KAIST

Facial Expression Recognition
Project Proposal

Abbas Mammadov **20200844** | **Zahra Bayramli** **20200812** | **Kaleb Mesfin Asfaw** **20200805** | **Tivan Varghese George** **20200829**

School of Computing
Korea Advanced Institute of Science and Technology

1 Introduction and Motivation

Human facial expressions reveal a lot of crucial, hidden, information, such as their emotional states and intentions. Recognizing the emotions of a person will help us understand the wishes and behavior of people. Considering the importance of interaction between humans and computers, it becomes necessary to develop a classifier to determine a person's emotion through their facial expression. Even though distinguishing the fundamental expressions by computers has been developed with a level of higher accuracy in simple environments, there is still room to improve on in terms of classifying emotions from facial expressions in a real world environment. Thus we aim to create an efficient, applicable model to work in real world environments, that classifies emotion based on facial expressions; independent of the age, ethnicity, or the appearance of a person.

2 Methods

2.1 Dataset

At the first stage, we are planning to use the FER2013 dataset, which was collected automatically by the Google image search API. This large-scale data, which is available on Kaggle, has been studied well and played a crucial role in various research papers. The dataset consists of 35887 grayscale images of faces, which have been normalized to the 48x48 pixel scale. The train, validation, and test set ratio is 80:10:10 (train- 28709, validation- 3589, test- 3589). Each grayscale image in the dataset is categorized into one of the seven categories- Angry, Disgust, Fear, Happy, Sad, Surprise, and Neutral.

2.2 Models

In this application project, we will propose several deep learning models in order to implement Facial Expression Recognition. As a base model, we are planning to refer to one of the most developed image recognition models- Residual Network (ResNet), which is an architecture that uses multiple CNN blocks. The main goal is to get higher accuracy compared to the existing published papers. In order to achieve this and avoid any kind of potential over-fitting, we will also use a data augmentation, as well as dropout for regularization. Moreover, we will try to apply a learning scheduling method to ensure the optimization and generalization of the model. As an optimizer, we plan to use Adam or SGD optimizers. To efficiently initialize weights we plan to use He Initialization. The output layer will use the softmax activation function, since this is a multi-classification problem. For the same reason we will use cross entropy as the loss function.

,

3 Intended Experiments

To validate our dataset, we are planning to use other emotional and facial expression depicting datasets, including JAFFE (The Japanese Female Facial Expression), CK+ (The Extended CohnKanade), and some proportion of TFD (The Toronto Face Database). These datasets also consist of 6 or 7 labels as in the main dataset. For error analysis, we are planning to add a confusion matrix from scikit learn library. By comparing the accuracies and F1 scores, we will be able to observe how the model performs in different environments. We are also planning to use the k-fold cross validation method with the help of computational resources (GPU).

References

[1]FER2013- Kaggle Dataset

[2]Prior Research Paper