International Journal on Document Analysis and Recognition (IJDAR) (2024) 27:195–208
https://doi.org/10.1007/s10032-023-00455-6

ORIGINAL PAPER

Attention-based multiple siamese networks with primary representation guiding for offline signature verification

Yu-Jie Xiong[1] · Song-Yang Cheng[1] · Jian-Xin Ren[1] · Yu-Jin Zhang[1]

Received: 20 January 2023 / Revised: 4 September 2023 / Accepted: 11 September 2023 / Published online: 9 October 2023

Abstract
In the area of biometrics and document forensics, handwritten signatures are one of the most commonly accepted symbols. Thus, financial and commercial institutions usually use them to verify the identity of an individual. However, offline signature verification is still a challenging task due to the difficulties in discriminating the minute but significant details between genuine and skilled forged signatures. To tackle this issue, we propose a novel writer-independent offline signature verification approach using attention-based multiple siamese networks with primary representation guiding. The proposed multiple siamese networks regard the reference signature images, query signature images, and their corresponding inverse images as inputs. These images are fed to four weight-shared parallel branches, respectively. We present an efficient and reliable mutual attention module to discover prominent stroke information from both original and inverse branches. In each branch, feature maps of the first convolution are utilized to guide the combination with deeper features, named as primary representation guiding, which guides the model into concerning the shallow stroke information. The four branches are concatenated in an ordered way and are put into four contrastive pairs, which is helpful to obtain useful representations by comparing reference and query samples. Four contrastive pairs generate four preliminary decisions independently. Then, the eventual verification result is created based on the four preliminary decisions using a voting mechanism. In order to assess the performance of the proposed method, extensive experiments on four widely used public datasets are conducted. The experimental results demonstrate that the proposed method outperforms existing approaches in most cases and can be applied to various language scenarios.

Keywords Attention mechanism · Multiple siamese networks · Contrastive pairs · Primary representation guiding · Offline signature verification

1 Introduction

Biometrics is to estimate an individual's identity using various characteristics, which is capable of making our life safer and more convenient. In general, biometrics can be classified into physiological biometrics and behavioral biometrics [1]. An ideal biometric should be universal and unique. Besides, the characteristics should be stable for a long period and be easy to obtain. Physiological biometrics considers fingerprints, iris and facial scans. However, behavioral biometrics concentrates on handwriting, voice and gait of an individual. Despite the differences among these biometrics, they both contribute to building a reliable system for personal authentication. As ones of the oldest accepted symbols to verify a person's identity, handwritten signatures have many advantages. Signatures are closely related to writing habits and are difficult to replicate. In addition, they are generally easy to access. Thus, signatures are extensively used in the field of finance, commerce and military. Signature verification aims to confirm or reject a personal claimed identity by analyzing his/her handwritten signatures. Manual signature verification requires specialized professional training which is time-consuming, and exists uncertainties during the authentication procedure. Powered by the rapid development of pattern recognition and artificial intelligence, automatic signature verification has attracted much attention recently.

Y.-J. Xiong, S.-Y. Cheng: These authors contributed equally to this work.

✉ Yu-Jie Xiong
xiong@sues.edu.cn

1 School of Electronic and Electrical Engineering, Shanghai University of Engineering Science, Shanghai 201620, China

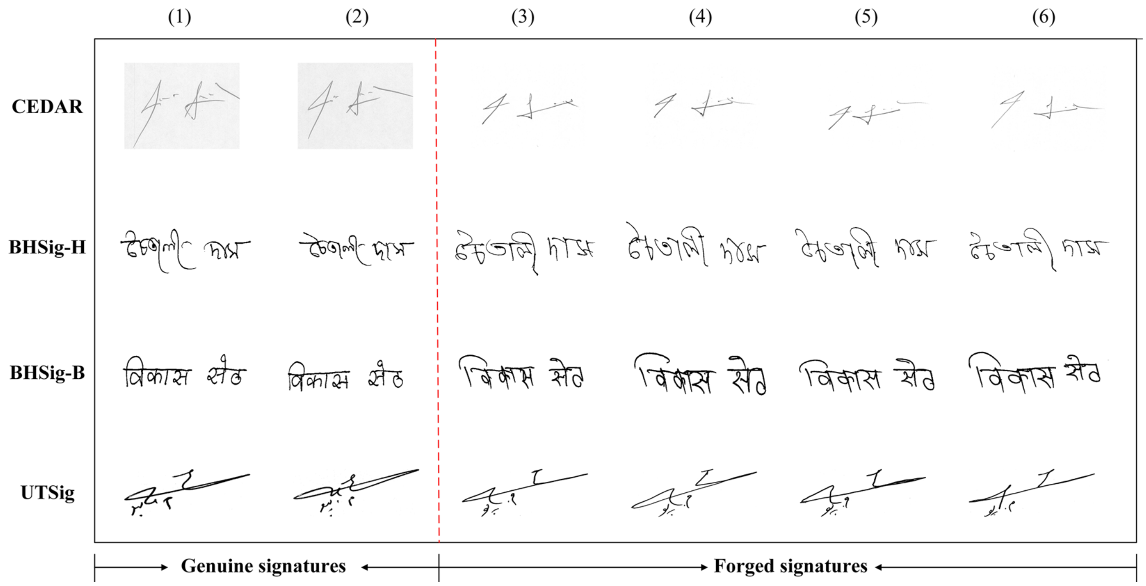


Fig. 1 Samples of CEDAR, BHSig-H, BHSig-B and UTSig dataset. Each row contains six signatures. Columns (1) and (2) are genuine signatures, and columns (3), (4), (5) and (6) are forged signatures

Generally, signature verification systems are divided into writer-dependent (WD) and writer-independent (WI) approaches [2]. For writer-dependent approaches, the model is trained for each user's signatures using some signatures of the user, while writer-independent approaches aim to conduct a uniform model to classify signatures from all users. There are two types of the input for an signature verification system based on data acquirement means, online (dynamic) and offline (static) data. For online data, chronological information such as the velocity, pressure, inclination and direction of pen movement will be recorded. The offline data are obtained with optical scanner devices and represented as static signature images. For realistic scenarios, offline signatures are easy to obtain, and writer independent approaches are capable of training and testing a verification system with a different subset of users. Therefore, writer-independent offline signature verification covers most practical application scenarios and is the most critical and challenging task.

For signature verification systems, handwritten signatures are classified as genuine or forged. The forgeries are the imitations of genuine signatures and are broadly classified into three types: random, simple and skilled forgery [3]. Random forgery is the forging of a signature without knowing the name of the signer and the signature patterns. As for simple forgery, the forger only knows his/her name but has never seen the genuine signatures. Skilled forgery is established by the forger who has seen genuine signatures and practiced the writing pattern repeatedly. Under such circumstances, there are high similarities between genuine and forged signatures. Figure 1 shows examples of signatures from four different widely used datasets: CEDAR [36], BHSig-H, BHSig-B [37] and UTSig [38]. Each row contains two genuine signatures of the same writer and four skilled forgeries. It is noticed that skilled forgeries are very similar to genuine signatures. Furthermore, genuine signatures of the same writer are not totally the same that indicates they exist high intra-class variability.

In practical applications, there are a small number of signatures used for training a verification system. In addition, the writer-dependent system needs to be retrained for every new writer, which is time-consuming. Thus, it is meaningful to construct a unified model which is capable of detecting the characteristic differences between genuine and forged signatures, and confirming the identity of unknown writers without extra training. In this paper, we propose attention-based multiple siamese networks (MSN) with primary representation guiding for offline signature verification. MSN receive the reference and query signatures and their corresponding inverse images. The input original reference and query signatures are grayscale images. These images are referred to four weight-shared branches. Mutual attention modules are proposed to ensure that the networks pay more attention to effective stroke information of signatures by connecting the original and inverse branches. Furthermore, we propose primary representation guiding to connect shallow features with deeper ones to realize feature reuse and avoid spatial crucial information lost problems. The features from four branches are formed into four contrastive pairs, which contribute to learning discriminative features by comparisons of signatures. The contrastive pairs are transmitted to four classifiers which make the preliminary decisions independently. At last, the final decision is generated using a voting mechanism in accordance with the four preliminary decisions.

The main contributions of this paper are as follows:

- We propose multiple siamese networks with primary representation guiding which strengthens feature propagation and encourages feature reuse. The spatial information of signature image is lost caused by the usage of pooling operation, the deployment of multiple siamese networks and primary representation guiding that is vital in improving the flow of stroke information.
- We propose a mutual attention module that considers the information between signatures and their corresponding inverse images in the multiple siamese networks. The attention module comprises spatial and channel attention mechanism, which enables the networks capture and focus prominent stroke information efficiently.
- Contrastive pairs are employed to maximize similarity between original and inverse images. Contrastive pairs ensure the networks extract meaningful representation by comparison of features from both original and inverse images discriminatively.
- Four widely used datasets are used to evaluate our proposed method. Benefiting from the above three strategies, our best EER on CEDAR, BHSig-B, BHSig-H and UTSig datasets reached 1.12%, 8.11%, 10.17% and 18.08%, which outperforms other existing approaches.

The remainder of this paper is organized as follows: Sect. 2 gives a literature review of related researches on signature verification; Sect. 3 describes the details of the proposed method; Sect. 4 provides experimental results and discussions. Finally, Sect. 5 concludes this paper and presents the future work.

2 Related work

Signature verification has attracted great attention in document analysis over the past decades and has played a significant role in numerous areas such as financial, commercial, security and judiciary [4–6]. Despite the remarkable progress, signature verification is still very challenging due to the high intra-class variety and low inter-class variety among signatures from different writers.

In developing an offline signature verification system, there are mainly three steps: pre-processing, feature extraction and classification. The pre-processing step often includes image enhancement and size normalization. During the step of feature extraction, a lot of unique handcrafted features extracted from signature images have been employed to verify a person's identity. The existing employed features can be roughly divided into two categories: functions or parameters [7]. For the classification step, writer-dependent and writer-independent classifiers are normally applied to judge the authenticity of acquired features. Meanwhile, with the emergence of deep neural networks (DNN), several approaches also tried to realize signature verification using DNN. Unlike traditional handcrafted features, DNN learn feature mapping from training data directly and accomplish the classification at the same time.

2.1 Traditional approaches

For traditional approaches, the crucial step of building a robust offline signature verification system is to extract high-quality stroke features. The feature extraction techniques can be mainly divided into function-based and parameter-based techniques. The function-based features consider the temporal information in the signature and are applied for online signature verification, while parameter-based features characterize the signature as a vector of elements [7, 8]. Moreover, parameter-based features can be divided into two categories: global and local parameters. For global parameters, Baltzakis et al. [9] proposed a series of grid and texture features for signature verification. Drouhard et al. [10] utilized the directional probability density function as a global characteristic to describe the signature. A. Dutta et al. [40] proposed a method for offline signature verification. Their approach combines local features and global statistics within signature images using histogram of oriented gradients (HOGs). Zois et al. [11] introduced a feature extraction scheme that relies on the detection of first-order transitions between asymmetrical lattice arrangements of simple pixel structures. In Ref. [12], a parameter-free feature was proposed to explore local and global information of visibility graphs.

Determined by the scale of local details, local parameters can be classified into component-oriented and pixel-based features. For component-oriented features, Oliveira et al. [13] investigated caliber and symmetry of signature as static graphometric features to automated signature verification. Aubin et al. [14] analyzed simple and individual strokes to verify the writer's identity. Rivard et al. [15] utilized gradient directions to obtain the high intra-personal variability of handwritten signatures. Bertolini et al. [16] introduced a new graphometric feature set which considers simulating the most important segments of the signature by using Bezier curves. For pixel-based features, Sabourin et al. [17] took the extended shadow code as a global feature vector for signatures which permits the local projection of the handwriting. Ferrer et al. [18] utilized local binary pattern (LBP) to measure the robustness of gray-level features when the signature background is complex. Solar et al. [19] utilized Scale Invariant Feature Transform (SIFT) features to extract local interest points in offline signature verification systems. Okawa [20] employed Fisher Vector (FV) as encoding of KAZE features from both genuine and forged signature images to extract discriminative features.

2.2 Deep learning approaches

Deep neural networks are able to extract different discriminative features and classify the object from end to end. It has become one of the most popular and powerful methods in computer vision and natural language processing [21, 22]. In the field of offline signature verification, Hafemann et al. [23] proposed a two-stage CNN to obtain the writer-independent feature representation. They also presented different effective strategies that use skilled forgeries to guide the feature representation process [5]. Furthermore, deep convolutional generative adversarial networks were applied in Ref. [25]; the networks were with multi-phase architecture and utilized unsupervised learning from feature extraction.

Siamese network is a typical structure of CNN. It usually contains two identical subnetworks and computes the distance metric between the highest level feature representations. It was first proposed by Bromley et al. [26] and has been very popular for verification tasks. In the area of signature verification, Dey et al. [4] proposed Signet based on the siamese network for writer-independent offline signature verification. Xian Zhang et al. [49] developed a pioneering deep learning model named the Multi-Path Attention Siamese Convolution Network (MA-SCN) to address the challenge of maintaining a delicate balance between achieving high accuracy and maintaining a low model capacity in the context of writer-independent offline signature verification. Xiao et al. [50] presents a two-stage siamese neural network model for offline handwritten signature verification. Amruta B. Jagtap et al. [51] introduced a method involving siamese neural networks and convolutional neural networks (CNNs) for signature verification. Wei et al. [27] proposed inverse discriminative network which is capable of intensifying the effective information of signatures. Mustafa et al. [28] utilized two-channel CNN as a feature extractor, where the two channels represent reference and query signatures, respectively. Similar work was extended in Ref. [29], Lin et al proposed a 2-channel-2-logit network, which uses logits as the output of convolutional layers to measure the similarity between reference and query signatures.

The stroke information is sparse in signature images, and the most of pixels are background pixels. Attention mechanism makes networks enhance weak information from the input data which is significantly relevant to the output. Up to now, many successful approaches have been reported. Wang et al. [30] proposed a residual attention network that captures mixed attention. Hu et al. [31] introduced a squeeze-and-excitation block by modeling the interdependencies between the channels of convolutional layers. Ren et al. [52] employed a transformer-based method called 2C2S for signature verification. This approach processes original and central-cropped signature pairs using two streams, with the incorporation of a squeeze-and-excitation operation and an upsampling enhancement module. The 2C2S model attains high verification accuracy across various datasets, showcasing its efficacy in the field of signature verification. Chen et al. [32] proposed a reverse attention block which guides the whole network sequentially discover complement object regions and details. Woo et al. [33] combined both channel and spatial attention modules for attention-based feature refinement. Except for attention mechanism, feature reuse also improves the verification performance. Huang et al. [34] proposed dense convolutional networks, which connect each layer to other layers and strengthen feature representation. Motivated by the above ideas, attention-based multiple siamese networks with primary representation guiding are proposed to capture the discriminative stroke information between signatures.

3 The proposed method

The architecture of our proposed attention-based multiple siamese networks with primary representation guiding is illustrated in Fig. 2, and the parameters of each branch of the proposed method are listed in Table 1.

At first, the signature images are passed through several preprocessings to remove extraneous information. Then, the proposed multiple siamese networks take these images (i.e., original signature images with white backgrounds and inverse signature images with black backgrounds) as inputs, which leads the model to highlight the stroke information instead of the background pixels. Note that the foregrounds of all signatures maintain grayscale strokes. These signature images are delivered to four weight-shared branches with the same structure. The two branches which receive original signature images are named as original branches, while the other two branches receiving inverse images are named as inverse branches. In every single branch, there are four convolutional modules, and each module consists of two convolutional layers followed by Rectified Linear Units (ReLU) function and a pooling layer. The channel number of four convolutional modules in each branch is 32, 64, 96 and 128, respectively.

The micro but significant stroke deformation information is easily lost in the pooling operation. In order to deal with this problem, primary representation guiding is proposed. For original branches, feature maps of the first layer in the first convolutional module are extracted. They are then added to the output of each other convolutional modules and are used as the input of following convolutional modules. In consideration of the mismatching problem of feature size and dimension during the addition operation, a transformation module composed of the nearest neighbor algorithm and 1×1 convolution layer is utilized to address this issue. As shown in

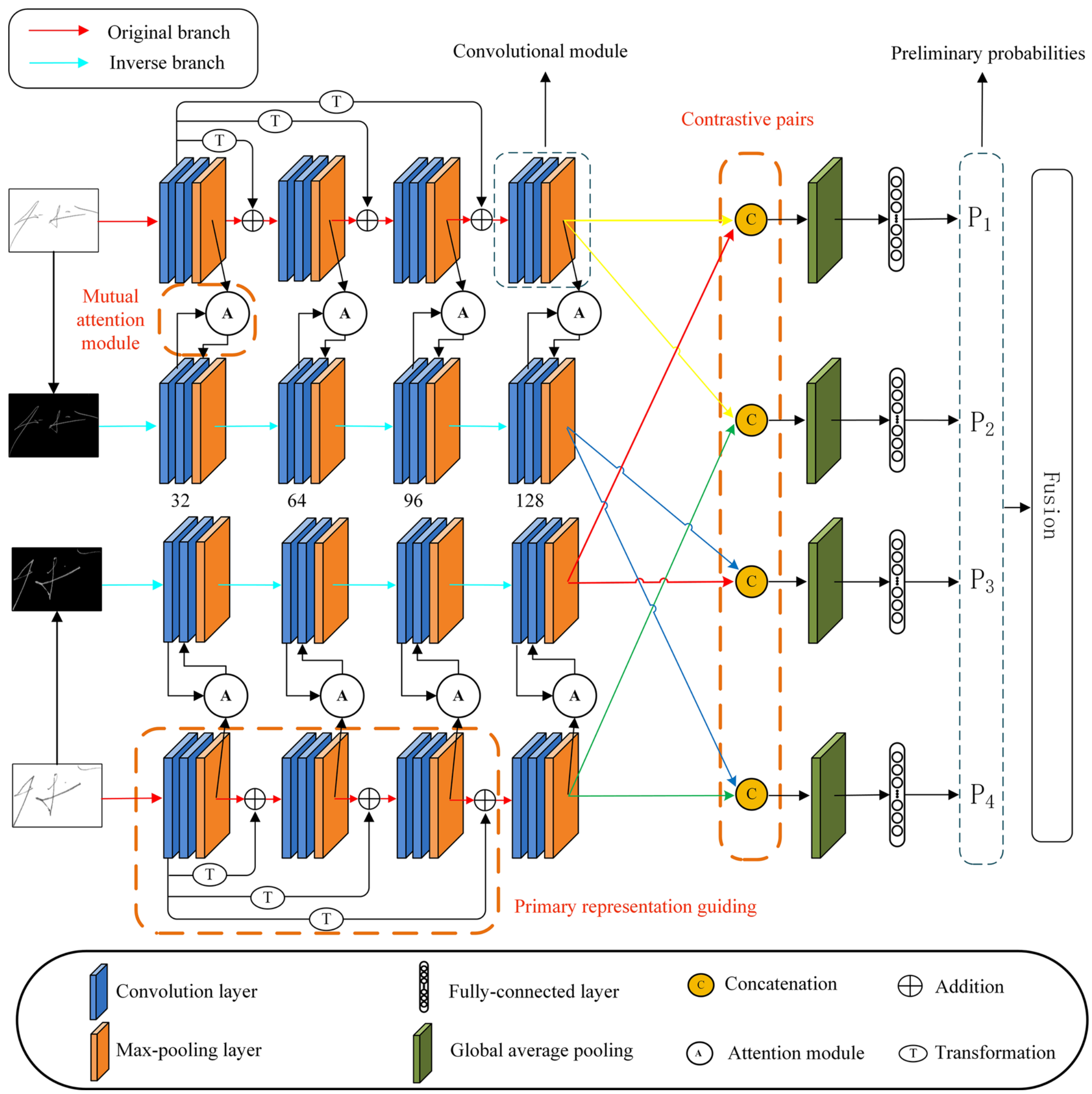


Fig. 2 Architecture of the proposed multiple siamese networks with primary representation guiding

Fig. 2, there are eight mutual attention modules in MSN. Each of them receives the output both from convolutional modules in the original branch and the first convolutional layer in the inverse branch. Then, feature maps are passed to the second layer of the convolutional module in the inverse branch via the attention module. It contributes to ensuring the networks focus on discriminative stroke information between original and inverse branches. The last feature maps from four branches are concatenated with each other and grouped into four contrastive pairs. With the help of global average pooling, these pairs are fed into fully connected layers and make independent preliminary decisions. The final results are generated from these decisions using a voting mechanism.

3.1 Pre-processing

Before the signature images are fed into the proposed MSN, it is necessary to enhance the quality of signature images. The following measures are adopted.

Remove the noise. In order to make the networks learn as many characteristic attributes of signatures as possible, we attempt to remove noise and keep foreground information about the signature itself. First, OTSU algorithm [35] is utilized to separate the foreground and background area and to obtain the binarized image. Second, for the binarized signature image, if the pixel value X_{ij} is not 0, it is reset to 1.

Table 1 Summary of the Parameters of Each Branch in MSN

Layer	Size	Other parameters
Input	$1 \times 155 \times 220$	–
Convolution	$32 \times 3 \times 3$	Stride = 1, pad = 1
Pooling	$32 \times 2 \times 2$	Stride = 2
Convolution	$64 \times 3 \times 3$	Stride = 1, pad = 1
Pooling	$64 \times 2 \times 2$	Stride = 2
Convolution	$96 \times 3 \times 3$	Stride = 1, pad = 1
Pooling	$96 \times 2 \times 2$	Stride = 2
Convolution	$128 \times 3 \times 3$	Stride = 1, pad = 1
Pooling	$128 \times 2 \times 2$	Stride = 2
GAP	–	–
Fully Connected	512	–
Fully Connected	1	–

$$X_{ij} = \begin{cases} 0 & X_{ij} = 0 \\ 1 & X_{ij} \neq 0 \end{cases} \tag{1}$$

Third, the original grayscale image $I_{original}$ is dot-multiplied with the mask X to obtain the signature grayscale image I.

$$I = I_{original} * X \tag{2}$$

However, there may still exist useless saturated noisy pixels at the image edges. The cropping of these borders contributes to reducing unnecessary calculations. The algorithm is implemented as follows:

$$CC = \begin{cases} saved, & S_{cc} > m \\ removed, & S_{cc} \leq m \end{cases} \tag{3}$$

The first step is to find all Connected Components (CC) in the image and count the size S_{cc} of them. Then, a threshold m is defined to remove these components. In this way, the remaining noisy pixels are removed.

Resize the images. Signature images have different sizes. As a typical CNN, the input size of our proposed MSN should be fixed. Thus, all images need to be resized into 155×220 using the linear interpolation.

Invert the images. All original signature images are inverted to generated inverted signature images. More specifically, the pixel I_{inverted} of the image is calculated as:

$$I_{\text{inverted}}(x, y) = 255 - I(x, y) \tag{4}$$

Figure 3 shows the pre-processing results of the signature images.

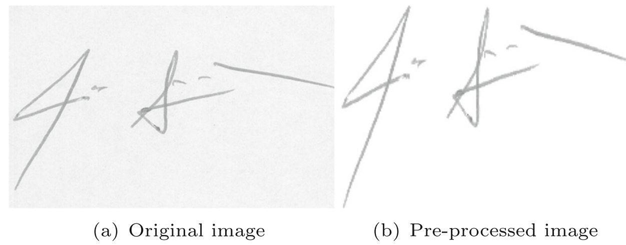


Fig. 3 Pre-processing results of signature images

3.2 Primary representation guiding

Downsampling reduces the number of parameters and model complexity, and it is a commonly used operation in CNN. For personal authentication, the crucial information is attached to the strokes of handwritten signatures. However, the micro but meaningful deformations of signatures are illegible after downsampling. To further enhance the information flow between convolutional modules and ensure the networks focus on low-level stroke features, direct connection of feature maps from the first convolutional layer to subsequent layers is employed. To reduce the calculation complexity, the features are not combined together through concatenation before they are fed into following layers; instead, they are combined to a single tensor using element-wise addition. In view of the mismatch problem of size and feature channel numbers, a downsampling operation and 1×1 convolutional layers are applied in our architecture. Figure 4 shows the layout of the proposed method schematically. The input of $(l+1)$-th $(l = 1, 2, 3)$ convolutional module is described as:

$$x_{l+1} = T_l(x_1) + H_l(x_l) \tag{5}$$

where x_1 refers to the output of first layer in the first convolutional module. x_{l+1} denotes the input of $(l+1)$-th convolutional module. T_l is a composite function of size and channel transform operation. H_l can be regarded as another composite function of operations such as convolution layer, rectified linear unit and pooling layer.

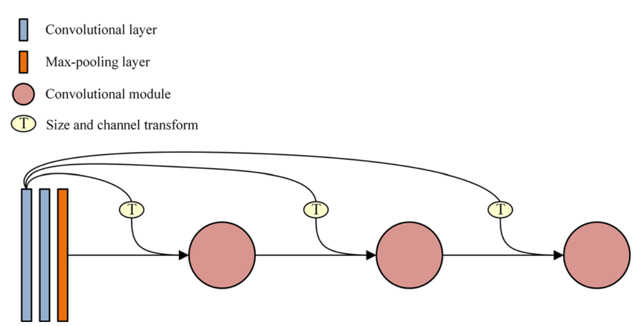


Fig. 4 Illustration of the primary representation guiding

3.3 Mutual attention module

Inspired by previous attention mechanisms [30–33], we present an efficient mutual attention module to make the networks capture discriminative stroke features between original and inverse signature images. By transferring information from original images to inverse images, the networks are capable of focusing on mutual details of signature images, thus improving the model performance. There are eight attention modules in MSN. In shallow layers, the attention modules enhance the representations of low-level features. In deep layers, the obtained high-level features have a more essential description of the semantic information, which is also conducive to verification. Figure 5 shows the structure of the proposed mutual attention module. In order to optimize the training process, residual connections are employed. Note that both spatial and channel attention are adopted in our work.

Spatial attention aims to discover which area of the signature image is informative. The output features of original branches are defined as $\mathcal{R}$. These features are fed into convolutional layers through upsampling operation with the nearest neighbor algorithm. Then a sigmoid activation function is used to transform features and obtain pixel-level attention scores. $\mathcal{S}$ indicates the output after the activation function, and $\mathcal{O}$ represents the output after the first convolutional layer of convolutional modules in inverse branches. The feature $\mathcal{S}$ is multiplied by $\mathcal{O}$ to generate weight distribution and added to $\mathcal{O}$ which can be expressed as $\mathcal{S} \cdot \mathcal{O} + \mathcal{O}$, and thus restrain unimportant information and enhance crucial features of signatures.

Channel attention is utilized to exploit the relationship among channel features. Global Average Pooling (GAP) and Global Max Pooling (GMP) layers receive the output features of spatial attention, respectively. GAP layers are able to collect global information of feature maps and avoid overfitting. GMP layers lead the model to capture more identifiable stroke features. The output features from GAP and GMP

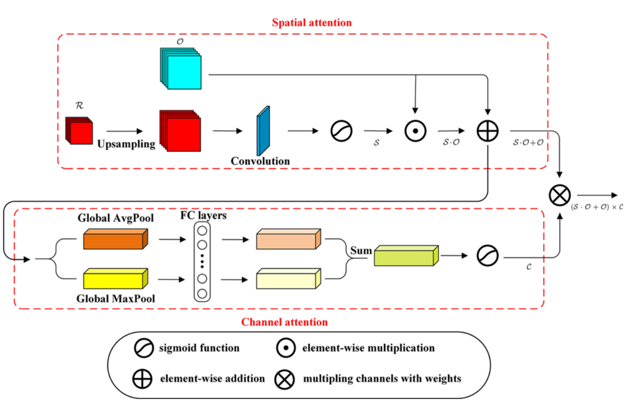


Fig. 5 Architecture of the proposed mutual attention module

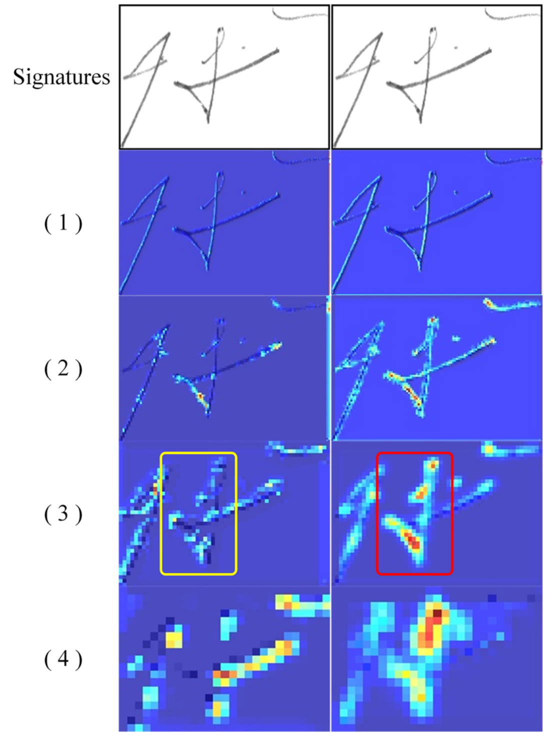


Fig. 6 Heatmap visualization from attention modules of Ref. [27] (left) and ours (right)

layers are then fed into FC layers. After the summation over output feature maps of FC layers, a weight vector $\mathcal{C}$ is produced through sigmoid activation. The channel attention mechanism is described as:

$$\begin{aligned} \mathcal{W}_t &= \sigma(\mathrm{FC}(\mathrm{GAP}(\mathcal{S} \cdot \mathcal{O} + \mathcal{O})) \\ &\quad + \mathrm{FC}(\mathrm{GMP}(\mathcal{S} \cdot \mathcal{O} + \mathcal{O}))) \end{aligned} \tag{6}$$

where σ represents the sigmoid function, and $\mathcal{W}_t$ denotes the attention weights of t-th channel. In the end, an attention mask $(\mathcal{S} \cdot \mathcal{O} + \mathcal{O}) \times \mathcal{C}$ is generated through element-wise multiplication and is fed into the second layer of convolutional modules in inverse branches.

The mutual attention module makes MSN possible to discover prominent stroke features from both original and inverse branches. Figure 6 shows some feature maps output from attention modules. Row 1 denotes the signature samples after pre-processing, and the remaining rows represent visualization results of the output from different-level attention modules. In order to demonstrate the effectiveness and efficiency of our attention module, the visualization results of proposed mutual attention module are compared with that of Ref. [27]. The red region corresponds high attention to strokes, while the dark blue region represents the model is indifferent to the background. Column 1 denotes the visualization results of Ref. [27], while column 2 represents the results of our proposed attention module. From the comparison between the red box and the yellow box in Fig. 6, it indicates that the our proposed attention module gives more concerns to the stroke features, which contributes to discovering more prominent strokes from both original and inverse branches.

3.4 Contrastive pairs

MSN contain four weight-shared branches and needs further processing to obtain the final result. It is meaningful to learn useful differentiated information by comparing reference samples with query samples. Therefore, combining reference and query features in an ordered way is considered. More specifically, the feature maps output from four branches are concatenated to four contrastive pairs. The four reference-query pairs are inverse reference-original query signatures, original reference-original query signatures, original reference-inverse query signatures and inverse reference-inverse query signature, respectively. The feature maps of each contrastive pair are spatially pooled using GAP which is possible to enforce correspondences between feature maps and categories. Then, the features are fed into FC layers to produce the class predictions. The value of c-th channel z_c after GAP layer can be calculated from c-th feature map p_c. The formula is defined as:

$$z_c = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} p_c(i, j) \tag{7}$$

where H and W are height and width of the feature map.

Each contrastive pair is capable of determining whether the input pair is genuine-genuine or genuine-forgery independently. MSN should make the same decision for all contrastive pairs no matter if the signature is inverted or not. In order to measure the difference between different pairs, a binary cross-entropy-based loss function is utilized:

$$Loss\,(X_i, Y_i) = -\sum_{i=1}^{4} w_i[y_i \log x_i + (1-y_i)\log(1-x_i)] \tag{8}$$

where y_i is the binary ground truth label, denoting whether the reference and query samples belong to the same writer or not. x_i denotes prediction probability which ranges from [0,1].

There are four decision results according to the structure of MSN. It is necessary to produce the final prediction results. Thus a voting mechanism is put forward here to solve the problem which can be calculated as follows:

$$Pre = \begin{cases} 1, & N_p \geq 3 \\ 0, & N_p, < 3 \end{cases} \tag{9}$$

where N_p is the number of contrastive pairs being predicted as genuine samples correctly. Pre indicates whether the two signatures are from the same writer or not.

4 Experiments and discussions

In this section, we demonstrate several experiments that reflect the various aspects of the proposed method for establishing the authorship of offline writer-independent handwritten signatures.

4.1 Datasets and experimental protocol

In order to evaluate the proposed method, we consider four publicly available datasets: CEDAR [36], BHSig-B, BHSig-H [37], and UTSig [38]. These datasets bring a unified solution to compare various approaches and evaluate their verification performance in a relative fair testing environment. Here is a brief description of the four datasets.

The CEDAR dataset is an English signature dataset that contains 55 individuals' samples [36]. Every writer is asked to sign 24 genuine signatures. There are 24 skilled forgeries per writer which are collected from about 20 skillful forgers. Therefore, the dataset comprises 24×55 = 1,320 genuine signatures and 1,320 forgeries. These signatures are scanned at 300 dpi resolution in grayscale format and stored as PNG images. The BHSig260 dataset [37] consists of two parts: BHSig-B and BHSig-H Dataset. BHSig-B is a Bengali dataset that contains 100 sets of offline handwritten signatures. For each writer, 24 genuine signatures and 30 skilled forgeries are available. This results in 24×100 = 2,400 genuine signatures and 30×100 = 3,000 forgeries. BHSig-H contains 160 individuals' samples which are written in Hindi. It consists of 24 genuine signatures and 30 forgeries for each writer, resulting in 24×160 = 2,840 genuine signatures and 30×160 = 4,800 skilled forgeries altogether. Both BHSig-B and BHSig-H dataset are collected from individuals with different educational backgrounds and ages. The signatures are scanned in grayscale format with 300 dpi resolution and stored as TIFF files. The UTSig dataset is a Persian language signature database and contains 115 individuals' signatures [38]. For each writer, there are 27 genuine signatures and 45 forgeries. As a result, 27×115 = 3,105 genuine signatures and 45×115 = 5,175 skilled forgeries are collected. The signatures are scanned at 600 dpi resolution and stored as TIF files.

Our proposed method is designed for writer-independent signature verification. Thus, the datasets need to be divided into training and testing samples independently. In columns 2 and 3 of Table 2, the number of writers used for training and testing is given. As an example, CEDAR dataset contains 24 genuine signatures for each writer; thus, there

Table 2 The Details of Experimental Protocol on Different Datasets

Dataset	$Train$	$Test$	Positive pairs	Negative pairs
CEDAR	50	5	276	276
BHSig-B	50	50	276	276
BHSig-H	100	60	276	276
UTSig	60	55	351	351

Table 3 Confusion Matrix for the Classification Results

True label	Predicted label 0	Predicted label 1
0	TN	FP
1	FN	TP

are $C_{24}^2 = 276$ genuine-genuine signature pairs. By combining all the genuine-forgery signatures of each writer, there are $24 \times 24 = 576$ negative pairs and 276 genuine-forgery pairs are randomly selected from each writer to balance the similar and dissimilar classes. Likewise, for BHSig-B and BHSig-H datasets, there are $C_{24}^2 = 276$ genuine-genuine and 276 genuine-forgery signature pairs for each writer. For UTSig, there are $C_{27}^2 = 351$ genuine-genuine and 351 genuine-forgery pairs per writer. Columns 4 and 5 in Table 2 give the number of positive and negative pairs of each writer used from the four datasets.

A global threshold T is used to verify whether the reference and query samples are a genuine-genuine or genuine-forgery pair. Table 3 provides the confusion matrix for the verification results based on threshold T. In Table 3, 0/1 represents the query signature is forged/genuine, respectively. TP (True Positive) represents the number of genuine signatures that have been predicted as genuine. TN (True Negative) represents the number of forged signatures that have been predicted as forgeries. FN (False Negative) represents the number of genuine signatures that have been predicted as forgeries. FP (False Positive) represents the number of forged signatures that have been predicted as genuine signatures.

Several metrics are considered to evaluate our approach according to Table 3. FAR (False Acceptance Rate) is defined as the ratio of forgeries accepted as genuine samples by the verification system, while FRR (False Rejection Rate) is defined as the ratio of genuine signature rejected by the system. EER (Equal Error Rate) and AER (Average Error Rate) have also been used to measure the performance of the proposed verification system. EER indicates the error rate when FAR and FRR are equal, while AER is the average value of both FAR and FRR. ACC (Accuracy) denotes the ratio of correct predictions over all test signatures.

$$FAR = \frac{FP}{TN + FP} \tag{10}$$

Table 4 Verification Performance of Different Contrastive Pairs on BHSig-B Dataset(%)

Number	FAR	FRR	EER	ACC
2 pairs	13.99	11.51	12.75	87.24
3 pairs	17.66	**7.56**	12.62	87.38
4 pairs	**13.63**	9.91	**11.77**	**88.22**

$$FRR = \frac{FN}{FN + TP} \tag{11}$$

$$AER = \frac{FAR + FRR}{2} \tag{12}$$

$$ACC = \frac{TP + TN}{TP + FP + TN + FN} \tag{13}$$

4.2 Analysis of contrastive pairs for multiple siamese networks

In order to investigate the effectiveness of the proposed multiple siamese networks, we begin investigating if the number of contrastive pairs for multiple siamese networks can be utilized. The input of multiple siamese networks is four signature images which two of them are original images and the others are inverse images. A contrastive pair is composed of two of those four parallel branches. The number of contrastive pairs has an influence on the effectiveness of multiple siamese networks. The local contrastive information is cracked when the number of contrastive pairs is too small. The more feature usually mean better preference. There are four parallel branches mentioned in Sect. 3. We choose 2, 3 and 4 contrastive pairs of them to obtain the final verification results, respectively. Two contrastive pairs consist of original reference-inverse query signatures and inverse reference-original query signatures. Three contrastive pairs consist of original reference-inverse query signatures, inverse reference-inverse query signatures and inverse reference-original query signatures. The verification results are summarized in Table 4. It can be seen that the 4 contrastive pairs of proposed method get the best experimental performance, and the accuracy is 87.24%. It proves that the proposed 4 contrastive pairs are capable of learning effective feature representations by comparisons of reference and query signatures. Note that the primary representation guiding and mutual attention modules are removed in the experiment.

4.3 Analysis of mutual attention module for multiple siamese networks

The attention module plays a significant role in discovering information from original branches and inverse branches. In this subsection, we evaluate the performance of the proposed

Table 5 Verification Performance of Attention Module on BHSig-B Dataset(%)

Model	FAR	FRR	EER	ACC
No attention module	**11.54**	11.89	11.71	88.28
Attention module in Ref. [27]	12.37	9.45	11.06	88.93
Mutual attention module (ours)	13.49	**7.55**	**10.52**	**89.47**

Table 6 Verification Performance of two strategies on attention module on BHSig-B dataset(%)

Strategy	FAR	FRR	EER	ACC
I-to-O	15.05	**4.53**	9.11	90.20
O-to-I	**8.07**	8.14	**8.11**	**91.88**

mutual attention module. Our module is compared with that of Ref. [27] to demonstrate that mutual attention focuses on more prominent stroke features. Table 5 gives the performance of the proposed mutual attention module on BHSig-B Dataset. It can be seen that our attention module exceeds that of Ref. [27]. Note that multiple siamese networks without attention module achieve the lowest FAR, which means the attention model increases the probability of false acceptance.

In our attention module, the output features from original branches are element-wise multiplied with the feature maps of the first layer from convolutional modules in inverse branches. In other words, we employ the information of original branches to energize the feature extraction of inverse branches(O-to-I). Theoretical, the information of inverse branches can be also used to energize the feature extraction of original branches (I-to-O). Comparative experiments are conducted on BHSig-B dataset to verify which strategy is better. Table 6 shows the experimental results of two strategies. The best verification performance is achieved by the strategy of O-to-I.

4.4 Analysis of primary representation guiding for multiple siamese networks

Primary representation guiding connects low-level features with higher ones to make the model focus on low-level stroke information. Three types of connections can be designed according to the deployment of primary representation guiding. In other words, primary representation guiding can be deployed in the original branches, inverse branches or both original and inverse branches. The performance results of three types of connections are shown in Table 7. The accuracy of proposed primary representation guiding in original branches is 0.07% percentage points higher than the inverse branches, which means the low-level stroke features contribute to the networks learning ability. However, when we

Table 7 Verification Performance of Different Deployments of Primary Representation Guiding on BHSig-B Dataset(%)

Model	FAR	FRR	EER	ACC
Original branches	**10.71**	9.52	**10.11**	**89.88**
Inverse branches	11.31	**9.05**	10.18	89.81
Both	10.73	10.49	10.61	89.38

deploy proposed primary representation guiding on both original and inverse branches, the results perform not well as expected, demonstrating that a simple stack of the primary representation guiding method may cause the performance reduction problem. In fact, there are only very small differences among the performance of three deployments.

4.5 Ablation analysis of the proposed method on various datasets

There are three main contributions of the proposed method: Contrastive Pair (CP), Mutual Attention (MA) and Primary Representation Guiding (PRG). The ablation experiments are conducted to discover the best combination of three contributions in different ways. There are totally four modes on the basis of baseline: CP+MA, CP+PRG, MA+PRG and CP+MA+PRG (baseline means the initial multiple siamese networks without mutual attention and primary representation guiding). Table 8 gives the verification results of different combinations on BHSig-B dataset, respectively. CP+MA+PRG achieves the best performance in ACC, FAR and EER metrics. CP+MA-based mode gets the lowest FRR, while its FAR is the highest among four combinations. To sum up, we get performance improvements, an ACC of 4.64% on BHSig-B dataset than baseline.

In order to prove applicability of our proposed method (CP+MA+PRG), it is also evaluated on other three datasets. From Table 9, we can see that the proposed method achieves performance improvements on the all three datasets. Compared with the baseline, the accuracy increased most on UTSig which is 7.73%, FAR decreased most on UTSig dataset which is 4.14%, FRR decreased most on UTSig dataset which is 11.12%, and EER decreased most on UTSig

Table 8 Verification Performance of Different Combinations on BHSig-B dataset(%)

Model	FAR	FRR	EER	ACC
Baseline	13.99	11.51	12.75	87.24
CP+MA	10.42	**6.44**	8.43	91.56
CP+PRG	8.23	10.88	8.96	90.43
MA+PRG	8.89	8.72	8.80	91.19
CP+MA+PRG	**8.07**	8.14	**8.11**	**91.88**

Table 9 Comparison With Baseline on CEDAR, BHSig-H and UTSig Dataset (%)

Dataset	Method	FAR	FRR	EER	ACC
CEDAR	Baseline	1.30	6.66	3.98	96.01
	MSN (ours)	**0.72**	**2.17**	**1.12**	**98.56**
BHSig-H	Baseline	**14.34**	12.55	13.45	86.54
	MSN (ours)	14.81	**5.54**	**10.17**	**89.82**
UTSig	Baseline	21.31	30.31	25.81	74.18
	MSN (ours)	**16.90**	**19.19**	**18.08**	**81.91**

Table 10 Comparison With Other Approaches on the BHSig-B Dataset(%)

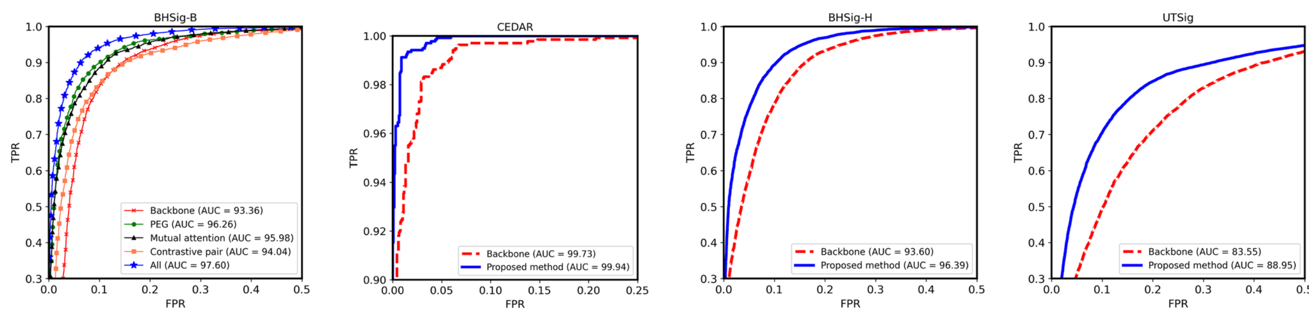
Method	Type	FAR	FRR	ACC
Dey et al. [4]	WI	13.89	13.89	86.11
Lin et al. [29]	WI	–	–	88.08
Pal et al. [37]	WD	33.82	33.82	66.18
Jadhav et al. [39]	WD	–	–	90.36
Jain et al. [41]	WI	–	–	76.03
Xiao et al. [50]	WI	14.25	6.41	90.64
MSN (ours)	WI	**8.07**	**8.14**	**91.88**

dataset which is 7.73%. The results on CEDAR achieve the best performance than other datasets. The performance on UTSig dataset only achieves 81.91%, and it is mainly because UTSig dataset is a more challenging dataset.

It can be concluded from Table 9 that the performance of the proposed method excels the baseline to varying degrees on different datasets. Figure 7 shows the ROC curves of the proposed method on the four datasets. For BHSig-B dataset, more curves are displayed since the ablation experiments are performed on it. The results prove the effectiveness and strength of MSN. Moreover, the average of Genuine-Genuine (G-G) and Genuine-Forgery (G-F) prediction probabilities produced for each writer on the four datasets is illustrated in Fig. 8. The prediction probabilities are expected to be kept as larger as possible for G-F, while the probabilities are supposed to be kept as smaller as possible for G-G. Nonetheless, it can be observed there remain probability values of G-G and G-F for the writer with same index is in close proximity, which reflects the forgeries are difficult to distinguish from genuine ones in the existing system.

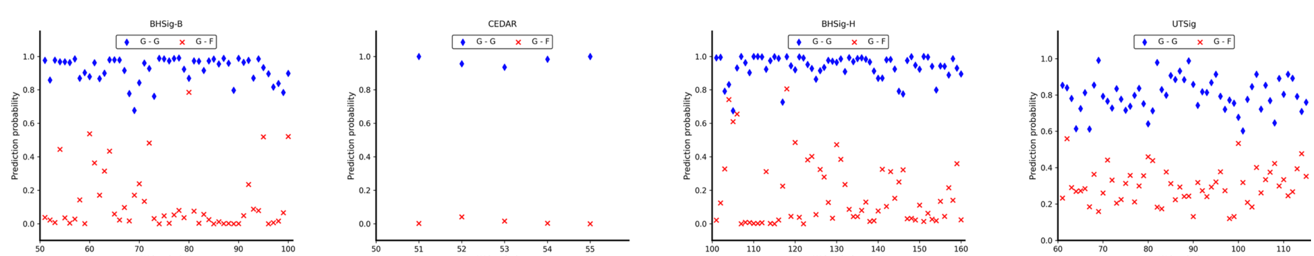
4.6 Performance comparison with the state-of-the-art methods

We also compare the proposed method with other state-of-the-art approaches on the four datasets. Table 10 shows the comparative analysis on the BHSig-B dataset. It is clear that the proposed method performs better than previous works which consider handcrafted features or deep learning-based approaches as feature extraction techniques.



(a) BHSig-B (b) CEDAR (c) BHSig-H (d) UTSig

Fig. 7 ROC curves of proposed method obtained on four datasets



(a) BHSig-B (b) CEDAR (c) BHSig-H (d) UTSig

Fig. 8 Average of Genuine-Genuine (G-G) and Genuine-Forgery (G-F) prediction probabilities computed for each writer on the four datasets

Table 11 Comparison With Other Approaches on the CEDAR Dataset (%)

Model	Type	FAR	FRR	EER	AER
Hafemann [5]	WD	–	–	4.63	–
Wei et al. [27]	WI	5.87	**2.17**	3.62	4.02
Kumar et al. [42]	WI	8.33	8.33	–	8.33
Bhunia et al. [43]	WD	–	–	–	1.64
Sharif et al. [44]	WD	4.67	4.67	–	4.67
Dey et al. [4]	WI	0	0	0	0
Dutta et al. [40]	WI	0	0	0	0
Jagtap et al. [51]	WI	0	0	0	0
Xiao et al. [50]	WI	6.78	4.20	–	–
MSN (ours)	WI	**0.72**	**2.17**	**1.12**	**1.44**

Table 12 Comparison With Other Approaches on the BHSig-H Dataset (%)

Method	Type	FAR	FRR	EER	ACC
Dey et al. [4]	WD	15.36	15.36	–	84.64
Lin et al. [29]	WI	–	–	13.34	86.66
Pal et al. [37]	WD	24.47	24.47	24.47	75.53
Dutta et al. [40]	WI	**13.10**	15.09	–	85.90
Jain et al. [41]	WI	–	–	–	83.50
Xiao et al. [50]	WI	12.29	9.60	–	88.98
MSN (ours)	WI	14.81	**5.54**	**10.17**	**89.82**

Table 13 Comparison With Other Approaches on the UTSig Dataset (%)

Method	Type	FAR	FRR	AER	ACC
Soleimani et al. [38]	WD	21.29	39.27	30.28	–
Soleimani et al. [45]	WD	**9.09**	32.42	20.28	–
Rezaei et al. [46]	WD	27.47	15.72	21.60	76.71
Mohapatra et al. [47]	WI	23.34	14.01	–	80.44
Maergner et al. [48]	WD	29.49	**7.88**	18.69	–
MSN (ours)	WI	16.90	19.19	**18.05**	**81.91**

Table 11 gives the evidence that the proposed method outperforms other approaches on CEDAR dataset. A possible reason for the higher performance on this dataset is the large number of signature samples for training. It can be observed from Table 11 that MSN achieve lower error rate in comparing with other approaches.

Tables 12 and 13 depict the comparison results on BHSig-H and UTSig dataset. The accuracies on the two datasets are 89.82% and 81.91%, respectively. Compared with previous work, the proposed MSN achieve better verification performance and are capable of distinguishing reference and query samples' features effectively.

4.7 Analysis of misclassified samples

In Fig. 9, we present signature samples from four distinct signature datasets: CEDAR, BHSig-B, BHSig-H and UTSig. Each sample consists of a reference signature and a query signature. Furthermore, we provide corresponding heatmaps for the cases to enhance the understanding of our experimental results.

For the CEDAR dataset, an instance is shown that our method predicts the signatures are from different writers, but it is actually created by the same person. The raw signature samples illustrate minor pen-stroke differences, leading to the model's erroneous prediction. This indicates that in certain scenarios, our models may struggle to capture subtle differences effectively. Similar to the first case, the scenario in the BHSig-B dataset predicts the signatures are from different writers improperly. Consistent with the initial case, the heatmap from BHSig-B reveals subtle differences alongside similarities. The model's predictions underscore the performance of a model lacking attention in this scenario. This emphasizes the challenge of traditional models in balancing discriminative and similarity aspects.

Moving to the BHSig-H dataset, an instance is displayed where the model predicts that the signatures are from the same person, but it is actually created by different writers. The signatures' similarity in this case heightens the complexity of authentication. Despite these similarities, the heatmap from the model does not adequately capture sufficient feature information for precise predictions. This instance underscores the limitations of the model when handling similar signatures. Lastly, in the UTSig dataset, another case reveals the model predicts that the signatures are from the same person, but it is actually created by different writers. The high similarity in the signature samples poses a challenge. The model fails to recognize subtle differences between these highly similar signatures, indicating an overemphasis on identical regions. This emphasizes the difficulty the model encounters in dealing with closely resembling signatures.

5 Conclusions

In this paper, we aim to solve the problem of writer-independent offline signature verification. To this end, we propose attention-based multiple siamese networks with primary representation guiding. The discriminative information can be learned through contrastive pairs by comparing reference and query signatures. Mutual attention modules are utilized to discover stroke details between original and inverse signatures. We propose the primary representation guiding which connects shallow features with deeper ones to avoid spatial information lost problems. Our best EER on CEDAR, BHSig-B, BHSig-H and UTSig datasets reaches

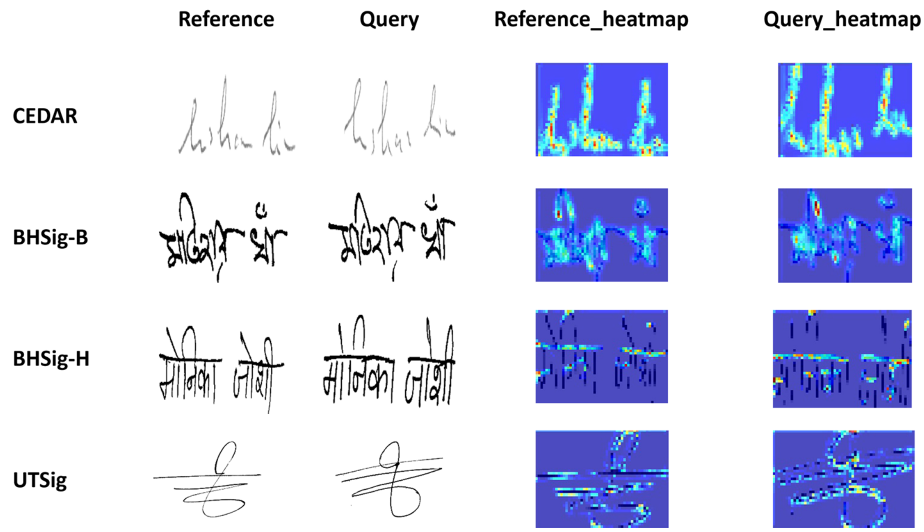


Fig. 9 Visualization of Misclassified Samples in Different Signature Detests

1.12%, 8.11%, 10.17% and 18.08%, respectively, which outperforms other previous approaches. Experimental results demonstrate that our proposed method is effective for writer-independent offline signature verification.

While multiple siamese networks exhibit monotonic improvement, it is an important direction for future work to better understand the potential cause of the balance between FAR and FRR when attention module is employed. On the other hand, it would be interesting to explore the use of primary representation guiding for other problems of handwritten document analysis.

Acknowledgements This work is jointly sponsored by National Natural Science Foundation of China (62006150) and Science and Technology Commission of Shanghai Municipality (21DZ2203100).

References

1. Alaei, A., Pal, S., Pal, U., Blumenstein, M.: An efficient signature verification method based on an interval symbolic representation and a fuzzy similarity measure. IEEE Trans. Inf. Forensic Secur. **12**(10), 2360–2372 (2017)
2. Hamadene, A., Chibani, Y.: One-class writer-independent offline signature verification using feature dissimilarity thresholding. IEEE Trans. Inf. Forensic Secur. **11**(6), 1226–1238 (2016)
3. Hafemann, L. G., Sabourin, R., Oliveira, L. S.: Offline handwritten signature verification - Literature review, in *Proc. Int. Conf. Image Process. Theory, Tools Appl.*, (2017), pp. 1-8
4. S. Dey, A. Dutta, J. Toledo, S. Ghosh, J. Llados and U. Pal, Signet: Convolutional siamese network for writer independent offline signature verification, 2017, Available: arXiv:1707.02131
5. Hafemann, L., Sabourin, R., Oliveira, L.: Learning features for offline handwritten signature verification using deep convolutional neural networks. Pattern Recognit. **70**, 163–176 (2017)
6. Zois, E., Theodorakopoulos, I., Economou, G.: Offline handwritten signature modeling and verification based on archetypal analysis, in *Proc. IEEE Int. Conf. Comput. Vision*, (2017), pp. 5515-5524
7. Impedovo, D., Pirlo, G.: Automatic signature verification: The state of the art, in IEEE Trans. Syst. Man Cybern. Part C-Appl. Rev. **38**(5), 609–635 (2008)
8. Diaz, M., Ferrer, M., Impedovo, D., Malik, M., Pirlo, G., Plamondon, R.: A perspective analysis of handwritten signature technology. ACM Comput. Surv. **51**(6), 1–39 (2019)
9. Baltzakis, H., Papamarkos, N.: A new signature verification technique based on a two-stage neural network classifier. Eng. Appl. Artif. Intell. **14**, 95–103 (2001)
10. Drouhard, J., Sabourin, R., Godbout, M.: A neural network approach to off-line signature verification using directional PDF. Pattern Recognit. **29**, 415–424 (1996)
11. Zois, E.N., Alexandridis, A., Economou, G.: Writer independent offline signature verification based on asymmetric pixel relations and unrelated training-testing datasets. Expert Syst. Appl. **125**, 14–32 (2019)
12. Zois, E. N., Zervas, E., Tsourounis, D., Economou, G.: Sequential motif profiles and topological plots for offline signature verification, in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, (2020), pp. 13245-13255
13. Oliveira, L. S., Justino, E., Freitas, C., Sabourin, R.: The graphology applied to signature verification, in *Proc. Conf. of the Int. Grap. Society*, (2005), pp. 286-290
14. Aubin, V., Mora, M., Santos-Penas, M.: Off-line writer verification based on simple graphemes. Pattern Recognit. **79**, 414–426 (2018)
15. Rivard, D., Granger, E., Sabourin, R.: Multi-feature extraction and selection in writer-independent off-line signature verification. Int. J. Doc. Anal. Recognit. **16**, 83–103 (2013)
16. Bertolini, D., Oliveira, L.S., Justino, E., Sabourin, R.: Reducing forgeries in writer-independent off-line signature verification through ensemble of classifiers. Pattern Recognit. **43**(1), 387–396 (2010)
17. R. Sabourin and G. Genest, An extended-shadow-code based approach for off-line signature verification. I. Evaluation of the bar mask definition, in *Proc. Int. Conf. on Pattern Recognit.*, 1994, pp. 450-453
18. Ferrer, M.A., Vargas, J.F., Morales, A., Ordonez, A.: Robustness of offline signature verification based on gray level features. IEEE Trans. Inf. Forensics Security **7**, 966–977 (2012)
19. Solar, J., Devia, C., Loncomilla, P., Concha, F.: Offline signature verification using local interest points and descriptors, in *Proc. Iberoamerican congress on Pattern Recognition*, (2008), pp. 22-29

20. Okawa, M.: Synergy of foreground-background images for feature extraction: Offline signature verification using fisher vector with fused kaze features, *Pattern Recognit.*. pp. 480-489, (2018)
21. Ren, S., He, K., Girshick, R., Sun, J.: Faster R-CNN: Towards real-time object detection with region proposal networks. IEEE Trans. Pattern Anal. Mach. Intell. **39**(6), 1137–1149 (2017)
22. Devlin, J., Chang, M., Lee, K., Toutanova, K.: BERT: Pre-training of deep bidirectional transformers for language understanding, (2018), arxiv: 1810.04805. [Online]. Available: arXiv:1810.04805
23. Hafemann, L., Sabourin, R., Oliveira, L.: Writer-independent feature learning for offline signature verification using deep convolutional neural networks, in *Proc. Int. Jt. Conf. Neural Networks*, (2016), pp. 2576-2583
24. Masoudnia, S., Mersa, O., Araabi, B.N., Vahabie, A., Sadeghi, M., Ahmadabadi, M.: Multi-representational learning for offline signature verification using multi-loss snapshot ensemble of CNNs. Expert Syst. Appl. **133**, 317–330 (2019)
25. Zhang, Z., Liu, X., Cui, Y.: Multi-phase offline signature verification system using deep convolutional generative adversarial networks, in *Proc. Int. Symp. Comput. Intell. Des.*, (2016), pp. 103-107
26. R. Shah, E. Sackinger, J. Bentz, I. Guyon, C. Moore, L. Bottou, J. Bromley and Y. Lecun, Signature verification using a Siamese time delay neural network, *Int. J. Pattern Recognit. Artif. Intell.*, pp. 737-744, 1993
27. Wei, P., Li, H., Hu, P.: Inverse discriminative networks for handwritten signature verification, in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, (2019), pp. 5757-5765
28. Yilmaz, M. B., Ozturk, K.: Hybrid user-independent and user-dependent offline signature verification with a two-channel CNN, in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, (2018), pp. 639-6398
29. Lin, C., Lin, F., Wang, Z., Yu, G., Yuan, L., Wang, H.: DeepHSV: User-independent offline signature verification using two-channel CNN, in *Proc. Int. Conf. Doc. Anal. Recognit.*, (2019), pp. 166-171
30. Wang, F., Jiang, M., Qian, C., Yang, S., Li, C., Zhang, H., Wang, X., Tang,X.: Residual attention network for image classification, in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, (2017), pp. 6450-6458
31. Hu, J., Shen, L., Albanie, S., Sun, G., Wu, E.: Squeeze-and-excitation networks. IEEE Trans. Pattern Anal. Mach. Intell. **42**, 2011–2023 (2020)
32. Chen, S., Tan, X., Wang, B., Hu, X.: Reverse attention for salient object detection, in *Proc. Eur. Conf. Comput. Vis.*, (2018), pp. 236-252
33. Woo, S., Park, J., Lee, J., Kweon, I.: CBAM: Convolutional block attention module, in *Proc. Eur. Conf. Comput. Vis.*, (2018), pp. 3-19
34. Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K. Q.: Densely connected convolutional networks, in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, (2017), pp. 2261-2269
35. Otsu, N.: A threshold selection method from gray-level histograms. *IEEE Trans. Syst., Man, Cybern.* 9(1), 62-66, (1979)
36. Kalera, M.K., Srihari, S., Xu, A.: Offline signature verification and identification using distance statistics. Int. J. Pattern Recognit. Artif. Intell. **18**(7), 1339–1360 (2004)
37. S. Pal, A. Alaei, U. Pal and M. Blumenstein. Performance of an off-line signature verification method based on texture features on a large indic-script signature dataset, in *Proc. IAPR Workshop Document Anal. Syst.*, 2016, pp. 72-77
38. Soleimani, A., Fouladi, K., Araabi, B.N.: Utsig: A persian offline signature dataset. IET Biom. **6**(1), 1–8 (2017)
39. Jadhav, S. K., Chavan, M. K.: Symbolic representation model for off-Line signature verification, in *Int. Conf. Comput., Commun. Netw. Technol.*,(2018), pp. 1-5
40. Dutta, A., Pal, U., Llados, J.: Compact correlated features for writer independent signature verification, in *Proc. Int. Conf. Pattern Recognit.*, (2016), pp. 3422-3427
41. Jain, A., Singh, S., Singh, K. P.: Signature verification using geometrical features and artificial neural network classifier, *Neural Comput. Appl.*, pp. 1-12, (2020)
42. Kumar, R., Sharma, J.D., Chanda, B.: Writer-independent off-line signature verification using surroundedness feature. Pattern Recognit. Lett. **33**(3), 301–308 (2012)
43. Bhunia, A., Alaei, A., Roy, P.: Signature verification approach using fusion of hybrid texture features, *Neural Comput. Appl.*, pp. 1-12, (2019)
44. Sharif, M., Khan, M., Faisal, M., Yasmin, M., Fernandes, S.L.: A framework for offline signature verification system: Best features selection approach. Pattern Recognit. Lett. **139**, 50–59 (2020)
45. Soleimani, A., Araabi, B.N., Fouladi, K.: Deep multitask metric learning for offline signature verification. Pattern Recognit. Lett. **80**, 84–90 (2016)
46. Rezaei, M., Naderi, N.: Persian signature verification using fully convolutional networks, 2019, arxiv: 1909.09720. [Online]. Available: arXiv:1909.09720
47. Ramesh Kumar Mohapatra, Kumar Shaswat and Subham Kedia, Offline handwritten signature verification using CNN inspired by Inception V1 architecture, in *Proc. IEEE Int. Conf. Image Inf. Process.*, 2019, pp. 263-267
48. Maergner, P., Pondenkandath, V., Alberti, M., Liwicki, M., Fischer, A.: Combining graph edit distance and triplet networks for offline signature verification. Pattern Recognit. Lett. **125**, 527–533 (2019)
49. Zhang, X., Wu, Z., Xie, L., et al.: Multi-Path Siamese Convolution Network for Offline Handwritten Signature Verification. *2022 The 8th International Conference on Computing and Data Engineering*, (2022), pp. 51-58
50. Xiao, W., Ding, Y.: A two-stage siamese network model for offline handwritten signature verification. Symmetry **14**(6), 1216 (2022)
51. Jagtap, A.B., Sawat, D.D., Hegadi, R.S., et al.: Verification of genuine and forged offline signatures using siamese neural network (SNN). Multimed. Tools Appl. **79**, 35109–35123 (2020)
52. Ren, J.X., Xiong, Y.J., Zhan, H., et al.: 2C2S: A Two-Channel and Two-Stream Transformer-Based Framework for Offline Signature Verification. Eng. Appl. Artif. Intell. **118**, 105639 (2023)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.