**Project Report Quickdraw Sketches**

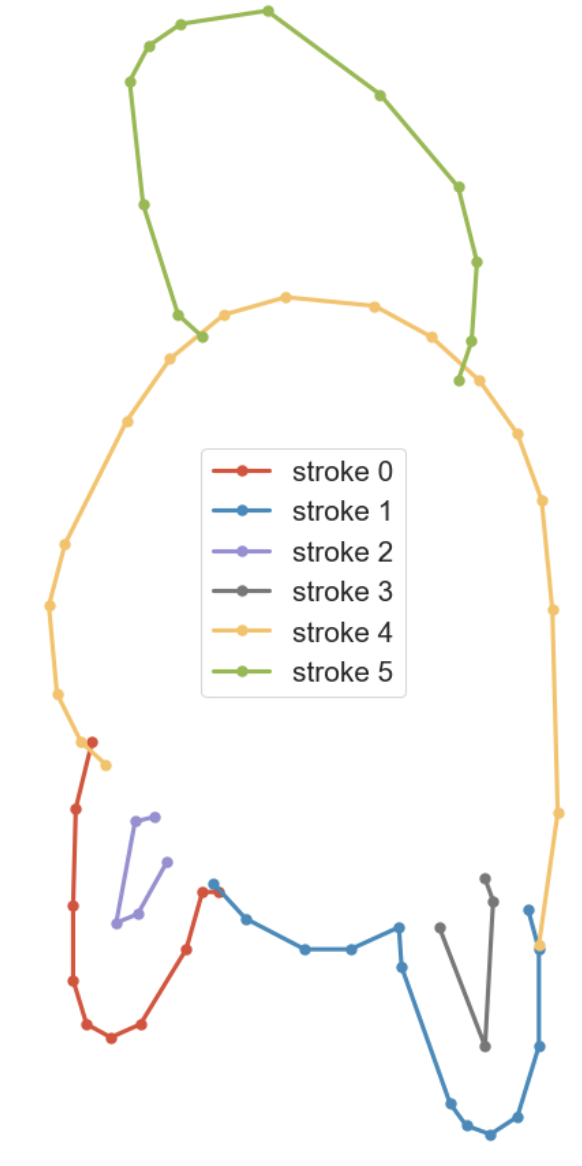
Data:-

<https://github.com/googlecreativelab/quickdraw-dataset>

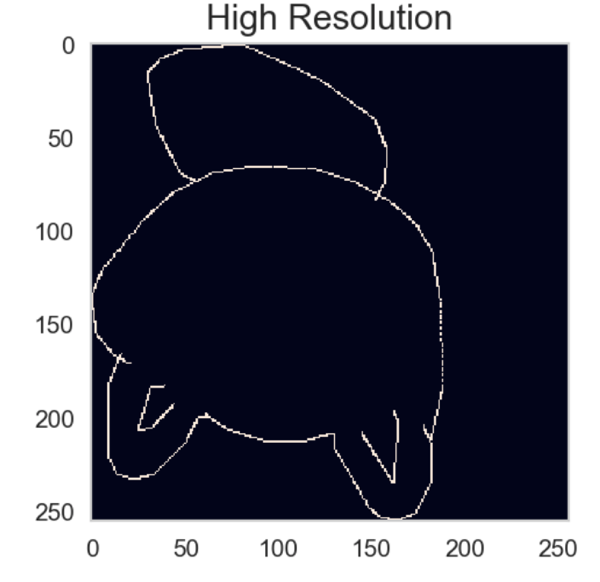
The Google Sketch data contains 517,500 json file in 345 Categories. Each json file signifies one sketch. Where there are arrays of how user draws an image is given in an array considering the canvas as 256\*256 images. They record how the cursor of the user how is drawing is moving and add apending that to list for every stroke.

Preprocessing:

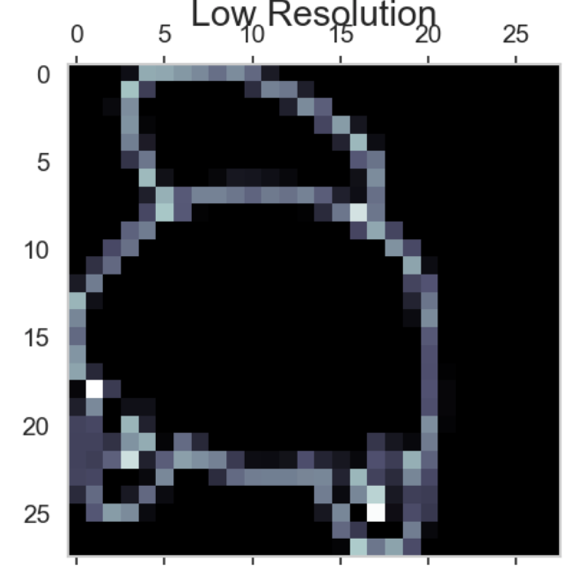
1. The first job is to convert these arrays of strokes to into a single multidimensional array and plotting it. This is a sketch of a bear.

****

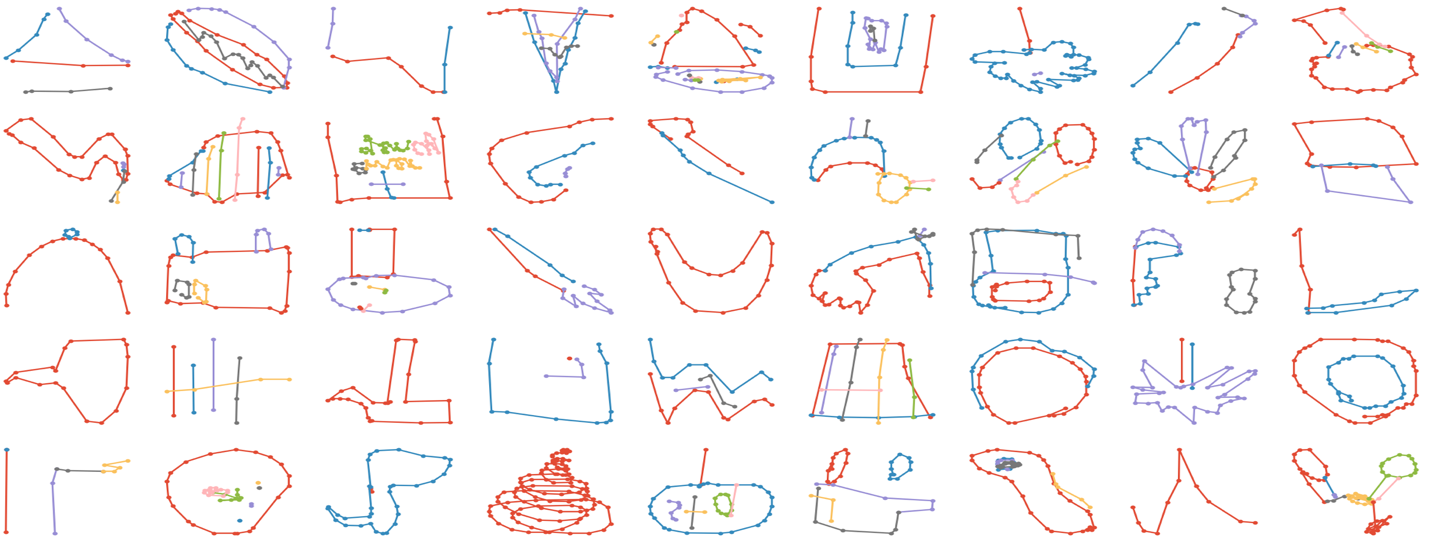
1. The next is to covert each of the array into a black and white image of 256\*256 pixels.

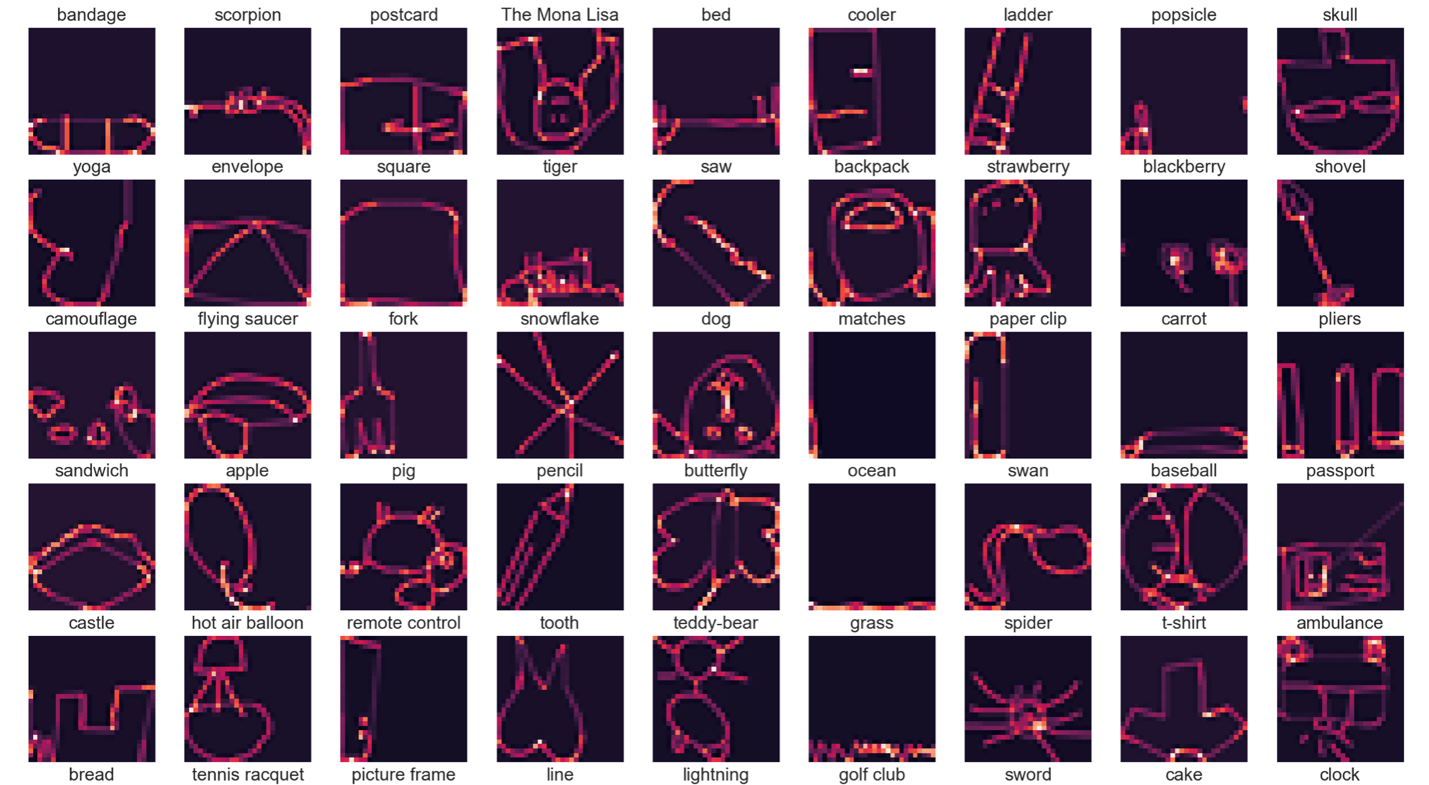
****

1. After that to reduce the dimensions we lower the resolution of the image to 28\*28 pixel image

****

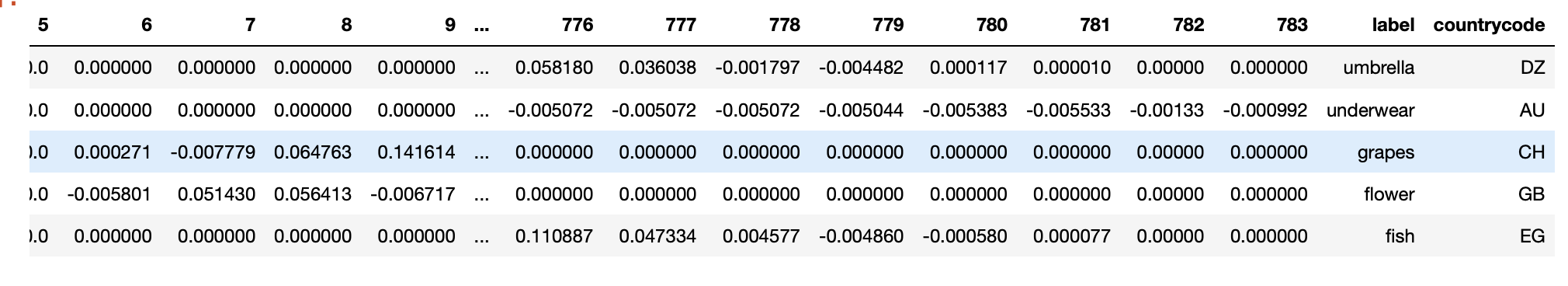
1. These are a set of plotted arrays and and their processed low resolution images

****

****

Initial Training

1. First flattening the multi-dimentional array of 28\*28 to create 784 features.

****

1. Reducing the no of features two half using PCA. Half of the original features are able to capture 90 percent of the variance.

****

1. Tried Algorithms like Naïve Bayes which gave an accuracy of very low as much as 15 percent.

Deep Learning

Seeing such low accuracy and never tried CNN before this project gave me a chance to explore convolution neural network and how they are used for Image classification

I used a small CNN due to processing power constrain and lack of GPU on my machine.

A screenshot of a social media post

Description automatically generated

A screenshot of a cell phone

Description automatically generated

**With that I got an validation accuracy of 47 percent and loss of 2.4252**

A close up of a map

Description automatically generated

This showed that model is not overfitting and a deeper model still can be used to improve the results.

The results also showed that some of the categories are very similar to each other like **undergarments and shorts OR pineapple and strawberry OR blackberry and grapes.**

**A screenshot of a cell phone

Description automatically generated**

These lead to majority of errors while classification. So, what should be done is these conflicting groups should be kept apart when playing the game. Pick us clusters of categories which are very different from each other and then pick random drawing from them.

The other step to improve the accuracy will be to ask the same user to draw both strawberry and pineapple so we can start observing the difference people make while the draw.