# Performance Comparison of Deep Learning Methods in Facial Emotion Recognition

Sai Krishna Karthikeya Dulla, Shahidhya Ramachandran, Sriram Sitharaman

## ABSTRACT


- Facial Emotion Recognition (FER) plays a vital role in Monitoring systems, Entertainment, Consumer Marketing, Education, Health, etc.
- This project presents a comprehensive analysis of the performance of various Deep Learning network architectures like AlexNet, VGGNet, ResNet and CapsuleNet.
- Attempted to minimize memory and runtime requirements by compressing the network using it's Dark Knowledge.


## DATA SOURCE

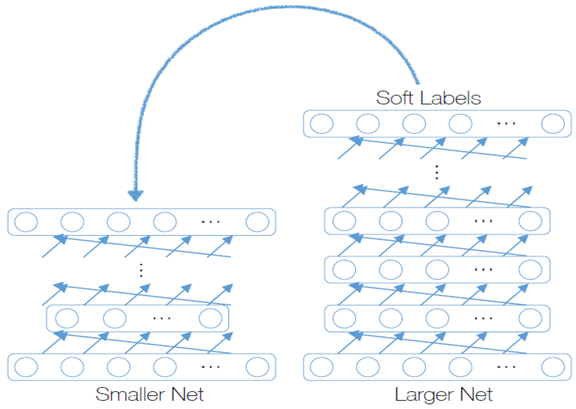
- AffectNet contains nearly One Million facial images collected from the Internet by querying three search engines using 1250 emotion related keywords
- 420K images were manually labelled with one emotion among Neutral, Happy, Sad, Surprise, Anger, Fear and Disgust
- Models were trained on a set of 50K images (resized to 96x96) created by sampling equal number of images from each class


## CHALLENGES

- Humans use a wider range of facial expressions than the seven basic expressions, with some expressions being combinations of the seven [1]
- Models developed for FER must be robust to changes in illumination, poses, ethnicity, race etc.
- Systems trained using posed indoor images do not generalize well on actual real-time images
- Deep Learning models can be memory and time consuming to deploy in hand-held devices

## METHODS

- Trained the following model architectures on Affectnet data:
  - AlexNet
  - VGGNet
  - Dark Knowledge (Ensemble)
  - ResNet
  - CapsuleNet
  - Representational Autoencoders
- For the Dark Knowledge [2] based network, the soft labels obtained from an ensemble of AlexNet, VGGNet and ResNet were used to train a smaller network having fewer parameters without affecting the accuracy



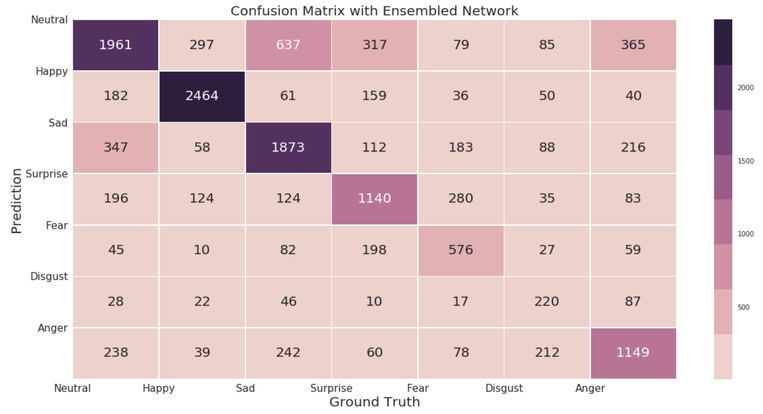

Dark-Knowledge Architecture (Image Source: ENGR-E 533 "Deep Learning Systems" Lecture 09: Network Compression)

## RESULTS

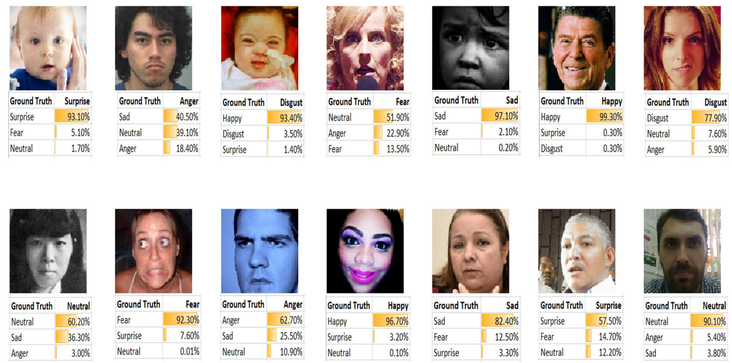
### PERFORMANCE

| MODEL | ACCURACY | TOP-2 |
|---|---|---|
| AlexNet | 59.09% | 79.24% |
| VGGNet | 61.40% | 81.38% |
| ResNet | 56.72% | 76.58% |
| CapsuleNet | 50.17% | 65.38% |
| Representational Autoencoders | 32.56% | 48.30% |
| Ensemble Network | 62.40% | 82.05% |
| Dark Knowledge | 62.04% | 82.01% |

## CONFUSION MATRIX




## PREDICTIONS




## CONCLUSION

- The performance of a Deeper network (VGGNet) is not significantly greater than a shallow network (Alexnet)
- The smaller net with Dark Knowledge is performing similar to the larger net with 91% reduction in network parameters
- Top-2 Accuracy is ~20% greater than the top-1 accuracy. Thus, Fine-grained classification models can be employed to capture subtle differences between expressions

## REFERENCES


1. S. Du et. al. Compound facial expressions of emotion. Proc. of the National Academy of Sciences, 111(15):1454–1462, 2014.
2. G. E. Hinton, O. Vinyals, and J. Dean. Distilling the knowledge in a neural network. arXiv:1503.02531v1, Mar. 2015


**Acknowledgements** : We would like to thank Professor David Crandall and the AIs for their invaluable guidance