Touchy Feely: An Emotion Recognition Challenge

Dhruv Amin (dhruv92@) | Patrick Chase (pchase@) | Kirin Sinha (ksinha@)

Introduction

- Goal is to classify the seven basic human emotions given images of facial expression
- Wide array of applications, ranging from improving livestreaming to aiding federal investigations
- For emotion recognition, most previous work reaches an accuracy of 60% using a variety of deep convolutional net architectures
- Using preprocessing with the Viola-Jones detector, transfer learning, and a simple SVM, some papers can reach an accuracy of 70%

Problem Statement

Given an input of a facial expression, we will classify it with one of the seven basic human emotions (Anger, Disgust, Fear, Happy, Sad, Surprise, Neutral). We will evaluate looking at the accuracy of our classifications, as well as looking at the confusion matrix to determine if certain expressions are more likely to be mislabeled as others

Dataset

- Kaggle challenge on Facial Expression Recognition data
- 28,709 48x48 pixel grayscale images of faces, labeled with the seven emotions

Methodology

We adjusted architectures and hyperparameters in response to our training and validation loss outputs overtime to increase learning and drive up accuracy. Our experiments, along with brief summaries of the takeaways, are listed below along with an overview of the final model.

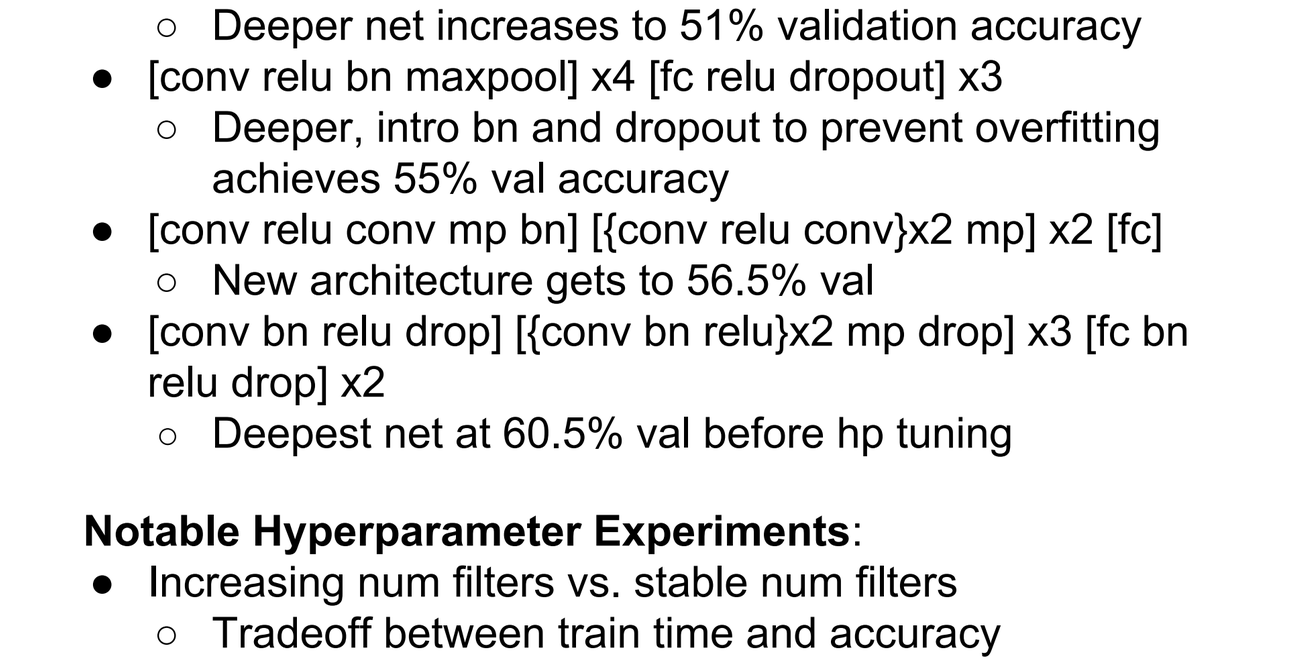
Notable Architecture Experiments:

- [conv relu] [fc relu] x2
  - Baseline achieves 40% accuracy on validation
- [conv relu] x3 [maxpool] [fc relu] x2
  - Deeper net increases to 51% validation accuracy
- [conv relu bn maxpool] x4 [fc relu dropout] x3
  - Deeper, intro bn and dropout to prevent overfitting achieves 55% val accuracy
- [conv relu conv mp bn] [{conv relu conv}x2 mp] x2 [fc]
  - New architecture gets to 56.5% val
- [conv bn relu drop] [{conv bn relu}x2 mp drop] x3 [fc bn relu drop] x2
  - Deepest net at 60.5% val before hp tuning

Notable Hyperparameter Experiments:

- Increasing num filters vs. stable num filters
  - Tradeoff between train time and accuracy
- Kernel sizes (4 ideal)
- Padding (alternate between 3 and 1)
- Dropout (.2 ideal)
- Affine layer sizes (1024 ideal)
- Layers (7 middle layers, 2 fully connected)

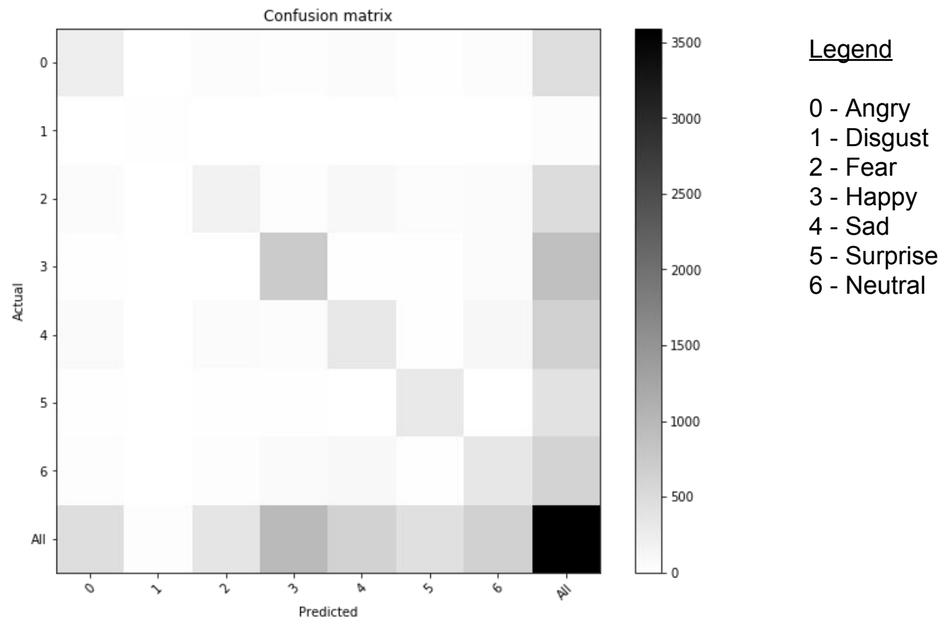
Final Architecture and Parameters:



Current Results

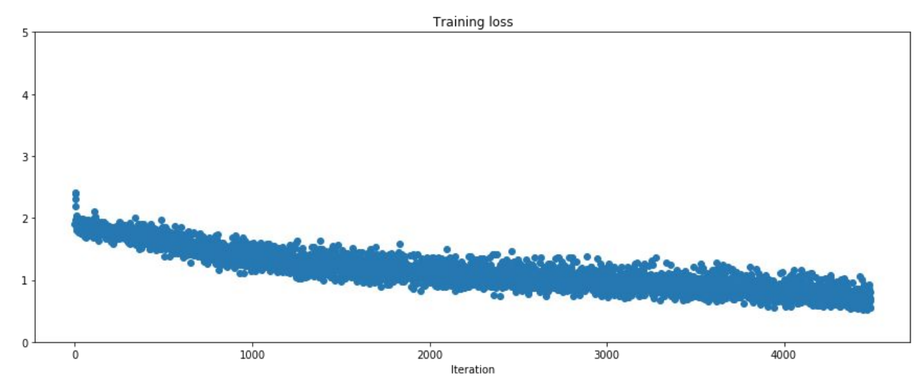
Our model gave the following accuracies:
Training Set: 78.49 % (averaged over 4 models)
Validation Set: 60.37% (averaged over 4 models)
Test Set: 61.05% (from our final model)

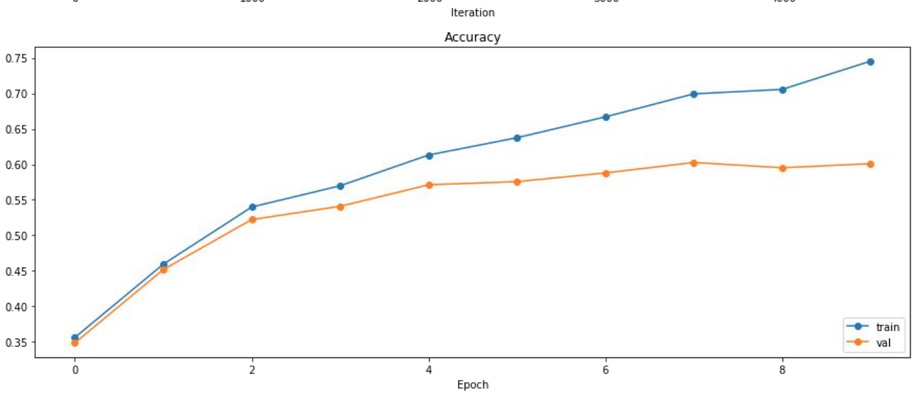
Confusion matrix for test set:



- Model struggles to determine between Sad (actual) and Fear (predicted), Fear (actual) and Sad (predicted)
- Model most successful at determining Happy

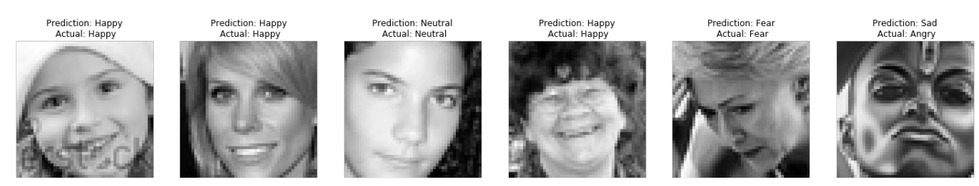
Training Loss and Train/Validation Accuracy over Time:

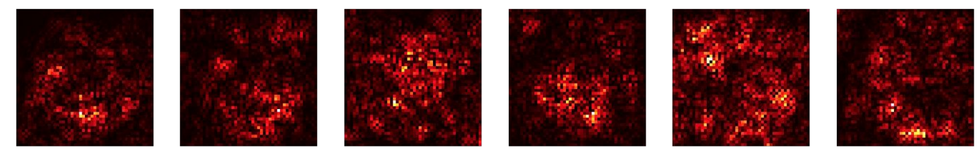




- Healthy loss curve although lack of learning in later epochs
- Gap between train and validation accuracy suggests more can be done for overfitting

Validation Set Images, Classifications, and Saliency:





- The saliency maps show which pixels are contributing the most to the classification
- We can see that the model is influenced by pixels we expect for various emotions (e.g. smile salient for happy)

Future Directions

- Real-time facial expression recognition using the model developed here
- Use transfer learning and try to increase validation accuracy to 70%