 0.4693 |
| Surprise (3): Precision | 0.4810 | 0.4708 | 0.4977 | 0.5177 | **0.5455** | 0.5000 |
| Surprise (3): Recall | 0.3907 | 0.3377 | 0.4966 | 0.5475 | 0.5033 | **0.6137** |
| Fear (4): Precision | 0.5880 | 0.5951 | 0.5867 | 0.5665 | **0.6181** | 0.5864 |
| Fear (4): Recall | 0.3501 | 0.3542 | 0.5744 | **0.5898** | 0.5597 | 0.5765 |
| Disgust (5): Precision | 0.6510 | 0.5679 | 0.5287 | 0.5600 | 0.5912 | **0.6655** |
| Disgust (5): Recall | 0.2702 | 0.3259 | **0.6156** | 0.5070 | 0.5599 | 0.5097 |
| Anger (6): Recall | 0.4645 | 0.4550 | **0.5552** | 0.4456 | 0.4941 | 0.4802 |
| Anger (6): Recall | 0.5413 | 0.4900 | 0.4587 | **0.5954** | 0.4758 | 0.5869 |
| Contempt (7): Precision | 0.3333 | **0.5384** | 0.3103 | 0.2827 | 0.3065 | 0.3407 |
| Contempt (7): Recall | 0.0192 | 0.0448 | 0.4038 | **0.5128** | 0.3654 | 0.2949 |

DenseNet surpasses all other architectures in “Anger” category recognition and is better in terms of recognition recall of “Fear” and “Contempt” categories. Also it has the highest precision for “Neutral” category.

The category “Contempt” is poorly recognized by all the types of neural networks considered, which speaks primarily of its ambiguity and similarity with other types of facial expressions, in particular “Neutral”.

As for the size of the network, here the smallest amount of memory is occupied by the weights for ResNetM (about 10.6 MB), the largest volume by the weights of the Xception network (83.8 MB).

For all considered types of neural networks, the representation of the input images in gray or color format did not lead to any significant difference in the values of the metrics accuracy, top-2 accuracy, processing time per image, and weights number. Thus, it can be concluded that for the facial recognition task it does not matter, the color or grayscale image is fed to the algorithm. This fact can save a significant amount of memory when storing datasets (about 65% of HDD space) and training of neural networks (about of 67% of operative memory).

1. Conclusions

It follows from the Table 3 that the applied architectures of a deep neural network for face expression recognition on AffectNet dataset show high quality indicators for the training set, but significantly worse results on the testing sample. This can be explained by the ambiguity of certain emotions on a person’s face, a variety of shooting angles and the presence of conflicting data in the training sample. The most accurate architecture is Xception. It gave classification accuracy 97.65% on training sample, 57.48% on testing sample 2 and top-2 accuracy 76.70% on testing sample 2.

The category “Contempt” is worst recognized by all the types of neural networks considered, which indicates its ambiguity and similarity with other types of facial expressions.

Experimental results show that for the considered task it does not matter, the color or grayscale image is fed to the input of the algorithm. This fact can save a significant amount of memory when storing data sets and training neural networks.

An important aspect for the further application of the considered approaches is the average classification time per image. It varies from 4 ms for ResNetM to 30 ms for DenseNet. This suggests that the all described approaches can be integrated into a real-time face recognition software.

To further studies on the paper topic it is necessary to expand the training and test samples to cover more images in “Fear”, “Disgust” and “Contempt” categories. Also, it will be promising to explore the emotion recognition on images with a face alignment based on key points, in order to reduce the impact of choosing bounding box of face detection algorithms.

1. Acknowledgment

The research was made possible by Government of the Russian Federation (Agreement № 075-02-2019-967).

References

1. Zeng, N., Zhang, H., Song, B., Liu, W., Li, Y., Dobaie, A. M.: Facial expression recognition via learning deep sparse autoencoders. Neurocomputing, vol. 273, pp. 643–649 (2018).
2. Yudin, D., Knysh, A.: Vehicle recognition and its trajectory registration on the image sequence using deep convolutional neural network. The International Conference on Information and Digital Technologies, pp. 435–441 (2017).
3. Yudin, D., Naumov, A., Dolzhenko, A., Patrakova, E.: Software for roof defects recognition on aerial photographs. Journal of Physics: Conference Series, 1015(3),032152 (2018).
4. Friesen, W., Ekman, P.: EMFACS-7: Emotional facial action coding system. Unpublished manuscript, University of California at San Francisco 2, 36 (1983).
5. Lyons, M. J., Akemastu, S., Kamachi, M., Gyoba, J. Coding Facial Expressions with Gabor Wavelets. 3rd IEEE International Conference on Automatic Face and Gesture Recognition, pp. 200–205 (1998).
6. Shan, C., Gong, S., McOwan, P. W.: Facial expression recognition based on Local Binary Patterns: A comprehensive study. Image and Vision Computing, vol. 27, no. 6, pp. 803–816. (2009).
7. Kotsia, I. Pitas, I. Facial expression recognition in image sequences using geometric deformation features and support vector machines. IEEE transactions on image processing (2006).
8. Wang, J., Yin, L., Wei, X., Sun, Y.: 3D Facial Expression Recognition Based on Primitive Surface Feature Distribution. IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06) (2006).
9. Lopes, A. T., de Aguiar, E., De Souza, A. F., Oliveira-Santos, T.: Facial expression recognition with Convolutional Neural Networks: Coping with few data and the training sample order. Pattern Recognition, vol. 61, pp. 610–628 (2017).
10. Mollahosseini, A., Chan, D., Mahoor, M. H.: Going deeper in facial expression recognition using deep neural networks. IEEE Winter Conference on Applications of Computer Vision (WACV) (2016)
11. Ding, H., Zhou, S. K., Chellappa, R.: FaceNet2ExpNet: Regularizing a Deep Face Recognition Net for Expression Recognition. 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017) (2017)
12. Zhang, K., Huang, Y., Du, Y., Wang, L. Facial Expression Recognition Based on Deep Evolutional Spatial-Temporal Networks. IEEE Transactions on Image Processing, vol. 26, no. 9, pp. 4193– 4203 (2017)
13. Zhang, T., Zheng, W., Cui, Z., Zong, Y., Yan, J., Yan, K.: A Deep Neural Network-Driven Feature Learning Method for Multi-View Facial Expression Recognition. IEEE Transactions on Multimedia, vol. 18, no. 12, pp. 2528–2536 (2016).
14. Face API от Microsoft Azure, <https://azure.microsoft.com/ru-ru/services/cognitive-services/face/#detection>, last accessed 2019/05/26.
15. Amazon Emotion API, [https://docs.aws.amazon.com/rekognition/latest/dg/ API\_Emotion.html](https://docs.aws.amazon.com/rekognition/latest/dg/%20API_Emotion.html), last accessed 2019/05/26.
16. Affectiva Emotion SDK. <https://www.affectiva.com/product/emotion-sdk/>, last accessed 2019/05/26.
17. Lucey, P., Cohn, J. F., Kanade, T., Saragih, J., Ambadar, Z., Matthews, I.: The Extended Cohn-Kanade Dataset (CK+): A complete expression dataset for action unit and emotion-specified expression. Proceedings of the Third International Workshop on CVPR for Human Communicative Behavior Analysis (CVPR4HB 2010), pp. 94-101 (2010).
18. Carrier, P.-L., Courville, A.: Challenges in Representation Learning: Facial Expression Recognition Challenge (2013), <https://www.kaggle.com/c/challenges-in-representation-learning-facial-expression-recognition-challenge/data>, last accessed 2019/05/26.
19. Facial expressions. A set of images for classifying facial expressions, <https://github.com/muxspace/facial_expressions>, last accessed 2019/05/26.
20. Afifi, M., Abdelhamed, A.,: AFIF4: Deep gender classification based on an AdaBoost-based fusion of isolated facial features and foggy faces. Journal of Visual Communication and Image Representation (2019)
21. Mollahosseini, A., Hasani, B., Mahoor, M. H.: AffectNet: A Database for Facial Expression, Valence, and Arousal Computing in the Wild. IEEE Transactions on Affective Computing, vol. 10 , no. 1, pp. 18–31 (2017)
22. Olson, D. L., Delen D.: Advanced Data Mining Techniques. Springer, 1st edition (2008).
23. Kaiming, H., Xiangyu, Z., Shaoqing, R., Jian S.: Deep Residual Learning for Image Recognition. ECCV, arXiv:1512.03385 (2015).
24. Yudin, D., Kapustina, E.: Deep Learning in Vehicle Pose Recognition on Two-Dimensional Images. [Advances in Intelligent Systems and Computing](https://www.scopus.com/sourceid/5100152904?origin=resultslist) 874, pp. 434–443 (2019).
25. Huang, G., Liu, Z., van der Maaten, L., Weinberger, K. Q.: Densely Connected Convolutional Networks. CVPR 2017[, arXiv:1608.06993](https://arxiv.org/abs/1608.06993) (2017).
26. Chollet, F.: Xception: Deep Learning with Depthwise Separable Convolutions. CVPR 2017, arXiv: 1610.02357 (2017)
27. Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna Z.: Rethinking the Inception Architecture for Computer Vision. ECCV, arXiv:1512.00567 (2016).
28. [Chollet](https://scholar.google.ru/citations?user=VfYhf2wAAAAJ&hl=ru&oi=sra), F.: Keras: Deep learning library for theano and tensorflow. <https://keras.io/>, last accessed 2019/05/26