Facial Expression Classification

Arda Kaya

Abstract—This paper is explaining two different projects. One of them is developing a classifier to recognize 6 different facial expressions which are anger, disgust, fear, happiness, sadness and surprise. It will show the data types, network choices for training and accuracy results. Finally it will demonstrate how to increase accuracy to 67 percent by using facial features. Another project is classifying products from given image dataset.

Index Terms—Robot, Facial Expression Classification.

1 Introduction

The first project's main goal is classifying series of the six universal emotions as defined by Psychologist Paul Ekman: Anger, Disgust, Fear, Happiness, Sadness and Surprise. This kind of classifier would be used with speech recognition models to check speaker's level of confidence, confusion or desperation while speaking. Some of the researches on this topic already done and turned into API services. Some of them Microsoft's Emotion API, Imotions and IBM Watson-uses voice tone to classify. Another project is classifying products from given image dataset. There are three classes which are Bottle, Candy Box and Nothing. The general use of this kind of classifiers is sorting products for storing or packaging.

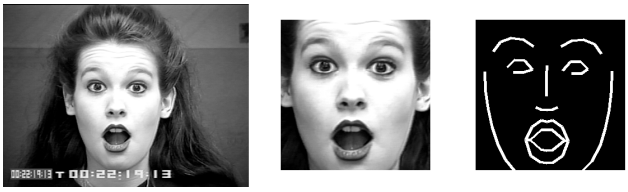

Fig. 1. Training Samples.

2 Background / Formulation

For emotion classifier two different networks were used ; AlexNet and LeNet. Both of them are convolutional networks. AlexNet is a deep convolutional network and designed for 224x224 ImageNet dataset. LeNet is desgined for grayscale 28x28 - MNIST dataset and main purpose was classifying handwritten letters. The similarity between handwritten letters and facial features gave the inspiration to use LeNet for this project. The epoch number was 20 because accuracy value was not progressing after that value. For product classifier GoogleNet network was used. It is designed for ImageNet dataset similar to AlexNet, however it is more accurate with 6.66 percent error rate and more optimized with 22 layers, which is 12x less parameters.

AlexNet parameters:

- Image size : 256x256
- Image type: Grayscale
- Training Image number: 292
- Epoch number: 20

LeNet parameters:

- Image size : 28x28
- Image type: Grayscale
- Training Image number: 315
- Epoch number: 20

Product classifier GoogleNet parameters:

- Image size : 224x224
- Image type: 3 Channels
- Epoch number: 8

3 Data Acquisition

The training data is provided by Cohn-Kanade. Originally images in database are sequences and they recorded while performers doing the facial expressions(Fig.2). To make the classification easier, just the last images used for the training(Fig.3).

Images in database have been cropped to get rid of unnecessary information in the background(Fig.4). As the face positions were different in every image, face recognition library is used to optimize the process.

Fig. 2. Image Sequence

Fig. 3. Last Image

To prepare data for LeNet network, cropped images transformed into facial features on black background(Fig.5). To automate this process face recognition library is used.

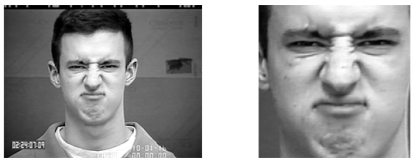

Fig. 4. Automated Cropped Image

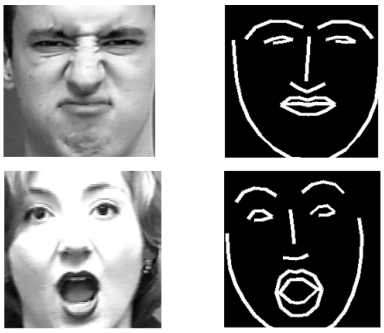

Fig. 5. Automated Cropped Image

4 Results

AlexNet's accuracy was 47. And LeNet's accuracy was 67. You can find training graphs in Fig.7 and Fig.8. LeNet's training time was 2 times faster than AlexNet's. Also,

Fig. 6. AlexNet Test Result

LeNet's accuracy values were progressing less spiky, according to AlexNet's.

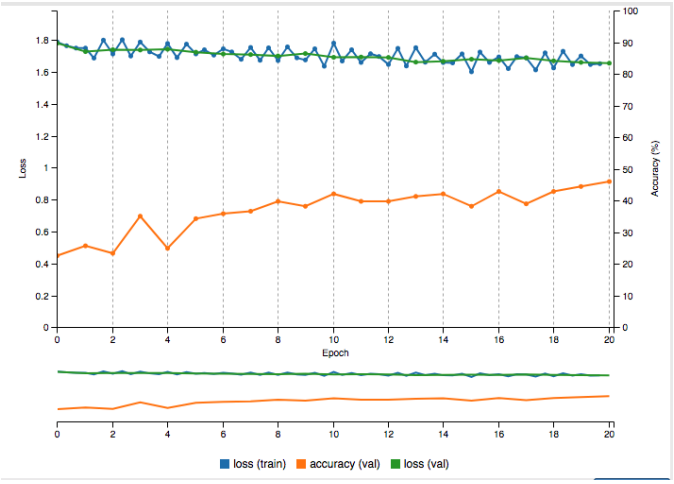

Fig. 7. AlexNet Training Result

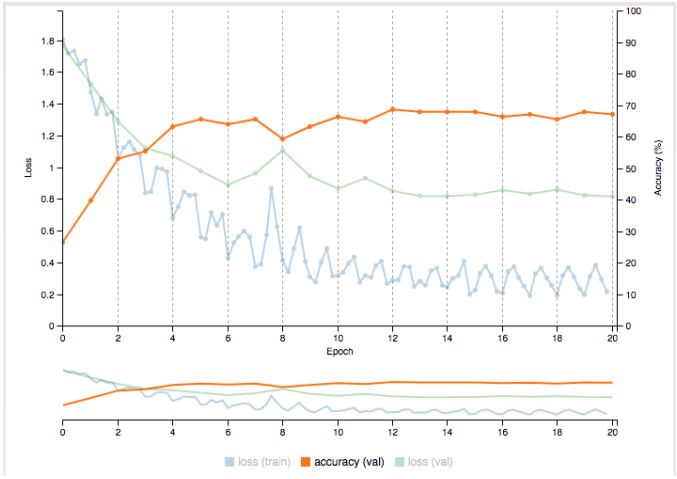

Fig. 8. LeNet Training Result

The results of the GoogLeNet network for product classification are:

- Inference time : 5.3
- Accuracy: 75.4

You can find watermarked screen-shot in Fig.9

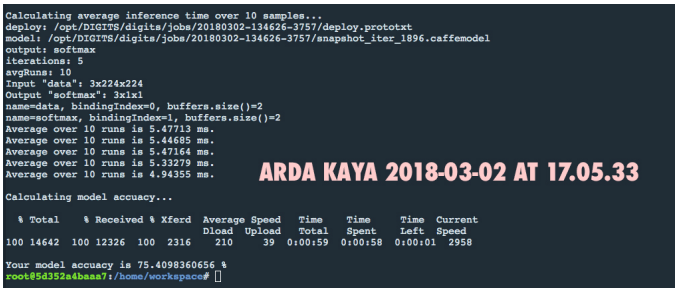

Fig. 9. Watermarked Screenshot

5 Discussion

The training images for LeNet has just lines that fits on facial features which are eyebrows, nose, eyelids and chin. This isolates most of the unnecessary data from the image and improves accuracy. However, this method has one problem. In the Fig 9, details on skin of chin and nose area gives important clues to differ anger from disgust expression. The isolated training images have no surface details, so the model does not use these clues for classification. The solution for this problem would be creating another model that can classify surface features and using with LeNet together.

6 Conclusion / Future work

This method classifies only images with 67 percent accuracy. Next stage will be adapting it to real-time videos. And experimenting with LSTM network and sequential training data to classify more complex emotional status such as anxiety level or detecting a lie.

Fig. 10. Anger and Disgust

REFERENCES

[1] T. . Kanade, Cohn and L. et al. (2010), *Cohn-Kanade Database*.
[2] A. GEITGEY, *face recognition*.

[1] [2]