

Approach for fine-tuning a model on the FER-2013 Dataset

1 Introduction to the dataset

Facial Expressions Recognition Dataset - 2013 contains 48x48 pixel grayscale images of faces with varied emotions. It contained seven emotions, such as Anger and Disgust. It contained 28k training and around 4k testing examples. The task was to classify the emotions expressed on faces accurately.

2 Model Selection and Other Parameters

2.1 Model Selection

I chose MobileNetV2 because the images were only 48x48 pixels, so using large models was not worth the training time. MobileNetV2 provided a stable foundation, and I was able to achieve decent results through it without any need for changes.

2.2 Partial Fine Tuning

I opted for partial fine-tuning with the model. After experimenting a little, unfreezing 40 layers from the top provided a stable option between training time and accuracy. The model improved substantially as a result of this.

2.3 Class Weights

For some emotions, such as disgust, the model exhibited relatively low recall due to data imbalance, which caused the model to lean more towards other categories. Implementing class weights resolved that problem (although at the cost of accuracy), resulting in a good and balanced model. I adopted its example from the previous DeepWeeds Notebook.

2.4 Image Size as 128

Although the provided image size was 48 x 48 pixels, since MobileNetV2 was initially designed and supported much larger images, scaling the image at the

input would help. I initially set it to 96, but later increased it to 128 to further improve the model. As a result, I achieved a relatively stable model.

2.5 Early Stopping

Early Stopping was almost mandatory at this point, as I had trained the model for 20 epochs, and training it entirely was just too time-consuming.

2.6 Reduce LR on Plateau

I implemented this as a test, and it ultimately improved accuracy at certain stages significantly. Although the learning rate was not changed in the final results till the last, it still proved helpful earlier.

3 Results

The model achieved an accuracy of 61% and a macro F1 score of 0.56, indicating a good performance compared to other models on FER-2013.