Signature detection & Recognition

Praveen Vijayan

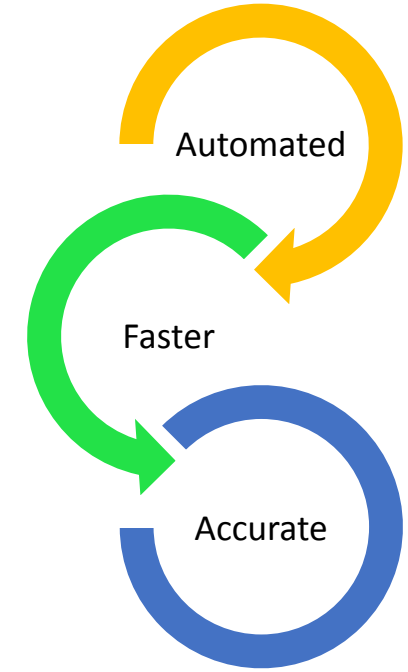
Official Email : praveen.vijayan@tcs.com

Personal Email : praveenkottayi@gmail.com

WHY part of Signature detection & Recognition

Where Scenarios

- Cheque clearance
- Legal document processing
- Invoice processing
- Bill processing



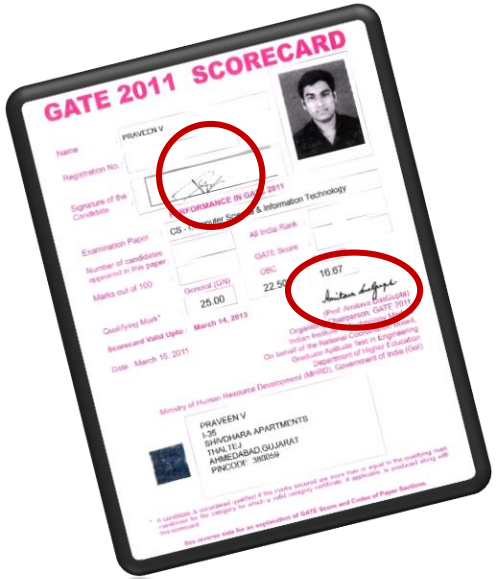

Anywhere in document


Multiple interests


Tag automatically


Any doc size / Font / Color / language


Any signature color / font

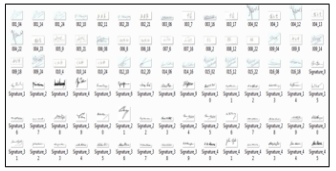


HOW part of Signature detection & Recognition

Creating : Signature classifier

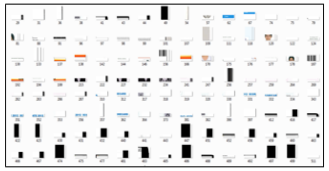
0

Get **Signature dataset** (PoC + sigcomp2011 dataset).



class

Create **Non – signature samples**(conditioned random sampling from the documents and filter those images which doesn't have traces of signature).



class

Transfer learning : Used **deep learning VGG-16** fully collected layer 2 to extract the features from images.

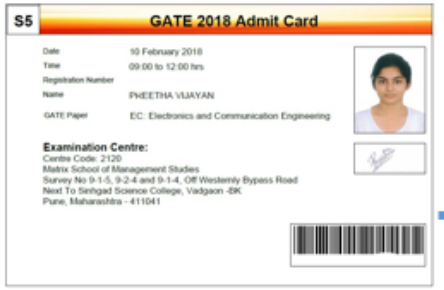
May use principal components to reduce dimension.

Create an optimized classifier using SVM.

Save the classifier using Joblib to use in the main pipeline.

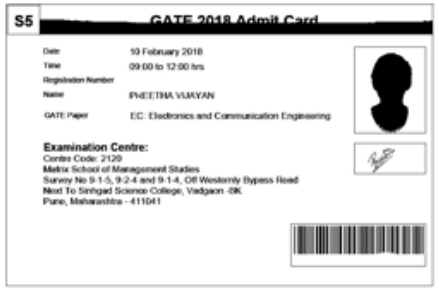
Original image

1



Thresholded image

2



Connected component/ blobs in the image

3

Blurring & Bounding box over big Connected components

4

Extracted probable Signatures

5

Signature Classifier results

6

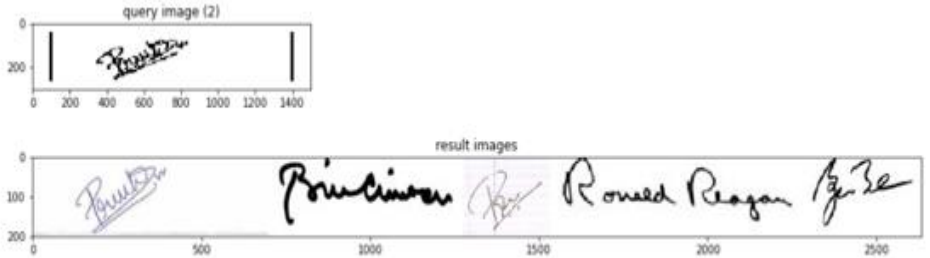
1

0

0

Most similar signatures

7



One more example

Original image

1

Thresholded image

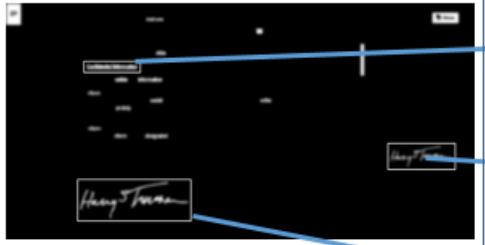
2

Connected component/ blobs in the image

3

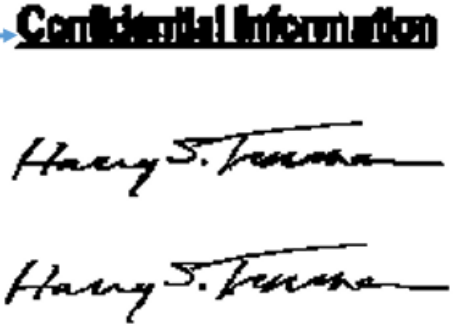
Blurring & Bounding box over big Connected components

4



Extracted probable Signatures

5



Signature Classifier results

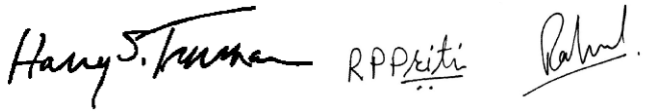
6

0

1

1

Similar Signatures based on similarity:
[**0.0**, **50.23**, **50.58**]



Similar Signatures based on similarity:
[**32.57**, **50.23**, **55.26**]

Pros of this method

Instead of going for a cascading scanning of probable signatures with a Classifier, preferred **image processing method**, which is way faster to find ROIs.

Created **dataset for non-signature examples** by conditional random sampling from the documents and removing those containing traces of signature.

Hence further reducing the number of probable signatures for comparison with Signature database.

Used **Deep learning(VGG-16)** to capture better features that can distinguish signatures.

At present system is designed in such a way that **False positive is better than False negative**. Because missing a valid signature may end up not getting tagged to the right person. Even if its false negative, the last stage of signature comparison would filter the wrong ones.

Further Improvements

- Need more data for Signature and non-signature set
- Padding the probable signature part instead of direct scaling
- Extracting better features from image
- Dynamic thresholding, blurring

Thank you