Third assigment
Emotion recognition

Živa Škof

Faculty of computer in information science, University of Ljubljana

Emotion recognition is an active research area because it is leading towards the improvement of human-computer interaction. Interest is growing in providing better services, neuromarketing, robots that are capable of sociable interaction etc... In this paper, facial expression recognition is tackled. CNN is used to recognize the emotions of faces that are detected with cascades. The goal is to recognize and distinguish seven different emotions. The accuracy of correctly recognized emotions is around 64 percent. Happiness is the most dominant emotion when you are using the live camera to recognize emotions.

1 Introduction

Interaction between people is through speech and body gestures both accompanied by facial expressions. Whit facial expressions our emotions are expressed which is an important part of communication and interaction. As there is a saying that a picture can tell more than a thousand words the same goes with nonverbal communication.

Computers and robots have become part of our everyday life. Their capability to recognize emotions improves the interaction and makes it more realistic. Emotions are evoked by various factors therefore they can be hard to predict. Prediction can be improved with machine learning.

In this paper, the training of CNN for emotion recognition is addressed. For recognition already existing CNN model is used [1].

2 Related work

Researchers have already done some remarkable models for facial expression recognition. In the article [2] they managed to get the accuracy of 93 percent. They got the bests results with a subset of Gabor filters using AdaBoost and training Support Vector on the outputs. In the following 3 articles, facial expression recognition was also made on the FER2013 database. In article [3] 57.1 percent accuracy is achieved, in article [4] 62.44 and 65.03 percent of accuracy in article [5]. The model used was also tested and trained on the FER2013 database and it is described in the following article [1] with the accuracy of 63.167 percent which ranked in the top 10 in Kaggle competition.

3 Methods

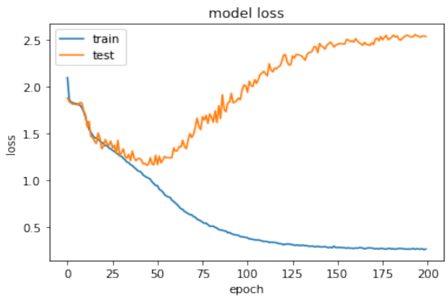
In this paper, a CNN using the FER2013 [6] dataset to differentiate between six groups of emotions - anger, disgust, fear, happiness, sadness, surprise, and neutral - will be trained. Dataset contains 35,888 images that are classified in previously mentioned six groups of emotions and represented with numbers from 0 - 6 in the same order. Each image is represented with 2,304 pixels intensity values and is denoted whether it will be used for training or testing. All images are gray-scale and 48x48. The extracted data is rearranged into an array of the same dimension, converted into unsigned integers, and normalized. After these steps are done all the labels are hot encoded.

For creating a Sequential CNN Keras is used - CNN will be a linear stack of layers. CNN has the following components:

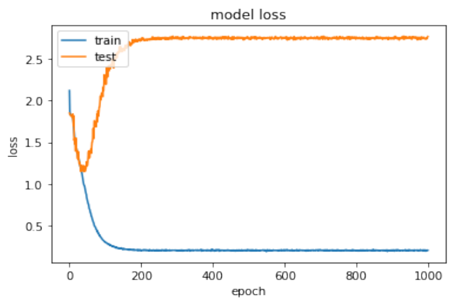
- Convolutional layers.
- Activation functions.
- Pooling layers.
- Dense layers.
- Dropout layers.
- Batch normalization.

Adam is used to compile our CNN. Data will pass the model 1000 times (epochs). For every epoch the following functions are called: ReduceLROnPlateau, EarlyStopping (while running the model multiple times early stopping is ignored since it means that in the majority of the time when progress stalls for some time there will not be any progress. Just to be sure there will not be any or at least not an important progress) and ModelCheckpoint. The batch size was 64 and the learning rate 0.001. 20 percent of training data were intended for model validation. In the end, the model and its weights were saved.

The code can be found on the following URL: https://github.com/sziva/SB_3

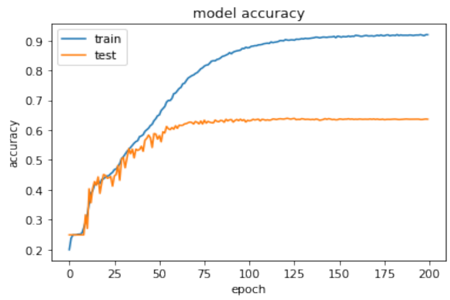


(a) epochs=200, batch size=64 and learning rate = 0.001

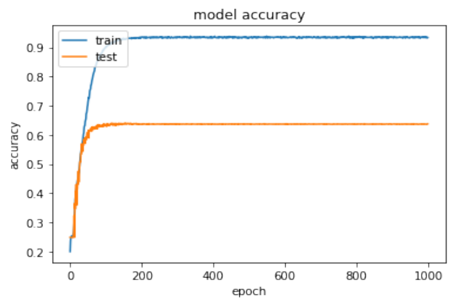


(b) epochs=1000, batch size=64 and learning rate = 0.001

Figure 1: Loss curves



(a) epochs=200, batch size=64 and learning rate = 0.001



(b) epochs=1000, batch size=64 and learning rate = 0.001

Figure 2: Accuracy curves

4 Results and Discussion

For evaluation of our CNN, the next performance curves were analyzed.

- Loss for train and test data.
- Accuracy for train and test data.
- CMC curve

From a loss curve that is shown in figure 1 improvement in number is a sign of a better model. It shows that the loss value stoped improving around 200 epoch what means that model will most likely not improve from this point on but it will get worse or stay the same.

From accuracy curve shown in a figure 2 it can be seen that the model reaches its maximal accuracy already before 100 epoch. In the following epoches only almost visiable improvements occure. The model can be stoped with EarlyStop but instead it ran on just to make sure the the future improvements will acctualy not accure.

The CMC (Cumulative Match Characteristic) curve on figure 3 shows that the score is improving with rank. From rank-2 to rank-4 CMC curve is increasing linearly. The accuracy of more than 90 percent occurs for rank-4, it means that wanted emotion is in top 4 detected emotions with the biggest weight.

5 Conclusion

Accuracy of 63 percent might not seem much and I agree. It gives a red flag that there is room for improvement. But on the other hand, if you take a look on the leaderboard of FER2013 the accuracy is not that bad since it came into the top 10. It is still 7 percent lower than the one in the first place but it is not as bad as it looks. To improve CNN different data set could be used. In some papers, there is information that some data from FER2013 are wrongly annotated.

For future work fusion of facial expression emotion recognition with speech emotion recognition could be tackled.

References

[1] J. Paul, "How i developed a c.n.n. that recognizes emotions and broke into the kaggle top 10." [Online]. Available: https://www.freecodecamp.org/news/facial-emotion-recognition-develop-a-c-n-n-and-break-into-kaggle-top-10-

[2] G. Littlewort, M. S. Bartlett, I. Fasel, J. Susskind, and J. Movellan, "Dynamics of facial expression extracted automatically from video," *Image and Vision Computing*, vol. 24, no. 6, pp. 615 – 625, 2006, face Processing in Video Sequences. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0262885605001654

[3] V. Tümen, F. Söylemez, and B. Ergen, "Facial emotion recognition on a dataset using convolutional neural network," in *2017 International Artificial Intelligence and Data Processing Symposium (IDAP)*, 2017, pp. 1–5.

[4] K. Liu, M. Zhang, and Z. Pan, "Facial expression recognition with cnn ensemble," in *2016 International Conference on Cyberworlds (CW)*, 2016, pp. 163–166.

[5] A. Agrawal and N. Mittal, "Using cnn for facial expression recognition: a study of the effects of kernel size and number of filters on accuracy," *The Visual Computer*, vol. 36, no. 2, pp. 405–412, 2020.

[6] Kaggle, "Fer2013." [Online]. Available: https://www.kaggle.com/c/

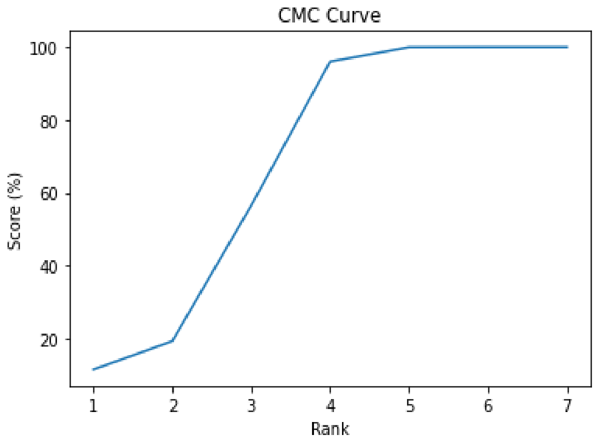


Figure 3: CMC curve for rank-1 ... rank-7, for a model with epochs=200, batch size=64 and learning rate = 0.001

challenges-in-representation-learning-facial-expression-recognition-challenge/overview