Methods for emotions, mood, gender and age recognition

D D Pribavkin[1], P Y Yakimov[1,2]

[1]Samara National Research University, Moskovskoe Shosse 34A, Samara, Russia, 443086
[2]Image Processing Systems Institute of RAS - Branch of the FSRC "Crystallography and Photonics" RAS, Molodogvardejskaya street 151, Samara, Russia, 443001

e-mail: pribavkindenis@gmail.com

Abstract. Recognition on images not only of shapes, but also of metadata is becoming increasingly popular among researchers in the field of convolutional neural networks and deep learning. This article provides an analytical overview of modern software solutions that recognize the images of emotions, mood, gender and age of a person. Enthusiasts invent all new and new architectures of convolutional neural networks, allowing to solve the tasks with considerable recognition accuracy.

1. Introduction

In recent years, there has been a rapid development of parallel data processing technologies, in particular due to the development of graphics processors, which are no longer intended only for computer graphics. This made it possible to train even the most complex neural networks in their architectures and opened up a whole horizon of unsolvable tasks [1], [2], [3]. Modern intellectual systems focus not only on pattern recognition from the input image, but also learn to isolate metadata from recognized objects, such as emotions, mood, gender, or a person's age.

Many researchers and enthusiasts in the field of machine learning and convolutional neural networks develop and offer their own, unique solutions that are different both ideologically and technically.

This article offers an analytical review of the following software solutions in the field of recognition of emotions, mood, sex and age of a person:

- Emotion recognition using Deep Convolutional Neural Network [4].
- A Compact Soft Stagewise Regression Network [5].
- Real-time Convolutional Neural Networks for Emotion and Gender Classification [6].
- Age Recognition using CNNs [7].

2. Review of existing solutions

2.1. Emotion recognition using Deep Convolutional Neural Networks

A solution that is a trained neural network that recognizes real-time emotions on a human face recognized from the input video stream.

It was built using the TFLearn programming library for the python programming language, based on the well-known TensorFlow machine learning framework developed by Google in 2015 [8]. This framework simplifies the development of the network, as it requires describing only the layers themselves instead of describing each neuron separately, and also simplifies network training by providing real-time process feedback and learning accuracy. Moreover, the library allows you to save the result of a trained model to use it later.

The resulting neural network model is shown in Figure 1.

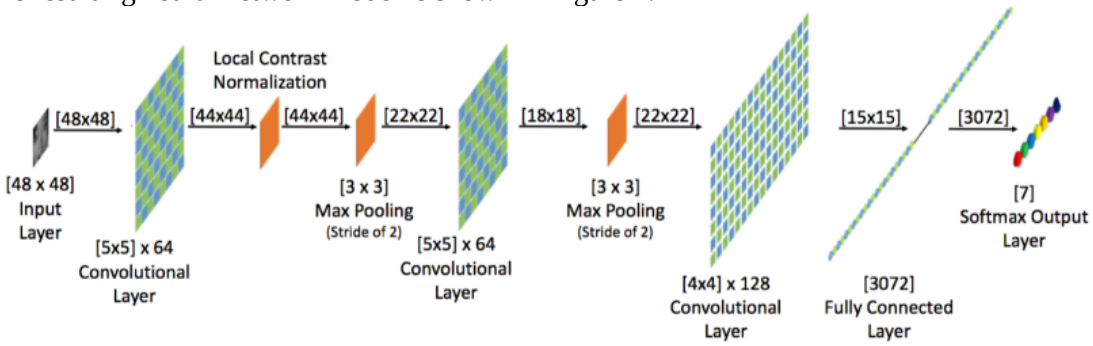

Figure 1. Neural network model.

In each frame of the video stream, an attempt is made to recognize a human face (s). This is achieved using the OpenCV open library recognition method [9]. Then, if a face was recognized in the image, that face is cut out and scaled to a size of 48x48 pixels. Only after that it is fed to the input of the neural network. Thus, we get optimized software that affects neural network resources only if there is at least one human face in the frame.

The model was trained with the help of dataset FERC-2013, which has about 20,000 images containing examples of the following emotions: anger, fear, happiness, sadness, surprise, indifference and disgust. The density of the distribution of emotions in this data is reflected in Figure 2.

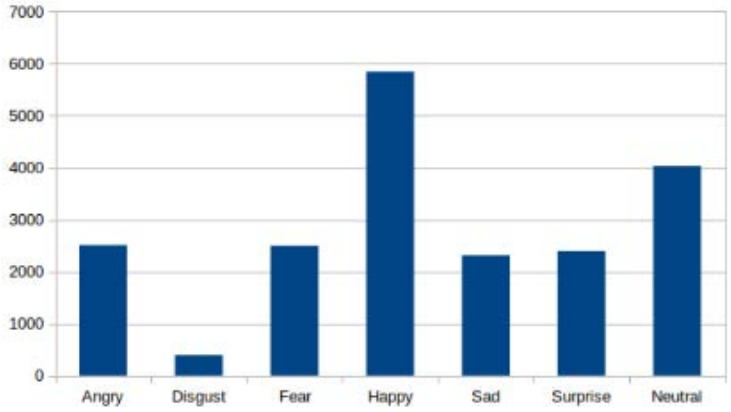

Figure 2. Density of distribution of emotions in dataset FERC-2013.

According to the results of training, the accuracy of recognition of emotions was achieved in 67%.

2.2. SSR-Net

This solution is an original neural network with soft stepwise regression (soft stagewise regression network) for recognizing age and sex. The network recognizes age and gender according to the following principle: images of 64x64 pixels are fed to the input of the network, a multi-level classification is made from several classes, where each level serves to refine the previous result, and then the result of the classification is processed using a regression.

The model itself is very compact and takes only 0.32 MB. But in spite of its compact dimensions, the performance of SSR-Net is close to the characteristics of the most modern methods, the sizes of models of which are 1500 times larger.

A model of this neural network with three levels and a pool size of 2 is shown in Figure 3.

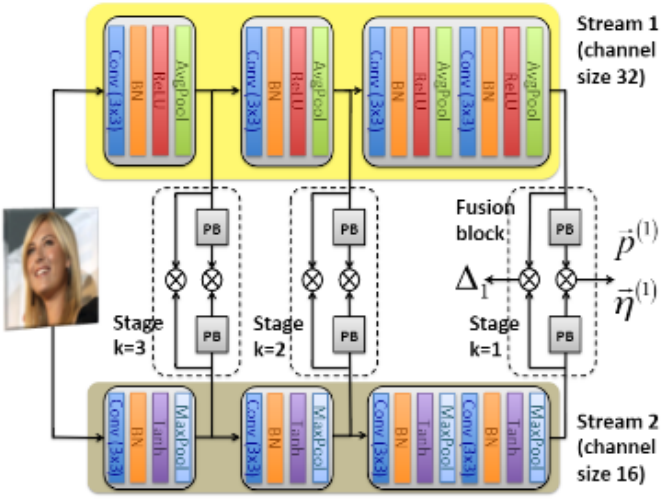

Figure 3. Neural network model SSR-Net.

For training this model, such datasets as IMDB, WIKI and Morph2 [11] were used. About 80% of randomly selected images from datasets were used to train the network, and the remaining 20% were used for testing.

An example of dependence of the number of SSR-Net, MobileNet and DenseNet network recognition errors trained in Morph2 data on the number of epochs is presented in Figure 4.

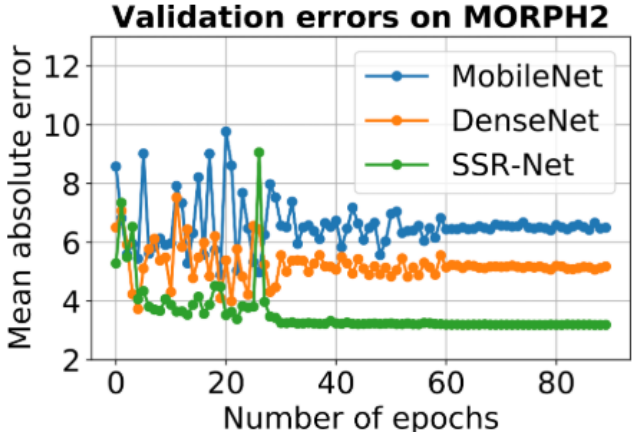

Figure 4. Graph of recognition errors versus the number of epochs.

2.3. Face classification and detection from the B-IT-BOTS robotics team

It is a real-time classifier of emotions and gender of a person, based on the convolutional neural network and the open image processing library openCV [9]. The model of this neural network is shown in Figure 5 and contains 600,000 parameters.

This model was trained in IMDB dataset, which has about 460,723 RGB images, each of which belongs to one of the classes: male or female [10]. At this dataset, recognition accuracy of 96% was achieved. Also, this model was validated on dataset FER-2013, which includes 35,887 images in gray tones, each of which belongs to one of the classes of emotions: anger, disgust, fear, joy, sadness, surprise and indifference. At this dataset, 66% accuracy was achieved.

2.4. Age and gender estimation

This solution is the implementation of a convolutional neural network for recognizing the sex and age of a person from the input image. The basis for the VGG-16 network architecture was taken due to its depth and controllability. This network accepts 256x256 pixel images as input.

The training was carried out on IMDB-WIKI datasets, and recognition accuracy of 64% was achieved [10].

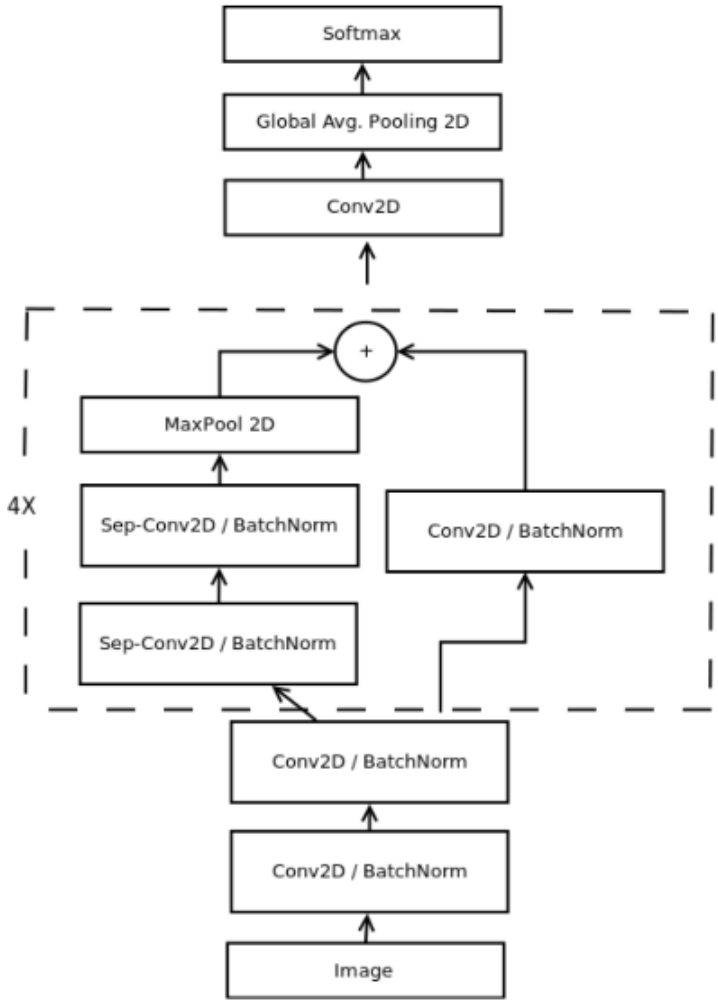

Figure 5. Neural network model.

2.5. General comparison of implementations

As a result of the analytical review, the decisions contained in the publications [6, 7] were selected:

1. A solution that recognizes a person's age, trained in dataset IMDB-WIKI with a recognition accuracy of 64%.
2. A solution that recognizes the sex of a person, trained in dataset IMDB with a recognition accuracy of 96%.
3. A solution that recognizes a person's emotions, trained in dataset FER-2013 with a recognition accuracy of 66%.

The source code of each solution was carefully analyzed and revised so that the digital image was provided as input to the software, and the result of the prediction of a convolutional neural network was obtained.

3. Conducting experimental studies

At the end of the previous section, software solutions were obtained, the main task of which is to recognize the age, gender and emotions of a person from a digital face image.

These solutions were chosen as objects for conducting an experimental study of their performance on 10 random images of the faces of people from the IMDB-WIKI dataset.

The following equipment and software were used during the pilot study:

1. Processor: intel Core i5-4570 3.2 GHz.
2. RAM: 8 Gb.
3. Operation system: Manjaro 18.0.4 «Illyria».

4. Programming language: Python 3.6.5.

Below (Figures 6–15)are the images used in the pilot study:

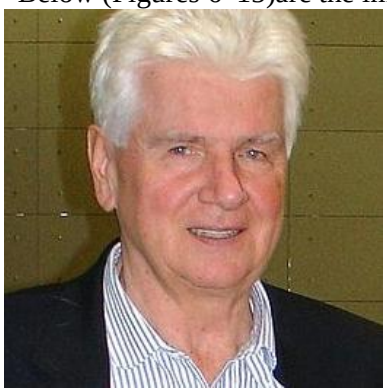

Figure 6. Image number 1.

Figure 7. Image number 2.

Figure 8. Image number 3.

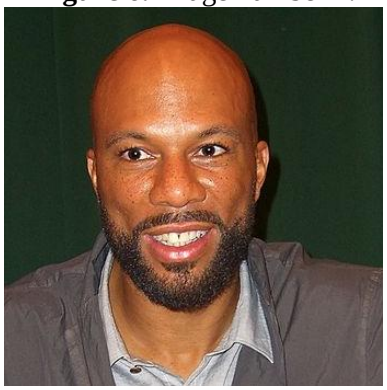

Figure 9. Image number 4.

Figure 10. Image number 5.

Figure 11. Image number 6.

Figure 12. Image number 7.

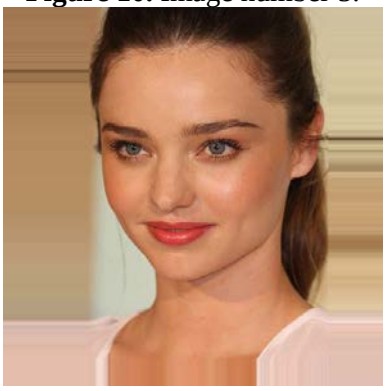

Figure 13. Image number 8.

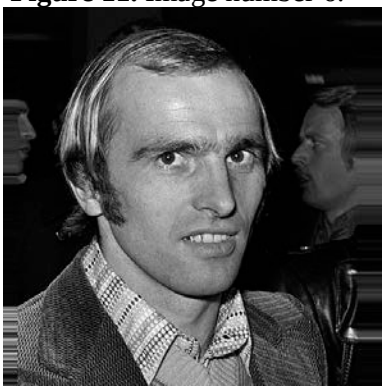

Figure 14. Image number 9.

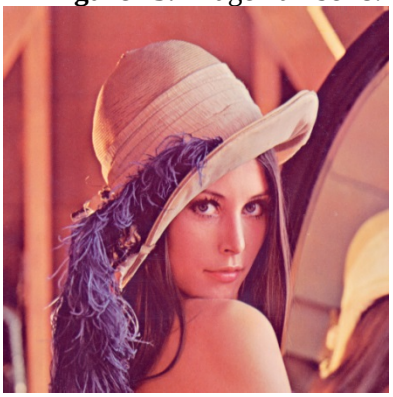

Figure 15. Image number 10.

Since the neural network with the same input data always produces the same result, performing multiple independent attempts at recognition in a row does not have much value, however, to measure a more accurate execution time of a convolutional neural network in a single digital image.

Tables 1, 2 and 3 present the results of recognizing age, gender and emotions, respectively, as well as the average execution time for 5 runs of a convolutional neural network in the next digital image.

Table 1. Final Age Recognition.

Image, №	Age	Average time, ms
1	64	523
2	31	642
3	39	627
4	34	510
5	47	481
6	17	524
7	56	537
8	25	450
9	34	583
10	25	592

Table 2. Final Gender Recognition.

Image, №	Gender	Average time, ms
1	Man	324
2	Woman	451
3	Man	430
4	Man	318
5	Man	293
6	Woman	305
7	Man	421
8	Man	326
9	Man	432
10	Woman	476

Table 3. Final Emotion Recognition.

Image, №	Emotion	Average time, ms
1	Happiness	389
2	Happiness	513
3	Surprise	527
4	Happiness	408
5	Fear	362
6	Happiness	385
7	Happiness	503
8	Happiness	476
9	Sadness	513
10	Neutral	563

As we can see from tables 1, 2 and 3, although the accuracy of recognition stated by the authors still requires some refinements (for example, a neural network that recognizes the sex of a person is clearly mistaken in image No. 8), these pre-trained convolutional neural networks are capable of producing meaningful results, reflecting reality.

4. Conclusion

As a result, we can conclude that such tasks as the recognition of emotions, mood, gender and age are very popular among researchers all over the world. Enthusiasts use different approaches to the

implementation of intelligent systems that can solve such problems and achieve good results in accuracy of recognition, even with limited resources. The main means of the implementation of the tasks are convolutional neural networks of various architectures, trained in well-known in the network dataset images.

5. References

[1] Bibikov S A, Kazanskiy N L and Fursov V A 2018 Vegetation type recognition in hyperspectral images using a conjugacy indicator *Computer Optics* **42(5)** 846-854 DOI: 10.18287/2412-6179-2018-42-5-846-854
[2] Shatalin R A, Fidelman V R and Ovchinnikov P E 2017 Abnormal behavior detection method for video surveillance applications *Computer Optics* **41(1)** 37-45 DOI: 10.18287/2412- 6179-2017-41-1-37-45
[3] Shustanov A, Yakimov P 2017 CNN Design for Real-Time Traffic Sign Recognition *Procedia Engineering* **201** 718-725 DOI: 10.1016/j.proeng.2017.09.594
[4] Correa E, Jonker A, Ozo M, Stolk R *Emotion Recognition using Deep Convolutional Neural Network* URL: https://github.com/isseu/emotion-recognition-neural-networks/blob/master/paper/Report_NN.pdf (1.11.2018)
[5] Tsun-Yi Y, Yi-Hsuan H, Yen-Yu L, Pi-Cheng Hu, Yung-Yu Ch *SSR-Net: A Compact Soft Stagewise Regression Network for Age Estimation* URL: https://github.com/shamangary/SSR-Net/blob/master/ijcai18_ssrnet_pdfa_2b.pdf (14.11.2018)
[6] Arriaga O, Plöger P G, Valdenegro M *Real-time Convolutional Neural Networks for Emotion and Gender Classification* URL: https://github.com/oarriaga/face_classification/blob/master/report.pdf (8.10.2018)
[7] Pakulich D, Alyamkin S, Yakimov S 2019 Age estimation using face recognition with convolutional neural networks *Avtometriya* **55(3)** 52-61 (in Russian) DOI: 10.15372/AUT20190307
[8] TFLearn library URL: http://tflearn.org/ (04.10.2018)
[9] OpenCV library URL: http://opencv.org (04.10.2018)
[10] IMDB-wiki dataset URL: https://data.vision.ee.ethz.ch/cvl/rrothe/imdb-wiki/ (04.10.2018)

Acknowledgements
This work was partly funded by the Russian Foundation for Basic Research – Project # 17-29-03112 ofi_m and the Russian Federation Ministry of Science and Higher Education within a state contract with the "Crystallography and Photonics" Research Center of the RAS under agreement 007-ГЗ/Ч3363/26.