

Hindi Signature Verification using Data generation with Generative Adversarial Networks

by

Jatin Kasera

Roll. No.: 2018IMT-042



विश्वजीवनामृतं ज्ञानम्

**ABV-INDIAN INSTITUTE OF INFORMATION
TECHNOLOGY AND MANAGEMENT GWALIOR (M.P.),
INDIA**

- Handwritten Signatures are one of the most prominent bio-metrics to authenticate documents in various institutions like banks, etc.
- Because of the huge increase in the quantity of documents signed, the frequency of frauds in signatures has increased significantly.
- Signature verification methods are split in 2 classes:
- **Online methods**, that uses dynamic processes such as digitizer signature to take account of speed and pressure utilising smart algorithms.

- **Offline** approach, signature on paper is signed and scanned to transform it into a digital signature using the optical scanner.
- In this thesis we concentrate on offline method because of their practical applications, no need for extra equipments.

Types of forgeries

- There are 4 types of sign forgeries as follows:
- **Simulation/Simple Forgery:** A type of forgery in which the forger has a sample of the falsified signature. Unskilled and skilled forgeries are classified based on the forger's experience.
- **Blind Forgery :** This is when the fraudster has no notion what the faked signature would be in appearance. This is the most difficult sort of forgery to identify since it does not have the appearance of a genuine signature.

Types of forgeries

- **Tracing** : It is done by arranging the reference document and the concerned document up to the light and tracing the lines of the reference sign over onto questioned document with a pen.
- **Optical Transmission** : It is a method of transferring an authentic signature onto a document using a xerox, scanner, replica machine, or photography. An examiner cannot definitively recognise a signature as authentic in this form of forgery unless having the Authentic to compare it against.

- Over the years, there has been a lot of work done in the domain of offline verification of English signatures, but signatures are signed in other languages too.
- Due to lack of development in the field of other languages like Hindi, etc., there has to be solution for Hindi Signature fraud detection too.
- Available methods lacks data set to train on , therefore we propose Hindi Signature Verification using Data generation with Generative Adversarial Networks.

- The examination of human signatures focuses in particular on improving the interaction between people and technology.
- Signature are often misunderstood as just group of words and letters but they are being treated as combination of pixels for verification.
- Some of the methods used previously were clustering-related approach to this problem. This process involves use of filtration methods to eliminate noise, image is further pre-processed, the ROI is compared with real signatures.

- The main concern in offline signature verification can be divided into 2 types:
 - ① **Identification**: it entails detecting whether a signer is already a user among the possible Database.
 - ② **Verification**: this is done after identification, if a signer is a user, then does its signatures matches with the one in Database.
- Despite significant progress in the area of offline verification of signature, but most of them were focused only on English Signatures.
- Therefore, a sustainable method has to be developed for offline Hindi signature validations.

Problem Statement

- Financial institutions such as banks etc. and companies recognise the main way in which transactions are authenticated. Cheques are signed, contracts and documents are permitted, transactions of credit cards are validated and verified using signatures. With availability and numbers of signed documents enormously expanded, this in turn causes the spike cases related to signature forgery.
- This project mainly focuses on using offline signature validation methods on Hindi Signature data sets. The unavailability of large data set of Hindi signature data to train model on, can be solved by studying Generative Adversarial Networks and then extracting their features and further classifying them as genuine and fake.

- In Writer Independent pattern, we propose method for learning features for the offline Hindi sign verification and identification. We present a rigorous method under which all the design judgements are based using from a set of predictors com-posed of a different set of users.
- To gain better insights of data, GAN is used to increase the size of the available data set for Offline Hindi Signature Verification. We offer a clear assessment of the learnt representations, demonstrating that authentic signs and competent replicas of Hindi Signs can be distinguished in separate sections of the feature set

- **Data Collection:** Handmade signs of Hindi script are captured, in order to construct a base of knowledge for each individual. Initially some authentic and fake signs were collected.

शिल्पा शर्मा
रौशनी चौपड़ा
महेश वर्मा
पुष्कर वर्मा

Figure: Sample Signatures

- **Pre-Processing:** The Signature used by the model is to be in digital images format. Thus signatures are collected by signing on fixed size of paper then scanning it using any scanning device. This may contain some irregularity which has to be removed by filtration and discarding images which are not useful.
- **Noise Abatement:** Noise is the outcome of errors induced by many sorts of acquisition processes, resulting in image pixels that do not correspond to genuine values. The picture obtained from a lens includes film grains, which are a bit noisy, To improve the noise separation process, a known type of noise is injected to the gray-scale image in a very small amount. This increases the threshold of the noise and hence can be filtered using a filter.

- **Binarization:** First the digitized image is transformed into binary image, this further converted into two tone images, Which can be used for further process. Below figures depicts the outcome of this operation:

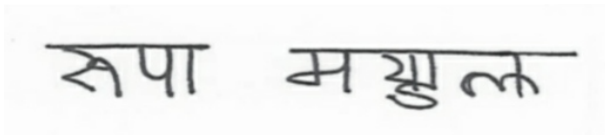


Figure: Grey-Scale Image

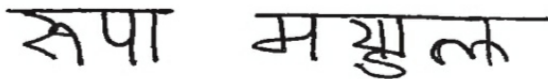


Figure: Converted Binary Image

Generative Adversarial Networks

- GANs is an algorithmic design composed of 2 neural networks, which are mutually competitive (hence "adversarial") in the generation of replicated sample data that can be passed on for real data.
- The GAN is an unsupervised form of machine learning, whereby the patterns or the regularities of the data input are discovered naturally.
- The model can be used to produce or generate new images (here images but can be other data) which could be taken from the original input image data set.
- It has two parts Generator and Discriminator.

Generative Adversarial Networks

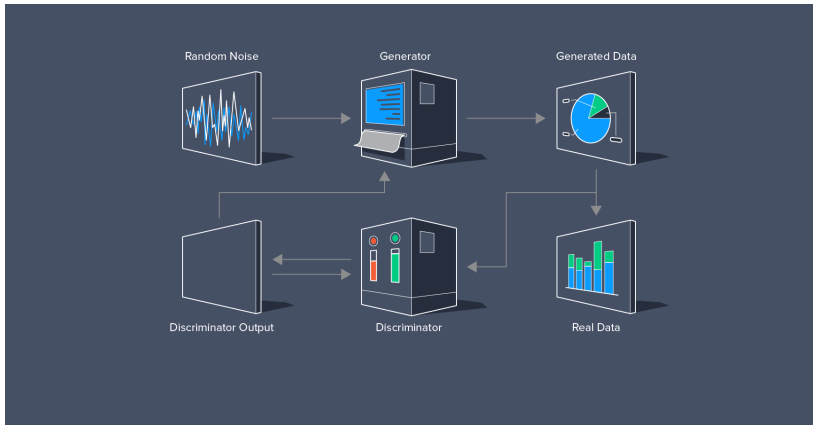


Figure: Basic block diagram of GAN

- The generator is analogous to the heart of the architecture.
- It is a component of model that is used to produce new similar data like images ,and to obtain really high performance at the end of the training process.
- The purpose of the generator is to be able to generate synthetic instances from a given input.
- The input is combined with distinct sets of random numbers, known as a noise vector, to ensure it can generate new realistic images every time.

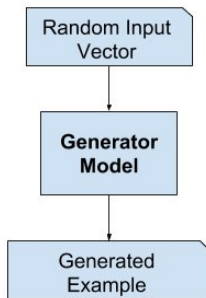


Figure: Generator component of GAN

Discriminator

- The discriminator is a sort of classifier whose goal is to distinguish between actual and created data that is real or fake.
- Classifiers aren't just for selecting image data, it can be video, audio.
- As a result, the discriminator is a form of classifier that learns to estimate the likelihood of a given example being real or fake using input data such as images' pixels values
- The discriminator's output probabilities assist the generator in learning to provide better samples over time

Discriminator

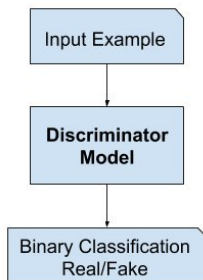


Figure: Discriminator component of GAN

Loss Function of GAN

The loss function of a generative adversarial network is:

$$E_x[\log(D(x))] + E_z[\log(1 - D(G(z)))] \quad (1)$$

where

- $D(x)$ is discriminator's output value for a given x .
- x is a real sample .
- z is a an instance of random noise.
- E_x means the probable value over all x .

Loss Function of GAN

- $D(G(z))$ is the probability given by discriminator's if a forged sample is genuine.
- $G(z)$ is the output of the generator for a given input z
- E_z means the expected value over all z .

The main objective of generator to reduce the value of the loss function, which equates producing samples that look like real ones. On the other hand, the discriminator tries to increase the loss function

Extraction of Features

The images from the GAN model is further processed to detect the authentic and replica of signs. This task is mainly accompanied by using the extraction of features. But before extracting features, the devanagiri script's properties must be studied carefully. In Devanagiri, a Hindi script, in which the specific direction of writing is from the left to right, in Hindi script, there is no such thing as upper characters and lower case characters.

vowels	
अ	आ
इ	ई
उ	ऊ
ऋ	
ए	ऐ
ओ	औ
अं	अः
Consonant	
क	ख
ग	घ
ङ	
च	छ
ज	झ
ञ	
ट	ठ
ड	ढ
ण	
त	थ
द	ध
न	
प	फ
ब	भ
म	
य	र
ल	व
श	ष
स	ह
क्ष	त्र
ज्ञ	

Figure: Vowels and Consonants of Devanagiri(Hindi) script

The extraction of characteristics from an image plays a vital role in any recognition task. For this purpose Gradient method of extracting characteristics is deployed. The Gradient method of extracting characteristics can be summarised as follows:

The handwritten element's grey-scale local-orientation histogram is utilised to obtain 576-d characteristics.

The steps followed to generate a 576-d based on gradient feature vector.

- **Step 1:** The images from GAN is filtered 4-times using mean filters of 2×2 .

Continued...

- **Step 2:** The output of the Step 1 i.e. grey scaled sign images is further gone through normalization till the mean of the grey images scale equals to 0 and max value is 1.
- **Step 3:** The images after going through normalization is further divided into blocks of size of 17 into 7. The size of block is determined empirically, balancing the trade-off between correctness and complication. The grey scaled sign image is turned into a two-tone sign image. This will result in removal of any extraneous additional details from the sign images.
- **Step 4:** The gradient's arc tangent (path of gradients) is gone through fragmented into 32 orientations, and the gradient's intensity is compounded with each of the quantified orientations.
- **Step 5:** The 17 into 7 is converted to block size of 9 into 4 across 16 orientation, thus $16 \times 9 \times 4 = 576$.

Result and Conclusion I

Results The extracted features were passed through SVM classifiers , first the sample training and testing was done on small dataset containing authentic and replica of signatures collected and passed through GAN.

Table: No. of Authentic and Duplicated Signs used for Sample Training

	Authentic Signatures	Duplicated Signatures
Training	15	21
Testing	12	11

But this is size of the basic training and testing. But the actual size on which the model trained were **860 real signs and 1040 fake signs** to begin from a public dataset. The signatures were passed through GAN model first it will generate fake images that will be used in further process.

Result and Conclusion II

Hindi Signatures	
<i>Genuine Signatures</i>	<i>Forged Signatures</i>
सुरेश वाजपेयी	सुरेश वाजपेयी
प्रदीप वर्मा	प्रदीप वर्मा
विकास मोदी	विकास मोदी
विवेक मटेश्वरि	विवेक मटेश्वरि

Figure: Authentic and Fake Signs after training on GAN

Result and Conclusion III

To check the performance of this models, 2 types of errors are exists first one is False Rejection, it is also known as Type-1 error. This is measured by amount of authentic signatures classified by the model as fake signatures. The rate at which the model does this work is known as False Rejection Rate, FRR.

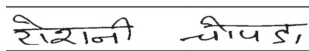


Figure: After gradient method, authentic Sign confused as fake sign

Second one is False Accepted, it is also known as Type-2 error. This is measured by amount of fake signatures classified by the model as authentic signatures. The rate at which the model does this work is known as False Acceptance Rate, FAR.

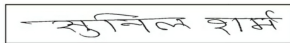









Figure: After gradient method, fake Sign confused as authentic sign

Conclusion The Hindi Signature verification system is useful for many practical areas like finance, institutional etc. So the offline verification has to be done at greater extent. We have tested gradient extraction method. But other methods etc can be used for extraction, but they will have more errors. The Dataset size issue can be resolved by using GAN. But in future, a large public dataset can be created for this purpose and modify the overall performance of the system.

-  S. J. Gideon, A. Kandulna, A. A. Kujur, A. Diana, and K. Raimond, “Handwritten signature forgery detection using convolutional neural networks,” *Procedia Computer Science*, vol. 143, pp. 978–987, 2018, 8th International Conference on Advances in Computing Communications (ICACC-2018)
-  S. Madabusi, V. Srinivas, S. Bhaskaran, and M. Balasubramanian, “On-line and off-line signature verification using relative slope algorithm,” 04 2020, pp. 11 –15.
-  J. Preez, B. Herbst, and J. Coetzer, “Offline signature verification using the discrete radon transform and a hidden markov model,” *EURASIP Journal on Advances in Signal Processing*, vol. 2004, 04 2019.

-  M. Ferrer, J. Alonso, and C. Travieso, “Offline geometric parameters for auto-matic signature verification using fixed-point arithmetic,”IEEE transactions on pattern analysis and machine intelligence, vol. 27, pp. 993–7, 07 2005.
-  I. Pottier and G. Burel, “Identification and authentication of handwritten signatures with a connectionist approach,” vol. 5, 07 1994, pp. 2948 – 2951 vol.5
-  S. Pal, M. Blumenstein, and U. Pal, “Automatic off-line signature verification systems: A review,”International Journal of Computer Applications, vol. 14, pp.Pages 20–27, 01 2011
-  S. Pal, U. Pal, and M. Blumenstein, “Hindi and english off-line signature identi-fication and verification,” vol. 174, 01 2012.



Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, “Generative adversarial nets,” in Advances in Neural Information Processing Systems, Z. Ghahramani, M. Welling, C. Cortes, N. Lawrence, and K. Q. Weinberger, Eds., vol. 27. Curran Associates, Inc., 2014.