TATA CONSULTANCY SERVICES
Experience certainty.

TATA

Handwritten Signature Extraction Using Connected Pixels and CNN

Suman Das - 324917
BFSI CDO ML/AI LAB
October 2020

Abstract

This paper explains how a novel way was adopted to read documents using Computer Vision - Deep Learning technique and using which any signature present in the document could be extracted.

It describes about a technique to identify areas in your document which are having connected texts, which normally is present in signatures, and how these regions can be fed into a classifier to detect if the region contains a signature or not. The generic steps taken to locate the region and classify the same, ensures that the solution can be applied to any kind of document.

The paper emphasizes on how the solution can be applied to all kinds of documents but it also describes on some of the limitation the approach has.

This signature extraction process could be used in a number of financial and legal scenarios where signatures are considered to be an important biometric.

TATA CONSULTANCY SERVICES
Experience certainty.

Contents

1. Introduction

Handwritten signatures are a primary source of biometric which is widely used in a lot of legal documents and verifying these signatures become a critical aspect of any legal transactions. Signature verification, when done electronically without any manual effect assumes that these signatures are segmented and clearly extracted. It therefore makes sense to have a solution which can look into any document and extract the underlying signature from it. This technique can be applied to any document for example a Driver's License, Passport, Cheques etc.

Some of the widely accepted techniques to detect signatures from documents are Hyperspectral Unmixing[1] which doesn't use Deep Learning, Deep CNN[2] which uses complex Deep Learning algorithms like YOLO or Faster-RCNN and some computer vision techniques like connected pixels detection which keeps the process simple and tries to encash om the fact that signature are characters in a document which has some unique features.

Some of the above techniques score in having a good accuracy while some others score based on their easy implementation. However, this paper builds on the ideas provided in the papers and topics mentioned above and try to combine the simplicity of connected pixel algorithm and the power of Deep Learning and come up with a novel way to apply the solution to all kinds of documents.

The approach, described in the subsequent sections, used to extract the region of interest (ROI), in this case the signature, is a novel technique. For every image we extract all such regions where the image pixels are connected to each other and then feed those connected pixels to a CNN Signature classifier. The classifier ignores all such regions and only spits out the signature if it finds any.

The paper is organized into four sections. Section 2 discusses the approach in detail. It is further subdivided into two sections where we touch upon the two components that are the heart of the solution. Section 3 is devoted to the discussion on how the extractor performs and what are the limitations. Section 4 concludes the paper describing why this solution is unique and what else can be tried to improve the solution.

2. Methodology

The signature extraction involves two different components; the first step is the connected pixels image extraction which extracts all the connected texts and thereby extracting all signature and non-signature images in the document. The second step involves the CNN signature classification which ingests all these connected texts and spits out the signature from the text.

The below flow diagram explains how the solution performs the signature extraction.

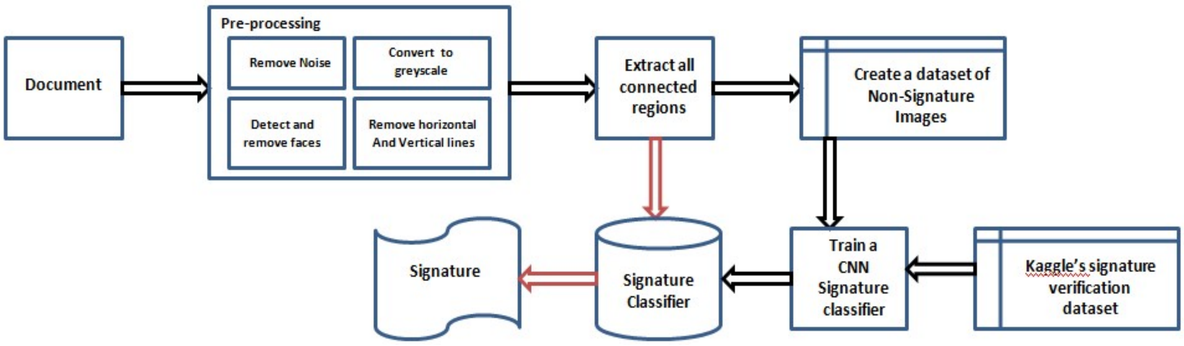

Fig 1. Flow diagram for signature extraction methodology. The black arrows show the connected pixels image extraction and the amber arrows show the classification steps.

2.1. Extracting Connected Pixels

Every document has a lot of text, most of them being machine printed. The characters in these texts may or may not be touching to each other. Also words do have a gap between them. Using some standard image processing tools the pixels in the image are morphed so that the characters within a word touch each other but still maintain gaps between words.

The image below shows a document which is converted to grayscale and connected pixels was applied to extract individual words from the image.

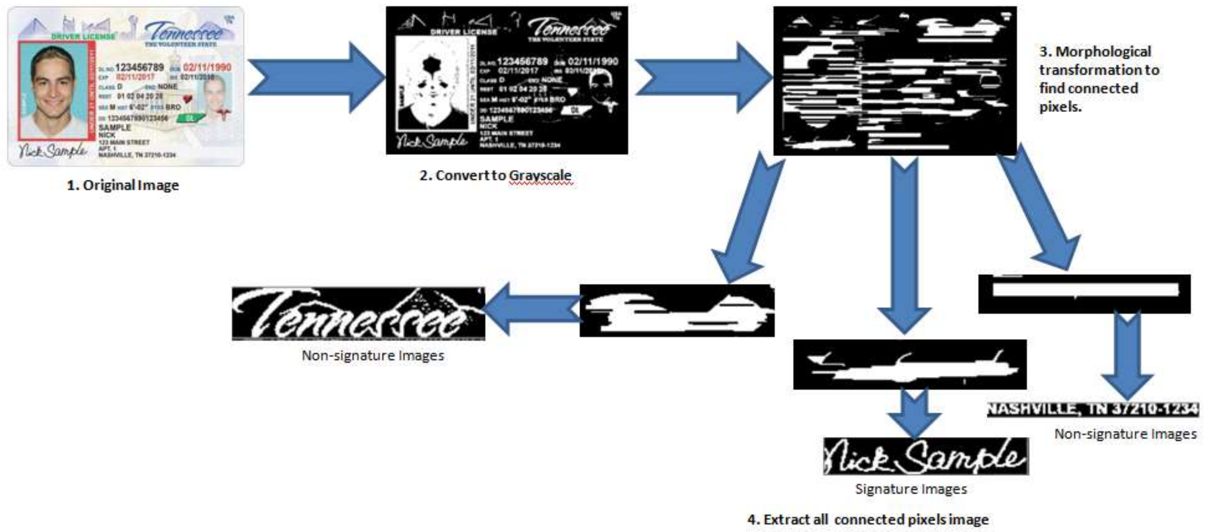

Fig 2. Connected pixels algorithm showing extraction of both signature and non-signature images

2.2. CNN Signature Classification

Another piece of the solution involves training a signature classifier. This classifier is supposed to be fed with all the connected ROIs received from the above algorithm and it is supposed to spit out only the signatures it detects.

To train such a classifier, a lot of labeled signatures are required. This can be found in some open source dataset like the Kaggle signature verification dataset[3]. This dataset has around 2000 odd images of different signatures broken into two classes, ‘original’ and ‘forged’. There is very minute difference between the two sets. Since we aren’t interested in signature verification we can ignore these differences and just use both sets to train our classifier. This way we obtain the ‘signature’ class.

To generate the ‘non-signature’ class the connected pixels algorithm comes to rescue. The algorithm can be applied on roughly around 100 document images collected online and all non-signature regions can be saved and labeled as ‘non-signatures’ class.

Finally using images from both the classes and by applying some amount of data augmentation to generate fabricated signatures a huge corpus of images can be created to train a Convolution Neural Network.

2.2.1. CNN Architecture

The images are resized to a dimension of 180*180 pixels and feed to the CNN architecture. The model has 3 layers of neurons of 32, 64 and 128 units each. Each layer is activated by a tanh function and followed by a max-pooling layer.

Finally the 3rd layer is then followed by a Dense layer of 128 units and activated using tanh function. The output is then fed to a sigmoid function which classifies the image either as a signature or non-signature. The model is trained with a batch size of 8, learning rate of 0.001 and for 100 epochs.

3. Results and Evaluation

The combination of the Kaggle signature dataset and the generated non-signature dataset is split in an 80-20 ratio into a train and validation sets while training. The classifier achieves an accuracy of 92% on the validation dataset.
Some of the results of the classification are shown below;

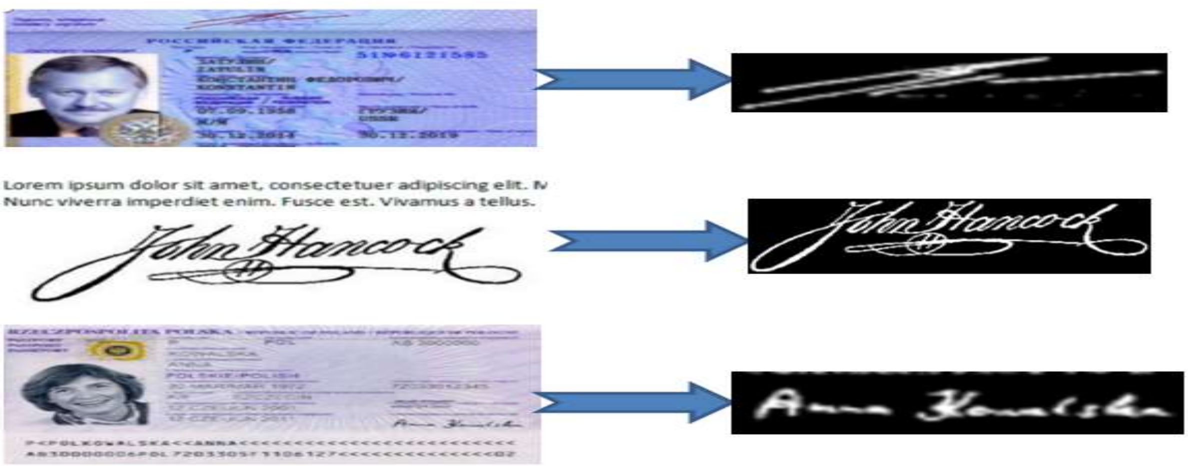

Fig 3. Performance of the model in some unknown images

It looks like the model does very well when the signature is well isolated from the other texts. However there are some limitations to the performance of the model which will be discussed in the next section.

3.1. Limitations

The solution though generic to all documents does have certain limitations. In certain scenarios the solution fails to extract the exact region of interest from the document. These limitations can be broadly classified to have two root causes.

3.1.1. Connected Pixels Algorithm extracts incorrect region of interest

It was found that when the signature in the document has a background image or another text very near it, the connected pixel algorithm connects the signature with these unwanted sections and end up having an extract which has other items along with the signature.

Since the solution is supposed to be very generic and applicable to all kinds of documents we cannot do a lot of parameter tuning to morph the documents properly. No matter how many different values of morphological parameters are tried it isn't possible to find a 'one-size-fits-all' parameter which can help the connected pixels algorithm to find the proper region of interest in any kind of image.

Fig 3. shows some of the scenarios where the signatures extracted aren't correct because the connected pixels algorithm wasn't able to find the proper region of interest.

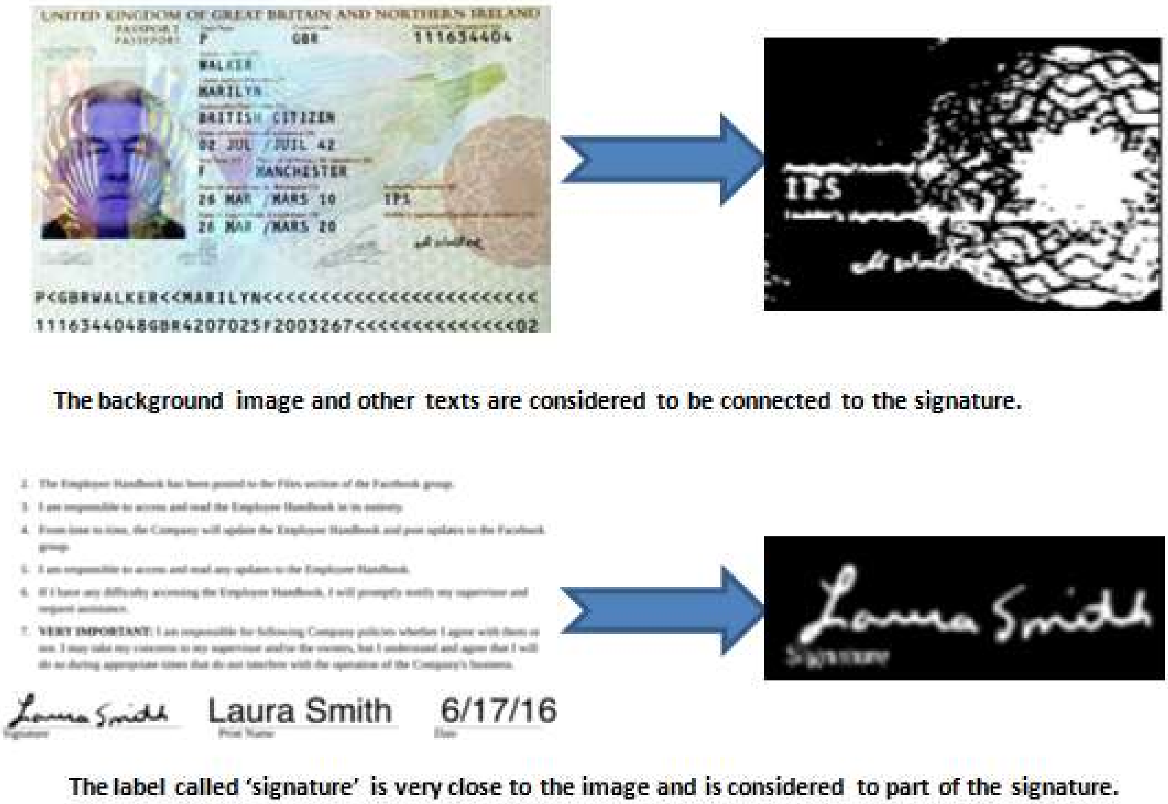

Fig 4. Limitation of connected pixels algorithm

3.1.2. Classifier returns a False-Positive

In case the document has other hand written texts, similar to a signature, the classifier sometimes considers that to be the ROI. There can be scenarios where there is a separate signature of a governing authority or there are some machine printed texts in the document which has a font similar to hand written texts. These regions when predicted by the classifier as a signature are called as a false-positive. The actual region of interest, the signature, is the true-positive.

An approach to filter out these false-positive records will be to compare their probability score with that of the true-positive. The probability score tells us how confident our classifier is when it labels a particular region of interest as a signature.

It is found that most of the time the true-positives have a higher probability score predicted by the classifier. However, our classifier has a 92% accuracy and there is a possibility that the reverse holds true too sometimes. This scenario of predicting an incorrect region of interest as a signature can be overcome if we consider the top two or three scores of the classifier and introduce a manual intervention to select the correct signature. But nevertheless this is still a limitation of the solution.

Fig 4. shows some of the scenarios where the classifier predicts 2 different regions of interest, one of them being a false-positive.

Fig 5. Classifier predicting false-positives with a higher probability than the true-positives

4. Conclusion

The signature extraction approach discussed in this paper tries to leverage open source technologies like python and using some of its popular packages like scikit-image, opencv and implementing the solution with Deep Learning framework.

There are many possible ways of achieving this and each one of these techniques has their own merits. Implementation of any of these solutions will depend on how familiar one is with the underlying technique.

The approach discussed in this paper can be very highly refined to improve a very high accuracy if the type of the document can be fixed. A certain type of document will have certain aesthetics which will help the connected pixels algorithm to extract the correct region of interest and hence mitigate one of the limitations we talked about.

Also the classifier can be retrained with more images and a different architecture which should improve the accuracy further. For example a transfer learning approach can be implemented to see if that improves the accuracy.

Finally, one can also try doing the preprocessing of the documents in a different way that should have a downstream effect.

5. References

1. Kashif Iqbal, and Khurram Khurshid: Automatic Signature Extraction from Document Images using Hyperspectral Unmixing.
2. Nabin Sharma; Ranju Mandal; Rabi Sharma; Umapada Pal; Michael Blumenstein: Signature and Logo Detection using Deep CNN for Document Image Retrieval
3. https://www.kaggle.com/robinreni/signature-verification-dataset

THANK YOU