Multi-task learning using GNet features and SVM classifier for signature identification

Anamika Jain | Satish Kumar Singh | Krishna Pratap Singh

Department of Information Technology, Indian Institute of Information Technology, Allahabad, Uttar Pradesh, India

Correspondence
Anamika Jain, Department of Information Technology, Indian Institute of Information Technology, Allahabad, Uttar Pradesh, India.
Email: anamika06jain@gmail.com

Abstract
Signature biometrics is a widely accepted and used modality to verify the identity of an individual in many legal and financial organisations. A writer and language-independent signature identification method that can distinguish between the genuine and forged sample irrespective of the language of the signature has been proposed. To extract the distinguishing features, a pre-trained model GoogLeNet, which is fine-tuned with the largest signature dataset present till date (GPDS Synthetic), has been used. The proposed method is tested over the BHSig260 (contains images from two regional languages, Bengali and Hindi) dataset. With the help of the above fine-tuned model, knowledge is transferred to the publicly available datasets – BHSig260 and MCYT-75. The features extracted using the fine-tuned model has been fed to the support vector machine (SVM) classifiers. With the proposed method, 96.5% and 95.7% accuracy on Bengali and Hindi datasets, and 93% on MCYT-75 with skilled forged samples have been achieved respectively.

1 | INTRODUCTION

Biometric-based authentication has been an emerging area for research in the last few decades. Biometrics was started initially with the body measurements, later with time and necessity, it involved many biometric properties related to the human body to provide authentication. Based on the properties of the human body, biometrics has been divided into two categories, that is, physiological and behavioural. Physiological biometrics contain face, iris, fingerprint etc. and behavioural biometrics contain gait, speech, signature, keyboard dynamics etc. [1,2]. Among all the biometric traits, signature is the most commonly used and cheapest form of biometrics with an added advantage of acquisition process familiarity among people. Verification of the signature has been divided into two categories based on the acquisition process, that is, online and offline. In the online process, signatures are collected on the electronic pad, which also captures additional information like coordinates, pressure, angle etc.; however, on the other hand, offline signatures are captured on paper which is then digitised with digital scanners. Offline mode of the signature does not contain auxiliary information. Absence of extra details makes the offline mode of the signature a challenging problem [3].

Despite being a simple biometric, it has been used in many organisations because of the simple and familiar acquisition process, but it is also prone to forgery. Forgery in signature biometrics can be divided into two types, that is, random forgery and skilled forgery. In random forgery, the forger does not have any information about a genuine user and the forger bluffs the genuine signature. On the other hand, in skilled forgery, the forger has the information about the genuine signature of the user and the forger tries to copy the genuine signature [4].

Figure 1 shows the different types of forgery. Figure 1a is the original image, Figure 1b is the random forged signature from the same dataset and Figure 1c is the skilled forged signature. It is shown in Figure 1 that skilled forged signature is approximately same as the original image.

Two approaches have been described in the literature for verification of the signature, that is, writer-dependent and writer-independent [5]. In writer-dependent process for each writer/signer, a binary classifier has trained, that discriminates between the positive (genuine) and negative (forged) sample. On the other hand, in writer-independent method, a single classifier has been trained for the discrimination of the negative and positive sample [5].

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2021 The Authors. IET Biometrics published by John Wiley & Sons Ltd on behalf of The Institution of Engineering and Technology.

IET Biometrics. 2021;10:117–126.

wileyonlinelibrary.com/journal/bme2 | 117
Signatures have intra-class variation, one sample from one class differs from the other samples of the same class and also there can be inter-class similarity, that is, two persons can sign in the same manner. The intra-class variation can happen because of many intrinsic (mood, mental state and cognitive behaviour of the person) and extrinsic (background, pen, ink colour, physiological style etc.) parameters. To solve this problem, we have to find the feature set that can easily distinguish between the intra-class and inter-class samples. Moreover, signature images do not have much textural information, and the handcrafted, descriptor-based methods do not perform well in the extraction of the good features. To extract the better features, good feature extraction model is required [6]. In today’s era, deep learning models are outperforming in terms of feature extraction, object detection and classification. With this motivation we have used deep learning model in our work.

Rest of the article is organised into six sections. Section 2 contains related literature to the proposed study. In Section 3 the proposed method has been described. Experiment and results have been discussed in Sections 4 and 5, respectively. Conclusion is discussed in Section 6.

2 | RELATED WORK

Although many offline signature verification schemes have been proposed [7–15]; mostly the schemes are either purely handcrafted feature based or conventional machine learning based. Because of very less textural information in the images, handcrafted features do not help much and feature engineering becomes vital [6]. In Ref. [6] authors have stated that, sometimes handcrafted features do not have much significance in the signature images, and suggested to use a good model for feature extraction. Recently many deep learning-based architectures for signature verification have been proposed [7,9,11,14].

In Ref. [7], authors have presented a method for signature verification that utilised Siamese network. Siamese network is a twin weight shared network, which uses distance-based loss function. In their work network accepts the genuine–genuine pair or genuine–forged pair. They have used contrastive loss function as a distance-based loss function. Contrastive loss function utilises the Euclidean distance. The distance-based loss function has been used to reduce the distance between genuine images and increase the distance between genuine and forged images. Neural Network has a limitation that all the input images should be of the same size, because of the fixed size neurons in the fully connected layer. To overcome the limitation, authors in Ref. [16] have presented a method that takes the different size of input images and to produce the fixed-length vector before fully connected layer, they have used Spatial Pooling Pyramid (SPP) [17]. SPP provides the fixed-length vector from the different sized input images.

For better generalisation, the system requires a large amount of data, but in case of real-time signature verification problem, we do not have an ample amount of data. To address this issue, authors in Ref. [9] presented a method that requires only one genuine signature of the signer. They identified the sequence of connected pixels of those having the same intensity in one direction. They have calculated the four sequences in four directions. Based on these four sequences, a feature vector was formed. One-class Support Vector Machine (SVM) was used for the classification. More information about the one-class SVM can be found in Ref. [18]. Authors in Ref. [14] have presented a method that uses writer-dependent and writer-independent process. They have used a two Channel CNN for feature extraction and verification. In the two channels, the first channel takes the reference signature and second takes the query signature image. This network does not focus on the learning of the distance metric like Siamese network. At the end of the two-channel CNN, two-way Soft-max layer has been added. To overcome the problem of the over-fitting, firstly authors train the network with the grey images, later with the binary images. In the end, they have calculated the matching score between the query image and the reference image and based on the decision threshold, verification has been performed. The extracted features by the two-channel network are fed to the SVM classifier with RBF kernel to perform the writer-dependent classification. In Ref. [11], authors have used four streams of the same network with the attention block. In the four streams, the two streams were taking the discriminative samples, and the other two collected the inverse samples. The attention block sends the information from inverse stream to discriminative stream and vice-versa, and forces in extracting the essential features of the signature.
image. Out of these four streams three pairs have been formed by merging the alternative pairs of the stream. Global Average pooling layer has been applied before the fully connected layer. At the last layer, the decision has been taken with the majority voting method.

This can be seen from the referred literature that authors have used SVM, Distance or Neural network based classifiers for the purpose of signature verification. Among those, SVM and Neural Network-based classifier give good performance. SVM has been chosen over Neural Network because it is not a solution for local minima, and has less chances for over-fitting [19,20].

Despite the advancements in the field of signature identification and verification, identification of forged sample is a challenging task. As shown in Figure 1 skilled forged signature is approximately same as the genuineness of the user. So it is required to use such a model that can provide robust feature set. With the motivation from the literature, we have used deep learning models for the extraction of the feature set.

To increase the performance of the system with skilled forged samples, we have proposed a writer-independent and language-independent multi-task method that uses transfer learning. The proposed method uses English dataset for the purpose of fine tuning of the network utilised in the paper, and also features have been extracted from different language dataset with transfer learning. This makes our method language independent, and also only one classifier has been trained for the whole dataset, so the proposed method is writer independent. The method is divided into two tasks:

1. The task one uses multi-class SVM classifier to predict the class of the query image. This task will be used for the identification of the user whether it belongs to the enrolled user or not. This can also be used as a verification with the help of the threshold.
2. Task two uses the binary class SVM classifier to determine whether the query sample is either genuine or forged. This task can be used for the identification of the genuineness of the user whether it is a forged user or the genuine user.

3 | PROPOSED METHOD

In this section, we have described the proposed multi-task method that uses a pre-trained network and SVM classifier. Figure 2 shows the block diagram of the proposed method. It has three phases, namely: (1) Fine-tuning, (2) Feature extraction and (3) Training and identification. To generate a discriminative feature set to distinguish between two different samples, we have used a pre-trained network GoogLeNet [21]. This network has proved its worth in the field of computer vision by providing less error on the ImageNet dataset.

3.1 | GoogLeNet architecture

GoogLeNet has been trained on the ImageNet dataset. This network is 22 layers deep architecture and has 4 000,000 parameters with two auxiliary classifiers and nine inception blocks. The auxiliary classifiers have been added to deal with the problem of the vanishing gradients and to provide better regularisation [21]. There can be variation in the content of the image. Sometimes the whole image is covered with the meaningful content and sometimes the part of the image is covered. To deal with this scenario, GoogLeNet uses inception module. With inception module, the width of the network has been increased, which helps to learn the relevant features according to the content of the image.

In this experiment, we have used GoogLeNet because it has relatively lower error rate as compared with the previously used models for signature images and also it has lesser number of parameters, which make our method to be used in lower configuration devices.

3.2 | Preprocessing

There can be variation in the size of the signature images due to many intrinsic and extrinsic factors. The input images should resized to $224 \times 224$ size, as the network (GoogLeNet) required the specific size of the input images to perform fine-tuning and transfer learning because of the specific number of neurons in the fully connected layer. There are basically three types of image re-sampling which have been used in image processing, that is, Nearest neighbour, Bilinear and Bicubic interpolation. Bicubic interpolation has been used over the other two methods because it retains the correctness of the image as compared with other two interpolation. So, here we have resized the images from each dataset to the size of $224 \times 224$ with bicubic interpolation.

3.3 | Training strategy

In deep networks, the initial layers are used to identify the edge and the high-frequency details from the images, and deeper layers learn the generalised shapes from the images. To keep the edge information from the starting layers of the GoogLeNet, we freeze some initial layers of the network. After performing experiments on freezing the number of layers (5,10,15), it has been found that by freezing 10 layers, we have achieved higher performance. The Figure 3 shows that features after the 10th layer, while freezing the initial 10 layers, provide more related features than the other two settings (i.e. freezing initial 5 and 15 layers).

In this article, we have utilised the concept of fine-tuning and transfer learning. In transfer learning, knowledge learned from one task has been utilised in other but related task. The pre-trained model was trained on a very different dataset (ImageNet) than the signature dataset, and it is suggested in the literature that when we want to use transfer learning on the different datasets, firstly it is required to fine-tune the pre-trained network with the large but related (signature) dataset. To perform fine-tuning of the GoogLeNet, we have used the largest available signature dataset (GPDS). Fine-tuned network...
with GPDS dataset provides the relevant features from other signature datasets.

While performing the experiment, there have been many parameters that were fine-tuned, and those parameters are described in Table 1.

The fine-tuning has been performed for the 20 epochs (after 20 epochs the validation accuracy of the system became saturated). The training has been performed using Stochastic Gradient Descent (SGD) with momentum optimiser [22]. The momentum helps the optimiser to move in the appropriate direction. The optimiser with the momentum is defined as:

\[
    w_t = \lambda w_{t-1} + \eta \nabla \theta f(\theta); \quad \theta = \theta - w_t
\]  

(1)

\[
    \nabla \theta f(\theta) = \frac{1}{m} \sum_{i=1}^{m} \nabla \theta L(x^i, y^i, \theta)
\]  

(2)

where, \( \lambda \) is momentum, \( L \) is the loss function, \( \eta \) is the learning rate, \( f(\theta) \) is the log-likelihood and \( m \) is the number of training samples.

**TABLE 1** Training parameters

| Parameter                  | Value  |
|----------------------------|--------|
| Initial learning rate      | 0.0003 |
| Momentum                   | 0.9    |
| Batch size                 | 10     |
| Gradient threshold         | Inf    |

**FIGURE 2** Block diagram of the proposed method

**FIGURE 3** Features after different freezing layers
3.4 Feature extraction and training

The experiment is divided into two tasks. In task one, the extracted features from the fully connected layer of the fine-tuned network are sent to the multi-class SVM with a linear kernel. This classifier generates the score for the query sample and based on the threshold value user is accepted or rejected in task one. We are performing this task for the purpose of initial scanning so as to check on the first stage whether a query sample is eligible for further analysis or not. While training the multi-class SVM for task one, we have not included features from the skilled forged samples of the signature dataset.

For task two also, the features have been extracted from fully connected layer of the fine-tuned model, and the extracted features are fed to the two-class SVM with the cubic kernel. This task involved skilled forged samples along with the genuine signature images of the signer. The genuine sample was assigned with label 1, whereas forged sample was assigned with label 0. For better generalisation, SVM required negative samples along with the positive samples, which is why we have used skilled forged samples while training [20]. We have used kernelisation because the feature sets of one class are superimposed on the features of the other class and it is not possible to discriminate them in this space. In kernel method of the SVMs, we transpose the feature sets in the higher dimension and in that dimension we fit the linear SVM. With the experiments, we have decided the use of the kernels in the respective task.

The task two identifies whether the given image is genuine or forged. Training of the SVM has been performed in fivefold cross-validation fashion for both the task.

Both SVMs (multi-class and two-class) use one-versus-one multi-class method. In one-versus-one approach, subset of the dataset has been used by the single classifier and also it splits the multi-class problem in the binary class problem for each class.

Function for linear and cubic kernel is shown in Equation (3).

\[
G(x_j, x_k) = x_j^T x_k \\
G(x_j, x_k) = (1 + x_j^T x_k)^3
\]

where, \( x \) represents the data points.

Both the SVMs (multi-class and binary) uses hinge loss. Hinge loss sum over all incorrect class and compare the output of the scoring function \( S = f(x_j \times W) \) for the \( j \)th (incorrect) class label and \( y_j^b \) (correct) class. Hinge loss [23] is defined in Equation (4).

\[
L_i = \max(0, S_j - S_{y_j} + \Delta)
\]

where, \( \Delta \) is the margin of the support vectors from the decision boundary; \( S \) is the scoring function \( y_j \) is the correct class label and \( j \) takes all the incorrect class labels.

Total loss (\( L \)) for the whole training set is defined as the sum over all the loss for each class. The total loss can be interpreted as Equation (5).

\[
L = \frac{1}{N} \sum_{i=1}^{N} L_i
\]

where, \( N \) is the number of classes in the dataset.

4 EXPERIMENT

In this section we have discussed the experimental setup, datasets and the parameters that have been used.

4.1 Experimental setup

The fine-tuning of the experiment has been performed on the workstation having configuration Intel(R) Core (TM) i7-7820 \( \times \) CPU @3.60 GHz, 16 GB RAM and have NVIDIA 1080 Ti of 11 GB.

4.2 Dataset

To check the robustness of the proposed method, we have performed this experiment on a different dataset. These datasets have a different language.

1. GPDS: This is the largest dataset available among the signature community. This dataset involved 4000 individuals [24,25]. Each person has 24 genuine and 30 forged signatures. GPDS dataset has been used for the fine-tuning of the GoogLeNet architecture and smaller version GPDS300 has been used for the testing of the fine-tuned network.

2. BHSig260: This dataset has two regional languages involved, that is, Hindi and Bengali [13]. In Hindi dataset 160 signers were involved and in Bengali dataset 100 users were involved. In both the dataset, each signer has 24 genuine and 30 forged samples.

3. MCYT-75: In this dataset, signatures of 75 users were collected. Each user has 15 genuine and 15 forged samples [26,27]. This dataset has been collected on paper sheets which were digitised using the digital scanners and cropped.

Table 2 shows the detailed description about the datasets used in the experiment.

4.3 Experimental protocol

In this experiment we have used fivefold cross validation. The whole dataset is divided into five-folds. Each fold was considered as a hold for the test set and the classifier is trained with the remaining fourfolds of the dataset. Later, the trained classifier is tested against the fold that was kept hold. The evaluation score has been extracted and the model is trained again for the next fourfold.
For task one, we have considered only the genuine signature of the dataset, and for task two, skilled forged signatures are also considered. Each dataset for both the task in every fold gets four images from each class randomly, so a total of 20 samples will be considered for training. Remaining samples have been considered for testing of the algorithm.

4.4 Performance measures

The results and comparison of the proposed method have been described in Tables 4 and 7. Accuracy, Average Error Rate (AER) and Equal Error Rate (EER) are used for the evaluation of the proposed method. AER has been calculated by the average of the False Acceptance Rate (FAR) and False Reject Rate (FRR). EER is defined where False Positive Rate (FPR) and FRR are equal.

\[
AER = \frac{FAR + FRR}{2}
\]  

5 RESULTS AND DISCUSSION

This section describes the results of the proposed method and the comparison of the proposed method with state of the art methods. We have performed the experiments on the fine-tuned GoogLeNet and the performances we obtained are better in comparison with the state of the art methods. The reason for better performance is that the features extracted using the model is quite discriminating for different samples. Figure 4 shows the features of different layers of fine-tuned GoogLeNet on two different and same Bengali signature samples. From the figure, it is evident that fine-tuned GoogLeNet extracted different features from samples of different class and similar features from the samples of the same class. The discrimination of the features can also be seen through Figure 5. The figure shows the distribution of the Euclidean distance of intra-class and inter-class feature set.

We have also performed the experiment with different number of training samples per user, that is, 12, 10 (as in some literature) and the remaining samples for testing of the algorithm. Results of different training samples have been shown in Table 3. The possible reason for the good performance with large number of samples is because with the large sample the model trained generalises well. It is clear from Table 3 that by using 20 samples per user, we have received the good performance. So for comparison with the state of the art methods, we have considered results which we get using 20 training samples per user.

Table 4 shows the results of the proposed method with and without using skilled forgery. We achieved 99.31% accuracy with random forgery on GPDS300 dataset. The possible reason for the higher accuracy on GPDS300 is that GoogLeNet is fine-tuned with the larger version of this dataset GPDS Synthetic. While on Bengali and Hindi dataset, we achieved
96.5% and 95.7% accuracy, respectively, with skilled forgery included in the system. We have achieved accuracy of 87.5% for GPDS-300 dataset when skilled forgeries have been included in the system. While on the MCYT-75 dataset we have recorded 98.6% and 93.9% accuracies on task one and task two respectively.

Table 5 and 6 shows the AER and EER on different datasets respectively. AER is the average of the False Accept Rate (FAR) and False Rejection Rate (FRR).

On the Bengali dataset we have received the minimal AER and we have also reported the maximum accuracy on the Bengali dataset with skilled forged samples; this justifies the performance of the proposed method.

Table 7 shows the comparative results in terms of accuracy and EER of the proposed method with the state of art methods.

It is clear from the table that our method gives comparative results on the publicly available datasets with the state of the art methods [7,11–13,28]. Inverse Discriminative Networks (IDN) [11] uses four parallel streams from two samples. From the four streams, the two are the inverse streams and two are the discriminative streams. They have shown that the accuracy on the Hindi and Bengali dataset are 93.04% and 95.32% respectively.

In Ref. [7], authors have used the Siamese network for the verification of the signature images. Siamese network is the twin network with the same weights and parameters. The network takes both genuine and forged signature samples for training and generalisation of the network. Authors have achieved the accuracy on the Bengali and Hindi dataset by 86.11% and 84.64%, respectively, but on GPDS-300 dataset the
The performance of their system is degraded. With our model we have achieved accuracy of 87.5% on GPDS dataset, which is slightly lower than the previously reported method [12], but it should be noted that we have reported results with the fivefold cross validation. Our method gives higher accuracy on Hindi and Bengali datasets that is, 95.7% and 96.5%, respectively.

This shows that our method is language independent. Also this is because of the pre-trained model which is fine-tuned with the GPDS dataset that has signature images in the English language. But we have tested the method with the signature of the regional languages (Bengali and Hindi) other than MCYT and GPDS-300 dataset that are in the English language.

Table 8 shows the AER comparison of the proposed method with the state of the art methods on GPDS and MCYT-75 dataset. For MCYT dataset our method gives lower AER as compared with Refs. [32–35]. Lower AER signifies that the True Acceptance Rate (TAR) is high and this determines that the system is having higher performance.
In IDN, authors have utilised the network in four streams along with eight attention blocks. The computational complexity of IDN is high as compared with the proposed method. Also, to train the four streams parallelly takes large time and huge computational power.

Figure 6 shows the ROC curve for all the datasets on the proposed method. It is clear from the figure that the area under the Bengali ROC curve is better and this signifies that the Bengali dataset gives comparatively good performance on the proposed method. This validates the results provided in Table 7.

6 | CONCLUSION

In this study, we propose a writer-independent and language-independent multi-task signature identification method that uses GoogLeNet as its base architecture. We have fine-tuned GoogLeNet with the largest signature dataset available. For the other datasets (MCYT-75, BHSig260 and GPDS-300) we have transferred the knowledge of the fine-tuned network and the extracted features are sent to the SVM classifier. The proposed method is divided into two tasks: in task one, we have identified the signer of the signature image and in task two we check whether the query image is genuine and forged. Experiments on the BHSig260 dataset show that the proposed method achieves considerably better performance than the state of the art methods irrespective of the language of the signature. The proposed method is language-independent because we have achieved 96.5% and 95.7% accuracy on Bengali and Hindi dataset respectively, when the pre-trained model is fine-tuned with the GPDS dataset that is basically a signature dataset in the English language.

REFERENCES

1. Ferrer, M.A., et al.: Robustness of offline signature verification based on gray level features. IEEE Trans. Inf. Forensics Secur. 7(5), 966–977 (2012)
2. Jain, A.K., Nandakumar, K., Ross, A.: 50 years of biometric research: Accomplishments, challenges, and opportunities. Pattern Recogn. Lett. 79, 80–105 (2016)
3. Diaz, M., et al.: A perspective analysis of handwritten signature technology. ACM Comput. Surv. 51(6), 1–39 (2019)
4. Hafemann, L.G., Sahouri, R., Oliveira, L.S. Offline handwritten signature verification—literature review. In: IEEE 2017 Seventh International Conference on Image Processing Theory, Tools and Applications (IPTA), pp. 1–8 (2017)
5. Dey, S.: Analyzing features learned for offline signature verification using deep cnns. In: 2016 23rd International Conference on Pattern Recognition (ICPR), pp. 2989–2994 (2016)
6. Dey, S.: Writer-independent feature learning for offline signature verification using deep convolutional neural networks. In: IEEE 2016 International Joint Conference on Neural Networks (IJCNN), pp. 2576–2583 (2016)
7. Dey S., et al.: Signet: convolutional siamese network for writer independent offline signature verification. CoRR, 1707, 02131, 2017
8. Okawa, M.: Synergy of foreground–background images for feature extraction: offline signature verification using Fisher vector with fused kaze features. Pattern Recogn. 79, 480–489 (2018)
