Automated recognition of facial expressions and gender in humans implemented on mobile devices

Romulus-Cristian Moraru
Department of Electronics and Computers
Faculty of Electrical Engineering and Computer Science
Brasov, Romania
morarucristian32@gmail.com

Angel Cațaron
Department of Electronics and Computers
Faculty of Electrical Engineering and Computer Science
Brasov, Romania
cataron@unitbv.ro, ORCID: 0000-0002-3986-5437

***Abstract*— This paper presents the implementation on mobile devices of an automated system which can recognize 7 basic facial expressions each of them associated to an emotion: happy, sad, angry, disgust, surprise, fear, and neutrality alongside a person's gender from a facial image using convolutional neural networks and transfer learning. The human faces are extracted from images generated by the camera of the device which runs the application in real time. The server side of the system was built in two versions: desktop which has good performances regarding recognition and processing speed and web which can run on any device which provides a browser and a camera.**

Keywords—mobile devices, neural networks, emotions, gender, transfer learning

I. INTRODUCTION

The automated recognition of facial expressions or gender of a person are activities that are trivial to human beings and have become more and more accessible to computers too thanks to the increasing popularity and development of convolutional neural networks. In this case, the need for systems that can perform on their own such complex operations has never been greater. There are numerous domains in which an automated system capable of recognizing facial expressions or gender is useful:

1) Software developement: it can help programmers to evaluate the impact of a of a website on a target-group of users, or to verify if and to what extent there is a connection between emotions, productivity, and the quality of developed applications.

2) In schools: to understand if and to what extent there is a connection between the efficiency of learning at home and certain emotional patterns. Also, there can be observed the emotional reaction to different activities in self learning in order to classify the efficiency of the methods and the materials suggested.

3) In improving video games: in observing different emotional patterns which may occur in interaction with different games or parts of the games, in particular which emotions are predominant, and which are not. In this case, the game developers can adapt their games to trigger the desired emotions in the minds of the players. [1]

4) In the automotive area: in developing of self-driving cars which can adapt to their driver's mood in order to offer assistance on the road and avoid any traffic incidents. In this case, based on the facial expressions of the driver (and perhaps other metrics like how much yawning is involved), the system can make suggestions like taking a break, changing the music or the temperature.

5) In advertising and market research: in receiving feedback from a target group to a commercial which is intended to be put on the market or from a new product that is intended to hit the market. This way, it would become easier to get a more real feedback rather than just creating log questionnaires. Another case would involve displaying in public places some personalized advertisements according to the passers-by's gender. [2]

6) In the medical field: in developing devices which can assist people with eyesight deficiencies in their day-to-day interactions. Another usage would be in creating automated systems which are unbiased by potential human error for examination of different cognitive deficiencies like schizophrenia or autism [3][4].

Our objective is to implement a tool which is able to recognize the facial expressions and gender in humans which has an interface for mobile devices, while the image processing is delegated to a powerful processing system. The challenge is to obtain a near real time data transfer along with a good recognition rate.

In this paper, we refer to gender as being the biological sex (male/female) based on anatomical markers that are generally easily recognizable by humans. We do not address the gender identity, since is not present in the two datasets that we use, this aspect being difficult to predict from images, because it depends generally on what a person identifies with. Furthermore, the dataset that we use for gender classification does not offer any information on intersex persons, just a label (male/female) for each image, this being the predominant gender. In practice, the system presented here displays two probabilities: one for male and one for female their sum being 1, which can correctly characterize any presence of both male and female features.

The rest of the paper is organized as follows. Section II presents a general description of the automated system for facial expressions and gender recognition with a detailed workflow. Section III introduces the two datasets used in the training process. Sections IV and V present the steps taken in developing and training of each neural network. Section VI presets the experimental results with the performances of the

two neural networks. The conclusions and future improvements are synthesized in Section VII.

II. Automated Recognition of Facial Expressions and Gender

The system presented here uses three neural networks: a pretrained one for face detection, one trained for classifying facial expressions and a third one which uses transfer learning from the previous one for classifying gender.

This system is created in two versions in order to decrease the processing time as much as possible and to be able to run on any platform that supports a browser.

The *client-server* architecture was used in order to be able to run the application on any platform that supports a browser, like phone or tablet which usually do not have processing power necessary for neural networks like a dedicated GPU. In this case, frames are collected from the client application (which runs on the phone or tablet etc.) and are sent to a powerful server which using a dedicated GPU does the processing and returns the predictions.

The clear advantage here is portability, but it also comes with a disadvantage meaning that the transmission of the frame back and forth from the server depends on the network's latency. This can lead to fewer frames to be collected and processed, influencing the desired real time experience.

The *desktop* application represents the version of the application that uses just Python code to run natively. In this case, a computer with a dedicated GPU was used for both training of the neural networks and making predictions in a desktop application to classify facial expressions and gender, leading to good performances regarding processing time.

The general process of the proposed automated system for classification of facial expressions and gender is displayed in Fig. 1. and consists of multiple steps for each application version:

A. Client-server architecture workflow:

- From the camera of the device on which the client application is running, each frame is collected every 500 milliseconds. This time interval was chosen to make sure that there is enough time for the server to process the frame request and return an answer without altering the desired real time user experience. A lower interval time between frame capturing would mean that the server will queue all the incoming frame requests in order to process them sequentially. This would lead to a delayed feedback for the end user of the application.
- A POST request is done to the server containing the frame encoded in base 64.
- When the server responds to the request, it gets the frame, decodes it from base 64 to matrix array representation and passes it to a neural network which is specialized in detecting faces. This network is pretrained and available for usage in OpenCV library [5]. This particular neural network was chosen because it offers better performances in detecting small faces than other selected methods like Haar Cascade, HoG face Detector and deep learning-based face detector from Dlib library [6].
- The pretrained neural network for face recognition returns the coordinates in which one or multiple faces can be found in the image, so their bounding boxes can be created. If there are multiple faces detected, the first found will be kept and its bounding box will be cut from the original frame.
- Then, the image containing the face will be preprocessed: converted to grayscale and to resized to 48x48 pixel dimension in order to pass it through the two neural networks: one for recognizing facial expressions and one for gender classification.
- Over the original image, there will be put a bounding box corresponding to the face's position in the image, alongside the probabilities for each of each emotion and gender.
- Finally, the image containing the answer of the two neural networks will be encoded back to base 64 and set back to the client application.

B. Desktop version workflow:

- In this case, the frame capturing of the image and the predictions are done in the same place, on the same machine, meaning that a higher number of images will be processed per second. This is because no extra time is required for the frame to be sent back and forth as in the client server architecture. And so, a new frame can be collected immediately after the current frame has been successfully processed.
- The steps performed in this case are almost the same as in the client-server case.
- Each frame is collected and passed through the OpenCV library [5].
- A smaller image containing just the face will be cropped from the original one, converted to grayscale and to 48x48 pixels dimension.
- The facial image will be passed through the two neural networks.
- The answer for each of them with the corresponding bounding box will be overlayed over the original frame which will be displayed immediately to the user.

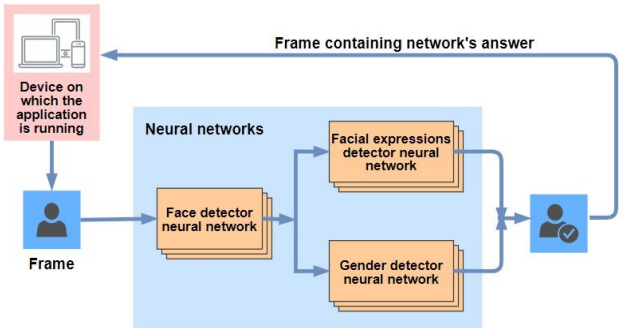

Fig. 1. General diagram of the process of facial expression and gender recognition

III. Training Datasets

Both neural networks use supervised learning, meaning that they learn the facial characteristics needed in facial expressions and gender recognition from training sets which

contain images with different facial expressions and genders labeled accordingly.

A. Fer 2013

This image set was used for training the facial expression recognition neural network which was used in the Kaggle competition called *Challenges in Representation Learning: Facial Expression Recognition Challenge* from 2013 [7]. It was created using Google image search API and has a collection of 32,298 grayscale images 48 by 48 containing faces which express one of the 7 basic emotions: happy, sad, angry, disgust, surprise, fear and neutrality as displayed in Fig. 2 [8].

Fig. 2. Images from FER2013 dataset

In the current example, FER2013 dataset was split into 3 parts:

a) Training set: containing 70% of the entire dataset which represents the set that was used by facial expressions recognition neural network to learn.

b) Validation set: containing 15% of the entire dataset, which did not interfere in the training process (the model did not use any of these images in order to be trained). Using this set, the accuracy and loss function were computed at the end of each epoch in order to observe the generalization capacity of the model. If we were to use just the training set as a marker, the model may have done overfitting and that would not have been easily observable. And so, the validation set helps to ensure the fact that the model keeps increasing its capacity of correct prediction on unseen data.

c) Test set: containing 15% of the entire dataset which just as the validation set, does not interfere in the training process. At the end of the training process, the accuracy and loss function on this set will be computed to better characterize the model performances on unseen data.

B. UTKFace

It is an image set used in training the gender classification neural network. It contains an approximate of 20,000 images of different gender, age and ethnicity labeled accordingly [9]. A snippet of this dataset is displayed in Fig. 4. In the current case, since we are interested in gender prediction, we used just the gender label in the training process.

Fig. 4. Images from UTKFace dataset

IV. Facial Expressions Classification

This paper, alongside a neural network for gender classification, presents a convolutional neural network trained to be capable of recognizing facial expressions. Due to the fact that we want to classify images, the neural network consists of 2 main parts:

- Convolutional layers with their adjacent layers (batch normalization and max pooling) needed in feature extraction from images. We used convolutional layers instead of just dense layers at this point, because since we work with images, the spacial component is needed to be kept. It has also been observed in practice the fact that using convolutional layers instead of dense ones (which

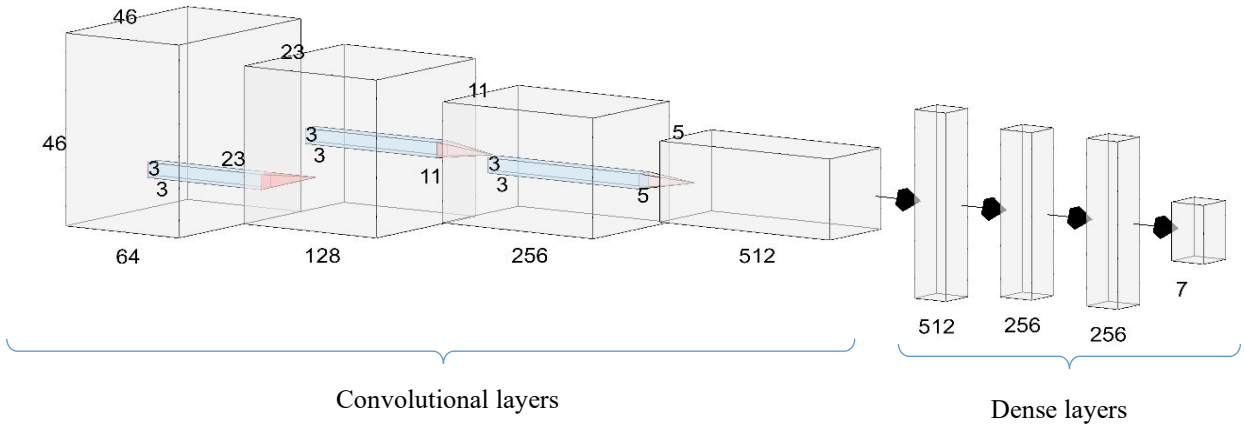

Fig. 3. General architecture of the neural network for classifying facial expressions

flattens the entire image in a long 1D vector before passing it through the network) results in better performances [10]

- Dense layers which perform the actual image classification, helps mapping the input (image) with the output (7 probability values, one for each class: happy, sad, angry, disgust, surprise, fear and neutrality).

The trained neural network for recognizing facial expressions represents an improvement of the model for recognizing facial expressions proposed in [11] with the purpose of increasing its performances, more precisely its accuracy on the validation and testing dataset. To make this possible, multiple methods were used:

a) Changing the neural network's architecture: by modifying the number of convolutional, batch-normalization, max pooling, dropout and dense layers. Also, each layer's parameters were modified in different configurations:

- For the convolutional layers, the number of filters (or kernels) and padding were modified.
- For the max pooling layers, the pool size and stride.
- The rate of dropout, to deactivate a smaller or greater number of neurons in order to prevent overfitting.

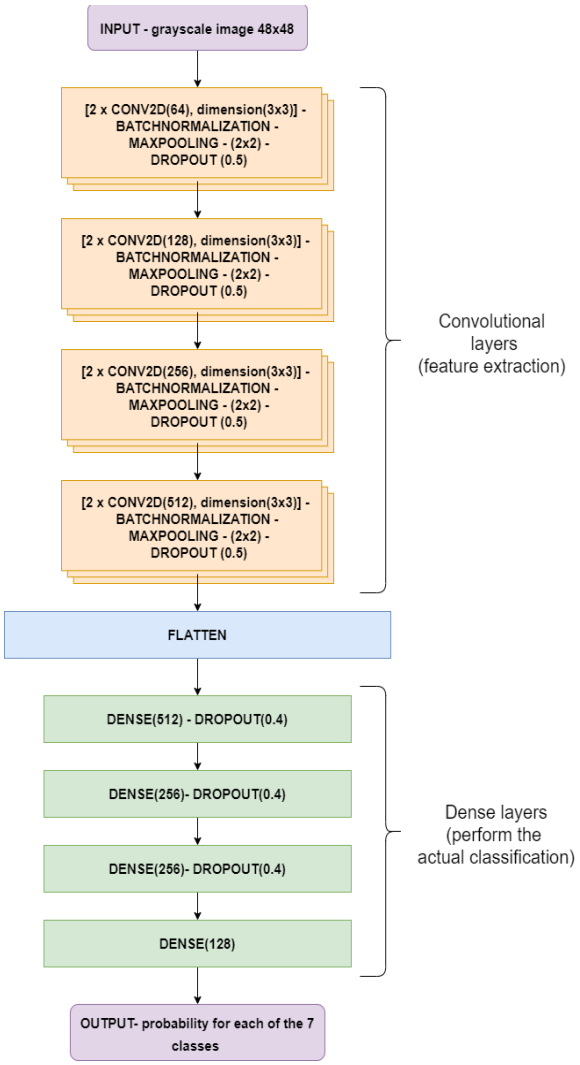

Fig. 5. Detailed architecture for the convolutional neural network for classifying facial expressions

b) Changing the learning rate: lowering or increasing it acccordingly

c) Increasing the number of epochs: to make a better observation on the neural network's ability to learn and generalize the learned data over a bigger period of time. A greater number of epochs on which the model was trained can show exactly the epoch on which it will start to overfit the training dataset (the values of accuracy and loss on the training set will show that the neural network improves, but the ones on the validation and testing set, will show performances reduction) [12].

d) Increasing the value of batch size parameter: which represents the number of images in a batch that are processed at a time before computing the loss function and adjustment of the weights. An increase of this value leads to fewer parameter updates in the training process [13].

e) Using a technique called cross-validation: because a deeper understanding of the model's ability to learn from data and generalize on unseen data is needed. In this case, splitting the entire training set in 3 parts: training, validation and testing is not enough to characterize the model as accurate as possible [14].

To get a beter understanding of the model's capabilities to correctly classify facial expressions, a technique called cross validation is needed in solving underfitting and overfitting problems. This is because from time to time splitting in training, validation and testing data, even randomly (as done in this case) may result in a biased split, leading to an incorrect evaluation of the model.

In the current case, cross validation was applied in the following way:

- The original dataset was split into K (in this case, 10) folds.
- Each of the folds above will become the validation set one at a time.
- The rest of the 9 folds will be used in the actual training and testing set.
- We wanted to also keep the 70%, 15%, 15% (training, validation, testing) split of the entire set, leading to a small disadvantage, meaning that some overlap may exist in the folds. This is because we wanted 2 conditions: 10 folds, and a percentage split.
- And so, we obtained 10 model configurations, each of them with a different combination of training, validation, and testing set.

This way, the influence of the dataset split upon the model could be better observed.

f) Using augmentation: in order to get more and diverse images from the current limited dataset, leading to a better ability to generalize on unseen data [15]. It was applied only for the training set (validation and testing sets were excluded) by mirroring each image on the vertical axis.

In the end, after numerous combinations of layer displacements and hyperparameter values, the network architecture looks like in Fig. 5.

V. Gender Classification

The neural network for classifying gender is trained to perform a binary classification (biological male or female) using facial images. Its input is the same as in the previous neural network for facial expressions: a 48x48 grayscale image containing just a face (the rest of the image being already cropped from the original image using the bounding box from generated by the neural network for face recognition).

A. Transfer learning

This method consists in using the information learned by a neural network in the process of training another one. For the method to be applicable, it is necessary that the two neural networks try to solve a related problem. For example, in developing a classification algorithm, a complex deep neural network with numerous layers and parameters like VGG16 or VGG19 trained on ImageNet dataset [16] can be used in classifying animals [17].

This method can be applied by using the pretrained neural network's weights as a starting point in training of the new one (fine-tuning), or by adding some pretrained neurons with its corresponding weights to the new neural network. The neurons from the base neural network will be "frozen", meaning that their weights will not be changed during the training [18]. In this case, only the new layers will adapt their weights to better classify the given dataset.

B. Application in this case

In this case, the pretrained weights of the convolutional layers from the facial expressions neural network were frozen and attached to the neural network specialized in gender recognition. Only the remaining dense layers are be trained. The method of transferring the knowledge from one neural network to another worked, because the weights of the convolutional layers that that were transferred, were tuned to just extract facial features from images regardless the scope of extraction (see Fig. 6). This facial image feature extraction represents necessary element in both classifying facial expressions and gender.

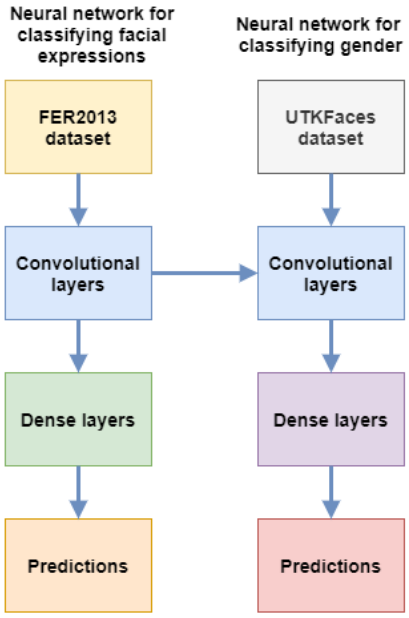

Fig. 6. General architecture of the convolutional neural network for classifying gender

In the case of the current network, the same steps in the search for better performances were done:

a) Changing the neural network's architecture: in this case, only a variation of number of dense and dropoout layers was performed alongside the parameters for each of them.

b) Changing the learning rate: different values for it as necessary

c) Increasing the number of epochs: as before, to observe the performances of the neural network and when it starts to overfit the training dataset

d) Increasing the value of batch size parameter: or decreasing it as needed

e) Using the technique called 10-fold-cross validation: as before, this technique was performed to be able to better observe how the model behaves with different configurations for training, validation and test set using the same (70%, 15%, 15% configuration for training, validation and test set, as for the other neural network).

In this case, augmentation for the training dataset wasn't performed since it is a binary classification, the need for extra training data was not considered necessary.

A practical advantage in using Transfer Learning was the fact that the necessary time to train the 10 models for the facial expression recognition neural network took on average an hour and 20 minutes, while for the second neural network, the one for classifying gender, only 12 minutes were necessary on average to train. This phenomenon was possible due to the fact that for the second neural network, only the dense layers were trained, and so, the number of computations necessary was considerably lower.

VI. Experimental Results

After completing the training and using cross-validation, 10 models for each neural network were obtained, each of them with a different dataset.

A. Neural network for classifying facial expressions

In the image bellow it can be observed the evolution of accuracy and loss on the training and validation set for each of the 10 models trained using the cross-validation method. The metric that captures the best the performance of the model here is the loss on the validation set.

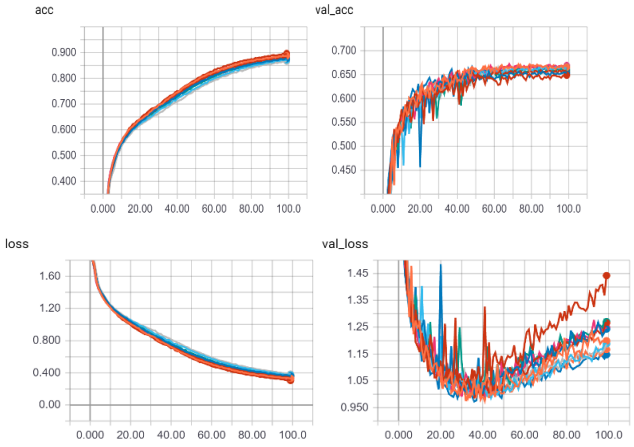

Fig. 7. The values for accuracy and loss function on the training set (left) and validation set (right) during training of the 10 neural netrock models for classifying facial expressions on 100 epochs.

In the training process the best model was saved, which is the model with the lowest value of the loss function on the validation set. In time, the loss value on the validation set increases (meaning that the model starts to overfit the training dataset), and so, the model with the best performances was obtained around epoch 30.

B. *Neural network for classifying gender*

On the same idea as above, for the second neural network 10 models were obtained. In this case, the best model can be achieved around epoch 90 out of 100.

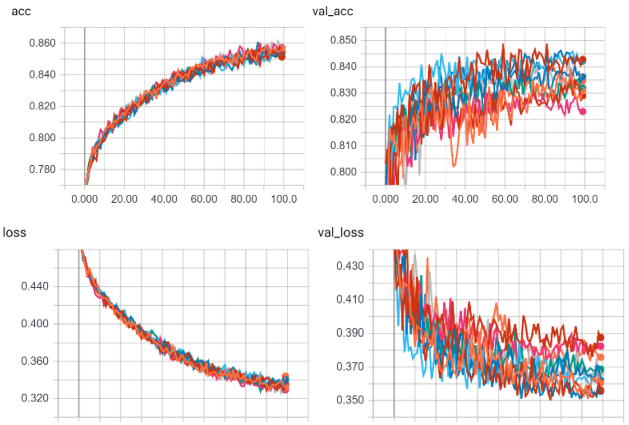

Fig. 8. The values of accuracy and loss function on the training set (left) and validation set (right) during training of the 10 neural networks for gender classification on 100 epochs.

C. *Unrelated training datasets*

The system is built to display to the user the predominant emotion and the gender from one facial frame.

In empirical tests performed in laboratory conditions (decent lightning, one face in the image), there has been observed that both neural networks perform well: a correct classification of a male or woman showing sadness, surprise, fear, disgust, angriness, or no emotion (neutrality).

An interesting phenomenon takes place when the predominant emotion displayed in the image is happiness. The neural network for classifying facial expressions makes a correct prediction with 100% confidence, while the gender classification neural network is tricked into predicting that there is a woman in the picture.

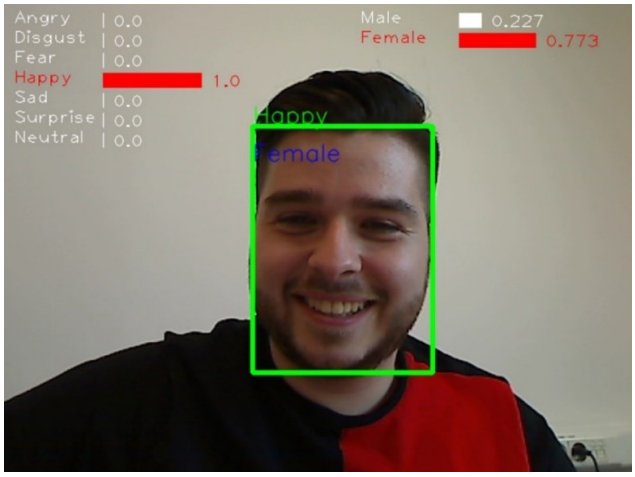

Fig. 9. The answer of the two neural networks when happiness is displayed by a male.

For no other facial expression, it has been observed that the gender classification neural network can be tricked this much, that it can predict incorrectly with a 73% confidence that in that picture is a woman instead of a man as in Fig. 9.

Although this phenomenon may seem weird, the explanation behind it is linked to the fact that the two neural networks have been trained on different datasets which are mainly unrelated.

The UTKFace dataset [9] that was used, does not contain any information regarding facial expressions, or any equal distribution of pictures by gender across all 7 pursued facial expressions. This way, the faces in the UTKFace dataset [9] do not offer any clear information about the facial expressions necessary to for the dense layers of second neural network to try to grasp in order to better classify how a surprised man or how a happy woman looks like. In this case, the best prediction of the gender classification neural network is possible when the displayed facial expression is neutrality.

D. *Summary of the two neural networks' performances*

In order to get some parameters who can characterize as good as possible both neural networks, an average of the performances of all 10 models was computed.

The table below synthesizes the performances of the two classifications models.

TABLE I. EXPERIMENTAL RESULTS OBTAINED BY AVERAGING THE ACCURACY AND LOSS PARAMETERS OF THE 10 MODELS OBTAINED FOR EACH OF THE NEURAL NETWORKS

	Neural network for classifying facial expressions		**Neural network for classifying gender**	
	Accuracy	***Loss***	***Accuracy***	***Loss***
Training set	0.88101	0.35288	0.85375	0.33466
Validation set	0.64281	0.99263	0.83929	0.36978
Test set	0.64195	0.99747	0.83297	0.36978

VII. CONCLUSION AND FUTURE IMPROVEMENTS

In this paper, we presented an automated system based on two convolutional neural networks which is capable to recognize 7 basic facial expressions and the gender of a person. It comes in two versions: a desktop one, and a web one in order to make it available on multiple platforms. The trained models behave well in the tested cases and can be used in developing a more complex mechanism capable of evaluating a series of new parameters like age or head and eyes orientation. It can be used in multiple domains, like programming, advertising, medicine and automotive to assist drivers. Therefore, neural networks as the ones that were presented are the starting points in more complex systems that can save human lives.

Although system presented here offers good performances in classifying facial expression and gender, there are several improvements that can be done in order to make it more robust:

1) For the client- server application:

a) Using sockets: to send/receive the image faster between client application and server for a better user experience

b) Compressing the frame that is sent to the server: to reduce its size and make a faster transfer of data between client and server

c) Cropping the face from the original frame: this can be done on the client side, in order to reduce the payload size before sending the it to the server, leading to a reduced transmission time

d) Adding neural networks capable of classification from video images: this would allow using UDP or DCP protocols for transfering data in real time between client and server with great speed and minor loss loss [19], leading to a noticeable decrease in latency for data transmittion.

2) For the neural networks:

a) Appending new datasets to the existing ones: for example: AFEW 7.0 [20], SFEW 2.0 [21], TFD [22] for the nerual network which classifies facial expresssions

b) Applying transfer learning: for both neural networks using layers from pretrained convolutional neural networks which use state-of-the-art architectures and datasets like InceptionV3 [23] or VGG [17].

c) Comparison with other implemented systems: further evaluation of effectiveness and efficiency of our current system in relation with the existing ones in order to establish its domain of performances.

REFERENCES

[1] A. Kolakowska, A. Landowska, M. Szwoch, W. Szwoch, M. R. Wróbel, "Emotion recognition and its applications." Human-Computer Systems Interaction: Backgrounds and Applications 3. Springer, Cham, 2014. 51-62.

[2] A. Singh "Facial Emotion Detection using AI: Use-Cases", 27 April 2018, https://blog.paralleldots.com/product/facial-emotion-detection-using-ai/ (accessed March 1, 2021).

[3] M. Leo, P. Carcagnì, C. Distante, P. Spagnolo, P. Mazzeo, A. Rosato, S. Petrocchi, C. Pellegrino, A. Levante, F. De Lumè, and F. Lecciso, "Computational Assessment of Facial Expression Production in ASD Children," Sensors, vol. 18, no. 11, p. 3993, Nov. 2018.

[4] M. Leo, P. Carcagnì, C. Distante, P. L. Mazzeo, P. Spagnolo, A. Levante, S. Petrocchi, and F. Lecciso, "Computational Analysis of Deep Visual Data for Quantifying Facial Expression Production," Applied Sciences, vol. 9, no. 21, p. 4542, Oct. 2019.

[5] V. Gupta, "Learn Opencv", Github Repository, 2018, https://github.com/spmallick/learnopencv/tree/master/FaceDetectionComparison/models (accessed March 12, 2021).

[6] V. Gupta, "Face Detection – OpenCV, Dlib and Deep Learning (C++ / Python)", 22 October 2018, https://learnopencv.com/face-detection-opencv-dlib-and-deep-learning-c-python/ (accessed March 2, 2021)

[7] Kaggle competition, https://www.kaggle.com/c/challenges-in-representation-learning-facial-expression-recognition-challenge/overview, 2013, (accessed March 12, 2021)

[8] I.J. Goodfellow, D. Erhan, P. L. Carrier, A. Courville, M. Mirza, B. Hamner, W. Cukierski, Y. Tang, D. Thaler, D.-H. Lee, Y. Zhou, C. Ramaiah, F. Feng, R. Li, X. Wang, D. Athanasakis, J. Shawe-Taylor, M. Milakov, J. Park, R. Ionescu, M. Popescu, C. Grozea, J. Bergstra, J. Xie, L. Romaszko, B. Xu, Z. Chuang, Y. Bengio "Challenges in representation learning: A report on three machine learning contests", International conference on neural information processing (pp. 117-124), Springer, Berlin, Heidelberg, 2013, November.

[9] Z. Zhang, Y. Song, and H. Qi, "Age progression/regression by conditional adversarial autoencoder.", Proceedings of the IEEE conference on computer vision and pattern recognition, 2017.

[10] A. Rosebrock, "Keras Tutorial: How to get started with Keras, Deep Learning, and Python", 10 September 2018, https://www.pyimagesearch.com/2018/09/10/keras-tutorial-how-to-get-started-with-keras-deep-learning-and-python/ (accessed March 1, 2021).

[11] F. Kinli, "[Deep Learning Lab] Episode-3: fer2013", April 6, 2018, https://medium.com/@birdortyedi_23820/deep-learning-lab-episode-3-fer2013-c38f2e052280 (acceses March 13, 2021) .

[12] D. Hawkins, "The Problem of Overfitting", Journal of chemical information and computer sciences. 44. 1-12. 10.1021/ci0342472, 2004

[13] L. S. Smith, P.-J. Kndermans, C. Ying, Q. V. Le "Don't decay the learning rate, increase the batch size.", arXiv preprint arXiv:1711.00489, 2017.

[14] D. Berrar, Cross-Validation, 10.1016/B978-0-12-809633-8.20349-X, 2018.

[15] C. Shorten, M. K. Taghi, "A survey on image data augmentation for deep learning.", Journal of Big Data 6.1 1-48, 2019.

[16] O. Russakovsky, J. Deng, H. Su, J.Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, L F. Fei,. "Imagenet large scale visual recognition challenge.", International journal of computer vision 115.3: 211-252, 2015.

[17] K. Simonyan, A. Zisserman. "Very deep convolutional networks for large-scale image recognition." arXiv preprint arXiv:1409.1556, 2014.

[18] D. Sarkar, "A Comprehensive Hands-on Guide to Transfer Learning with Real-World Applications in Deep Learning", November 15, 2018, https://towardsdatascience.com/a-comprehensive-hands-on-guide-to-transfer-learning-with-real-world-applications-in-deep-learning-212bf3b2f27a (accessed March 13 2021).

[19] S. A. Nor, R. Alubady, W. A. Kamil, "Simulated performance of TCP, SCTP, DCCP and UDP protocols over 4G network",Procedia Computer Science,Volume 111, 2017, ISSN 1877-0509

[20] A. Dhall, R. Goecke, S. Ghosh, J. Joshi, J. Hoey, and T. Gedeon, "From individual to group-level emotion recognition: Emotiw 5.0," in Proceedings of the 19th ACM International Conference on Multimodal Interaction. ACM, 2017, pp. 524–528.

[21] A. Dhall, O. Ramana Murthy, R. Goecke, J. Joshi, and T. Gedeon, "Video and image based emotion recognition challenges in the wild: Emotiw 2015," in Proceedings of the 2015 ACM on International Conference on Multimodal Interaction. ACM, 2015, pp. 423–426.

[22] J. M. Susskind, A. K. Anderson, and G. E. Hinton, "The toronto face database," Department of Computer Science, University of Toronto, Toronto, ON, Canada, Tech. Rep, vol. 3, 2010.

[23] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, Z. Wojna, "Rethinking the inception architecture for computer vision." Proceedings of the IEEE conference on computer vision and pattern recognition, 2016.