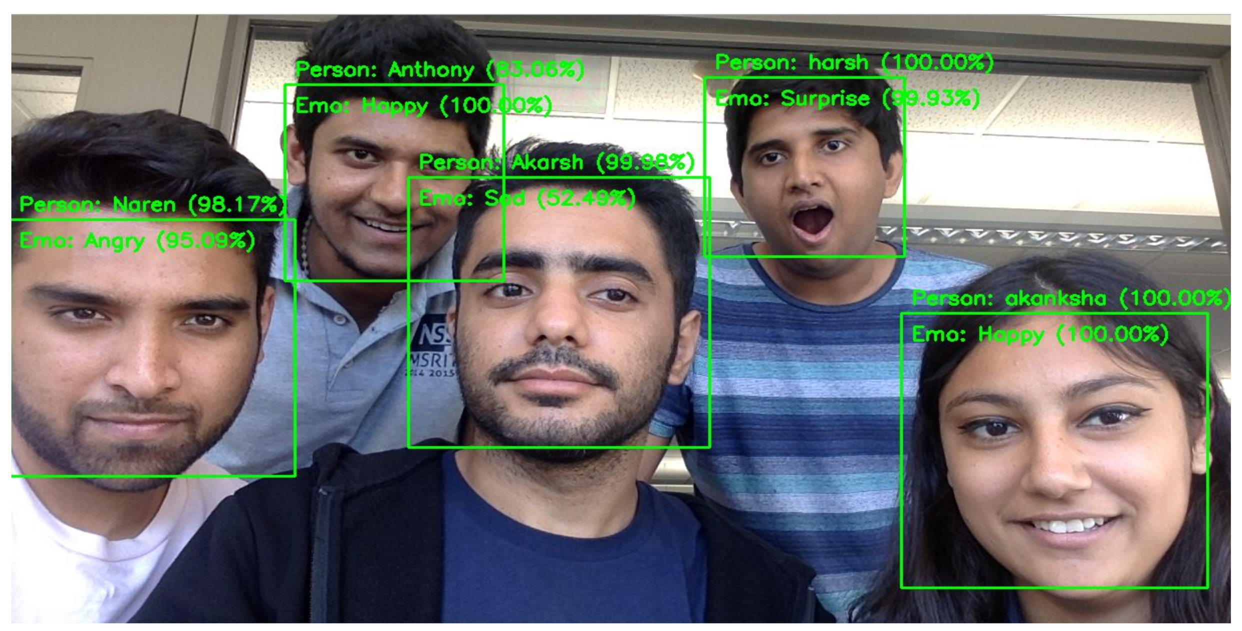
Crowd Analysis by Face Recognition and Expression Detection

Akanksha Diwedy[1], Akarsh Dang[1], Anthony Prajwal Prakash[1], Harsh Waghela[1], Naren Divvala[1]

[1] Department of Computer Science

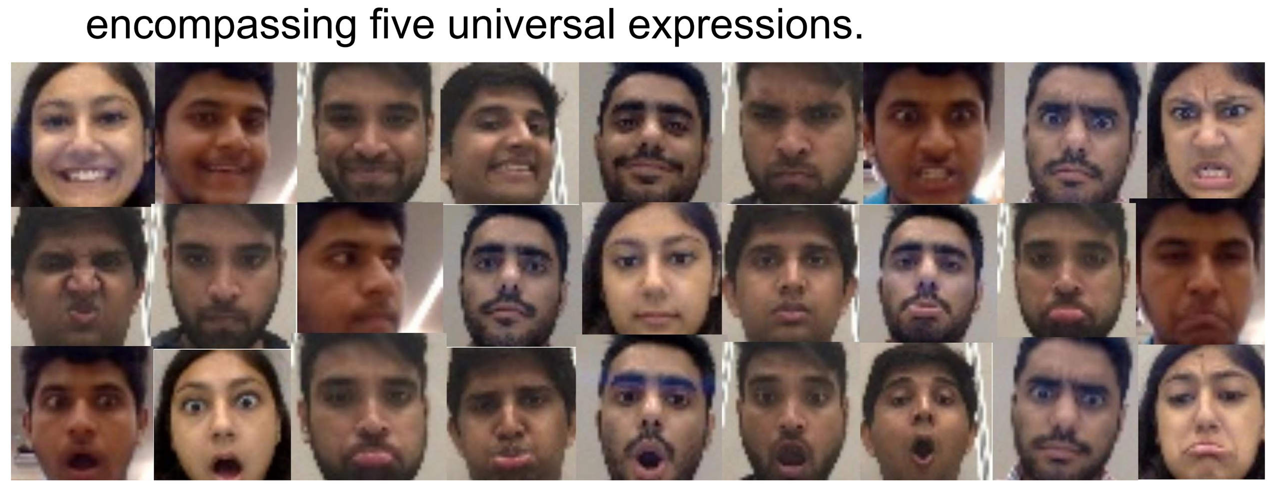
Abstract

- *We aim to successfully design a Multi-Label Convolutional Neural Network which performs two tasks:*
  - *Detecting the emotion on a person's face from one of five universal expressions (i.e. happy, sad, anger, surprise and neutral)*
  - *Recognize the identity of the person*
- *Face identification and Expression recognition have not been explored independently.*
- *Many theoretical models discussing this possibility and effects on the accuracy of the classification but were not backed by some practical or experimental figures.*

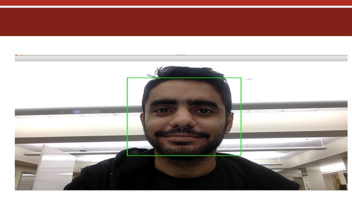


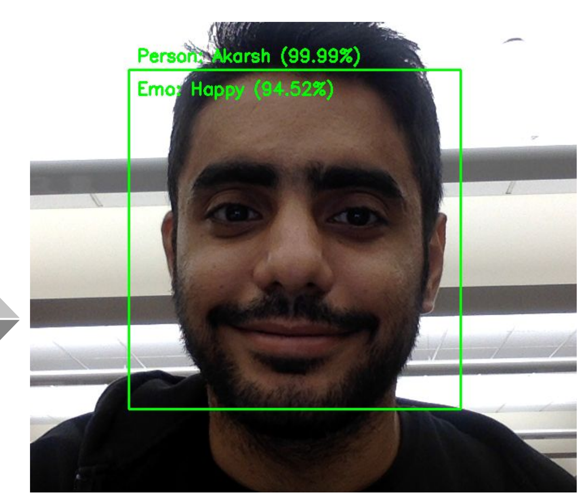
Data and Data Preprocessing

- We curated our own dataset of approximately 5000+ images of five people.
- We clicked and collected images of all five subjects in different lighting conditions and labelled every image based on subject name within emotion directories.
- The dataset consists of 1000+ images per person encompassing five universal expressions.

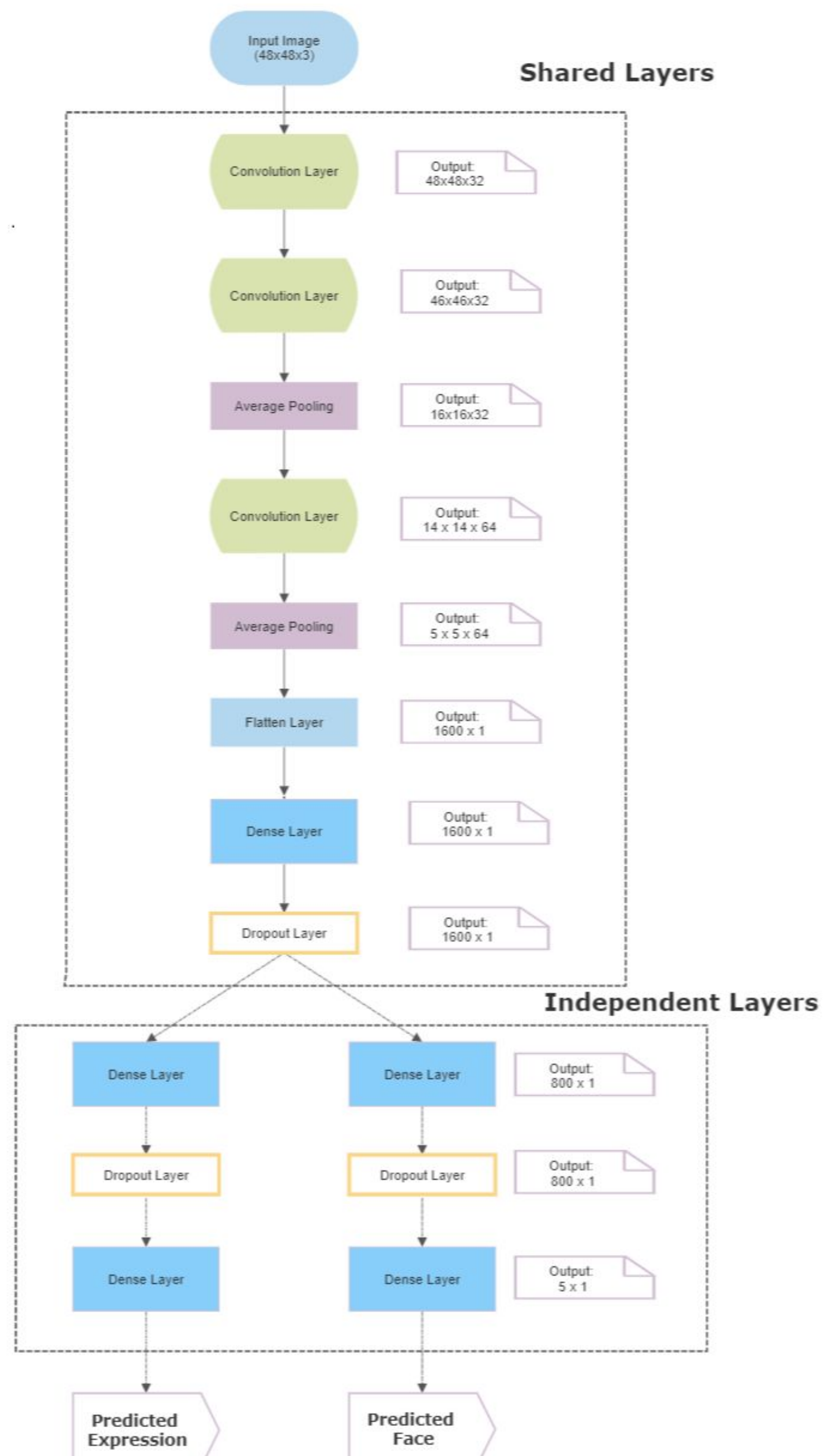


- **Taking Real time video feed**
- **Image Capture(Frame by frame)**
- **Face Detection**
- Scaling Images (48 x 48 pixels)
- Normalizing
- One hot encoding
- ZCA Whitening
- Horizontal Flip
- Rotation
- Zooming
- Feeding resulting image to network





Model

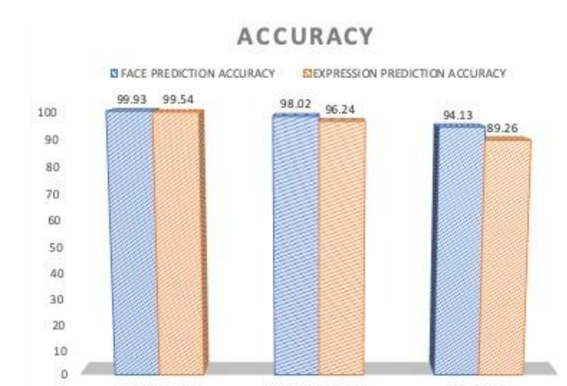


Approach

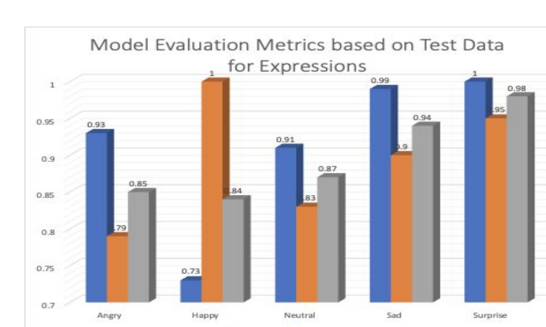
- As Face Recognition and Expression Detection involve learning some common features of the human face, we have proposed a model that uses shared layers to identify and learn these features.
- In addition to these shared layers, our model contains parameters independent to each task.
- We use two separate Loss Functions for each task. Our network trains by using the Adam Optimizer to minimize the sum of the two losses.
- We use Softmax with Categorical Cross entropy as our loss function because its proven to work well with Multi-Label classification.
- Using this loss, we train a CNN to output a probability over five expressions in addition to the subjects identity.

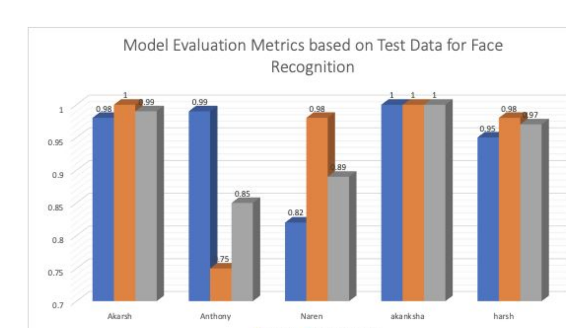
Results

- The model performed well on test data achieving an accuracy of 94% for face prediction and 89% for expression prediction.
- The shared model approach achieved the following accuracy for training, validation and test datasets.



- The performance of the test data was evaluated using Precision, Recall and F1-score. Our observations can be summarized as below.





- We can see above that Surprise and Sad are the easiest expression to identify for our model.
- Happy was the most incorrectly predicted class in the testset with the lowest precision value of 0.73.
- High precision was observed for all our subjects and Akanksha's identity was most easily recognizable by the model.
- Anthony was the subject with most incorrect predictions having 127 misclassified images.

- The below tables help us visualize the misclassifications made by the model on test data.

Confusion Matrix for Expression Detection (Test Data)

	Angry	Happy	Neutral	Sad	Surprise
Angry	395	79	20	6	0
Happy	0	500	1	0	0
Neutral	3	82	415	0	0
Sad	28	4	14	442	1
Surprise	1	18	4	0	474

Confusion Matrix for Face Recognition (Test Data)

	Akarsh	Anthony	Naren	Akanksha	Harsh
Akarsh	499	0	1	0	0
Anthony	7	373	104	0	16
Naren	0	0	486	0	11
Akanksha	0	0	0	489	0
Harsh	1	2	5	0	493

Discussion

- We started with comparing the pros and cons of having a separate network for each of the two tasks or having a single network performing multi-label classification.
- We observed that the model can be sensitive to hue, saturation and brightness. Thus, we trained the model with the images captured in different lighting conditions.
- We observed the effect of different epoch values on the model's accuracy. Optimal performance was achieved by early stopping at the 50 epoch mark.

Conclusion

- Our model successfully predicts the identity and expression of a subject with high accuracy.

In the future we plan on:

- Improving our dataset by capturing images of more expressions such as fear, disgust, contempt etc.
- Enable our model to categorize unidentified subjects to an 'unknown' class.
- Learning the body language of subjects which can give a better understanding of a crowd's behaviour.

References

- *CPU Real Time face detection - Towards Data Science (https://towardsdatascience.com/faced-cpu-real-time-face-detection-using-deep-learning-1488681c1602)*
- *Multi-label classification with Keras - PyImageSearch (https://www.pyimagesearch.com/2018/05/07/multi-label-classification-with-keras/)*
- *Introduction to Emotion Recognition - Algorithmia blog (https://blog.algorithmia.com/introduction-to-emotion-recognition/)*
- *Octavio Arriaga, Paul G. Ploger, Matias Valdenegro: Real-time Convolution Neural Networks for Emotion and Gender Classification*
- *A Brief Review of Facial Emotion Recognition Based on Visual Information (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5856145/)*
- *How brands are using emotion-detection technology (https://econsultancy.com/how-brands-are-using-emotion-detection-technology/)*