Real Time Emotion Recognition using Convolutional Neural Networks

Prepared under the guidance of :
Dr. M.P.S. Bhatia

Submitted by :
Deepak Rathi(244/CO/14) Devesh Saini(247/CO/14) Harshil Yadav(258/CO/14)

Problem Statement

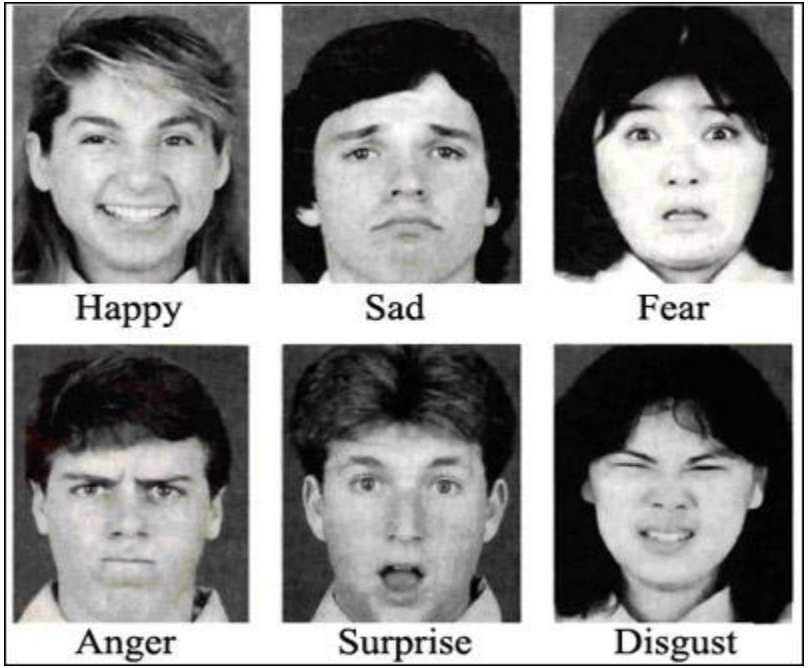
The problem that we attempt to tackle in this project is Real time detection of Emotions using Convolutional Neural Networks (CNN).

Introduction

- As humans, we classify emotions all the time without knowing it. Humans are well-trained in reading the emotions of others, in fact, at just 14 months old, babies can already tell the difference between happy and sad. But can computers do a better job than us in accessing emotional states?
- Emotion is any conscious experience characterized by intense mental activity and a certain degree of pleasure or displeasure.Emotion is often intertwined with mood, temperament, personality, disposition, and motivation.
- Emotion classification, the means by which one may distinguish one emotion from another, is a contested issue in emotion research and in affective science.

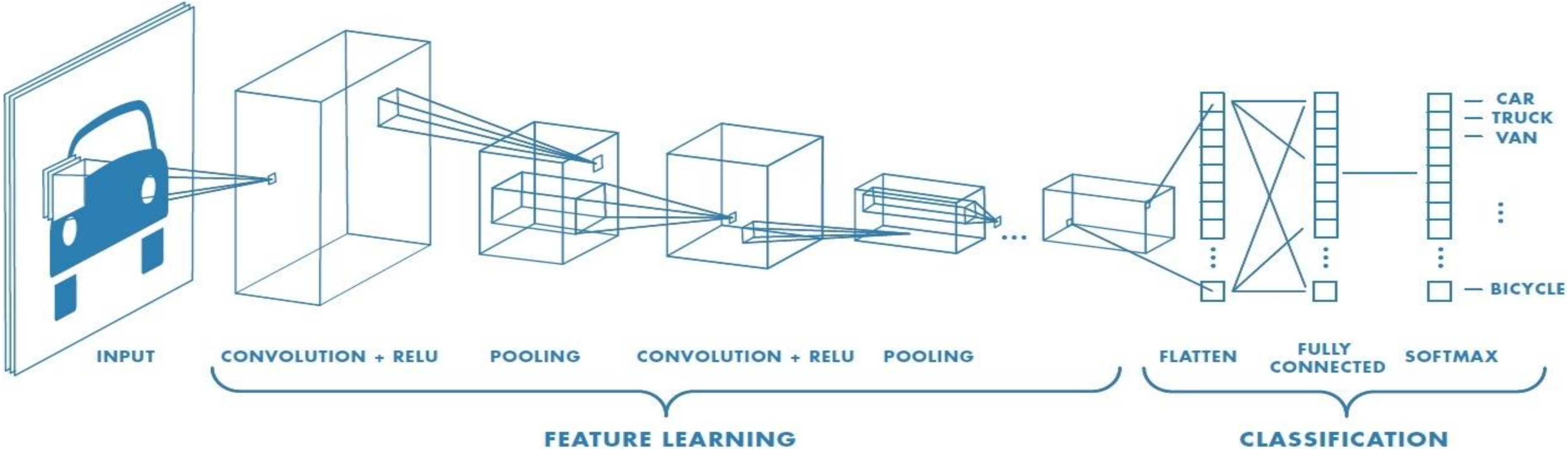
Basic Emotions

- The six basic emotions we wish to classify are : anger, disgust, fear, happiness, sadness and surprise.



Convolutional Neural Networks

- Convolutional Neural Networks (ConvNets or CNNs) are a category of Neural Networks that have proven very effective in areas such as image recognition and classification. ConvNets have been successful in identifying faces, objects and traffic signs apart from powering vision in robots and self driving cars.



Literature Review

1. Deep Convolutional Neural Network were used for Emotion Recognition on Kaggle Dataset [1].
2. Existing architectures such as AlexNet, VGG-16, Inception, and Inception-Resnet [2] were used to classify images and Saliency maps were used to analyse results.
3. AlexNet, GoogLeNet, and ResNet were used on the TinyImageNet dataset[3].
4. Bimodal Classification[4][6] studied.
5. Powell's Direction Set Method[5] was used for back propagation on Cohn-Kanade dataset.

Objective

The main objective is to develop Emotion Recognition Application using the best architecture .

1) Achieve state of the art Convolutional Neural Network for emotion classification problem.
2) Perform emotion classification on the Kaggle ILSVRC2016 dataset with good accuracy.
3) Develop an application that should:
   - Take live video feed as input.
   - Process the video frame by frame using CNN and identify various relationships in the image.
   - The features extracted should be able to classify emotions in real time.

User Interface

User interface of the project is made using a web application. The web application is developed using Ruby as a language and Rails as the platform. Rails is a model–view–controller (MVC) framework, providing default structures for a database, a web service, and web pages. It encourages and facilitates the use of web standards such as JSON or XML for data transfer, and HTML, CSS and JavaScript for display and user interfacing.

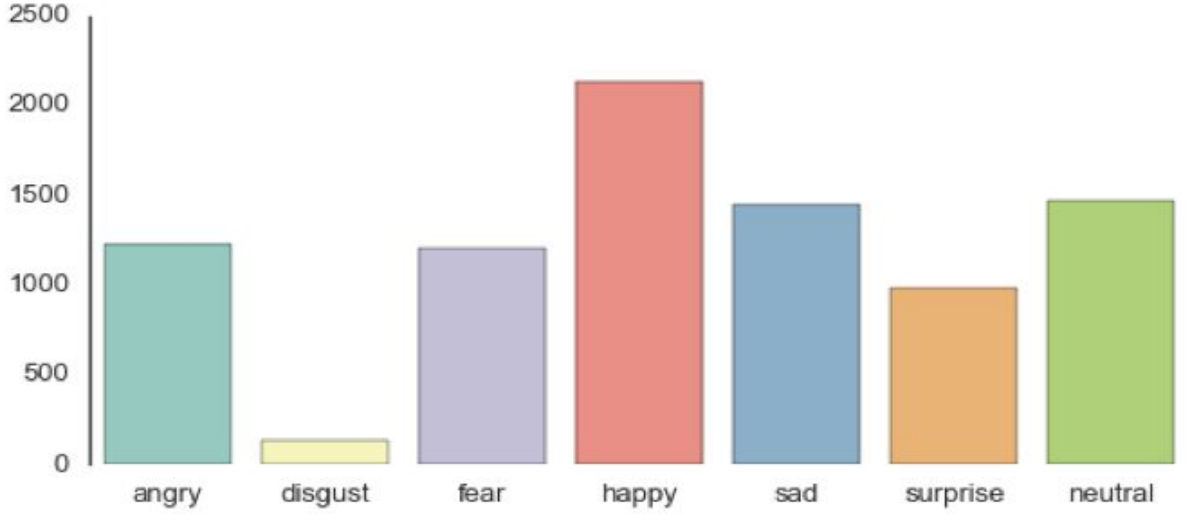
- Front-end consists of layout and buttons.
- Back-end consists of routes, controller and process handling.

Demonstration

Implementation

Dataset Collection

We used the dataset from the Kaggle challenge on Facial Expression Recognition, which gives 48x48 pixel grayscale images of faces.The dataset contains images that vary in viewpoint, lightning and scale. It consists of 35,887 images.

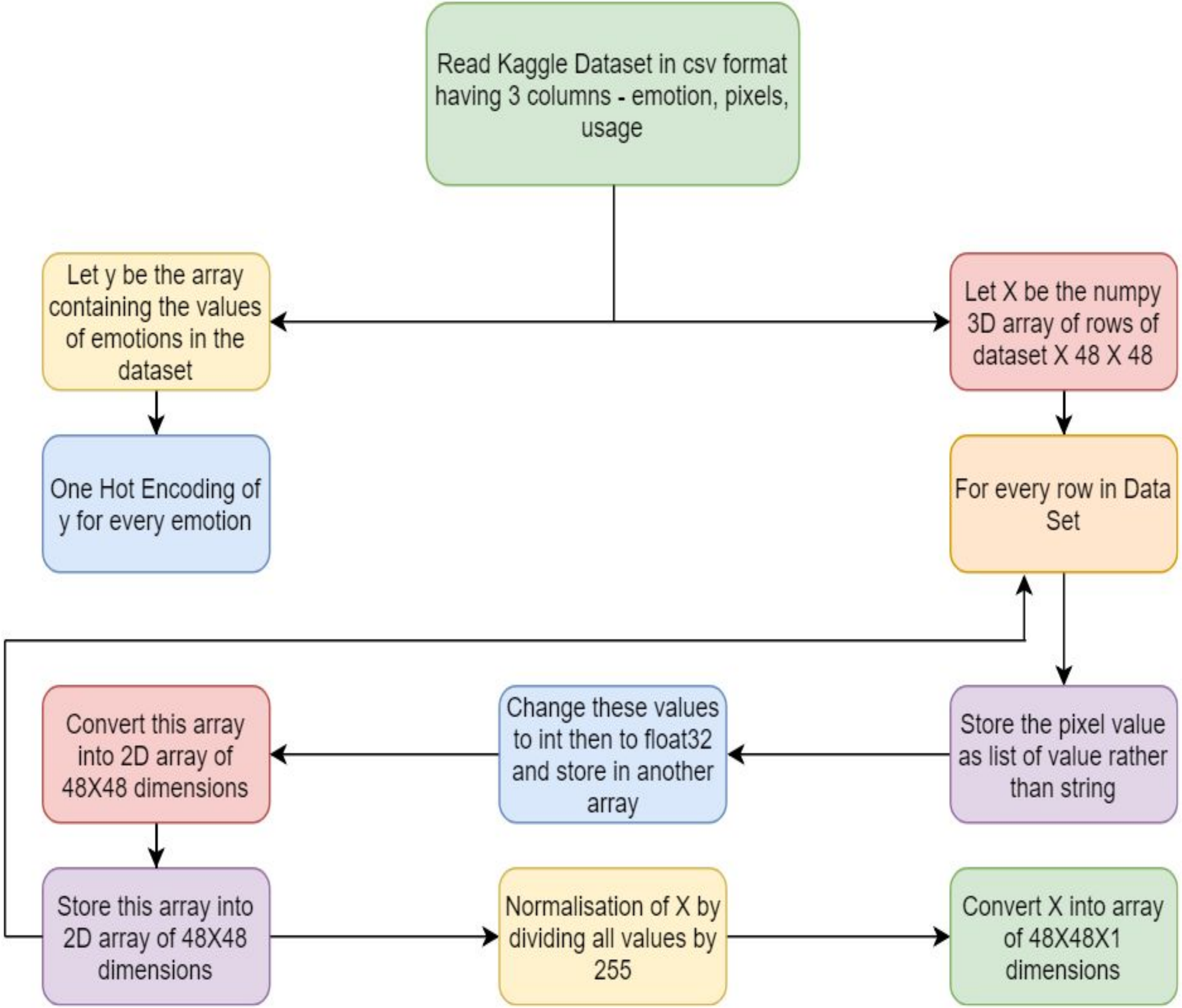


Data Preprocessing

1) The train.csv file contains three columns, "emotion", "pixels" and usage. The "emotion" column contains a numeric code ranging from 0 to 6, inclusive, for the emotion that is present in the image.
2) The "pixels" column contains a string surrounded in quotes for each image.
3) The contents of this string a space-separated pixel values in row major order.
4) The data set is split into training , validation and test set in the ratio 80:10:10.

The steps followed are shown in the next figure

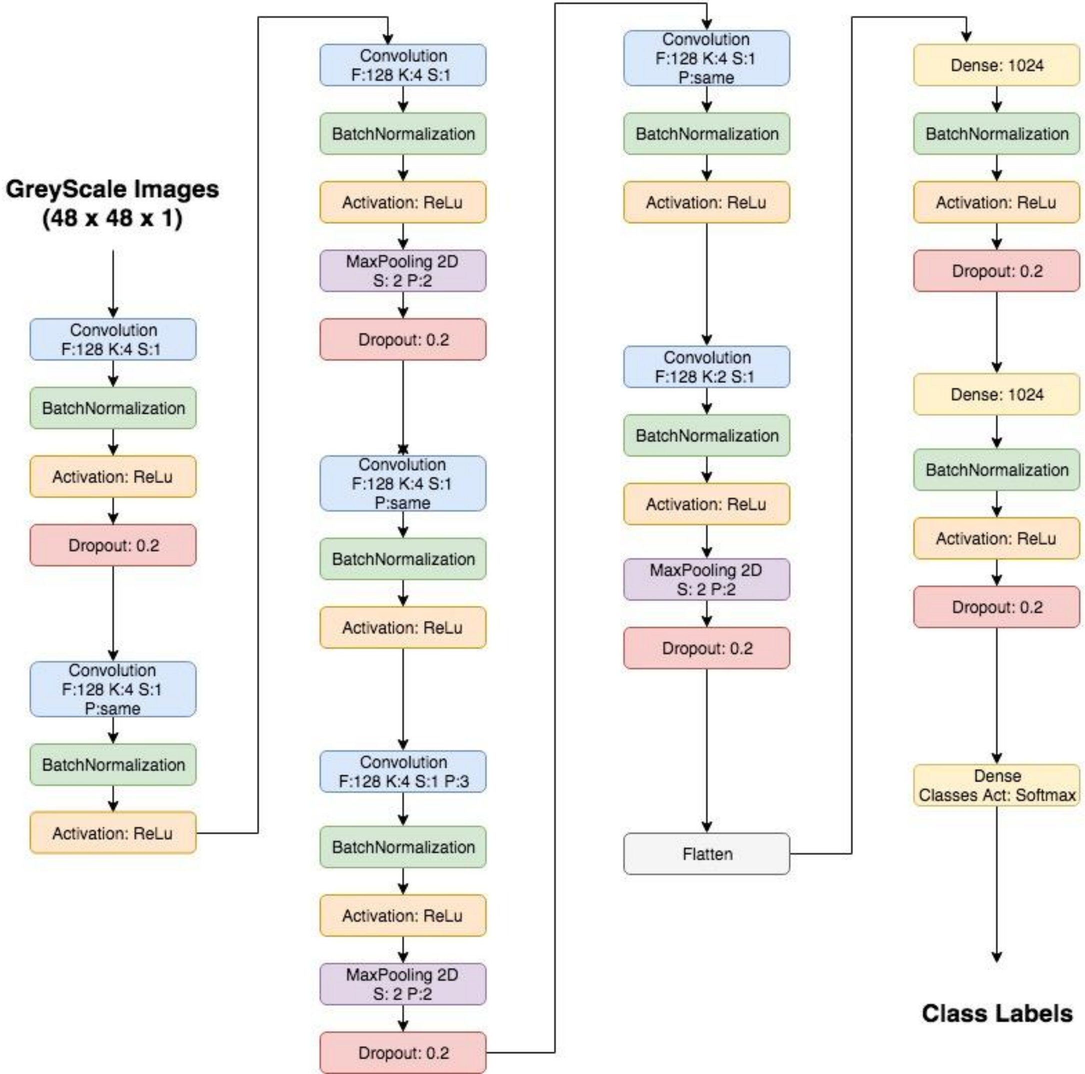
Steps in Data Processing



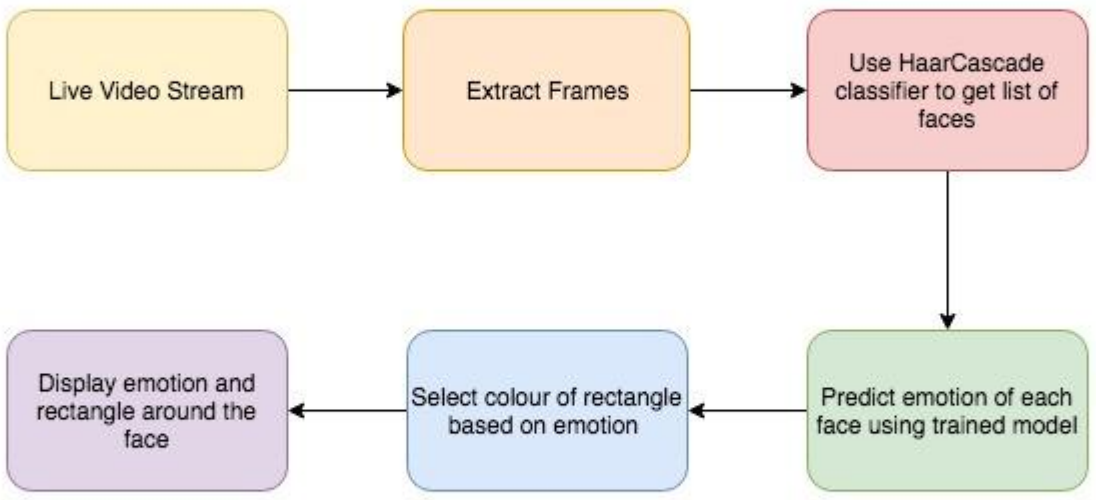
PROPOSED MODEL

Model Used

The model was built entirely using Keras. We would run each of our models for 16 epochs on the cloud after which the performance seemed to level off.



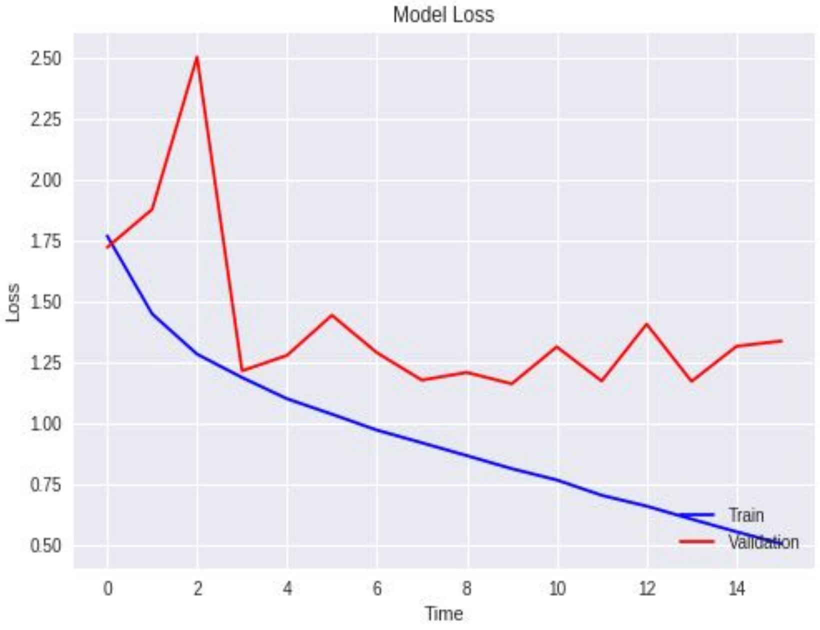
Video Processing



Result Visualization and Analysis

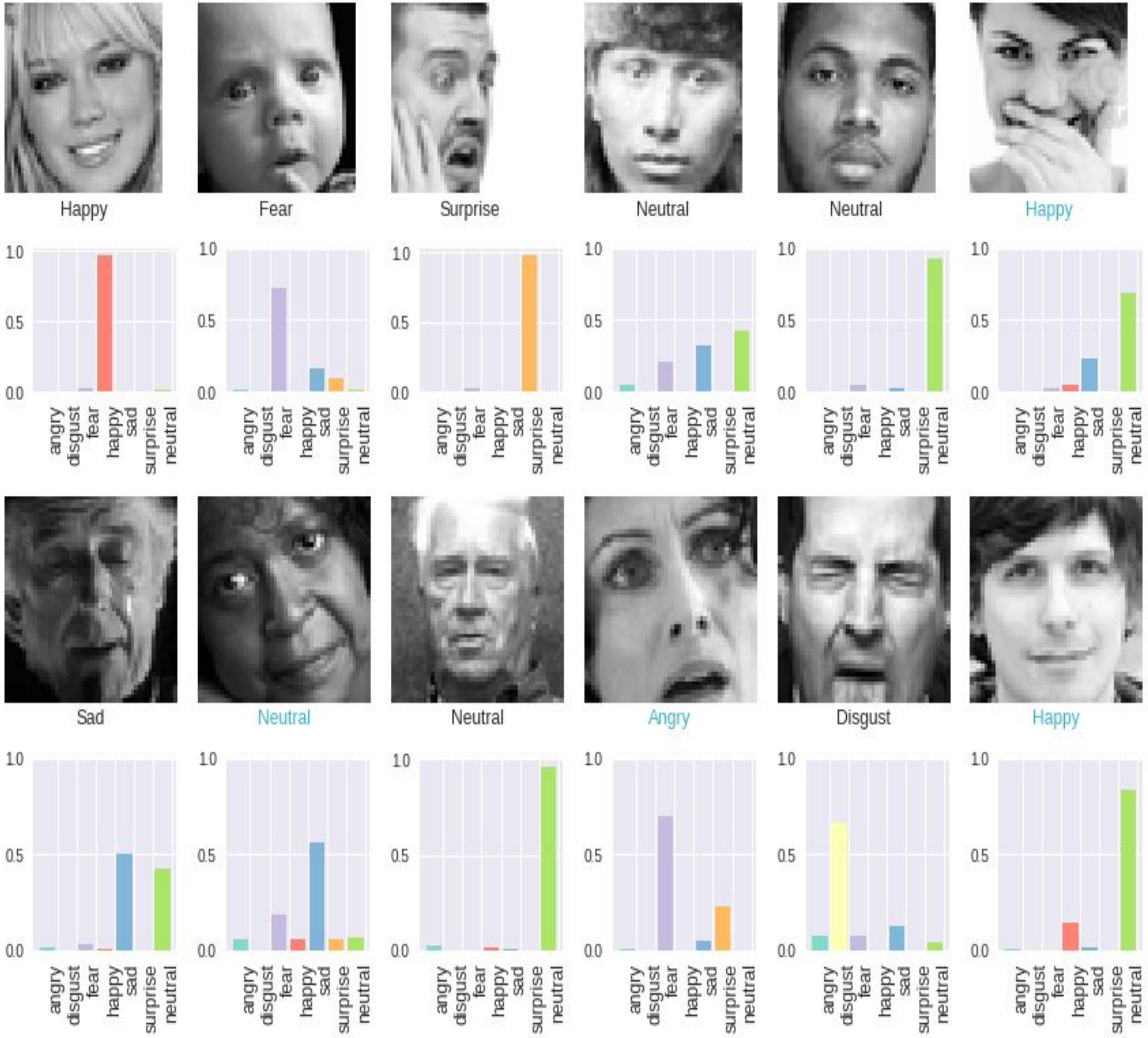
Set	Accuracy	Loss
Training Set	84.78	0.434
Validation Set	61.52	1.170
Test Set	62.72	1.157

Training v/s Validation Accuracy Graph

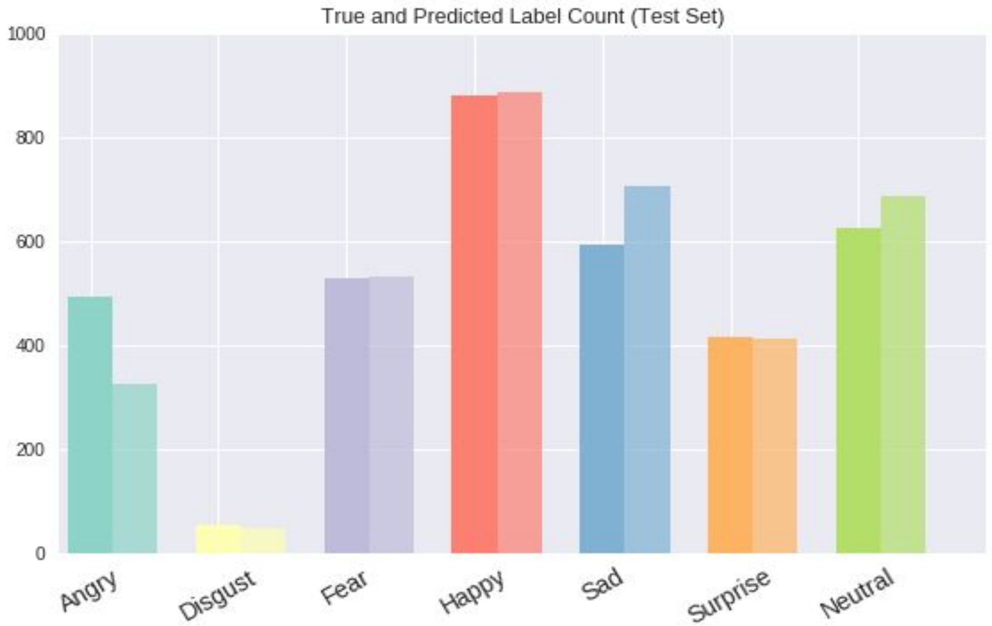


Training v/s Validation Loss Graph

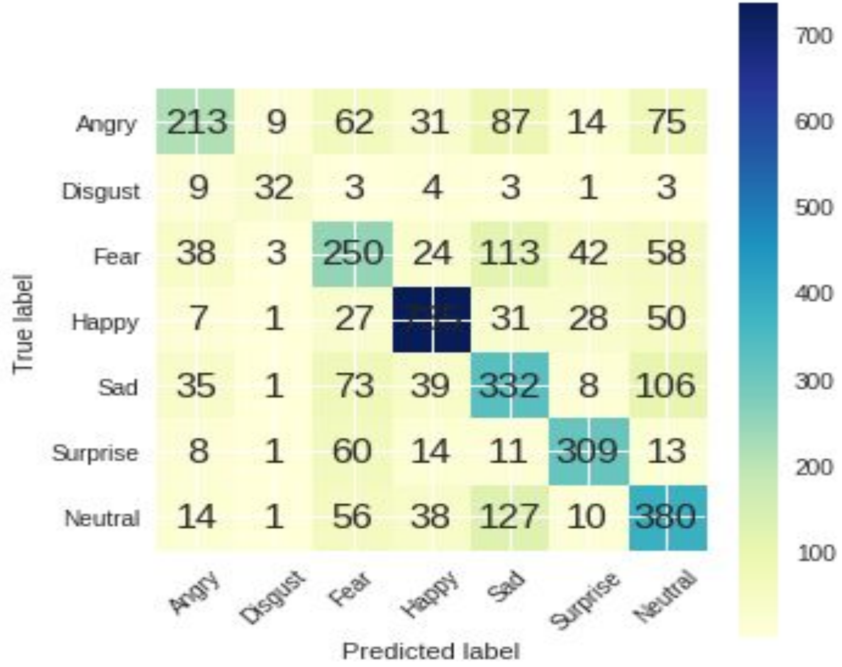
Model Predictions on Dataset Images



True and Predicted Label Count in Test Set

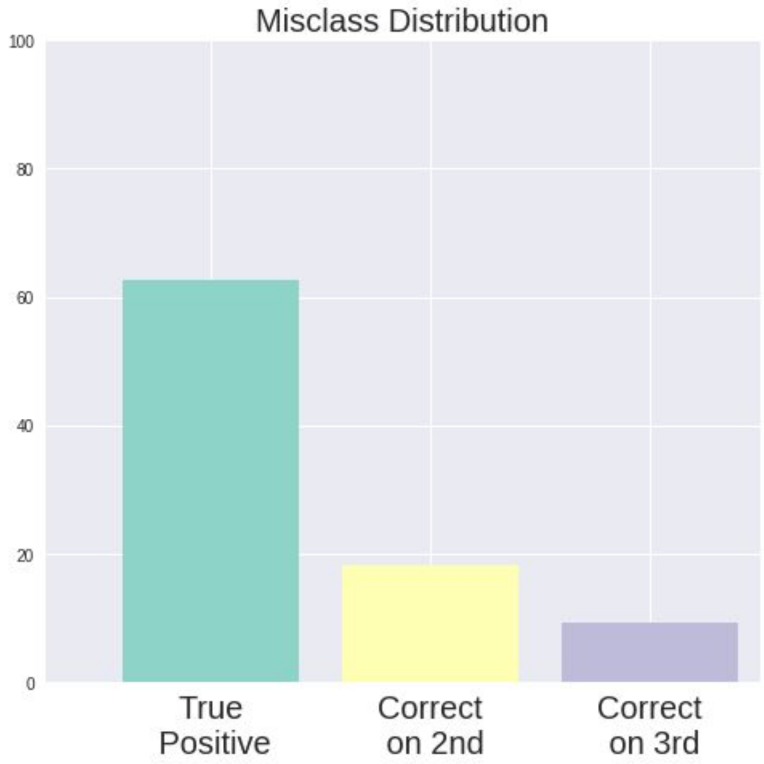


Confusion Matrix



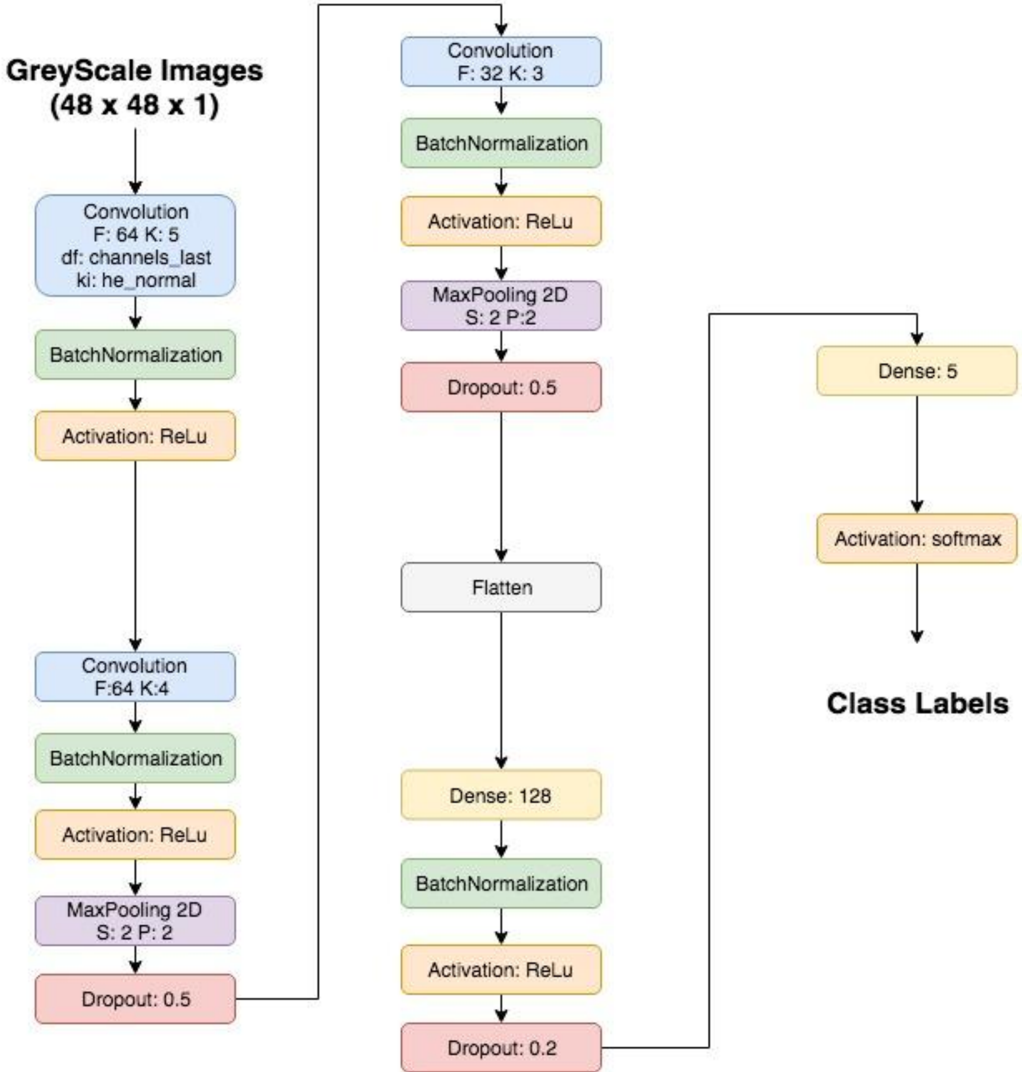
Classification Report

	precision	recall	f1-score	support
angry	0.57	0.53	0.55	491
disgust	0.95	0.38	0.55	55
fear	0.58	0.33	0.42	528
happy	0.83	0.84	0.84	879
sad	0.45	0.62	0.52	594
surprise	0.80	0.76	0.78	416
neutral	0.57	0.65	0.60	626
avg / total	0.65	0.64	0.63	3589

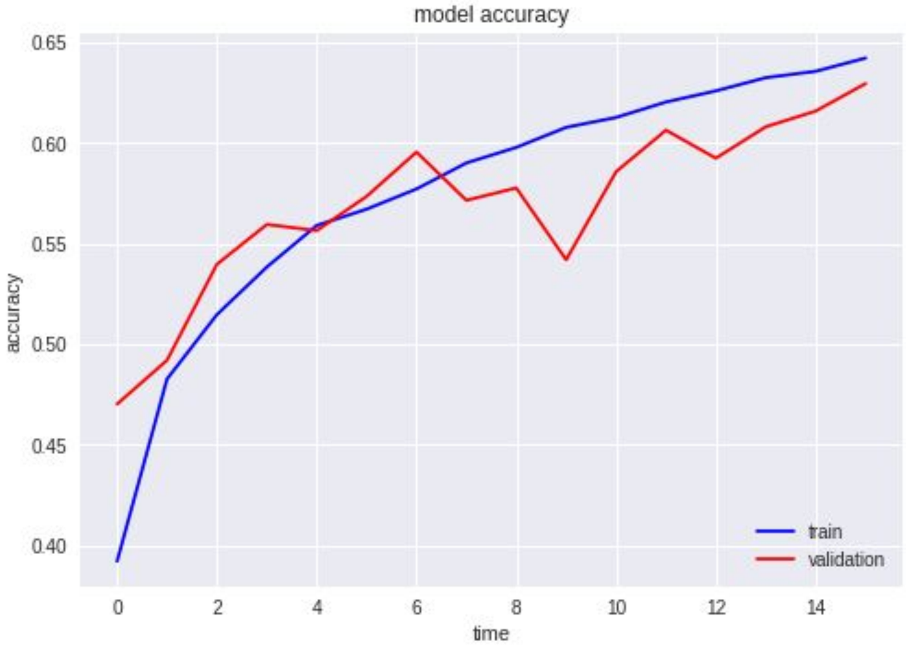


Model 2

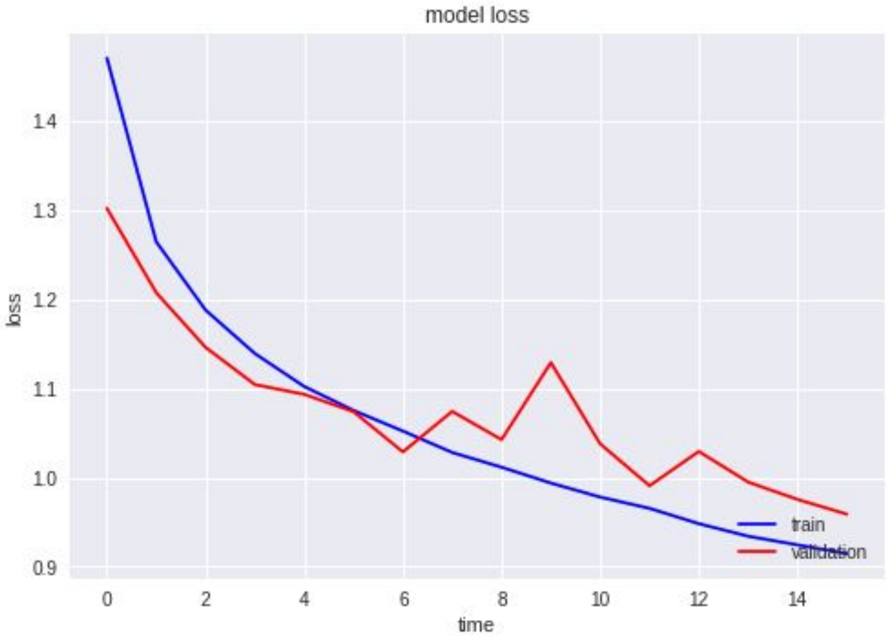
Architecture of 5-classes Recognition Model



Accuracy and Loss Graphs of Model 2



Training v/s Validation Set Accuracy Graph



Training v/s Validation Set Loss Graph

Result Comparison

Challenges in Representation Learning: Facial Expression Recognition Challenge was hosted on Kaggle.com in 2013. The best solution had an accuracy of 71% and top 10 submissions had an accuracy of 60%. We were able to achieve an accuracy of 64.59% on the test set by using model 2.

Metrics	Model 1	Model 2
Validation Set Accuracy	61.52	62.94
Test Set Accuracy	62.72	64.59
Classification Classes	7	5
Trainable Parameters	4,540,039	358,533
Total Parameters	4,545,927	359,173

Conclusion

- In this project, we sought to classify the image of a face into the seven basic human emotions. We developed and experimented with the architecture of a deep convolutional neural network ourselves, and performed a hyperparameter search to optimize our results. We compared different architectures to classify emotions and we were able to reach the state-of-the-art test accuracy of approximately 64.59%.
- However, we believe if we addressed the overfitting of the training data, we could reach even higher test accuracies.
- We also developed a web application to demonstrate our project and provide interactive interface to the user.
- We have described a holistic, non intrusive approach to emotion detection, by checking for user's facial expressions. These are powerful measures, even for low image resolution and in-the-wild circumstances such as bad illumination, facial expressions, non-frontality etc.

Future Work

1) **Build a REST API and add functionalities to Website :** REST API can be built that finds human faces within images and make prediction about each facial emotions. User must be able to paste the url of an image or drag-and-drop an image file.,have the API return with annotated faces and cropped thumbnail. The API must return probabilities and an unique ID for image in JSON format.

2) **Extend to Bimodal Classification :** Ensemble Learning can be used to obtain predictions by combining both audio and video. This will improve the accuracy of emotion detection.

3) **Transfer Learning :** It is storing knowledge gained while solving one problem and applying it to a different but related problem. Our model trained on ImageNet dataset can be fine tuned and results can be analyzed.

References

- Touchy Feely: An Emotion Recognition Challenge Authors : Dhruv Amin Stanford University, Patrick Chase Stanford University, Kirin Sinha Stanford University
- Facial Expression Recognition for wild images with analysis from Saliency maps Authors: Priyanka Rao ,Ling Li
- Deep Convolutional Neural Networks for Tiny ImageNet Classification Hujia Yu Stanford University
- Recognising Facial Expressions Using Deep Learning Authors : Alexandru Savoiu Stanford University, James Wong Stanford
- Neural Networks for Emotion Classification by Yafei Sun
- Analysis of Emotion Recognition using Facial Expressions, Speech and Multimodal Information Carlos Busso, Abe Kazemzadeh, Sungbok Lee, Ulrich Neumann , Shrikanth Narayanan

THANK YOU