Deep Metric Learning with Cross-Writer Attention for Offline Signature Verification

Lu-Rong Ling[1,2][0009−0005−0000−0291], Heng Zhang[1][0000−0001−9448−4031], Fei Yin[1][0000−0002−6412−9140], and Cheng-Lin Liu[1,2][0000−0002−6743−4175]

[1] State Key Laboratory of Multimodal Artificial Intelligence Systems (MAIS), Institution of Automation, Chinese Academy of Sciences, Beijing 100190, China
[2] School of Artificial Intelligence, University of Chinese Academy of Sciences, Beijing 100049, China
linglurong2022@ia.ac.cn, heng.zhang@ia.ac.cn, fyin@nlpr.ia.ac.cn, liucl@nlpr.ia.ac.cn

Abstract. Signature verification is a biometric and document forensics technology useful for personal identification in various security applications. Signature verification in the writer-independent scenario remains a challenge, particularly in distinguishing between genuine signatures and skilled forgeries. In this paper, we propose a writer-independent signature verification method based on deep metric learning with cross-writer attention. Our cross-writer attention module includes two parts: SimAM (a Simple, Parameter-Free Attention Module), as well as the cross-attention mechanism. SimAM is combined with each DenseBlock to interact information of two inputs, which makes the learned weights better account for the difference between two input signatures. Cross-attention aligns global and local information in learned feature representations of two input signatures. Further, we introduce a focal contrast loss function for deep metric learning to overcome the sample imbalance. Extensive experiments demonstrate the effectiveness of the proposed method, which achieves superior performance on several public datasets and also indicates the effectiveness of each module.

Keywords: Signature Verification · SimAM · Cross-Attention · Deep Metric Learning.

1 Introduction

Signature verification is an essential component of biometric authentication systems. It plays a crucial role in securing sensitive transactions and critical information access. Signatures, which have long served as a unique and personalized means of identity verification in legal, financial, and administrative domains, are now facing increased demand for accurate and reliable verification methods due to the advent of digital technologies. The main objective of signature verification is to verify the authenticity of a given signature by comparing it with a reference signature. The technology is dichotomized into online versus offline verification

depending on whether the stroke trajectory is recorded or not. Offline signature verification is more challenging because dynamic writing features are not available on signature images.

The performance of offline signature verification is critically dependent on the power of feature extraction algorithms. Manual features such as Histogram of Oriented Gradient (HOG) [1], Scale-Invariant Feature Transform (SIFT) [2], and Speeded Up Robust Features (SURF) [3] were typically employed in conventional machine-learning-based systems. Based on feature extraction, a similarity measure or discriminant function is learned to make the decision of verification. However, these methods have strict restrictions on the format and content of input signatures, and do not guarantee the desired verification performance due to their reliance on manual feature engineering [4]. As a result, despite considerable efforts by researchers to develop traditional approaches, little improvement in performance has been achieved [5].

In recent years, more studies have utilized convolutional neural networks (CNNs) to extract signature features, dramatically improving the performance of signature verification systems compared to previous hand-crafted features. In the framework of deep learning, the prevailing approach for signature verification is the Siamese network, which adjusts weights of the feature extraction network via learning a distance metric for judging whether two input signatures are written by the same person or not. However, despite its effectiveness, its improvement is limited because it fails to consider the interaction information between reference and query signatures during feature extraction. Many improved methods based on the Siamese network have been proposed to enhance the performance. Xiong and Cheng [6] proposed a multiple Siamese network with an attention module. Some works tried to combine global and local features by manual region feature extraction, segmenting the signature image into regions [7,8]. Based on the vision transformer framework, TransOSV [9] presented a new holistic-part unified model, which significantly improved performance and produced competitive outcomes by capturing the relationships among signature strokes from the holistic signature image.

Despite successful feature extraction, mismatches between the feature vectors of query and reference signatures might still arise. Consequently, directly calculating the distance between these vectors can lead to errors due to misalignment issues. Therefore, aligning the extracted features is equally crucial but often overlooked in the past. In this paper, we propose a model for offline signature verification to enhance the feature extraction and address the misalignment problem using cross-writer attention. The model is built on a Deep Convolutional Siamese Network for end-to-end feature representation learning on sample data of signature pairs. To bolster the model's capacity to discern differences between inputs, we integrate a modified SimAM module. This mechanism promotes interaction between the two branches and underscores distinctions between the two signatures during feature extraction. Furthermore, we use an interactive metric learning module with cross-attention to compute the distance between the learned representations. This module addresses the issue of

misalignment between features, providing a straightforward and effective solution that significantly enhances the overall performance. Together, the SimAM and cross-attention form our cross-writer attention module. Extensive experiments demonstrate that both the SimAM and the cross-attention mechanism are effective to improve the signature performance and the proposed method performs competitively to state-of-the-art methods.

The main contributions of the paper are as follows:

- We propose a cross-writer attention mechanism to improve feature representation and metric learning in the signature verification model, which is built on the Deep Convolutional Siamese Network framework.
- We propose a focal contrast loss function for deep metric learning to further improve the performance of the signature verification model.
- Experiments on public benchmarks show the superior performance of our method and also demonstrate the effectiveness of each module.

2 Related Work

2.1 Signature Verification

Early signature verification methods relied mainly on template matching and model learning based on manual features. In the template matching approach, the test signature is compared with templates already stored in the database using the dynamic time warping (DTW) algorithm [10]. In contrast to template matching, statistical models were widely used in signature verification with manually extracted features. Using multiple types of manual features, Baltzakis et al. [11] adopted multi-layer perceptron neural networks and radial base function neural networks in a two-stage manner. Yilmaz et al. [12] combined two types of features, namely, histogram of oriented gradients and histogram of local binary patterns, and fused global and user-dependent classifiers for classification. In the absence of forged signatures as counterexamples, Guerbai et al. [13] modified the decision function of the One-Class Support Vector Machine (OC-SVM) by adjusting the optimal threshold through combining different distances in the OC-SVM kernel to reduce misclassification. To improve the discriminative power, Okawa et al. [14] proposed a new feature extraction approach based on a Fisher vector with fused KAZE features using a multilevel fusion strategy. Other classification models such as adaboost [15] have also played a significant role in signature verification.

Recent work has taken advantage of deep learning, in particular, CNNs [16, 17]. Siamese neural network, first proposed for online signature verification [18], has been widely used in offline signature verification [19,20]. Following the Siamese neural network, many superior methods with improved feature representation and metric learning have been proposed. Wei et al. [21] designed a four-stream network and a multi-path attention mechanism to explore the local signature stroke information. Liu et al. [22] designed a region-based deep

convolutional Siamese network by segmenting each image into a series of overlapping regions for feature representation and metric learning. For capturing the global contextual relationships among signature strokes, Li et al. [9] proposed a novel holistic part unified model based on the vision transformer (ViT) framework [23]. Hafemann et al. [24] investigated the impact of adversarial examples on signature verification. By actively varying existing data and generating new data, adversarial variation network [25] could help signature verification tasks mine more effective features for signature verification.

2.2 Attention Mechanism

Inspired by human visual processing systems, attention mechanisms [26] have been widely used in deep neural networks to refine feature maps. One representative work adopted in signature verification [22] is Squeeze-and-Excitation (SE) [27], which captures some context cues from a global view and then uses two fully connected layers to model interactions between channels. Global context attention [28] incorporates long-range dependencies and effectively models global contexts in a lightweight manner. Attentive Normalization (AN) [29] learns a mixture of affine transformations and utilizes their weightedsum as the final affine transformation applied to re-calibrate features. However, these one-channel attention mechanisms cannot effectively refine image features. Convolutional Block Attention Module (CBAM) [30] infers attention maps along both channel and spatial dimensions, and then the attention maps are multiplied with the input feature map for adaptive feature refinement. This two-step manner in CBAM involves very high computation. In contrast to existing attention modules, SimAM [31] is a simple and parameter-free attention module, which inspires us to design a more efficient signature verification model. Compared to ViT-based model [9, 32, 33], our SimAM-based module can achieve comparable performance but consume less computing resources.

3 The Proposed Method

As shown in Fig. 1, our proposed signature verification model consists of modules of preprocessing, feature extraction and metric learning enhanced with cross-writer attention. The pair of input signature images are first preprocessed to normalize size and gray scale, then fed into the convolutional feature extractor network for feature representations. The similarity score between extracted feature vectors is calculated by the metric model to decide whether the original signatures are written by the same writer or not. For effective feature representation and metric learning, we propose the cross-writer attention module that includes two parts: one is the modified SimAM [31] and the other is the cross-attention [34]. The SimAM is combined with DenseBlock for better feature representation. The cross-attention can capture the highly correlated and salient points in feature space for similarity metric. In training procedure, a focal contrast loss is used to learn the parameters of the whole model.

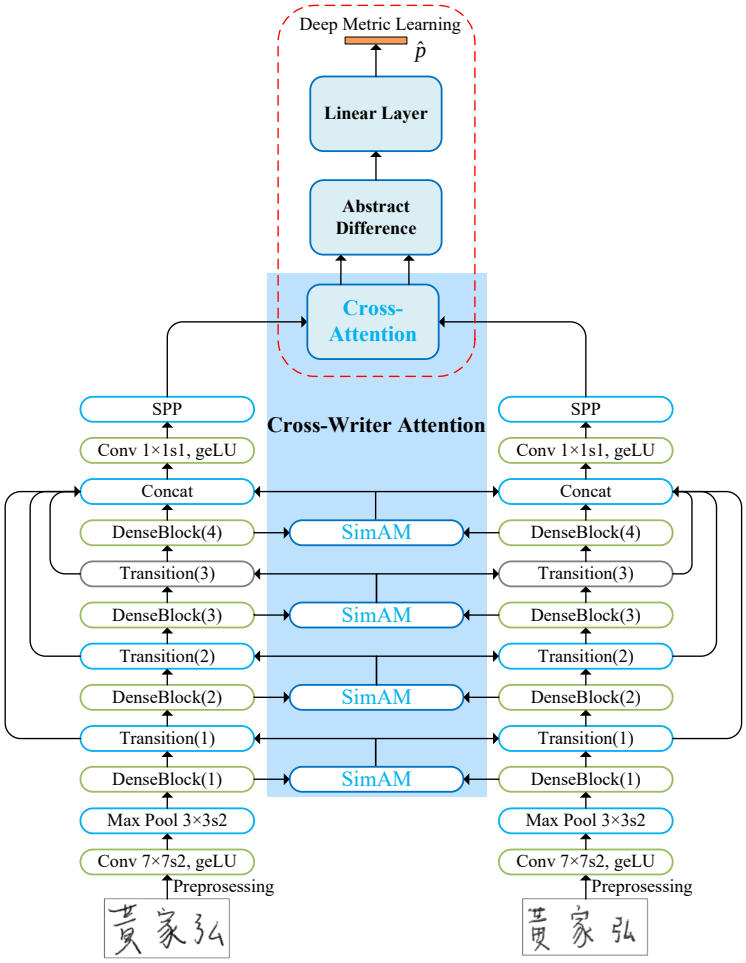


Fig. 1. Network of the proposed Siamese Network.

3.1 Preprocessing

In preprocessing, grayscale normalization is employed to alleviate the effects of illumination discrepancies and diverse pen types utilized by writers. This normalization aims to ensure uniformity in pixel values within the signature regions across different signatures. By standardizing the foreground pixels, grayscale normalization aids in streamlining subsequent processing stages like feature extraction and classification. This normalization procedure is conducted as follows:

$$g'_f = \frac{(g_f - E(g_f)) * 10}{\delta(g_f)} + 30, \tag{1}$$

where g_f and g'_f represent the original and normalized grayscale, respectively. $E(g_f)$ and $\delta(g_f)$ indicate the mean and standard deviation of the original grayscale in the foreground.

To standardize the sizes and positions of signatures within the diverse images, we use the moment normalization method to partially align the locations of signature strokes. Here, $f(x, y)$ and $f(x', y')$ represent the pixel value of the coordinates (x, y) and (x', y') in the original and normalized images, respectively. Subsequently, we map $f(x, y)$ to $f'(x', y')$ with the following formulation:

$$x = \frac{(x' - x'_c)}{\alpha} + x_c, \tag{2}$$

$$y = \frac{(y' - y'_c)}{\alpha} + y_c, \tag{3}$$

In this context, (x'_c, y'_c) denotes the center of the normalized signature, while (x'_c, y'_c) represents the centroid of the original signature. α represents the ratio of the normalized signature size to the original signature size. It can be computed through the central moments μ_{pq} of an inverted image where signature strokes are depicted in gray, and the background is black. For well-fitting the signature foreground within the plane of the normalized image, the scaling ratio is calculated by:

$$\alpha = 0.6 \cdot min(\frac{H_{norm}\sqrt{\mu_{00}}}{2\sqrt{2\mu_{02}}}, \frac{W_{norm}\sqrt{\mu_{00}}}{2\sqrt{2\mu_{20}}}), \tag{4}$$

where H_{norm} and W_{norm} represent the pre-defined height and width of the normalized image, and μ_{pq} denotes the center moments:

$$\mu_{pq} = \sum_x \sum_y (x - x_c)^p (y - y_c)^q [255 - f(x, y)], \tag{5}$$

where we set H_{norm} and W_{norm} as 224 and 224 in experiments, which means that we normalized the size of signature images as 224×224 for the following feature extraction.

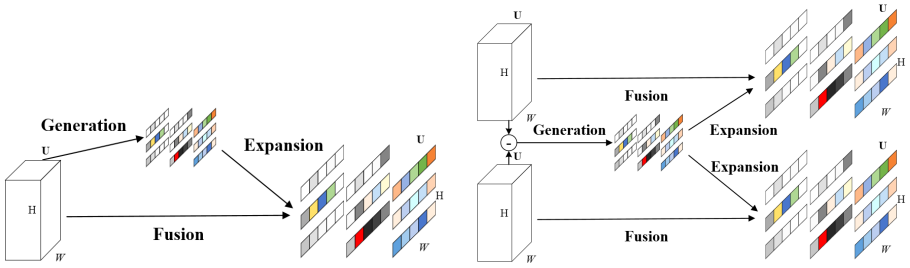
3.2 Feature Extraction

After conducting an experimental comparison of several well-known architectures of CNN and visual transformer including AlexNet [35], VGG [36], ResNet [37], DenseNet [38], ViT [23], and T2T-ViT [39], we chose DenseNet-36 to construct the Deep Convolutional Siamese Network for feature extraction. The dense connectivity of DenseNet is essential for promoting robust feature reuse and effectively capturing intricate patterns in handwriting. This high parameter efficiency is crucial for addressing the gradient vanishing problem, making DenseNet suitable for scenarios with limited data, unlike ViT network, which requires a large volume of data. The dense connectivity mechanism is highly effective in reducing information loss during the training process, which is crucial for preserving the

nuanced details of handwriting. Furthermore, the design of DenseNet's global connections contributes to maintaining a holistic view of the input data, which is especially important in tasks like signature verification, where the entire signature is essential for accurate assessment. The architecture of DenseNet-36 follows the same structure as described in [22].

Based on the DenseNet backbone, Convolutional Siamese architecture, comprising two branches of CNNs with shared weights, is specifically designed to learn feature representations of signature images. And we employ the geLU activation function [40] which is used in BERT and GPT-2. As shown in Fig. 1, the cross-writer attention and SPP (Spatial Pyramid Pooling) are combined into two CNNs branches to interact with the information of two inputs. In this way, the characteristics of two signature images can be fully exploited.

3.3 Cross-Writer Attention



(a) Original SimAM module. (b) Improved SimAM module.

Fig. 2. Comparison diagram of the original SimAM and improved SimAM.

SimAM Modules

Revisiting SimAM module As shown in Fig. 2(a), SimAM [31] is a highly effective attention module designed for CNNs and infers 3-D attention weights for the feature map in a layer without adding parameters to the original networks. The module is based on well-known neuroscience theories and conceptualized as an energy function intended to determine the significance of each neuron. The proposal of SimAM as the attention module with unified weights is based on recent observations suggesting that the two types of attention in the human brain tend to work in harmony. In contrast to conventional attention modules, such as BAM and CBAM, which combine space attention and channel attention in parallel or serial, respectively, SimAM introduces a novel approach by unifying the weights. This addresses the growing understanding of the interconnected nature of different attention mechanisms and seeks to capitalize on the synergy between space and channel attention. An essential aspect of understanding attention is evaluating the importance of each neuron. According to neuroscience,

neurons that are rich in information often exhibit distinct firing patterns from surrounding neurons and are capable of suppressing the activities of other neurons, known as spatial suppression. Therefore, the simplest way to identify these important neurons is by measuring the linear separability between one target neuron and the others. This process leads to the definition of an energy function, which serves the purpose of identifying significant neurons and measuring the linear separability between neurons. The energy function is given by:

$$e_t(\omega_t, b_t, \mathbf{y}, x_i) = \frac{1}{M-1} \sum_{i=1}^{M-1} (-1 - (\omega_t x_i + b_t))^2 + (1 - (\omega_t t + b_t))^2 + \lambda {\omega_t}^2, \quad (6)$$

where t and x_i are the target neuron and other neurons in a single channel of the input feature $\mathbf{X} \in \mathbb{R}^{C \times H \times W}$. i is index over spatial dimension and $M = H \times W$ is the number of neurons on that channel. ω_t and b_t are weight and bias of linear transforms of t and x_i.

Improved SimAM module For our purpose of signature verification, we present an enhancement to the SimAM module aimed at optimizing the utilization of differential information between inputs. As depicted in Fig. 2(b), the feature extraction network processes two inputs to produce feature maps of identical size. Subsequently, we calculate the difference between these two feature maps and further operations. This modification endows the improved SimAM module with the capability to enhance features essential for signature verification by harnessing the disparate information present in the feature maps.

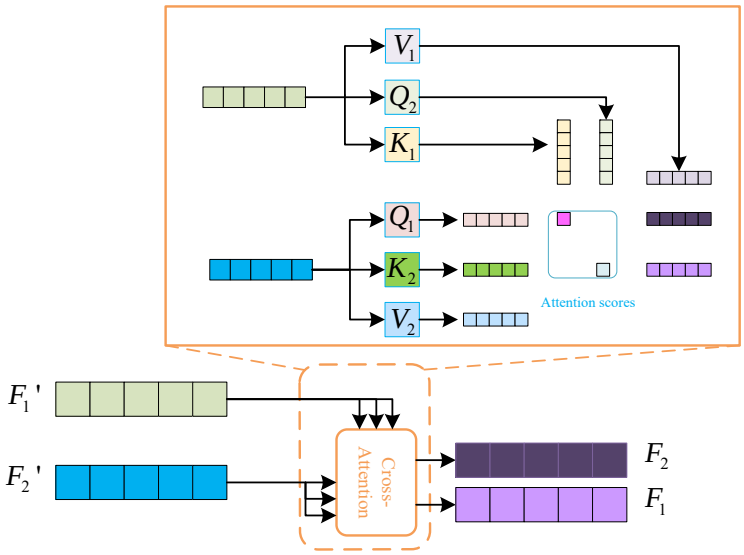


Fig. 3. The details of cross-attention.

Cross-Attention

Cross-attention [34], a mechanism integrated into the architecture of certain contemporary natural language processing (NLP) tasks, such as the Transformer model, enables one sequence to selectively focus on another. This proves beneficial in various NLP tasks, particularly in machine translation, where aligning portions of the input sequence with corresponding parts of the output sequence is essential. Cross-attention mechanism closely resembles the self-attention mechanism employed in the Transformer model; however, in the case of cross-attention, the focus is on one sequence attending to another sequence rather than attending to itself. A detailed depiction of the cross-attention module used in our network can be found in Fig. 3.

After inputting two signature images into the Deep Convolutional Siamese Network, two feature vectors are obtained as F_1' and $F_2' \in \mathbb{R}^d$, with d representing the dimensionality of the feature vector. To align F_1' and F_2', which are the feature vectors of the two branches, the key K, value V, and query Q are initially generated. Specifically, the key, value, and query vectors for F_1' are expressed as $K_1 = M(F_1')$, $V_1 = N(F_1')$, and $Q_1 = L(F_2')$, respectively. Similarly, the key, value, and query vectors for F_2' are denoted as $K_2 = M(F_2')$, $V_2 = N(F_2')$, and $Q_2 = L(F_1')$, where M, N, L are linear mappings employed to project the input sequences into a shared hidden space of the same dimensionality. The subsequent step involves attention weight computations as detailed below:

$$Attention_1 = softmax(Q_1 \cdot K_1^{\mathrm{T}}), \tag{7}$$

$$Attention_2 = softmax(Q_2 \cdot K_2^{\mathrm{T}}), \tag{8}$$

After computing attention weights, matrix multiplication is used to adjust the input sequences, resulting in cross-attention-adjusted outputs:

$$F_1 = Attention_2 \cdot V_2, \tag{9}$$

$$F_2 = Attention_1 \cdot V_1, \tag{10}$$

The signature verification task traditionally emphasized refining the network structure, with less consideration given to the issue of misalignment in feature vector pairs extracted by the network. Feature vectors with global and local features may encounter misalignment problems when directly calculating their distances. To overcome this challenge, one potential solution is cross-attention, which involves treating the pairs of feature vectors extracted by the network as two sequences and feeding them into a cross-attention mechanism. This process aligns global and local information in both sequences, producing two aligned sequences denoted as feature vectors F_1 and F_2. Subsequently, the absolute distance between these aligned feature vectors is computed, mitigating the misalignment issues.

3.4 Deep Metric Learning

We compared several distance measures, including the Cosine, Euclidean distance, and the absolute value of feature vector pairs. Our findings revealed that the "absolute value" denoted as $F = |F_1 - F_2|$ yielded the most favorable results. Following this discovery, a linear layer is incorporated to project the feature vector F into a 2-dimensional space using base vector $(\hat{p_1}, \hat{p_2})^{\mathrm{T}}$. Here, $\hat{p_1}$ signifies the anticipated probability that both signatures belong to the same user, while $\hat{p_2}$ denotes the anticipated probability of the opposite scenario ($\hat{p_1} + \hat{p_2} = 1$). This process allows for the treatment of signature verification as a binary-class classification problem, and to optimize our model, the focal contrast loss is employed as the objective function.

Contrastive loss The contrastive loss [41] can be defined as follows:

$$Loss(p, \hat{p}) = -[p \cdot \ln(\hat{p_1}) + (1 - p) \cdot \ln(\hat{p_2})] = \sum_{i=1}^{2} -p_i \cdot \ln(\hat{p_i}), \tag{11}$$

In the context of comparing two signatures to determine whether they are written by the same user or not, the target class, denoted as p, indicates the similarity between the signatures. Specifically, if the two signatures are from the same user, then $p_1 = 1$ and $p_2 = 0$; otherwise, $p_1 = 0$ and $p_2 = 1$. Therefore, the predicted probability, $\hat{p}$, is used to approximate the similarity measure, particularly using $\hat{p_1}$ to approximate the similarity between the two signatures.

Focal contrast loss Although contrastive loss has been shown effective in signature verification tasks, significant challenges still impede further enhancements in performance in end-to-end signature verification patterns, because of the serious imbalance of positive/negative samples and easy/difficult samples.

Drawing inspiration from the focal loss [42], we introduce a new loss called focal contrast Loss (FCLoss), formulated as follows:

$$FCLoss = -\alpha p^{\gamma} \log(\hat{p_1}) - (1 - \alpha)(1 - p)^{\gamma} \log(\hat{p_2}), \tag{12}$$

where α is an adjusting factor to balance the weights of positive and negative classes. γ is a tunable exponent parameter, typically taken as a positive value. It adjusts the weights between easily classified samples (p large) and challenging samples (p small).

The design of focal loss aims to improve the model's handling of class imbalance issues and difficulty sample problems. This expression incorporates the loss calculation for both positive and negative class scenarios, reducing the loss for easily classified samples by decreasing their weights. This adjustment directs the model to focus more on challenging samples, thereby enhancing the model's ability to effectively address class imbalances and difficulty samples.

4 Experimental Results

4.1 Datasets

We evaluate the signature verification performance on four datasets: CEDAR [43], BHSig-B, BHSig-H [44] and HanSig [45]. HanSig is a large-scale public Chinese handwritten signature database. Examples of these datasets are shown in Fig. 4.

CEDAR. The CEDAR dataset comprises signatures from 55 users, each contributing 24 genuine signatures and 24 skilled forgeries. This results in $C_{24}^2 = 276$ genuine-genuine pairs and $C_{24}^1 \times C_{24}^1 = 576$ genuine-forged pairs per user. For training, 50 users were selected randomly, with the remaining 5 users allocated to the testing set. The training set contains 42,600 samples, while the testing set contains 4,260 samples.

BHSig-B and BHSig-H. BHSig-B and BHSig-H datasets consist of signatures from 100 Bengali and 160 Hindi users, respectively. Each user contributed 24 genuine signatures and 30 forgeries. This yields 276 genuine-genuine pairs and 720 genuine-forged pairs per individual. BHSig-B was split into training and testing sets: 80 users for training (79,680 samples) and 20 users for testing (19,920 samples). BHSig-H followed a similar split: 100 users for training (99,600 samples) and 60 users for testing (59,760 samples). This partitioning ensures thorough model evaluation across different user signatures.

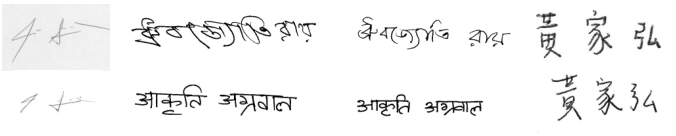


Fig. 4. Examples of signature images in CEDAR, BHSig-B, BHSig-H, and HanSig with genuine signatures in the first row and corresponding forged signatures in the second row.

HanSig. The HanSig dataset is a collection of 885 candidate names, chosen based on the frequency distributions of name occurrences in real-world contexts, and signatures have been gathered from 238 writers. Each name has been signed 20 times in three styles – neat, normal, and stylish – to introduce greater signing variability. Consequently, the dataset consists of 17,700 genuine signatures and an equivalent number of skilled forgeries. This allows for the creation of 190 genuine-genuine pairs of signatures as positive samples and 400 genuine-forged pairs as negative samples for each individual (computed as $C_{20}^2 = 190$ and $C_{20}^1 \times C_{20}^1 = 400$). The dataset has been randomly divided into a training set, which includes 795 names, and a test set, which consists of 90 names. The training set

contains 469,050 samples, while the test set comprises 53,100 samples, enabling an extensive evaluation of the dataset.

4.2 Evaluation Metrics

The signature verification performance is measured in three indices: False Acceptance Rate (FAR), False Rejection Rate (FRR) and Accuracy (Acc). FAR represents the ratio of false acceptances to all forged samples, while FRR reflects the ratio of false rejections to all genuine samples. Accuracy is the ratio of correctly predicted samples to all predicted samples. The EER, a widely used metric in biometric system evaluation, indicates the point where FRR equals FAR. The specific calculations for these metrics are as follows:

$$FAR = \frac{FP}{TN + FP}, \tag{13}$$

$$FRR = \frac{FN}{TP + FN}, \tag{14}$$

$$Acc = \frac{TP + TN}{TN + FN + TP + FP}, \tag{15}$$

where TP, TN, FN, FP are defined as follows::

True Positive (TP): Number of genuine signatures predicted as genuine.

True Negative (TN): Number of forged signatures predicted as forgeries.

False Negative (FN): Number of genuine signatures predicted as forgeries.

False Positive (FP): Number of forged signatures predicted as genuine signatures.

4.3 Experimental Settings

We implement the proposed model using the PyTorch deep learning framework, employing the widely adopted Adam optimizer [46], configured with a learning rate set to 0.001 to facilitate effective convergence during training. The parameters of the FCLoss are set to $\alpha = 0.25$, $\gamma = 2$. To strike a balance between computational efficiency and model convergence, each training iteration involves mini-batches of 64 pairs of signature images. To enhance the model's generalization capability and mitigate overfitting, a dropout rate of 0.1 is applied to the linear layer as a regularization technique. This dropout mechanism randomly drops a fraction of the connections during training, preventing the model from relying too heavily on specific pathways and enhancing its robustness. The experiments are performed on a dedicated workstation equipped with a formidable 12GB Nvidia GeForce TITAN Xp GPU, which accelerates the training process, enabling swift computation of complex model operations. The system's efficiency is further underscored by its remarkable speed in signature verification, taking an average of only 10 milliseconds to process a pair of signatures. This efficient execution is crucial for real-time applications, where rapid decision-making is imperative. The system takes only 10 ms to verify a pair of signatures on average.

4.4 Results and Discussions

The signature verification results are shown in Table 1, with a comparative analysis of the proposed method with networks based on CNNs and Transformers. Our proposed method and baseline both achieve state-of-the-art results on the CEDAR dataset. TransOSV has achieved state-of-the-art performance on the BHSig-B dataset with a transformer as a holistic encoder and CNNs as a part decoder. It is worth noting that our proposed method has outperformed all approaches that only use CNNs. Besides, on the BHSig-H dataset, our method has demonstrated superior performance compared to the similar CNNs+Attention network, MSN+Attention. TransOSV achieved a state-of-the-art performance, which is because they used a pre-trained ViT-based model. When evaluated on the HanSig dataset, our method outperforms MGRNet, which utilizes a multi-scale global and regional feature learning network with co-tuplet loss. Additionally, it is worth noting that our proposed method has the added advantage of being adaptable for any newly added writer without requiring system retraining. In a word, our proposed method outperforms all the compared methods on CEDAR, BHSig-B, and HanSig databases.

Table 1. Comparison of signature verification performance on the four datasets (%). "DenseNet+cross-writer attention" is the proposed method.

Dataset	Method	FRR	FAR	Acc
CEDAR	MSN+Attention [6]	0	3.18	98.41
	MGRNet [45]	3.55	3.33	96.56
	MSDN [22]	6.74	6.74	93.26
	DenseNet+cross-writer attention	**0**	**0**	**100**
BHSig-B	MSN+Attention [6]	6.44	10.42	91.56
	MGRNet [45]	6.20	5.93	93.93
	ViT(Pre-trained) [23]	18.48	18.48	81.52
	T2T-vit(Pre-trained) [39]	10.83	5.08	93.33
	TransOSV [9]	3.56	3.56	96.44
	DenseNet+cross-writer attention	**2.14**	2.95	**97.27**
BHSig-H	MSN+Attention [6]	5.16	17.06	88.88
	MGRNet [45]	6.56	6.76	93.34
	ViT(Pre-trained) [23]	20.18	20.18	79.82
	T2T-vit(Pre-trained) [39]	24.23	9.94	86.10
	TransOSV [9]	**3.24**	**3.24**	**96.76**
	DenseNet+cross-writer attention	12.14	7.11	91.42
HanSig	MGRNet [45]	**7.69**	11.85	90.23
	T2T-vit(Pre-trained) [39]	38.56	10.69	80.34
	DenseNet+cross-writer attention	16.33	**6.61**	**90.26**

4.5 Ablation Studies

Effectiveness of SimAM module We evaluate the model with the SimAM module removed to test its impact on signature verification performance. The obtained results of these ablation studies are reported in Table 2. The findings indicate that the removal of the SimAM module resulted in performance degradation, with a reduction of 2.10 percentage points on the HanSig dataset. This implies that our improved SimAM module effectively emphasizes differences between feature pairs. Furthermore, we conduct experiments by incorporating different self-attention mechanisms into our baseline to demonstrate the effectiveness of the mechanism we used. The experimental results are shown in Table 3. The results show that the proposed improved SimAM module outperforms SE and CBAM on the datasets except BHSig-H.

Effectiveness of cross-attention In our investigation of the significance of the cross-attention module in our deep metric learning approach, we found that its incorporation resulted in a noticeable improvement in signature verification performance, as evidenced by the results presented in Table 2. Our analysis indicates that the cross-attention module effectively addresses the issue of misalignment in feature vectors extracted from the Siamese Network. Specifically, it aligns the global information and local features between vector pairs, thereby mitigating the potential negative impacts of misalignment on performance.

Effectiveness of FCLoss The effects of the FCLoss are validated through a comparison with the original contrastive loss. The results of this comparison are displayed in Table 4. It is evident from the results that the FCLoss outperforms the CE (Cross-Entropy) loss on our method. These findings provide compelling evidence that focal contrast loss effectively prioritizes hard samples, leading to superior performance.

Table 2. Ablation study on the HanSig dataset to evaluate each module of the proposed cross-writer attention (Acc in %).

Model	**SimAM**	**cross-attention**	**Acc**
Baseline	-	-	87.48
	✓	-	89.81
	-	✓	88.16
Our method	✓	✓	**90.26**

Table 3. Performances of different self-attention module (%).

	SE			**CBAM**			**SimAM**		
	FRR	**FAR**	**Acc**	**FRR**	**FAR**	**Acc**	**FRR**	**FAR**	**Acc**
CEDAR	0.00	0.00	100.0	0.00	0.00	100.0	0.00	0.00	100.0
BHSig-B	**3.68**	2.62	97.09	4.38	2.55	96.94	3.77	**2.30**	**97.29**
BHSig-H	17.30	6.11	**90.79**	**17.00**	6.72	90.43	24.21	**5.88**	89.76
HanSig	22.35	11.01	85.34	22.01	9.05	86.78	**13.91**	**8.42**	**89.81**

Table 4. Effectiveness of FCLoss (%).

Loss	**CELoss**			**FCLoss**		
	FRR	**FAR**	**Acc**	**FRR**	**FAR**	**Acc**
CEDAR	0.00	0.00	100.0	0.00	0.00	100.0
BHSig-B	11.34	3.08	94.63	**9.31**	3.10	**95.18**
BHSig-H	15.18	8.31	89.78	18.83	**6.00**	**90.45**
HanSig	30.33	10.36	83.21	**17.84**	**9.99**	**87.48**

5 Conclusions

This paper proposes a novel offline signature verification method utilizing a Deep Convolutional Siamese Network with cross-writer attention. To enhance the discrimination ability, an improved SimAM module is proposed to be inserted between the Siamese Network, thereby rendering the networks more attentive to the differences between the feature maps. Furthermore, the proposed deep metric learning module with cross-attention addresses the problem of feature misalignment between feature vectors. Experimental results on benchmark datasets CEDAR, BHSig-B, and HanSig demonstrate the effectiveness of the proposed method, achieving accuracy rates competitive with state-of-the-art methods. We also plan to test the feasibility of cross-writer attention in online signature verification task.

Acknowledgements

This work has been supported by the National Key Research and Development Program under Grant No. 2022YFC3301703 and the National Natural Science Foundation of China (NSFC) grant 62136001.

References

1. C.V Aravinda, Lin Meng, K.R Uday Kumar Reddy, and Amar Prabhu: Signature recognition and verification using multiple classifiers combination of Hu's and HOG features. International Conference on Advanced Mechatronic Systems, 2019: 63-68.
2. Bhushan S. Thakare, and Hemant R. Deshmukh: A combined feature extraction model using SIFT and LBP for offline signature verification system. International Conference for Convergence for Technology (2018): 1-7.
3. Li Wen Goon, and Swee Kheng Eng: Offline signature verification system using SVM classifier with image pre-processing steps and SURF algorithm. Journal of Physics: Conference Series 2107.1 (2021).
4. Kamran Shaukat, Suhuai Luo, Vijay Varadharajan: A novel method for improving the robustness of deep learning-based malware detectors against adversarial attacks. Eng. Appl. Artif. Intell. 116: 105461 (2022).
5. Talha Mahboob Alam, Kamran Shaukat, Ibrahim A. Hameed, Wasim Ahmad Khan, Muhammad Umer Sarwar, Farhat Iqbal, Suhuai Luo: A novel framework for prognostic factors identification of malignant mesothelioma through association rule mining. Biomed. Signal Process. Control. 68: 102726 (2021).
6. Yu-Jie Xiong, Song-Yang Cheng: Attention based multiple Siamese network for offline signature verification. ICDAR 2021: 337-349.
7. Muhammad Imran Malik, Marcus Liwicki, Andreas Dengel, Seiichi Uchida, Volkmar Frinken: Automatic signature stability analysis and verification using local features. ICFHR 2014: 621-626
8. Muhammad Sharif, Muhammad Attique Khan, Muhammad Faisal, Mussarat Yasmin, Steven Lawrence Fernandes: A framework for offline signature verification system: best features selection approach. Pattern Recognit. Lett. 139: 50-59 (2020)
9. Huan Li, Ping Wei, Zeyu Ma, Changkai Li, Nanning Zheng: TransOSV: offline signature verification with transformers. Pattern Recognit. 145: 109882 (2024).
10. Michael Stauffer, Paul Maergner, Andreas Fischer, Rolf Ingold, Kaspar Riesen: Offline signature verification using structural dynamic time warping. ICDAR 2019: 1117-1124.
11. H. Baltzakis, N. Papamarkos: A new signature verification technique based on a two-stage neural network classifier, Eng. Appl. Artif. Intell. 14 (1): 95–103 (2001).
12. Mustafa Berkay Yilmaz, Berrin A. Yanikoglu, Caglar Tirkaz, Alisher Kholmatov: Offline signature verification using classifier combination of HOG and LBP features. IJCB 2011: 1-7.
13. Yasmine Guerbai, Youcef Chibani, Bilal Hadjadji: The effective use of the one-class SVM classifier for handwritten signature verification based on writer-independent parameters. Pattern Recognit. 48(1): 103-113 (2015).
14. Manabu Okawa: Synergy of foreground-background images for feature extraction: offline signature verification using Fisher vector with fused KAZE features. Pattern Recognit. 79: 480-489 (2018).
15. Juan Hu, Youbin Chen: Offline signature verification using real adaboost classifier combination of pseudo-dynamic features. Proceedings of the International Conference on Document Analysis and Recognition, 2013: 1345-1349.
16. L.G. Hafemann, R. Sabourin, L.S. Oliveira, Learning features for offline handwritten signature verification using deep convolutional neural networks, Pattern Recognit. 70: 163–176 (2017).
17. Xi Lu, Linlin Huang, Fei Yin: Cut and compare: end-to-end offline signature verification network. ICPR 2020: 3589-3596.

18. Jane Bromley, Isabelle Guyon, Yann LeCun, Eduard Säckinger, Roopak Shah: Signature verification using a Siamese time delay neural network. NIPS 1993: 737-744.
19. Sounak Dey, Anjan Dutta, Juan Ignacio Toledo, Suman K. Ghosh, Josep Lladós, Umapada Pal: SigNet: convolutional Siamese network for writer independent offline signature verification. CoRR abs/1707.02131 (2017).
20. Victoria Ruiz, Ismael Linares, Ángel Sanchez, Jose Francisco Velez: Off-line handwritten signature verification using compositional synthetic generation of signatures and Siamese Neural Networks. Neurocomputing 374: 30-41 (2020).
21. Ping Wei, Huan Li, Ping Hu: Inverse discriminative networks for handwritten signature verification. CVPR 2019: 5764-5772.
22. Li Liu, Linlin Huang, Fei Yin, Youbin Chen: Offline signature verification using a region based deep metric learning network. Pattern Recognit. 118: 108009 (2021).
23. Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, Neil Houlsby: An image is worth 16x16 Words: transformers for image recognition at scale. ICLR 2021.
24. Luiz G. Hafemann, Robert Sabourin, Luiz S. Oliveira: Characterizing and evaluating adversarial examples for offline handwritten signature verification. IEEE Trans. Inf. Forensics Secur. 14(8): 2153-2166 (2019).
25. Huan Li, Ping Wei, Ping Hu: AVN: an adversarial variation network model for handwritten signature verification. IEEE Trans. Multim. 24: 594-608 (2022).
26. Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, Illia Polosukhin: Attention is all you need. NIPS 2017: 5998-6008.
27. Jie Hu, Li Shen, Samuel Albanie, Gang Sun, Enhua Wu: Squeeze-and-excitation networks. IEEE Trans. Pattern Anal. Mach. Intell. 42(8): 2011-2023 (2020).
28. Yue Cao, Jiarui Xu, Stephen Lin, Fangyun Wei, Han Hu: Global context networks. IEEE Trans. Pattern Anal. Mach. Intell. 45(6): 6881-6895 (2023).
29. Xilai Li, Wei Sun, Tianfu Wu: Attentive normalization. ECCV (17) 2020: 70-87.
30. Sanghyun Woo, Jongchan Park, Joon-Young Lee, In So Kweon: CBAM: convolutional block attention module. ECCV (7) 2018: 3-19.
31. Lingxiao Yang, Ru-Yuan Zhang, Lida Li, Xiaohua Xie: SimAM: a simple, parameter-free attention module for convolutional neural networks. ICML 2021: 11863-11874.
32. Wenhai Wang, Enze Xie, Xiang Li, Deng-Ping Fan, Kaitao Song, Ding Liang, Tong Lu, Ping Luo, Ling Shao: Pyramid vision transformer: a versatile backbone for dense prediction without convolutions. ICCV 2021: 548-558.
33. Haiping Wu, Bin Xiao, Noel Codella, Mengchen Liu, Xiyang Dai, Lu Yuan, Lei Zhang: CvT: introducing convolutions to vision transformers. ICCV 2021: 22-31.
34. Dzmitry Bahdanau, Kyunghyun Cho, Yoshua Bengio: Neural machine translation by jointly learning to align and translate. ICLR 2015.
35. Alex Krizhevsky, Ilya Sutskever, Geoffrey E. Hinton: ImageNet classification with deep convolutional neural networks. Commun. ACM 60(6): 84-90 (2017).
36. Karen Simonyan, Andrew Zisserman: Very deep convolutional networks for large-scale image recognition. ICLR 2015.
37. Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun: Deep residual learning for image recognition. CVPR 2016: 770-778
38. Gao Huang CVPR 2017 Gao Huang, Zhuang Liu, Kilian Q. Weinberger: Densely connected convolutional networks. CVPR 2017: 2261-2269

39. Li Yuan, Yunpeng Chen, Tao Wang, Weihao Yu, Yujun Shi, Francis E. H. Tay, Jiashi Feng, Shuicheng Yan: Tokens-to-token ViT: training vision transformers from scratch on ImageNet. ICCV 2021: 538-547
40. Dan Hendrycks, Kevin Gimpel: Bridging Nonlinearities and Stochastic Regularizers with Gaussian Error Linear Units. CoRR abs/1606.08415 (2017).
41. Raia Hadsell, Sumit Chopra, Yann LeCun: Dimensionality reduction by learning an invariant mapping. CVPR (2) 2006: 1735-1742.
42. Tsung-Yi Lin, Priya Goyal, Ross B. Girshick, Kaiming He, Piotr Dollár: Focal loss for dense object detection. ICCV 2017: 2999-3007.
43. Meenakshi K. Kalera, Sargur N. Srihari, Aihua Xu: Offline signature verification and identification using distance statistics. Int. J. Pattern Recognit. Artif. Intell. 18(7): 1339-1360 (2004).
44. Srikanta Pal, Alireza Alaei, Umapada Pal, Michael Blumenstein: Performance of an off-line signature verification method based on texture features on a large indic-script signature dataset. DAS 2016: 72-77.
45. Fu-Hsien Huang, Hsin-Min Lu: Multiscale global and regional feature learning using co-tuplet loss for offline handwritten signature verification. CoRR abs/2308.00428 (2023).
46. Diederik P. Kingma, Jimmy Ba: Adam: a method for stochastic optimization. ICLR (Poster) 2015.