

Introduction:

Human emotion is defined by a large number of factors including mood, environment, and personal factors. These emotions may manifest in several physiological observable features of the individual, including factors of the person’s brain activity, heart rate, and facial expressions [1]. Predicting or estimating emotions based on these observable parameters can have great value in fields such as mental health monitoring, multimedia recommendation, and human-computer interaction. Here, we present methods for predicting emotion based on images taken as people experience various emotions. Facial recordings can capture emotional changes that manifest in facial expressions of individuals [2]. Specific emotions can be linked to changes in facial expressions corresponding to the muscles located in the face. These changes can be recognized using neural network models and linked to unique emotions based on facial muscle movements. In this project, we analyze several neural network architectures capable of recognizing input images of facial expressions and predicting the corresponding emotions. The neural network predicts 7 emotions (anger, disgust, fear, happiness, sadness, surprise, and neutral) based on 48 x 48 greyscale images of facial expressions. In this paper, we look at the efficacy of several different deep neural networks models in predicting emotions based off of these facial expressions.

Literature Review:

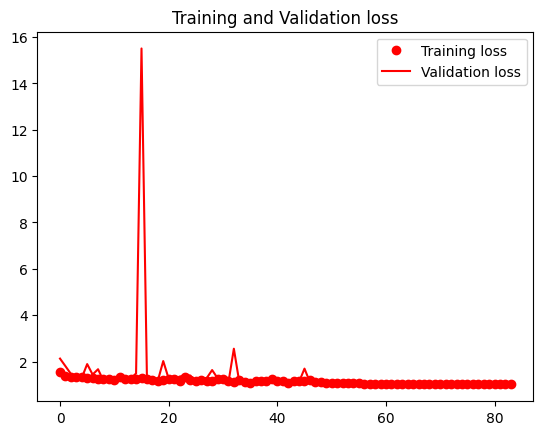
Human emotions are complex in that they are driven by many variables and expressed in a person-specific manner. This quality makes predicting emotions difficult due to their ill-defined and volatile nature that is unique to each person. Emotions are typically defined as categories (e.g. happy, sad, excited) but they can also be defined through quantitative measures of the emotion’s intensity and pleasantness [3]. These emotions can manifest themselves in a number of different ways either physiologically or visually. Physiologically, emotions can cause specific, observable reactions in several signals including the person’s heart rate, brain activity, temperature, or even in skin conductance [4]. Though changes in physiology can be observed, linking these changes to specific emotions has proven to be a difficult task. Similar to the plethora of physiological signals that can be indicative of emotion, facial expressions and movements of facial muscles are also highly correlated with specific emotional changes [2]. More specifically, certain movements and positions of key facial features can be indicative of an underlying emotion experienced by a person. For example, if a person is happy, they may move their lips and eyes in a specific direction to indicate their happiness. However, the specific changes are highly subjective, making it a difficult task to complete in the field of computer vision.

Convolutional neural networks have greatly improved the capabilities of emotion prediction based on images of people experiencing emotions. These neural networks are optimal for image recognition tasks and capable of learning features in images that correspond to specific labels given to them. A number of models have previously been used for facial expression recognition tasks successfully, including ResNet and VGG16 [5]. The accuracies for these tasks are typically much lower than similar tasks such as image recognition (i.e. MNIST) because of the similarity between facial images and high subject-subject variability. Two people may express the same emotion in unique ways that has to be captured by these models. Facial emotion recognition can benefit a number of fields ranging from healthcare to education. For example, facial images can be used to assess patients' emotion or mood and alert medical professionals in the case of anomalies. This can be important for the monitoring of those with mental illness such as post-traumatic stress disorder or major depressive disorder. In a workplace or education setting, emotion recognition can be used to optimize productivity by warning people about various factors such as stress. Optimal facial emotion recognition can have a profound effect on the fields of medicine and societal wellbeing [6].

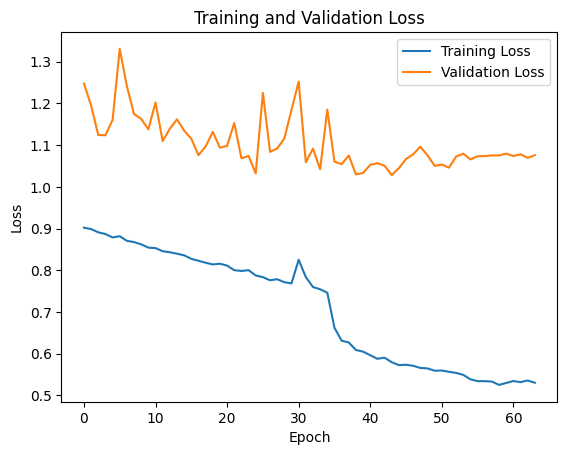
In this paper, we analyze several different models (ResNet, VGG, and MobileNet) in a facial expression recognition task. Our dataset comes from a public repository of a facial expression recognition challenge [7]. The dataset consists of 35,887 grayscale images consisting of 48 x 48 pixels. Each image contains a facial expression from one of seven categories: Angry, Disgust, Fear, Happy, Sad, Surprise, and Neutral. For each image the face of the person is centered and the true label is provided. The best performing models in the original challenge ranged between 60-70% accuracy, which is significantly lower than other image recognition tasks. As a basis of our model, we adopt code developed for a simpler facial expression task consisting of 5 emotions and apply it to the dataset. The goal of this project is to gain a deeper understanding of facial expressions and compare methods in predicting emotions associated with them.

Training details and Models:

Results:  
 We first established a baseline by running the custom Resnet18 model provided in the repository, which yielded around 63% accuracy. The plan was to use prebuilt keras models to train on the dataset, and compare the results to the baseline. As the images were grayscale images, we needed to transform the images to RGB by extending the dimensions of the images before attempting to use pretrained models provided by keras. We attempted to use the pretrained VGG16 and Resnet-50 models. With the pretrained ResNet-50 model, we attained around 60% accuracy on the data using a plateau learning rate scheduler. Following this approach, we opted to replicate the approach taken in the repository and define a custom ResNet-50.



With this custom Resnet-50, we specified regularization in the convolutional layers and as well as a dropout layer and a similar learning rate scheduler, which yielded slightly better results of 65%. It is important to note that the validation and training loss curves differ quite a bit, indicating a degree of overfitting towards the training set, which is likely due to the setup of the custom ResNet-50 being more complex than the pretrained ResNet-50.



Conclusion:

In this project, we attempted to use various neural network architectures to classify facial expressions into categories of emotions. Originally, we based our architecture on previous work based on emotion recognition with fewer emotions (5) and expanded the model to include 7 emotions and train with a larger dataset than originally intended. We hoped to try several models on the task of facial expression recognition and understand more about the nature of facial expressions and the benefits of each type of model. We were able to successfully train neural networks to recognize facial expressions. The strongest performing model ended up being the VGG16 model adopted for our uses. The MobileNetV2 model has the poorest performance of all models and appeared to

References:

[1] Bradley, M. M., & Lang, P. J. (2000). Measuring emotion: Behavior, feeling, and physiology.

[2] Ko, B. C. (2018). A brief review of facial emotion recognition based on visual information. *sensors*, *18*(2), 401.

[3] Russell, J. A. (1980). A circumplex model of affect. *Journal of personality and social psychology*, *39*(6), 1161.

[4] Shu, L., Xie, J., Yang, M., Li, Z., Li, Z., Liao, D., ... & Yang, X. (2018). A review of emotion recognition using physiological signals. *Sensors*, *18*(7), 2074.

[5] Mellouk, W., & Handouzi, W. (2020). Facial emotion recognition using deep learning: review and insights. *Procedia Computer Science*, *175*, 689-694.

[6] Boughanem, H., Ghazouani, H., & Barhoumi, W. (2023). Facial Emotion Recognition in-the-Wild Using Deep Neural Networks: A Comprehensive Review. *SN Computer Science*, *5*(1), 96.

[7] Dumitru, Ian Goodfellow, Will Cukierski, Yoshua Bengio. (2013). Challenges in Representation Learning: Facial Expression Recognition Challenge. Kaggle. https://kaggle.com/competitions/challenges-in-representation-learning-facial-expression-recognition-challenge