Facial Emotion Recognition

Manasa Donepudi

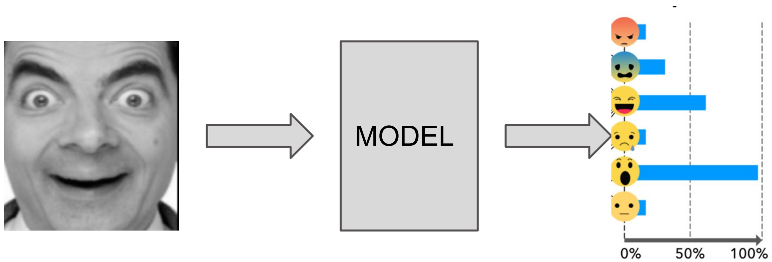
Nitin Chakravarthy Gummidela

Shruthi Reddy Rodda

Outline/Agenda

- Introduction
- Dataset overview
- Model
- Experimental Results
- Visualization
- Conclusion
- Future work
- References

Objective



- Prediction of emotion from facial expressions
- Compare different models and techniques
- Analyse the prediction by diving into the model

Related Work

1) In order to recognize and categorize facial expressions accurately, based on our knowledge in class, we first thought of using basic machine learning techniques such as Support Vector Machine using PCA.
2) Krizhevsky et.al.[4] Simonyan et.al.[5] Y. LeCun et.al[6] used VGG, AlexNet and LeNet architecture for ImageNet dataset
3) Tang et.al. used SVM over global features extracted from the CNNs

Overview of FER2013

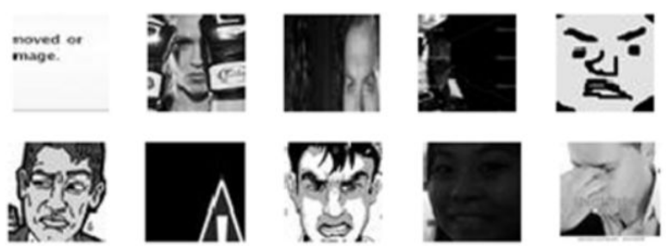
Training data: 28,709
Validation data: 3,589
Test data: 3,589

48x48 Grey scale Images

0 – angry
1 – disgust
2 – fear
3 – happy
4 – sad
5 – surprise
6 – neutral

FER2013

- Incorrectly Labeled samples
- Samples which are not faces
- Difficult because the model has to generalize for incorrect data
- Human accuracy on the dataset is 65+-5 %

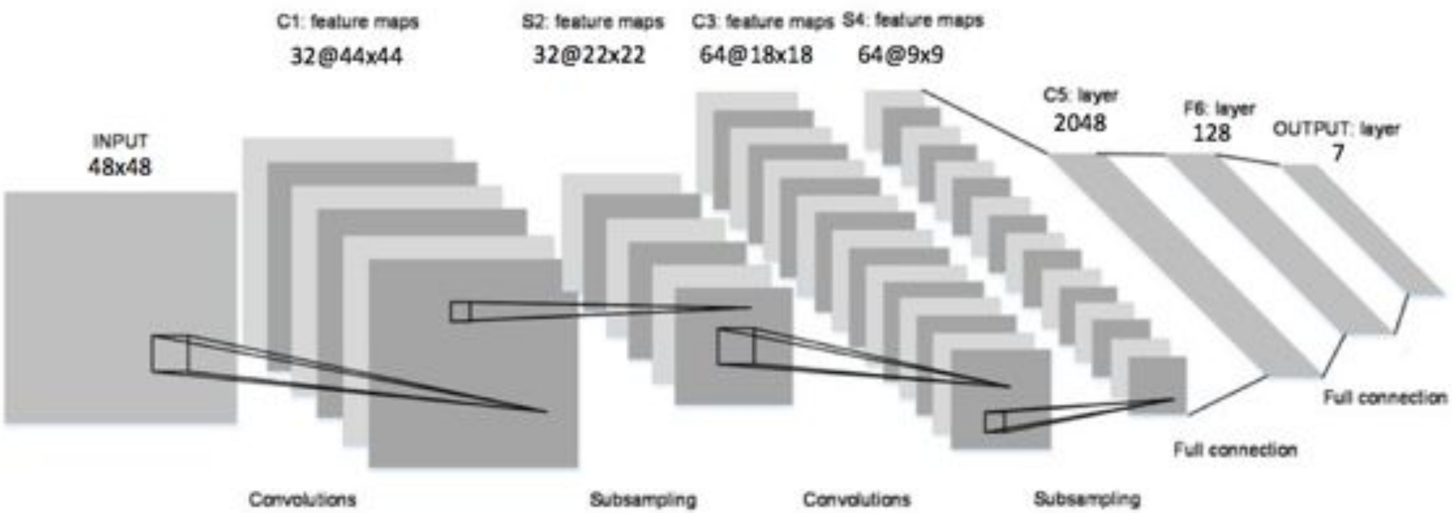


CNNs

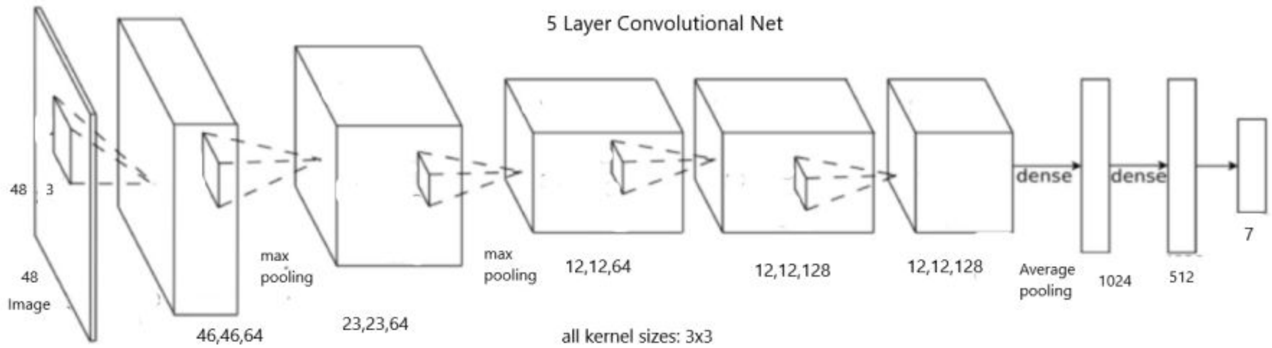
- Shallow CNNs to deeper CNNs
- LeNet
  - 2 Convolution layers
- AlexNet
  - 5 Convolution layers
- VGG12
  - 12 Convolution layers

2-Layer ConvNet

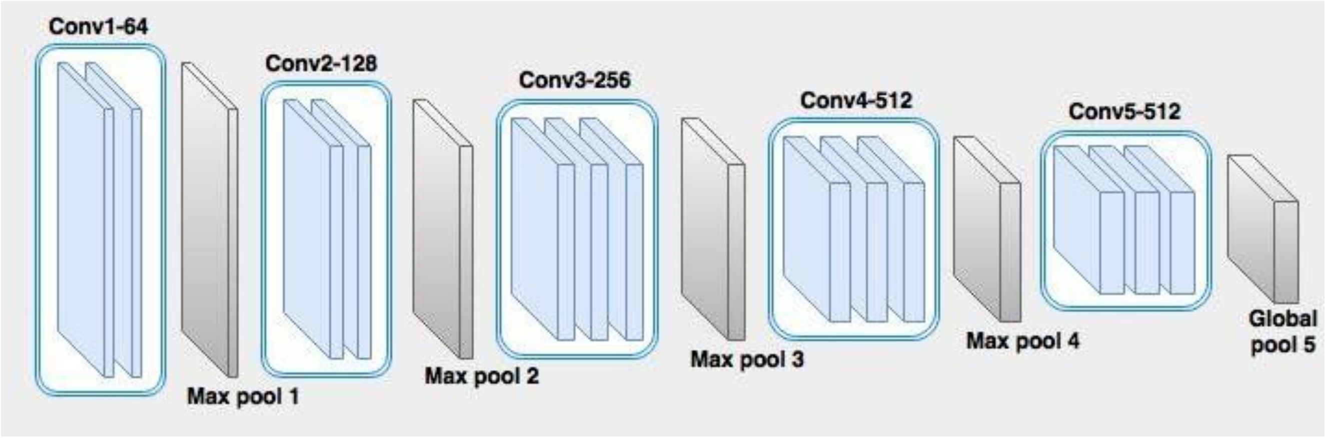
2-layer convolution net - LeNet Architecture



Deep-ConvNets



5 Layer CNN

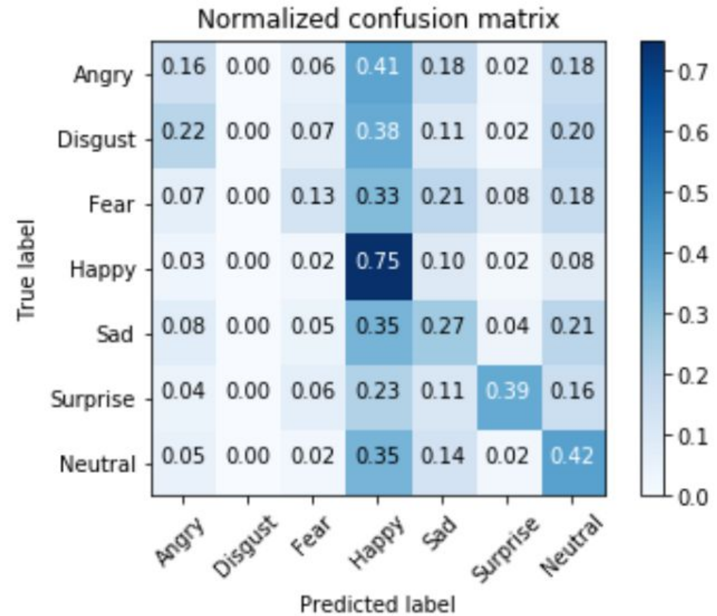


16 layer CNN (adapted to 12 layers)

SVM with PCA

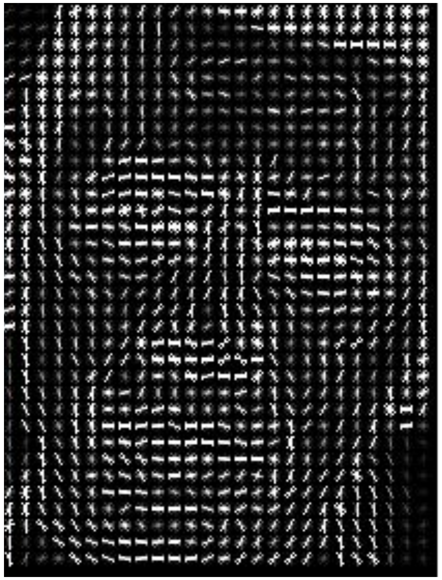
- # principal components = 200
- radial Basis function
- C = 1
- gamma = 0.0001
- accuracy = 39.98 %

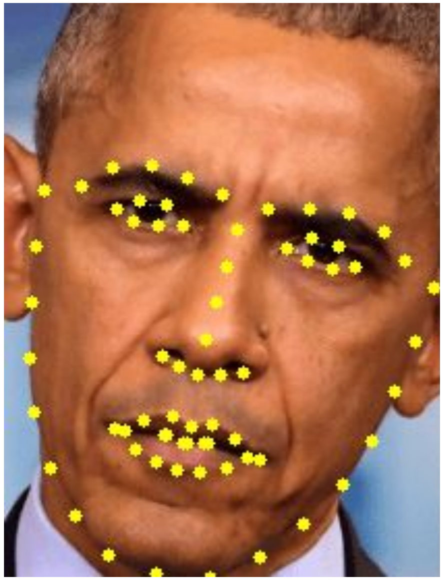
	precision	recall	f1-score	support
0	0.29	0.14	0.19	491
1	0.00	0.00	0.00	55
2	0.35	0.13	0.19	528
3	0.41	0.75	0.53	879
4	0.27	0.26	0.27	594
5	0.61	0.40	0.48	416
6	0.35	0.41	0.38	626
avg / total	0.37	0.38	0.35	3589



SVM on HOG and Facial Landmarks

accuracy = 48.2%





Softmax v/s SVM

- Multiclass SVM instead of softmax
- consistent increase in accuray

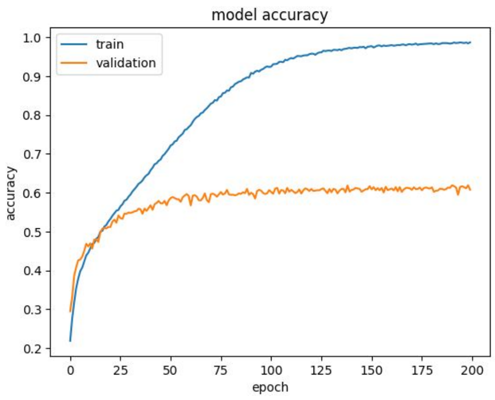
model	Softmax	MultiClass SVM
LeNet	56.2	56.9
AlexNet	61.1	61.8
VGG12	63.9	64.5

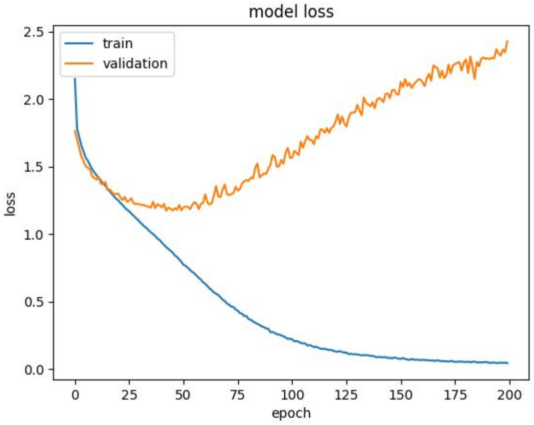
Results

Model	Avg. Precision	Avg. Recall	Avg. f1 score	Accuracy (%)
SVM with PCA	0.37	0.38	0.35	39.98
SVM on HOG and Face Landmarks	1.00	0.48	0.65	48.2
AlexNet	0.62	0.62	0.62	61.1
LeNet	0.55	0.56	0.55	56.2
VGG12	0.63	0.64	0.63	63.9
SVM on CNN	0.64	0.64	0.64	64.4

Analysis

Training Curves:

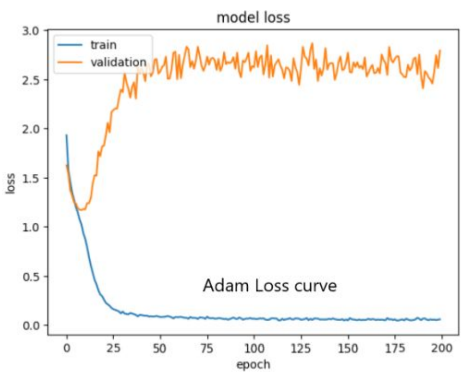


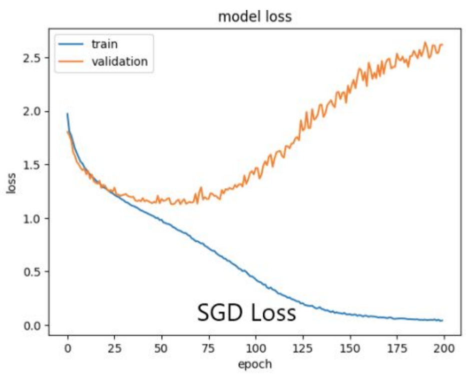


Hyperparameter Tuning on 5-Layer Net

- Optimizers:

  We have tried several optimizers like SGD, Adam, Adadelta, RMS Prop.





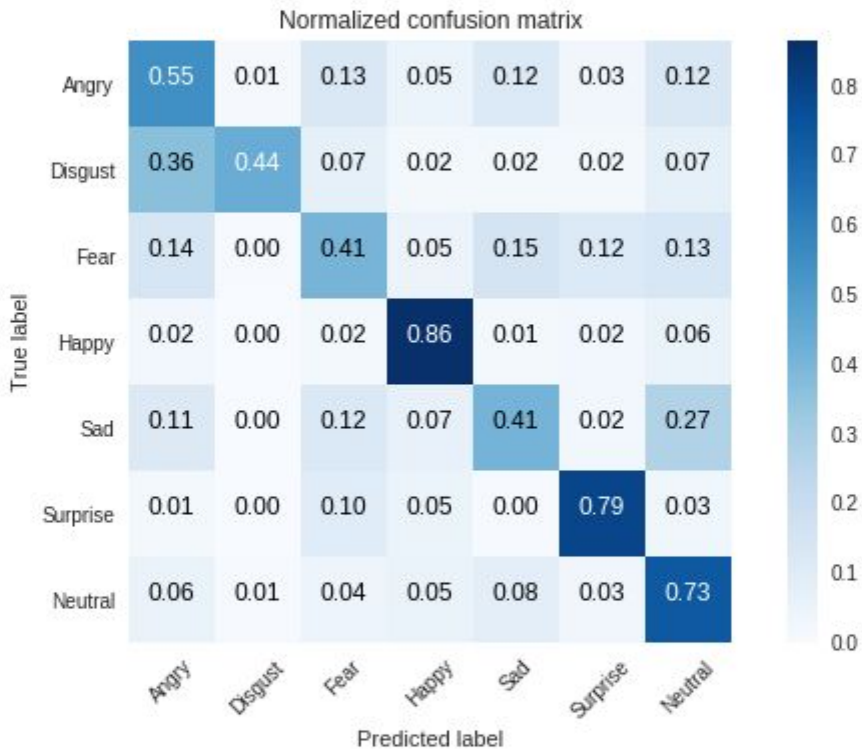
- Number of filters:
    - All layers have same number of filters(64)
    - Top three layers have 64 filters and rest two have 128 filters
    - Top three layers have 128 filters and rest two have 64

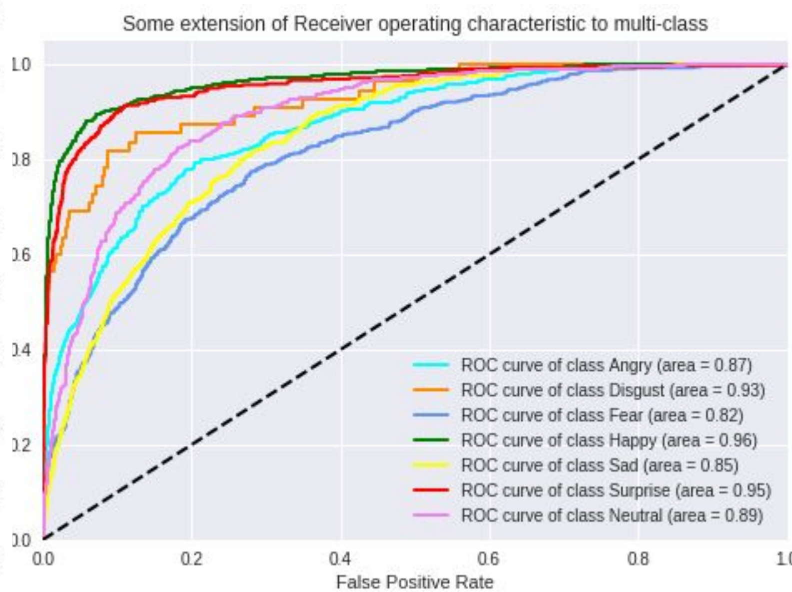
- Dropout rates:
    - Without dropout, accuracy 0.55
    - With optimal dropout, accuracy 0.61

- Fully Connected Layers

- Number of Convolutional Layers
    - Shallow to deep CNNs

Confusion matrix, ROC curves





Misclassifications:
Disgust -> Angry
Sad -> Neutral

Visualization -Layer activations

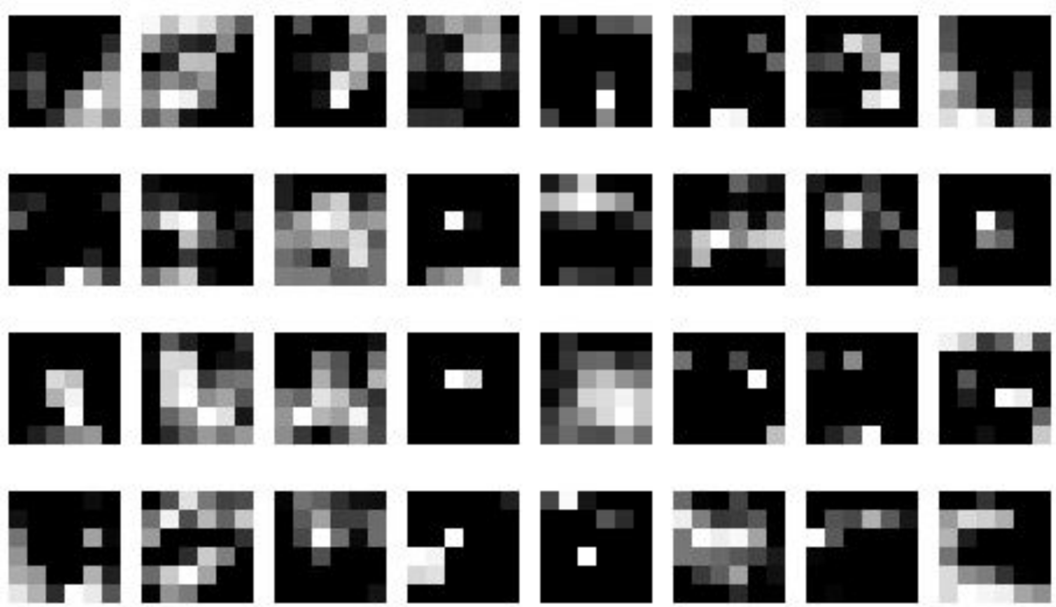
Layer -1 activations

More visualization...

Layer -3

1 max pooling

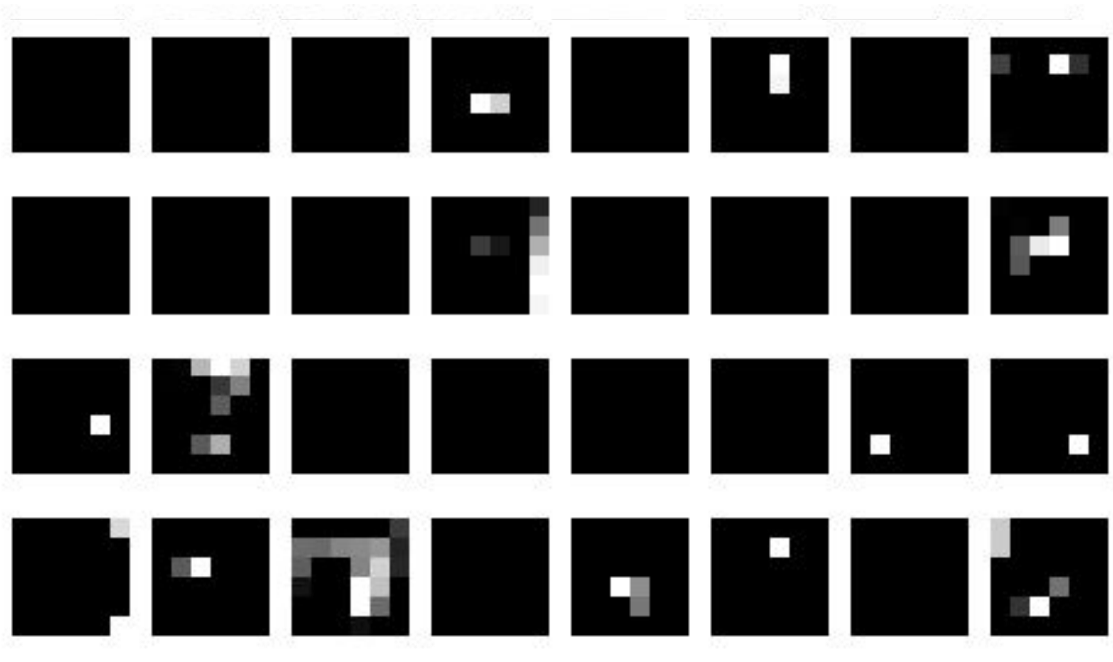
And more …



Layer -8

3 max pooling layers

And more…

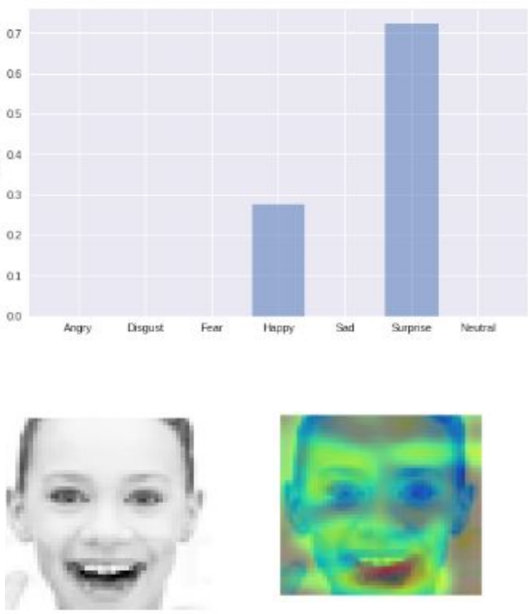


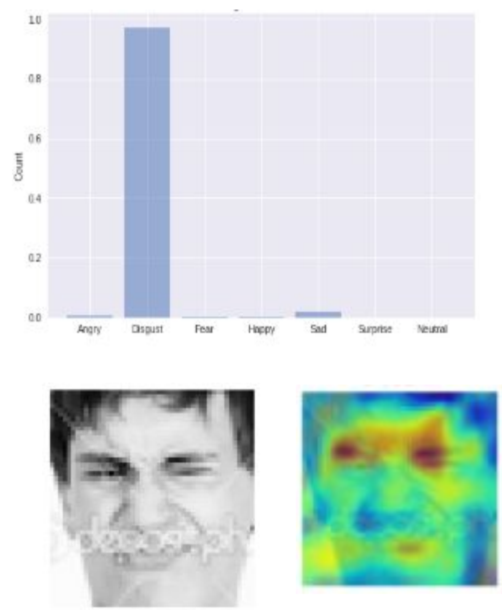
Layer -9

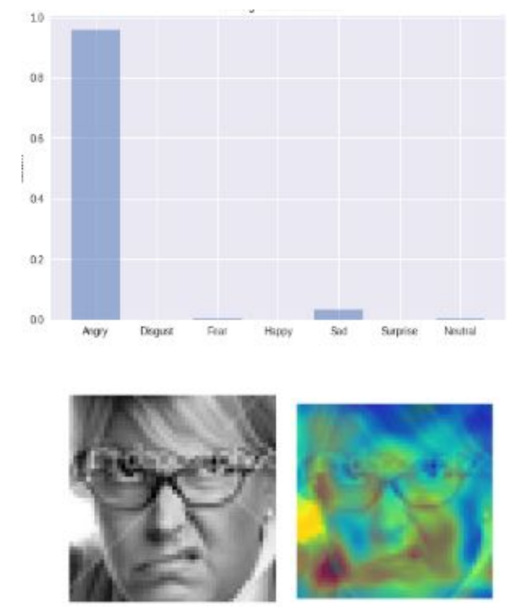
Final Activation map

inputs to dense layers that predict the probabilities

Images and Prediction probabilities





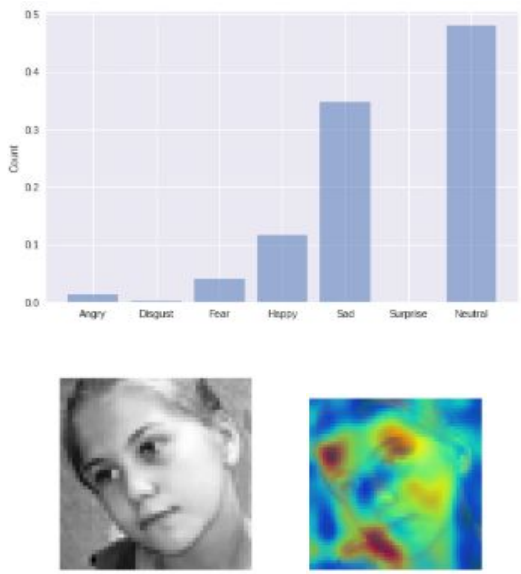


Correctly classified
Label: 5, Surprise

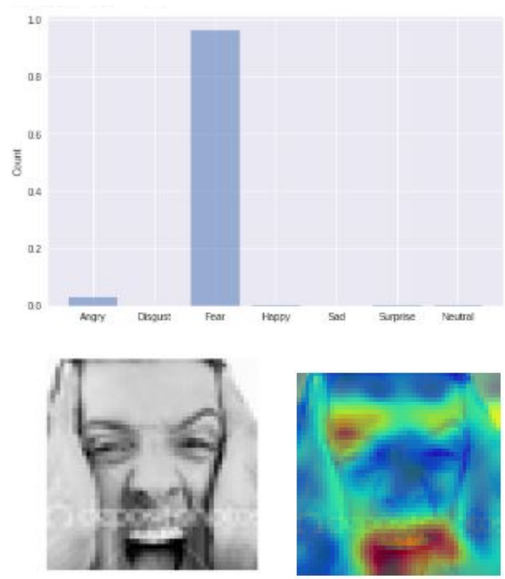
Correctly classified
Label: 1, Disgust

Correctly classified
Label: 0, Angry

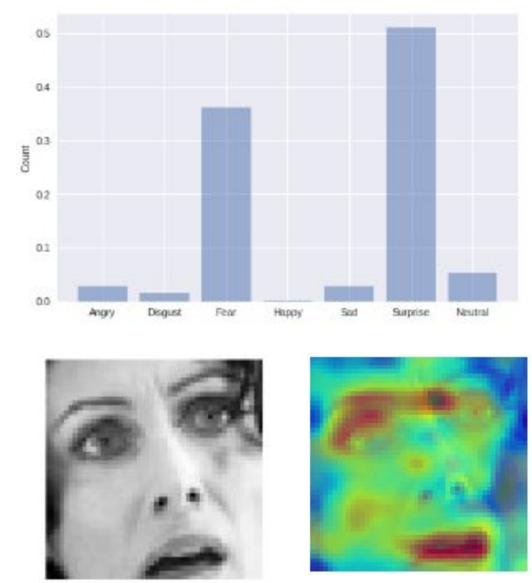
More Images ...



Classified Neutral (6)
Original Sad(4)



Classified Fear (2)
Original Angry(1)



Classified Surprise (5)
Original Angry(1)

Conclusion

- SVMs perform better over global features (global activation, HOG and Face landmarks) rather than dimensionally reduced images
- Increasing dropout improves accuracy to a certain extent (bias-variance trade off)
- Deep networks perform better for classification than shallow networks.
- CNNs performed better for a class with less samples database where SVM was biased towards the class with more samples

Future Work

- Exploring bigger networks - GoogLE Net, ResNet
- Using pretrained models
- Realtime classification

References

1. Arriaga, Octavio, Matias Valdenegro-Toro, and Paul Plöger. "Real-time Convolutional Neural Networks for Emotion and Gender Classification." *arXiv preprint arXiv:1710.07557* (2017).
2. https://www.kaggle.com/c/challenges-in-representation-learningfacial-expression-recognition-challenge/leaderboard
3. Mollahosseini, Ali, David Chan, and Mohammad H. Mahoor. "Going deeper in facial expression recognition using deep neural networks." *Applications of Computer Vision (WACV), 2016 IEEE Winter Conference on*. IEEE, 2016.
4. Simonyan, K., & Zisserman, A. (2014). Very Deep Convolutional Networks for Large-Scale Image Recognition. Arxiv.org. Retrieved 18 November 2018, from https://arxiv.org/abs/1409.1556
5. Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. "Imagenet classification with deep convolutional neural networks." In *Advances in neural information processing systems*, pp. 1097-1105. 2012.
6. Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner. Gradient-based learning applied to document recognition. Proceedings of the IEEE, november 1998
7. Tang, Yichuan. "Deep learning using linear support vector machines." arXiv preprint arXiv:1306.0239 (2013).

Thank You