# Emotion Recognition on AffectNet via CNNs

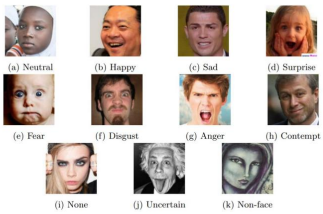
Oseas Ayerdi, Christopher Bucknell
CS230 Project — Stanford University
{oseas, bucknell}@stanford.edu

## Motivation

- Emotion recognition has a variety of applications:
  - Healthcare.
  - Testing user experience of software, products and services.
- Through Convolutional Neural Networks we seek to recognize emotions from pictures of faces.
- Ecosystem to success in the task:
  - Availability of large annotated datasets.
  - Large-scale computational power.

## Dataset

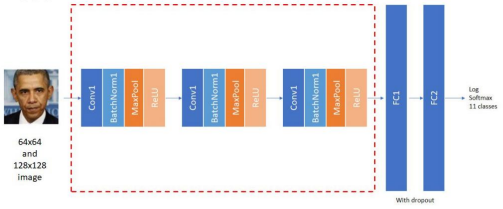
- We used 390,969 manually annotated images from AffecNet (Mollahosseini et al., 2017), labeled into 11 categories:



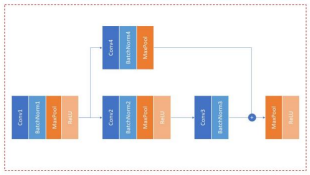

- Resized images to 128x128 px.
- Divided data into:
  - Train: 90% [351,874]
  - Validation: 5% [19,542]
  - Test: 5% [19,553]
- Types of data augmentation:
  - Flip & Color Jitter

## Architectures

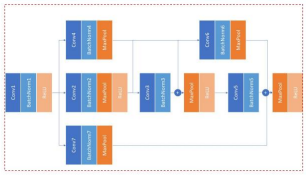
Base model:



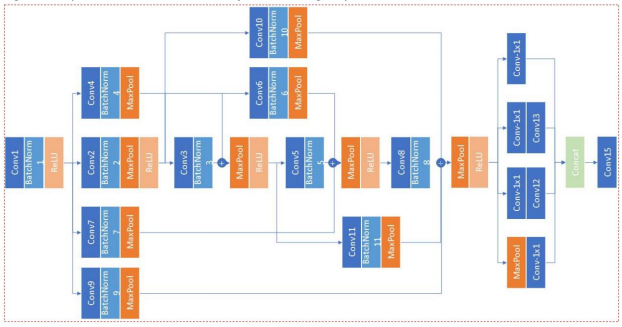

ResNet:



DenseNet:



Hybrid (DenseNet + Inception Layer):

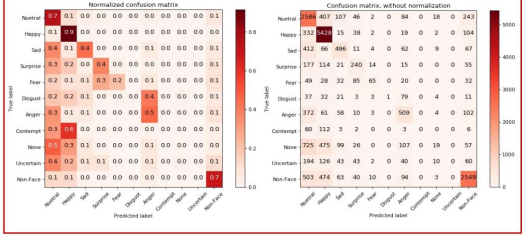


## Results

Accuracy on Validation Set

| Model | Accuracy (%) | # of Parameters |
|---|---|---|
| Human Error w/o Training | 57.0 | N/A |
| Baseline (64) | 60.2 | 1,143,307 |
| Baseline (128) | 60.7 | 4,289,803 |
| ResNet (64) | 60.7 | 2,613,227 |
| ResNet (128) | 61.1 | 9,691,115 |
| Baseline (128, DA) | 60.2 | 4,289,803 |
| Human Error w/ Training | 63.0 | N/A |
| DenseNet (128, DA) | 63.2 | 22,320,971 |
| Hybrid | 63.4 | 11,961,579 |

Error Analysis: Confusion Matrix (DenseNet)




## Conclusions

- Conclusions
  - While changing the network architecture, we achieved human-level performance on overall accuracy at a huge computational cost.
  - Fear category was the most improved category, changing weights in loss function could improve others.
- Future work:
  - Salient maps to understand mechanics of network.
  - Reduce # of parameters with 1x1 convolutions.
  - Increase number of inception layers or better cost function.