Identification and Grading of Freehand Sketches Using Deep Learning Techniques

Chandan C G
Student, Department of ISE
PES Institute of Technology
Bangalore
chandan.cg12@gmail.com

Deepika M M
Student, Department of ISE
PES Institute of Technology
Bangalore
deepikamm12@gmail.com

Suraksha S
Student, Department of ISE
PES Institute of Technology
Bangalore
surakshas27197@gmail.com

Dr. Mamatha H R
Professor, Department of CSE
PES University
Bangalore
mamathahr@pes.edu

***Abstract*—A great deal of work has been done in the zone of computer vision on images. Be that as it may, the region of freehand sketches stays to be an unexplored zone. Freehand sketches fluctuate in light of masterful style. In spite of the fact that they contain minimum details, people can predict the class to which the sketch has a place. Concentrating on such negligibly itemized sketches can help us in understanding the neurobiological procedures that occur in people. In this paper, a Convolutional Neural Network (CNN) has been developed which can classify the freehand sketches in view of specific highlights. Classified freehand sketches are graded relative to the prototype being considered. Publicly available dataset of Eitz et al. is considered for identification and grading. Evaluating the sketches will help in surveying the advance of a client who is figuring out how to draw outlines.**

Keywords— Keras, TensorFlow, Rectified Linear Unit (ReLu), Scale Invariant Feature Transform (SIFT), Adaptive Moment (ADAM) optimizer

I. INTRODUCTION

Sketching is a common place of verbal exchange given that pre-historic instances, pictographs predate the arrival of language through tens of lots of years and these days the capability to draw and understand sketched objects is ubiquitous.

FIG 1 : SAMPLE DATA

Analysing freehand sketches allows comprehension of our neuro-cognitive strategies. Sketches as we understand are freehand drawings that are accomplished that may not usually be completed work.

Sketches serve several purposes: 1. Artists document what they see 2. It's a brief way of representing something. The visible subtle factors in freehand sketches are typically sparse. As said before, sketches of a comparable element can trade in mild of aesthetic style and drawing potential. Moreover, they may be harder to understand than photographs considering the fact that they are less detailed. They delineate characterising features of actual global entities. They push aside facts which seem like much less imperative or hard to attract. In this paper, we make use of convolutional neural network to categorise the sketches and grade them. Convolutional neural networks are neurobiologically inspired feed - forward artificial neural networks. They incorporate multiple layers of neurons. Neurons in every layer are amassed into units.

Our CNN model is trained utilizing the dataset of Eitz et al. [1] containing 20,000 freehand sketches across 250 different classes. Sample data is shown in FIG 1. Our model takes in the features in the sketches. At a point when a freehand sketch is given to the model as an input, it classifies utilizing the features learnt. Likewise, when a sketch belonging to a category is given to the model, it grades the sketch in respect to the first legitimate outline outfitted to the model to start with and it creates a platform where freehand sketches will be classified and graded automatically.

Grading of free hand sketches has been introduced in this paper. Grading is based on the CNN score and SIFT-matching score.

This paper is organized as follows: The related literature is reviewed briefly in section II. Problem definition is discussed in Section III. In section IV, system architecture is presented. The methodology is specified in section V. In section VI, results obtained are described. In section VII, conclusions are discussed and in section VIII, future enhancements are talked about.

II. RELATED WORK

Paper [2] has explored human sketches in large scale. Sketches are being represented as bag of features and multi-class SVM (support vector machine) is used to train and classify sketches.

Freehand sketch recognition framework based on deep features extracted from convolutional neural networks is presented in [3]. Imagenet CNN and modified version of LeNet CNN are the two models of CNN used. The framework is evaluated on a publicly available database of Eitz et al containing thousands of freehand sketches. The results show an improvement by 3 % to 11 % over the state - of - the - art accuracies .

Freehand sketches are drawn as a composition of hand-drawn curves that are added sequentially over time. Freehand sketch recognition methods exhibit two broad themes : domain-specific and general. In domain-specific category, the system will be able to identify sketches belonging to a particular domain. Here, domain specific assumptions are made. In general category, general nature of freehand sketches will be captured.

Convolutional neural systems have been utilized to build the precision of acknowledgment of sketches drawn by various individuals. Authors explain the major difference between sketch recognition problem and customary photographic image classification : First, freehand representations are less visually complex than photos, this is on the grounds that, on account of photos, there are three shading channels for each pixel whereas sketches are Black and White in colour. Sketches essentially contain clear space, though photos have data all through the picture. Second contrast is about the variation. In photographic images, obstacles could be illumination, camera angle etc, whereas, sketches vary based on artistic style.

Paper [4] clarifies about ADAM optimizer, an algorithm for first-order gradient-based optimization of stochastic target functions in light of versatile evaluations of lower-order moments. It has also been stated that this method is straightforward to implement, computationally efficient, has little memory requirements, is invariant to diagonal rescaling of the gradients and is well suited for problems that are large in terms of data and/or parameters.

Paper [5] presents a novel technique for picture closeness measure. The proposed methodology here makes encompassing insight regarding the assessment of non-precise, simple to include portrayed data. The framework would be able to give the client choices of either recovering comparable pictures in the database or positioning the nature of the sketch against a given standard.
The proposed technique can adapt to pictures containing a few complex questions in an homogeneous foundation.

III. PROBLEM FORMULATION

The aim of this paper is to identify and grade freehand sketches. Though a lot of work is being done in computer vision on images, the area of freehand sketches continues to be an inferior area. Also, since sketches contain minimal details, it is harder to distinguish them than colored images. Therefore, it is a challenge to extract features and classify sketches.

The classified sketches are further needed to be graded. In some cases, a few sketches of the same object get a higher score over others because of some angular variation with respect to the prototype being considered as a reference. However, applying some transformations on these sketches could reduce all these sketches to a similar view, thus becoming similar sketches and getting similar scores. All these transformations have to be carried out on the test dataset before the final grades are given for the sketches. This becomes a challenging part again.

IV. SYSTEM ARCHITECTURE

A. IDENTIFICATION

The dataset of Eitz et al. is used to train the model. Many preprocessing steps have been carried out on the sketches in the dataset. Once the model learns the features in the dataset, it is ready to classify sketches given to it as inputs. It classifies the sketches by displaying the class name to which the sketch belongs along with the accuracy.

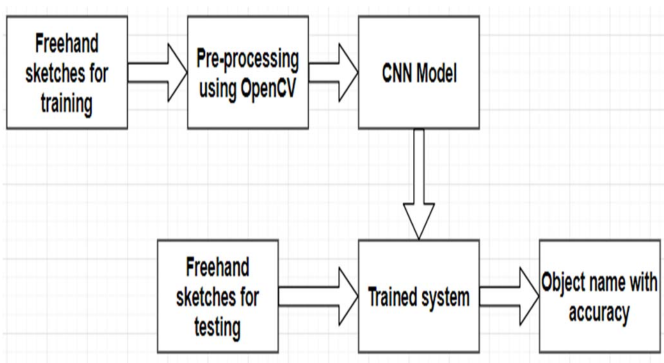

FIG 2 : SYSTEM ARCHITECTURE FOR IDENTIFICATION

B. Grading

The grading of test images solely depends on the output of softmax layer. Grading is done relatively by comparing the sketch with the set of sketches collected for each category. We use SIFT (Scale invariant feature transform) to grade the sketches. SIFT extracts local features using keypoints and descriptors. We find the matches using keypoints and descriptors. We use a function which uses key points of both test image and scoring image and number of matches to grade the test image.

score = max of score for each scoring image over test image

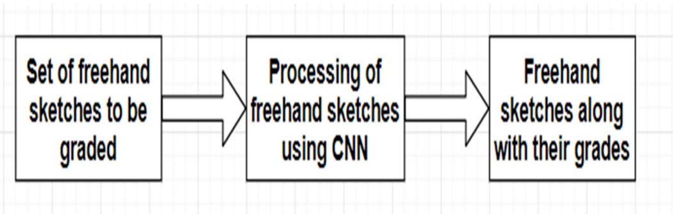

FIG 3 : SYSTEM ARCHITECTURE FOR GRADING

V. Methodology

4.1 Dataset

Publicly available dataset of Eitz et al.[1] is being considered for our experiments. It consists of 20,000 sketches across 250 different classes with 80 sketches per class. The sketches are arranged in separate subfolders by name of the classes. We trained our model with varying number of classes and making observation each time for the accuracy obtained. Keras with tensorflow backend is used to build, train and test our model.

4.2 Preprocessing

Preprocessing is a vital part of any Computer Vision project. In this work, input sketches of size 128 x 128 pixels have been considered. Pre-processing on these sketches involves a few steps :

4.2.1 Uniform aspect ratio

The overwhelming majority of the neural networks acknowledge square pictures, thus we want to ensure that the sketches being considered contain sustained area unit of same size and perspective proportion. Every sketch should be checked for their sizes and trimmed properly.

4.2.2 Scaling

This is an important step where we ensure that input data has similar data distribution. Hence training the network converge faster comparatively. We likewise reduce our sketches from [0, 255] to [0, 1.0] which is the most regular scaling system.

4.2.3 Augmentation

This is again a most typical preprocessing technique that involves augmenting sketches within the dataset that virtually increases the sparse dataset. It usually scales, rotates and does wide variety of changes so that the neural network is bestowed to a large assortment of sorts. This makes the neural system more averse to learn undesirable attributes.

4.3 Lenet

As high resolution sketches are not dealt with, a 7 - level convolutional neural network has been chosen. The original Lenet architecture uses Tanh activation function. But, in our model, ReLu has been chosen over Tanh since it tends to give much better classification accuracy.

Our Lenet model consists of input layer followed by Convolutional (CONV), Rectified Linear Unit (RELU) and max pool (POOL) layers which are followed by another set of CONV, RELU and POOL and then by fully connected (FC) layer, RELU and another FC layer. [6]

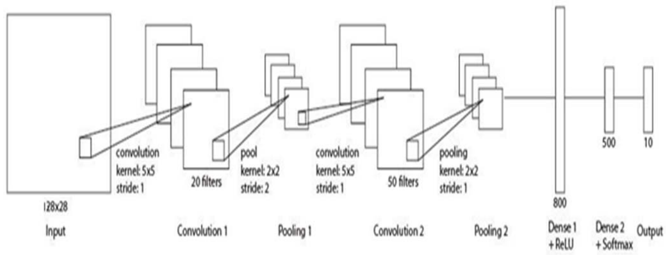

FIG 4 : LENET MODEL

INPUT => CONV => RELU => POOL => CONV => RELU => POOL => FC => RELU => FC

Input size of 128 x 128 is fed into the model

First Convolutional layer will learn 20 convolution filters, where each filter is of size 5 x 5, applied on ReLU activation function followed by 2 x 2 max-pool layer in both the *x* and *y* direction with a stride of 2.

Second Convolutional layer will learn 50 convolution filters, where each filter is of size 5 x 5, followed again by ReLU and 2 x 2 max-pool layer.

Later a fully-connected layer with 500 units with Relu.

Finally a fully-connected layer with number of units equal to number of class is used.

4.4 Training & Testing

Eitz dataset has been used to train our model.

Training and testing ratio is 80:20. This ratio is obtained based on trial and error. Both training and validation accuracy is better for this ratio.

Google Colab platform [7] is used to train the model.

Trained model is serialized to disk and can be used to test on different sketches.

4.5 SIFT

Scale Invariant Feature Transform (SIFT) is used to extract distinctive image features from keypoints. This is used to grade the input sketches.

SIFT uses DoG (Difference of Gaussian) to extract important key-points in an image.

(x, y, α) values implies there is a keypoint (x, y) at α scale.

After the DoG's are found, images are found for features over space and scale.

One image pixel is compared with eight neighbours in same scale, nine neighbours in previous scale and nine neighbours in next scale.

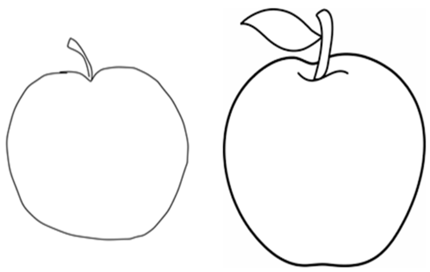

FIG 5 : TESTING AND SCORING SKETCHES FOR APPLE RESPECTIVELY

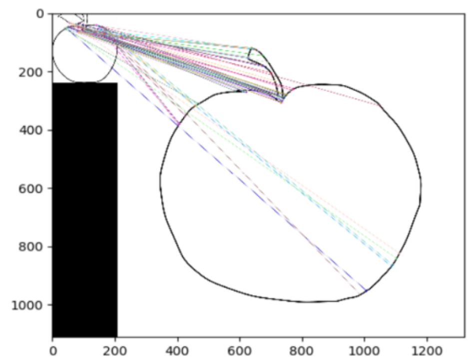

FIG 6 : SIFT MATCHING FOR APPLE

SIFT is both rotation invariant and scale invariant. In all the cases, SIFT extracts same features.

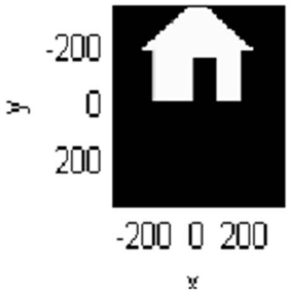

FIG 7 : NORMAL IMAGE

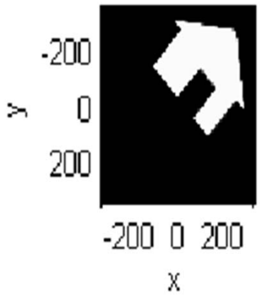

FIG 8 : ROTATED IMAGE

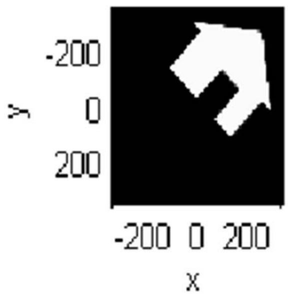

FIG 9 :NORMAL IMAGE

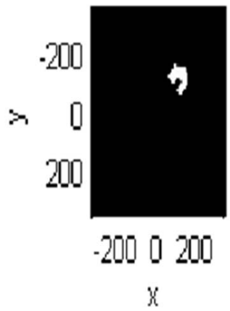

FIG 10 : SCALED IMAGE

4.6 Scoring

Scoring of the sketches is the key idea in our paper.

Scoring is mainly based on the CNN score and SIFT keypoint matching score.

(1) CNN score obtained after applying softmax layer helps to classify the sketch and maps to its respective category.

(2) Based on the category obtained in (1), SIFT keypoint matching is done for all the images in that category with the test sketch.

Score = number of keypoint matches / number of keypoint descriptors

The maximum score obtained among all images for a category is considered.

The Final score is the average of (1) and (2) which is displayed in percentage

VII. RESULTS AND DISCUSSIONS

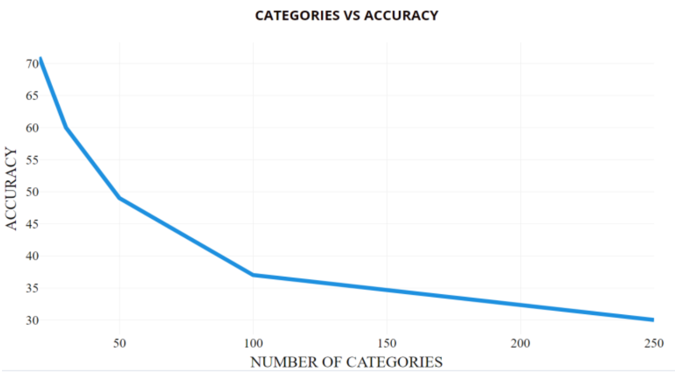

FIG 11 : NUMBER OF CATEGORIES VS ACCURACY

We trained our model with varying number of classes. We obtained 70% accuracy for 15 classes. Accuracy obtained for 20 classes is 64% and for 30 classes, it is 60%. When 50 classes are considered, we obtained an accuracy of 49% and for 250 classes, accuracy is 35%.

Accuracy gradually decreases with increasing number of classes because black and white images (sketches) have a single channel and it is difficult to classify sketches as some categories like ant and bee look almost similar. The sketch images of ant and bee is shown in FIG 12.

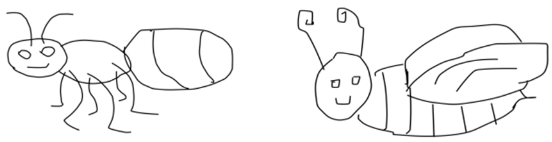

FIG 12 : SKETCH IMAGES OF ANT AND BEE RESPECTIVELY

As the score of the sketch solely depends on the classification part, classification accuracy has to be higher. Considering less categories gives more accuracy. As said earlier, classification accuracy for 15 categories is 70%. Categories are considered at random and average accuracy is been calculated.

The scoring is mainly based on features extracted by SIFT. These key points extracted are used to compute descriptors. Based on these keypoints and generated descriptors, we find a match between the scoring sketches over test sketch . Based the number of matches, number of keypoints of scoring sketch and number of keypoints of testing sketch, we output the score of test image versus scoring image. The maximum of this score is the final score.

a) Classification results:

FIG 13 : CLASSIFICATION OUTPUT FOR ALARM-CLOCK INPUT SKETCH

Our CNN model predicts top 3 classes that an input sketch resembles. However, if a sketch has more than 80% probability of belonging to a particular class, only that particular class is given in the output.

FIG 14 : CLASSIFICATION OUTPUT FOR ANGEL INPUT SKETCH

b) Scoring results:

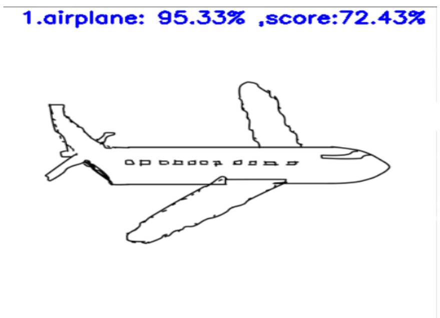

FIG 15 : SCORING OUTPUT FOR AIRPLANE INPUT SKETCH

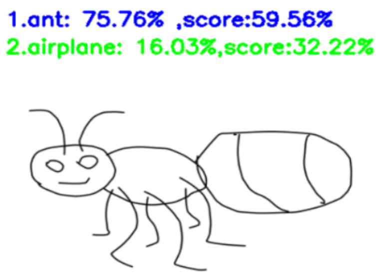

FIG 16 : SCORING OUTPUT FOR ANT INPUT SKETCH

VIII. CONCLUSION

In this paper, identifying and grading of freehand sketches is proposed. Through the proposed technique, the framework can recognize the freehand outline and grade it with likelihood of freehand draw coordinating the portrayals in the informational index. The order and programmed evaluating framework is executed utilizing keras and tensorflow as backend. Actualized result demonstrates this work can accomplish to a greater degree and higher precision with more unambiguous data.

IX. FUTURE DIRECTIONS

One of the enhancements that could be done is to compute the performance of the system using different Convolutional Neural Networks.

Another improvement could be to make use of the temporal stroke information for each of the sketches.

All the ambiguous data could be removed from the dataset and the system could be tested on the remaining dataset. This would show up an increase in accuracy of the system in classifying a sketch but at the same time, dataset would contain lesser sketches than before. Moreover, it requires a lot of manual work. Also, ambiguity is subjective. Hence, it may vary from person to person.

Another enhancement could be to tweak the model to get better and accurate scores for a sketch.

REFERENCES

1. http://cybertron.cg.tu-berlin.de/eitz/projects/classifysketch/
2. M. Eitz, J. Hays, and M. Alexa. How do humans sketch objects, ACM Trans. Graph. (Proc. SIGGRAPH), 31(4):44:1– 44:10, 2012
3. Ravi Kiran Sarvadevabhatla, R a Venkatesh Babu, “Freehand sketch recognition using deep features”,cited: arXiv:1502.00254[cs.CV],2015
4. D. Kingma and J. Ba. Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980, 2014.
5. Chalechale, A., Naghdy, G., and Mertins, A. 2005. Sketch-based image matching using angular partitioning. IEEE Trans. Systems, Man and Cybernetics, Part A 35, 1, 28–41.
6. https://www.pyimagesearch.com/2016/08/01/lenet-convolutional-neural-network-in-python/
7. https://colab.research.google.com/notebooks/welcome.ipynb