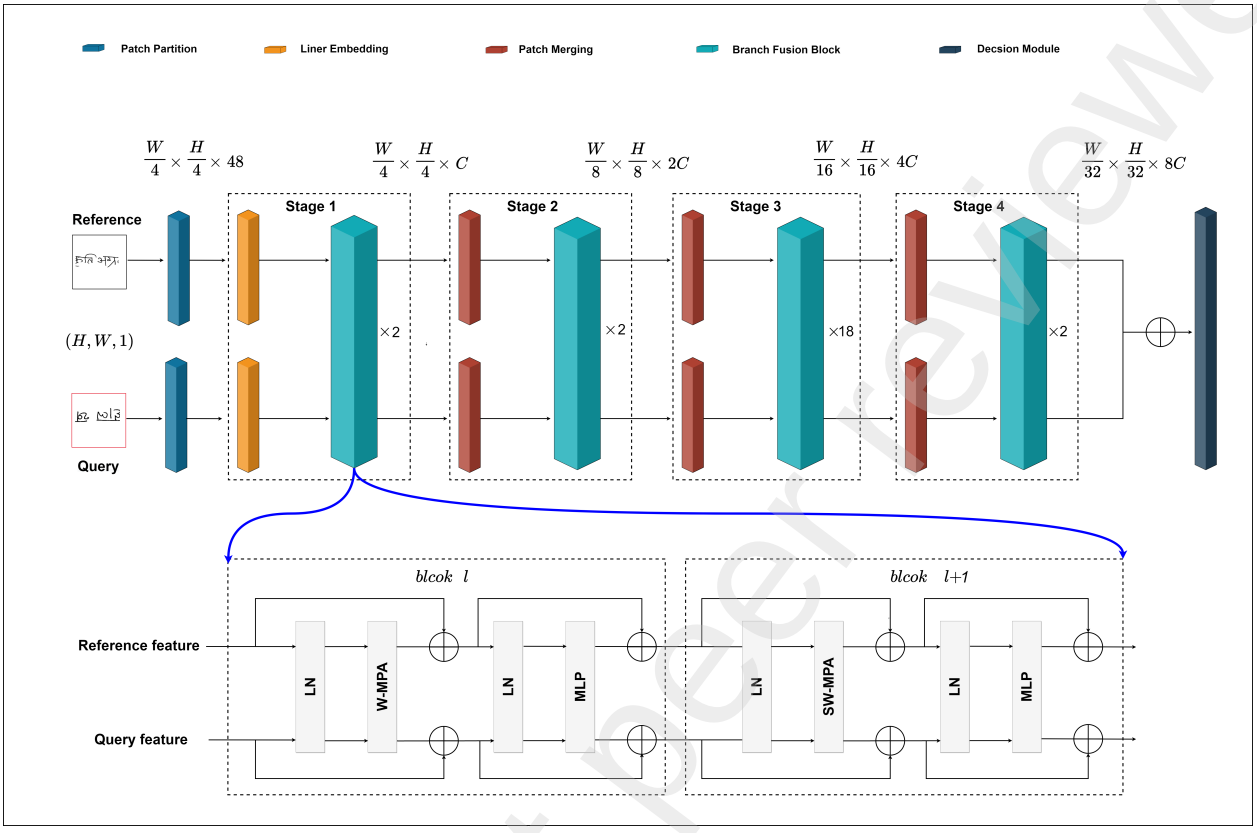
Graphical Abstract

PAST: Pairwise Attention Swin Transformer for Offline Signature Verification

Yu-Jie Xiong, Jian-Xin Ren, Dong-Hai Zhu, Xi-Jiong Xie, Xi-He Qiu



Highlights

PAST: Pairwise Attention Swin Transformer for Offline Signature Verification

Yu-Jie Xiong, Jian-Xin Ren, Dong-Hai Zhu, Xi-Jiong Xie, Xi-He Qiu

- We propose a Pairwise Attention (PA) mechanism that is highly efficient for pairwise signature verification. Pairwise attention facilitates bidirectional information exchange between reference and query signatures without introducing any additional assumptive temporal information.
- We adopt the $\mathcal{Q} - \mathcal{R}$ branches approach to establish input symmetry, ensuring that the input order is not affected for both sequences. This innovated method involves leveraging the $\mathcal{Q}$ and $\mathcal{R}$ branches to create a balanced and symmetrical input structure, thereby preserving the integrity of the input order in both reference and query sequences.
- We conduct extensive comparative and ablation experiments, demonstrating that our proposed method significantly outperforms other state-of-the-art (SOTA) methods on existing datasets. This indicates the generalizability of our approach to signature verification across different languages. Furthermore, we investigated the influence of background factors in the CEADAR dataset.

PAST: Pairwise Attention Swin Transformer for Offline Signature Verification

Yu-Jie Xiong[a,*,1], Jian-Xin Ren[a,1], Dong-Hai Zhu[a], Xi-Jiong Xie[b] and Xi-He Qiu[a]

[a]School of Electronic and Electrical Engineering, Shanghai University of Engineering Science, 333 Longteng Road, Songjiang District, Shanghai, 201620, Shanghai, China
[b]School of Information Science and Engineering, Ningbo University, 818 Fenghua Road, Jiangbei District, 315211, Ningbo, Zhejiang, China

ARTICLE INFO

Keywords:
Offline Signature Verification
Symmetry of Input
Swin Transformer

ABSTRACT

Signature verification has shown tremendous potential as a reliable biometric in both academic research and industrial applications. With the advent of deep learning, signature verification has made remarkable progress in the past decade. However, despite significant progress, the challenge of detecting subtle differences between genuine and forged signatures, leading to concerns over privacy protection and data security of the signature verification system. Recently, the tremendous success of transformers in Natural Language Processing has led to their extension to computer vision, resulting in significant advancements. The multi-head self-attention mechanism is considered crucial for the success of transformer. As the name implies, its query, key, and value all originate from the same sequence, rendering it suitable for single input tasks. However, pairwise signature verification treats reference and query signature images equally as two independent inputs. Regarding this matters, the mere amalgamation of the two independent inputs in the form of a single sequence inevitably gives rise to potential inherent issues. To tackle this problem, we present a Pairwise Attention (PA) mechanism that keeps the symmetry of inputs. Unlike the original attention mechanism, pairwise attention facilitates bidirectional information exchange between reference and query signatures without introducing any additional assumptive temporal information. Subsequently, combing with the architecture of Swin Transformer, we propose Pairwise Attention Swin Transformer(PAST). Our method fundamentally solves the problem of introducing false assumptive temporal information during the process of input fusion, but also performs impressively on several public datasets. Experimental results show that PAST outperform most existing methods. In addition, we investigated the impact of background information from the CEDAR database on the results. The study revealed that including background information in the training data significantly improved the results compared to when background information was not included.

1. Introduction

Pairwise signature verification plays an essential role in biometrics. Instead of comparing a single signature against a writer-dependent reference model, as the traditional signature verification, pairwise signature verification involves comparing two signatures equally to determine whether they are generated by the identical writer.

Signature verification is a vital area in biometrics with broad practical applications, including finance, justice, insurance, and criminal investigations [14]. In particular, incorporating swarm intelligence-based task scheduling [38] and learning-based cloud server configuration [6], the consideration of handwritten signature authentication is explored, providing an additional layer of security for IoT devices. However, it is a challenge due to difficulties in signature sample collection, sparse features, and small interclass variability. Moreover, changes in the writing style of the same person over time further complicate the task. The objective of signature verification is to distinguish between genuine signatures and forged ones. According to the degree of imitation, there are three different types of forged signatures: random forgery, simple forgery, and skilled forgery [11]. Random forgery involves using a signature sample from a different individual, whereas simple forgery denotes a signature sample that mimics the handwriting of the genuine author's name. On the other hand, skilled forgery entails a practiced imitation of the authentic signature.

Depending on the signature acquisition mode, signature verification can be categorized into two types: online and offline [18]. The online approach is not widely available due to the complex application scenario and the requirement of specialized equipment, although it contains abundant state and positional information such as location, velocity, and pressure. On the other hand, offline signature verification involves obtaining signature samples by scanning or photographing paper documents. The offline approach has the advantages of low equipment requirements and content limits compared to the online approach, making it more practical for a broader range of applications and research studies. However, the lack of dynamic information in offline signatures poses a challenge in achieving good verification performance [10]. This paper focuses mainly on the offline

*Corresponding author
xiong@sues.edu.cn (Y. Xiong); Suesrenjianxin@outlook.com (J. Ren); m325123217@sues.edu.cn (D. Zhu); xiexijiong@nbu.edu.cn (X. Xie); qiuxihe1993@gmail.com (X. Qiu)
ORCID(s): 0000-0002-2769-022X (Y. Xiong); 0000-0003-1629-8422 (J. Ren); 0009-0000-7599-8045 (D. Zhu); 0000-0002-5288-1861 (X. Xie); 0000-0003-4024-925X (X. Qiu)
[1]Co-first authors

approach. For convenience, all signature verification methods mentioned here refer to offline signature verification unless otherwise specified.

Nowadays, the prevailing approaches for signature verification often simplify multiple one-to-one match into a one-to-many recognition or classification. This simplification works well in scenarios involving single-input verification and enhances the system speed. However, when it involves pairwise inputs, it poses a serious issue that is easily overlooked. In theory, an ideal verification model should yield identical output results for pairwise inputs $(\mathcal{A}, \mathcal{B})$ and pairwise inputs $(\mathcal{B}, \mathcal{A})$. However, for the multiclassification model, pairwise inputs $(\mathcal{A}, \mathcal{B})$ (denoted as $\{a_1, \ldots, a_N\}, \{b_1, \ldots, b_N\}$) are often treated as two-channel images, which are concatenated to form a composite tensor (denoted as $\{a_1, \ldots, a_N, b_1, \ldots, b_N\}$) for subsequent computations. This process introduces additional sequence information that is not originally present. As a result, with the invented temporal sequence, the symmetry of pairwise inputs is disrupted ($\{a_1, \ldots, a_N, b_1, \ldots, b_N\} \neq \{b_1, \ldots, b_N, a_1, \ldots, a_N\}$), leading to potential differentials in output results, which are unreasonable and unacceptable.

To address this issue, we propose an pairwise-attention mechanism. It is performed to steer the model to constantly exchange information between reference and query signatures for producing the new respective feature maps. The new feature maps contain not only the characteristics of reference and query signatures but also reflect their importance relative to each other. As a result, it is more applicable to signature verification. Our experiments on five datasets demonstrate that Pairwise Attention Swin Transformer achieves significant performance enhancements compared to existing methods, making it a promising approach for signature verification. The main contributions of this paper are as follows:

- We propose a Pairwise Attention (PA) mechanism that is highly efficient for pairwise signature verification. Pairwise attention facilitates bidirectional information exchange between reference and query signatures without introducing any additional assumptive temporal information.
- We adopt the $\mathcal{Q} - \mathcal{R}$ branches approach to establish input symmetry, ensuring that the input order is not affected for both sequences. This innovated method involves leveraging the $\mathcal{Q}$ and $\mathcal{R}$ branches to create a balanced and symmetrical input structure, thereby preserving the integrity of the input order in both reference and query sequences.
- We conduct extensive comparative and ablation experiments, demonstrating that our proposed method significantly outperforms other state-of-the-art (SOTA) methods on existing datasets. This indicates the generalizability of our approach to signature verification across different languages. Furthermore, we investigated the influence of background factors in the CEADAR dataset.

2. Related Work

2.1. Signature Verification

Signature verification is a complex task that involves differentiating genuine signatures from forged ones. Traditional approaches typically consist of three main steps: preprocessing, feature extraction, and classification. Preprocessing is the first step in signature verification, where operations such as edge detection, binarization, and skew correction are performed to preprocess the raw signature images and obtain a more suitable form. Feature extraction plays a vital role in signature verification, as it aims to learn relevant representations that can effectively distinguish complex stroke features. In the final step, learning based classifiers are used to determine whether a signature is genuine or forged, based on extracted features. Traditional handcrafted feature-based methods heavily rely on the prior knowledge, struggling to identify signatures written in different styles or orientations. Deep learning methods can automatically learn discriminative features from raw data, including signatures with varying styles and structures.

There are two mainstream methods for signature verification in the field of deep learning: Siamese-network-based and two-channel-based architectures. Siamese neural networks, initially proposed by Bromley et al. [2], employ a weight-sharing mechanism where two identical subnetworks share parameters. They are particularly useful for comparing the similarity or dissimilarity between two inputs. One of the key advantages of the Siamese architecture is its order independence, as swapping the order of the signatures being compared does not impact the final result. In the context of signature verification, Siamese networks can assess the similarity between two signatures. For example, Xiong et al. [36] proposed Multiple Siamese Network (MSN) with four parallel branches, incorporating an attention module to extract salient features from handwriting. Similarly, Ghosh and Rajib [9] employed a two-branch Recurrent Neural Network (RNN) with Long Short-Term Memory (LSTM) to extract rich structural and directional features for signature verification. Likewise, Shen et al. [27] combined a one-dimensional Multi-scale Residual-based Siamese Neural Network (1D-MSNet) and adaptive boosting softmax classification, makeing the network pay more attention to the information of important feature sequences. Additionally, Victoria Ruiz et al. [25] used Siamese Neural Networks to address off-line handwritten signature verification with random forgeries, augmenting the training set with synthetic data.

2.2. Vision Transformers

The Transformer architecture, originally developed for natural language processing, has been adapted for computer vision tasks. Vision Transformer (ViT) [8] is one of the most significant developments in this area. Several approaches have emerged in the area of Vision Transformers. The Data-efficient Image Transformer (DeiT) [31] aims to reduce reliance on large datasets through distillation techniques and has achieved state-of-the-art performance on benchmark

datasets like ImageNet. The Swin Transformer [19] is a hierarchical ViT that reduces computation by partitioning windows and has shown competitive performance on various benchmarks. RegionViT [4] utilizes region-based attention mechanisms to capture spatial dependencies in images and outperforms other methods on benchmark datasets. DeepViT [37] a depth-wise block to replace the traditional multi-head self-attention mechanism, resulting in improved accuracy and reduced parameters. Pyramid ViT [33] proposes a pyramid-style architecture with multi-scale feature representation and performs well on benchmarks, especially for tasks involving smaller objects. The evolution of Vision Transformers is an exciting and innovative research area in computer vision, offering the potential to propel the field forward in new and groundbreaking ways. In the field of signature verification, vision transformers are playing an increasingly important role. Li et al. [17] proposes a model based on vision transformers, TransOSV, which significantly enhances offline signature verification by effectively integrating global and discriminative local features. Chu et al. [5] proposes a novel Multi-Size Assembled-Attention Swin-Transformer network that leverages self-attention and cross-attention mechanisms for authenticating signature handwriting. Wei et al. [34] proposes the inverse discriminative network (IDN) for handwritten signature verification, employing a novel multi-path attention mechanism across discriminative and inverse streams to enhance focus on signature strokes. However, there are still several challenges that need to be addressed, such as improving the interpretability and robustness of Vision Transformers, developing more efficient training and optimization techniques, and exploring their potential for other computer vision tasks beyond object recognition.

In the task of signature verification, the self-attention mechanism has certain limitations. The original self-attention mechanism is designed to handle individual input sequences and may face challenges when dealing with paired signature inputs.These limitations include: (1) The self-attention mechanism treats each element in the input sequence as queries, keys, and values, and computes attention scores between them. However, this unidirectional attention mechanism is limited in capturing the interdependencies between input sequences. In signature verification, where the mutual dependencies and correlations between signatures are crucial, the self-attention mechanism falls short in directly modeling such interdependencies. (2) The self-attention mechanism independently processes each input sequence without direct information exchange or communication. As a result, important spatial or sequential patterns between input signatures may not be effectively utilized during the feature extraction stage. This limitation hampers the self-attention mechanism's ability to fully leverage significant spatial or sequential patterns in signature verification tasks, leading to suboptimal feature extraction. In contrast, our proposed Pairwise Attention Swin Transformer (PAST) method addresses these limitations by introducing the pairwise-attention mechanism.

To address the issues caused by the framework and self-attention mentioned above, we propose the Pairwise Attention Swin Transformer (PAST). It introduces the pairwise-attention mechanism, which allows both inputs to participate in the attention mechanism and better focuses on their correlation. This attention mechanism allows the model to selectively focus on different regions of the input feature image and facilitate information exchange between input feature signatures during the feature extraction process. Compared to unidirectional self-attention mechanism's, the pairwise attention mechanism captures the interdependence between input sequences more effectively, improving the performance of signature verification tasks. Additionally, PAST facilitates information exchange and weight sharing between input feature signatures, enhancing feature extraction by utilizing spatial and sequential patterns within input signatures. PAST improves the handling of correlation and dependency between input signatures in signature verification by using the pairwise-attention mechanism. In comparison, we developed a Swin-Siamese and a Swin-2-channel architecture using the Swin Transformer. Our results from several signature datasets demonstrate that PAST outperforms these models and traditional self-attention mechanisms, showcasing its efficacy in handling dependencies and spatial-sequential patterns more effectively. This innovation not only optimizes feature extraction but also mitigates the negative impacts of signature order variations inherent in two-channel systems.

3. Method

We present Pairwise Attention Swin Transformer (PAST), with a designed architecture for signature verification, which utilizes the pairwise-attention mechanism to keep the symmetry of pairwise inputs. An overview of the Pairwise Attention Swin Transformer (PAST) is presented in Figure 1. The network comprises two weight-shared branches, $\mathcal{R}$ and $\mathcal{Q}$, dedicated to processing reference and query signatures, respectively. For the $\mathcal{R}$-branch, an input reference signature with a size of $H \times W \times 1$ is first split into non-overlapping patches with 4×4 pixels by the patch partition module, transforming the inputs into sequence embedding. During this process, the dimension of the feature map is extended to 48, and the feature spatial is reduced by 16×.

Our proposed pairwise attention swin transformer draws structural inspiration from the Swin Transformer, maintaining the use of four stages for propagation. In the Stage 1, each patch undergoes linear projection to a higher-dimensional embedding space using a learnable linear projection matrix (linear embedding). The resulting patch embeddings, with dimensions $(\frac{H \times W}{16}, C)$, where C is the embedding dimension, are then processed by a series of branch fusion blocks. Each block involves a multi-head pairwise-attention mechanism followed by a position-wise feed-forward network. This mechanism facilitates bidirectional information transfer between reference and query signatures, while the feed-forward network introduces non-linearity to the attended

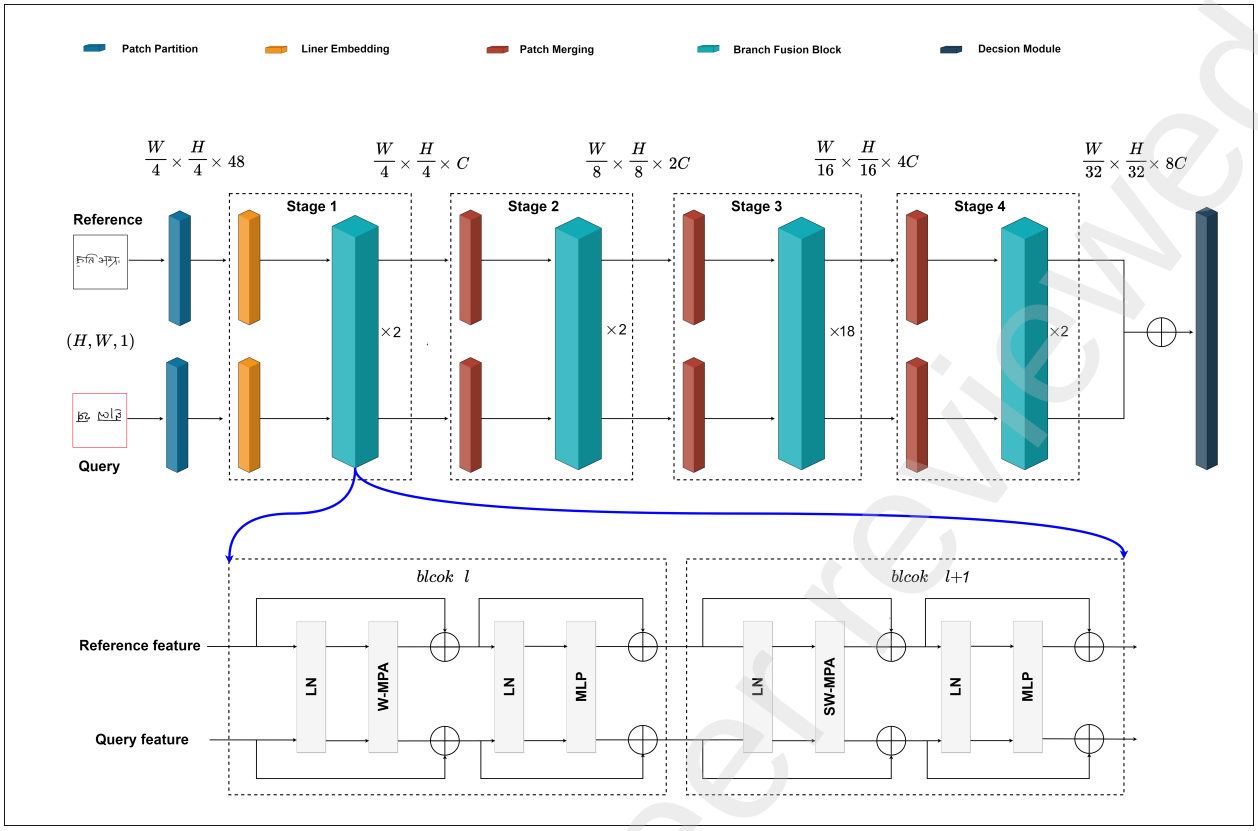


Figure 1: The structure of the proposed Pairwise Attention Swin Transformer.

features. Each block outputs a sequence of patch embeddings with the same dimension as the input, resulting in a feature map with reduced spatial resolution and increased feature dimension. Stage 2 is akin to Stage 1, but with the patch partition module replaced by a patch merging module. This module aggregates neighboring patches, producing a smaller feature map with increased spatial resolution and reduced feature dimension. The feature map is then processed by another set of branch fusion blocks to extract more complex features. In Stage 3, the input feature map is again processed by a patch partition module to obtain a new set of patches. These patches are then processed by a set of branch fusion blocks to extract multi-scale features. The final stage of the proposed architecture is similar to Stage 3, except that the number of the branch fusion blocks. The output of this stage is a set of high-level features that can be fused with the output of the $\mathcal{Q}$-branch and eventually fed into the decision layer for determining whether two signatures belong to the same person.

In contrast to the Swin Transformer, PAST diverges significantly in the following aspects: (1) PAST employs $\mathcal{Q}-\mathcal{R}$ branches approach to ensure input symmetry, providing a foundational possibility for both input orders: reference signature + query signature and query signature + reference signature (as the fusion in the early stage makes it impossible to distinguish later on); (2) In the case of independent inputs, our proposed pairwise attention is utilized to achieve symmetric attention computation; (3) Building upon the discussed aspects, the branch fusion block is formulated by combining these elements. Within the block, features from the $\mathcal{Q}-\mathcal{R}$ branches undergo processing through LayerNorm (LN) layers, Window Multi-head Pairwise-Attention (W-MPA), Multilayer Perceptron (MLP), and Shift Window Multi-head Pairwise-Attention (SW-MPA), each using equal weight.

3.1. Patch partition

In the PAST architecture, the step of patch partitioning involves segmenting the input image, which is represented as $I \in \mathbb{R}^{H \times W \times C}$, where H, W, and C respectively represent the image's height, width, and number of channels. This segmentation process transforms the image into a collection of smaller, discrete patches, each measuring $P \times P$, arranged to ensure there is no overlap, thus preparing the image for sophisticated analysis. This segmentation creates a grid of fixed-size patches, each acting as an individual unit, streamlining the image's complex structure into a simplified, organized collection of data points.

3.2. Linear embedding

After the input image is partitioned into smaller, non-overlapping patches of fixed dimensions $P \times P$, each patch

undergoes a transformation through linear embedding. Linear embedding is achieved by applying a linear transformation to the flattened pixel vectors of each patch. Specifically, Linear Embedding projects the tensor with dimensions $(H/4 \times W/4) \times 48$ onto an arbitrary dimension C, resulting in a tensor with dimensions $(H/4 \times W/4) \times C$. Furthermore, linear embedding imbues the patches with positional information. In the absence of inherent sequential data within images, unlike text, the pairwise transformer relies on this embedding process to incorporate positional encodings, enabling the model to understand the spatial relationships between different patches of the image.

3.3. Patch merging

The Patch Merging layer serves as a downsampling mechanism designed to reduce resolution and adjust the number of channels, promoting a hierarchical structure while conserving computational resources and minimizing information loss. Each downsampling step reduces the sample size by a factor of two, effectively halving the dimensions in both the row and column directions. This reduction is achieved by selecting elements at intervals of two, creating a new patch from these elements. Subsequently, all the newly formed patches are concatenated to form a single tensor, which is then flattened. At this stage, the channel dimension increases to four times its original size due to the reduction in the height and width dimensions by half. A fully connected layer then processes this expanded tensor to adjust the channel dimension back to twice its initial size. Patch Merging efficiently compacts the data and preserves essential information through careful patch selection and recombination. By expanding and then strategically reducing the channel dimension, the model maintains critical features necessary for performance, optimizing both the data structure and computational efficiency without significant loss of information.

3.4. Pairwise-attention

The pairwise-attention mechanism is designed for pairwise verification tasks. It addresses the limitations of the self-attention mechanism by taking into account both the reference and query signatures in the attention mechanism. This enables effective capture of correlations and interdependencies between the inputs. Figure 2 illustrates the internal structure diagram of the complete pairwise-attention mechanism. The dashed lines in the left half represent the attention generation process from the reference signature to the query signature, while the solid lines in the right half represent the attention generation process from the query signature to the reference signature. Similar to self-attention, each attention generation requires three values, q, k, and v, to characterize the relationship between the reference and query signatures. In the context of our proposed pairwise-attention, we focus on the dependencies between the query signature and the reference signature. As depicted by the dashed lines in the left half of Figure 2, the input features (reference feature R and query feature Q) are first transformed into sequences of q, k, and v using a linear layer. Similarly, the right half of Figure 2 undergoes a transformation, but in this case, it converts the reference feature and query feature into sequences of k, q and v. The reference and query features are generated from the respective feature maps of the reference and query signatures. The feature vectors for R and Q can be represented as $[r_1, r_2, r_3, ..., r_N]$ and $[q_1, q_2, q_3, ..., q_N]$, respectively, where N represents the size of the spatial dimension and C represents the number of channels. In this scenario, the three vectors can be formalized as follows:

$$q^l = L_1(R) \qquad q^r = L_1(Q) \tag{1}$$

$$k_l = L_1(Q) \qquad k_r = L_1(R) \tag{2}$$

$$v_l = L_2(R) \qquad v_r = L_2(Q) \tag{3}$$

Where, L_1 and L_2 represent the linear layers, and R and Q ($\in R^{C \times N}$) represent the feature maps of the reference and query signatures, respectively.

Next, the associations between the input R and Q are modeled based on the interactions among attention q, k and v. In the proposed pairwise-attention mechanism, the dot product is used to establish the link between q and k. After applying a softmax layer, the attention map of q for k can be formulated as follows:

$$Sig^l_{qk} = \text{Softmax}(q_l \cdot {k_l}^T) \tag{4}$$

$$Sig^r_{kq} = \text{Softmax}(k_r \cdot {q_r}^T) \tag{5}$$

$$S^l_{ij} = \frac{\exp(L_1(R_i) \cdot L_1(Q_j)^T)}{\sum_{i=1}^{N} \exp(L_1(R_i) \cdot L_1(Q_j)^T)} \tag{6}$$

$$S^r_{ij} = \frac{\exp(L_1(Q_i) \cdot L_1(R_j)^T)}{\sum_{i=1}^{N} \exp(L_1(Q_i) \cdot L_1(R_j)^T)} \tag{7}$$

where Sig^l_{qk} represents the significance of q for k, Sig^r_{kq} represents the significance of k for q, and S^r_{ij} represents the importance of the i^{th} vector in Q for the j^{th} vector in R, while S^l_{ij} represents the importance of the i^{th} vector in R for the j^{th} vector in Q.

Finally, the interaction between the attention map matrix Sig^l_{qk} or Sig^r_{kq} and the global vector v is considered. To capture the interactions, an element-wise product is used. The formulation is as follows:

$$Atten^l_{RQ} = \frac{\text{Softmax}(L_1(R) \cdot L_1(Q)^T)}{\sqrt{d_k}} \cdot L_2(R) \tag{8}$$

$$Atten^r_{QR} = \frac{\text{Softmax}(L_1(Q) \cdot L_1(R)^T)}{\sqrt{d_k}} \cdot L_2(Q) \tag{9}$$

where $Atten^l_{RQ}$ and $Atten^r_{QR}$ represents the attention between R and Q, and $\sqrt{d_k}$ is the scaling factor to stabilize the gradients.

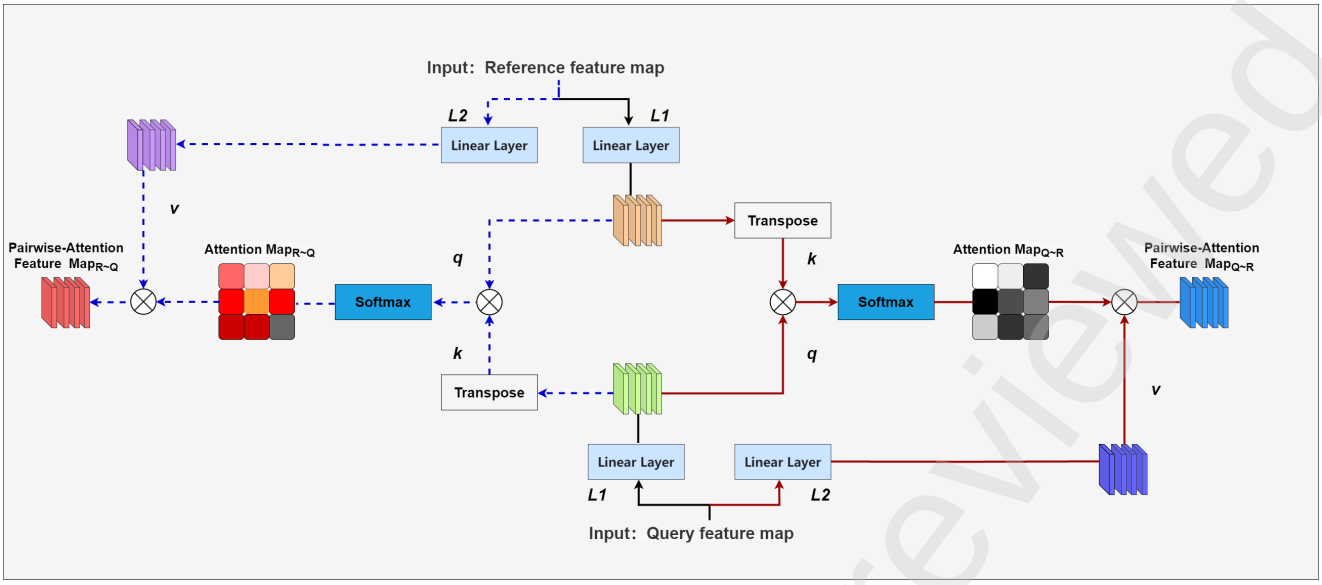


Figure 2: The process details of pairwise-attention.

3.5. Branch fusion block

The branch fusion block follow the window shifted and partitioning strategy of the Swin Transformer while replacing the self-attention mechanism with the proposed pairwise-attention mechanism. This ensures a fair comparison with the Swin Transformer. The proposed window multi-head pairwise-attention (W-MPA) and shift window multi-head pairwise-attention (SW-MPA) correspond to the window multi-head self-attention (W-MSA) and shift multi-head self-attention (SW-MSA).

The lower part of Figure 1 shows the branch fusion blocks. In Figure 1, we can observe two consecutive branch fusion blocks. Block l consists of LayerNorm (LN) layers, W-MPA, residual connections, and a 2-layer MLP. Block $l+1$ is similar to Block l, except that W-MPA is replaced with SW-MPA. It is important to note that Block l and Block $l+1$ are executed alternatively and sequentially.

The computational procedure for an input feature in the branch fusion blocks is as follows:

$$R_1 = \text{W-MPA}(LN(R)) + R \tag{10}$$

$$\hat{R} = \text{MLP}(LN(R_1)) + R_1 \tag{11}$$

$$\hat{R}_1 = \text{SW-MPA}(LN(\hat{R})) + \hat{R} \tag{12}$$

$$\hat{R}_2 = \text{MLP}(LN(\hat{R}_1)) + \hat{R}_1 \tag{13}$$

where R_1 and $\hat{R}$ represent the output features of the (S)W-MPA and the MLP module for Block 1, respectively. W-MPA and SW-MPA represent window-based multi-head pairwise-attention using regular and shifted window partitioning strategies, respectively.

4. Experiments

In this section, we present several experiments that reflect the various aspects of the proposed method for establishing the authorship of offline handwritten signatures. Additionally, we analyze the impact of signature background on verification performance.

4.1. Datasets and experimental protocol

To demonstrate the effectiveness of the proposed PAST, a series of experiments are conducted on five signature datasets: CEDAR [30], BHSig-H & BHSig-B [22], UTSig [29] and MCYT-75 [21]. The CEDAR dataset is a well-known English offline signature dataset that includes 1,320 genuine and 1,320 forged signatures obtained from 55 writers, where each writer has 24 genuine and 24 forged signatures. To simulate realistic conditions, all samples are collected at different periods of time and in a fixed 2×2 inches space. The signatures are then digitized at 300 dpi resolution and saved as PNG files. The BHSig260 database comprises 6,240 genuine and 7,800 forged signatures from 260 individuals, which can be divided into two parts: BHSig-H, a challenging Hindi dataset with 160 writers, and BHSig-B, a publicly available Bengali dataset with 100 writers. Each writer in both datasets provided 24 genuine signatures and 30 forgeries, and the samples are obtained from people with diverse educational backgrounds and ages to simulate a real application scenario. UTSig is a Persian offline signature dataset with 8,280 signatures from 115 writers. Each writer has 27 genuine and 45 forged signatures. The dataset considers variables such as signing period, writing instrument, signature box size, and observable samples for forgers. All

signatures are completed on A4-sized white forms, scanned at 600 dpi in eight-bit grayscale, and saved in TIF format. MCYT-75 is a Spanish offline signature dataset that includes 2,250 signatures from 75 individuals, with each individual providing 15 genuine and forged signatures. The signature samples are scanned at 600 dpi and saved as BMP files.

Table 1
Details of experimental protocol on different datasets

Dataset	Writers of Training	Writers of Test	positive and negative pairs trained per writer(P/N)
CEDAR	50	5	276/276
BHSig-H	100	60	276/276
BHSig-B	50	50	276/276
UTSig	60	55	351/351
MCYT-75	75	75	66/66

NOTE: To ensure the comparability of the data, our data partitioning format is kept fully consistent with that of the reference literature.

Our proposed method receives signature pairs as input, which consist of genuine and forged signatures from writers. The signature pairs can be further partitioned into two categories: positive pair and negative pair. The positive pair comprises two genuine signatures, and the negative pair consists of genuine and forged signatures. Both pairs are composed of positive and negative samples, respectively. In this case, for each writer with 24 genuine signatures and 30 forged signatures included in the BHSig-B dataset, there are $C_{24}^2 = 276$ genuine-genuine signature pairs (positive pairs) and $24 \times 30 = 720$ genuine-forged signature pairs (negative samples). To balance the positive and negative samples, 276 genuine-forgery pairs are randomly selected from each writer to balance the similar and dissimilar classes. Likewise, for BHSig-H datasets, there are $C_{24}^2 = 276$ genuine-genuine signature pairs and $24 \times 30 = 720$ genuine-forged signature pairs for each writer. For CEADR dataset, there are $C_{24}^2 = 276$ genuine-genuine signature pairs and $24 \times 24 = 576$ genuine-forged signature pairs for each writer. For UTSig, there are $C_{27}^2 = 351$ genuine-genuine and 351 genuine-forgery pairs per writer. The MCYT-75 dataset differs from others in that 80% of signatures are randomly chosen from each writer for training, while the remaining signatures are utilized as testing samples. In the first case, there are $C_{12}^2 = 66$ genuine-genuine and 66 genuine-forgery signature pairs each writer for training. In the second case, there are twice (132) genuine-forgery signature pairs each writer for training. The specific division is presented as Table 1. In order to minimize the bias of the results, all experiments are repeated 5 times, and the corresponding average values & standard deviations are also reported.

Evaluation of the proposed technique is measured through four indices: false acceptance rate (FAR), false rejection rate (FRR), Equal Error Rate (EER), and Accuracy (ACC). The specific calculations are shown in the following formula:

$$FAR = \frac{FP}{TN + FP} \tag{14}$$

$$FRR = \frac{FN}{TP + FN} \tag{15}$$

$$ACC = \frac{TP + TN}{TN + FN + TP + FP} \tag{16}$$

where TP (True Positive) is the number of correctly identified legitimate signatures, TN (True Negative) is the number of correctly identified forgeries, FN (False Negative) is the number of legitimate signatures misclassified as forgeries, and FP (False Positive) is the number of forgeries misclassified as legitimate signatures.

4.2. Ablation analysis of the proposed PAST

In this section, we ablate important design elements in the proposed Pairwise Attention Swin Transformer (PAST), focusing on the impact of different combinations of transformer blocks in PAST, different associated object values, balanced and unbalanced training datasets, and two strategies for fusing the two branches of output features. Moreover, all our ablation experiments were conducted on the BHSig-B dataset for both training and testing.

Impact of the allocation of PAST Transformer blocks:

Because the PAST structure is divided into four stages, we distribute the pairwise attention a mechanism across these stages, resulting in configurations of 2-2-18-2, 6-6-6-6, and 2-18-2-2, where each number in the sequence represents the number of blocks assigned to the four stages in the network. The results, as shown in Table 2, suggest that the network achieves the highest performance with a 2-2-18-2 block allocation, with a verification accuracy of 96.43%. 2-2-18-2 block allocation, with a verification accuracy of 96.43%. In the subsequent experiments, we keep the default configuration of 2-2-18-2.

Table 2
Impact of allocating the number of branch fusion blocks to each of the four stages in PAST on verification performance

Stage 1	Stage 2	Stage 3	Stage 4	ACC
2	2	18	2	**96.43**
6	6	6	6	95.37
2	18	2	2	93.88

Impact of the value in the context of the interactive attention mechanism:

Specifically, based on the self-attention calculation method, we represent the features generated in the attention pairing as the actual input object. Here, Q represents the features generated by the reference signature, while K represents the features generated by the query signature. We studied three different sequences of matrix-vector operations: (1) Q vector multiplied by K matrix, then multiplied by Q matrix; (2) Q vector multiplied by K matrix, then multiplied by K matrix; (3) Q vector multiplied by K matrix, then multiplied

by the average of Q and K matrices, $(Q + K)/2$. The three different operation sequences involve changing the content of values, representing combinations of value vectors from Q, K, and both Q and K. The results of this analysis, as shown in Table 3, indicate that the network achieved the highest performance when the value originated from the reference signature feature map.

Table 3
Impact of different values on verification performance in interaction attention mechanism

Q	K	V	ACC
Reference	Query	Reference	**96.43**
		Query	95.06
		(Reference+Query)/2	92.17

Impact of the sample balance on verification performance:

The results are presented in Table 4, which compares the use of balanced and unbalanced training samples. The Balanced Samples (BS) method trained with an equal number of positive and negative samples exhibited superior performance compared to the Unbalanced Samples (US), which trained with negative samples twice as abundant as positive samples. These findings suggest that the balance between positive and negative samples significantly affects verification performance, and utilizing balanced samples during the training phase can lead to improved results.

Table 4
Verification performance comparison between Balanced Samples (BS) and Unbalanced Samples (US) training methods

Training Method	FAR	FRR	ACC	EER
BS	**4.28**	**2.86**	**96.43**	**3.57**
US	4.33	5.41	95.13	9.74

Impact of the fusion on two branches:

The first strategy, referred to as 'concat', involves concatenating both branches prior to layer normalization. The second approach, referred to as 'sum', involves summing the outputs of both branches. As illustrated in the Table 5, the 'sum' strategy demonstrates superior performance compared to the 'concat' strategy. In general, the 'sum' strategy generates a new feature that encapsulates some of the key characteristics of the input features, although this process may result in some information loss. On the other hand, the 'concat' strategy concatenates the input features directly, allowing the model to learn how to effectively handle them. However, this strategy is computationally demanding. Despite the varied results of these approaches in the literature, we find that the 'sum' strategy are appropriate for our method.

Table 5
Comparison of two strategies for fusing two streams of output features

Fusion Strategy	ACC	FAR	FRR	ERR
Concat	95.81	4.02	4.36	4.19
Sum	**96.43**	**2.86**	**4.28**	**3.57**

Table 6
Performance comparison of pairwise-attention and self-attention on multiple signature datasets

Model	DATASET	FAR	FRR	ACC
Swin-Siamese		25.42	4.08	85.25
Swin-2-channel	BHSig260-B	8.89	11.24	89.93
PAST		**4.28**	**2.86**	**96.43**
Swin-Siamese		13.06	17.08	84.93
Swin-2-channel	BHSig260-H	12.45	14.71	86.42
PAST		**5.50**	**4.20**	**95.26**
Swin-Siamese		21.08	4.64	87.14
Swin-2-channel	CEDAR	9.48	12.11	89.21
PAST		**5.00**	**3.35**	**95.83**
Swin-Siamese		23.69	24.47	75.77
Swin-2-channel	UTSIG	27.34	**19.29**	76.68
PAST		**22.08**	22.18	**78.87**
Swin-Siamese		5.83	8.78	92.69
Swin-2-channel	MCYT-75	5.42	6.34	94.11
PAST		**0.84**	**1.48**	**98.83**

4.3. Comparative study of PAST and Swin Transformer under different frameworks

We further investigate the performance of pairwise attention and self-attention in signature verification. For fair comparison and optimal performance evaluation, we established fixed configurations in both PAST and Swin Transformers. Different numbers of transformer blocks were allocated at each stage, and the final transformer block allocation was set to 2-2-18-2. Additionally, the dimensionality of the input features after the linear embedding layer in Stage 1 was set to 128. In the models using self-attention, we incorporate two commonly used methods in signature verification, namely the 2-channel and Siamese approaches for data modeling. In the experiment, we maintain both Siamese-network-based and two-channel-based architectures, employing the Swin Transformer Block from the Swin Transformer and configured as described above. This configuration aligns with our PAST, having the same reference signature and query signature inputs. As a result, we obtain two derived models based on the Swin Transformer, namely Swin-Siamese and Swin-2-channel. The experimental results, as demonstrated in Table 6, based on five public datasets, show that the PAST model outperforms the Swin-Transformer model on all datasets. For example, on the BHSig260-B dataset, our model achieves FAR of 4.28%, FRR of 2.86%, and ACC of 96.43%, significantly outperforming the Swin-Siamese

and Swin-2-channel models. Only on the UTSig dataset, the Swin-2-channel model performs better than our method in terms of FRR, but our method achieves superior results in terms of FAR and ACC. These results indicate that the proposed PAST model with pairwise attention performs better than the Swin-Transformer model with self-attention and can be applied to multiple signature datasets.

4.4. Comparisons with the state-of-the-art

We conducted a comparative analysis of the proposed PAST against SOTA across five datasets. On each dataset, PAST delivered outstanding verification performance.

Table 7
Comparison of the proposed PAST with existing methods on CEDAR

Model	FAR	FRR	ACC	EER
SigNet[10]	**0**	**0**	**100.00**	**0**
MSN[36]	3.18	**0**	98.40	1.63
2C2S[23]	**0**	**0**	**100.00**	**0**
2C2L[15]	-	-	**100.00**	**0**
LQP[1]	5.01	6.12	-	-
BFS[26]	4.67	4.67	-	-
PAST	**0**	**0**	**100.00**	**0**

Table 8
Comparison of the proposed PAST with existing methods on BHSig-H

Model	FAR	FRR	ACC	EER
SigNet[10]	15.36	15.36	84.64	15.36
MSN[36]	17.06	5.16	88.88	11.31
LBP and ULBP[22]	24.47	24.47	75.53	24.47
2C2S[23]	8.66	9.98	90.68	9.32
2C2L[15]	-	-	86.66	13.34
SURDS[3]	12.01	8.98	89.50	-
IDN[35]	8.99	4.93	93.04	-
AVN[16]	-	-	94.32	5.65
PAST	**5.50**	**4.20**	**95.26**	**4.85**

As shown in the Table 7, almost all previous methods achieved 100% ACC on CEDAR. Of course, we obtained perfect results (zero error rate) on this database as well. However, this situation forces us to consider the issue of data bias or overfitting. We conducted an in-depth analysis of this phenomenon in conjunction with the dataset (to the best of our knowledge, this is the first time such a detailed analysis has been conducted), and the details of the analysis will be explained in the next section. The results in Table 8 demonstrate that the proposed method outperforms all other existing methods on BHSig-H. Specifically, the previously best-performing method, AVN, has an accuracy that is 0.94% lower than our method. According to Table 9, on BHSig-B, for the FAR, IDN has a slight advantage over our method, with a margin of just 0.16%. However, our method outperforms IDN in terms of ACC and FRR, with

Table 9
Comparison of the proposed PAST with existing methods on BHSig-B

Model	FAR	FRR	ACC	EER
SigNet[10]	13.89	13.89	86.11	13.89
MSN[36]	10.42	6.44	91.56	8.43
TransOSV[17]	9.90	9.90	-	9.90
2C2S[23]	5.37	8.11	93.25	6.75
2C2L[15]	10.44	9.37	88.08	11.92
LBP and ULBP[22]	33.82	33.82	66.18	33.82
SURDS[3]	19.89	5.42	87.34	-
Ref.[12]	-	-	90.36	-
IDN[35]	**4.12**	5.24	95.32	-
AVN[16]	-	-	93.80	6.14
PAST	4.28	**2.86**	**96.43**	**3.57**

Table 10
Comparison of the proposed PAST with existing methods on UTSIG

Model	FAR	FRR	ACC	EER
Ref.[29]	21.29	39.27	-	-
Ref.[28]	**9.09**	32.42	-	-
Ref.[24]	32.43	32.50	-	-
Ref.[20]	29.49	**7.88**	-	-
PAST	22.08	22.18	**78.87**	**22.13**

Table 11
Comparison of the proposed PAST with existing methods on MCYT-75

Model	FAR	FRR	ACC	EER
Ref.[29]	-	-	-	8.50
Ref.[1]	6.13	12.71	85.63	9.86
Ref.[13]	5.20	6.45	88.49	5.82
Ref.[7]	-	-	-	9.12
Ref.[32]	6.54	8.69	-	7.08
PAST	**0.84**	**1.48**	**98.83**	**1.16**

improvements of 2.38% and 1.11% respectively. Table 10 provides a comparison between the proposed PAST method and existing methods on UTSIG. The results indicate that while previous methods show good performance in terms of FAR and FRR, our method still maintains a leading position in each metric. Based on Table 11, our method on MCYT-75 achieves a FAR of 0.84%, an FRR of 1.48%, an accuracy of 98.83%, and an EER of 1.16%, significantly outperforming previous methods.

4.5. The secret of achieving 100% accuracy on the CEDAR

This section examines the impact of signature background on the performance of the CEDAR dataset. An interesting phenomenon is observed during experiments on the CEDAR dataset: when the model is trained on the CEDAR dataset that includes backgrounds, it achieves 100% accuracy. However, when the background is removed, the

Table 12
Cross-Language validation results of PAST on different signature datasets

DATASET	BHSig-B	BHSig-H	CEDAR	UTSig	MCYT-75
BHSig-B	**94.29**	76.40	51.63	54.24	46.88
BHSig-H	82.96	**95.83**	52.54	53.90	41.24
CEDAR	62.53 (50.00)	64.49 (50.00)	**95.83** (100.00)	54.96(50.00)	44.75(50.00)
UTSig	56.50	51.22	82.46	**78.87**	55.64
MCYT-75	73.67	70.34	75.04	71.39	**98.83**

NOTE: The accuracy value in parentheses indicates the model's performance on the CEDAR dataset with background information included, while the value outside parentheses shows performance without it.

Table 13
The impact of signature background on CEDAR dataset performance

Model	CEDAR	FAR	FRR	ACC
Swin-Siamese		21.08	4.64	87.14
Swin-2-channel	Without background	9.48	12.11	89.21
PAST		**5.00**	**3.35**	**95.83**
Swin-Siamese		0	0	100
Swin-2-channel	With background	0	0	100
PAST		0	0	100

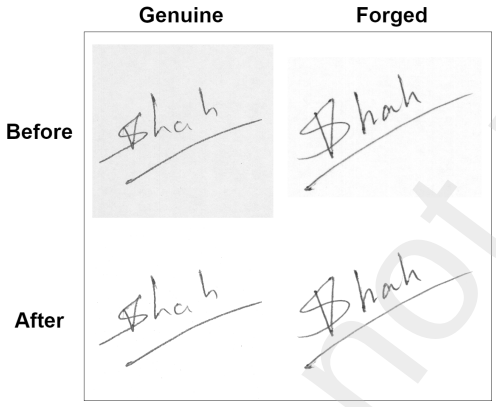


Figure 3: The comparison of signatures before and after processing.

validation performance of the model decreases. Table 13 is divided into two parts, one for models trained without backgrounds and the other for models trained with backgrounds. It shows the results of the experiment exploring the impact of signature background on the performance of the proposed PAST method and two methods using Swin Transformer (Swin-Siamese and Swin-2-channel) on the CEDAR dataset. For models trained without backgrounds, the proposed PAST achieves the best performance with a 5.00% FAR, 3.35% FRR, and 95.83% accuracy. Swin-Siamese achieved a 21.08% FAR, 4.64% FRR, and 87.14% accuracy, while Swin-2-channel achieved a 9.48% FAR, 12.11% FRR, and 89.21% accuracy. However, when models are trained with backgrounds, an interesting phenomenon is observed. All models achieve perfect performance with a 0% FAR, 0% FRR, and 100% accuracy. This indicates that the presence of background information in the training data greatly benefits the model's performance in recognizing signatures. It should be noted that Swin-2-channel achieves a significantly lower accuracy of 10% when trained with backgrounds, which may be due to overfitting. An example of the signature images before and after pre-processing is presented in the Figure 3.

4.6. Cross-dataset validation

This experiment evaluates the generalization ability of the proposed PAST across different languages. To assess this, we conducted cross-language verification experiments using five different language-specific signature datasets: CEDAR, BHSig-B, BHSig-H, UTSig, and MCYT-75. The experimental results were then represented using a confusion matrix, where each row represents training on one dataset and testing on different datasets. As shown in Table 12, the results indicate that PAST performs well in intra-language testing but experiences a decrease in performance in cross-language testing due to dataset variations.In the case of the CEDAR dataset, the model achieves the highest accuracy of 100% when trained with background information (indicated in parentheses). However, when trained on the CEDAR dataset without background and tested on other datasets, the model fails to maintain a high level of accuracy, indicating that it fails to make effective judgments and does not learn any meaningful information about the signatures. This is because the model only learns to differentiate the presence or absence of background in signature images, rather than learning any meaningful information about the signatures themselves.

5. Conclusion

This paper introduces Pairwise Attention Swin Transformer (PAST), a novel approach for signature verification that leverages the pairwise-attention mechanism tailored for pairwise verification tasks. The experimental results demonstrate that PAST outperforms both the baseline Swin Transformer and existing methods across all five signature datasets, achieving remarkable performance. These findings underscore the significance of pairwise-attention in signature verification tasks, validating the effectiveness of PAST

in capturing subtle differences between genuine and forged signatures. The proposed PASTs exhibits tremendous potential for practical signature verification applications, addressing the limitations of existing methods and leveraging the innovative pairwise-attention mechanism. PAST paves the way for precise and dependable signature verification, with potential impacts on fields including document authentication, identity verification, and financial transactions.

Furthermore, our experiments highlight the substantial impact of the pairwise-attention mechanism on the model's performance, with an increase in the number of branch fusion blocks and the embedding dimension further enhancing accuracy. While the exploration of larger models was constrained by hardware limitations, our research suggests that investigating larger models holds the potential to further amplify the efficacy of our pairwise-attention mechanism.

6. ACKNOWLEDGMENT

This work was supported by the National Natural Science Foundation of China (62006150); Science and Technology Commission of Shanghai Municipality (21DZ2203100).

CRediT authorship contribution statement

Yu-Jie Xiong: Conceptualization of this study, Methodology, Investigation, Supervision. **Jian-Xin Ren:** Software, Writing - Original draft preparation, Methodology, Validation. **Dong-Hai Zhu:** Methodology, Supervision, Writing – review editing. **Xi-Jiong Xie:** Methodology, Validation, Writing – review editing. **Xi-He Qiu:** Data curation, Investigation, Supervision.

References

[1] Bhunia, A.K., Alaei, A., Roy, P.P., 2019. Signature verification approach using fusion of hybrid texture features. Neural Computing and Applications 31, 8737–8748.

[2] Bromley, J., Guyon, I., LeCun, Y., Sckinger, E., 1993. Signature verification using a "siamese" time delay neural network, in: Advances in neural information processing systems(NIPS), pp. 737–744.

[3] Chattopadhyay, S., Manna, S., Bhattacharya, S., Pal, U., 2022. Surds: Self-supervised attention-guided reconstruction and dual triplet loss for writer independent offline signature verification. 2022 International Conference on Pattern Recognition (ICPR) , 1600–1606.

[4] Chen, C.F., Panda, R., Fan, Q., 2021. Regionvit: Regional-to-local attention for vision transformers. arXiv preprint arXiv:2106.02689 .

[5] Chu, J., Zhang, W., Zheng, Y., Ahmad, R., 2023. Signature verification by multi-size assembled-attention with the backbone of swin-transformer.

[6] Cong, P., Zhou, J., Wang, J., Wu, Z., Hu, S., 2023. Learning-based cloud server configuration for energy minimization under reliability constraint. IEEE Transactions on Reliability , 1–13.

[7] Diaz, M., Ferrer, M.A., Eskander, G.S., Sabourin, R., 2016. Generation of duplicated off-line signature images for verification systems. IEEE Transactions on Pattern Analysis and Machine Intelligence 39, 951–964.

[8] Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., 2021. An image is worth 16x16 words: Transformers for image recognition at scale URL: https://openreview.net/forum?id=YicbFdNTTy.

[9] Ghosh, R., 2021. A recurrent neural network based deep learning model for offline signature verification and recognition system. Expert Systems with Applications 168, 114249.

[10] Hafemann, L.G., Sabourin, R., Oliveira, L.S., 2017. Learning features for offline handwritten signature verification using deep convolutional neural networks. Pattern Recognition 70, 163–176.

[11] Hafemann, L.G., Sabourin, R., Oliveira, L.S., 2019. Characterizing and evaluating adversarial examples for offline handwritten signature verification. IEEE Transactions on Information Forensics and Security 14, 2153–2166.

[12] Jadhav, S.K., Chavan, M., 2018. Symbolic representation model for off-line signature verification, in: International Conference on Computing, Communication and Networking Technologies (ICCCNT), IEEE. pp. 1–5.

[13] Jagtap, A.B., Sawat, D.D., Hegadi, R.S., Hegadi, R.S., 2020. Verification of genuine and forged offline signatures using siamese neural network (snn). Multimedia Tools and Applications 79, 35109–35123.

[14] Kaur, H., Kumar, M., 2021. Signature identification and verification techniques: State-of-the-art work. Journal of Ambient Intelligence and Humanized Computing , 1–19.

[15] Li, C., Lin, F., Wang, Z., Yu, G., 2019. Deephsv: User-independent offline signature verification using two-channel cnn, in: International Conference on Document Analysis and Recognition (ICDAR), IEEE. pp. 166–171.

[16] Li, H., Wei, P., Hu, P., 2021. Avn: An adversarial variation network model for handwritten signature verification. IEEE Transactions on Multimedia 24, 594–608.

[17] Li, H., Wei, P., Ma, Z., Li, C., Zheng, N., 2024. Transosv: Offline signature verification with transformers. Pattern Recognition 145, 109882.

[18] Liu, L., Huang, L., Yin, F., Chen, Y., 2021a. Offline signature verification using a region based deep metric learning network. Pattern Recognition 118, 108009.

[19] Liu, Z., Lin, Y., Cao, Y., Hu, H., 2021b. Swin transformer: Hierarchical vision transformer using shifted windows, in: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), pp. 10012–10022.

[20] Maergner, P., Pondenkandath, V., Alberti, M., Liwicki, M., 2019. Combining graph edit distance and triplet networks for offline signature verification. Pattern Recognition Letters 125, 527–533.

[21] Ortega-Garcia, J., Fierrez, J., Simon, D., Gonzalez, J., 2003. Mcyt baseline corpus: a bimodal biometric database. iee proc vis image signal process spec issue biom internet. IEE Proceedings - Vision Image and Signal Processing , 395 – 401.

[22] Pal, S., Alaei, A., Pal, U., Blumenstein, M., 2016. Performance of an off-line signature verification method based on texture features on a large indic-script signature dataset, in: International Workshop on Document Analysis Systems (DAS), IEEE. pp. 72–77.

[23] Ren, J.X., Xiong, Y.J., Zhan, H., Huang, B., 2023. 2c2s: A two-channel and two-stream transformer-based framework for offline signature verification. Engineering Applications of Artificial Intelligence 118, 105639.

[24] Rezaei, M., Naderi, N., 2019. Persian signature verification using fully convolutional networks. arXiv preprint arXiv:1909.09720 .

[25] Ruiz, V., Linares, I., Sanchez, A., Velez, J.F., 2020. Off-line handwritten signature verification using compositional synthetic generation of signatures and siamese neural networks. Neurocomputing 374, 30–41.

[26] Sharif, M., Khan, M.A., Faisal, M., Yasmin, M., 2020. A framework for offline signature verification system: Best features selection approach. Pattern Recognition Letters 139, 50–59.

[27] Shen, Q., jun Luan, F., Yuan, S., 2022. Multi-scale residual based siamese neural network for writer-independent online signature verification. Applied Intelligence 52, 14571–14589.

[28] Soleimani, A., Araabi, B.N., Fouladi, K., 2016. Deep multitask metric learning for offline signature verification. Pattern Recognition Letters 80, 84–90.

[29] Soleimani, A., Fouladi, K., Araabi, B.N., 2017. Utsig: A persian offline signature dataset. IET Biometrics 6, 1–8.
[30] Srinivasan, H., Srihari, S.N., Beal, M.J., 2006. Machine learning for signature verification, in: Conference on Computer Vision, Graphics and Image Processing (ICVGIP), Springer. pp. 761–775.
[31] Touvron, H., Cord, M., Douze, M., Massa, F., 2021. Training data-efficient image transformers amp; distillation through attention, in: Proceedings of the 38th International Conference on Machine Learning, pp. 10347–10357.
[32] Vargas, J., Ferrer, M., Travieso, C., Alonso, J.B., 2011. Off-line signature verification based on grey level information using texture features. Pattern Recognition 44, 375–385.
[33] Wang, W., Xie, E., Li, X., Fan, D.P., 2021. Pyramid vision transformer: A versatile backbone for dense prediction without convolutions, in: IEEE International Conference on Computer Vision (ICCV), IEEE. pp. 568–578.
[34] Wei, P., Li, H., Hu, P., 2019a. Inverse discriminative networks for handwritten signature verification, in: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 5757–5765.
[35] Wei, P., Li, H., Hu, P., 2019b. Inverse discriminative networks for handwritten signature verification, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 5764–5772.
[36] Xiong, Y.J., Cheng, S.Y., 2021. Attention based multiple siamese network for offline signature verification, in: International Conference on Document Analysis and Recognition (ICDAR), Springer. pp. 337–349.
[37] Zhou, D., Kang, B., Jin, X., Yang, L., 2021. Deepvit: Towards deeper vision transformer. arXiv preprint arXiv:2103.11886 .
[38] Zhou, J., Shen, Y., Li, L., Zhuo, C., 2023. Swarm intelligence-based task scheduling for enhancing security for iot devices. IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems 42, 1756–1769.