Intelligent Automation & Soft Computing
DOI: 10.32604/iasc.2023.033578
Article

Tech Science Press

Multi-Path Attention Inverse Discrimination Network for Offline Signature Verification

Xiaorui Zhang[1,2,3,4,*], Yingying Wang[1], Wei Sun[4,5], Qi Cui[6] and Xindong Wei[7]

[1]School of Computer and Software, Nanjing University of Information Science & Technology, Nanjing, 210044, China
[2]Wuxi Research Institute, Nanjing University of Information Science & Technology, Wuxi, 214100, China
[3]Engineering Research Center of Digital Forensics, Ministry of Education, Jiangsu Engineering Center of Network Monitoring, Nanjing, 210044, China
[4]Jiangsu Collaborative Innovation Center of Atmospheric Environment and Equipment Technology (CICAEET), Nanjing University of Information Science & Technology, Nanjing, 210044, China
[5]School of Automation, Nanjing University of Information Science & Technology, Nanjing 210044, China
[6]Department of Electrical and Computer Engineering, University of Windsor, Windsor, N9B 3P4, Canada
[7]School of Teacher Education, Nanjing University of Information Science & Technology, Nanjing, 210044, China
*Corresponding Author: Xiaorui Zhang. Email: zxr365@126.com

Received: 21 June 2022; Accepted: 22 September 2022

Abstract: Signature verification, which is a method to distinguish the authenticity of signature images, is a biometric verification technique that can effectively reduce the risk of forged signatures in financial, legal, and other business environments. However, compared with ordinary images, signature images have the following characteristics: First, the strokes are slim, i.e., there is less effective information. Second, the signature changes slightly with the time, place, and mood of the signer, i.e., it has high intraclass differences. These challenges lead to the low accuracy of the existing methods based on convolutional neural networks (CNN). This study proposes an end-to-end multi-path attention inverse discrimination network that focuses on the signature stroke parts to extract features by reversing the foreground and background of signature images, which effectively solves the problem of little effective information. To solve the problem of high intraclass variability of signature images, we add multi-path attention modules between discriminative streams and inverse streams to enhance the discriminative features of signature images. Moreover, a multi-path discrimination loss function is proposed, which does not require the feature representation of the samples with the same class label to be infinitely close, as long as the gap between inter-class distance and the intra-class distance is bigger than the set classification threshold, which radically resolves the problem of high intra-class difference of signature images. In addition, this loss can also spur the network to explore the detailed information on the stroke parts, such as the crossing, thickness, and connection of strokes. We respectively tested on CEDAR, BHSig-Bengali, BHSig-Hindi, and GPDS Synthetic datasets with accuracies of 100%, 96.24%, 93.86%, and 83.72%, which are more accurate than existing signature verification methods. This is more helpful to the task of signature authentication in justice and finance.

Keywords: Offline signatures; biometric verification; multi-path discrimination loss; attention mechanisms; inverse discrimination

1 Introduction

The handwritten signature is one of the most important behavioral biological characteristics [1]. Every day, a large number of important documents are signed all over the world, so signature verification is an unsupervised identification task [2] that is widely used. Signature verification aims at extracting discriminative information from signature images to verify authenticity without identifying the content or meaning of signature images. Signature authentication is a legal means of biometric authentication [3]. Signature authentication is often used in judicial authentication, finance, commerce, and other fields. Once the verification is wrong, it may cause serious consequences such as judicial justice and property loss. Therefore, it is particularly necessary to develop a model with higher authentication accuracy.

The task of author-independent offline signature verification has been a challenge in computer vision, and domestic and international scholars have proposed many different methods to perform this task. Author-independent verification is an approach with different training and test samples that requires the network to learn generic information for distinguishing between genuine and forged images. Offline signatures refer to signatures captured by scanners or any other imaging devices [4]. In the early stage, traditional manual features were mainly used for authenticity classification, e.g., directional gradient features, HOG [5], and LBP [5], and contour shape features, sparse coding [6]. These features need to be carefully designed according to the characteristics of the data. If the data source changes, it needs to be redesigned, and the extraction process is time-consuming. In contrast, features extracted by networks do not need manual participation. We only need to design the network model to extract features, which is conducive to feature extraction from large-scale datasets [7]. The latest development in signature verification [8–11] solves the problem of signature verification only by image classification, rather than modeling signature images themselves. However, due to the particularity of signature images, different from ordinary images, the existing classification methods by networks are not very accurate. First, the strokes in the signature image are slim and most of the area is background, so the signature image feature vector extracted by networks contains less valid information and most of the information is invalid in the background area. Second, the signature images may change a little with time, place, and the physical and mental state of signers, which has high intra-class variability.

This study designs a unique multi-path attention inverse discrimination network. First, considering the high intra-class variability of signature images, this study designs multi-path attention modules based on two pooling strategies and proposes a multi-path discrimination loss function to ensure that the network can tolerate some intra-class variability on the premise of correct classification. The module acts between the features extracted from different layers of the inverse stream and the discriminative stream. The output features of the inverse stream transmit the channel and spatial attention information to the convolutional layer of the discriminative stream; therefore, the network can learn the discriminative information of signature images and enhance important features. The multi-path discrimination loss measures the difference between input samples. As long as the gap between inter-class distance and the intra-class distance is bigger than the set classification threshold, it does not require that the feature representation of the samples with the same class label is infinitely close. Therefore, the multi-path discrimination loss allows the feature of the samples with the same class label to be different within a certain range, and the multi-path discrimination loss performs well in distinguishing samples that have different class labels but are particularly similar. Secondly, this study proposes a method of inverse discrimination. By making the same prediction on the original signature image and the inverse signature image, the difference in background will not affect the results predicted by the model. In this way, the feature of the blank background part, which does not obviously affect the verification, is weakened, and the feature of the stroke parts is strengthened. Therefore, our model can extract more discriminative features by focusing on strokes, even if there is a large blank background in images. This method can solve the problems of less effective information and low accuracy caused by a large blank background in the signature image.

In summary, our contributions are as follows.

i) This study proposes a method of inverse discrimination that makes the network perform the same prediction for the original signature image and the inverse signature image with different backgrounds. This study weakens the feature of the blank background part so that the network can automatically focus on the stroke parts to extract the detailed features, which effectively alleviates the problem of a wide range of blank background yet less effective information in the signature images.

ii) This study has designed the multi-path attention modules, which use the feature pooling results obtained by deeper convolution in the inverse stream to guide the extraction of shallow features in the discriminative stream. By passing attention information based on channel and spatial in the inverse stream to the discriminative stream, we can enhance the features of important channels and spaces, so that the network can learn the discriminative information.

iii) This study proposes a multi-path discrimination loss function. The loss consists of triple loss and inverse discrimination loss. Triple loss allows the difference to some extent by setting the classification threshold, which can overcome the problem that signature images have high intra-class differences. Through the inverse discrimination loss, the network punishes the signature pairs that make different predictions for the original signature image and the inverse signature image. Therefore, the inverse discrimination loss can constrain the network to make the same prediction for the original signature image and the inverse signature image with the same stroke and different color backgrounds. In this way, the influence of a large blank background on the prediction is reduced, the feature of blank background is weakened, and the feature of stroke parts is relatively strengthened.

The rest of this paper is structured as follows. The second part introduces the related work; The third part introduces the data set used, the technology involved, and the proposed method; The fourth part shows the results of this study and the comparison and conclusion with other similar work; The fifth part summarizes the paper and prospects the future research direction.

2 Related Work

Due to its importance in finance, legal, and business environments, signature verification has been widely studied in the past decades. This chapter will introduce the related work involved in our method in detail, from three aspects: network structure, attention mechanism, and inverse discrimination.

2.1 Network Structure

The existing models in the signature verification community mainly include a two-channel network and a Siamese network. Zagoruyko et al. [8] compared various network architectures and stated that two-channel networks are flexible and fast to train, but difficult to test. Yilmaz et al. [12] proposed a two-channel network to compare two signature images. The two signature images are concatenated into a two-channel image as the input of the network. Unlike the Siamese network [13], the two-channel network has no shared branches, and thus it is faster to train.

However, two-channel networks are time-consuming in the test stage, and it needs to calculate the similarity of each corresponding channel of the two input images, and then combine the similarity of all channels as the final classification result.

In contrast, the Siamese network is fast and simple in the test stage [14], and it has been successfully used for face verification tasks by weakly supervised metric learning. The Siamese network is a coupled architecture, and it contains two same sub-networks with the same parameters and weights. Siamese networks were first proposed by LeCun et al. [15] for verifying signature images and achieved the highest accuracy at that time. Dey et al. [16] proposed a SigNet model based on the Siamese network for exploring the small differences between real signature images and forged signature images.

The signature images of the same signer will change a little over time, requiring the network to allow for a range of differences between the feature of samples of the same class. Because a triple loss function allows differences to some extent in classes, it is more appropriate to use in signature image verification. The proposed method inverses the signature image for inverse discrimination, that is, there are six input images. However, commonly, Siamese networks allow only two input samples. Therefore, this study uses a variant of the Siamese network, which is an architecture of six-stream with six input samples. The details can be found in Section 3.1 Network Architecture in this paper.

2.2 Attention Mechanism

The attention mechanism can not only make the network focus on a specific space [17] but also enhance the feature of this space. For example, in the signature verification task, the attention mechanism is used to make the network focus more on the stroke parts and enhance the features of stroke parts. There are three types of attention mechanisms: channel attention, spatial attention [18], and channel-spatial attention. Channel attention models the interdependence among channels and determines the importance of each channel. Hu et al. [19] proposed the SENet model to improve accuracy by modeling correlations among channels and assigning more weights to the important channels, which won the championship of the 2017 Large Scale Visual Recognition Challenge (ILSVRC) competition. Yue et al. [20] proposed a Progressive Channel Attention Network (PCANet), where a novel channel-attention module (CAM) is used to estimate channel parameters by weighted average calculation instead of the global average calculation.

However, channel attention pays no attention to the spaces that need to be focused on in each feature map [21]. In contrast, spatial attention extracts a spatial attention matrix to determine the spaces that need to be focused on. Channel-spatial attention combines the advantages of channel attention and spatial attention. Chen et al. [22] proposed the SCA-CNN model that incorporates the spatial and channel-wise attention in a CNN. Woo et al. [23] proposed the CBAM model, which is based on the feature map to sequentially derive attention maps along two independent dimensions of the channel and spatial, and then the attention map is multiplied by the input feature map to refine the features. This module can be integrated into any CNN architecture.

However, the channel-spatial attention only acts on the feature map of the same stream and does not fuse the information of different streams. Therefore, this study proposes multi-path attention modules that act on different streams. The deeper layers of the inverse stream, transmit attention information based on channel and spatial to the shallow layers of the discriminative stream. In this way, we enhance the feature of important channels and spaces and guide the network to learn distinguishing features. The details can be found in Section 3.3 of this paper.

2.3 Inverse Discrimination

To improve the prediction ability of the network in the confusion area for which it is difficult to judge its classification, Huang et al. [24] propose an inverse attention network (RAN). The first branch learns the probability of the confusion area belonging to each class; The second branch learns the probability of the confusion area does not belong to each class. The third branch combines the prediction results of the first two branches and outputs the combined prediction. The performance of the network is improved by combining the probability that the confusing space belongs to each class and the probability that it does not belong to each class. Chen et al. [25] proposed a deep salient object detection network. By erasing the current predicted salient regions from the feature maps, the network can eventually explore the missing object parts and details, which results in high accuracy. The proposed network can explore the details of the missing parts of objects by removing the current predicted significant spaces to reduce the weight of these spaces. Inspired by this, this study proposes a method of inverse discrimination. Through training, the network can make the same prediction on signature images with the same strokes and different backgrounds, reducing the weight of the feature of backgrounds. We force the network to focus more on the stroke parts, to meet the challenges of slim strokes and less effective information. The details can be found in section 3.2 inverse discrimination mechanism of this paper.

3 Proposed Method

In this section, we design a novel multi-path attention inverse discrimination network (MAIDNet), which is referred to as MAIDNet. To extract the features of the signature image better, this study proposes an inverse discrimination method and adds multi-path attention modules to the sextuplet network, which is constrained by the multi-path discrimination loss function. Next, we will present details from four aspects: architecture, inverse discrimination module, multi-path attention module, and loss function.

3.1 Architecture

The input of the MAIDNet is anchor, positive and negative triplet pairs, in which anchor, positive and negative are image pairs composed of an original image and an inverse image. The original image refers to the signature image in the dataset, with a white background and gray strokes. The inverse image refers to the image with a black background and gray strokes obtained from the original image by $Inverse_{S_i} = 255 - S_i$. As shown in Fig. 1, the MAIDNet contains six streams, three discriminative streams, and three inverse streams. Each row represents a vgg16 network which is called a stream. There are nine attention modules between discriminative and inverse streams. The images in the datasets are processed into a size of 3 * 148 * 148 after preprocessing. As shown in Fig. 1, each red rectangle represents an attention model. The nine attention models are called multi-path attention models. The three discriminative streams take the original images in anchor, positive, and negative image pairs as input, respectively, and extract the features through the convolution module. Each stream contains five convolution modules. The first two convolution modules contain two convolution layers activated by the Rectified Linear Unit (ReLU) function (convolution kernel size 3 × 3. String is 1, padding is 1) and a max pooling layer (the pooling matrix is 2 × 2, the stripe is 2). The last three convolution modules contain three convolution layers activated by the ReLU function (convolution kernel size 3 × 3. String is 1, padding is 1) and a max pooling layer (the matrix of pooling is 2 × 2, the stripe is 2). The number of convolution cores of five modules in each stream is 64, 128, 256, 512, and 512, respectively. The three inverse streams take inverse images of anchor, positive, and negative image pairs as input, respectively. Inverse streams have the same network structure as a discriminative stream.

The features of different streams are merged into four feature maps corresponding to four pairs: the inverse image in positive and the original image in anchor, the original image in positive and the original image in anchor, the inverse image in negative and the original image in anchor, the original image in negative and the original image in anchor. Through a global average pooling layer (GAP), the four merged features are input to each of the four fully connected layers to calculate the inverse discrimination loss. The features of original images and inverse images which are from the anchor, positive and negative are cascaded, respectively, and the cascaded three sets of features are used to calculate the triplet loss. The sum of the triple loss and inverse discrimination loss is the total multi-path discrimination loss.

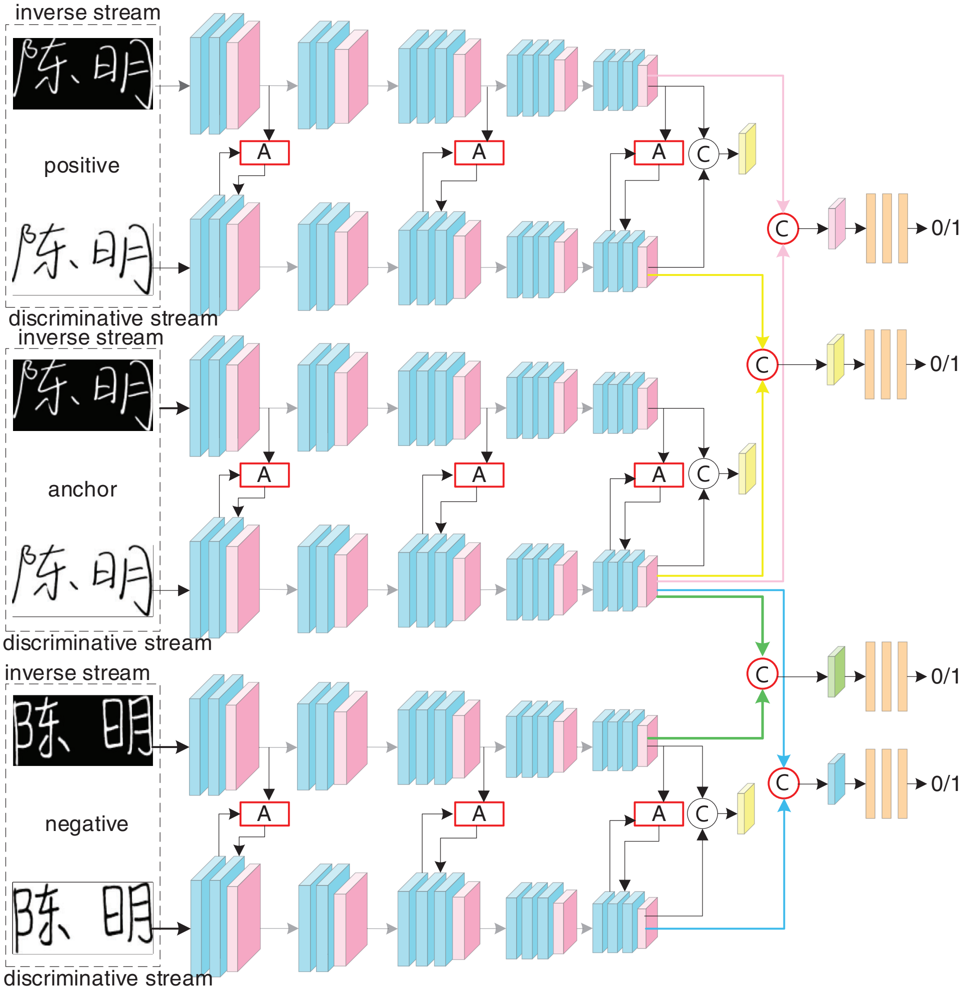

Figure 1: Overview of our method

3.2 *Inverse Discrimination Module*

As shown in Fig. 2, the signature images in anchor and positive are from the same signer, while the signature images in negative and anchor are from different signers. That is, using the signature image in the anchor as a reference, the network should discriminate the image in the positive as a genuine signature and the image in the negative as a forged signature. If the model correctly describes the signature stroke parts, then verification prediction should be independent of the color of the signature images, that is, the model should make the same prediction for the original image and the inverse image. Therefore, the model makes the same prediction for the inverse image in positive and the original image in the anchor (pair 1), and the original image in positive and the original image in the anchor (pair 2). The positive images are judged as true signature images regardless of original or inverse images. Similarly, the model makes the same prediction for the inverse image in the negative and the original image in the anchor (pair 3), the original image in the negative, and the original image in the anchor (pair

4). The negative is judged as a forged signature regardless of the original image or the inverse image. The negative images are judged as false signature images regardless of original or inverse images.

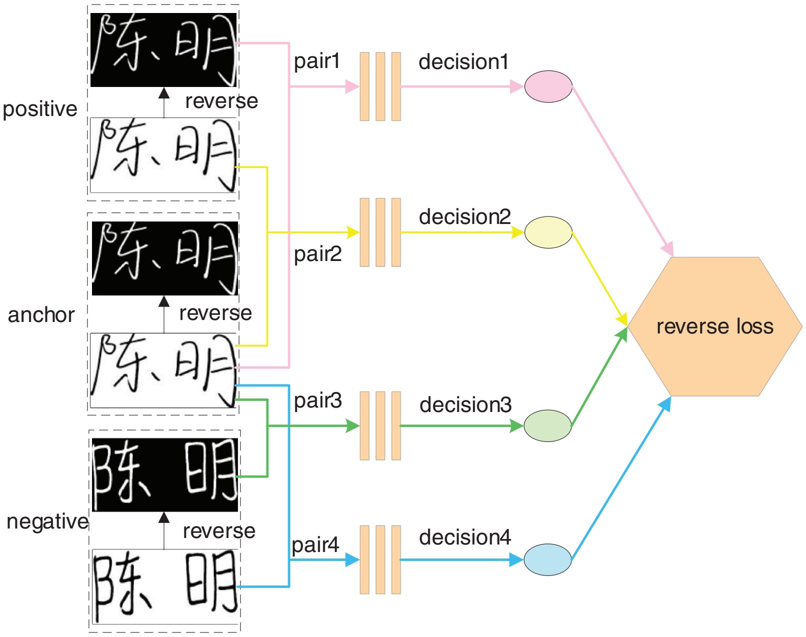

Figure 2: Illustration of inverse verification mechanism

During the training, the network penalizes the inconsistent signature pairs through the reverse authentication loss, forcing the model to make the same prediction on the original signature image and the reverse signature image, so that different backgrounds will not affect the prediction results of the model. That is, the feature representation of the blank background is weakened, and the feature representation of the stroke parts is strengthened, urging the network to automatically focus on the stroke part to extract features. Therefore, this inverse discrimination approach effectively alleviates the problem of less effective information, which is caused by slim strokes of signature images and mitigates the effect of ink dot noise in the background space.

3.3 Multi-Path Attention Module

Attention, a theory of simulating human cognitive behavior, has shown excellent performance in image processing. Each channel contributes differently to critical information, and different spaces of the image have different importance. This study designs multi-path attention modules based on channel and spatial which connect the discriminative stream and the inverse stream. As shown in Fig. 3, there are nine paths of attention modules between discriminative streams and inverse streams. C, H and W respectively represent the number of channels, width and height of the feature map. As shown in Fig. 3, each attention module contains both forward and backing processes. The forward process receives the first layer of features from each convolutional module of the discriminative stream. The backing process is mainly the output features of the inverse stream to pass attention information to the second convolutional layer of the discriminative stream.

We bilinearly interpolate the output feature $F_{inverse}$ of each module of the inverse stream to the same dimension as an output feature of the first convolution layer in the discriminative stream. Global average and global maximum pooling are performed for each channel along the spatial axis, so each two-dimensional feature map is compressed into a real number. We obtain the $C \times 1 \times 1$ feature vectors F^c_{avg} and F^c_{max} and sum element by element to obtain F^c_{sum}, the feature value in F^c_{sum} represents the importance of each channel, and the larger the feature value, the more important the corresponding channel. The output feature $F_{discrimnative}$ of the first convolution layer of each module of the discriminative stream is multiplied by F^c_{sum}. The attention information learned from the inverse stream is transferred to the discriminative stream. Add the result of multiplication to the output feature $F_{discrimnative}$ of the discriminative stream to get the refined feature $F_{channel}$, which is mainly to prevent the gradient from disappearing when the feature value becomes smaller after multiplication. The more important

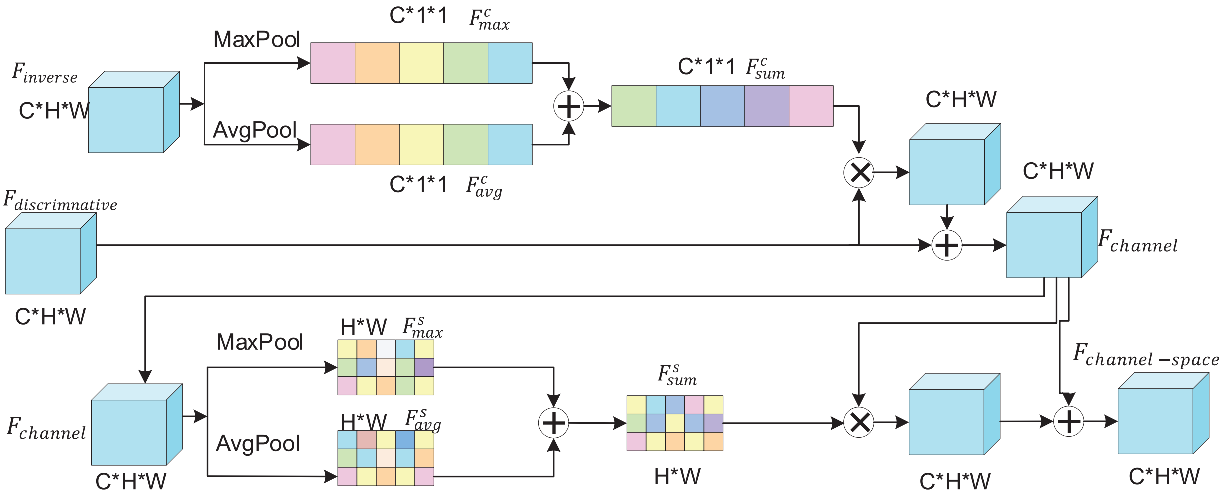

Figure 3: Description of the multi-path attention module

The $F_{channel}$ is subjected to a global average pooling and the global maximum pooling along the channel axis, and the feature vectors F^s_{avg} and F^s_{max} of H * W are obtained. We sum F^s_{avg} and F^s_{max} element by element to get F^s_{sum}, and each eigenvalue in F^s_{sum} represents the importance of each position. The refined feature $F_{channel}$ is multiplied by F^s_{sum} and added with $F_{channel}$ to obtain the refined feature of channel and spatial $F_{channel-space}$. $F_{channel-space}$ weighted the important features in the two dimensions of the channel and spatial, assigned a larger weight to the important features, and enhanced important features, so this feature is more discriminating.

3.4 Loss Function

In order to constrain the network to allow a certain range of differences between different signatures of the same signer and to supervise the network to make the same predictions for the original images and inverse images, this study proposes a multi-path discrimination loss function adapted to our model. The loss consists of triple loss and inverse discrimination loss, as shown in the following equation:

$$L = L_{triplet} + L_{reverse} \tag{1}$$

The first part is the triplet loss, as shown in the following equation:

$$L_{\text{triplet}} = \max(d(a,p) - d(a,n) + margin, 0) \tag{2}$$

$d(a,p)$ and $d(a,n)$ are the Euclidean distances between the eigenvectors of samples and positive samples, samples, and negative samples respectively. Margin is a super parameter used to measure the distance difference between the eigenvectors of positive samples and negative samples. After experimental adjustment, $margin = 0.05$. The loss function punishes the triplet pairs whose distance between classes is larger than the *margin* within classes but does not penalize the triplet pairs whose distance between classes is smaller than the *margin* within classes. In this way, the network won not misjudge when identifying the signatures of the same signer but with subtle differences. This coincides with the characteristic that the signature image may change slightly with time and place.

The second part is the inverse discrimination loss, as shown in the following equation.

$$L_{inverse} = -\sum_{i=1}^{4} \alpha_i \left[y ln y_i + (1-y)\ln(1-y_i) \right] \tag{3}$$

y is a binary label for judging whether a signature image belongs to a real signature or a forged signature, where 1 indicates that the test signature is a real signature and 0 indicates that it is a forged signature. $\hat{y}_i(i = 1, 2, 3, 4)$ are the predicted probability that the test signatures in the four image pairs pair1, pair2, pair3, and pair4 in Fig. 2 are genuine, respectively α is the hyperparameter to adjust the weights of the four pairs. The loss is the sum of four pair's image losses. The network punishes signature pairs that make inconsistent decisions about the original image and the inverse image. The model makes the same prediction for the original signature image and the inverse signature image. In this way, the feature of a largely blank background is weakened, and the feature of stroke parts is promoted, guiding the network to focus on the stroke part to extract features.

4 Experiments

In this section, we evaluate the performance of our proposed method. We conduct ablation experiments to verify the effectiveness of our proposed approach and compare it with the current advanced methods. In addition, we also validate the generalization of our model through cross-language experiments. Next, we will introduce it in detail from the four aspects: datasets, evaluating indicators, implementation details, and experimental results.

4.1 Datasets

There are three datasets that this study used: CEDAR (https://github.com/wk-ff/IDN), BHSig260 (https://github.com/wk-ff/IDN), and GPDS Synthetic (https://gpds.ulpgc.es/). The CEDAR dataset contains the signatures of 55 signers, and each signer has 24 real signatures and 24 forged signatures, with a total of 1320 real signatures and 1320 forged signatures. Each signer was asked to forge the signatures of the other three signers, and each signer forged them 8 times. Signatures were forged randomly on the spot without special training. Therefore, the CEDAR data set is simple and easy to distinguish. The BHSig260 dataset contains the signatures of 260 signers, of which 100 are signed in Bengali and 160 in Hindi. Each signer has 24 real signatures and 30 forged signatures, and there are a total of 6,240 real signatures and 7,800 forged signatures. The GPDS Synthetic datasets consist of 4,000 signatures, each of which has 24 real signatures and 30 forged signatures, with a total of 96,000 real signatures and 120,000 forged signatures. The GPDS Synthetic dataset contains many signature images, and some signers have received special signature training, so this dataset contains a large number of skilled forged signatures. The skilled forged signature is very similar to the real signature, which is even more difficult to distinguish.

We divide each dataset as follows and randomly select M signers from the K (K > M) signers of each dataset. The signatures of these M signers are used for training, and the signatures of the remaining K-M signers are used for testing. Since BHSig260 and GPDS synthetic datasets contain 30 forged signatures for each signer, 720 (real and forged) signature pairs and 576 (real and real) signature pairs can be obtained for each signer. To balance the same and different classes, we randomly select 576 (real and fake) signature pairs from each signer. Table 1 shows the K and M values of different datasets used in our experiments.

Table 1: Datasets partition

Datasets	K	M
CEDAR	55	50
BHSig260-B	100	80
BHSig260-H	160	128
GPDS Synthetic	4000	3200

4.2 Evaluating Indicators

This study uses accuracy (ACC), false rejection rate (FRR), and false acceptance rate (FAR) to comprehensively evaluate our method. The accuracy is the proportion of the number of correct predictions of the number of all samples. The false acceptance rate (FAR) is the proportion of forged signatures misidentified as genuine signatures to all forged signatures, that is, the proportion of false judgment. The false rejection rate (FRR) is the proportion of genuine signatures misjudged as forged signatures to all genuine signatures, that is, the proportion of missed reports. In practical applications, the real signature is misjudged as a forged signature, which will affect the user's experience. The calculation formula is shown below.

$$ACC = \frac{n_{correct}}{n} \tag{4}$$

$$FAR = \frac{n_{false-true}}{n_{false}} \tag{5}$$

$$FRR = \frac{n_{true-false}}{n_{true}} \tag{6}$$

where $n_{correct}$ is the number of correctly predicted samples, and n is the total number of samples. $n_{false-true}$ is the number of samples that are forged signatures and interpreted as the real signatures. n_{true} and n_{false} are the number of real samples and forged samples respectively. $n_{true-false}$ is the number of samples that are real signatures and interpreted as forged signatures.

4.3 Implementation Details

Our training is based on the PyTorch 1.0 framework, and the experimental platform is equipped with NVIDIA RTX3050Ti and i7-8700 CPU. We set the input image size to 224×224 pixels and use mini-batch SGD with an initial learning rate of 1e-4 for training. A total of 8 epochs are trained for CEDAR and BHSig260, and 20 epochs are trained for GPDS Synthetic. We set the batch size to 32 and the boundary value margin to 0.005. To reduce accidental error, we conduct five tests on each dataset and take the average of five test results.

4.4 Experimental Results

4.4.1 Ablation Experiment

We conduct ablation experiments to analyze the importance of each component in the whole model. Table 2 lists the test results of the accuracy of the model on each data set when the corresponding component is missing. In the second and third columns, the attention module and inverse discrimination module are reduced respectively, and the accuracy decreases in different degrees, which proves the effectiveness of our proposed module.

Table 2: Ablation experimental results in units of 10^{-2}

Datasets/ablation module	Normal	Off-attention	Off-inverse learning
CEDAR	100	94.60	98.86
BHSig260-B	96.24	93.72	94.37
BHSig260-H	93.86	91.62	92.42
GPDS	83.72	81.64	82.56

First, we remove the multi-path attention module that acts between the discriminative stream and inverse stream and let each network extract the features of images independently, instead of using the results of a deep layer in the inverse stream to guide the extraction of a shallow layer in the discriminative stream. The experimental results show that the accuracy of the model decreases by 5.4%, 2.52%, 2.24%, and 2.08% on the CEDAR, BHSig260-B, BHSig260-H, and GPDS datasets, respectively. We found that the smaller the number of signature images in the dataset, the higher the accuracy obtained by our module.

Secondly, we remove the inverse discrimination process, cascade the features of the original image and the inverse image, respectively, directly calculate triplet loss instead of inverse supervision loss, and no longer constrain the network to focus on stroke part extraction information. Analyzing the experimental results, we find that the accuracy of the model decreases by 1.14%, 1.87%, 1.44%, and 1.16% on the CEDAR, BHSig260-B, BHSig260-H, and GPDS datasets, respectively. Again, we find that the smaller the number of signature images in the dataset the higher the accuracy obtained by our module. Furthermore, by comparing the ablation experiments, we found that the attention module was more effective than the inverse discrimination module in improving accuracy, which leads us to conclude that using the attention mechanism to enhance discriminative information is a more effective way to improve accuracy for visual classification tasks.

4.4.2 Comparative Experiments

Specifically, our method models the high intra-class variability and less effective information of signature images and can address the problems in signature verification tasks more pertinently. We compared our method with the current methods, and Table 3 shows the results of different methods on three datasets CEDAR, BHSig260, and GPDS Synthetic. The experimental results show that for Accuracy evaluation metrics, our model outperforms the other methods in Table 3 Especially on the CEDAR dataset, we achieve 100% accuracy. This is because we add multi-path attention modules and the inverse discrimination module to the model, which not only enhance the discriminative features of the images, but also motivate the network to focus on the stroke parts to extract detailed features. As a result, the accuracy of validation is improved on all datasets.

The comparison with the method proposed by Ping Wei et al. reveals that the accuracy of our method is higher than their method; however, our FAR metric is higher, that is, our model misclassifies forged signatures as genuine ones more.

Given this result, we guess that the reason may be that our network has relaxed the limit of classification considering that there is a certain degree of difference in the allowed classes, and the boundary value margin of judgment in the loss function is set to be larger, which leads the network to misjudge some forged signatures as real ones.

Table 3: The comparison between our method and other methods in units of 10^{-2}

Databases	Methods	#Singers	Accuracy	FAR	FRR
CEDAR	Hafemann et al. [26]	55	94.32	5.74	3.22
CEDAR	Dutta et al. [27]	55	**100.00**	**0.00**	**0.00**
CEDAR	Dey et al. [16]	55	**100.00**	**0.00**	**0.00**
CEDAR	Wei et al. [28]	55	95.98	5.87	2.17
CEDAR	Zois et al. [1]	55	96.62	5.12	2.03
CEDAR	**Ours**	55	**100.00**	**0.00**	**0.00**
Bengali	Pal et al. [29]	100	66.18	33.82	33.82
Bengali	Dutta et al. [27]	100	84.90	15.78	14.43
Bengali	Dey et al. [16]	100	86.11	13.89	13.89
Bengali	Wei et al. [28]	100	95.32	4.12	5.24
Bengali	Zois et al. [1]	100	95.44	4.02	3.98
Bengali	**Ours**	100	**96.24**	**6.42**	**0.43**
Hindi	Pal et al. [29]	160	75.53	24.47	24.47
Hindi	Dutta et al. [27]	160	85.90	13.10	15.09
Hindi	Dey et al. [16]	160	84.64	15.36	15.36
Hindi	Wei et al. [28]	160	93.04	8.99	4.93
Hindi	**Ours**	160	**93.86**	**9.04**	**3.04**
GPDS	Dey et al. [16]	4000	77.76	22.24	22.24
GPDS	Wei et al. [28]	4000	81.64	18.44	12.26
GPDS	Zois et al. [1]	4000	81.92	18.02	12.31
GPDS	**Ours**	4000	**83.72**	**19.32**	**9.88**

For this result, we guess that the reason may be that our network relaxes the limit of classification considering that a certain degree of variation within classes is allowed. The margin in the loss function is set bigger, which causes the network to misclassify some forged signatures as real ones. We also find that our FRR metric is lower; i.e., the proportion of otherwise genuine signatures misclassified as forgeries are much lower. This shows that our model has learned more robust features about distinguishing genuine signatures from forged signatures and will not easily misclassify genuine signatures for forged signatures.

4.4.3 Cross Language Test

The CEDAR, BHSig260, and GPDS datasets used in this study belong to three different languages. CEDAR and GPDS are English signature datasets, BHSig260-Bengali part of BHSIG260 dataset is Bengali signature, and BHSig260-Hindi part is Hindi signature. To test the accuracy of signature verification across different languages, this study conducted a cross-language experiment. The model was trained on one dataset and tested on another dataset in different languages. For example, we train on Hindi signature dataset and test on the English dataset. Table 4 shows the accuracy of cross-language tests, in which rows correspond to training datasets and columns correspond to test datasets. The data in the Table 4 are the verification accuracies for various combinations of experiments. The experimental results show that for all datasets, the highest accuracy can be obtained by training and testing with the same dataset and the accuracy of cross-language signature verification is greatly reduced. It is concluded that the signature is closely dependent on the language. The accuracy of BHSig260-B and BHSig260-H datasets in cross-language testing is not as severe as that of other datasets, which may be due to the similarity in style between Bengali and Hindi handwritten signatures. Compared with other cross-language tests, the accuracy of the CEDAR dataset and the GPDS dataset decreases more slowly, probably because both datasets are English signature datasets. In addition, we also find that the model is more robust and has higher verification accuracy when trained on a relatively large and diverse dataset than when trained on a small data set. For example, the accuracy with the training set of GPDS synthesis and the test set of CEDAR is 80.16%, while the accuracy with the training set of CEDAR and the test set of GPDS synthesis is only 56.84%, which is 23.32 percentage points lower than the first case. We guess that expanding the number of samples in the training set can improve the accuracy of cross-language testing.

Table 4: Cross language test accuracy results in units of 10^{-2}

Train/Test	CEDAR	Bengali	Hindi	GPDS synthetic
CEDAR	100	64.15	55.61	54.26
CEDAR	100	65.20	57.33	56.84
Bengali	50.00	86.81	64.57	52.66
Bengali	53.52	96.24	74.35	56.32
Hindi	59.57	60.65	84.64	52.78
Hindi	63.84	72.20	93.86	53.64
GPDS	79.13	66.65	63.77	77.76
GPDS	80.16	67.80	64.52	83.72

As can be seen from Table 4, when cross-language tests are conducted, the accuracy of the verification drops sharply. However, by comparing our method with that proposed by Bromley et al. [15], it is found that the accuracy of our model is higher, with a relatively small decrease. Therefore, our model has better generalization ability and better robustness.

5 Conclusion

This study proposes a multi-path attention inverse discrimination network (MAIDNet). Compared with existing methods, instead of simply image classification, this study proposes a solution based on the characteristics of signature images. By using the inverse discrimination method, multi-path attention module, and multi-path discrimination loss function, we effectively solve the problem that most of the signature images are blank background spaces with less valid information and signature images change slightly with time. The experimental results show that our method can extract more robust features and improve the accuracy of verification. The effectiveness of our proposed method was proved by ablation experiments, and the superiority of our method was proved by comparative experiments.

In the future, the proposed method can be extended in the following directions. (i) Explore the method of using anchor generation and non-maximum suppression technology [30] to locate information-rich and discriminative regions in the signature image to obtain local features with clearer details. Combine global features with local features instead of only using global features for signature verification. (ii) We can calculate the dot product of the features of the original image and the inverse image to calculate the high-order feature [31], instead of just cascading. (iii) Our work can be extended to the field of face authentication with a large blank background.

Acknowledgement: We would like to thank the support of the School of Nanjing University of Information Science & Technology.

Funding Statement: This work was supported, in part, by the National Nature Science Foundation of China under grant numbers 62272236; in part, by the Natural Science Foundation of Jiangsu Province under grant numbers BK20201136, BK20191401; in part, by the Priority Academic Program Development of Jiangsu Higher Education Institutions (PAPD) fund.

Conflicts of Interest: We declare that we have no conflicts of interest to report regarding the present study.

References

[1] E. N. Zois and E. Zervas, "Sequential motif profiles and topological plots for offline signature verification," in *Proc. CVPR*, Seattle, WA, USA, pp. 13248–13258, 2020.

[2] Y. Dai and Z. Luo, "Review of unsupervised person re-identification," *Journal of New Media*, vol. 3, no. 4, pp. 129–136, 2021.

[3] Y. Guerbai, Y. Chibani and B. Hadjadji, "The effective use of the one-class SVM classifier for handwritten signature verification based on writer-independent parameters," *Pattern Recognition*, vol. 48, no. 1, pp. 103–113, 2015.

[4] V. L. Souza, A. L. Oliveira and R. Sabourin, "A writer-independent approach for offline signature verification using deep convolutional neural networks features," in *Proc. BRACIS*, São Paulo, SP, Brazil, pp. 212–217, 2018.

[5] M. B. Yilmaz, B. Yanikoglu, C. Tirkaz and A. Kholmatov, "Offline signature verification using classifier combination of HOG and LBP features," in *Proc. IJCB*, New York, NY, USA, pp. 1–7, 2011.

[6] E. N. Zois, M. Papagiannopoulou, D. Tsourounis and G. Economou, "Hierarchical dictionary learning and sparse coding for static signature verification," in *Proc. CVPR workshops*, Salt Lake City, UT, USA, pp. 432–442, 2018.

[7] T. Jiang, "A review of person re-identification," *Journal of New Media*, vol. 2, no. 2, pp. 45–60, 2020.

[8] S. Zagoruyko and N. Komodakis, "Learning to compare image patches via convolutional neural networks," in *Proc. CVPR*, Boston, Massachusetts, USA, pp. 4353–4361, 2015.

[9] M. B. Yilmaz and K. Ozturk, "Hybrid user-independent and user-dependent offline signature verification with a two-channel CNN," in *Proc. CVPR*, Salt Lake City, UT, USA, pp. 526–534, 2018.

[10] Y. Taigman, M. Yang, M. Ranzato and L. Wolf, "Closing the gap to human-level performance in face verification," in *Proc. CVPR*, Columbus, Ohio, USA, pp. 1701–1708, 2014.

[11] S. Dey, A. Dutta, J. I. Toledo, S. K. Ghosh and J. Lladós, "Signet: Convolutional siamese network for writer independent offline signature verification," *Pattern Recognition*, vol. 53, no. 1, pp. 93–102, 2017.

[12] M. B. Yilmaz and K. Ozturk, "Hybrid user-independent and user-dependent offline signature verification with a two-channel CNN," in *Proc. CVPR*, Salt Lake City, UT, USA, pp. 526–534, 2018.

[13] J. M. Zhang, J. Sun, J. Wang, Z. Li and X. Chen, "An object tracking framework with recapture based on correlation filters and Siamese networks," *Computers & Electrical Engineering*, vol. 98, no. 1, pp. 107730–107737, 2022.

[14] W. Sun, L. Dai, X. R. Zhang, P. S. Chang and X. Z. He, "RSOD: Real-time small object detection algorithm in UAV-based traffic monitoring," *Applied Intelligence*, vol. 52, no.8, pp. 8448–8463, 2022.

[15] J. Bromley, I. Guyon, Y. LeCun, E. Säckinger and R. Shah, "Signature verification using a "siamese" time delay neural network," *Advances in Neural Information Processing Systems*, vol. 7, no. 4, pp. 669–688, 1993.

[16] M. K. Kalera, S. Srihari and A. Xu, "Offline signature verification and identification using distance statistics," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 18, no.7, pp. 1339–1360, 2004.

[17] J. Chen, Z. Zhou, Z. Pan and C. Yang, "Instance retrieval using region of interest based CNN features," *Journal of New Media*, vol. 1, no. 2, pp. 87–99, 2019.

[18] W. Sun, G. Z. Dai, X. R. Zhang, X. Z. He and X. Chen, "TBE-Net: A three-branch embedding network with part-aware ability and feature complementary learning for vehicle re-identification," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 9, pp. 14557–14569, 2022.

[19] J. Hu, L. Shen and G. Sun, "Squeeze-and-excitation networks," in *Proc. CVPR*, Salt Lake City, UT, USA, pp. 7132–7141, 2018.

[20] H. J. Yue, S. Shen, J. Y. Yang, H. F. Hu and Y. F. Chen, "Reference image guided super-resolution via progressive channel attention networks," *Journal of Computer Science and Technology*, vol. 35, no. 3, pp. 551–563, 2020.

[21] W. Sun, X. Chen, X. R. Zhang, G. Z. Dai, P. S. Chang *et al.,* "A multi-feature learning model with enhanced local attention for vehicle re-identification," *Computers, Materials & Continua*, vol. 69, no. 3, pp. 3549–3560, 2021.

[22] L. Chen, H. Zhang, J. Xiao, L. Nie, J. Shao *et al.,* "Spatial and channel-wise attention in convolutional networks for image captioning," in *Proc. CVPR*, Honolulu, HI, USA, pp. 5659–5667, 2017.

[23] S. Woo, J. Park, J. Y. Lee and I. S. Kweon, "Cbam: Convolutional block attention module," in *Proc. ECCV*, Munich, MUC, Germany, pp. 3–19, 2018.

[24] Q. Huang, C. Xia, C. Wu, S. Li, Y. Wang *et al.,* "Semantic segmentation with reverse attention," arXiv, preprint arXiv: 1707. 06426, 2017.

[25] S. Chen, X. Tan, B. Wang and X. Hu, "Reverse attention for salient object detection," in *Proc. ECCV*, Munich, MUC, Germany, pp. 234–250, 2018.

[26] L. G. Hafemann, R. Sabourin and L. S. Oliveira, "Offline handwritten signature verification—literature review," in *Proc. IPTA*, New York, NY, USA, pp. 1–8, 2017.

[27] A. Dutta, U. Pal and J. Lladós, "Compact correlated features for writer independent signature verification," in *Proc. ICPR*, Cancun, CUN, Mexico, pp. 3422–3427, 2016.

[28] P. Wei, H. Li and P. Hu, "Inverse discriminative networks for handwritten signature verification," in *Proc. CVPR*, Long Beach, CA, USA, pp. 5764–5772, 2019.

[29] S. Pal, A. Alaei, U. Pal and M. Blumenstein, "Performance of an offline signature verification method based on texture features on a large indic-script signature dataset," in *Proc. IAPR Workshop*, New York, NY, USA, pp. 72–77, 2016.

[30] Z. Yang, T. Luo, D. Wang, Z. Hu, J. Gao *et al.,* "Learning to navigate for fine-grained classification," in *Proc. ECCV*, Munich, MUC, Germany, pp. 420–435, 2018.

[31] T. Y. Lin, A. RoyChowdhury and S. Maji, "Bilinear CNN models for fine-grained visual recognition," in *Proc. ICCV*, Santiago, SCU, Chile, pp. 1449–1457, 2015.