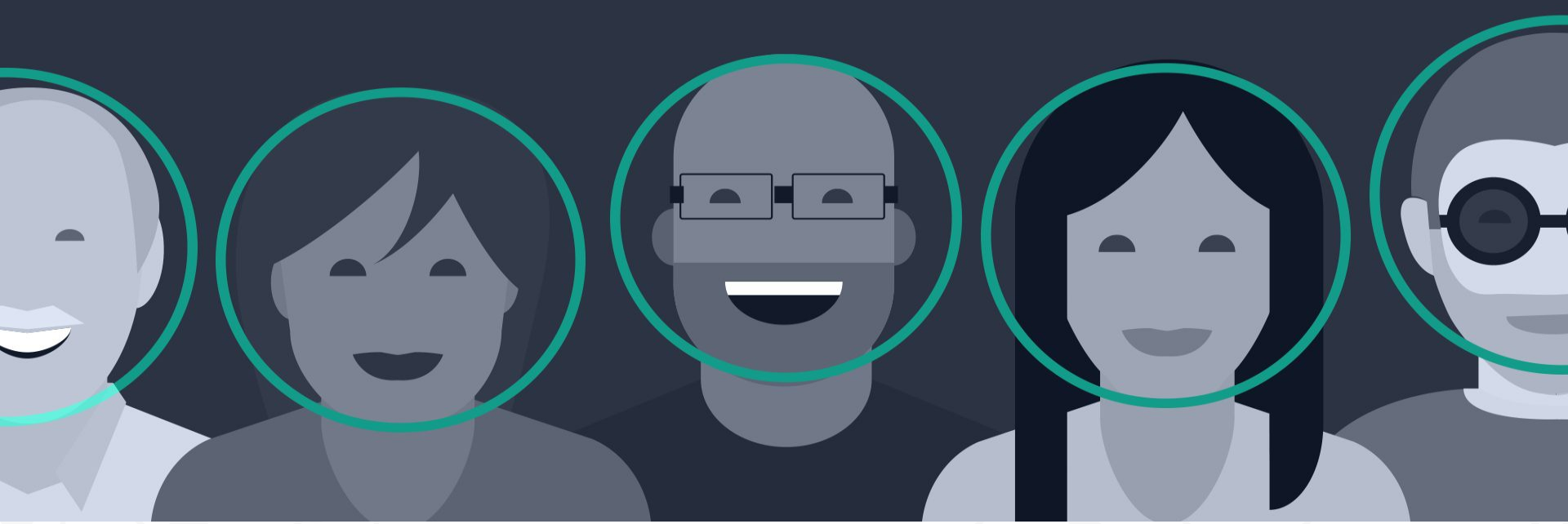


Facial Expression Recognition Using Convolutional Neural Networks

Presented By Joey Navarro, Data Scientist

Image References:
https://www.kairos.com/blog/face-detection-explained

Agenda

- Introduction
- Problem Statement
- Required Toolkit
- Dataset Used
- Neural Network Framework Used
- Data Preprocessing and Model Preparation
- Model Training Loop
- Model Architectures
- Model Results
- Prediction Examples

Introduction.

- Individuals on the autism spectrum often struggle understand others emotions through their facial expressions and other features.
- What if there was a way to detect emotion based off of facial expressions to help individuals on the autism spectrum identify how others feel?

Image Source: freepik.com

Problem Statement

Utilize pre-classified images to train a model for the purpose of facial expression recognition using a deep convolutional neural network.

Why use a deep convolutional neural network for this task?

Deep learning excels in recognizing objects in images as it's implemented using 3 or more layers of artificial neural networks where each layer is responsible for extracting one or more feature of the image.

What I Need - Toolkit

- Pre-Classified Data (CK+ Data Set)
- Neural Network Framework (Pytorch)
- OpenCV (Computer Vision Library)
- Matplotlib library
- Numpy library
- Model Architecture (VGG19 & ResNet18)
- NVIDIA CUDA Capable Graphics Card with at least 6GB Memory
- Windows or Linux OS Computer

Pre-Classified Data - Cohn-Kanade+ (CK+) Data Set

- Compiled by Jeffrey Cohn, and Takeo Kanade at Carnegie Mellon University Robotics Institute and The University of Pittsburgh Department of Psychology Respectively for Disney Research.
- The CK+ dataset is an extension of the CK dataset.
- The original dataset contains 327 labeled facial videos.
- The dataset used here is composed of the last three frames from each video in the CK+ dataset, therefore it contains a total of 981 facial expressions.
- Not original dataset used in project modeling.

Neural Network Framework - Pytorch

1. PyTorch is Pythonic
2. Easy to learn
3. Higher developer productivity
4. Easy debugging
5. Data Parallelism

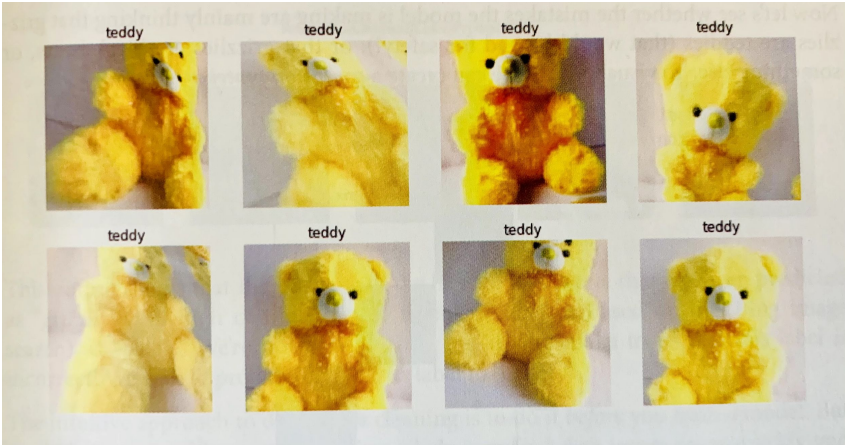
PyTorch

NVIDIA CUDA Capable Graphics Card with at least 6GB Memory

EVGA GeForce GTX 1660 SC Ultra

Data Preprocessing and Preparing Model

- Read images as pixel arrays for 48 x 48 pixel images for each emotion.
- Flattened the arrays into an integer list.
- Scaled data by a factor of 3 (limited by GPU memory).
- Create training dataset from a random ⅓ of all images.
- Setup augmentation by creating random variation transformations.
- Set up 10 Fold K-Fold cross validation.



Example of random image data transformations

Model Training Loop

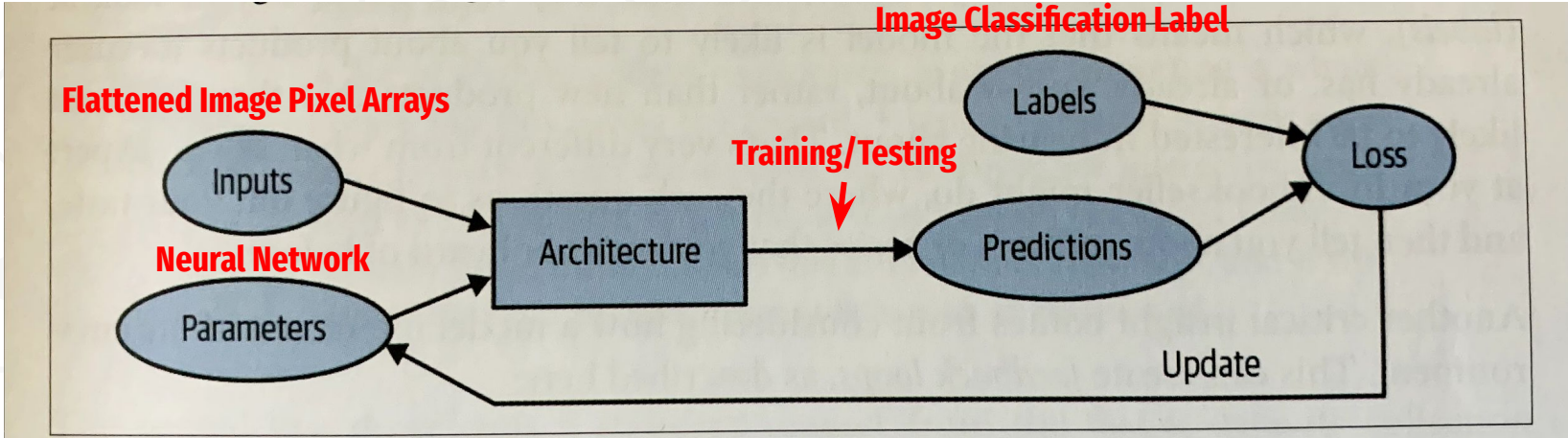


Figure 1-8. Detailed training loop

Model Architecture - VGG19

- This architecture's name comes from it's founding group at Oxford, namely Visual Geometry Group (VGG).
- The 19 after the name stands for the number of layers in the network.
- This network can be used for transfer learning, therefore it may also be used for facial recognition tasks.
- It is built by stacking convolutions together but the model's depth is limited because of an issue called diminishing gradient. This issue makes deep convolutional networks difficult to train.

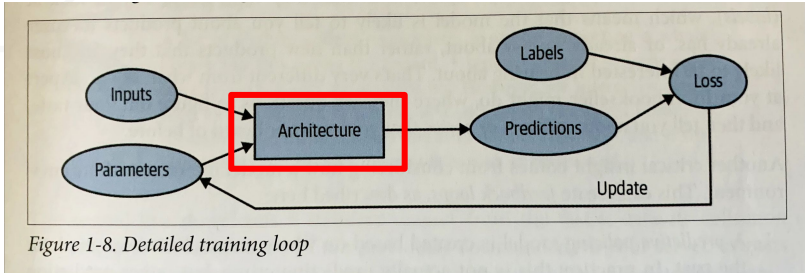
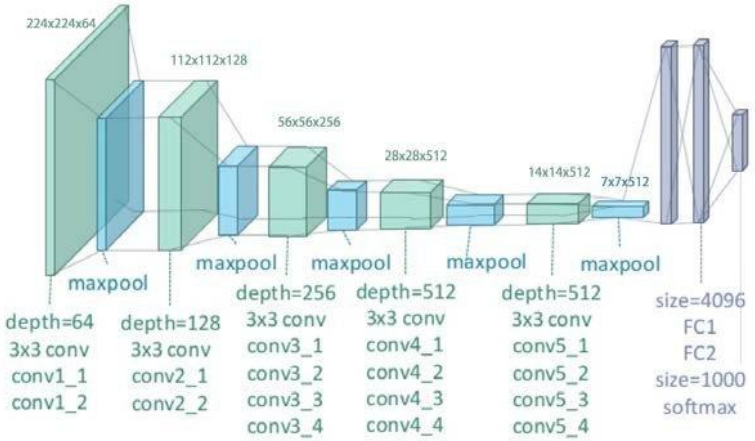


Figure 1-8. Detailed training loop



VGG19 Architecture

Model Architecture - ResNet18

- This architecture's name comes from the functions used by the network, namely residual learning functions or Residual learning Network (ResNet).
- The 18 after the name stands for the number of layers in the network. Resnet model was proposed to solve the issue of diminishing gradient.
- The idea is to skip the connection and pass the residual to the next layer so that the model can continue to train.
- With Resnet models, CNN models can go deeper and deeper.

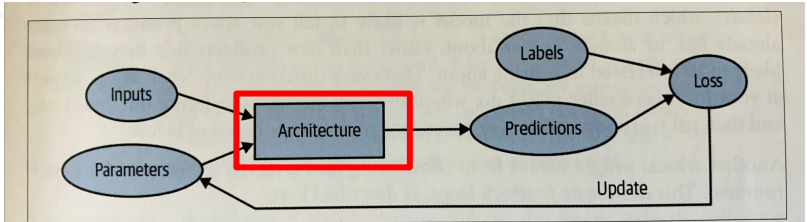
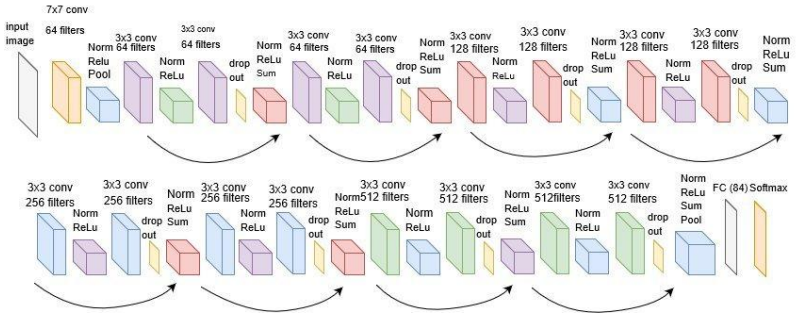
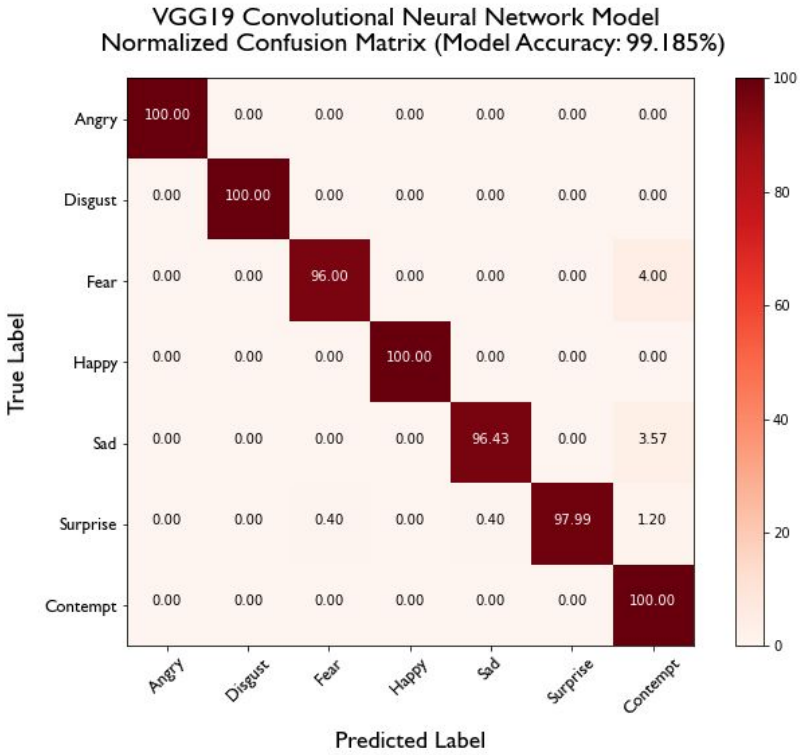


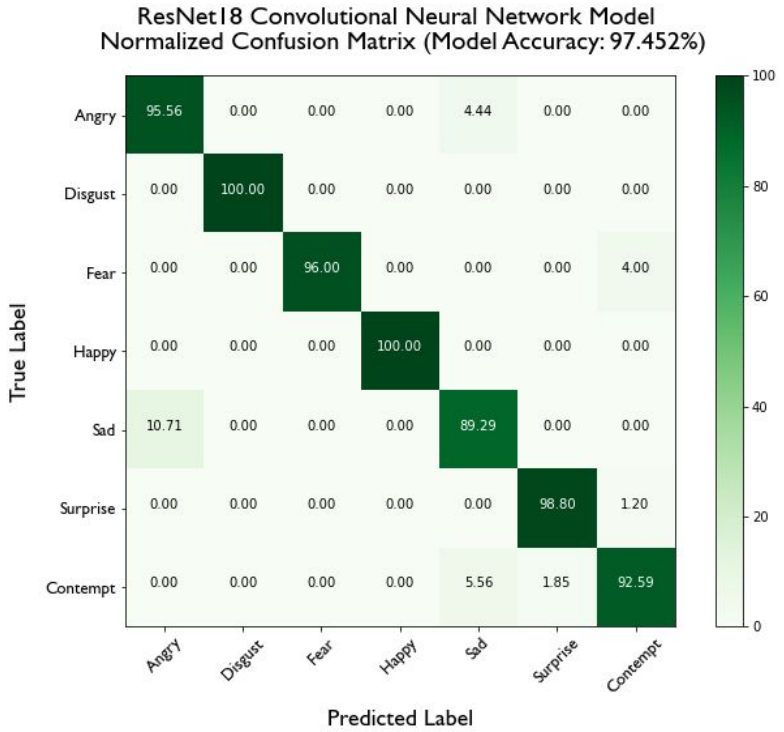
Figure 1-8. Detailed training loop



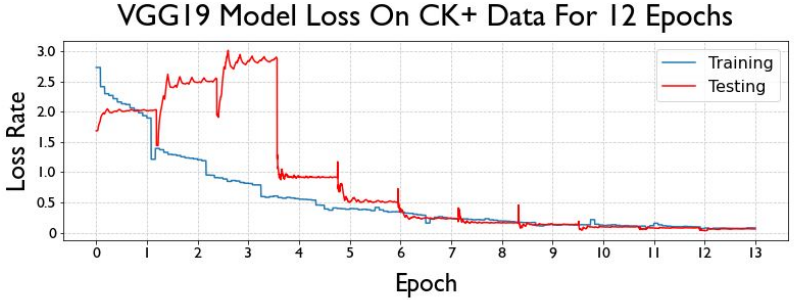
ResNet18 Architecture

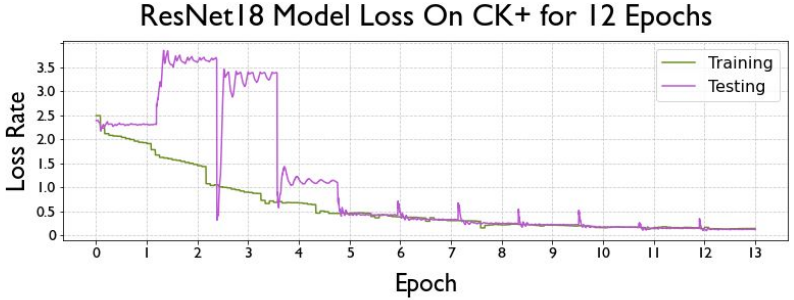
Running the Models - Confusion Matrices

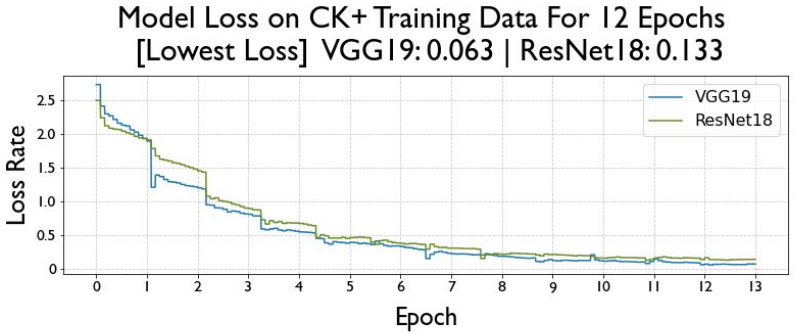


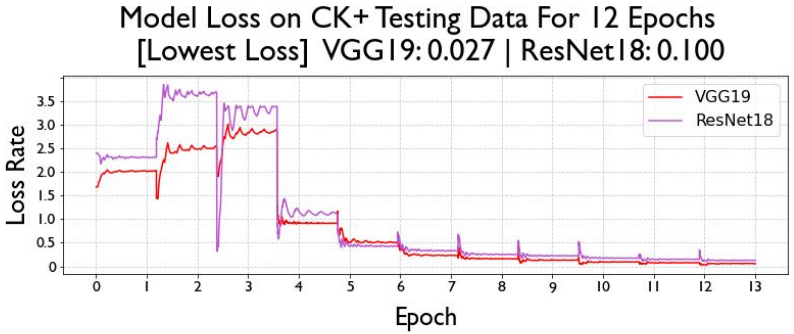


Running the Models - Loss

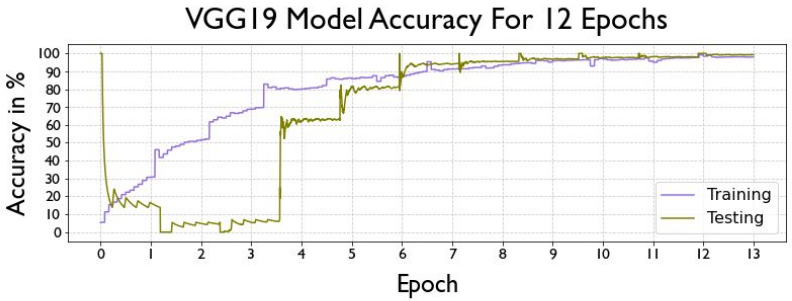


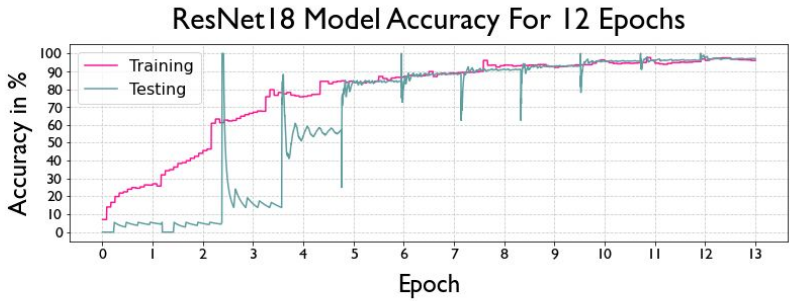


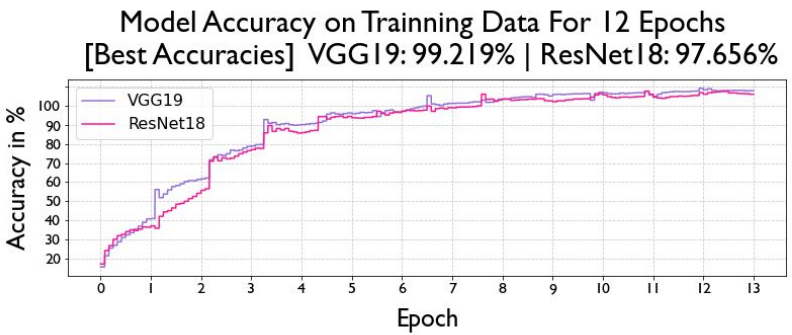


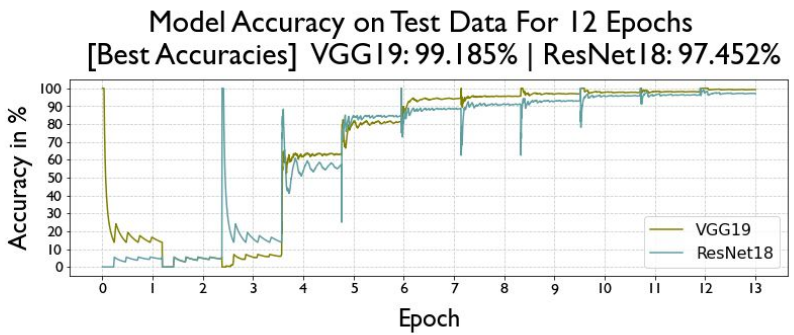


Running the Models - Accuracy





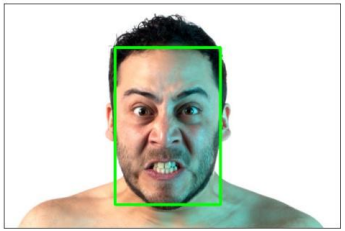




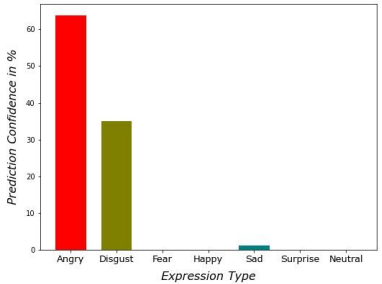
Prediction - Example

Original Image

Image Face Detection



Expression Prediction Scores



Emoji Expression

Angry Face is Being Expressed

Original Image

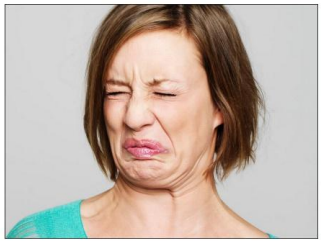
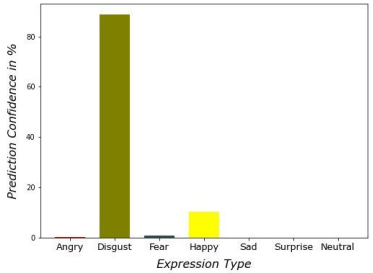


Image Face Detection

Expression Prediction Scores



Emoji Expression

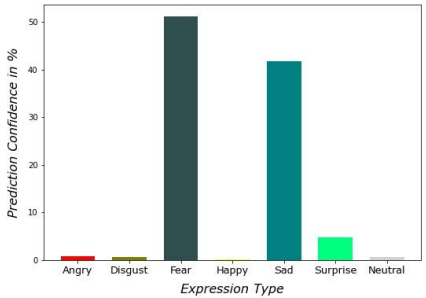
Disgust Face is Being Expressed

Prediction - Example

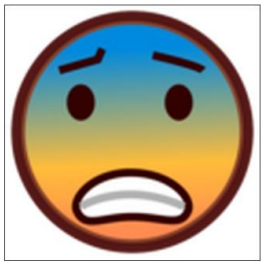
Original Image

Image Face Detection

Expression Prediction Scores



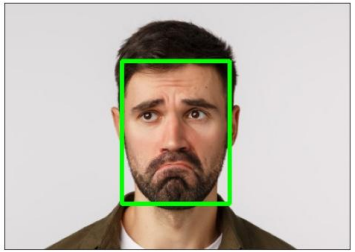
Emoji Expression



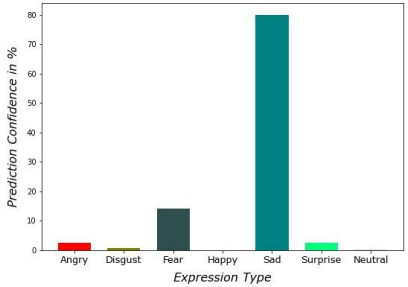
Fear Face is Being Expressed

Original Image

Image Face Detection



Expression Prediction Scores



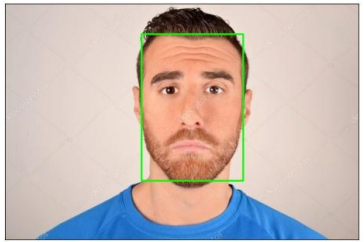
Emoji Expression

Sad Face is Being Expressed

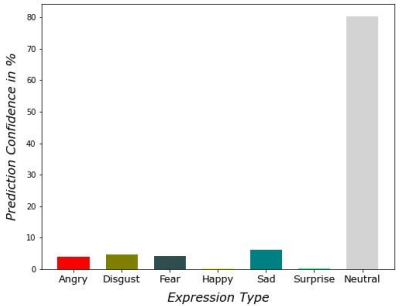
Prediction - Example

Original Image

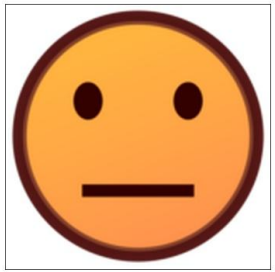
Image Face Detection



Expression Prediction Scores



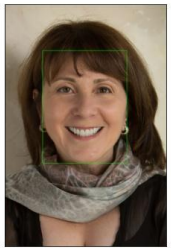
Emoji Expression



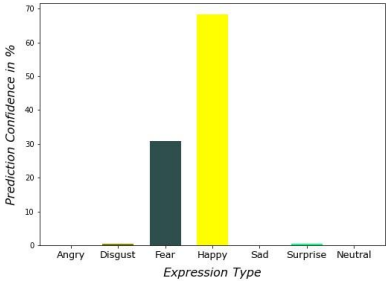
Neutral Face is Being Expressed

Original Image

Image Face Detection



Expression Prediction Scores



Emoji Expression

Happy Face is Being Expressed

Conclusion

- The CK+ dataset proved to be a very useful dataset to use for creating a model using deep convolutional neural networks.
- Both VGG19 and ResNet18 Model Architecture led to the creation of very accurate models with low loss rates.
- The models were not perfect as both neural networks struggled slightly when distinguishing between some emotions. VGG19 struggled slightly when learning disgust, sadness, and surprise.
- ResNet18 struggled slightly when learning anger, disgust, fear, sadness, surprise, and contempt. Therefore, when deciding which model to use for making expression predictions, VGG19 was the natural choice.
- When predicting expressions the model struggled when the image did not have a plain background and it also struggled to distinguish between subtly different emotions when in practice on real world images.
- Future work would focus on obtaining a larger dataset with more training data that shows a bit more variation in the expressions for each emotion and also creating a webcam app to use in real time for individuals such as my sister-in-law.

Questions?