Facial Expression Recognition using Deep Learning

Snigdha Bhagat
ITCS-5156 Spring 22 - Lee

February 28, 2022

Abstract

It is said that even though words are powerful means of communication, a more sophisticated, age-old, biological and deep rooted way of communication is using facial expressions. In this paper I want to achieve the highest possible accuracy using the FER2013 dataset, but also studying how we can improve the performance of CNN's in the real world.

Referred paper- Khanzada, A., Bai, C., & Celepcikay, F. T. (2020). Facial Expression Recognition with Deep Learning. CoRR, abs/2004.11823(2020), 7. Arxiv. https://doi.org/10.48550/arXiv.2004.11823

1. Introduction

Facial expression is one of the most powerful, natural and universal signals for humans to convey their emotional state. Communication can be both, in verbal and non-verbal form. Facial expression is used in the non-verbal communication. Recognizing emotions in an image with the help of facial data is known as Facial Expression Recognition. Facial Expression Recognition uses algorithms to instantaneously detect faces, code facial expressions, and recognize emotional states. In recent years various applications are using Facial Expression Recognition, and today we need to improve the past work.

1.1. Problem Statement

Most works on Facial Expression Recognition have used fake or simulated data that human models have simulated. This leads to an increased efficiency as those simulated emotions are well formed and are evident. The problem lies when the same is used on real world data. Real world data, as compared to simulated data, has a greater noise, is from a different distribution, and has subtle changes that are not easy to perceive. The goal for this project is to train highly efficient models for the real world.

1.2. Motivation

There is an increased amount of human interaction in today's modern world. Where applications are arranged from image filters to robot assistance. So with the increase in AI applications and image-based contents, there is a surplus need for better expression recognizers.

Also there's an increase in use of AI in healthcare where patient monitoring mechanisms and systems use FER on a daily basis which once again has gained a top spot for future research.

1.3. Solution

This project aims to solve the problem of real world facial expression recognition using transfer learning and novel data augmentation approaches on modified VGG16 and RESNET50 architectures. For transfer learning we are using weights of these models trained on the VGGFace dataset, and to benchmark the results the FER2013 dataset is used.

2. Related Researches

2.1. MicroExpNet: An Extremely Small and Fast Model For Expression Recognition From Face Images

With the increase in use of multimedia, IOT devices require onsite processing of data. There is a dire need of research which will help in creating smaller models that can fit in these devices. Research of Cugu et al., 2017 aims at doing exactly that. They use a knowledge distillation method to reduce the size of the model. The resulting model is smaller than 1mb and although is less accurate than state-of-the art deep models, is a turning point in developing micro-architectures.

2.2. Facial Expression Recognition using Convolutional Neural Networks: State of the Art

Researchers Pramerdorfer & Kampel, 2016 sheds light on the then state-of-the art Facial Expression Recognition techniques and highlight the major algorithmic differences between these researches. They also point out the potential bottlenecks in these methods. They also aim to give direction on how to solve these bottlenecks. Their ensemble of various models is the highest ever accuracy in their time.

2.3 Relation with related works

Both of these works demonstrate the use of CNN's and data augmentation to solve the problems in FER. Using the work of Pramerdorfer & Kampel, 2016, Khanzada et al.,2020 try to resolve some of the bottlenecks in their work. The ensemble of models used in the Pramerdorfer & Kampel, 2016 inspired this project to use models with multiple architectures. We use a custom classifier in the models which is developed by studying the above ensemble. The work of Cugu et al., 2017 helped understand how to make our models faster. We have implemented dropout and frozen some of the layers to prevent overfitting and improve the speeds of models.

3. Methods

3.1. My approach

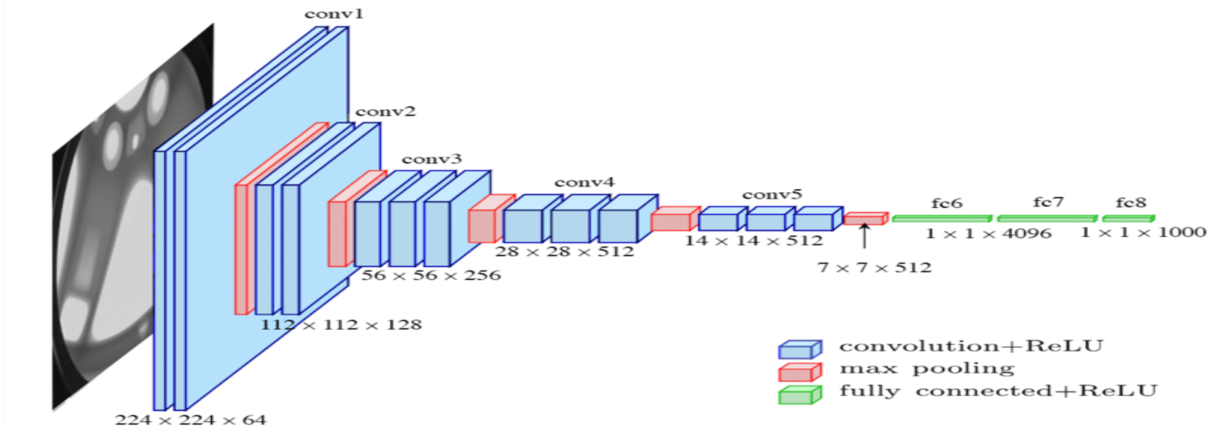
I used 2 very popular image recognition CNN architectures namely VGG16 and RESNET50. These models were pre trained on the VGGFace dataset. I used the pre-trained models of Refik Can Malli who pre trained the models on the VGGFace dataset and achieved the highest ever accuracy.

To use transfer learning I trained these models on the FER2013 dataset by Goodfellow et al. The FER2013 dataset is available as a CSV file containing the attributes as- emotion, pixels(comma separated value of each individual pixel in the matrix), test_train distribution. The data needed to be pre-processed to fit the model's layers.

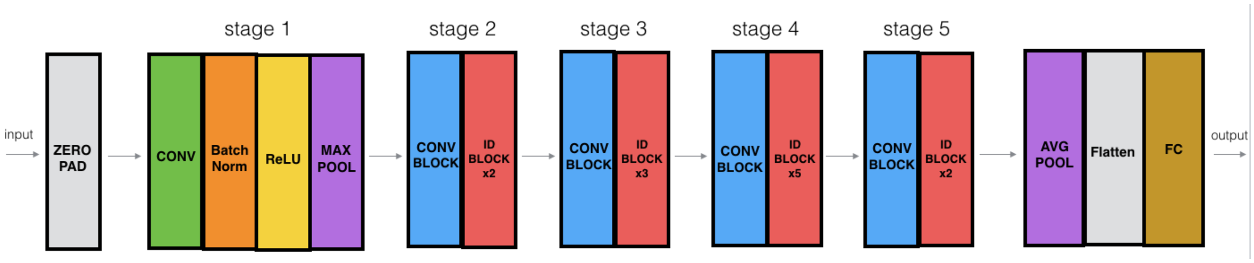
To pre-process each individual row was converted to a png image with the given label and distribution. These images were then converted to grayscale and a size of 194 x 194 to fit the model kernels. Data augmentation techniques like Vertical rotation, horizontal rotation, angle rotation, cropping, zooming, shifting etc were used to augment data. This augmentation was done directly through the keras imageGenerators which create flow images while training and apply these augmentations to the flow of images.

Now, comes the training part. The models were trained for 100 epochs with standard gradient descent and Adam optimizer. A variable learning rate reducer was used in order to decrease the learning rate on plateau. The training was done on the training distribution of FER2013 with validation testing using test_public distribution of the FER2013. It was observed that due to pre-training most models gained a pretty good accuracy in the first 10 epochs. I tried to implement style-augmentation as suggested in the original paper, but it was not clear which style transfer model the authors used and what was the source image for the styles. They had given the augmented data as a downloadable, which no longer exists. In all the only training on a single dataset and naive augmentation, transfer learned models produced comparable results.

3.2. Architecture



VGG16 model architecture with a customer classifier at the end for output classes of 7.



RESNET50 architecture

4. Experiments

4.1 Setup

The trained models were evaluated on test_private distribution of the FER2013 dataset. Each model was also trained using class weights to remove the bias. Each model was separately trained with and without class weights to compare the results. After that for each model, a graph for training, validation accuracy against epochs was plotted, then the models were tested on the test dataset to find the final accuracy. Also, model weights were saved after every 10 epochs while training. The best weights were selected using the criteria of train loss and validation loss. The final weights are stored for future experiments.

Besides the above experiments the models were also tested for class error rate. This experiment helped in understanding the importance of class weights. Also, the models were compared to other non pre-trained models of the same architecture.

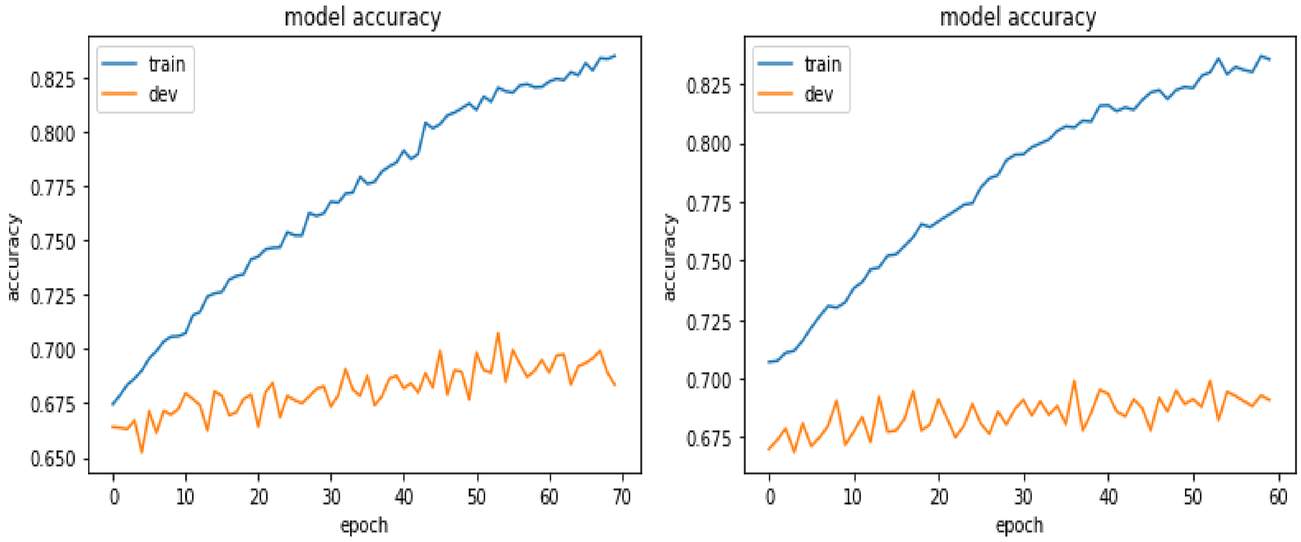
4.2 Results

The models had an average accuracy of 71% which is amongst the top five accuracies for the FER2013 dataset.

Models	**Training Accuracy**	**Testing Accuracy**
VGG16	83.5%	70.5%
VGG16 with class weights	74.6%	69%
RESNET50	88%	71.4%
RESNET50 with class weights	85.09%	73.10%

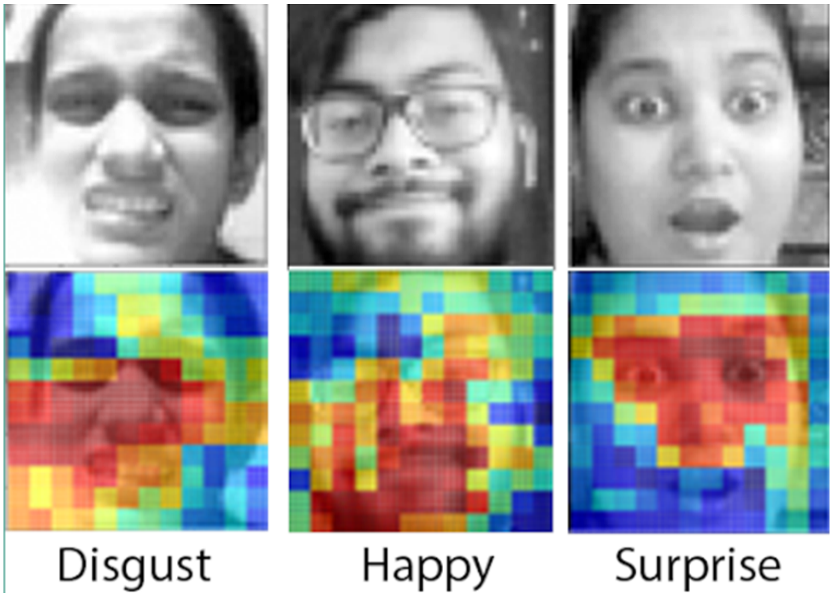
Comparing the results from the above table, RESNET50 with class weights has achieved the highest accuracy. When comparing the above results with the author's results, every result besides the RESNET50 with class weights matches the author’s results. My reasoning for this is that I used smaller batch sizes which reduced the variation that the model received. Another more likely reason can be that the pre-trained weights have been updated since the author’s use which has led to increased accuracy.

Below is the training graph of VGG-16 and VGG-16 with class weights.



5. Conclusion

At the start of the project, my main goal was to get the highest accuracy for the FER2013 dataset without training on excessive data. I explored several models from shallow CNNs to top image recognition models like VGG and RESNET. In addition to this, I was exposed to transfer learning which is a very powerful tool for training on small datasets and increasing durability to other distributions. To alleviate FER2013 dataset's inherent class imbalance, I applied class weights and data augmentation. In all the models achieved comparable accuracy with the author with less data. Given more computing power, a bigger or additional datasets might help further. Using occlusion maps, I discovered that the models were detecting correct facial features for each emotion.



6. Future Works

To further improve the accuracy of the models, I hope to utilize attention networks which auto adjusts the feature map of the convolution layer for each data sample. Additionally by using extra datasets like the CK+ and JAFFE, I plan to increase the resiliency of the models. The authors have ensembled seven models to achieve the highest ever accuracy of 75.7% on FER2013 dataset. I plan to duplicate this method and then create my own ensemble to compete with the mentioned ones.

References

Khanzada, A., Bai, C., & Celepcikay, F. T. (2020). Facial Expression Recognition with Deep Learning. *CoRR*, *abs/2004.11823*(2020), 7. Arxiv. https://doi.org/10.48550/arXiv.2004.11823

Cugu, I., Sener, E., & Akbas, E. (2017). MicroExpNet: An Extremely Small and Fast Model For Expression Recognition From Frontal Face Images. *CoRR*, *abs/1711.07011*(2017), 0. Arxiv. http://arxiv.org/abs/1711.07011

Pramerdorfer, C., & Kampel, M. (2016). Facial Expression Recognition using Convolutional Neural Networks: State of the Art. *CoRR*, *abs/1612.02903*(2016), 0. Arxiv. http://arxiv.org/abs/1612.02903

Goodfellow I.J. et al. (2013) Challenges in Representation Learning: A Report on Three Machine Learning Contests. In: Lee M., Hirose A., Hou ZG., Kil R.M. (eds) Neural Information Processing. ICONIP 2013. Lecture Notes in Computer Science, vol 8228. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-42051-1_16

Minaee, Shervin, Mehdi Minaei, and Amirali Abdolrashidi. 2021. "Deep-Emotion: Facial Expression Recognition Using Attentional Convolutional Network" Sensors 21, no. 9: 3046. https://doi.org/10.3390/s21093046