Facial Emotion Recognition with Convolutional Neural Networks

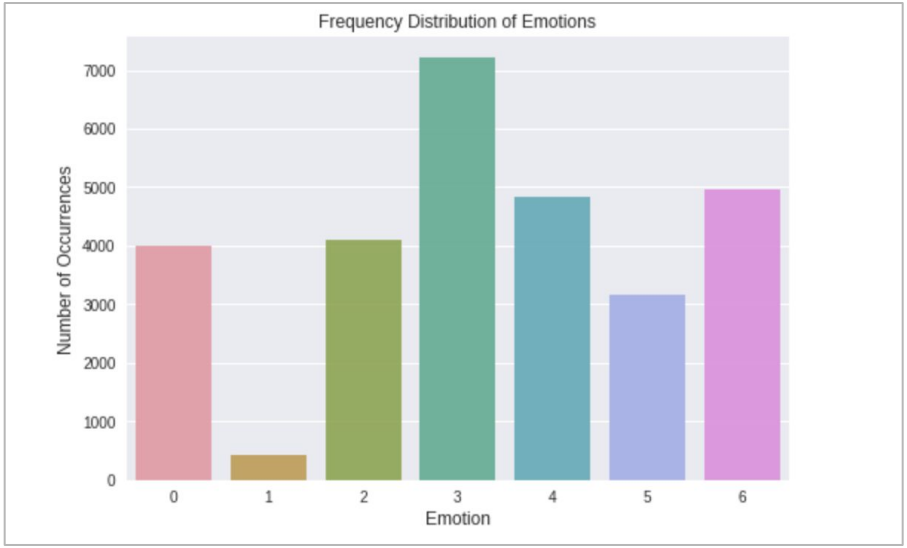
Justin Littman
Matthew Martin
DS 4440

Background

- Automatic Facial Emotion Recognition (FER)
  - Human Computer Interaction
  - Virtual Reality
  - Augmented Reality
  - Driver Assistant Systems
  - Entertainment
- Find more efficient and smooth connections between humans and machines

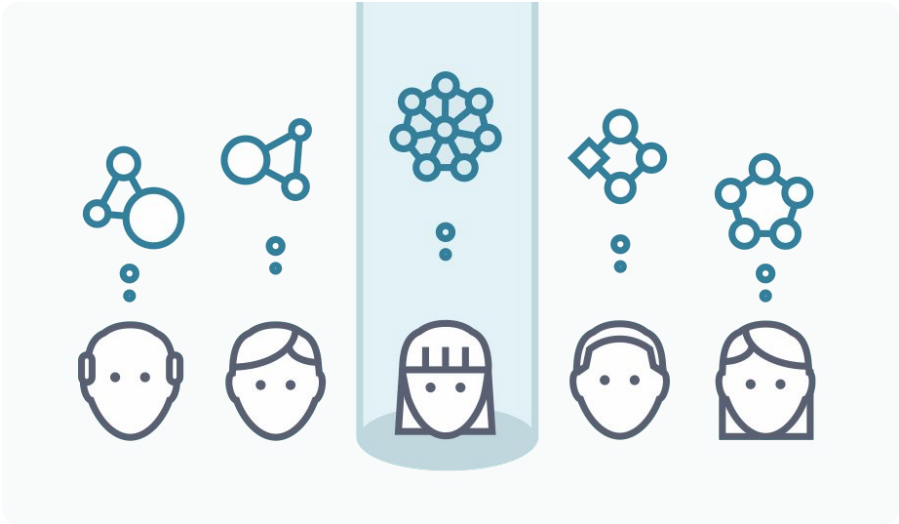
Dataset

- FER2013 Dataset
  - 35887 sample images
  - Seven labels
    1. Anger
    2. Disgust
    3. Fear
    4. Happy
    5. Sad
    6. Surprise
    7. Neutral



Past Results with FER2013

- Kaggle Competition
  - Challenges in Representation Learning: Facial Expression Recognition Challenge (2013)
  - Top accuracy ~ 0.71
- Github Paper
  - Used Sequential CNN and Xception inspired architectures
  - Accuracy ~ 0.66



Difficulties

- FER has many challenges
  - Non-uniform nature of human face
  - Limitations due to
    - Lighting
    - Face orientation
    - Shadows
- Nature of facial expressions
  - Subtle movements
- Need accurate and robust models
  - Deep learning can help achieve this

Implementations

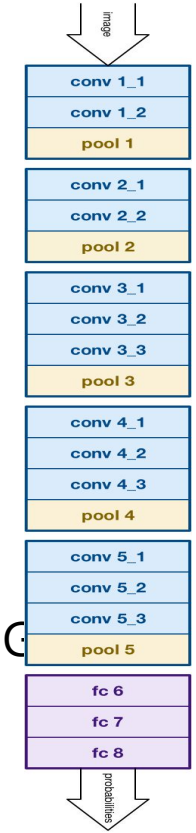
- Visual Geometry Group (VGG)
- Residual Neural Network (ResNet)
- Self-created model

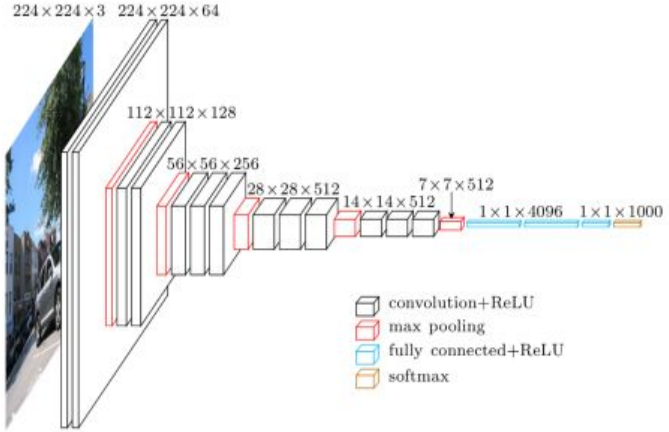
VGG (2014)

- Developed by Oxford University's Visual Geometry Group (VGG)
- Created model for ImageNet competition
  - Placed second for classification
- Wanted to see how ConvNet depth affects accuracy
- Best performing models
  - VGG16
  - VGG19

Architecture

- VGG Block
  - Convolution Layer
  - Nonlinearity
  - Maximum Pooling
- Kernel Size
  - 3x3 used in each Convolution
- Pooling Size
  - 2x2 with stride of 2 in each Pool
- Different implementations of VGG





VGG Implementations

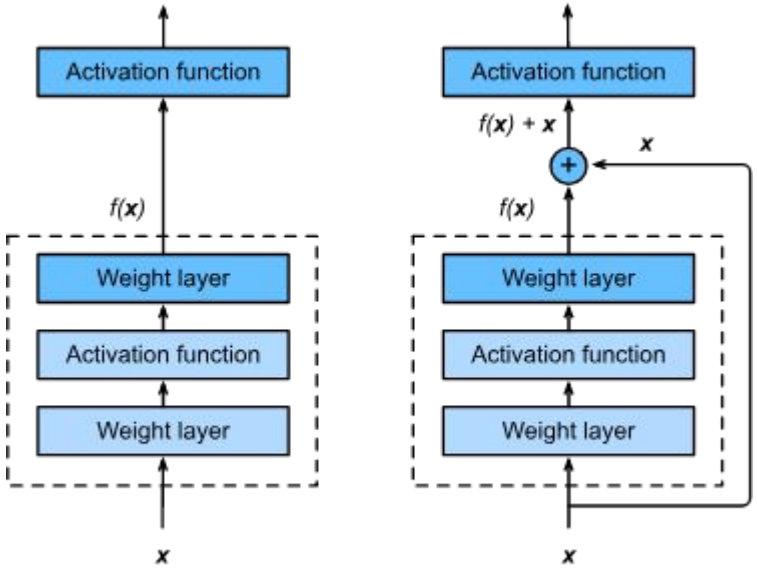
- Implemented VGG11 and VGG16
  - Tested batch sizes of 32, 64, and 128 for each
- Tested Dropout with VGG11
  - More Dropout
    - 50% dropout after every layer
  - Less Dropout
    - 50% dropout after every other layer

Resnet (2015)

- Developed by Microsoft Research
- Placed first in ImageNet 2015 competition for classification
  - Achieved error of .357
- Introduced deep residual learning framework
  - Eases training of deep networks
- Combats increasing depth problem

Architecture

- Idea: Allow information to easily propagate through the network
- Residual Blocks
  - Convolution Layer
  - Batch Normalization
  - Nonlinearity
  - Skip Connection
- Kernel Size
  - Uses VGG's 3x3 design



Resnet Implementation

- Implemented ResNet50, ResNet18, and ResNet10*
  - Tested batch sizes of 32, 64, and 128 for each
  - Tested Adam and RMSProp optimizers for each
- ResNet10 was an attempt to simplify the larger architectures
  - Used half the blocks of ResNet18

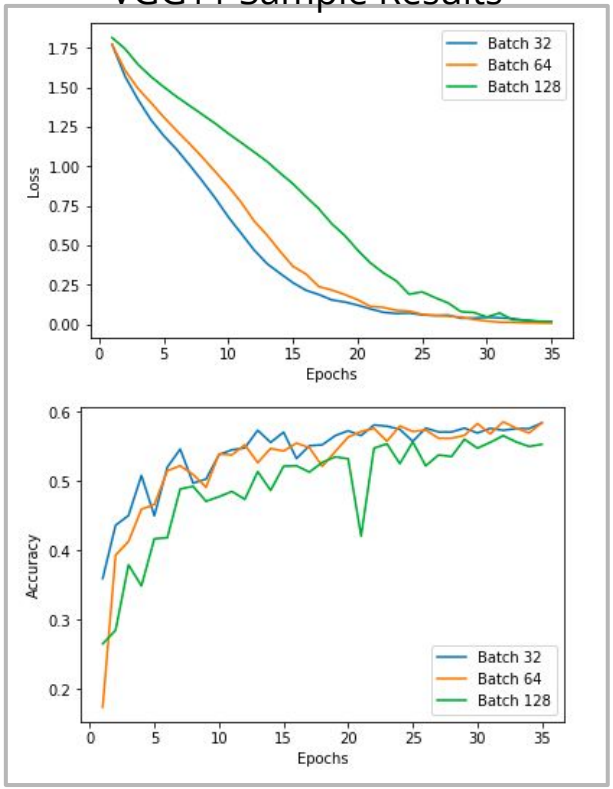
Our Model Architecture

- Followed the VGG Block architecture
  - Added Batch Normalization after each layer
  - Increase number of filters in each block
    - Number of filters: 3,5,7
- Added an attention component
  - Tells model what areas of the image are more important

VGG Results

- VGG11
  - Batch size of 64 had best performance
  - Test accuracy of 58.9%
- VGG16
  - Batch size of 32 had the best performance
  - Test accuracy of 57.8%
- VGG11 with More Dropout
  - Test accuracy of 38%
- VGG11 with Less Dropout
  - Test accuracy of 58.7%

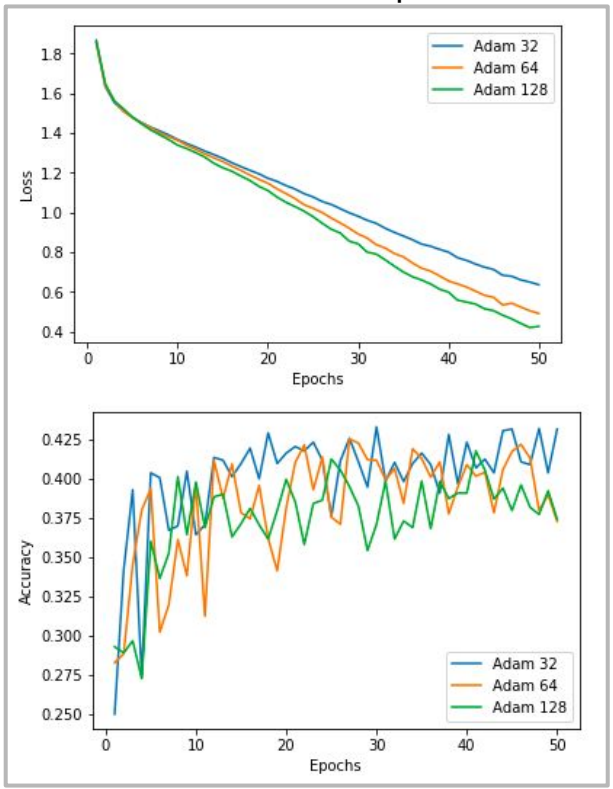
VGG11 Sample Results



ResNet Results

- ResNet50 and ResNet18
  - Accuracies were low for both Adam and RMSProp optimizers
  - No accuracy above 25%
- Modified ResNet (10 Layers)
  - Best performance was with batch size 32 and Adam optimizer
  - Accuracy of 43.2%

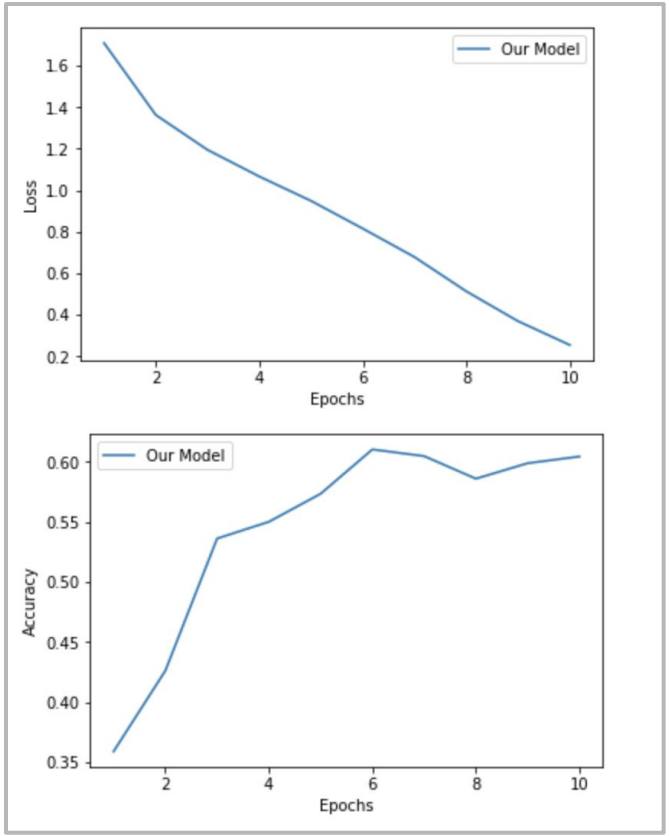
Modifed ResNet Sample Results



Our Model Results

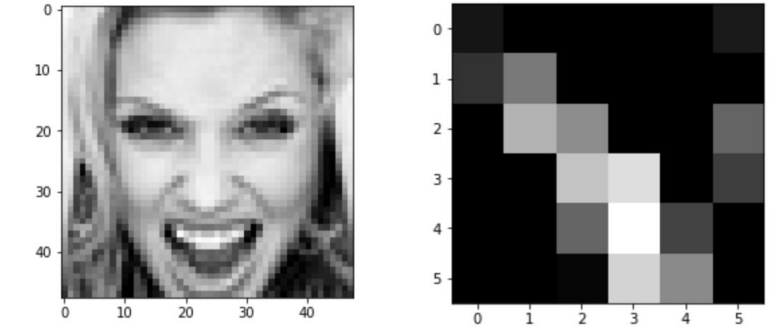
- Batch size of 64 with an Adam optimizer
  - Test Accuracy of 61.4%
- Best performance on Happy (3) and Surprise (5)

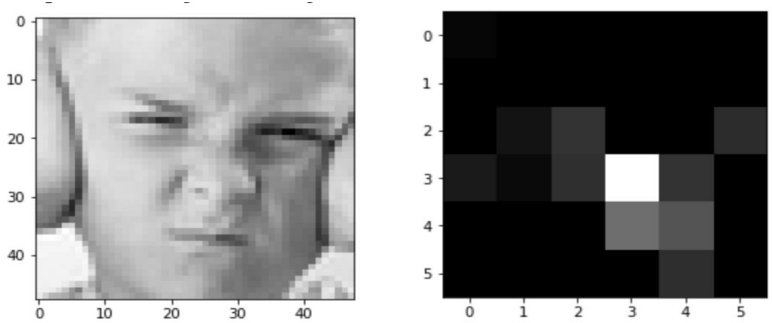
	precision	recall	f1-score	support
0	0.68	0.41	0.51	958
1	0.79	0.49	0.60	111
2	0.41	0.54	0.47	1024
3	0.84	0.81	0.82	1774
4	0.45	0.57	0.50	1247
5	0.83	0.65	0.73	831
6	0.58	0.58	0.58	1233



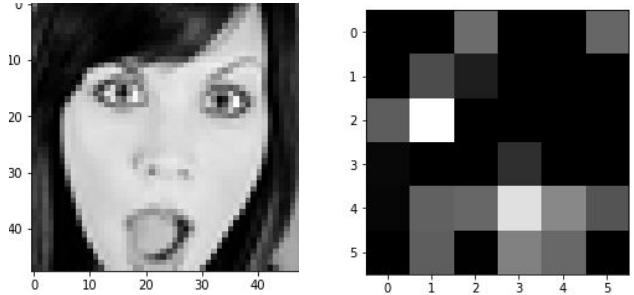
Attention Areas

Anger Attention

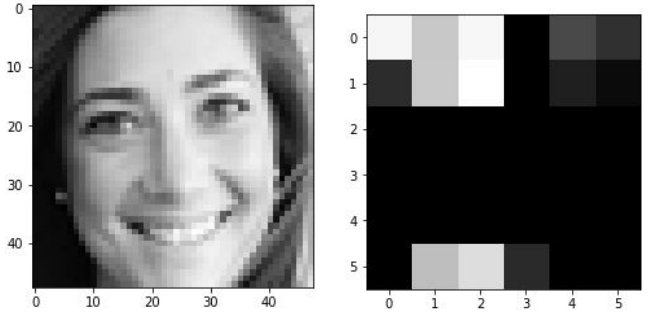




Surprise Attention



Happy Attention



Demo