Int. J. Appl. Math. Comput. Sci., 2026, Vol. 36, No. 1, 129–140
DOI: 10.61822/amcs-2026-0010

RESEARCH ON FEW–SHOT HANDWRITING IDENTIFICATION BASED ON SIAMESE NETWORKS

XUYANG WANG [a], CHENGZHI XU [a,*], LIUYUAN DONG [a], RUIZHEN XIE [a], WANLI YANG [b]

[a] Hubei Provincial Engineering Research Center
for Digital & Intelligent Manufacturing Technologies and Applications
Hubei University of Technology
No. 28, Nanli Road, Hongshan District, Wuhan, Hubei, China
e-mail: xcz911@hbut.edu.cn,
{609224869,1439075938,2267372148}@qq.com

[b] Key Laboratory of Image Information Processing and Intelligent Control,
of the Ministry of Education
Huazhong University of Science and Technology
No. 1037, Luoyu Road, Hongshan District, Wuhan, Hubei, China
e-mail: 873744617@qq.com

With the rise of the digital era, handwriting examination has become crucial for identity verification and document provenance. However, determining whether samples from different texts are written by the same person remains challenging. The challenge is greater in few-shot settings, where data are scarce and writing styles vary widely. Traditional methods often lack sufficient accuracy and robustness. We propose a dual-branch Siamese network for handwriting verification. It fuses attention mechanisms with a feature-bank matching strategy. This design improves adaptation and generalization under few-shot conditions. It also suppresses background noise and emphasizes key writing traits. We evaluate the method on CCSbC, the mixed Chinese–English hard-pen dataset named MDC, and CCD-CQU. The model attains high accuracy on multi-class few-shot classification tasks. It shows strong robustness and adaptability. With data augmentation and feature optimization, it could deliver more efficient handwriting identification in real-world applications.

Keywords: handwriting identification, attention mechanism, Siamese network, feature library, few-shot learning.

1. Introduction

Handwriting examination is a technical procedure that systematically analyzes handwriting to assess document authorship. It falls under questioned document examination (QDE) and is a key branch of forensic science. It is widely used in criminal investigations and civil litigation, and it is crucial in investigating white-collar crimes such as impersonation, fraud, and document forgery. However, traditional handwriting examination is centered on the expert's reasoning process (Abbas *et al.*, 2021). Its outcomes cannot guarantee absolute precision. It also suffers from low efficiency and high cost. In addition, many factors—such as the writer's age, psychological state, and physical condition—affect handwriting features (Van Drempt *et al.*, 2011). These factors can introduce uncertainty into the conclusions. They limit the applicability of traditional methods in complex cases and increase the risk of reliance on subjective expert judgment.

With advances in information technology, handwriting identification is moving from expert-driven practice to automation and intelligence (Zhao and Li, 2023). In handwriting recognition, cross-script signature recognition is an emerging topic. Traditional methods focus on a single script. In real applications, a signature may appear in different scripts. For example, a user enrolled with an English signature may be verified using a Chinese one. This cross-script need drives research into more robust methods (Hossain *et al.*, 2025). Recent progress in deep learning has improved automated

*Corresponding author

handwriting systems, which enhance accuracy, efficiency, and objectivity. However, most studies still target a single script and controlled writing conditions. They offer limited solutions for few-shot scenarios with many writers and diverse texts. Generalization often degrades when writer styles vary widely.

To address these issues, we study multi-class few-shot handwriting identification. We propose a dual-branch Siamese network that integrates attention and feature-bank matching. We validate the model in mixed Chinese–English settings. Our main contributions are as follows:

1. *Improved Siamese feature-extraction framework.* We design a four-layer feature extractor to better capture key stroke regions. We introduce a dual-path EMA (enhanced mixed attention) module to replace the conventional decision network. This optimizes feature fusion and refinement. It significantly improves adaptability and robustness in few-shot settings.

2. *Proposed training and data-generation strategies for few-shot learning.* We devise a novel pair-generation scheme and training pipeline. It maintains sample diversity and mitigates data scarcity. The model converges stably with limited data.

3. *Designed feature-bank-based decision method.* We construct a high-quality feature bank and perform similarity matching. This enables efficient multi-class classification. It markedly improves accuracy and robustness in multi-writer, multi-script tasks.

2. Related work and a background

Handwriting identification has evolved from traditional visual inspection to modern biometric techniques (Ali and Abdulrazzaq, 2023). With deep neural networks, automation is entering an intelligent phase. Researchers have proposed diverse methods for varied scenarios.

In handwriting classification, traditional approaches are inefficient and inaccurate. On English handwriting datasets, VGG-16 performs well for image classification (Simonyan and Zisserman, 2014), yet its strictly sequential architecture leads to long training times and difficult tuning. It also struggles to capture fine details. To address this, Kai *et al*. (2020) replace part of VGG-16's convolutions with composite convolutional layers. This speeds up feature extraction, shortens training, and improves sensitivity to subtle features. In the work of Huang *et al*. (2023), a new convolutional neural network (CNN) is designed with multiple convolution, pooling, and fully connected layers. It classifies calligraphic styles from four Tang-dynasty masters: Ouyang Xun, Chu Suiliang, Yan Zhenqing, and Liu Gongquan.

In signature verification, Wei *et al*. (2019) propose an inverse discriminative network (IDN). It uses four weight-sharing streams (two discriminative and two inverse) and a multi-path attention module (Vaswani *et al*., 2017). The attention strengthens stroke information and mitigates the sparsity of signature images. TransOSV (Li *et al*., 2022) builds on the vision transformer (ViT) framework (Dosovitskiy *et al*., 2020). It encodes signatures into global features to capture long-range stroke relations. A local decoder learns local patterns. With contrastive learning, it extracts subtle differences between genuine and forged signatures. SigNet (Dey *et al*., 2017) is a convolutional Siamese network inspired by Krizhevsky *et al*. (2012). It takes a pair of signature images and computes the Euclidean distance between their embeddings to measure similarity. Ren *et al*. (2023) introduce a dual-channel, dual-stream transformer. It uses an improved Swin Transformer backbone (Liu *et al*., 2021) to extract multi-scale features and model inter-channel relations. An upsampling enhancement module focuses on informative regions. The fused features support the final real–fake decision, and experiments show superior performance.

In text-independent handwriting identification, few-shot learning (FSL) has attracted broad attention in image recognition and text classification. Kleber *et al*. (2013) provide a comprehensive survey, formally defining FSL and distinguishing it from related machine-learning problems. Askari *et al.* (2025) propose a few-shot method that combines prototype rectification with self-attention; by optimizing class-prototype representations, it markedly improves classification performance under scarce data (Zhao *et al*., 2024). Nonetheless, directly applying such methods to handwriting identification remains challenging, especially in complex settings with many writers and multiple scripts. Zhao and Li (2023) summarize three core FSL paradigms: metric-based, model-based, and optimization-based approaches. Singla and Mittal (2025) review recent offline signature verification techniques and note limitations when handling complex signatures and few-shot data. In few-shot classification, the goal is to train a classifier with acceptable performance using limited samples (Sánchez-DelaCruz and Loeza-Mejía, 2024). However, traditional metric-based methods still face performance bottlenecks. Whether for handwriting classification or signature verification, deep models typically rely on large datasets. In handwriting identification specifically, data acquisition is inherently constrained—one person's writing samples rarely number in the thousands and are more often only a few hundred characters. This constraint challenges model generalization and training efficacy. Thus, despite strong results on generic text data,

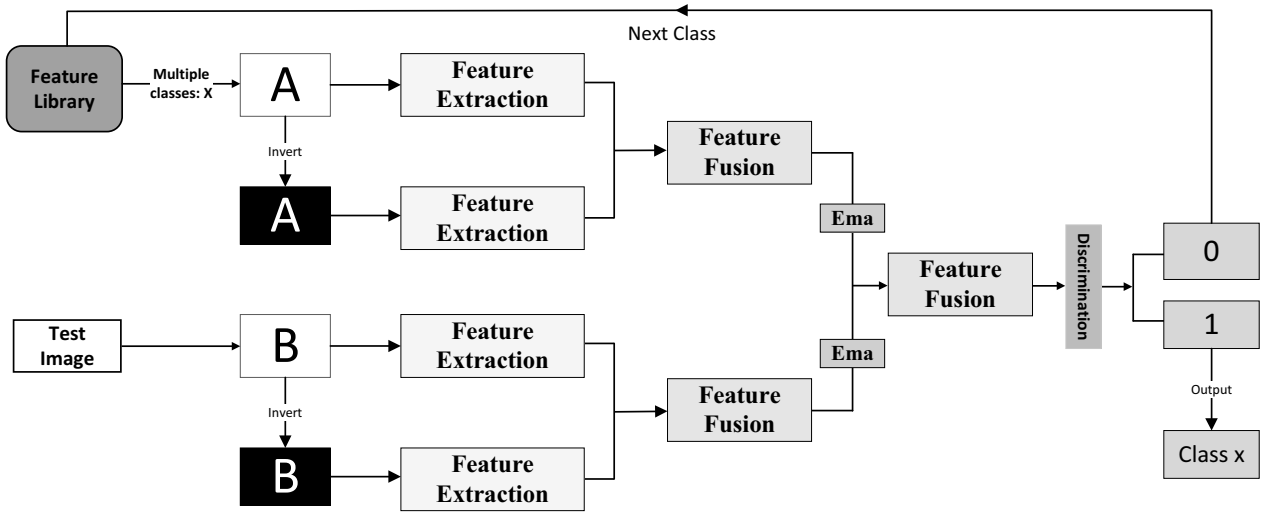

Fig. 1. Structure diagram of different handwriting identification systems.

effectively transferring and adapting these techniques to handwriting identification remains an open problem (Aljehani *et al*., 2024).

The performance of neural networks depends heavily on dataset coverage and sample quality. In recent years, as style recognition has advanced, researchers have released several public datasets. Along those, CVL (Wang *et al*., 2020) contains 50 classes with about 450 samples per class, while CCD-CQU by Huang *et al*. (2023) includes 8,282 samples spanning four classes. Zhang *et al*. (2022) introduce a new handwriting verification benchmark—Multimodal Signature and Digit String (MSDS)—with two subsets: Chinese Signatures (MSDS-ChS) and Token Digit Strings (MSDS-TDS). It includes contributions from 402 users, with 20 genuine and 20 skilled-forgery samples per user per subset. The IAM Handwriting Database (Marti and Bunke, 2002) is a widely utilized dataset for handwriting research, comprising 1,539 pages of text from 657 writers, with a total of 13,353 text lines and 115,320 individual word samples. The dataset is captured at a resolution of 300 DPI in grayscale PNG format. Although numerous datasets are available for handwriting recognition and writer identification studies, the number of handwriting image samples per writer per category is generally limited to 300–500. In conventional training, small datasets make it difficult to build highly accurate neural models; consequently, the scarcity of samples substantially limits generalization.

3. Research methods

Inspired by the dual inverse discriminative attention modules proposed by Wei *et al*. (2019) and Chao-Qun *et al*. (2024), this study designs a dual-channel Siamese network for feature extraction. The overall architecture is shown in Fig. 1. The model consists of four main components: feature extraction, feature fusion, an EMA module, and a discrimination module. A handwriting image is first converted into a grayscale-inverted version. Both the original and inverted images are fed into the network. The dual inverse discriminative attention mechanism then applies weighted processing to extract fine-grained features. These features are compressed and merged through channel fusion to form compact representations. A subset of these representations is stored in a feature library as category references, which are later used to match new handwriting samples for high-similarity discrimination.

3.1. Dataset and generation of extended signature image pairs. To build a high-quality dataset of signature image pairs, we design an automatic generation method. It produces positive pairs (different signatures from the same author) and negative pairs (signatures from different authors) based on a given dataset, and stores them as training path files.

First, signature images are loaded from a designated folder. Suppose the dataset has five classes, each with 100 images. Positive pairs are created by randomly selecting two different images from the same class and assigning a label of 1. The number of possible positive pairs per class is $(n(n-1))/2$. For example, with $n = 100$, there are 4,950 pairs per class, or 5×4950 pairs in total. In practice, we sample only a subset of these pairs to control dataset size and improve generation efficiency.

Negative pairs are formed by selecting images from two different classes and assigning a label of 0. Their total number is set equal to that of positive pairs. These pairs account for 80% of the entire sample set to maintain

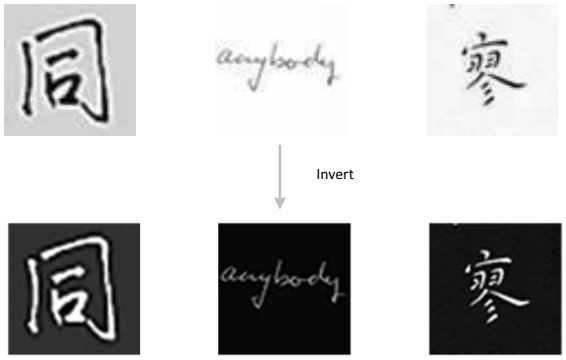

Fig. 2. Inverse (top: positive, bottom: negative).

balance. For example, if class A is the reference, the second image is chosen from any non-A class.

To improve diversity and generalization, positive and negative pairs are randomly shuffled before saving. The training and test sets are stored separately as text files, with each line containing the image paths and the label. This method uses random pairing without requiring identical signature content, making it suitable for diverse signature classification tasks.

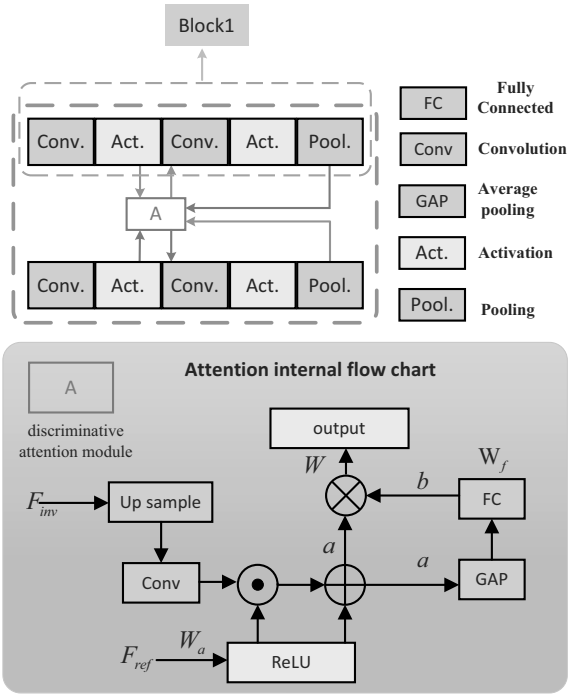

Fig. 3. Dual inverse discriminative attention module.

3.2. Grayscale processing. As pointed out by Bhattacharya *et al*. (2013), the number of channels in the input image has a significant impact on model performance. In RGB images, redundant color information may introduce noise and cause unstable training. In contrast, grayscale images, with only one channel, reduce complexity and computational cost while improving verification accuracy.

Our experiments show that using grayscale images significantly boosts learning performance compared with color images, especially in offline signature verification. Therefore, we adopt single-channel inputs throughout this study to enhance training efficiency and accuracy, as illustrated in Fig. 2.

3.3. Dual inverse discrimination attention module. To achieve information guidance and interactive fusion between cross-modal features (original and inverted images), we integrate a cross-branch attention mechanism (Chao-Qun *et al*., 2024) after each convolutional stage (as illustrated in Fig. 3). The core objective of this module is to use complementary features from one branch to guide discriminative feature generation in the other. This process enables mutual information exchange and enhances features at multiple levels.

This attention mechanism takes the reverse branch feature map W_{inv} and the reference branch feature map W_{ref} as inputs to generate a new weighted feature map F_{out}. First, the reverse feature W_{inv} is upsampled to the same spatial size as W_{ref}, resulting in $F_{inv}^{\uparrow} = UpSample(F_{inv})$. Subsequently, a 1×1 convolutional layer is then applied to compress the channels, generating a preliminary spatial attention map. After sigmoid normalization, this produces the spatial attention weight $W_a = \sigma(Conv(F_{inv}^{\uparrow}))$. The weight is multiplied with W_{ref} and added to the original feature to obtain the intermediate feature $F_{tmp} = W_a \cdot F_{ref} + F_{ref}$ (element-wise or channel-wise multiplication).

For channel attention compression and amplification, the enhanced feature map W_{tmp} first undergoes global average pooling (GAP) to extract a channel-wise vector $f = GAP(F_{tmp})$. This vector is passed through a fully connected (FC) layer to model channel dependencies, producing the channel attention vector $W_f = \sigma(FC(f))$. After reshaping W_f to match the original channel dimensions, it is multiplied with the intermediate features to produce the final fused output $F_{out} = F_{tmp} \cdot W_f$.

3.4. Enhanced mixed attention. To further enhance the model's capability to represent local structures and spatial dependencies, we propose a group-wise enhanced mixed attention (EMA) module (Ouyang *et al*., 2023). Inspired by the multi-scale embeddings and attention mechanisms of Askari *et al*. (2025), EMA integrates multi-scale feature extraction with attention to more effectively capture key stroke characteristics while suppressing background noise. Specifically, EMA partitions the channels into groups, enabling lightweight

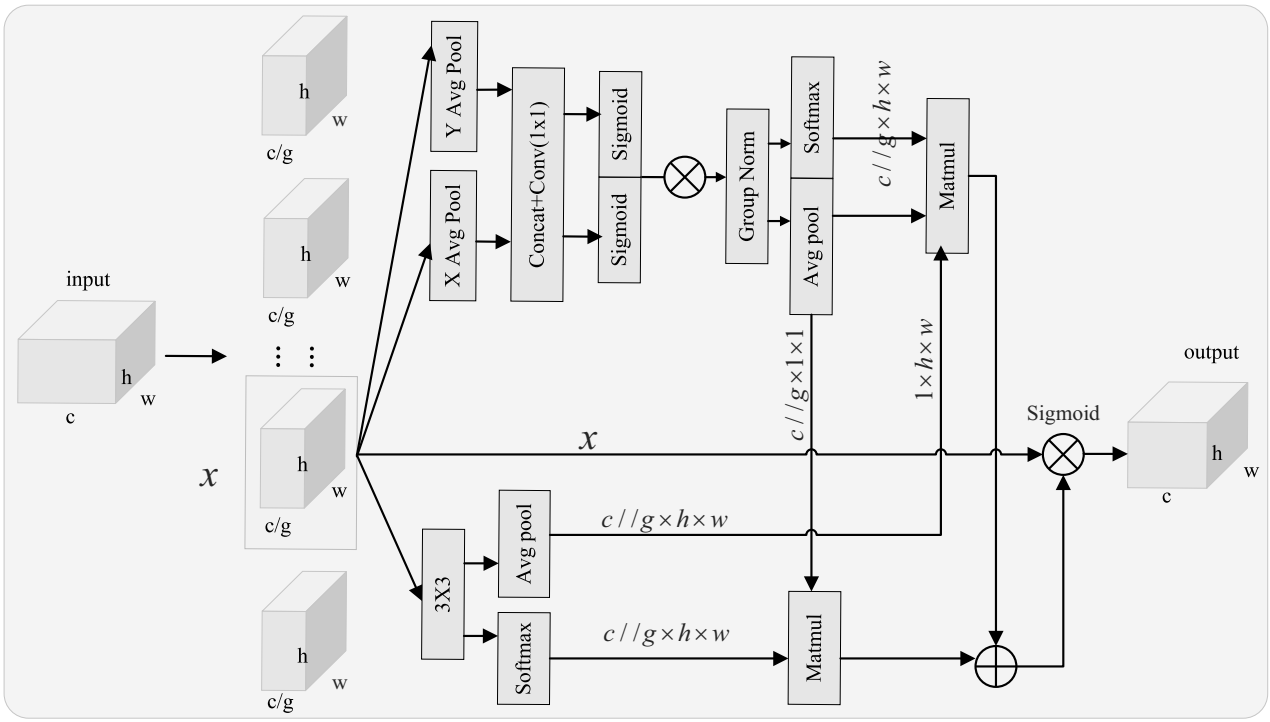

Fig. 4. EMA module. Here, "g" denotes the number of groups, "X Avg Pool" represents 1D horizontal global pooling, and "Y Avg Pool" stands for 1D vertical global pooling.

intra-group attention and cross-scale aggregation. This design enriches both local and global descriptors without introducing substantial computational overhead.

Given an input feature map $X \in \mathbb{R}^{B\times C\times H\times W}$ (where B denotes the batch size, C the number of channels, and $H \times W$ the spatial resolution), EMA first divides the channels into G groups (default value omitted here), each containing C/G channels. Each group is processed independently through attention operations to reduce computational complexity and improve the diversity of feature representations. The subdivided feature maps are denoted as $X_g \in \mathbb{R}^{B\times C/G\times H\times W}$, as illustrated in Fig. 4.

For each group feature X_g, the module applies adaptive average pooling along the horizontal and vertical directions, producing $X_h \in \mathbb{R}^{B\times C/G\times H\times 1}$ and $X_w \in \mathbb{R}^{B\times C/G\times 1\times W}$, respectively. These features are concatenated along the channel dimension and fused via a 1×1 convolution to yield the intermediate representation $Z \in \mathbb{R}^{B\times 2C/G\times H\times W}$, where Z denotes the fused intermediate feature map. Subsequently, Z is split along the channel dimension into $X'_h \in \mathbb{R}^{B\times C/G\times H\times W}$ and $X'_w \in \mathbb{R}^{B\times C/G\times H\times W}$. These two components are then multiplied element-wise with the original group feature map to generate the spatially enhanced feature X_1:

$$X_1 = GN\big(X_g \cdot \sigma(X'_h) \cdot \sigma(X'^{\top}_w)\big), \tag{1}$$

where GN represents group normalization and σ denotes the sigmoid activation function.

While incorporating spatial enhancement, the EMA module also constructs two mutually symmetrical attention paths. The statistical descriptors F_1 and F_2 are extracted from the two branches as follows:

$$F_1 = \mathrm{Softmax}(\mathrm{GAP}(X_1)), \tag{2}$$

$$F_2 = \mathrm{Softmax}(\mathrm{GAP}(X_2)), \tag{3}$$

$$X_2 = \mathrm{Conv}_{3\times 3}(X_g), \tag{4}$$

where GAP denotes global average pooling.

The two attention vectors are then interchanged, multiplied with the corresponding feature vectors, and combined through a weighted fusion process to obtain the final hybrid attention weights:

$$W = \sigma\big(\mathrm{Reshape}(\mathrm{MatMul}(F_1, X_2) + \mathrm{MatMul}(F_2, X_1))\big). \tag{5}$$

Finally, the generated weights W are applied to the original grouped features X_g, producing the module's weighted output:

$$Y = \mathrm{Reshape}(X_g \cdot W), \tag{6}$$

which restores the original feature dimensions (B, C, H, W) and is fully compatible with subsequent network layers.

In summary, the EMA module captures both spatially guided features and global dependencies. This design integrates local and global context, enhancing feature representation. By dividing channels into multiple groups, EMA reduces computational complexity through grouped processing and improves scalability. Its bidirectionally symmetrical attention paths enhance stability and reduce overfitting risks common in single-path designs. In practice, EMA can be embedded in the feature fusion stages of multi-scale backbone networks to amplify responses in critical regions, significantly improving the model's discriminative power and generalization.

3.5. CNN and a Siamese network. The non-specific handwriting identification system consists of two main components: network identification and category matching. In the network identification stage, a Siamese network takes paired samples as inputs. Two neural network branches with shared parameters extract features independently. At the top layer, the weighted L1-norm distance between feature pairs is computed, and its sigmoid response is used to measure sample similarity (Koch *et al*., 2015). Following prior work (Wei *et al*., 2019; Dey *et al*., 2017), we adopt a quadruple Siamese network structure. It comprises the BiFuseStream module, feature fusion, the EMA module, the feature compression module, and a fully connected classifier (Fig. 5). Images A and B are first converted to grayscale. They are then inverted to generate A1 and B1, yielding four images in total.

To improve discrimination under complex backgrounds and handwriting variations, we propose a dual-branch feature extraction and fusion module, BiFuseStream. This module contains four consecutive convolutional stages. Each stage has two convolutional layers, a max-pooling layer, and a cross-branch attention mechanism (Fig. 5).

The input images consist of original signature images (denoted as A and B) and their inverted versions (denoted as A_r and B_r), which are fed into the reference branch and inversion branch, respectively. At each stage, the dual-stream symmetric structure extracts features and fuses information, thereby enhancing the model's sensitivity to fine-grained image differences.

Step 1: *Shallow feature extraction and initial attention guidance.* In the first stage, the input image passes through two 3×3 convolutional layers (conv1_1 and conv1_2). The channel number increases from 1 to 32, and ReLU activation enhances non-linear representation. A 2×2 max-pooling layer then reduces spatial dimensions. Next, the two branches interact via the attention module. Pooled features from the reverse branch are fused with shallow convolutional features from the reference branch, and vice versa. The attention module upsamples features using nearest-neighbor interpolation, applies a 1×1 convolution, and generates spatial attention maps via sigmoid activation. Finally, channel-wise weighting—computed from global average pooling, a fully connected layer, and Sigmoid activation—reinforces the fused results.

Step 2: *Mid-level semantic enhancement and cross-branch guidance.* The fused feature maps enter the second-stage convolutional module (conv2_1 and conv2_2). The channel dimension expands to 64. Another pooling operation follows. At this stage, features capture richer stroke structures and edge orientations. These are important for signature style discrimination. The dual-branch attention interaction is repeated. This enhances dependency modeling between features from different branches. The output is a new feature pair, att_f_2 and att_t_2.

Step 3: *High-level abstract semantic modeling.* The fused features are processed by conv3_1 and conv3_2. Channels increase to 96. This stage focuses on modeling global structures and texture patterns of handwriting. Pooling-based downsampling follows. Guided by the attention mechanism, the two branches perform third-stage semantic fusion. This extracts more discriminative high-level features.

Step 4: *Global semantic integration and final fusion.* In the final stage, features pass through conv4_1 and conv4_2. The channel number expands to 128 to capture the deepest high-dimensional semantics. The attention mechanism strengthens inter-branch correlations. It also builds on the features obtained from previous stages. The enhanced feature maps from both branches are concatenated along the channel dimension. This forms the ultimate fused representation for similarity computation or classification.

As shown in Fig. 5, the model outputs four 128-dimensional feature tensors after Step 4. Feature fusion in two vertically stacked Siamese networks produces two 256-dimensional feature vectors. These are sent to two EMA (efficient multi-scale attention) modules for independent attention computation. The outputs are fused to form a 512-dimensional vector. An additional EMA module refines this vector. Multi-stage attention computation enriches the information in each channel. This significantly improves the expressive power of the feature maps. The model captures detailed features and contextual information.

The fused 512-channel feature map from the EMA module is passed through a convolutional layer to reduce the channel dimension to 256. This lowers computational cost and helps prevent overfitting. ReLU activation strengthens non-linear representation. Batch normalization standardizes features and stabilizes

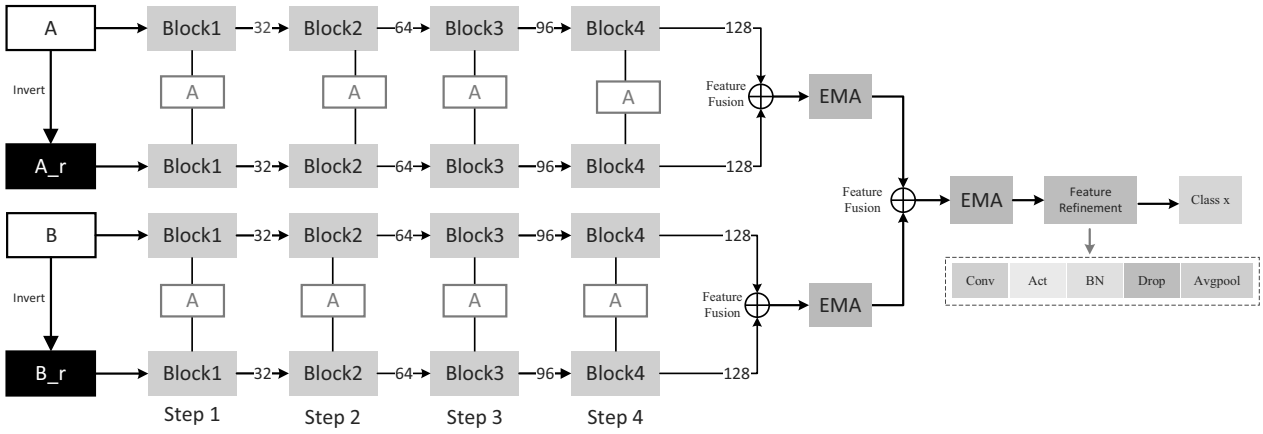

Fig. 5. Dual-branch Siamese network.

training. A dropout layer (default rate: 0.5) prevents neuronal co-adaptation. Next, global average pooling compresses spatial dimensions into a 256-dimensional vector. This vector enters the classifier, which has two fully connected layers. Each layer uses ReLU activation, batch normalization, and dropout (default rate: 0.2). These extract high-level discriminative features. Finally, a sigmoid activation normalizes the two output values to the range [0, 1] for binary classification.

Figure 5 illustrates the complete architecture and feature propagation process. Through this hierarchical, symmetric, and interactive framework, the model captures stylistic differences and structural characteristics of signatures at multiple scales. This design improves authentication robustness and generalization.

3.6. Feature repository construction and optimization. In the previous section, the neural network model produced only a binary output. Clearly, such a model alone cannot meet the multi-class requirements of a handwriting classification task. To enable multi-class classification, multiple experiments were conducted to verify the feasibility of a feature library-based classification approach. In this method, several images are randomly selected from each category to form a "reference image feature library", which assists the network in the final classification stage.

Specifically, representative samples are randomly chosen from each class and fed into the model for forward inference. The feature vector from the GAP layer is extracted, and only samples whose predicted labels match their ground-truth labels are retained. This ensures both feature accuracy and class discriminability. For each valid sample, its features are added to the library according to its label: if the label has not appeared before, a new data structure is initialized to store that class's feature vectors and image paths. Before a new sample is added, its feature vector is compared with the existing vectors of the same class using cosine similarity. If the number of samples for that class has not reached the preset similarity threshold (a manually defined minimum cosine similarity), the sample is added to the library. For efficient retrieval and similarity computation, all compressed features and their labels are serialized into binary files in .joblib format, forming an initial feature library with compactness, representativeness, and strong discriminative power.

Since the quality of a randomly generated feature library is unstable—for example, some images may have many strokes while others have few—two optimizations were introduced to improve both the library quality and matching accuracy for test samples. First, high-quality samples with rich handwriting features and clear strokes were selected to replace low-quality samples. As shown in Fig. 6, a simple Chinese character such as "one" was replaced with a more complex character "leather", or an English word like "as" with "brought". Increasing the stroke count and complexity of reference samples provides the network with richer and more useful features during inference.

Second, the number of reference images per class in the library was increased. This enhances the completeness of feature representation and prevents classification errors caused by relying on a single, possibly atypical reference image. For instance, an initial library may contain three images per class; if performance is unsatisfactory, more images are added. Experiments showed that increasing to about 25 images per class—after applying the first optimization—yields the best results, with no further accuracy gains beyond that point. This strategy maximizes coverage of potential features and stroke patterns, including those in unforeseen characters (e.g., the handwriting characteristics of all 26 English letters or common Chinese strokes), ensuring that each class's

feature representation is both comprehensive and robust.

Through the first optimization, the compressed features in the refined library have higher quality, and the selected images have greater clarity, enabling more precise feature matching in the classification stage. In the experiments, the optimized feature library improved the model's accuracy on both 16-class and 20-class datasets. It should be noted that, due to class imbalance in the dataset, the number of stored features per category is not identical. After one full cycle of comparisons through the library, the category with the largest number of model outputs equal to 1 is taken as the final classification result.

The optimized feature library plays a crucial role in improving model accuracy. Through this construction and optimization approach, the library can assist the model in performing multi-class handwriting identification tasks.

4. Experiment

4.1. Dataset. In this study, to assess the feasibility of handwriting encryption, we first selected the Chinese Calligraphy Styles by Calligraphers (CCSbC) dataset released on the Alibaba Cloud Tianchi platform. The dataset contains handwritten Chinese character images from 20 renowned Chinese calligraphers. Each calligrapher contributes between 1,000 and 7,000 images. Each image has a resolution of 64×64 pixels in JPG format, with an average of about 5,251 images per class. For ease of training and label standardization, we replaced the original abbreviated name labels of the calligraphers with numeric codes.

Although CCSbC provides abundant brush-written samples, it is primarily composed of Chinese calligraphy with relatively fixed styles. Moreover, English handwriting features tend to emphasize dynamic characteristics, whereas Chinese character handwriting features emphasize static characteristics (Han *et al.*, 2007). Therefore, the CCSbC dataset cannot fully meet the modeling needs for the stylistic diversity of everyday pen-based writing. To improve the model's adaptability in practical applications, we introduce a mixed Chinese–English handwriting dataset, MDC (Multilingual Handwriting Dataset), to extend generalization across multiple languages and writing scenarios, as shown in Fig. 7. The dataset includes approximately 9,000 images across eight classes for English and about 10,000 images across eight classes for Chinese. It covers various writing instruments and contexts, with rich stylistic diversity, better satisfying modeling needs in multilingual and multi-writing scenarios. MDC provides a more robust foundation for subsequent model training and evaluation, helping to improve performance in traditional calligraphy recognition, practical handwriting verification, and encryption tasks.

Fig. 6. Replace "one" with "leather" and replace the English word "as" with "brought" in the image.

Fig. 7. Partial sample examples of the MDC dataset.

4.2. Learning process. The experiments were conducted on a workstation equipped with an NVIDIA RTX 3060 GPU, enabling efficient model training and inference through GPU acceleration. The experimental environment includes Python 3.8.19, PyTorch 2.3.1, OpenCV 4.10.0, and NumPy 1.24.1. The model was optimized using the Adam optimizer, which is an adaptive learning rate optimization algorithm that combines the advantages of the momentum and RMSProp algorithms. RMSProp is an optimization algorithm that adjusts the learning rate to accelerate the gradient descent process. It can automatically adjust the learning rate, thereby achieving faster convergence and better stability during the training process (Kingma and Ba, 2015). During training, the initial learning rate is set to 0.0001 and the batch size is 24, and the model is trained for 200 epochs. The model was optimized using the Adam optimizer, with an initial learning rate set to 0.0001, a batch size of 24, and was trained for 200 epochs. For all datasets, such as CCSbC, MDC, and CCD-CQU, 80% of the data was used for training and 20% for testing. The cross-entropy loss function was employed for optimization, and model performance was evaluated using accuracy (ACC), the false acceptance rate (FAR), and the false rejection rate (FRR).

During the preprocessing stage, all images were loaded and converted into grayscale to reduce computational complexity and focus on structural features. Subsequently, Otsu's thresholding method was applied to binarize the images, enhancing contrast and improving the model's robustness to low-quality samples. To preserve the aspect ratio of the images, a customized resizing approach was implemented: images were scaled to a target resolution of 64×64 pixels and

padded with white pixels, thus avoiding distortion of structural features. Finally, pixel values were normalized to the [0,1] range to ensure consistency across inputs and accelerate model convergence. Each training instance consisted of two grayscale images and a corresponding label.

The experiment first tests on the CCSbC dataset (20 categories) and evaluates the model's performance under different parameter configurations. The results are shown in Table 1.

As seen in the table, changes in the data volume and dropout rate significantly affect the model's performance. When the data volume is small (200) and the dropout rate is 0.2, the model's FAR (false acceptance rate) and FRR (false respection rate) on the test set are 0.80% and 0.93%, respectively, with an accuracy of 98.46%. When the dropout rate is increased to 0.5, the FAR and FRR rise to 0.88% and 0.98%, respectively, with a slight improvement in classification accuracy to 98.86%. The model demonstrates a certain level of robustness.

As the data volume increases to 300, the model performance improves further. With a dropout rate of 0.2, the FAR and FRR decrease to 0.68% and 0.85%, respectively, and accuracy significantly increases to 99.33%. When the dropout rate is 0.5, the FAR and FRR are 0.82% and 0.92%, respectively, with a slight decrease in accuracy to 99.13%. As the data volume further increases to 400, the model's FAR and FRR at the lower dropout rate (0.2) decrease to 0.66% and 0.83%, respectively, with accuracy reaching 99.31%. Overall, the increase in the data volume effectively reduces both the false acceptance rate and false rejection rate, thereby improving classification accuracy of the model. However, the rise in the dropout rate may slightly increase the error rate, but it has a limited impact on classification accuracy, demonstrating a certain degree of robustness to noise and interference. The test results on the MDC dataset (16 categories) are shown in Table 2.

The results show that, when the data volume is small (150) and the dropout rate is low (0.2), the model's FAR and FRR are relatively high, with an accuracy of 94.69%. As the data volume increases to 200 and the dropout rate remains at 0.2, the FAR and FRR decrease to 10.00% and 10.33%, respectively, and accuracy improves to 95.69%. When the dropout rate increases to 0.5, although the data volume is still 200, both the FAR and FRR increase significantly, and accuracy drops to 90.25%, indicating that a higher dropout rate has a negative impact on the model performance. When the data volume is further increased to 300 and the dropout rate is 0.2, the FAR and FRR decrease to 9.33% and 10.00%, respectively, and accuracy increases to 96.14%. However, when the dropout rate increases to 0.5, the FAR and FRR rise to 11.00% and 11.00%, respectively, and accuracy drops to 90.17%.

Table 1. Performance evaluation of the CCSbC dataset.

Num	Drop	FAR [%]	FRR [%]	Acc [%]
200	0.2	0.82	0.93	98.46
200	0.5	0.88	0.98	98.86
300	0.2	0.68	0.85	99.33
300	0.5	0.82	0.92	99.13
400	0.2	0.66	0.83	99.31

Table 2. Performance evaluation of the MDC dataset.

Num	Drop	FAR [%]	FRR [%]	Acc [%]
150	0.2	12.83	12.67	94.69
200	0.2	10.00	10.33	95.69
200	0.5	11.50	12.00	90.25
300	0.2	9.33	10.00	96.14
300	0.5	11.00	11.00	90.17

Table 3. Performance evaluation of the CCD-CQU dataset.

Num	Drop	FAR [%]	FRR [%]	Acc [%]
150	0.2	2.85	2.67	90.83
200	0.2	2.50	2.15	91.25
200	0.5	2.92	2.63	90.25
300	0.2	2.67	1.33	98.00
300	0.5	2.78	1.92	90.17

In addition, this study uses the trained model for the feature library classification task. The test results are shown in Table 3.

When the data volume is small (150) and the dropout rate is low (drop $= 0.2$), the model's FAR is 2.85%, the FRR is 2.67%, and accuracy is 90.83%. When the data volume increases to 200 and the dropout rate remains at 0.2, the FAR and FRR decrease to 2.50% and 2.15%, respectively, and accuracy slightly improves to 91.25%. However, when the dropout rate increases to 0.5, the FAR and FRR rise to 2.92% and 2.63%, respectively, and accuracy drops to 90.25%.

When the data volume is further increased to 300 and the dropout rate is 0.2, the FAR and FRR decrease to 2.67% and 1.33%, respectively, and accuracy significantly improves to 98.00%. Even under the condition of drop $=$ 0.5, the FAR and FRR are 2.78% and 1.92%, respectively, with accuracy remaining at 90.17%. These results show that the model performance on the CCD-CQU dataset significantly improves as the data volume increases, especially with a low dropout rate, where the model performs best.

Based on the above experimental results, it is evident that increasing the data volume significantly improves the model's classification performance, especially under low dropout conditions (drop $= 0.2$). However, even with limited sample sizes, the training method proposed in this paper can still train the model effectively. The FAR and FRR of the model decrease significantly

Table 4. Matching classification data.

Dataset	FAR [%]	FRR [%]	Acc [%]
CCSbC	–	–	97.13
MDC	–	–	92.86
CCD-CQU	–	–	95.63

Table 5. Comparative experiment.

Class	Type1(C1) [%]	Type1(C2) [%]	Our [%]
O	93.8	92.1	94.56
C	98.3	96.2	96.67
L	97.0	95.2	96.03
Y	95.7	93.0	95.26

on all datasets, and accuracy increases substantially, demonstrating a strong generalization ability. However, a higher dropout rate (drop $=$ 0.5) may cause some information loss, which negatively impacts classification performance, particularly on the MDC and CCD-CQU datasets. Overall, sufficient data volume and reasonable dropout rate configuration are key to optimizing model performance. Although accuracy slightly improves with further increases in the data volume, the model is already able to meet the requirements in small sample scenarios. In the future, data augmentation and feature optimization can further improve the model's robustness and classification ability. This study utilizes the trained model and feature library to perform classification tasks by matching the trained model to the test sets of multiple datasets. Each test sample is matched with every image in the feature library for classification. The results are shown in Table 4. According to the experimental data, accuracy is generally lower than when the model performs inference alone, which is a normal phenomenon (the model cannot guarantee correct inference, and classification will fail when the inference is incorrect). Despite the accuracy being limited, this experiment validates that through feature library matching, within an acceptable error range, the model can not only output 0 or 1 but also directly provide the final inference result. For example, this inferred image belongs to category 18, and the next one belongs to category 4.

In comparison with the experiment by Huang *et al*. (2023), where the highest accuracy after parameter tuning was 96%, this paper tests the CCD-CQU dataset composed of data from Ouyang Xun (O), Chu Siliang (C), Yan Zhenqing (Y), and Liu Gongquan (L) using the proposed method. The highest accuracy achieved in this study is 95.63%. Although the final classification accuracy in this paper is lower than that obtained by Huang *et al*. (2023), this work focuses more on maintaining high accuracy by identifying differences between two images, even with a small number of samples. This approach is particularly suitable for small sample scenarios, and the accuracy per category is quite consistent, demonstrating its robustness and practical value in limited data situations. In the future, further optimization of feature extraction and data augmentation techniques may help achieve higher classification performance. A comparison with the selected test accuracies of configuration types 1 (C1) and 2 (C2) by Huang *et al*. (2023) at $K = 32$ is shown in Table 5.

In the experiments, our method achieved significant performance improvements across multiple datasets. Compared with the multi-scale bimodal fusion network proposed by Xu *et al*. (2024), our model demonstrated greater robustness and adaptability when handling small-sample data. This indicates that the handwriting identification system designed in this paper holds substantial application value in identification tasks.

5. Conclusion

This study focused on the challenges of handwriting identification in few-shot scenarios, specifically addressing the problem of accurately determining whether different texts are written by the same individual. A dual-branch feature fusion network was designed, combined with a feature library matching method, to enhance recognition accuracy and generalization capability under limited data conditions. By constructing a dual-branch feature fusion architecture and introducing an attention mechanism, the proposed model significantly improves the ability to capture key stroke features from handwriting images while effectively reducing the interference caused by background noise. Moreover, by employing a training strategy based on comparing original and inverted images, the model gains a deeper understanding of fine-grained stroke regions, further enhancing its adaptability to different writing contents. Experimental results on the CCSbC, MDC, and CCD-CQU datasets demonstrated that the model achieves high accuracy and robustness in few-shot scenarios, showcasing strong adaptability to diverse handwriting styles. To address the scarcity of images per class in the dataset, this paper proposed a training strategy and data generation method tailored to small-sample scenarios. Signature images were read from a designated directory to automatically construct positive pairs (different signatures from the same writer) and negative pairs (signatures from different writers), which were then randomly shuffled and saved as the training and test sets. By using uniform sampling to control the data size and to ensure equal numbers of positive and negative samples, the approach increased data diversity and improves the model's generalization ability, thereby alleviating data insufficiency and enabling stable convergence under limited-sample conditions. Despite the promising results

achieved in this study, there are still some limitations. On the one hand, although the current datasets cover Chinese calligraphy and mixed-script handwritten texts in Chinese and English, the diversity and scale remain limited, which restricts the broader application of the model. On the other hand, for handwriting with highly unique styles or extremely complex strokes, the model's ability to capture and recognize fine details may decline. Future research should focus on expanding the variety and scale of the datasets and incorporating more advanced feature extraction and matching techniques to further enhance the model's adaptability to complex handwriting styles, thereby better meeting the demands of real-world applications.

Acknowledgment

This research was made possible with the strong support of the Hubei Province Engineering Research Center for Digitalization and Intelligent Manufacturing Technology and Applications. We are deeply grateful for their assistance, as the resources and support provided by the center were crucial to our research on the application of deep learning and intelligent analysis technologies in handwriting identification. Their contribution played a decisive role in the successful completion of this project.

References

Abbas, F., Gattal, A., Djeddi, C., Siddiqi, I., Bensefia, A. and Saoudi, K. (2021). Texture feature column scheme for single-and multi-script writer identification, *IET Biometrics* **10**(2): 179–193.

Ali, S. and Abdulrazzaq, M. (2023). A comprehensive overview of handwritten recognition techniques: A survey, *Journal of Computer Science* **19**(5): 569–587.

Aljehani, A., Hasan, S.H. and Khan, U.A. (2024). Advancing text classification: A systematic review of few-shot learning approaches, *International Journal of Computing and Digital Systems* **16**(1): 1–14.

Askari, F., Fateh, A. and Mohammadi, M.R. (2025). Enhancing few-shot image classification through learnable multi-scale embedding and attention mechanisms, *Neural Networks* **187**: 107339, DOI: 10.1016/j.neunet.2025.107339.

Bhattacharya, I., Ghosh, P. and Biswas, S. (2013). Offline signature verification using pixel matching technique, *Procedia Technology* **10**: 970–977, DOI: 10.1016/j.protcy.2013.12.445.

Chao-Qun, L., Da-Han, W., Shun-Xin, X., Xue-Ke, C., Chi-Ming, W., Xu-Yao, Z. and Shun-Zhi, Z. (2024). Offline handwriting verification based on Siamese network and multi-channel fusion, *Acta Automatica Sinica* **50**(8): 1–11.

Dey, S., Dutta, A., Toledo, J.I., Ghosh, S.K., Lladós, J. and Pal, U. (2017). SigNet: Convolutional Siamese network for writer independent offline signature verification, *arXiv* 1707.02131.

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J. and Houlsby, N. (2020). An image is worth 16×16 words: Transformers for image recognition at scale, *arXiv* 2010.11929.

Han, D.-y., Wang, J.-w., Li, Z. and Li, Y. (2007). Comparative studies on handwriting features of Chinese and english scripts, *Criminal Technology* **4**: 16–18, DOI: 10.16467/j.1008-3650.2007.04.006.

Hossain, S.G.S., Ghosh, M., Obaidullah, S.M. and Roy, K. (2025). Writer identification using cross-script signature images, *in* S. Kumar *et al.* (Eds), *5th Congress on Intelligent Systems*, Springer Nature Singapore, Singapore, pp. 45–55.

Huang, Q., Li, M., Agustin, D., Li, L. and Jha, M. (2023). A novel CNN model for classification of chinese historical calligraphy styles in regular script font, *Sensors* **24**(1): 197.

Kai, H., Hongyue, M., Xu, F. and Kun, L. (2020). English handwriting identification method using an improved VGG-16 model, *Journal of Tianjin University (Science and Technology)* **53**(9): 984–990.

Kingma, D. and Ba, J. (2015). Adam: A method for stochastic optimization, *3rd International Conference on Learning Representations (ICLR), San Diego, USA*.

Kleber, F., Fiel, S., Diem, M. and Sablatnig, R. (2013). CVL-DATABASE: An off-line database for writer retrieval, writer identification and word spotting, *2013 12th International Conference on Document Analysis and Recognition, Washington DC, USA,* pp. 560–564, DOI: 10.1109/ICDAR.2013.117.

Koch, G., Zemel, R. and Salakhutdinov, R. (2015). Siamese neural networks for one-shot image recognition, *ICML Deep Learning Workshop, Lille, France,* Vol. 2, pp. 1–30.

Krizhevsky, A., Sutskever, I. and Hinton, G.E. (2012). ImageNet classification with deep convolutional neural networks, *Advances in Neural Information Processing Systems* **25**: 1097–1105.

Li, H., Wei, P., Ma, Z., Li, C. and Zheng, N. (2022). Offline signature verification with transformers, *2022 IEEE International Conference on Multimedia and Expo (ICME), Taipei, Taiwan,* pp. 1–6, DOI: 10.1109/ICME52920.2022.9859886.

Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S. and Guo, B. (2021). Swin transformer: Hierarchical vision transformer using shifted windows, *Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, Canada*, pp. 10012–10022.

Marti, U.-V. and Bunke, H. (2002). The IAM-database: An English sentence database for offline handwriting recognition, *International Journal on Document Analysis and Recognition* **5**(1): 39–46, DOI: 10.1007/s100320200071.

Ouyang, D., He, S., Zhang, G., Luo, M., Guo, H., Zhan, J. and Huang, Z. (2023). Efficient multi-scale attention module with cross-spatial learning, *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Rhodes. Greece*, pp. 1–5.

Ren, J.-X., Xiong, Y.-J., Zhan, H. and Huang, B. (2023). 2C2S: A two-channel and two-stream transformer based framework for offline signature verification, *Engineering Applications of Artificial Intelligence* **118**: 105639, DOI: 10.1016/j.engappai.2022.105639.

Sánchez-DelaCruz, E. and Loeza-Mejía, C.-I. (2024). Importance and challenges of handwriting recognition with the implementation of machine learning techniques: A survey, *Applied Intelligence* **54**(8): 6444–6465.

Simonyan, K. and Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition, *arXiv* 1409.1556.

Singla, A. and Mittal, A. (2025). Exploring offline signature verification techniques: A survey based on methods and future directions, *Multimedia Tools and Applications* **84**(6): 2835–2875.

Van Drempt, N., McCluskey, A. and Lannin, N.A. (2011). A review of factors that influence adult handwriting performance, *Australian Occupational Therapy Journal* **58**(5): 321–328.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł. and Polosukhin, I. (2017). Attention is all you need, *Advances in Neural Information Processing Systems* **30**: 6000–6010.

Wang, Y., Yao, Q., Kwok, J.T. and Ni, L.M. (2020). Generalizing from a few examples: A survey on few-shot learning, *ACM Computing Surveys* **53**(3): 1–34.

Wei, P., Li, H. and Hu, P. (2019). Inverse discriminative networks for handwritten signature verification, *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, USA*, pp. 5764–5772.

Xu, Z., Chen, Z., Wu, Y., Li, H., Lv, W., Jin, L. and Wang, Q. (2024). A multi-scale bimodal fusion network for robust and accurate online handwriting recognition, *ICASSP 2024: IEEE International Conference on Acoustics, Speech and Signal Processing, Seoul, South Korea*, pp. 6460–6464.

Zhang, P., Jiang, J., Liu, Y. and Jin, L. (2022). MSDS: A large-scale Chinese signature and token digit string dataset for handwriting verification, *Advances in Neural Information Processing Systems* **35**(2645): 36507–36519.

Zhao, H. and Li, H. (2023). Handwriting identification and verification using artificial intelligence-assisted textural features, *Scientific Reports* **13**(1): 21739.

Zhao, P., Wang, L., Zhao, X., Liu, H. and Ji, X. (2024). Few-shot learning based on prototype rectification with a self-attention mechanism, *Expert Systems with Applications* **249**(A): 123586, DOI: 10.1016/j.eswa.2024.123586.

Xuyang Wang received his BS degree in computer science and technology from the Engineering & Technology College of the Hubei University of Technology, China, in 2023. He is currently pursuing his MS degree in computer science and technology at the School of Computer Science, Hubei University of Technology. His research interests include deep learning, computer vision, and information processing. He is engaged in related research work.

Chengzhi Xu received his BS and MS degrees in computer application technology from the Wuhan University of Technology, China, in 2004 and his PhD degree in computer software and theory from Wuhan University, China, in 2010. He worked as a visiting scholar at Oxford Brookes University, UK, in 2015. He is currently a lecturer at the Hubei University of Technology. His research interests include machine learning, computer vision, and SLAM. His research focuses on the application of deep learning in computer vision for complex scenarios, as well as the integration of multimodal data and the application of attention mechanisms in system optimization.

Liuyuan Dong received his BS degree in computer science and technology from the School of Computer Science, Hubei University of Technology, China, in 2023. He is now pursuing his MS degree in computer science and technology at the Hubei University of Technology. His research interests include deep learning, EEG signal processing, and information processing. He is engaged in related research at the School of Computer Science, Hubei University of Technology.

Ruizhen Xie received his BS degree in biomedical engineering from the School of Automation, Nanjing University of Aeronautics and Astronautics, China, in 2024. He is currently pursuing his MS degree in computer science and technology at the School of Computer Science and Technology, Hubei University of Technology. His research interests include deep learning, EEG signal processing, and information processing. He is engaged in related research at the School of Computer Science and Technology, Hubei University of Technology.

Wanli Yang received his BE degree in automation from the School of Artificial Intelligence and Automation, Huazhong University of Science and Technology, China, in 2023. He is currently pursuing his MS degree in control engineering at the School of Artificial Intelligence and Automation, Huazhong University of Science and Technology. His research focuses on deep learning, computer vision, and information processing.

Received: 29 April 2025
Revised: 13 August 2025
Re-revised: 22 September 2025
Accepted: 23 October 2025