Journal of Intelligent System and Applied Data Science (JISADS)

Journal homepage : https://www.jisads.com
ISSN (2974-9840) Online

DEEP LEARNING FOR OFFLINE SIGNATURE VERIFICATION: A NOVEL MULTI-CHANNEL FEATURE FUSION NETWORK

Harshal Hemane[1] *, *Anuradha Kasangottuwar*[2]

[1] * *E&TC Engineering Department, DACOE, Karad, India*
[2]*PES Modern COE in department of E&TC Engineering, Pune, Maharashtra, India*

ABSTRACT

This paper addresses the critical challenge of offline signature verification, a task crucial for authenticating documents and identities. Existing deep learning approaches, primarily deep metric learning with Siamese networks and two-channel discriminative methods, face limitations. While Siamese networks excel at feature extraction, their reliance on Euclidean distance can overlook subtle directional and scaling information, hindering the capture of intricate feature relationships. Conversely, two-channel discriminative methods, though effective in initial dissimilarity assessment, often suffer from significant feature loss due to early image fusion. To overcome these challenges, we propose the Multi-channel Feature Fusion Network , a novel writer-independent model for handwritten signature verification. The proposed framework leverages a quadruple Siamese network and a dual inverse discriminative attention mechanism for robust feature extraction and enhancement from both original and inverse grayscale images. These rich, multi-dimensional features are then integrated through an innovative channel fusion process. Finally, an ACMix-based discriminative module is employed to determine image similarity with high precision. Comprehensive experiments on four diverse language signature demonstrate the superior efficacy and promising potential of the framework, affirming its advantages over current methodologies.

Keywords: Offline handwritten signature verification, deep learning, channel fusion

1. INTRODUCTION

In contemporary society, signature handwriting verification, as one of the crucial forensic methods, is widely applied in various fields such as law, insurance, and culture [15,20,10,19,41]. Due to the uniqueness, stability, and reliability of signature handwriting, it serves as an important basis for authenticating documents and confirming identities. However, with the continuous advancement of technology, signature handwriting examination also faces numerous challenges. The origin of signature handwriting can be traced back to ancient times when people used various symbols and graphics to sign. With the development of paper and ink, people began to use handwritten signatures. As early as 439 AD, the Roman Empire used signatures to verify the authenticity of documents. However, it was not until the early 20th century that signature handwriting began to attract research attention. During this period, disciplines such as psychology and statistics began to be applied to the study of signature handwriting, providing a theoretical basis for signature handwriting examination.

Signature handwriting plays an important role in multiple fields. In the legal field, signature handwriting is a crucial basis for confirming the authenticity of documents and is also part of the evidence in court. In the insurance field, signature handwriting is used to identify the authenticity of policies and prevent insurance fraud. In the cultural field, signature handwriting reflects the artist's style and personality, holding significant value for in-depth research in graphology. With technological progress, signature handwriting examination faces many challenges. Signature handwriting is susceptible to factors such as writing habits, emotions, and environment, making the accuracy of handwriting examination complex. Furthermore, the development of signature forgery techniques also brings certain difficulties to the examination work. Considering that the authenticity of most current documents still relies on handwritten signatures for verification, and the cost of manual judgment is too high, there is an urgent need to develop an accurate and efficient signature verification technology.

Signature verification technologies are divided into online signature verification technology and offline signature verification technology based on the input method. For online signature verification, researchers can obtain dynamic information about the signing process, such as stroke trajectories, inclination, and pen pressure [16,31,8,9]. In offline signature verification technology, researchers can only obtain static information, which is signature images captured by scanners or cameras [34,17,1,42]. Because static information provides less information than dynamic information, offline signature verification is more challenging than online signature verification. In today's environment, where paper documents are widely used, offline signature verification has a more widespread application space. Signature verification technology is also divided into writer-dependent and writer-independent methods based on whether it is related to the writer. In writer-dependent methods, researchers' test samples depend on training samples, meaning that each signatory in the test set has a certain amount of signature samples in the training set [21,22,2]. In practical applications, it is impractical to collect and train a large number of samples for each user. In writer-independent methods, the users in the training set and the test set are independent of each other [36,32], thus, they are more valuable in practical applications.

Signature forgery methods are classified into three types based on the proficiency of forgery: random forgery, simple forgery, and skilled forgery [11]. Random forgery signatures have no information about the imitated person, so they differ greatly from genuine samples. Simple forgery involves forged samples that do not follow the writing style of the imitated person, having some similarity to genuine samples. Skilled forgery is performed by professionals who analyze the signature characteristics of the imitated person, resulting in highly similar forged signatures. For skilled forged samples, non-professionals generally cannot distinguish them. Therefore, if criminal organizations obtain relevant information about the imitated person and meticulously forge signatures for criminal activities, this will have adverse effects on the original signatory. Furthermore, for the writer themselves, signatures written in different environments can also vary greatly. Therefore, finding the differences between genuine and forged samples will be a challenging task. To facilitate researchers' study of offline signature verification methods, many public offline signature verification datasets are currently available in academia, such as the English CEDAR dataset [28], GPDS dataset [24], the BHSig260 dataset [7] which includes Bengali and Hindi, and the Chinese MSDS dataset [42], ChiSig dataset [37].

Before the rise of deep learning, researchers typically used traditional image processing methods such as feature matching for signature verification. For example, references [6] and [30] developed the first offline and online signature verification systems; reference [12] utilized the stroke directionality of characters for directional decomposition, then performed band decomposition on the sub-images of each direction, using the decomposed sampled signal values as handwriting features, and employed feature matching methods for writer identification; reference [25] performed identity discrimination through multi-channel two-dimensional Gabor filtering and other methods. Nowadays, researchers are continuously exploring new methods for signature handwriting examination, and with the rise of deep learning and related technologies, reference [3] adopted a Siamese network to extract features from two input sample images separately, and then used metric learning methods to determine the similarity distance between the two signatures, selecting a threshold to determine if they were written by the same person. This metric learning method has significant limitations: on one hand, most metric learning methods use Euclidean distance for calculation, and Euclidean distance only considers the absolute distance between two points, easily overlooking direction and scaling information, and not considering the correlation between data, thus ignoring the relationships between values within feature vectors; on the other hand, its metric threshold is solved through an iterative process, which, although it can obtain the optimal solution for the current dataset, has low generalization ability, and the same threshold will have completely different effects on different datasets. Therefore, reference [4] proposed DeepHSV to address this drawback, using a two-channel discriminative method for offline handwriting verification. By image fusion, two images to be compared are fused into a single image for model input, which can effectively solve the limitations of metric learning. However, they directly fuse the images before model input, at which point the features of the two compared images are not yet very distinct, leading to the loss of a large number of fine features between different images, thus making it impossible to distinguish meticulously forged signatures. Reference [33] proposed an inverse discriminative octuple attention mechanism, where inverse discriminative images are attached as attention to the original images, making the model focus more on stroke features, and achieved good results on multiple datasets. The limitation of this method is that it focuses too much on the features of the original image and only uses inverse discriminative features as auxiliary judgment information. This paper believes that handwriting features can be obtained not only from the original image but also from inverse grayscale images, which contain a large amount of image features.

Offline handwriting signature verification technology can be regarded as a binary classification task, but it differs significantly from traditional image classification. The differences are: 1) The similarity between the two input images in a handwriting verification system is much higher than in other fields, and the detailed differences between the two images are too sparse; 2) The images are grayscale single-channel images; 3) The essence of handwriting verification system discrimination is style comparison, and improper design can easily lead to overfitting. To address these issues, different scholars have proposed different solutions, such as IDN [33], TransOSV [18], LGR [23], etc. The above methods use CNN or self-attention [33–34] techniques, which are generally classified into different types. However, ACMix proposed in reference [13] proves that the two methods have a strong potential relationship. This paper uses it as the discrimination module of the model, which will make the model focus more on the sparse information features of the fused images to achieve higher discrimination accuracy.

This paper addresses the limitations of two-channel discriminative methods by designing a Multi-channel Feature Fusion Network framework. It employs dual inverse discriminative attention for feature extraction and enhancement of original and inverse grayscale images, integrates the extracted multi-dimensional vectors through channel fusion, and finally uses ACMix for image similarity judgment. This network model has achieved good results on four datasets: CEDAR, BHSig-B, BHSig-H, and ChiSig, demonstrating the effectiveness and generality of the proposed method.

The main contributions of this paper are as follows: 1) Proposed the framework, which enhances the differences between genuine and forged images by fusing multi-Siamese networks to extract multi-dimensional detailed features of input images; 2) Improved the inverse discriminative attention module, strengthening the ability to extract signature features through a dual inverse discriminative attention mechanism; 3) Conducted experiments on CEADR, BHSig-B, BHSig-H, and ChiSig datasets, achieving excellent results superior to baseline papers and most existing methods.

1.1 Siamese Network

Deep metric learning methods primarily involve two samples passing through the same network to generate sample vectors, after which the distance between these two samples is calculated to determine if they belong to the same class. This network is known as a Siamese network. Siamese networks, also called twin networks, are a special neural network structure that can input two images for feature extraction, with the two models sharing weights. In 1993, Siamese networks were first proposed for signature recognition on American checks [18].

Due to their simple structure and ease of implementation, Siamese networks are widely used in image similarity measurement. After passing through the same feature extractor, the extracted features have strong image representativeness. Generally, this network is often used to handle verification problems where the two inputs do not differ significantly. The network takes a pair of samples as input and is trained to make samples with the same label closer in the feature space, and samples with different labels further apart. Therefore, this network has promoted the development of offline signature verification. For example: SigNet proposed in reference [37], MSDN proposed in reference [23], TransOSV proposed in reference [12], etc. The basic network framework of a Siamese network is shown in Figure 1: where A and B are the two input samples, Network 1 and Network 2 are feature extraction networks, and the two networks share parameters. After inputting images, feature vectors a and b are generated through the feature extraction network, and the metric distance between samples a and b is calculated using a metric function. Finally, the network parameters are optimized using a contrastive loss function or other loss functions.

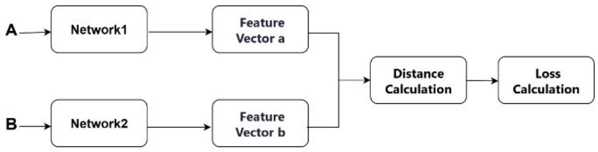

Figure 1: Network Structure Diagram

By generating two feature vectors through a Siamese network, and adhering to the principle that images of the same category are closer and different images are further apart, an optimal threshold can be found by traversing the range between the minimum and maximum distances. However, the method of traversing to find the optimal threshold has a significant limitation: the current threshold is obtained by traversing the current training and test sets, resulting in very low algorithm scalability. Moreover, using Euclidean distance to judge the similarity of different images, while Euclidean distance only considers the absolute distance between two points, easily overlooks information about direction and scaling, and does not consider the correlation between data, thus ignoring the relationships between values within feature vectors. Therefore, a new method is needed to solve this problem.

1.2 Two-Channel Discriminative Network

Another mainstream offline signature verification method is the two-channel discriminative method, which fuses two images and directly outputs 0/1 to determine if they belong to the same class. The biggest difference between this method and the Siamese network is that the Siamese network generates vectors from two samples through the same network structure and then makes a judgment, while the two-channel discriminative network fuses the two images into a single two-channel image before inputting them into the network, and then inputs this single image into a monolithic network to obtain the result of whether they are of the same class. In a two-channel discriminative network, the network does not explicitly extract the input features, but measures their distance in the first step. This design greatly reduces the search parameter space, making two-channel networks particularly suitable for signature verification. The image similarity calculation method based on two channels was proposed by reference [38], and since its proposal, it has achieved considerable results in the field of offline signature verification. For example, reference [6] used two-channel fusion and dual logit output as supervision conditions for training in offline signature verification. Reference [12] proposed an offline signature framework based on two channels and dual Transformers, etc. The basic network diagram of a two-channel discriminative network is shown in Figure 2: where A and B are the two input samples, and the network model is a feature extraction network. After inputting two images, they are first fused into a new image C through image preprocessing before entering the monolithic network. Then, C is input into the monolithic network, and the network output directly indicates whether they were written by the same person, i.e., 0 or 1.

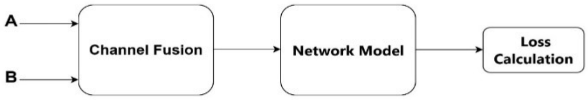

Figure 2: Structure of 2-channel network

Essentially, the two-channel discriminative method treats image similarity judgment as a binary classification method. Through the two-channel discriminative network, the calculation of similarity distance is performed in the first step of the network, and the network directly outputs whether the signatures were written by the same person. Compared to the Siamese network method, this method significantly reduces the search parameter space, effectively speeding up network training; on the other hand, the method of directly outputting results avoids the limitations of the Siamese network's threshold, and the accuracy will not be significantly affected when changing training datasets or adding data. Current networks for two-channel discriminative methods directly perform fusion on original images or after image cropping, i.e., measuring the distance on the initial two images. At this point, the image features are not yet obvious, and simply fusing them will result in the loss of a large number of fine features, ultimately leading to poor model performance.

1.3 Grayscale Processing

In offline signature verification, this paper inputs two single-channel images. Reference [14] also attempted to train with three-channel color images, but the effect was not as good as grayscale images. In grayscale images, different grayscale distributions will have a significant impact on the model's results. For example, black text on a white background and white text on a black background, different inputs will have a significant impact on the training of the same model. This is because in signature verification images, the data model only needs the feature information of the handwriting strokes, and most background information is invalid or even harmful. If the background information consists of pixels with a value of 0, the result after convolution will not change, which has a considerable impact on feature extraction and even the model's output. However, this does not mean that white-on-black images are all invalid information; they also contain detailed and important information. Addressing this issue, reference [30] proposed an inverse discriminative network, where the network input is a black-on-white image. This network enhances the effective information for signature verification through grayscale processing and a multi-path attention module. The attention module of this method extracts features from inverse grayscale images and creates an attention module loaded onto the original grayscale images, making the model focus more on the stroke information of the image. This method innovatively extracts features from both black-on-white and their inverse grayscale images. However, the focus of feature extraction in this method is on the original grayscale image, neglecting that its inverse grayscale image is not only a tool for auxiliary attention but also contains a large amount of handwriting stroke information.

1.4 ACMix

Convolutional kernels and self-attention are two powerful techniques for representation learning, and there is a strong potential relationship between them because most of the computations in these two paradigms are actually performed through the same operations. Specifically, a convolution with kernel size k × k can be divided into k2 individual 1 × 1 convolutions, followed by shifting and summing operations. In ACMix, 1 × 1 convolutional kernels are first used to project input features into queries, keys, and values, and then the attention weights and the aggregation of value matrices, i.e., the aggregation of local features, are calculated. Therefore, ACMix can elegantly integrate these two seemingly different paradigms, enjoying the benefits of both self-attention and convolution, while having smaller overhead compared to pure convolution or self-attention [33].

This paper proposes a network structure to address the feature loss problem in two-channel discriminative networks. It employs a quadruple Siamese network and a dual inverse discriminative attention mechanism for feature extraction, fuses the extracted multiple subtle features through channel fusion, and finally uses a combination of self-attention and convolutional networks to determine whether it is a genuine sample pair.

2. METHODOLOGY

As an end-to-end signature verification system, it consists of feature extraction, channel fusion, and an ACMix module. A pair of signature images first undergo inverse grayscale acquisition, generating a total of four images, which are then input into a quadruple Siamese network. Then, the grayscale and inverse grayscale images of the same image are weighted and calculated through a dual inverse discriminative attention module to extract a large number of detailed features. Finally, the extracted different image representations are fused through channel fusion, and a combination of convolutional neural networks and self-attention is used for discriminative processing to achieve high-similarity image discrimination.

2.1 Dual Inverse Discriminative Attention Module

The feature extractor of this network adopts a quadruple Siamese network structure. This network consists of two convolutional blocks, each containing two convolutional layers activated by ReLU function. Each convolutional layer has a size of 3 × 3, stride of 1, and padding of 1. The dimension of each convolutional block is 64, 128. The reference image and its inverse grayscale image, and the test image and its inverse grayscale image are respectively input into the feature extraction network, and the networks share weights. Between the grayscale image convolutional block and the corresponding inverse grayscale image convolutional block, four dual stroke attention modules are connected. Each attention module connects the convolutional module in the discriminative flow and the convolutional module in the inverse flow, as shown in Figure 3.

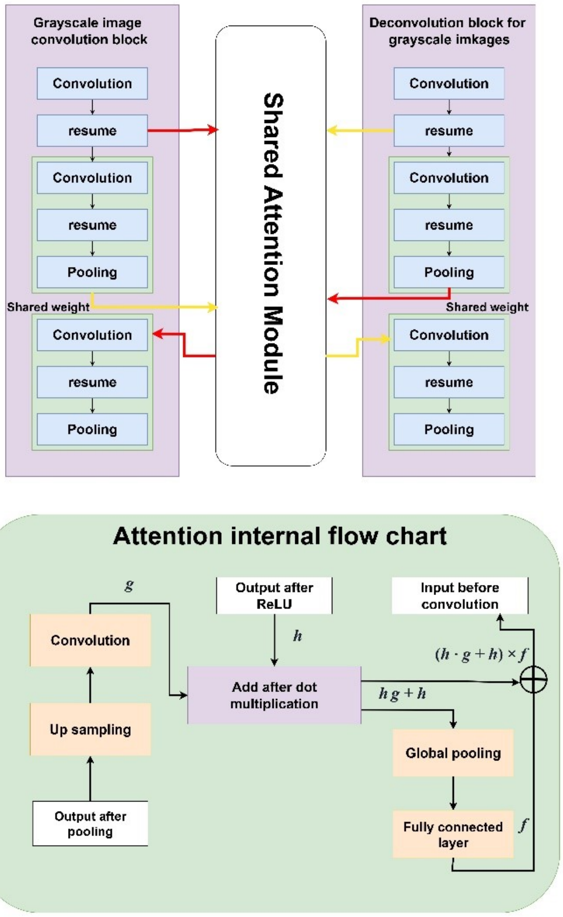

Figure 3: Dual reverse forensic attention module

The grayscale image convolutional block and the inverse grayscale image convolutional block of the same image are simultaneously input into the dual inverse discriminative attention module. Among them, according to the different moments of output from the grayscale image convolutional block and the inverse grayscale image convolutional block, two data streams enter the shared attention. The internal flowchart of the attention module is shown in the right part of Figure 3, taking the yellow data stream in the left part as an example: The feature vector output from the grayscale image convolutional block is input into the upsampling structure, which uses the nearest neighbor algorithm for upsampling and performs convolution operations with Sigmoid activation, outputting g. The inverse grayscale image convolutional block outputs h after ReLU. h is multiplied by the elements of g, and then h is added to produce the intermediate attention measurement h · g + h, where "·" denotes element-wise multiplication. The subsequent Global Average Pooling (GAP) layer and a fully connected layer (FC) with Sigmoid activation receive the intermediate attention measurement and output a weight vector f. Each channel is multiplied by each element of f to generate the final attention (h · g + h) × f, which is then output to the second layer of the inverse grayscale image convolutional block for convolutional processing. This method has two data streams, red and yellow, depending on the input and output, and the shared attention module parameters are shared between them. The dashed boxes in the two convolutional blocks also share weights, and the two convolutional blocks also share weights. Assuming the ReLU output of the grayscale image is x1, and the output of the grayscale image convolutional block is y1; the ReLU output of the inverse grayscale image is x2, and the output of the inverse grayscale image convolutional block is y2; the shared convolutional blocks in the dashed part are collectively referred to as w. Therefore, different data streams have different formulas, specifically as shown in equations (1) to (4):

$out1 = h(x1\ )\ g[w(x2\ )] + h(x1\ )$ (1) $y1 = w[out1 \times f\,(out1\ )]$ (2) $out2 = h(x2\ ) \cdot g[w(x1\ )] + h(x2\ )$ (3) $y2 = w[out2 \times f\,(out2\ )]$ (4)

Among them, equations (1) and (2) are for the red data stream, and equations (3) and (4) are for the yellow data stream.

In the attention module, this paper processes grayscale images and inverse grayscale images separately for attention, forming dual inverse discriminative attention.. By comparing with IDN's attention, it is found that although IDN has quadruple attention, as the convolution operation deepens, the focus of the attention module becomes more abstract. In contrast, the framework's attention module, on one hand, introduces dual features, creating constraints between attentions, enabling accurate focus on stroke edge information even in the second layer of attention feature maps; on the other hand, by reducing the number of layers, it extracts sufficiently detailed features during channel fusion. Through the multi-path attention mechanism, the important features for signature verification are enhanced.

Since the attention module in this paper connects the original grayscale image and the inverse grayscale image, the final attention mask will guide the network to learn discriminative features for signature verification and suppress misleading information. The entire framework has 4 attention modules connecting different convolutional modules, applying the attention mechanism at different scales and resolutions.

2.2 Multi-channel Fusion

The two-channel discriminative method fuses two input single-channel images into a two-channel image and directly outputs whether they are similar to quickly obtain results. However, because the two-channel discriminative network performs channel fusion in the original state of the images, the image features are not yet obvious, and simply fusing them will result in the loss of a large number of fine features, ultimately leading to poor model performance. Therefore, in the framework, four images, each with 128-dimensional feature information after feature extraction, are fused, totaling 512-dimensional feature information. Compared to traditional two-channel discriminative networks, our framework not only includes the convolutional features of the two images to be discriminated but also includes the convolutional features of the inverse grayscale images of the two images. This method considers more channel information during the fusion process, allowing the network to capture more information and increasing the diversity of fused features. Therefore, in the subsequent discrimination stage, this network can achieve higher accuracy.

The proposed framework connects the extracted features of the reference image, reference image inverse grayscale image, test image, and test image inverse grayscale image. Each image has 128 dimensions. After passing through the discriminative module, it finally outputs 0/1 to determine whether they are the same person. Compared to the two-channel discriminative method, the multi-channel image formed by multi-channel fusion has more fused image features. The distance calculation between images changes from one positive and one negative sample, totaling two images, to 256 dimensions for each positive and negative sample to calculate the difference. Because there are more image features, the calculated distance is more accurate.

2.3 Discriminative Module

In the discriminative module, this paper primarily uses the ACMix module, supplemented by two small convolutional blocks for discrimination. After feature fusion, a 512-dimensional feature vector is obtained, with many and large differences between features. To accurately extract features, the ACMix module, which combines convolution and self-attention, is used for feature extraction.

The 512-dimensional feature representation after channel fusion is not directly input into a fully connected layer for classification. Instead, it first passes through a discriminative module based on a monolithic network model [33]. This module performs overall feature learning and judgment based on self-attention and convolutional networks, finally outputting a 0/1 binary classification result. The structure of the discriminative module consists of two small convolutional modules and one ACMix module. The input of the first small convolutional module is the integrated 512-dimensional features after channel fusion. Then, the ACMix's convolution and self-attention mechanisms are used for feature extraction. Finally, a small convolutional module is used for feature summarization, and then it enters a multi-layer perceptron for classification. At this point, the features entering the multi-layer perceptron are the 512-dimensional image features extracted by the discriminative network, which contain the overall difference information between the reference image, reference image inverse grayscale image, test image, and test image inverse grayscale image.

Global average pooling is introduced in the multi-layer perceptron to reduce network redundancy. To avoid overfitting, this paper uses 0.5 Dropout. Finally, the entire network will output a Sigmoid-activated feature value, generating a judgment probability between 0 and 1. In the accuracy judgment process, this paper sets a probability less than or equal to 0.5 as a forged signature, and a probability greater than 0.5 as a genuine signature. The loss function uses binary cross-entropy loss, and its formula is:

$$L = -(1/n) \sum[yi \lg(pi) + (1 - yi)\lg(1 - pi)] \quad (5)$$

Where yi represents the true label of sample i, 1 for positive class, 0 for negative class. pi represents the probability that sample i is predicted as positive, and similarly, 1 − pi is the probability that the sample is predicted as negative.

3. EXPERIMENT

The quantity and quality of datasets have a significant impact on the model. Currently, with the in-depth research of domestic and foreign scholars in the field of offline handwriting verification, many public offline datasets have been proposed. This paper will use the English CEDAR dataset, the BHSig260 dataset (including Bengali and Hindi), and the Chinese ChiSig dataset for model testing and evaluation. Statistical information for various datasets is shown in Table 1.

The CEDAR dataset is a signature sample dataset in English. It consists of samples from 55 signers, with each signer having 24 genuine signature samples and 24 forged signature samples. According to previous work, this paper selects samples from 50 individuals for training and samples from the remaining 5 signers for testing. For each signer, this dataset has 276 reference-genuine sample pairs and 576 reference-forged sample pairs.

Table 1: Offline signature verification dataset

Data set name	Language	Signature type	Number of pictures	Real to fake sample ratio
CEDAR	English	55	2624	24/24
BHSig-B	Bengali	100	5400	24/30
BHSig-H	Hindi	160	8640	24/30
ChiSig	Chinese	102	10242	-/-

To ensure a balance of positive and negative samples, this paper will randomly draw reference-forged sample pairs based on the number of reference-genuine sample pairs. Therefore, for each signer, this paper will have 276 reference-genuine sample pairs and 276 reference-forged sample pairs for training and testing.

The BHSig260 dataset includes Bengali and Hindi datasets, which are treated as two different datasets in this paper. The BHSig-B dataset contains Bengali signature images from 100 signers. Each signer has 24 genuine signatures and 30 forged signatures. Based on previous experience, this paper randomly selects signatures from 50 signers for training, and signatures from the remaining signers for testing. The BHSig-H dataset contains Hindi signature images from 160 signers. Each signer has 24 genuine signatures and 30 forged signatures. Similarly, this paper will randomly select signatures from 100 signers as the training set to train the model, and signatures from the remaining 60 signers as test data. For each signer in the above two datasets, this paper also randomly draws 276 reference-genuine sample pairs and 276 reference-forged sample pairs for training and testing.

Reference [11] constructed a novel Chinese document offline signature forgery detection benchmark dataset, ChiSig, which includes all tasks such as signature detection, restoration, and verification. The dataset consists of clean handwritten signatures, synthetically interfered handwritten signatures, and synthetic documents with handwritten signatures. The authors randomly generated 500 names and then asked volunteers to sign according to certain rules to obtain clean signature data, which can be used for signature verification tasks. Because the number of volunteers is greater than the number of names, there are cases where different writers have the same name, which poses a great challenge for signature verification. Afterwards, the authors obtained scanned documents that can be used as synthetic backgrounds from public resources such as the XFUND dataset, Chinese national standards, and patents. For this dataset, this paper randomly draws 250 signatures as the training set and 250 signatures as the test set. For each name, signatures written by the same volunteer are treated as genuine sample pairs, and signatures written by different volunteers are treated as forged sample pairs. For dedicated forged data, they are only treated as forged sample pairs, and forged data are not treated as genuine sample pairs. To ensure data balance between genuine and forged sample pairs, this paper removes redundant sample pairs.

3.2 Evaluation Metrics

For the CEDAR and BHSig260 datasets, this paper will follow the settings in reference [30] and use False Rejection Rate (FRR), False Acceptance Rate (FAR), and Accuracy (ACC) to comprehensively evaluate the framework and compare it with other existing methods.

FRR is defined as the ratio of the number of false rejections to the number of genuine samples. FAR is defined as the ratio of the number of false acceptances to the number of forged samples. ACC is defined as the ratio of the number of correctly judged samples to the total number of samples.

For the ChiSig dataset, this paper uses the evaluation metrics proposed by the dataset authors: Accuracy, Equal Error Rate (EER), and True Acceptance Rate (TAR) for comparison. EER evaluates the balance point where FRR equals FAR; the lower the EER, the better the model performance. The calculation method for TAR is shown in equations (6) to (8), and TAR is only calculated when the False Acceptance Rate (FAR) equals 10−3:

FAR = (Number of False Acceptances) / (Number of Forgeries) (6) FRR = (Number of False Rejections) / (Number of Genuine Samples) (7) TAR = 1 − FRR (8)

3.3 Comparative Experiments

To verify the model’s effectiveness, this paper selects the latest deep learning models for comparison based on the current development of handwriting verification tasks, namely SigNet (2017arXiv) [37], IDN (2019CVPR) [30], DeepHSV (2019ICDAR) [6], SDINet (2021AAAI) [13], SURDS (2022ICPR) [39], 2C2S (2023EAAI) [40], TransOSV (2022ICME) [12]. These models include methods combining Siamese networks with metric learning, as well as methods using two-channel discrimination. The comparison results are sufficient to illustrate the advantages of the proposed model proposed in this paper. For convenience of observation, the optimal solution is bolded, the suboptimal solution is underlined, and the third best solution is wavy. CEDAR, BHSig-B, and BHSig-H are shown in Table 2, Table 3, and Table 4, respectively. The results for the ChiSig dataset will be introduced in Section 3.4.

In the experimental results on the CEDAR dataset, the proposed model achieved 100% accuracy. The main reason is that this dataset has a small number of samples, a simple structure, and large differences, so many methods have achieved good results on this dataset. Comprehensive analysis shows that model’s ACC improved by 3.62% and 1.75% compared to IDN and SDI, respectively, and achieved 100% like SigNet, DeepHSV, and 2C2S. In the BHSig-B dataset, the experimental results show that the model has a greater advantage than current mainstream offline handwriting verification algorithms, achieving an accuracy of 95.61%, and this is also proven in the comparison of FRR and FAR, reaching optimal or suboptimal. Compared to IDN, model’s ACC improved by 0.29%. Compared to the latest algorithms 2C2S and TransOSV, it improved by 2.36% and 5.56%, respectively. This is sufficient to prove the superiority of the model proposed in this paper.

Table 2: Comparison on CEDAR dataset (%)

Model name	FRR	FAR	ACC
SigNet (2017arXiv)	0	0	100.00
DeepHSV (2019ICDAR)	-	-	100
IDN (2019CVPR)	2.17	5.87	96.38
SDINet (2021AAAI)	3.42	0.73	98.25
2C2S (2023EAAI)	0	0	100.00
OURS	0	0	100.00

Table 3 Comparison on BHSig-B dataset (%)

Model Name	FRR	FAR	ACC
SigNet (2017arXiv)	13.89	13.89	86.11
DeepHSV (2019ICDAR)	—	—	88.08
IDN (2019CVPR)	5.24	4.12	95.32
SDINet (2021AAAI)	7.86	3.30	94.42
SURDS (2022ICPR)	5.42	19.89	87.34
2C2S (2023EAAI)	8.11	5.37	93.25
TransOSV (2022ICME)	9.95	9.95	90.05
OURS	3.86	3.84	95.61

Table 4 Comparison on BHSig-H dataset (%)

Model Name	FRR	FAR	ACC
SigNet (2017arXiv)	15.36	15.36	84.64
DeepHSV (2019ICDAR)	—	—	86.66
IDN (2019CVPR)	4.93	8.99	93.04
SDINet (2021AAAI)	3.77	6.24	95.00
SURDS (2022ICPR)	8.98	12.01	89.50
2C2S (2023EAAI)	9.98	8.66	90.68
TransOSV (2022ICME)	3.39	3.39	96.61
OURS	4.89	4.89	95.70

Similar to CEDAR and BHSig-B, our model also achieved good results on the BHSig-H dataset. Compared to the latest algorithms, model achieved an accuracy of 95.7% on the BHSig-H dataset, although it is not the optimal result, its FRR is third best, and the others are suboptimal. Furthermore, compared to the optimal, model's accuracy is only 0.89% lower, while in BHSig-B, compared to the optimal model TransOSV in BHSig-H, our model achieved a 5.56% lead in accuracy. This is sufficient to show that model's generalization ability is superior to TransOSV.

3.4 Ablation Experiment

In addition, this paper conducted ablation experiments on the ChiSig dataset. InceptionResnet is the baseline model provided in the dataset paper [11]. This paper conducted comparative experiments by reproducing SigNet and IDN code.

As shown in Table 5, the baseline IDN compared with its channel fusion method, the channel fusion method improved the accuracy by 0.9% compared to the original method; the dual inverse discriminative attention expanded the information of the inverse grayscale image, providing more detailed information during channel fusion, which improved the accuracy to 88.96%, an increase of 3.24% compared to channel fusion. The ACMix discriminative structure further improved the model's accuracy to 95.23%.

Table 5 Ablation experiment on ChiSig dataset (%)

Model Name	EER	TAR	ACC
InceptionResnet	6.60	28.10	93.60
SigNet	—	—	82.28
IDN (Baseline)	17.91	10.50	84.82
IDN (Channel Fusion)	14.81	9.61	85.72
IDN (Channel Fusion + Attention)	11.38	7.82	88.96
OURS (No Inverse Gray, No Attention)	11.78	32.49	88.09
OURS (No Inverse Gray, Single Attention)	10.83	—	89.20
OURS (Inverse Gray, No Attention)	7.84	—	92.14
OURS (Full Model)	5.19	28.96	95.23

To demonstrate the impact of inverse grayscale images and corresponding attention on the results, this paper also conducted experiments by removing grayscale images and attention. 'No inverse grayscale image' means the model only inputs reference images and test images. 'Single attention' means that in the dual attention module, the input for dot product and upsampling is provided by itself, and everything else is consistent with the final model.

Table 6 Main parameters on ChiSig dataset (%)

Model Name	FRR	FAR	ACC	Notes
IDN (Baseline)	10.46	17.91	84.82	Original implementation
IDN (Channel Fusion)	9.61	18.97	85.72	+Feature combination
IDN (Channel Fusion + Attention)	7.82	14.27	88.96	+Attention mechanism
OURS (No Grayscale Inversion, No Attention)	21.91	17.26	88.09	Basic version
OURS (No Grayscale Inversion, Single Attention)	15.59	16.30	89.20	+Attention layer
OURS (Grayscale Inversion, No Attention)	6.90	17.18	92.14	+Image preprocessing
OURS (Full Model)	5.34	5.34	95.23	Complete configuration

For no inverse grayscale image, after introducing single attention, the accuracy increased by 1.11%, while introducing inverse grayscale images increased the accuracy by 4.05%. Experimental results show that the addition of attention and inverse grayscale images is feasible, and the addition of inverse grayscale images has a greater improvement effect than the addition of attention.

This ablation experiment proves the rationality of the proposed method. In addition, to facilitate future researchers to compare using FRR and FAR metrics, this paper also calculated the FRR and FAR metrics of our proposed model on the ChiSig dataset, as shown in Table 6.

3.5 Cross-Language Experiment

Furthermore, this paper also conducted cross-language tests. In this work, CEDAR, BHSig-B, BHSig-H, and ChiSig, four different languages, were used for testing. This paper trained the model using the training set of one language and tested it on the training sets of the remaining languages. For example, this paper trained the model on the BHSig-B training dataset and tested it on the BHSig-H test dataset. The division of training and test data is the same as in the experiments on each independent dataset. Table 7 shows the accuracy of cross-language tests, where rows correspond to training languages and columns correspond to test languages.

Table 7 shows that cross-language signature verification performance rapidly declines. This paper believes that the essence of an offline signature verification system is style feature matching.

Each person's signature is closely related to their writing style habits, and different language styles have different writing habits, leading to the inability of the current dataset's learned style to be applied to other datasets. The accuracy of the BHSig-B dataset and BHSig-H dataset is higher than other datasets, possibly because the writing styles of Hindi and Bengali are more similar.

Table 7 Cross-language test (%)

Training Set → Test Set	CEDAR	BHSig-B	BHSig-H	ChiSig
CEDAR	100.00	48.76	49.89	57.48
BHSig-B	64.86	95.61	82.79	63.71
BHSig-H	50.11	86.27	95.70	20.00
ChiSig	54.60	70.02	55.37	95.23

4. CONCLUSION

This paper proposes a novel offline handwriting verification model, for handwritten signature verification in writer-independent scenarios. This model first extracts features through two layers of convolutional networks and a dual attention module, then performs feature fusion through channel fusion, and finally uses the ACMix discriminative module to determine the similarity of multiple images. It uses an inverse supervision mechanism and a dual attention mechanism to solve the problem of insufficient detailed feature information in traditional channel fusion methods. In testing, by inputting reference signature images and test signature images, the model directly outputs whether the test signature is genuine or forged. Experimental results demonstrate the advantages and potential of the proposed method. Future work will focus on research into cross-language signature verification and recognition.

REFERENCES

1. Soleimani A, Araabi B N, Fouladi K. Deep multitask metric learning for offline signature

verification. Pattern Recognition Letters, 2016, 80: 84−90

2. Ferrer M A, Alonso J B, Travieso C M. Offline geometric parameters for automatic signature verification using fixed-point arithmetic. IEEE Transactions on Pattern Analysis And Machine Intelligence, 2005, 27(6): 993−997
3. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez A N, et al. Attention is all you need. In: Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach, USA: Curran Associates Inc., 2017. 6000−6010
4. Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X H, Unterthiner T, et al. An image is worth 16×16 words: Transformers for image recognition at scale. In: Proceedings of the 9th International Conference on Learning Representations. Austria: OpenReview.net, 2021.
5. A Transformer Based Handwriting Recognition System Jointly Using. arXiv, 2025.
6. Li C, Lin F, Wang Z Y, Yu G, Yuan L, Wang H Q. DeepHSV: User-independent offline signature verification using two-channel CNN. In: Proceedings of the International Conference on Document Analysis and Recognition (ICDAR). Sydney, Australia: IEEE, 2019. 166−171
7. Liu Cheng-Lin, Liu Ying-Jian, Dai Ru-Wei. Writer identification by multichannel decomposition and matching. Acta Automatica Sinica, 1997, 23(1): 56−63
8. Hafemann L G, Oliveira L S, Sabourin R. Fixed-sized representation learning from offline handwritten signatures of different sizes. International Journal on Document Analysis and Recognition (IJDAR), 2018, 21(3): 219−232
9. Hafemann L G, Sabourin R, Oliveira L S. Learning features for offline handwritten signature verification using deep convolutional neural networks. Pattern Recognition, 2017, 70: 163−176
10. Xia X H, Song X Y, Luan F G, Zheng J G, Chen Z L, Ma X F. Discriminative feature selection for on-line signature verification. Pattern Recognition, 2018, 74: 422−433
11. Yan K H, Zhang Y, Tang H R, Ren C K, Zhang J, Wang G A, et al. Signature detection, restoration, and verification: A novel Chinese document signature forgery detection benchmark. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). New Orleans, USA: IEEE, 2022. 5163−5172
12. Li H, Wei P, Ma Z Y, Li C K, Zheng N N. Offline signature verification with transformers. In: Proceedings of the IEEE International Conference on Multimedia and Expo (ICME). Taipei, China: IEEE, 2022. 1−6
13. Li H, Wei P, Hu P. Static-dynamic interaction networks for offline signature verification. In: Proceedings of the 35th AAAI Conference on Artificial Intelligence. Vancouver, Canada: AAAI Press, 2021. 1893−1901
14. Bhattacharya I, Ghosh P, Biswas S. Offline signature verification using pixel matching technique. Procedia Technology, 2013, 10: 970−977
15. Bromley J, Bentz J W, Bottou L L, Guyon I, Lecun Y, Moore C, et al. Signature verification using a “Siamese” time delay neural network. International Journal of Pattern Recognition and Artificial Intelligence, 1993, 7(4): 669−688
16. Hu J, Chen Y B. Offline signature verification using real adaboost classifier combination of Pseudo-dynamic features. In: Proceedings of the 12th International Conference on Document Analysis and Recognition. Washington, USA: IEEE, 2013. 1345−1349
17. Xing Z J, Yin F, Wu Y C, Liu C L. Offline signature verification using convolution Siamese network. In: Proceedings of SPIE 10615, 9th International Conference on Graphic and Image Processing (ICGIP). Qingdao, China: SPIE, 2017. 415−423
18. Bromley J, Guyon I, LeCun Y, Säckinger E, Shah R. Signature verification using a “Siamese” time delay neural network. In: Proceedings of the 6th International Conference on Neural Information Processing Systems. Denver, Colorado: Morgan Kaufmann Publishers Inc., 1993. 737−744
19. Zou Jie, Sun Bao-Lin, Yu Jun. Online handwriting matching algorithm based on stroke features. Acta Automatica Sinica, 2016, 42(11): 1744−1757
20. Cpałka K, Zalasiński M, Rutkowski L. New method for the online signature verification based on horizontal partitioning. Pattern Recognition, 2014, 47(8): 2652−2661
21. Jain A K, Ross A, Prabhakar S. An introduction to biometric recognition. IEEE Transactions on Circuits and Systems for Video Technology, 2004, 14(1): 4−20
22. Kalera M K, Srihari S, Xu A H. Offline signature verification and identification using distance statistics. International Journal of Pattern Recognition and Artificial Intelligence, 2004, 18(7): 1339−1360
23. Liu L, Huang L L, Yin F, Chen Y B. Offline signature verification using a region based deep metric learning network. Pattern Recognition, 2021, 118: Article No. 108009
24. Herbst N M, Liu C N. Automatic signature verification based on accelerometry. IBM Journal of Research and Development, 1977, 21(3): 245−253
25. Cairang X M, Zhaxi D J, Yang X L, Hou Y, Zhao Q J, Gao D G, et al. Learning generalisable representations for offline signature verification. In: Proceedings of the International Joint

Conference on Neural Networks (IJCNN). Padua, Italy: IEEE, 2022. 1−7

26. Enhancing Signature Verification Using Triplet Siamese Similarity. MDPI, 2024.
27. Combining Multi-Scale Fusion and Attentional Mechanisms for. MDPI, 2025.
28. Nagel R N, Rosenfeld A. Steps toward handwritten signature verification. In: Proceedings of the 1st International Joint Conference on Pattern Recognition. 1973. 59−66
29. Learning features for offline handwritten signature verification using. Nature, 2025.
30. Wei P, Li H, Hu P. Inverse discriminative networks for handwritten signature verification. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach, USA: IEEE, 2019. 5764−5772
31. Kumar R, Sharma J D, Chanda B. Writer-independent off-line signature verification using surroundedness feature. Pattern Recognition Letters, 2012, 33(3): 301−308
32. Zhang P R, Jiang J J, Liu Y L, Jin L W. MSDS: A large-scale Chinese signature and token digit string dataset for handwriting verification. In: Proceedings of the 36th International Conference on Neural Information Processings Systems. New Orleans, USA: 2022. 36507−36519
33. Pan X R, Ge C J, Lu R, Song S J, Chen G F, Huang Z Y, et al. On the integration of self-attention and convolution. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New Orleans, USA: IEEE, 2022. 815−825
34. Okawa M. Synergy of foreground-background images for feature extraction: Offline signature verification using Fisher vector with fused KAZE features. Pattern Recognition, 2018, 79: 480−489
35. A Survey of Offline Handwriting Signature Verification. ResearchGate, 2025.
36. Pal S, Alaei A, Pal U, Blumenstein M. Performance of an offline signature verification method based on texture features on a large Indic-script signature dataset. In: Proceedings of the 12th IAPR workshop on Document Analysis Systems (DAS). Santorini, Greece: IEEE, 2016. 72−77
37. Dey S, Dutta A, Toledo J I, Ghosh S K, Llados J, Pal U. SigNet: Convolutional Siamese network for writer independent offline signature verification. arXiv preprint arXiv: 1707.02131, 2017.
38. Zagoruyko S, Komodakis N. Learning to compare image patches via convolutional neural networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Boston, USA: IEEE, 2015. 4353−4361
39. Chattopadhyay S, Manna S, Bhattacharya S, Pal U. SURDS: Self-supervised attention-guided reconstruction and dual triplet loss for writer independent offline signature verification. In: Proceedings of the 26th International Conference on Pattern Recognition (ICPR). Montreal, Canada: IEEE, 2022. 1600−1606
40. Ren J X, Xiong Y J, Zhan H J, Huang B. 2C2S: A two-channel and two-stream transformer based framework for offline signature verification. Engineering Applications of Artificial Intelligence, 2023, 118: Article No. 105639
41. Guerbai Y, Chibani Y, Hadjadji B. The effective use of the one-class SVM classifier for handwritten signature verification based on writer-independent parameters. Pattern Recognition, 2015, 48(1): 103−113
42. Zhu Yong, Tan Tie-Niu, Wang Yun-Hong. Writer identification based on texture analysis. Acta Automatica Sinica, 2001, 27(2): 229−234
43. EID, A. A., Miled, A. B., Mahmoud, A. F., Abdalla, F. A., Jabnoun, C., Dhibi, A., ... & Belhaj, S. (2024). Leveraging Arabic Text Embedded in Images: Challenges and Opportunities in NLP Analysis. Journal of Intelligent Systems and Applied Data Science, 2(1).