

SigRec - Signature Recognizer

A machine learning approach to automate signature verification



About me

- Ujjwal Gupta (Slapbot) :

Hi I am 20 years old, living in Kanpur with more than 2 years of professional experience in a remote workplace, working for a data-centric startup as a lead developer based in Perth, Australia.

I have had previous experience with data science and back-end systems with various projects involving facial detection, recognition and NLP tasks like speech synthesis and intent-recognition.

Design Principles

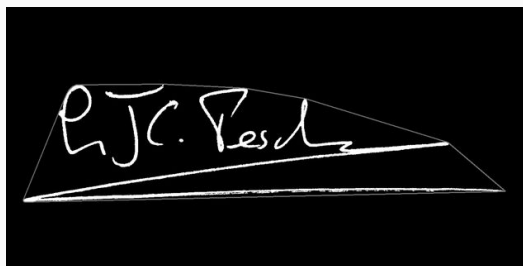
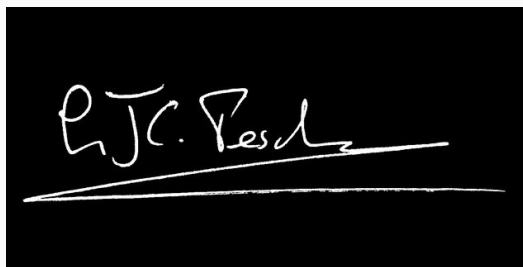
1. Cleaning the image and normalizing dimensions as well as reducing any extra noises to improve consistency using OTSU algorithm.
2. Developing a Deep Learning model to compute high dimensional feature vectors of an image.
3. Creating a simple method to acquire descriptors and keypoints metrics via one of the standard computer vision algorithm. (BRISK/SIFT)
4. Setting up a writer-independent classifier like SVM to classify the minute differences between genuine and forged signatures.

Architecture

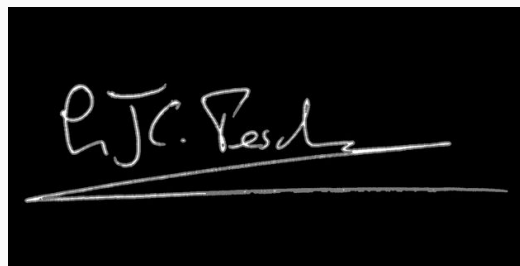
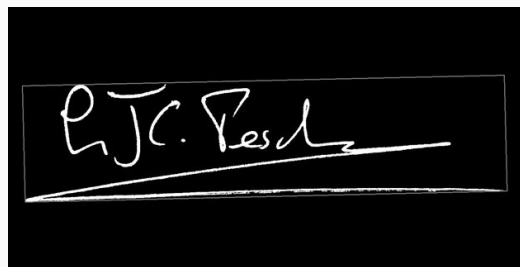
1. The entire platform follows a pattern of microservice such as it is completely decoupled from the main/parent platform.
2. High dimensional 2048 feature vectors are created using SigNet pretrained on CEDAR, GPDS, MYCT, etc Datasets.
3. Small dimensional vectors like contour features such as minimum bounding area of signature, aspect ratio, contour area, convex hull area are computed.
4. Feature Extractor like BRISK/SIFT is used to find keypoints (edges) and descriptors of a signature as histogram of features.

Contour Approximation

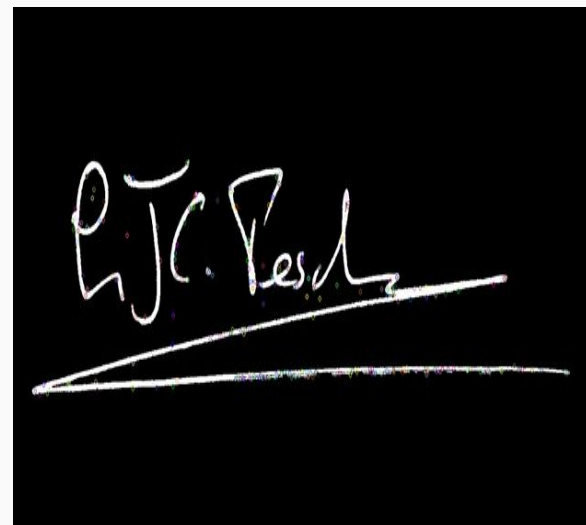
Min Bounding Box



Contour Area
Approximation

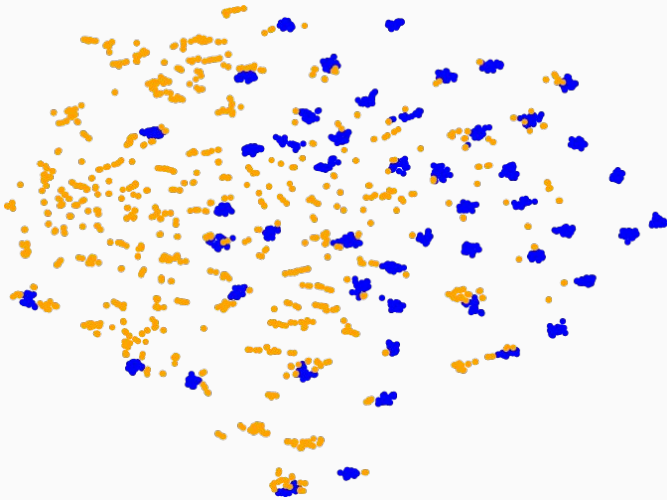


Keypoints Retrieval

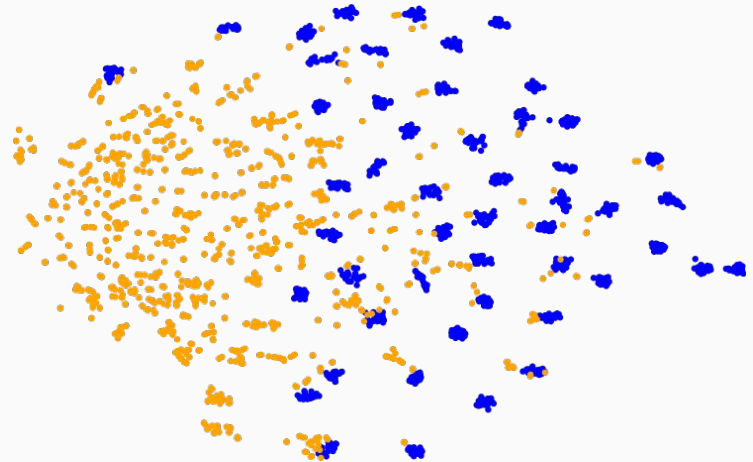


SigNet High Dimensional Feature Extraction

Using CNN to compute high feature dimensional data which can further be compacted by some dimensionality reduction like PCA (Principal Component Analysis) and Standard Scaling. Blue point -> Genuine Signature, Yellow Point -> Forged Signature



Model trained with genuine signatures.



Model trained with genuine signatures and skilled forgeries

Keypoints and Descriptors Comparison

Quick and Efficient Feature Matching Using FLANN (Fast Approximate Nearest Neighbour Search Library)

genuine vs genuine comparison

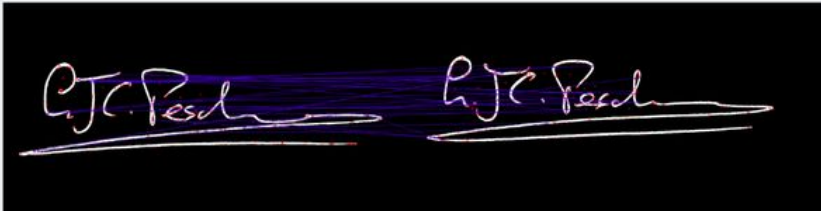
Signature Verification Demo

Verifying Signature's Authenticity Using Smart Feature Extraction And Linear SVC Modelling.

Confidence: 0.83285530244459

Prediction: Genuine

Feature Analysis:



genuine vs forged comparison

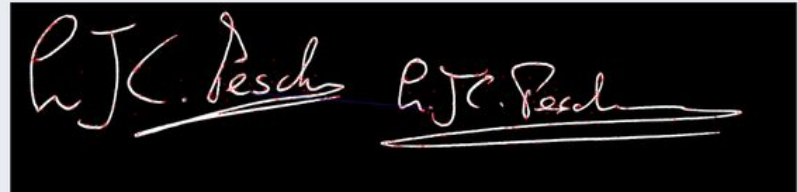
Signature Verification Demo

Verifying Signature's Authenticity Using Smart Feature Extraction And Linear SVC Modelling.

Confidence: 0.67817301731949

Prediction: Forged

Feature Analysis:



Key Challenges Faced

1. Cropping and Centering Signature to ensure it fits the canvas properly before sending to learning/inferencing task.
2. Augmenting Data to ensure the change in angle of signature doesn't skew the model too much.
3. Learning a SVC classifier $O(n^3)$ polynomial time complexity with limited resources by tuning hyper-parameters and trying alternate approaches such as Logistic Regression or a SGD Classifier.

Scaling And Ease Of Use

1. There is already a running web-app to test out the model at <https://sigrec.herokuapp.com/> which is setup with a custom dataset to demonstrate the POC in a RESTFUL API format.
2. CNN Model only being used as a feature extractor (transfer learning) rather than relying on training the model and sinking down in another problem.
3. Images can be saved in S3 buckets while the progressive web-app is running on any of the compute instances such as EC2 using Nginx/Apache web server.
4. Decoupling from main platform allows more efficient work sprints with Continuous Integration (testing) and Continuous Delivering without messing with the main platform.

Accuracy Achieved

1. An accuracy of about 97% reached in given dataset for the challenge with limited resources and training images.
2. Model can be tuned specifically for better precision or recall scores allowing better prediction of True Positives and hence saving time of the bank officials.
3. Overfitting is another concern which can be addressed with more diverse and large datasets.

```
Set 26/30
Status: Original, Predicted: Original, isCorrectlyPredicted: True
Status: Original, Predicted: Original, isCorrectlyPredicted: True
Status: Forgery, Predicted: Forgery, isCorrectlyPredicted: True
Status: Forgery, Predicted: Forgery, isCorrectlyPredicted: True
Status: Forgery, Predicted: Forgery, isCorrectlyPredicted: True
Status: Forgery, Predicted: Forgery, isCorrectlyPredicted: True
Status: Forgery, Predicted: Forgery, isCorrectlyPredicted: True
Set 27/30
Status: Original, Predicted: Original, isCorrectlyPredicted: True
Status: Original, Predicted: Original, isCorrectlyPredicted: True
Status: Forgery, Predicted: Forgery, isCorrectlyPredicted: True
Status: Forgery, Predicted: Forgery, isCorrectlyPredicted: True
Status: Forgery, Predicted: Forgery, isCorrectlyPredicted: True
Status: Forgery, Predicted: Forgery, isCorrectlyPredicted: True
Status: Forgery, Predicted: Original, isCorrectlyPredicted: False
Set 28/30
Status: Original, Predicted: Original, isCorrectlyPredicted: True
Status: Original, Predicted: Original, isCorrectlyPredicted: True
Status: Forgery, Predicted: Forgery, isCorrectlyPredicted: True
Status: Forgery, Predicted: Forgery, isCorrectlyPredicted: True
Status: Forgery, Predicted: Forgery, isCorrectlyPredicted: True
Status: Forgery, Predicted: Forgery, isCorrectlyPredicted: True
Status: Forgery, Predicted: Forgery, isCorrectlyPredicted: True
Set 29/30
Status: Original, Predicted: Original, isCorrectlyPredicted: True
Status: Original, Predicted: Original, isCorrectlyPredicted: True
Status: Forgery, Predicted: Forgery, isCorrectlyPredicted: True
Status: Forgery, Predicted: Forgery, isCorrectlyPredicted: True
Status: Forgery, Predicted: Forgery, isCorrectlyPredicted: True
Status: Forgery, Predicted: Forgery, isCorrectlyPredicted: True
Status: Forgery, Predicted: Forgery, isCorrectlyPredicted: True
Accuracy of Genuine Predictions: 1.0
Accuracy of Forged Predictions: 0.95
Precision: 0.9
Recall: 1.0
F1 score: 0.95
Genuine+Forge Relative Accuracy: 0.97
All Accuracy: 0.97
slapbot@slapbot-GL553VD:~/my_side_projects/sigrec_final$
```

Some links about me?

->

<https://github.com/slapbot>

<http://slapbot.github.io>

ugupta41@gmail.com

Thank You