

# AuthSig: Safeguarding Scanned Signatures Against Unauthorized Reuse in Paperless Workflows

Ruiqiang Zhang<sup>1</sup>, Zehua Ma<sup>1\*</sup>, Guanjie Wang<sup>1</sup>, Chang Liu<sup>1</sup>, Hengyi Wang<sup>1</sup>, Weiming Zhang<sup>1\*</sup>

<sup>1</sup>Anhui Province Key Laboratory of Digital Security, University of Science and Technology of China

## Abstract

With the deepening trend of paperless workflows, signatures as a means of identity authentication are gradually shifting from traditional ink-on-paper to electronic formats. Despite the availability of dynamic pressure-sensitive and PKI-based digital signatures, static scanned signatures remain prevalent in practice due to their convenience. However, these static images, having almost lost their authentication attributes, cannot be reliably verified and are vulnerable to malicious copying and reuse. To address these issues, we propose AuthSig, a novel static electronic signature framework based on generative models and watermark, which binds authentication information to the signature image. Leveraging the human visual system’s insensitivity to subtle style variations, AuthSig finely modulates style embeddings during generation to implicitly encode watermark bits-enforcing a One Signature, One Use policy. To overcome the scarcity of handwritten signature data and the limitations of traditional augmentation methods, we introduce a keypoint-driven data augmentation strategy that effectively enhances style diversity to support robust watermark embedding. Experimental results show that AuthSig achieves over 98% extraction accuracy under both digital-domain distortions and signature-specific degradations, and remains effective even in print-scan scenarios.

## Introduction

Handwritten signatures have long been relied upon for identity authentication in financial, legal, and other official contexts, and they continue to play a role in modern social credit systems. Yet, with ongoing digital transformation and the shift toward paperless workflows (Sellen and Harper 2003), electronic signatures are increasingly preferred over ink-on-paper ones. According to the E-SIGN Act (E-SIGN Act 2000), an electronic signature is “an electronic sound, symbol, or process” that is attached to or logically associated with a contract or other record and executed by a person with the intent to sign. Common implementations include digital certificates, email confirmations, and password-protected workflows. Consequently, numerous online signature platforms have emerged, such as DocuSign, Adobe Sign, and Signaturit (Alexander 2019), which fall into three main paradigms: dynamic signatures (Okawa 2021, 2020), which

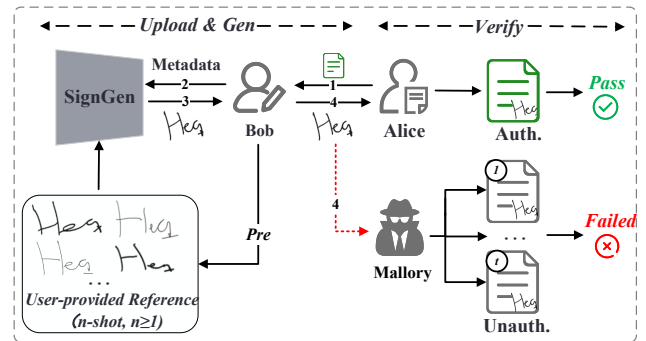


Figure 1: System overview. System overview: Bob pre-registers  $n$  signature samples. The system generates a metadata-bound signature for Alice’s request, valid only for this verification. Any reuse with mismatched documents will fail authentication. The numbered arrows show the chronological sequence of operations.

leverage dedicated hardware to capture writing dynamics (e.g., pen pressure, velocity) as used by platforms like Signaturit to validate authenticity via behavioral biometrics; PKI-based digital signatures, where services such as DocuSign and Adobe Sign employ certificates issued by Certificate Authorities to bind user identity and ensure document integrity; and static scanned-signature images, where users capture an image of their handwritten signature via scanning or photography and paste it into a digital document.

Despite their technical merits, current digital-signature methods generally suffer from a detachment between signature appearance and authentication attributes. In the first two paradigms, dynamic signatures rely on biometric handwriting features and PKI digital signatures rely on cryptographic algorithms; in both cases, the authentication process depends on externally derived attributes (dynamic features or PKI certificates) rather than on the visual presentation of the signature. Furthermore, these methods are often encumbered by expensive proprietary hardware and poor cross-platform compatibility. The third paradigm, static scanned signature images, further degrades security. Verification of these images lacks reliable biometric support, and their reusability severs the essential link between the signing act and the spe-

\*Corresponding author.

cific document content. Nonetheless, their operational simplicity and visual clarity sustain their prevalence in low-risk contexts (e.g., internal approvals, non-binding accords). Yet, if misappropriated, a static signature image can be illicitly reused to forge documents, with no post-hoc mechanism to conclusively attribute authorship, thereby exposing organizations and individuals to dispute and liability.

To overcome these limitations, we pursue a watermark-based static electronic signature framework that enables signature-verification binding and natural sign-authenticate workflows. As illustrated in Fig. 1, upon Alice’s authorization request, Bob encodes metadata (document, signatory, expiration, etc.) into a watermark bitstream. He then combines the watermark with his pre-registered signature samples to generate a one-time-use watermarked signature image, which is delivered to Alice. During verification, the watermark is extracted and validated, and associated metadata are queried to enable granular checks (e.g. document binding, expiration). When an attacker illicitly reuses the signature image, it won’t pass system verification, enforcing the “One Signature, One Use” mechanism.

However, the widely used watermarking paradigms, primarily designed for natural images with post-processing methods, are ill-suited to the target framework, as they often cause visible artifacts in smooth regions and lack robustness to typical document transformations when applied to signature images. Thus, we propose a generative signature watermarking framework that fuses watermark with dynamic signature style features and employs a generative reconstruction mechanism for implicit embedding, establishing a strong link between authorization and the signature. Furthermore, to overcome the limitation posed by the small scale of existing handwritten signature datasets on generative model training, a keypoint-driven data augmentation strategy is designed. For embedding via dynamic features without altering the signature’s character structure, we employ supervised contrastive learning to extract signer-invariant content features and jointly train a style encoder with our generator for fine-grained, implicit modeling of signature style.

In summary, the contributions of this paper are as follows:

- We propose AuthSig, a electronic signature framework that leverages watermarking to equip static signatures with verifiability and document-binding properties. This enables an intuitive, natural signing-and-verification process closely resembling traditional ink signatures.
- We devise a keypoint-driven data augmentation strategy tailored for glyph data. This approach effectively mitigates the scarcity of training data by enriching signature sample diversity while meticulously preserving the topological structure of the handwriting.
- Based on a generative model, we constructed a end-to-end watermarking architecture: it fuses watermark information with signature style features in the feature space and leverages generative reconstruction for implicit watermarking. This approach delivers strong robustness against both common digital-domain distortions and those specific to signature scenarios, even including print-scan process.

## Related Work

### Handwriting Imitation

Handwriting synthesis has progressed via GANs (Goodfellow et al. 2020) and diffusion models (Sohl-Dickstein et al. 2015; Ho, Jain, and Abbeel 2020). HiGAN+ (Gan et al. 2022) disentangles content and style for realistic generation but is constrained by training bias. Diffusion-based models—WordStylist (Nikolaidou et al. 2023), CTIDiffusion (Zhu et al. 2023), and One-DM (Dai et al. 2024)—enable cross-lingual and few-shot synthesis with improved fidelity. Recent works like DiffusionPen (Nikolaidou et al. 2024) and StylusAI (Riaz et al. 2024) offer fine-grained control and stylistic generalization. However, these models rely on labeled datasets (e.g., IAM (Marti and Bunke 2002)) and assume consistent writer styles, while signature data often lack content labels and exhibit high intra-writer variability, challenging generalization.

### Signature Image Verification

Offline signature verification has traditionally been framed as a biometric classification task, using methods like SVMs (Pal, Pal, and Blumenstein 2013), Siamese networks (Dey et al. 2017; Avola et al. 2021; Chattopadhyay et al. 2022), and graph-based models (Maergner et al. 2019) to extract writer-specific features from static signature images. While effective on benchmarks like CEDAR, these techniques are often brittle under cross-domain shifts and high intra-writer variability. More critically, they focus mainly on visual or biometric similarity, neglecting signature intent and the risk of unauthorized reuse in digital contexts. Digital watermarking could address this by binding content and enabling usage tracking, but existing methods—designed for documents or natural images—struggle with handwritten signatures. For example, document-based methods rely on structured layouts (Yang et al. 2023), absent in free-form handwriting, while image watermarking techniques target high-frequency textures. Signature images, however, concentrate meaningful pixels within continuous strokes—only about 14%—making such methods less effective. Techniques like HiDDeN (Zhu et al. 2018) and CIN (Ma et al. 2022) use high-frequency embeddings, but watermarks often end up in background regions, compromising imperceptibility. These challenges highlight the difficulty of applying traditional watermarking methods to signature verification.

## Method

AuthSig comprises three sequential stages. First, we abstract each signature’s topology through skeleton extraction and keypoint filtering, followed by TPS-based data augmentation to simulate realistic handwriting variations. Second, as shown in Fig. 2, we employ supervised contrastive learning to derive signer-invariant content features and architect a latent-space diffusion model that integrates style features and structural cues for high-fidelity signature reconstruction. Third, we introduce a diffusion-based implicit watermark framework that injects watermark into style features via an adaptive gated residual fusion mechanism, ensuring robust, imperceptible watermarking within the generated signature.

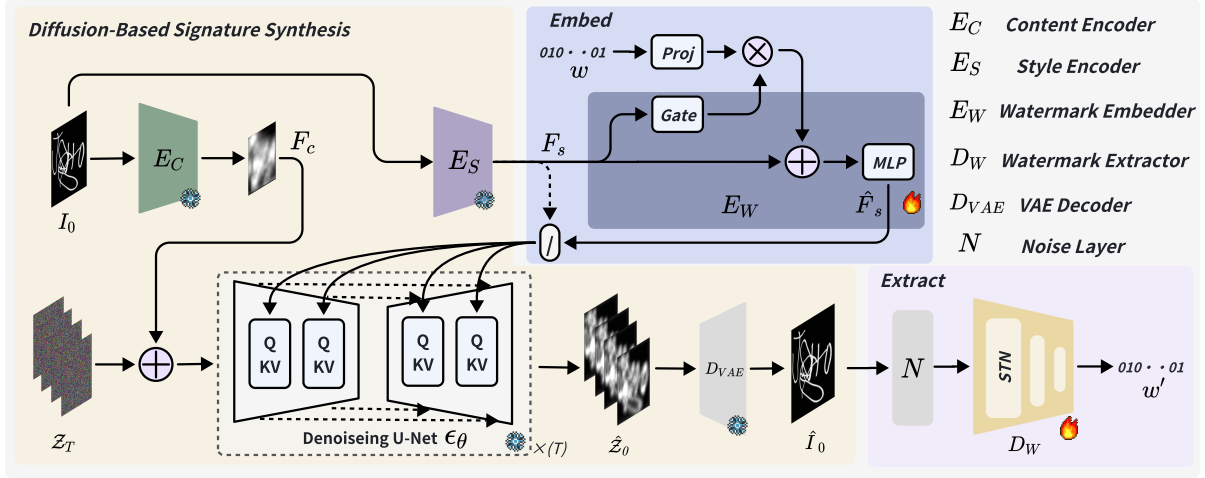


Figure 2: The training process of the proposed watermarking model. **Embedding:** Content feature  $F_c$  is concatenated with noise latent  $z_T$ , while style feature  $F_s$  is fused with watermark bits  $w$  via a gated-residual block; this conditioned input guides the diffusion model to generate the watermarked signature. **Extraction:** Extractor  $E_D$  employs STN for geometric alignment and a multi-layer MLP to recover  $w$ . A distortion layer  $N$  enhances training robustness.

### Keypoint-Based Structural Augmentation

Considering that existing image-augmentation strategies applied directly to signature images easily cause stroke discontinuities and overlaps. We abstract the signature from its spatial pixel form to a more essential topological structure using signature skeletonization and a keypoint filtering mechanism. By subsequently applying a Thin Plate Spline (TPS) transformation to the keypoints, as shown in Fig. 3, we ensure that the resulting augmented samples both retain the structural integrity of the original signature and effectively model the natural variability of authentic signatures.

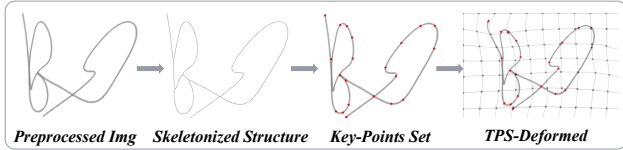


Figure 3: Structural augmentation pipeline: the input signature is skeletonized, keypoints are extracted, and a Thin Plate Spline (TPS) transformation is applied for topologically consistent deformation.

Specifically, for an input signature image  $I \in \mathbb{R}^{H \times W \times 3}$ , we first convert it to grayscale ( $I_{\text{gray}} \in \mathbb{R}^{H \times W}$ ), binarize it ( $I_{\text{bin}} \in \{0, 1\}^{H \times W}$ ), and apply morphological opening to obtain the preprocessed signature  $I_{\text{pre}}$ . Then, the Zhang-Suen thinning algorithm is adopted for skeleton extraction on  $I_{\text{pre}}$ , which provides a set of skeleton points  $\mathcal{S} \subset \mathbb{Z}^2$ . From this set, we extract a simplified set of critical points  $\mathcal{P}$  using the Ramer-Douglas-Peucker (RDP) algorithm:

$$\mathcal{P} = \text{RDP}(\mathcal{S}, \varepsilon) = \{(x_1, y_1), (x_2, y_2), \dots, (x_k, y_k)\} \quad (1)$$

where  $\varepsilon$  governs the density of the extracted points, and each  $(x, y)$  denotes the coordinates of a skeleton keypoint.

Finally, the signature image  $I$  is warped using a Thin Plate Spline (TPS) transformation. We randomly sample a subset of the keypoints:

$$\mathcal{C} = \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\} \subset \mathcal{P} \quad (2)$$

to serve as control points. Each control point  $(x_i, y_i) \in \mathcal{C}$  is perturbed by a random displacement from a truncated Gaussian distribution:  $\delta_{x_i}, \delta_{y_i} \sim \mathcal{N}_{\text{trunc}}(0, \sigma^2, [-k\sigma, k\sigma])$  with  $\sigma = 1$  and a truncation parameter  $k \in \mathbb{R}^+$ . The destination coordinates  $(x'_i, y'_i)$  are clamped within the image bounds:

$$\begin{aligned} x'_i &= \max(0, \min(W, x_i + \delta_{x_i})) \\ y'_i &= \max(0, \min(H, y_i + \delta_{y_i})) \end{aligned} \quad (3)$$

In this manner, the signature image undergoes a smooth, non-rigid deformation that preserves the overall topological structure of the strokes. This process effectively simulates variations in writing dynamics, such as angle and ligature, generating highly realistic samples.

### Signature Synthesis Network

**Feature Decoupling** Let the signature dataset be  $\mathcal{D} = \{(I_k, id_k)\}_{k=1}^N$ , where  $I_k \in \mathbb{R}^{H \times W \times C}$  is an input image and  $id_k$  is the corresponding signer identity label. We hypothesize that a signature's feature representation can be decomposed into two components: a *content feature* and a *style feature*. The former captures semantic information such as character structure and stroke composition, which should remain consistent across all the user's signatures; the latter reflects idiosyncratic signing dynamics and variations unique to the writer. Following this hypothesis, the set of features for all  $m$  samples belonging to a single signer  $id_i$  can be

formally represented as:  $\{C_{s_1}^k, C_{s_2}^k, \dots, C_{s_j}^k\} \subset \mathbb{R}^d$ . In this notation,  $C^k$  denotes the content feature of the  $k$ -th signer, while  $s_j$  denotes the instance-specific style feature of sample  $j$ . As existing datasets lack explicit annotations for signature content, we propose to learn a content feature encoder,  $E_c$ , using a Supervised Contrastive Learning framework.

Given a mini-batch of  $B$  samples drawn from the signature dataset  $\mathcal{D}$ , let

$$c_k = \frac{\text{Proj}(E_c(I_k))}{\|\text{Proj}(E_c(I_k))\|} \in \mathbb{R}^d \quad (4)$$

denote the normalized, projected content embedding of the input image  $I_k$ . Here  $E_c$  is the content-feature encoder and  $\text{Proj}(\cdot)$  is the learnable projection head onto the unit hypersphere. For  $k \in \{1, \dots, B\}$ , define  $P(k) = \{p \in A(k) \mid id_p = id_k\}$  as the set of positive examples sharing the same signer label  $id_k$ , and  $A(k) = \{1, \dots, B\} \setminus \{k\}$  as the set of all other indices in the batch. Therefore, the training objective for  $E_c$  is

$$\mathcal{L}_{SupCon} = \sum_{k=1}^B \frac{-1}{|P(k)|} \sum_{p \in P(k)} \log \frac{\exp(c_k \cdot c_p / \tau)}{\sum_{a \in A(k)} \exp(c_k \cdot c_a / \tau)} \quad (5)$$

where  $\tau$  is a temperature hyperparameter that controls the concentration of similarity scores. By minimizing  $\mathcal{L}_{SupCon}$ , it promotes encoder  $E_c$  to learn feature representations that possess high cohesion for same-identity signatures and high separability for different-identity signatures, compelling  $E_c$  to extract content features related to character structure and semantic information.

**Diffusion-Based Signature Synthesis** Diffusion models typically operate in a low-dimensional latent space to ensure computational efficiency, employing a Variational Autoencoder (VAE) to map images from pixel space to this latent representation (Rombach et al. 2022). However, standard pre-trained VAEs often fail to preserve the fine-grained details of signatures. This stems from the under-representation of intricate glyph data in their original training corpora, leading to significant information loss during reconstruction. To address this, we perform domain-specific fine-tuning on the VAE before using it in our diffusion pipeline. The process is guided by a composite objective function  $\mathcal{L}_{VAE}$ , which combines three distinct loss terms:

$$\mathcal{L}_{VAE} = \mathcal{L}_{rec} + \mathcal{L}_{masked\_rec} + \lambda \mathcal{L}_{KL} \quad (6)$$

Here,  $\mathcal{L}_{rec}$  penalizes pixel-wise reconstruction errors,  $\mathcal{L}_{masked\_rec}$  emphasizes regions marked by a binary mask  $m$ , and  $\mathcal{L}_{KL}$  regularizes the latent space. This targeted tuning enables the VAE to better preserve critical stroke-level details essential for high-fidelity diffusion generation.

Building upon this optimized latent representation, our proposed algorithm for signature image reconstruction leverages a Denoising Diffusion Probabilistic Model (Ho, Jain, and Abbeel 2020). To achieve high-fidelity and identity-consistent reconstruction, two distinct types of conditional features are incorporated to guide this denoising

process. As shown in Fig.2,  $F_s$  is the style feature of the signature image, extracted by the style encoder  $E_s$  to capture the signer’s dynamic and individualistic characteristics and integrated via a cross-attention mechanism, whereas  $F_c$ , obtained from the content encoder  $E_c$ , encodes the structural information of the signature strokes and is fused with the timestep embedding vector  $t_{emb}$  at each denoising block. Accordingly, the training objective at this stage is defined as

$$\mathcal{L}_{denoise}(\theta) = \mathbb{E}_{Z_0 \sim q(Z_0), t \sim \mathcal{U}[1, T], \epsilon \sim \mathcal{N}(0, \mathbf{I})} \left[ \|\epsilon - \epsilon_\theta(Z_t, t, F_s, F_c)\|^2 \right] \quad (7)$$

Here,  $Z_t$  is obtained via the reparameterization trick. During inference, we initiate the reverse diffusion process from a pure noise sample  $Z_T \sim \mathcal{N}(0, \mathbf{I})$  and iteratively apply the reverse steps. To accelerate this sampling, we employ Denoising Diffusion Implicit Models (DDIM) (Song, Meng, and Ermon 2020). Once the denoised latent representation  $\hat{Z}_0$  is obtained, a VAE decoder  $\mathcal{D}$  reconstructs the final signature image  $\hat{I}_0 = \mathcal{D}(\hat{Z}_0)$ . This framework balances computational complexity with generation quality by performing diffusion in the latent space, while the integration of conditional features critically ensures both the identity consistency and structural accuracy of the generated signatures.

## Watermark Embedding & Extraction

To overcome the challenges of quality degradation and limited capacity associated with directly embedding bit-level information into signature images, we introduce a diffusion-based framework for robust bit embedding and extraction, as illustrated in Fig. 4. By fusing the binary information with semantic image features and injecting it as guidance into the diffusion model, our method achieves a balanced trade-off among embedding capacity, robustness, and visual fidelity.

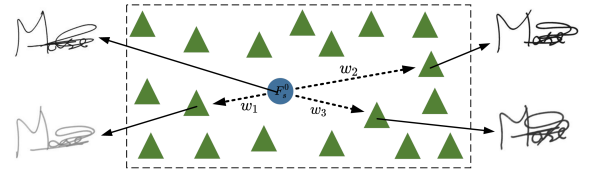


Figure 4: Schematic diagram of watermark embedding process.

The process begins by decomposing the input image  $I$  into content features  $F_c$  and style features  $F_s$  using content encoder  $E_c$  and style encoder  $E_s$ .  $F_c$  captures intrinsic signature characteristics (e.g., strokes and structure), while  $F_s$  reflects signer-specific stylistic variations. In practice,  $F_c$  can be extracted from any signature image of the signer, and the reference style feature  $F_s^0$  is computed from a few prior signature samples. We embed watermark  $w$  (where  $w_i \in \{0, 1\}$ ) into  $F_s^0$  to preserve semantic via a gated residual fusion in watermark embedder  $E_W$ :

$$F'_s = F_s^0 + g \cdot \text{Proj}(w) \quad (8)$$

where  $\text{Proj}(\cdot)$  is a projection layer and  $g$  is a learned gating parameter. Then  $F'_s$  serves as the conditioning input for the

diffusion model to reconstruct an image  $\hat{I}$  that carries the embedded information.

The extraction process reverses this pipeline. The watermark extractor  $D_W$  first applies a Spatial Transformer Network (STN) (Jaderberg et al. 2015) to correct the input image  $\hat{I}$ . It then recovers the fused feature  $\hat{F}'_s$ , concatenates it with a pre-defined  $F_s^0$ , which serves as a signer-specific prior. Unlike generic watermark verification, signature verification assumes a known identity, making it natural and feasible to access the corresponding reference style feature. The combined features are fed into an MLP to predict the embedded bits:

$$w' = MLP([\hat{F}'_s \oplus F_s^0]) \quad (9)$$

During training, we freeze the parameters of  $E_c$ ,  $E_s$ , the denoising model  $\epsilon_\theta$ , and the VAE, optimizing only  $E_W$  and  $D_W$ . The total loss is defined as:

$$\mathcal{L}_{\text{total}} = \lambda_1 \mathcal{L}_{\text{secret}} + \lambda_2 \mathcal{L}_{\text{recover}} + \lambda_3 \mathcal{L}_{\text{style}} + \lambda_4 \mathcal{L}_{\text{content}} \quad (10)$$

The individual losses are formulated as:

$$\mathcal{L}_{\text{secret}} = BCE(w, w'), \quad \mathcal{L}_{\text{recover}} = MSE(F_s^0, \hat{F}'_s) \quad (11)$$

$$\mathcal{L}_{\text{style}} = MSE(F_s^0, F'_s) + \lambda_t \cdot ReLU(m - \|F_s^0 - F'_s\|_2) \quad (12)$$

Finally, to maintain the structural integrity of the signature,  $\mathcal{L}_{\text{content}}$  is formulated using the Contextual Loss (Mechrez, Talmi, and Zelnik-Manor 2018):

$$\mathcal{L}_{\text{content}} = -\log \left( CX \left( E_c(I), E_c(\hat{I}) \right) \right) \quad (13)$$

where  $CX(\cdot, \cdot)$  measures the similarity between feature sets.

## Experiments

### Implementation Details

We train our model on the GPDS 1–150 dataset (Vargas et al. 2007), which comprises 150 signers with 54 signature images each. We randomly allocate 120 signers’ signatures to the training set and reserve the remaining 30 for evaluation. Training samples undergo our proposed augmentation pipeline, hereafter denoted as GPDS-Aug, which is then used for training the generative model. All experiments are implemented in PyTorch and executed on an NVIDIA RTX A6000 GPU. Prior to training and testing, all images are uniformly resized to  $224 \times 224 \times 3$  pixels. We train with a batch size of 32 using the Adam optimizer. The learning rate is initialized at  $5e-5$  and annealed via a cosine schedule.

For quantitative benchmarking, we establish a multi-facet comparison: besides the conventional LGDR watermarking method (Ma et al. 2021), we evaluate three deep-learning approaches (HiDDeN (Zhu et al. 2018), CIN (Ma et al. 2022), and WAM (Sander et al. 2024)). The baseline methods are implemented according to their original configurations as described in the respective papers. For our method, AuthSig embeds 48 bits of information. Robustness is measured by extraction accuracy (Acc.). Visual fidelity is assessed using HWD (Pippi et al. 2023), FID (Heusel et al. 2017), and KID (Bińkowski et al. 2018); HWD-tailored for handwriting-computes the Euclidean distance between the mean feature embeddings of genuine and synthesized signatures, sensitively reflecting stylistic deviations.

### Visualization Results

We qualitatively evaluate the visual quality of our scheme against four baseline methods: HiDDeN, WAM, LGDR, and CIN. As illustrated in Fig.5, HiDDeN introduces a pronounced background hue shift and blurs signature strokes due to pixel-level perturbations. WAM induces a global color drift and conspicuous green edge artifacts that are easily detected by the human eye. LGDR and CIN, while visually better, still exhibit compression artifacts that degrade fine-stroke details. Such distortions stem from their global embedding strategies and reliance on high-frequency image components. For instance, HiDDeN’s adversarial training enhances robustness on natural images, yet handwritten signatures constitute only roughly 14% of pixels, compelling watermark embedding into extensive background regions and precipitating visual artifacts. Similarly, WAM’s local embedding strategy, though designed to minimize background interference, is ill-suited for the sparse nature of signature data, leading to edge artifacts and color anomalies.

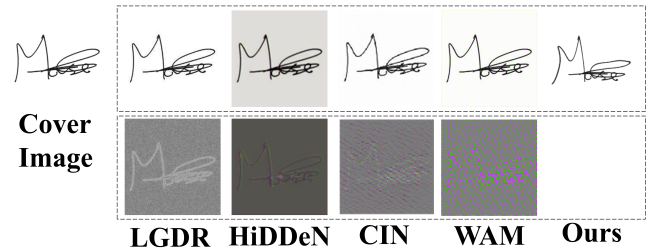


Figure 5: Visual examples of watermarked images  $I_w$  and residuals of the proposed scheme and the compared methods. left: cover image  $I_0$ .

In contrast, AuthSig embeds watermark information by modulating subtle stylistic features of the signature (e.g., local strokes weight, turn curvature). This approach avoids perceptible distortions in the pixel domain, such as blurring, noise, or color shifts. The imperceptibility of our method is rooted in the natural tolerance of the Human Visual System (HVS) to minor variations in signatures. The stylistic appearance of any individual’s signature—including its stroke morphology, ligature flow, and slant—inherently varies due to factors like the writing environment and the signer’s emotional state. This intrinsic variability effectively masks the minute alterations introduced by the watermarking process, rendering them unnoticeable in practical applications. Since our scheme is based on a generative model rather than post-processing techniques used by the baselines, there is no conventional reference image for the watermarked signature, rendering residual computation infeasible.

### Robustness to Common Image Distortions

We evaluate the robustness of AuthSig against six representative image distortions—JPEG compression, Gaussian blur, Gaussian noise, brightness variation, contrast variation, and hue shift. As illustrated in Fig. 6, our approach consistently outperforms baselines across all distortion types and

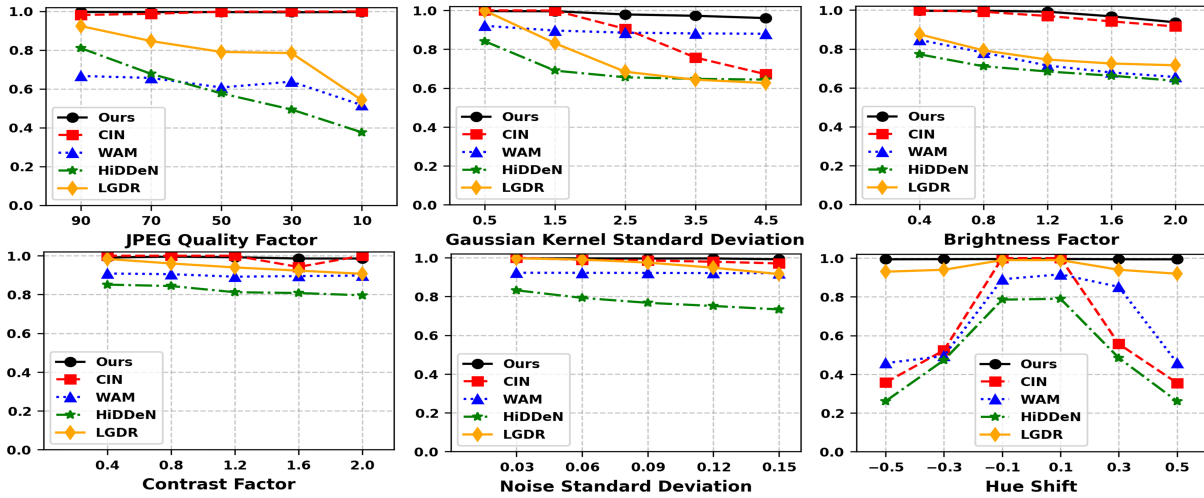


Figure 6: Robustness of our method versus competing watermarking schemes under common digital distortions.

severity levels, maintaining over 98% extraction accuracy in most cases. This robustness arises from embedding watermark within the dynamic topological features of signatures rather than pixel intensities. In contrast, directly applying existing watermarking algorithms to handwritten data presents serious challenges. LGDR and HiDDeN suffer marked Acc drops under moderate to heavy JPEG compression, due to inherent conflicts between their high-frequency embedding strategies and JPEG’s DCT-based quantization. Similarly, existing methods show reduced performance under Gaussian blur, Gaussian noise, and Salt-and-Pepper noise. For instance, HiDDeN and CIN lose over 20% accuracy under strong Gaussian blur. This may be because existing watermark algorithms, which assume embedding in high-frequency details of natural images, fail in such sparse data. The sparse feature distribution characteristics of the handwritten signature dataset (effective pixel ratio about 14%) make existing schemes unable to maintain their robustness.

It is important to note that the algorithm’s performance exhibits a degree of degradation when subjected to high-intensity brightness adjustments. Specifically, when the Brightness Factor reaches 2.0 (indicating a twofold increase in brightness), the Acc. decreases to 93.7%. The underlying cause for this decline is overexposure, which results in the loss of strokes in certain signature samples with lighter handwriting. In extreme cases, this can lead to an image that is almost entirely white. Such an extreme scenario represents a critical state of signature image failure and is highly unlikely to occur in practical applications. Concurrently, comparative results indicate that existing methods degrade even more severely under such conditions.

### Robustness to Office Operations

In digital office workflows, signature images are often subject to ROI-preserving cropping and background transparency, see Appendix A for details. To assess robustness under these common manipulations, we benchmarked AuthSig against state-of-the-art baselines. As show in Table

| Metric              | CIN    | LGDR   | HiDDeN | WAM    | Ours          |
|---------------------|--------|--------|--------|--------|---------------|
| BG. Transparency    | 0.7195 | 0.5632 | 0.5225 | 0.4612 | <b>0.9921</b> |
| ROI-Preserving Crop | 0.7448 | 0.5814 | 0.5516 | 0.5936 | <b>0.9943</b> |

Table 1: Comparison of Different Methods on Background Transparency and Cropping.

1, AuthSig consistently exceeds 99% extraction accuracy for both operations, substantially outperforming competing methods. These results confirm the practical applicability of our scheme in real-world electronic document scenarios.

### Robustness to Cross-Media Distortions

To assess our method’s viability for cross-media signature authentication in real-world office settings, we conducted a three-phase physical experiment. First, we produced 205 watermarked signature images from the GPDS test set. Second, we printed them using two consumer printers (HP M233dw and M180n). Third, we re-digitized the prints with four devices—a Canon LiDE 300 scanner, an iPhone XR, an iPhone 15, and a vivoX100. To reduce device-specific variations, each print-and-scan trial was replicated three times. We report watermark extraction accuracy as our main metric. During mobile captures, phones were tripod-mounted and sheets fixed to a stand; all images were manually cropped and deskewed before decoding.

As shown in Table 2, AuthSig consistently exceeds 97% accuracy across most device pairs. Crucially, it remains robust under challenging conditions—laser-induced ink spread and uneven lighting during handheld capture—matching the 97% accuracy achieved with the uniformly lit LiDE scanner. Standard deviation of  $\pm 0.007$  indicates stability across lighting and handling conditions. These findings demonstrate that our electronic signature scheme preserves the signature’s authentication integrity even when the document is transferred from digital to physical medium (i.e., printed on paper). This capability overcomes limitations of conven-



Figure 7: Our method generates smooth transitions (top), while direct interpolation in the VAE latent space leads to superimposed artifacts from the two input images (bottom).

tional PKI digital signatures, which are restricted to electronic media, addressing the need for robust cross-media document authentication.

| Devices   | iPhoneXR | iPhone15 | vivoX100 | LiDE   |
|-----------|----------|----------|----------|--------|
| HP M233dw | 0.9662   | 0.9841   | 0.9724   | 0.9793 |
| HP M180n  | 0.9712   | 0.9822   | 0.9796   | 0.9879 |

Table 2: Extraction Acc. Across Various Printing and Scanning Device Configurations

### Effectiveness of Feature Decoupling

To evaluate the feature decoupling mechanism, we conducted an ablation study with three variants: Full-Model (content + style), Style-Only (style without content), and Content-Only (content without style). Detailed visual results are provided in Appendix B, and quantitative evaluations are in Table 3. Results show that high-fidelity reconstruction is achievable only when both features are jointly used. In contrast, single-feature models fail to reconstruct meaningful content, exhibiting severe semantic distortion. These findings highlight the necessity of combining content and style features for accurate signature image reconstruction.

Meanwhile, compared to directly using the latent variables of a Variational Autoencoder (VAE) to represent watermark messages, the proposed decoupling strategy models changes in the signature’s style features at a finer granularity, achieving robust watermark embedding while preserving the natural characteristics of handwriting. As shown in Fig. 7, entangled features in the latent spaces of a Variational Autoencoder (VAE) cause ghosting artifacts during feature interpolation. The proposed method effectively mitigates semantic conflicts and produces smooth, intermediate transition effects. Notably, the watermarked signatures are style-guided generations constrained within the distribution of genuine handwriting, ensuring realistic and authentic outputs. This capability further demonstrates that our generative model can freely sample within the feature space, enabling high-quality reconstruction of signature images and thereby providing ample capacity for watermark embedding; see Section D in Appendix.

### Effectiveness of VAE Finetuning

We investigate the effect of fine-tuning a variational autoencoder (VAE) for handwritten signature reconstruc-

| Model        | HWD ↓         | FID ↓          | KID ↓         |
|--------------|---------------|----------------|---------------|
| Style-Only   | 1.7905        | 370.5272       | 0.4785        |
| Content-Only | 1.7865        | 223.3530       | 0.2356        |
| Full-Model   | <b>0.2047</b> | <b>40.8138</b> | <b>0.0418</b> |

Table 3: Evaluation of the feature decoupling mechanism.

tion. Pretrained on generic datasets like LAION, existing VAEs poorly capture the fine-grained stroke patterns and high-frequency details of handwriting. To address these limitations, we fine-tune the VAE backbone from runwayml/stable-diffusion-v1-5 using domain-specific signature data. This adaptation improves latent space expressiveness, enabling more faithful detail preservation. We denote the pretrained model as PT-VAE and the fine-tuned variant as FT-VAE. As shown in Table 4, the Reconstruction Handwritten Wasserstein Distance (rHWD) reduces from 0.1446 to 0.0344 and the Reconstruction Fréchet Inception Distance (rFID) on GPDS 1-150 drops from 5.76 to 1.63. Similar improvements are observed on the CEDAR, GPDS-Aug, and IAM datasets. Qualitative assessments further confirm enhanced continuity and fluency of strokes, surpassing the baseline in stylistic fidelity.

| Dataset  | Model | Evaluation Metrics |               |               |             |
|----------|-------|--------------------|---------------|---------------|-------------|
|          |       | PSNR               | SSIM          | rHWD ↓        | rFID ↓      |
| GPDS-150 | PT    | 32.54              | 0.9878        | 0.1446        | 5.76        |
|          | FT    | <b>44.37</b>       | <b>0.9970</b> | <b>0.0344</b> | <b>1.63</b> |
| CEDAR    | PT    | 33.24              | 0.9814        | 0.1471        | 5.44        |
|          | FT    | <b>43.12</b>       | <b>0.9942</b> | <b>0.0513</b> | <b>1.97</b> |
| GPDS-Aug | PT    | 33.35              | 0.9860        | 0.1564        | 6.18        |
|          | FT    | <b>41.01</b>       | <b>0.9951</b> | <b>0.0414</b> | <b>1.72</b> |
| IAM      | PT    | 34.42              | 0.9906        | 0.1611        | 5.02        |
|          | FT    | <b>40.68</b>       | <b>0.9968</b> | <b>0.0713</b> | <b>0.80</b> |

Table 4: Performance Comparison of Pre-trained VAE (PT) and Fine-tuned VAE (FT)

## Conclusion

We propose a novel electronic signature framework that integrates watermark technology with generative models, imbuing static signatures with verifiability and document-binding capabilities. Extensive evaluations show strong robustness against common digital distortions and signature-specific degradations, even including print-scan process. Future work will focus on increasing watermark capacity and enhancing visual quality.

## Acknowledgments

This work was supported in part by the Natural Science Foundation of China under Grant 62402469, 62121002, 62472398, U2336206, by Fundamental Research Funds for the Central Universities under Grant WK2100000041.

## References

- Alexander, H. S. 2019. Easy Automation. *Law Prac.*, 45: 32.
- Avola, D.; Bigdello, M. J.; Cinque, L.; Fagioli, A.; and Marini, M. R. 2021. R-SigNet: Reduced space writer-independent feature learning for offline writer-dependent signature verification. *Pattern Recognition Letters*, 150: 189–196.
- Bińkowski, M.; Sutherland, D. J.; Arbel, M.; and Gretton, A. 2018. Demystifying mmd gans. *arXiv preprint arXiv:1801.01401*.
- Chattopadhyay, S.; Manna, S.; Bhattacharya, S.; and Pal, U. 2022. Surds: Self-supervised attention-guided reconstruction and dual triplet loss for writer independent offline signature verification. In *2022 26th International Conference on Pattern Recognition (ICPR)*, 1600–1606. IEEE.
- Dai, G.; Zhang, Y.; Ke, Q.; Guo, Q.; and Huang, S. 2024. One-dm: One-shot diffusion mimicker for handwritten text generation. In *European Conference on Computer Vision*, 410–427. Springer.
- 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 preprint arXiv:1707.02131*.
- E-SIGN Act. 2000. Electronic Signatures in Global and National Commerce Act (ESIGN Act). Pub. L. No. 106–229, 114 Stat. 464. Accessed: 2025-07-14.
- Gan, J.; Wang, W.; Leng, J.; and Gao, X. 2022. HiGAN+: handwriting imitation GAN with disentangled representations. *ACM Transactions on Graphics (TOG)*, 42(1): 1–17.
- Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; and Bengio, Y. 2020. Generative adversarial networks. *Communications of the ACM*, 63(11): 139–144.
- Heusel, M.; Ramsauer, H.; Unterthiner, T.; Nessler, B.; and Hochreiter, S. 2017. Gans trained by a two time-scale update rule converge to a local nash equilibrium. *Advances in neural information processing systems*, 30.
- Ho, J.; Jain, A.; and Abbeel, P. 2020. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33: 6840–6851.
- Jaderberg, M.; Simonyan, K.; Zisserman, A.; et al. 2015. Spatial transformer networks. *Advances in neural information processing systems*, 28.
- Ma, R.; Guo, M.; Hou, Y.; Yang, F.; Li, Y.; Jia, H.; and Xie, X. 2022. Towards blind watermarking: Combining invertible and non-invertible mechanisms. In *Proceedings of the 30th ACM International Conference on Multimedia*, 1532–1542.
- Ma, Z.; Zhang, W.; Fang, H.; Dong, X.; Geng, L.; and Yu, N. 2021. Local geometric distortions resilient watermarking scheme based on symmetry. *IEEE Transactions on Circuits and Systems for Video Technology*, 31(12): 4826–4839.
- Maergner, P.; Howe, N. R.; Riesen, K.; Ingold, R.; and Fischer, A. 2019. Graph-based offline signature verification. *arXiv preprint arXiv:1906.10401*.
- 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: 39–46.
- Mechrez, R.; Talmi, I.; and Zelnik-Manor, L. 2018. The contextual loss for image transformation with non-aligned data. In *Proceedings of the European conference on computer vision (ECCV)*, 768–783.
- Nikolaidou, K.; Retsinas, G.; Christlein, V.; Seuret, M.; Sfikas, G.; Smith, E. B.; Mokayed, H.; and Liwicki, M. 2023. Wordstylist: styled verbatim handwritten text generation with latent diffusion models. In *International Conference on Document Analysis and Recognition*, 384–401. Springer.
- Nikolaidou, K.; Retsinas, G.; Sfikas, G.; and Liwicki, M. 2024. DiffusionPen: Towards Controlling the Style of Handwritten Text Generation. In *European Conference on Computer Vision*, 417–434. Springer.
- Okawa, M. 2020. Online signature verification using single-template matching with time-series averaging and gradient boosting. *Pattern Recognition*, 102: 107227.
- Okawa, M. 2021. Time-series averaging and local stability-weighted dynamic time warping for online signature verification. *Pattern Recognition*, 112: 107699.
- Pal, S.; Pal, U.; and Blumenstein, M. 2013. A two-stage approach for English and Hindi off-line signature verification. In *New Trends in Image Analysis and Processing—ICIAP 2013: ICIAP 2013 International Workshops, Naples, Italy, September 9-13, 2013. Proceedings 17*, 140–148. Springer.
- Pippi, V.; Quattrini, F.; Cascianelli, S.; and Cucchiara, R. 2023. HWD: A Novel Evaluation Score for Styled Handwritten Text Generation. In *Proceedings of the British Machine Vision Conference*.
- Riaz, N.; Saifullah, S.; Agne, S.; Dengel, A.; and Ahmed, S. 2024. StylusAI: Stylistic Adaptation for Robust German Handwritten Text Generation. In *International Conference on Document Analysis and Recognition*, 429–444. Springer.
- Rombach, R.; Blattmann, A.; Lorenz, D.; Esser, P.; and Ommer, B. 2022. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 10684–10695.
- Sander, T.; Fernandez, P.; Durmus, A.; Furon, T.; and Douze, M. 2024. Watermark anything with localized messages. *arXiv preprint arXiv:2411.07231*.
- Sellen, A. J.; and Harper, R. H. 2003. *The myth of the paperless office*. MIT press.
- Sohl-Dickstein, J.; Weiss, E.; Maheswaranathan, N.; and Ganguli, S. 2015. Deep unsupervised learning using nonequilibrium thermodynamics. In *International conference on machine learning*, 2256–2265. pmlr.
- Song, J.; Meng, C.; and Ermon, S. 2020. Denoising diffusion implicit models. *arXiv preprint arXiv:2010.02502*.
- Vargas, F.; Ferrer, M.; Travieso, C.; and Alonso, J. 2007. Off-line handwritten signature GPDS-960 corpus. In *Ninth International Conference on Document Analysis and Recognition (ICDAR 2007)*, volume 2, 764–768. IEEE.

Yang, X.; Zhang, J.; Fang, H.; Liu, C.; Ma, Z.; Zhang, W.; and Yu, N. 2023. AutoStegaFont: Synthesizing vector fonts for hiding information in documents. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, 3198–3205.

Zhu, J.; Kaplan, R.; Johnson, J.; and Fei-Fei, L. 2018. Hidden: Hiding data with deep networks. In *Proceedings of the European conference on computer vision (ECCV)*, 657–672.

Zhu, Y.; Li, Z.; Wang, T.; He, M.; and Yao, C. 2023. Conditional text image generation with diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 14235–14245.