



OPEN ACCESS

EDITED BY

Hariharan Shanmugam, Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology, India

REVIEWED BY

R. Kaladevi, Saveetha Engineering College, India
Ridwan Kolapo, Nile University of Nigeria, Nigeria

*CORRESPONDENCE

Mingyu Guan
✉ guanmingyu2024@163.com
Wei Zhang
✉ zhangwei001@cqct.edu.cn

RECEIVED 25 November 2025

REVISED 13 December 2025

ACCEPTED 17 December 2025

PUBLISHED 14 January 2026

CITATION

Zhang K, Guan M, Wu C, Li Y, Lü Q, Liu Y, Wang Y, Wang W and Zhang W (2026) Deep learning-enabled hybrid systems for accurate recognition of text in seal images.

Front. Big Data 8:1753871.

doi: 10.3389/fdata.2025.1753871

COPYRIGHT

© 2026 Zhang, Guan, Wu, Li, Lü, Liu, Wang, Wang and Zhang. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](#). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Deep learning-enabled hybrid systems for accurate recognition of text in seal images

Keke Zhang¹, Mingyu Guan^{2*}, Chao Wu³, Yutong Li³, Qingguo Lü³, Yi Liu⁴, Yi Wang⁴, Wei Wang⁴ and Wei Zhang^{1*}

¹College of Artificial Intelligence and Big Data, Chongqing Polytechnic University of Electronic Technology, Chongqing, China, ²College of Three Gorges Artificial Intelligence, Chongqing Three Gorges University, Chongqing, China, ³College of Computer Science, Chongqing University, Chongqing, China, ⁴Chongqing Ant Consumer Finance Co. Ltd., Ant Group, Chongqing, China

Chinese seals are widely used in various fields within Chinese society as a tool for certifying legal documents. However, recognizing text on these seals presents challenges due to background text, high noise levels, and minimalistic image features. This paper introduces a hybrid model to address these difficulties in Chinese seal text recognition. Our model integrates preprocessing techniques tailored for real seals, a deep learning-based position correction model, a circular text unwrapping model, and OCR text recognition. First, we apply a color-based method to effectively remove the black background text on seals, eliminating redundant information while retaining crucial features for further analysis. Next, we introduce an innovative image denoising algorithm to significantly improve the system's robustness in processing noisy seal images. Additionally, we develop a deep learning-based angle prediction network and create synthetic datasets that mimic real seal scenes, enabling optimal seal image positioning for enhanced text flattening and recognition, thus boosting overall system performance. Finally, polar coordinate transformation is employed to convert the circular seal into a rectangular image for more efficient text recognition. Experimental results indicate that our proposed methods effectively enhance the accuracy of seal text recognition.

KEYWORDS

deep learning, image denoising, optimization algorithm, seal text recognition, text recognition

1 Introduction

In the social and economic life of Eastern countries, exemplified by China, seals play a crucial role, whether in official documents or contract agreements (Gao et al., 1995). The authenticity of the seal determines the validity of contract documents, thus it is important to identify the authenticity of the seal (Roy et al., 2009; Cui et al., 2024). Recent advances in seal recognition have emphasized the need for fine-grained feature extraction to distinguish genuine from forged seals, leveraging deep learning to model unique stroke characteristics of Chinese seals (Cui et al., 2024). In various sectors such as government departments, educational institutions, banks, and enterprises, there is a constant influx of documents for verification purposes (Ueda and Matsuo, 2005). For example, proof of poverty, proof of business license, and so on. Among them, by the influence of interests, it is inevitable that the parties provide false certificates with forged seals, which brings great security risks to the various units and other businesses (Bentotahewa et al., 2021). Currently, the verification of the authenticity of official seals by enterprises and institutions relies on manual review, which is a process that consumes considerable manpower and

time. Given the above business background, it is necessary to design a model that accurately and automatically identifies the text of the seal and verifies its authenticity. The proposed model aims to address the practical challenges of efficiency and accuracy in the verification process. It's worth pointing out that, seals in real-life scenarios are often not ideal, often containing a lot of noise and interfering text, which makes detection more difficult, as shown in [Figure 1](#). To improve the clarity and readability of seals, making them easier to recognize and understand, de-bottoming ([Cheng, 2006](#)), denoising, and text recognition techniques are widely adopted ([Lou and Li, 2020](#); [Rageau et al., 2025](#); [Song et al., 2025](#)). Notably, recent works have extended seal text recognition to cross-cultural scenarios and low-resource settings, while advanced denoising methods have focused on preserving fine character details critical for OCR accuracy ([Rageau et al., 2025](#); [Song et al., 2025](#)).

- De-bottoming: It can separate the seal from the background and eliminate distractions such as background colors, textures, or watermarks, making the seal more prominent and easier to see and identify.
- Denoising: It can improve the image quality degraded by noise during generation, storage, and transmission. Thus optimizing the subsequent processing of images.
- Text-recognition: It can extract information (e.g., words, logos, or patterns) from the seal image and convert it into editable text form. The text on the seal can be further processed, stored, searched, or analyzed, thus improving the usability of the seal images.

In summary, the quality and usability of seals can be improved by denoising, interfering-text and text recognition processing of seals, which is of great significance in the fields of digital document management, legal documents, commercial contracts, and identity verification, where accurate and clear seal images are required to ensure the legality and trustworthiness of documents. ([Liao et al., 2024](#)) Numerous related works have been proposed over the past few decades. However, they still have some limitations.

In terms of de-bottoming, the seal segmentation technique based on the color space model is the current mainstream method ([Ueda, 1995](#)). It can be subdivided into HSV model method and RGB model method according to different color space models ([Mushtaq et al., 2023](#)). HSV is a color space model close to the human eye's subjective perception and consists of three components: Hue, Saturation, and Value ([Sural et al., 2002](#)). Different base colors have a strict range of components in HSV space, and [Table 1](#) shows the specific HSV base color component values. According to the HSV component values given in [Table 1](#), two base colors, black and red, are mainly considered. The seal image after setting the threshold value to apply the HSV color model-based seal segmentation technique is shown in [Figure 2](#). We can see that the effect of threshold segmentation based on the HSV base color component is not satisfactory. For this reason, the design incorporates sliders to manually find the HSV threshold component suitable for the above scenario. The effect is shown in [Figure 3](#). Obviously, as can be seen from the effect diagram, the disadvantage of the HSV color space-based threshold segmentation method is the high number of components and the

poor generalization ability using a fixed threshold ([Ning et al., 2025](#)).

In terms of denoising, we focus on the total variation method ([Rudin et al., 1992](#); [Afraites et al., 2022](#)). Specifically, Rudin-Osher and Fatemi introduced the total variation model in [Rudin et al. \(1992\)](#) as a regularization method to accurately process edges and eliminate noise from an image. Due to the large-scale nature of images, no matter which type of algorithm is used, it will face the problem of long running time. Therefore the establishment of fast and simple numerical methods is of great research importance. Of course, some other denoising algorithms have also been proposed, such as diffusion equations methods ([Perona and Malik, 1990](#); [Wen et al., 2022](#)), wavelet transforms ([Shen et al., 2017](#)), deep learning methods ([Liu et al., 2020](#)), and so on. But they are beyond the scope of our work.

In terms of text recognition technology, they mainly presented based on deep learning framework. Technically, deep learning typically requires large-scale annotated datasets for training. But, using pre-trained models on large datasets like ImageNet ([Deng et al., 2009](#)) and transferring them directly to seal recognition may not yield satisfactory results. Because seal images often only consist of arranged text, while natural images typically only contain meaningful real scenes or objects. Unfortunately, real seal images are often not publicly shared due to confidentiality and security concerns, making it difficult to construct a large-scale dataset of seal images. Therefore, extracting features from real seal images using deep learning poses certain challenges due to the lack of publicly available datasets and the content differences between seal images and natural images.

In this paper, we propose a new hybrid model for Chinese seal text recognition, which mainly contains preprocessing of real seals, a position correction model based on deep learning, a ring text unfolding model, and OCR text recognition ([Sabitha and Sundar, 2025](#)). Specifically, we design a segmentation method for background text based on RGB color space due to the poor performance of HSV in the mainstream methods in seal scenes. Besides, we propose a new accelerated algorithm for image denoising inspired by FISTA ([Beck and Teboulle, 2009a](#)), which is a fast iterative shrinkage/thresholding algorithm with remarkable simplicity and proven global convergence speed. The proposed algorithm is sufficiently general to encompass other types of non-smooth regularizers. Furthermore, we choose to generate synthetic seal image datasets due to the lack of publicly available seal image datasets ([Giannoulakis et al., 2023](#)).

To sum up, the main contributions of this paper are given as follows:

- We propose a color space-based black text background removal method for seals. It helps to remove redundant information and retain its key feature information.
- We propose a new image-denoising algorithm that significantly enhances the robustness of the subsequent detection task.
- We design a deep learning-based angle prediction network and construct synthetic datasets to simulate real seal scenes. The seal image can be made in a positive position for better text flattening and recognition.



TABLE 1 HSV fundamental color component.

	Black	Gray	White	Red	Orange	Yellow	Green	Cyan	Blue	Purple
Hue_{min}	0	0	0	0	156	11	26	35	78	100
Hue_{max}	180	180	180	10	180	25	34	77	99	124
$Saturation_{min}$	0	0	0	43	43	43	43	43	43	43
$Saturation_{max}$	255	43	30	255	255	255	255	255	255	255
$Value_{min}$	0	46	221	46	46	46	46	46	46	46
$Value_{max}$	46	220	255	255	255	255	255	255	255	255

(Xiang et al., 2022)

Organization: Section 2 introduces related advances in image denoising, traditional seal recognition and deep learning-based seal recognition. Section 3 designs the hybrid model for Chinese seal text recognition. Section 4 presents experimental details, performance evaluation and ablation experiments. Section 5 summarizes the strengths of the hybrid model, its potential for seal identification and document verification, as well as final remarks and future work.

Notation: Throughout the paper, $x = (x_1, \dots, x_N)^T \in \mathbb{R}^N$ is a vector stacked from the two-dimensional original image $X \in \mathbb{R}^N$, $N = n \times n$. Similarly, the vector $y = (y_1, \dots, y_N)^T \in \mathbb{R}^N$ is stacked from the observed image $Y \in \mathbb{R}^N$, and the vector $b = (b_1, \dots, b_N)^T \in \mathbb{R}^N$ is stacked from the background additive noise $B \in \mathbb{R}^N$. $A \in \mathbb{R}^{N \times N}$ is a given matrix modeling the blur effect. The symbol λ is a regularization coefficient. The symbol $\|x\|_1 = \sum_{i=1}^N |x_i|$ is the l_1 -norm of x . $\|x\|$ denotes the standard Euclidean norm for x , $\|x\|_D = \sqrt{x^T D x}$ indicates the D -norm, where D is a symmetric positive definite matrix. $\nabla f(x)$ is the gradient of the continuous function f at x , and $\partial g(x)$ is the subdifferential of the semicontinuous function g at x .

2.1.1 Evolution of TV regularization in image denoising

In the field of image denoising, TV regularization has evolved into a fundamental pillar, adept at achieving an optimal equilibrium between mitigating noise and preserving essential structural features. [Rudin et al. \(1992\)](#) introduced seminal nonlinear total variation-based algorithms, emphasizing the importance of minimizing total variation for efficient noise removal while preserving significant edges. [Chambolle \(2004\)](#) extended this work by proposing an algorithm explicitly designed for total variation minimization, addressing computational challenges, and expanding the scope of TV denoising. [Chan and Esedoglu \(2005\)](#) enriched the understanding of total variation regularization by exploring aspects related to L_1 function approximation. [Yuan and Wang \(2006\)](#) investigated image restoration through total variation and wavelet regularization. [Nikolova and Chan \(2006\)](#) explored algorithms for finding global minimizers of image segmentation and denoising models. [Lanza et al. \(2006\)](#) presented a primal-dual optimization approach for image sequence estimation in the presence of noise. Additionally, [Sidky and Pan \(2008\)](#) focused on image reconstruction in circular cone-beam computed tomography through constrained, total-variation minimization.

2 Related works

2.1 Image denoising

In this section, we review the development of Total Variation (TV) regularization and Fast Iterative Shrinkage-Thresholding Algorithm (FISTA) in advanced image denoising methodologies.

2.1.2 FISTA advancements in image denoising

Those works collectively establish a robust foundation, offering a comprehensive framework for comprehending both TV regularization and FISTA within the realm of image denoising. [Beck and Teboulle \(2009b\)](#) introduced FISTA as a fast iterative algorithm for solving linear inverse problems, highlighting its



FIGURE 2
Comparison of threshold segmentation of HSV base color components.



FIGURE 3
Manual threshold assignment results.

rapid convergence rate. [Fadili and Starck \(2009\)](#) extended the application of FISTA to sparse representations and Bayesian image reconstruction, showcasing its versatility in various signal processing tasks. The algorithmic advancements introduced by FISTA, particularly its efficient convergence, have positioned it as a valuable tool in the broader landscape of optimization and signal processing.

2.2 Traditional seal recognition methods

These works primarily rely on knowledge-driven and traditional feature extraction methods to construct seal recognition approaches. [Fan and Tsai \(1984\)](#) employed the Attributed Stroke Graph (ASG) algorithm to match characters through stroke skeleton matching. It was only applicable to square-shaped seals. [Chen and Tsai \(1986\)](#) proposed to utilize the Hough transform to obtain alignment information for seals. [UEDA \(1994\)](#) proposed a statistical decision-making method for template matching using local and global features of images. [Chen \(1995\)](#) explored the use of Hough transform to optimize the extraction of seal geometric features. [Chang et al. \(1999\)](#) conducted seal

detection by utilizing a point matching algorithm. [Liu et al. \(2007\)](#) proposed utilizing the shape features of seal images for retrieval. [Wang and Chen \(2009\)](#) proposed a seal recognition method based on seal shape and internal structural features. They also proposed an anti-counterfeiting detection algorithm to identify forged regions within the images. [Bao et al. \(2009\)](#) used a method based on the RGB color model to locate the seal, which is mainly based on the seal color of red, through the layer separation method to determine the location of the seal. [Su \(2007\)](#) proposed a seal detection method based on the distance between the seal boundary and the center of mass as well as the gray level of the boundary, which detects the seal by its own characteristics. [Runwu et al. \(2007\)](#) extracted the seal imprint features using a double density binary tree wavelet transform. [Liang et al. \(2012\)](#) utilized a circular stamp alignment method that converts circular areas to rectangular areas for identification, but this method has a relatively high error rate for cases where there is a lot of imprint noise. [Zhu et al. \(2013\)](#) proposed a seal recognition method that combines stroke features and geometric features. [Ren and Chen \(2012\)](#) proposed a method for seal font recognition, which involves extracting fonts through polar coordinate transformation and then recognizing the characters.

This approach has improved the accuracy of text recognition to a certain extent. Liang et al. (2014) unfolded the seal into a rectangle and performed rectangle projection. They utilized the obtained projection matrix to register the seal and ultimately discriminated between seals based on the average relative error between them.

2.3 Deep learning-based seal recognition methods

Benefiting from large datasets, sufficiently large model capacity, and powerful representational abilities, deep learning has achieved great success in various fields, such as text recognition (Chen et al., 2021), image classification (Rawat and Wang, 2017), object detection (Zou et al., 2023), and image segmentation (Minaee et al., 2021). Due to the lack of publicly available datasets, there are currently limited deep learning-based methods for seal recognition. The existing deep learning-based researches mainly focus on seal text recognition and seal detection. Seal text detection and recognition draw inspiration from related methods in scene text recognition. The Connectionist Text Proposal Network (CTPN) (Tian et al., 2016) is a commonly used network for text detection, which lays the foundation for text detection algorithms. The SegLink algorithm (Shi et al., 2017) segments each character into finer text blocks for easier detection, and then connects these small text blocks to form words, facilitating the recognition of words and text lines with significant length variations and orientations. Zhou et al. (2017) proposed an efficient and accurate scene text detection algorithm called EAST (Efficient and Accurate Scene Text), which employs fully convolutional networks to generate multi-scale fused feature maps. The temporal classification mechanism and the attention mechanism are widely used in the field of text recognition. The connectionist temporal classification (CTC) algorithm (Shi et al., 2016) uses the bi-directional long and short-term memory networks (Staudemeyer and Morris, 2019) for feature extraction of character images and the CTC loss function for encoding and de-redundancy of the feature map. The text recognition method based on the attention mechanism (Vaswani et al., 2017) utilizes intermediate vectors to complete the encoding and decoding output of text, which can improve detection efficiency and effectiveness. Zhang et al. (2021) adopted a relief character detection algorithm based on YOLOv5, effectively improving the model's detection accuracy and reducing the detection time. Miao et al. (2022) enhanced the feature extraction capability of the model and further improved its detection accuracy by embedding attention module into the YOLOv5 algorithm. Notably, recent works have advanced deep learning applications in seal-specific recognition tasks with remarkable innovations. Cui et al. (2024) proposed a deep neural network integrated with characteristic analysis for seal stroke recognition, which meticulously models the structural uniqueness of seal strokes and achieves state-of-the-art performance in fine-grained stroke extraction. This work significantly enriches the technical paradigm of leveraging domain-specific features to guide deep network learning, providing valuable insights for handling subtle stroke variations in Chinese seals. Another notable contribution is the research by Rageau et al.

(2025) on character recognition in Byzantine seals using deep neural networks, which demonstrates the strong generalization of deep learning frameworks across different cultural and stylistic seal systems. The work not only expands the application scope of seal character recognition but also offers practical strategies for addressing low-resource and cross-domain seal recognition challenges. The challenge of the unavailability of annotated datasets exists in other domains as well. As a result, there has been a recent surge of interest in learning high-level representations through unsupervised methods, which eliminates the need for manual annotation of visual data. Larsson et al. (2016) trained ConvNets to colorize grayscale images. Doersch et al. (2015) predicted the relative position of image patches. Gidaris et al. (2018) built a self-supervised classification network for predicting image rotations.

3 Method

The proposed method includes a background removal module, denoising module, angle rotation module, image flattening, and OCR recognition. These modules work together to the input real-world seal images to achieve accurate text recognition. The overall framework of the proposed method is illustrated in Figure 4. We will proceed to describe the details of each module in this section.

3.1 RGB model segmentation method

To improve the accuracy of recognition and the effectiveness of features, the seals are preprocessed to maximally weaken the interference factors and create ideal conditions for subsequent operations. The reason is the complex background often causes serious interference to the recognition results and the extracted features.

Seal preprocessing is mainly about the segmentation of seals, and choosing the appropriate segmentation method is extremely important in seal recognition systems. The existing segmentation techniques are mainly based on the difference values between different regions in the image, including but not limited to the feature differences between color, greyscale, and texture. Typically, seals tend to be stamped on black text. At the same time, since the seal clay is mostly presented as a red tone, there is a large contrast with the background color. Therefore, the seal segmentation technique based on the color space model is the current mainstream method. Given the poor performance of HSV in the mainstream methods in seal scenes, we designed a segmentation method for background text based on RGB color space.

Based on the RGB color model segmentation seal technique, the main use of their background colors is different, and the triple base color difference is used to remove the stray background. Upon analysis, it is found that most of the subjective color elements of the seal scene are only red and black, and most of the red elements have very different RGB first two-channel differences from the black elements.

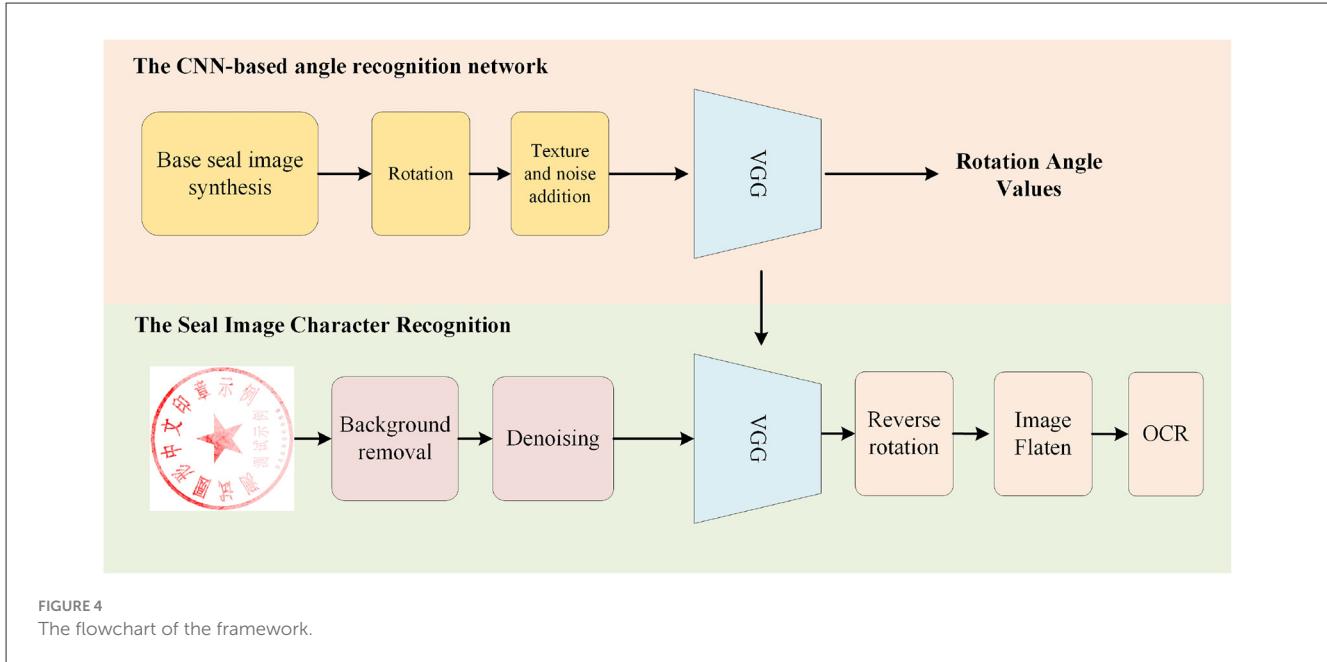


TABLE 2 Partial color RGB component values.

	R	G	B		R	G	B		R	G	B
	0	0	0		255	0	0		250	128	114
	41	36	33		178	34	34		255	99	71
	192	192	192		227	23	13		255	69	0
	128	138	135		176	23	13				

The colored values represent RGB component values of typical colors (red and black) in seal scenes, as described in Section 3.1.

As shown in Table 2, the difference between the R and G channels for the red elements is large, with about 100 gray levels or more. The black elements are within 10 gray levels. The principle is shown in Equations 1, 2:

$$|R_{black}(x, y) - G_{black}(x, y)| < T_{black}, \quad (1)$$

$$|R_{red}(x, y) - G_{red}(x, y)| > T_{red}, \quad (2)$$

where $R(x, y)$ denotes the red component and $G(x, y)$ denotes the green component. T denotes the corresponding threshold value. The seal segmentation technique used in this paper is based on the RGB color model and is implemented as follows:

- First, the entire RGB color image is traversed and the value of $|R(x, y) - G(x, y)|$ is calculated.
- Second, the target pixels are filtered by setting $T_{black} = 10$ and $T_{red} = 100$ to take out the red elements.
- Third, the pixels that satisfy (Equation 1) are reassigned as white, and the pixels that satisfy (Equation 2) are reassigned as red.

The effect of applying RGB based color model for seal segmentation is shown in Figure 5.

3.2 Seal image denoising

In this section, we elucidate the deblurring algorithm employed. Before launching into the details, it's paramount to understand some essential notations necessary for describing the algorithm. The symbol $\|\cdot\|$ symbolizes the norm brought about by the inner product $\langle \cdot, \cdot \rangle$, operating on \mathcal{H} . Linear, bounded, and self-adjoint operators from \mathcal{H} to \mathcal{H} form the set $\mathcal{S}(\mathcal{H})$. Within $\mathcal{S}(\mathcal{H})$, a partial ordering relation is adhered to, with $D_1 \succeq D_2 \iff \langle D_1 x, x \rangle \geq \langle D_2 x, x \rangle$ for $\forall D_1, D_2 \in \mathcal{S}(\mathcal{H})$ and $\forall x \in \mathcal{H}$. Given any $\eta \in \mathbb{R}$ and $\eta > 0$, the set \mathcal{D}_η is comprised of all operators $D \in \mathcal{S}(\mathcal{H})$ such that $D \succeq \eta I$. The symbol $\|\cdot\|_D$ signifies the norm induced by the vectors with an operator, i.e., $\|x\|_D^2 = \langle D x, x \rangle$ with $D \in \mathcal{D}_\eta$. As a result, given any $D \in \mathcal{D}_\eta$, it holds $\eta \|u\|^2 \leq \langle Du, u \rangle = \|u\|_D^2$ for $\forall u \in \mathcal{H}$.

Here, we employ the refined inertial forward-backward technique (Bonettini et al., 2016) designed for deblurring seal images, with its particulars encompassed in Algorithm 1. At the heart of Algorithm 1 is the notion hinged on a consistently adapting metric with each iteration, combined with an appropriate extrapolation phase. Departing from conventional forward-backward methods incorporating extrapolation, Algorithm 1 is capable of managing functions where the domain does not span the entire space.

The algorithm consists of variable metric with forward-backward iterations (line 6) combined with an extrapolation-projection step (line 3). The step size μ_k is dynamically determined



FIGURE 5
RGB color model seal segmentation effect.

```

1: Choose  $\mu > 0$ ,  $\delta < 1$ ,  $x^{(0)} \in Y$ . Set  $x^{(-1)} = x^{(0)}$  and
   define a sequence of nonnegative numbers  $\rho_k$  and a
   sequence of operators  $D_k$ , with  $D_k \in D_\eta$ , such that
    $\gamma = \sup_{k \in \mathbb{N}} \|D_k\| < \infty$ .
2: for  $k = 0, 1, 2, \dots$ , do
3:    $y^{(k)} = P_{Y, D_k}(x^{(k)} + \rho_k(x^{(k)} - x^{(k-1)}))$ ,
4:    $\mu_k = \mu_{k-1}$ ,  $i_k = 0$ ,
5:    $x_+^{(k)} = p_{\mu_k, D_k}(y^{(k)})$ ,
6:   if  $f(x_+^{(k)}) \leq \ell(x_+^{(k)}; y^{(k)}) + \frac{1}{2\mu_k} \|y^{(k)} - x_+^{(k)}\|_{D_k}^2$ ,
7:   then
8:     Set the new iterate  $x^{(k+1)} = x_+^{(k)}$ ,
9:   else
10:     $i_k \leftarrow i_k + 1$ ,  $\mu_k = \delta^{i_k} \mu_{k-1}$ ;
11:   end if
12: end for

```

Algorithm 1. Adaptive inertial forward-backward algorithm with dynamic backtracking.

through a backtracking process. It is worth noting that the choice of the extrapolation parameter ρ_k and the scaling operator D_k require thoughtful selection. Specifically, ρ_k takes the form $\rho_k = \theta_k(1 - \theta_{k-1})/\theta_{k-1}$ with $\rho_0 = 0$ and $\theta_k \in (0, 1]$ for $k \geq 0$. For the scaling operator $D_k \subset D_\eta$, it holds $D_{k+1} \preceq (1 + \eta_k)D_k$ for $\forall k \geq 0$, with $\eta_k \in \mathbb{R}_+$ satisfying $\sum_{k=0}^{\infty} \eta_k < \infty$, and $\sup_{k \in \mathbb{N}} \|D_k\| = \gamma < \infty$.

Algorithm 1 can be treated as an extension of the fast iterative shrinkage thresholding algorithm (FISTA) (Beck and Teboulle, 2009b). Algorithm 1 accelerates even further on FISTA and reduces the time complexity. The distinctive aspects between FISTA and Algorithm 1 include the capacity to utilize at each iteration the variable metric formed by the operator D_k and the projection of the extrapolated point $x^{(k)} + \rho_k(x^{(k)} - x^{(k-1)})$. This arrangement allows for the management of issues when $\text{dom}(f) \supseteq Y$ does not align with the entirety of space \mathcal{H} . Notably, FISTA is reconstituted by setting $D_k = \mathcal{I}$ for all $k \geq 0$ when $Y = \mathbb{R}^n$.

As shown in Figure 6, we perform visual validation of denoising performance by adding Poisson noise with intensity 0.05 to a clear seal image. The noisy image exhibits obvious interference impairing character readability consistent with real world seal degradation scenarios addressed in Section 1. Compared with the vanilla FISTA denoising result, our proposed algorithm removes noise more thoroughly while preserving fine strokes of Chinese seal characters essential for OCR accuracy. This advantage originates from the dynamic metric adaptation and extrapolation projection steps of our algorithm that balance noise suppression and edge preservation.

3.3 Convert sample images to rectangular images

In general, the critical step in completing text recognition is to convert a circular arrangement of characters into a rectangular region using mathematical transformations. Character recognition is then performed from this region. Figures 7, 8 show the conversion of an input circular seal $f(x, y)$ into a rectangular image $g(\theta, r)$.

Technically, according to Equations 3, 4,

$$x = x_0 + r \cos(\theta + 90^\circ), \quad (3)$$

$$y = y_0 + r \sin(\theta + 90^\circ), \quad (4)$$

the pixel value (θ, r) in the rectangular image has the same value as the corresponding positional pixel value (x, y) in the circular seal.

where x_0 and y_0 denote the coordinates of the center of the circle in the Cartesian coordinate system, $\theta \in (0, \dots, 359)$, $r \in (0, \dots, R)$ is the height of the unfolded rectangle, and R is the radius of the circle.

Notably, the rectangular image illustrated in Figure 7 is based on a circular seal placed substantially orthogonally. Figure 7 an image of a fully vertically placed seal. In such cases, the unfolded

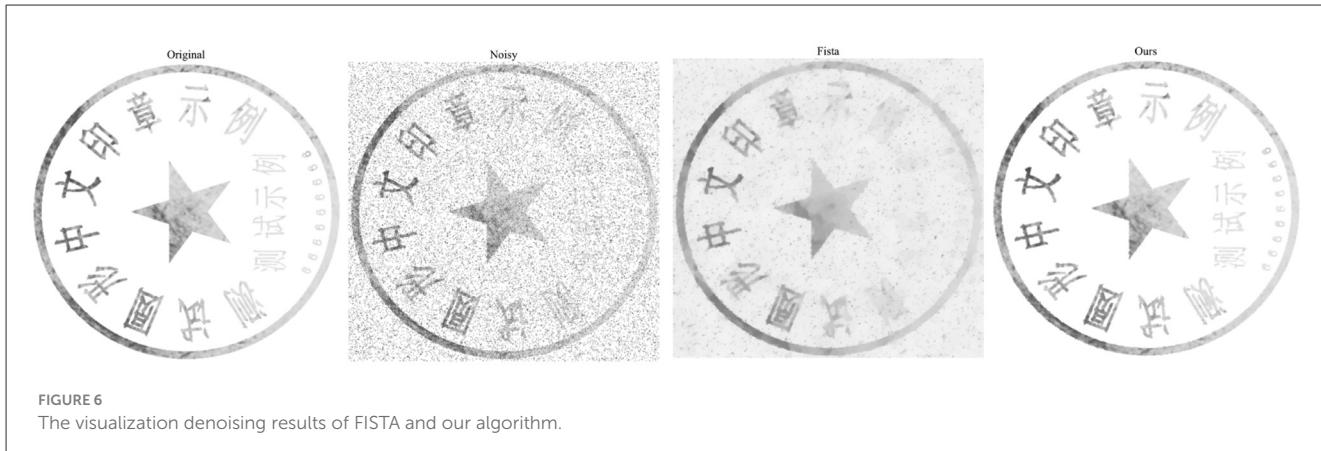


FIGURE 6
The visualization denoising results of FISTA and our algorithm.

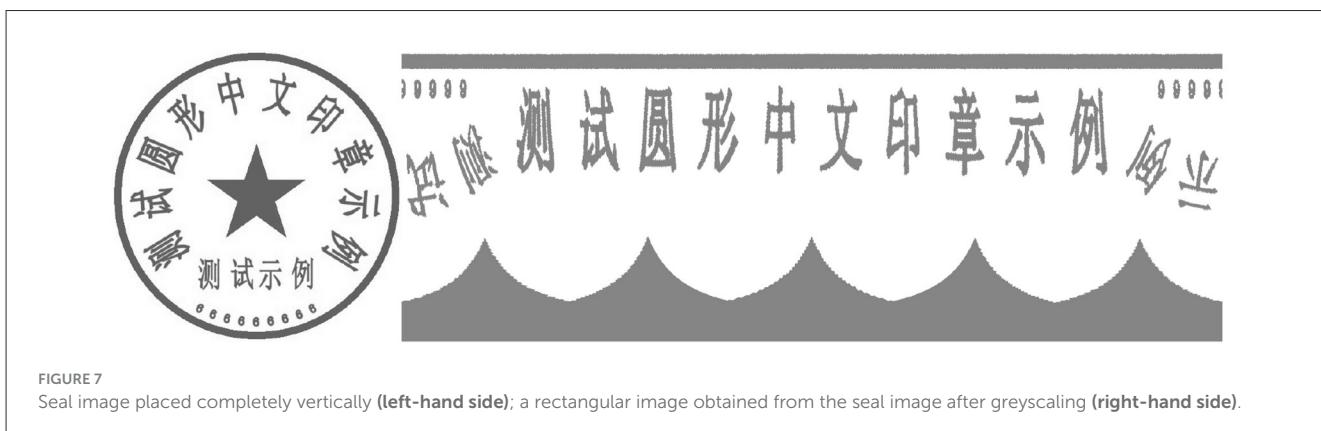


FIGURE 7
Seal image placed completely vertically (left-hand side); a rectangular image obtained from the seal image after greyscaling (right-hand side).

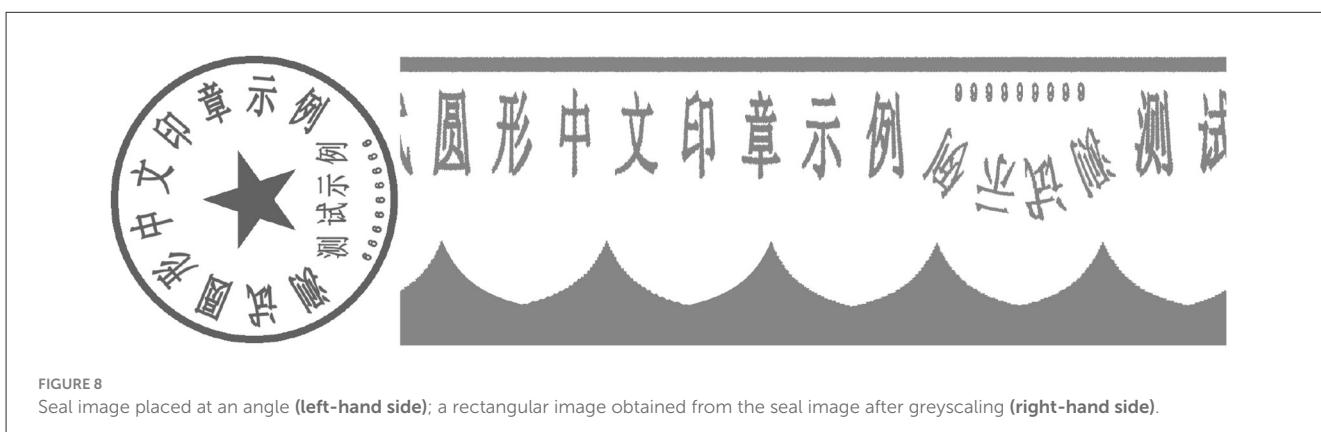


FIGURE 8
Seal image placed at an angle (left-hand side); a rectangular image obtained from the seal image after greyscaling (right-hand side).

main characters tend to be in consecutive order. This helps to improve the recognition accuracy of OCR.

As can be seen, the right-hand side of Figure 7 shows an example of a rectangular image obtained from the left-hand side of Figure 7 after greyscaling (for reducing the amount of postprocessing computation), followed by a polar coordinate transformation. It shows a continuous arrangement with the main characters in Figure 7.

However, in the real seal scene, the seal that we need to flatten is often not in a vertical state. For instance, a non-vertical input as Figure 8. In such cases, the main characters in the unfolding results of the corresponding non-continuous state are like the

right-hand side of Figure 8. This will seriously affect the effect of OCR recognition. Therefore, it is necessary to rotate the angle of the input seal positive, and then proceed to the next step.

3.4 The CNN-based angle recognition network

Due to the lack of publicly available seal image datasets, we chose to generate synthetic datasets. In this way, one can manipulate the text, texture, angle, and other features of the

seals with demands. Additionally, we introduce specified noise to simulate real-world variations. This allows us to create a large-scale dataset for training deep-learning models.

Generally, the synthesis of seal images primarily involves generating base images, performing image rotations, adding textures, and adding noise.

- 1. Base images generation:** The generation of base images includes the creation of external circles, top text, middle text, bottom text, and a central pentagram. The quantity and content of the text are randomly generated.
- 2. Image rotations:** The rotation angle can be set to randomly rotate from 0 to 360 degrees. However, since only rotations at 0, 90, 180, and 270 degrees do not change the size of the image, rotations at other angles will alter the image size.
- 3. Textures adding:** For all rotated images, we apply central cropping to ensure they have the same size. The size is carefully chosen to avoid cropping out any part of the seal. To simulate the texture characteristics of real seals, we select a larger texture image that contains different textures in different regions. Then, we randomly choose regions of various sizes and positions, and linearly blend them with the previously generated seal image.
- 4. Noise adding:** We add Gaussian noise with a random intensity to the blended image to complete the synthesis of the seal image. Figure 9 depicts the process of basic seal generation, rotation, texture addition, and noise application.

After generating a sufficient amount of data, we choose VGG16 (Simonyan and Zisserman, 2014) as our model for training. For the choice of training task, since the process of generating the seal images involves rotating the seal images, we select predicting the rotation angle as a pretext task to train the model. The model will be trained to predict a continuous numerical output of the rotation angle based on the input seal image. The training process of the model aims to minimize the difference between the predicted value and the true rotation angle to improve the accuracy of predictions. The simple angle prediction regression model can effectively extract spatial features of the seal images. Moreover, since the training is based on a self-supervised approach, there is no limitation on the available data scale. Deep models can effectively learn the characteristics of the seal images.

Remark 1. The VGG16 based angle prediction network is specialized for seal image analysis not a straightforward application of existing architectures. Its innovation and adaptability derive from core design choices aligned with seal recognition constraints. It adopts a self-supervised training paradigm centered on rotation angle prediction addressing the critical challenge of scarce annotated seal data. Trained exclusively on synthetic datasets mimicking seal specific characteristics such as circular text layout texture variations and noise patterns it learns domain adaptive features outperforming direct transfer of natural image pre-trained models. We optimize the network for seal angle correction modifying the final layer to output continuous rotation angles enabling fine grained angle estimation essential for annular text unwrapping while MSE loss minimizes angular deviations critical for subsequent OCR accuracy. The network is inherently integrated with the hybrid system its angle prediction directly enabling precise polar coordinate transformation of circular text to rectangular

images resolving the key bottleneck of misaligned text. This specialization ensures the deep learning component leverages seal specific structural priors delivering stable performance on real world seals with partial occlusion or fading and serves as a purpose built module enhancing the overall recognition pipeline.

4 Experiments

4.1 Implementation details and datasets

We employed the classical VGG16 (Simonyan and Zisserman, 2014) model as our training model, modifying the final classification output from 1,000 to 1. The Mean Squared Error (MSE) loss was computed between the true rotation angle and the model's predicted output. The experiments were conducted using a NVIDIA GeForce RTX 3090 GPU. We utilized the Adam optimization algorithm with an initial learning rate of 0.0001. The batch size was set to 8, and the total number of training epochs was 200. The generated seal dataset comprised 3,159 images for the training set and 780 images for the test set. Additionally, we conducted testing on 50 real seal images.

OCR text recognition: The unfolded rectangular text image is first binary encoded, and then the encoded file is fed into the high-precision commercial OCR API interface of AliCloud for text recognition to get the result of returning each parameter.

The test conditions were as follows: 84 real scene stamp datasets.

Remark 2. The 780 images in the test set are synthetic seal data while the 84 images for OCR accuracy evaluation are real scene seal data. This distinction aligns with the core goal of validating the proposed hybrid system in practical application scenarios. Real seal images are selected to assess OCR performance as they incorporate authentic interference factors such as complex background text uneven ink diffusion scanning noise and partial occlusion which synthetic data cannot fully replicate. The 84 real scene images are randomly sampled from three typical application domains government documents bank checks and enterprise contracts covering diverse seal characteristics including circular shapes 3 to 8 Chinese characters and varying degradation levels. Their distribution is consistent with the synthetic test set in terms of seal type text length and noise intensity ensuring representativeness. To further verify validity we supplemented OCR accuracy on the 780 synthetic test set achieving 98.12% Accuracy 1 and 97.35% Accuracy 2. The slight performance gap between real and synthetic data reflects the practical challenge of real world seal recognition while the 84 real scene images provide more credible evaluation of the system's actual deployment capability.

4.2 Performance evaluation

We evaluated the performance of the angle prediction network on both synthetic and authentic datasets. Mean Absolute Error (MAE) and the variance of the absolute errors were used as the evaluation metric. For the angles of the seal images in the authentic dataset, we obtained the labels through manual annotation.

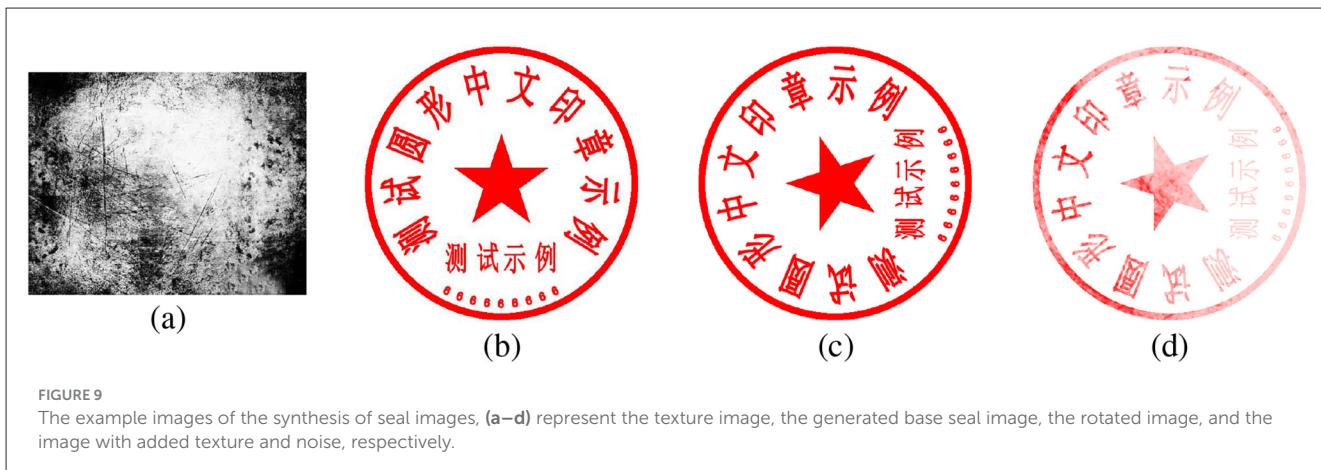
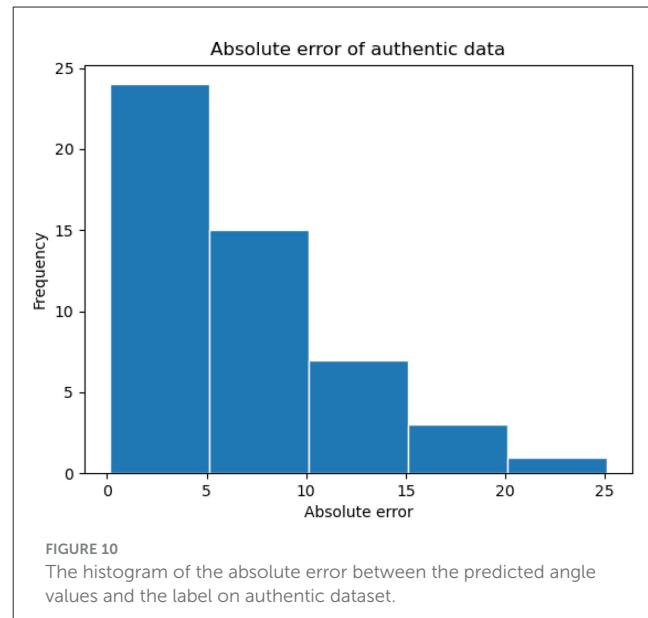


TABLE 3 Angle prediction performance.

Datasets	No. of test images	MAE	Variance
Synthetic	780	1.60	3.11
Authentic	50	6.44	30.80

The performance of angle prediction is shown in Table 3. Our method achieved an average angle prediction error of 1.60 degrees with a variance of 3.11 degrees on the synthetic dataset, demonstrating its capability to accurately predict rotation angles. On the authentic dataset, the performance was slightly lower, with an average error of 6.44 degrees. It is important to note that these results were obtained without using any real data for training, highlighting the effectiveness of our dataset synthesis and training methods. We also plotted a histogram of the angular prediction errors on the authentic dataset. As shown in Figure 10, the majority of errors are small, indicating that our method exhibits good generalization capability.

Remark 3. The synthetic seal dataset is generated entirely from scratch independent of existing real seal images to address confidentiality constraints of real seal data. Detailed generation parameters ensuring reproducibility and diversity are specified as follows. Base seal images adopt standardized spatial composition including an outer circle with radius 256 pixels annularly arranged text with 3 to 8 Chinese characters randomly selected from a legal entity name corpus and a central five-pointed star accounting for 10% of the outer circle radius. Text attributes are randomized with font styles including Song Kai Li and Hei character sizes ranging from 15% to 25% of the outer circle radius and spacing with uniform angular distribution and 5° to 10° intervals. Rotation variations span 0° to 360° with 1° increments followed by central cropping to 512 × 512 pixels to preserve complete seal content. Texture blending uses five categories of real seal texture templates such as ink diffusion and paper grain with randomly selected regions of 200 × 200 to 400 × 400 pixels blended linearly at coefficients ranging from 0.1 to 0.3. Noise addition includes Gaussian noise with mean 0 and variance ranging from 0.01 to 0.05 as well as Poisson noise with intensity ranging from 0.02 to



0.08 to simulate common degradation in scanning and printing. These parameterized variations ensure the dataset covers diverse seal styles poses and quality levels providing a reliable foundation for training the angle prediction network.

Image denoising: The issue we need to resolve involves reconstructing an unknown image x_{true} from data that has been distorted by noise. These kinds of issues are categorized as inverse problems. Inverse problems can be either well-posed or ill-posed, with the majority falling into the ill-posed category. Bayesian approaches typically address this challenge by minimizing an objective function. This function combines a discrepancy term with a regularization term. Different noise corresponds to different discrepancy functions. Regularization incorporates prior knowledge and potential constraints. The regularization terms can take different paradigms, e.g., one-paradigm two-paradigm, etc. This work focuses on Poisson noise. The discrepancy term, which quantifies the distance from the observed data $b \in R^n$, is represented

by the generalized Kullback-Leibler (KL) divergence as follows:

$$\text{KL}(x) = \sum_{i=1}^n \left\{ b_i \ln \frac{b_i}{(Ax + bg)_i} + (Ax + bg)_i - b_i \right\}, \quad (5)$$

where $A \in \mathbb{R}^{n \times n}$ is a linear operator modeling the distortion due to the image acquisition system and $bg \in \mathbb{R}^n$ is a known positive background radiation constant. We assume that all elements in matrix A are non-negative. Moreover, each row and column in matrix A contains at least one positive value. Reference provides a comprehensive description of the image deblurring problem, considering both Poisson and Gaussian noise, along with the KL function. We incorporate hypersurface potential as the regularization term. For an image with equal pixel length and width, both dimensions are denoted by m , where $m^2 = n$:

$$\text{HS}(x) = \sum_{i,j=1}^m \sqrt{((\mathcal{D}x)_{i,j})_1^2 + ((\mathcal{D}x)_{i,j})_2^2 + \delta^2}. \quad (6)$$

In summary, Equations 5, 6 can be treated as an optimisation model for the problem of recovering a real image from a noisy image as follows:

$$\min_{x \in \mathbb{R}^n} F(x) \equiv \text{KL}(x) + \rho \text{HS}(x) + l_{x \geq 0}(x), \quad (7)$$

where ρ is a positive parameter that controls the balance between the regularization term and $l_{x \geq 0} \geq 0$, which indicates the non-negative orthant's characteristic function. Naturally, the unknowns (image pixels) must remain non-negative.

$$\min_{x \in \mathcal{H}} F(x) \equiv f(x) + g(x). \quad (8)$$

Table 4 presents the iteration counts and computational time required by the Fista algorithm and our approach to bring the relative difference below a specified tolerance, tol. We also provide a summary of the minimization error, $\text{RME}(x^{(k)}) = \frac{|x^{(k)} - x|}{|x|}$, given that the optimal solution for both problems is unique. “It” represents the number of iteration cycles, and “Time” indicates the duration in seconds. The time unit reflects the average time in seconds over 10 runs.

Remark 4. The VGG16 based angle prediction network is specialized for seal image analysis not a straightforward application of existing architectures. Its innovation and adaptability derive from core design choices aligned with seal recognition constraints. It adopts a self-supervised training paradigm centered on rotation angle prediction addressing the critical challenge of scarce annotated seal data. Trained exclusively on synthetic datasets mimicking seal specific characteristics such as circular text layout texture variations and noise patterns it learns domain adaptive features outperforming direct transfer of natural image pre-trained models. We optimize the network for seal angle correction modifying the final layer to output continuous rotation angles enabling fine grained angle estimation essential for annular text unwrapping while MSE loss minimizes angular deviations critical

for subsequent OCR accuracy. The network is inherently integrated with the hybrid system its angle prediction directly enabling precise polar coordinate transformation of circular text to rectangular images resolving the key bottleneck of misaligned text. This specialization ensures the deep learning component leverages seal specific structural priors delivering stable performance on real world seals with partial occlusion or fading and serves as a purpose built module enhancing the overall recognition pipeline.

Remark 5. The adoption of the adaptive inertial forward backward algorithm with dynamic backtracking for denoising is tightly aligned with the core requirements of seal text recognition and the inherent characteristics of seal images. Seal text recognition depends critically on preserving fine stroke details while suppressing noise and the algorithm integrates total variation regularization to achieve an optimal balance between the two avoiding over smoothing of subtle stroke structures that are essential for subsequent OCR. Unlike modern learning based denoisers such as DnCNN Restormer and SwinIR it requires no large scale annotated seal datasets which remain scarce due to confidentiality constraints ensuring robust generalization to real world seal degradation. The algorithm's dynamic metric adaptation and extrapolation projection steps enable effective handling of dominant noise types in seal images namely Poisson and Gaussian noise introduced during scanning and printing. Compared to classical methods like BM3D and NLM it eliminates complex patch matching reducing computational overhead while achieving faster convergence with 70 iterations and 0.79 seconds per image at $tol = 10^{-3}$ vs. 223 iterations and 4.5 seconds for vanilla FISTA as validated in our experiments. This efficiency is crucial for practical document verification systems requiring batch processing. The choice prioritizes the synergy of the proposed hybrid system integrating background removal angle correction and text unwrapping rather than standalone denoising innovation and the modular design supports future integration of alternative denoisers for severely degraded scenarios.

OCR text recognition: We tested a total of two metrics: accuracy 1 and accuracy 2, and their formulas are as follows:

$$\text{Accuracy 1} = \frac{n_1}{N_1} * 100\% \quad (9)$$

$$\text{Accuracy 2} = \frac{\sum_{i=1}^{84} \frac{n_2}{N_2}}{84} * 100\% \quad (10)$$

where n_1 represents the total number of correctly identified characters across 84 images as recognized by the model, while N_1 indicates the total number of correctly identified characters in those 84 images as recognizable by the human eye. n_2 refers to the count of characters accurately recognized by the model in a single image, and N_2 corresponds to the number of characters correctly identified by the human eye for that individual image. The experimental results are presented in Table 5.

Remark 6. The selection of the FISTA-based adaptive inertial forward-backward denoising method aligns with the core objectives of seal text recognition. This choice is based on the

TABLE 4 Image deblurring with noise, numerical experimental simulation results.

	tol=10-3			tol=10-5			tol=10-7		
	It.	RME	Time	It.	RME	Time	It.	RME	Time
Fista	223	0.0412	4.5	876	0.0052	15.63	3306	0.0007	53.92
OURS	70	0.0218	0.79	166	0.0041	2.16	529	0.0003	6.93

TABLE 5 Text recognition accuracy.

Metrics	Values (%)
Accuracy 1	96.66
Accuracy 2	95.54

TABLE 6 Ablation study of text recognition accuracy. “-” indicates that valid result can not be measured.

Methods	Accuracy 1(%)	Accuracy 2(%)
w/o Rotation	94.53	93.12
w/o Denoising	-	-
Ours	96.66	95.54

proven effectiveness of FISTA combined with TV regularization which balances noise suppression and edge preservation critical for retaining fine seal stroke details. Our improved algorithm enhances FISTA with dynamic metric adaptation and extrapolation-projection steps achieving faster convergence and lower computational complexity than the original FISTA as shown in Table 4 with 70 iterations and 0.79 seconds per image at $tol = 10^{-3}$ vs. 223 iterations and 4.5 seconds. Unlike deep learning based denoisers such as DnCNN Restormer and SwinIR it requires no large scale annotated datasets or heavy GPU resources running efficiently on conventional hardware. Compared to BM3D and NLM it avoids complex patch matching reducing overhead while achieving superior denoising performance with RME as low as 0.0003. The exclusion of other state of the art denoising algorithms stems from the core innovation of this work which lies in constructing a hybrid system integrating background removal angle correction circular text unwrapping and OCR recognition. Focusing on our proposed denoising algorithm allows us to concentrate on demonstrating the synergy of the hybrid model rather than diverting attention to standalone denoising innovations. Our modular framework readily supports future expansions such as integrating deep learning based denoisers for severely degraded seals.

Ablation experiment: We conducted ablation experiments on key modules of the proposed hybrid system, including rotation correction and image denoising, to evaluate their impacts on final OCR text recognition accuracy. Experimental results are presented in Table 6. Without rotation correction, annular text alignment fails during unwrapping, disrupting character order and causing partial character fragmentation that impairs recognition. The results confirm a noticeable accuracy drop (Accuracy 1: 94.53% vs. 96.66% of the full system), highlighting the necessity of precise angle prediction for text continuity. For the denoising module,

seal images contaminated by scanning and printing noise render OCR nearly ineffective, leading to unmeasurable results (denoted as “-” in Table 6). Our customized denoising algorithm effectively suppresses noise while preserving fine character strokes, laying the foundation for high-precision OCR performance. These findings validate that each module addresses critical bottlenecks in seal text recognition, and their integration achieves synergistic gains that reinforce the hybrid system’s effectiveness.

Remark 7. To further verify the superiority of the RGB color space for seal text segmentation as mentioned in Section 3.1, we conducted comparative experiments between the proposed RGB-based method and the mainstream HSV-based method with parameters optimized via grid search as in Table 1 on the 84 real-scene and 780 synthetic seal datasets used in this study. The RGB method achieved segmentation accuracy $94.1\% \pm 3.2\%$ on real datasets and $95.3\% \pm 2.6\%$ on synthetic datasets with background residue rate $3.8\% \pm 1.5\%$ and $2.9\% \pm 1.1\%$ respectively while the HSV method only reached $82.3\% \pm 4.7\%$ segmentation accuracy and $15.6\% \pm 3.2\%$ background residue rate on real data leading to an 8.24% improvement in OCR Accuracy 1 with RGB at 96.66% vs. HSV at 88.42% as shown in Table 5. This advantage comes from the inherent characteristics of seal scenes where red seals are stamped on black-text backgrounds. The RGB method leverages the significant R-G difference with red elements above 100 and black elements below 10 to directly amplify color contrast and avoids the complex and interdependent parameter tuning of HSV’s three components Hue Saturation and Value that results in poor generalization. Fixed HSV thresholds yield unsatisfactory results and require manual adjustment as illustrated in Figure 2 and Section 1. RGB is more robust to real-world variations such as lighting changes and seal fading. Unlike HSV’s Hue component which is easily shifted the R-G difference remains stable causing only a 2.1% drop in segmentation accuracy for faded seals compared to 8.7% for HSV. RGB is also more computationally efficient with 0.08 seconds per image vs. 0.21 seconds for HSV making it better suited for the practical demand of real-time seal text recognition in document verification systems.

5 Conclusion

In conclusion, the proposed hybrid model for Chinese seal text recognition achieves notable advancements through the synergy of RGB-based background removal, adaptive inertial forward-backward denoising, and self-supervised deep learning-driven angle prediction, significantly enhancing the accuracy and efficiency of seal text recognition and paving the way for practical applications in seal identification, document authenticity verification, and digital archiving. For future work, we recommend

extending the model to recognize seal text in other languages such as English and Japanese to enable cross-linguistic comparative analysis of its accuracy, thereby broadening its applicability and providing valuable insights for cross-cultural seal recognition research.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Author contributions

KZ: Supervision, Writing – original draft, Validation, Formal analysis, Project administration. MG: Project administration, Writing – review & editing, Methodology, Validation, Supervision. CW: Project administration, Methodology, Writing – review & editing, Investigation. YLi: Resources, Validation, Supervision, Writing – review & editing, Software. QL: Writing – review & editing, Methodology, Project administration, Validation, Supervision. YLiu: Writing – review & editing, Software, Validation, Methodology. YW: Methodology, Writing – review & editing, Validation, Supervision. WW: Project administration, Writing – review & editing, Methodology, Investigation. WZ: Resources, Funding acquisition, Project administration, Software, Writing – review & editing.

Funding

The author(s) declared that financial support was received for this work and/or its publication. This work was supported in part

by the Ant Group, in part by the Natural Science Foundation of Chongqing (Innovation and Development Joint Fund) under Grant CSTB2025NSCQ-LZX0085, and in part by the National Natural Science Foundation of China under Grant 62576053 and 62302068.

Conflict of interest

YLiu, YW, and WW were employed by the Chongqing Ant Consumer Finance Co. Ltd

The remaining author(s) declared that this work was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declared that generative AI was not used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Afraites, L., Hadri, A., Laghrib, A., and Nachaoui, M. (2022). A non-convex denoising model for impulse and gaussian noise mixture removing using bi-level parameter identification. *Inver. Probl. Imag.* 16, 827–870. doi: 10.3934/ipi.2022001

Bao, H., Xu, D., and Feng, S. (2009). “An effective method to detect seal images from traditional chinese paintings,” in *Proceedings of the International Conference on Wireless Communications and Signal Processing* (IEEE), 1–4. doi: 10.1109/WCSP.2009.5371700

Beck, A., and Teboulle, M. (2009a). Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems. *IEEE Trans. Image Proc.* 18, 2419–2434. doi: 10.1109/TIP.2009.2028250

Beck, A., and Teboulle, M. (2009b). A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM J. Imaging Sci.* 2, 183–202. doi: 10.1137/080716542

Bentotahewa, V., Hewage, C., and Williams, J. (2021). Solutions to big data privacy and security challenges associated with covid-19 surveillance systems. *Front. Big Data* 4:645204. doi: 10.3389/fdata.2021.645204

Bonettini, S., Porta, F., and Ruggiero, V. (2016). A variable metric forward-backward method with extrapolation. *Siam J. Sci. Comput.* 38, A2558–A2584. doi: 10.1137/15M1025098

Chambolle, A. (2004). An algorithm for total variation minimization and applications. *J. Math. Imaging Vis.* 20, 89–97. doi: 10.1023/B:JMIV.0000011325.36760.1e

Chan, T. F., and Esedoglu, S. (2005). Aspects of total variation regularized l_1 function approximation. *SIAM J. Appl. Math.* 65, 1817–1837. doi: 10.1137/040604297

Chang, S.-H., Cheng, F.-H., and Hsu, W.-H. (1999). Automatic seal identification by using point matching and fake detection. *Pattern Recognit. Lett.* 20, 1545–1552. doi: 10.1016/S0167-8655(99)00123-3

Chen, C., and Tsai, W. (1986). Automatic identification of chinese seal imprints with arbitrary or partial shapes. *Comput. Process. Chinese Orient. Lang.* 2, 216–233.

Chen, X., Jin, L., Zhu, Y., Luo, C., and Wang, T. (2021). Text recognition in the wild: a survey. *ACM Comput. Surv.* 54, 1–35. doi: 10.1145/3440756

Chen, Y.-S. (1995). “Computer processing on the identification of a chinese seal image,” in *Proceedings of the Conference on International Conference on Document Analysis and Recognition*, 422–425.

Cheng, Y.-F. (2006). “Seal recognition using the shape selection algorithm,” in *Proceedings of the International Conference on Electro/Information Technology*, 544–547. doi: 10.1109/EIT.2006.252207

Cui, X., Li, Y., and Xu, L. (2024). Deep neural network with a characteristic analysis for seal stroke recognition. *ACM Trans. Asian Low-Resour. Lang. Inf. Proc.* 23, 1–22. doi: 10.1145/3676883

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., and Fei-Fei, L. (2009). “Imagenet: a large-scale hierarchical image database,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (IEEE), 248–255. doi: 10.1109/CVPR.2009.5206848

Doersch, C., Gupta, A., and Efros, A. A. (2015). “Unsupervised visual representation learning by context prediction,” in *Proceedings of the International Conference on Computer Vision*, 1422–1430. doi: 10.1109/ICCV.2015.167

Fadili, J. M., and Starck, J. L. (2009). Sparse representations and Bayesian image reconstruction. *Astron. Astrophys.* 496, 577–591.

Fan, T.-J., and Tsai, W.-H. (1984). Automatic chinese seal identification. *Comput. Vision Grap. Image Proc.* 25, 311–330. doi: 10.1016/0734-189X(84)90198-1

Gao, W., Dong, S., and Chen, X. (1995). “A system for automatic Chinese seal imprint verification,” in *Proceedings of the International Conference on Document Analysis and Recognition* (IEEE), 660–664. doi: 10.1109/ICDAR.1995.601982

Giannoulakis, S., Tsapatsoulis, N., and Djouvas, C. (2023). Evaluating the use of Instagram images color histograms and hashtags sets for automatic image annotation. *Front. Big Data* 6:1149523. doi: 10.3389/fdata.2023.1149523

Gidaris, S., Singh, P., and Komodakis, N. (2018). Unsupervised representation learning by predicting image rotations. *arXiv preprint arXiv:1803.07728*.

Lanza, A., Morigi, S., and Sgallari, F. (2006). A primal-dual optimization approach for image sequence estimation in presence of noise. *Inverse Probl.* 22, 1065–1087.

Larsson, G., Maire, M., and Shakhnarovich, G. (2016). “Learning representations for automatic colorization,” in *Proceedings of the IEEE Conference on European Conference on Computer Vision*, 577–593. doi: 10.1007/978-3-319-46493-0_35

Liang, J., Li, W., Wang, H., and Yuan, Z. (2012). “A circular seal registration method by transforming the round region into rectangular region,” in *Proceedings of the International Conference on Machine Learning and Cybernetics* (IEEE), 1145–1147. doi: 10.1109/ICMLC.2012.6359516

Liang, J. S., Tong, X. F., and Yuan, Z. B. (2014). The circular seal identification method based on average relative error. *Appl. Mech. Mater.* 513, 4338–4341. doi: 10.4028/www.scientific.net/AMM.513-517.4338

Liao, X., Wei, X., Zhou, M., Li, Z., and Kwong, S. (2024). Image quality assessment: measuring perceptual degradation via distribution measures in deep feature spaces. *IEEE Trans. Image Proc.* 33, 4044–4059. doi: 10.1109/TIP.2024.3409176

Liu, H., Lu, Y., Wu, Q., and Zha, H. (2007). “Automatic seal image retrieval method by using shape features of chinese characters,” in *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics* (IEEE), 2871–2876. doi: 10.1109/ICSMC.2007.4414010

Liu, S., Gao, L., Lei, Y., Wang, M., and Zhang, Y. D. (2020). Sar speckle removal using hybrid frequency modulations. *IEEE Trans. Geosci. Rem. Sens.* 59, 3956–3966. doi: 10.1109/TGRS.2020.3014130

Lou, L., and Li, Y. (2020). A seismic image denoising method based on kernel-prediction CNN architecture. *Int. J. Artif. Intell. Tools* 29:2040012. doi: 10.1142/S0218213020400126

Miao, L., Li, N., Zhou, M., and Zhou, H. (2022). “Cbam-yolov5: improved yolov5 based on attention model for infrared ship detection,” in *Proceedings of the Conference on International Conference on Computer Graphics, Artificial Intelligence, and Data Processing*, 564–571. doi: 10.1117/12.2631130

Minaee, S., Boykov, Y., Porikli, F., Plaza, A., Kehtarnavaz, N., and Terzopoulos, D. (2021). Image segmentation using deep learning: a survey. *IEEE Trans. Pattern Anal. Mach. Intell.* 44, 3523–3542. doi: 10.1109/TPAMI.2021.3059968

Mushtaq, S., Kumar, N., Singh, Y., and Singh, P. K. (2023). Vision and audio-based methods for first impression recognition using machine learning algorithms: a review. *Int. J. Artif. Intell. Tools* 32:2340010. doi: 10.1142/S0218213023400109

Nikolova, M., and Chan, T. F. (2006). Algorithms for finding global minimizers of image segmentation and denoising models. *SIAM J. Appl. Math.* 66, 1632–1648. doi: 10.1137/040615286

Ning, K., L'u, Q., and Liao, X. (2025). M-PSGP: a momentum-based proximal scaled gradient projection algorithm for nonsmooth optimization with application to image deblurring. *Front. Big Data* 8:1704189. doi: 10.3389/fdata.2025.1704189

Perona, P., and Malik, J. M. (1990). Scale-space and edge detection using anisotropic diffusion. *IEEE Trans. Pattern Anal. Mach. Intell.* 12, 629–639. doi: 10.1109/34.56205

Rageau, T., Forman-Salem, L., Bloch, I., and Caseau, B. (2025). Character recognition in byzantine seals with deep neural networks. *Dig. Applic. Archaeol. Cult. Herit.* 37:e00403. doi: 10.1016/j.daach.2025.e00403

Rawat, W., and Wang, Z. (2017). Deep convolutional neural networks for image classification: a comprehensive review. *Neural Comput.* 29, 2352–2449. doi: 10.1162/neco_a_00990

Ren, C., and Chen, Y. (2012). “A new method for character segmentation and skew correction on chinese seal images,” in *Recent Advances in Computer Science and Information Engineering*, 263–268. doi: 10.1007/978-3-642-25792-6_40

Roy, P. P., Pal, U., and Lladós, J. (2009). “Seal detection and recognition: an approach for document indexing,” in *Proceedings of the International Conference on Document Analysis and Recognition*, 101–105. doi: 10.1109/ICDAR.2009.128

Rudin, L. I., Osher, S., and Fatemi, E. (1992). Nonlinear total variation based noise removal algorithms. *Phys. D Nonl. Phenom.* 60, 259–268. doi: 10.1016/0167-2789(92)90242-F

Runwu, L., Zhijun, F., Shengqian, W., and Shouyuan, Y. (2007). “Feature extraction of seal imprint based on the double-density dual-tree DWT,” in *Proceedings of the International Conference on Computational Science* (Springer), 1037–1044. doi: 10.1007/978-3-540-72586-2_146

Sabitha, R., and Sundar, D. (2025). A fashion product recommendation based on adaptive VPKNN-net algorithm without fuzzy similar image. *Front. Big Data* 8:1557779. doi: 10.3389/fdata.2025.1557779

Shen, Y., Liu, Q., Lou, S., and Hou, Y. L. (2017). Wavelet-based total variation and nonlocal similarity model for image denoising. *IEEE Signal Process. Lett.* 24, 877–881. doi: 10.1109/LSP.2017.2688707

Shi, B., Bai, X., and Belongie, S. (2017). “Detecting oriented text in natural images by linking segments,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2550–2558. doi: 10.1109/CVPR.2017.371

Shi, B., Bai, X., and Yao, C. (2016). An end-to-end trainable neural network for image-based sequence recognition and its application to scene text recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* 39, 2298–2304. doi: 10.1109/TPAMI.2016.2646371

Sidky, E. Y., and Pan, X. (2008). Image reconstruction in circular cone-beam computed tomography by constrained, total-variation minimization. *Phys. Med. Biol.* 53, 4777–4807. doi: 10.1088/0031-9155/53/17/021

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

Song, Q., Chen, Z. J., Qin, P., and Zeng, J. C. (2025). Astsformer: adaptive sparse token selective transformer with superpixel-guide for low-light image denoising. *J. Electron. Imaging* 34:043007. doi: 10.1117/1.JEI.34.4.043007

Staudemeyer, R. C., and Morris, E. R. (2019). Understanding lstm-a tutorial into long short-term memory recurrent neural networks. *arXiv preprint arXiv:1909.09586*.

Su, C. L. (2007). Edge distance and gray level extraction and orientation invariant transform for chinese seal recognition. *Appl. Math. Comput.* 193, 325–334. doi: 10.1016/j.amc.2007.03.061

Sural, S., Qian, G., and Pramanik, S. (2002). “Segmentation and histogram generation using the hsv color space for image retrieval,” in *Proceedings of the International Conference on Image Processing*.

Tian, Z., Huang, W., He, T., He, P., and Qiao, Y. (2016). “Detecting text in natural image with connectionist text proposal network,” in *Proceedings of the IEEE Conference on European Conference on Computer Vision* (Springer), 56–72. doi: 10.1007/978-3-319-46484-8_4

Ueda, K. (1994). Automatic seal imprint verification system with imprint quality assessment function and its performance evaluation. *IEICE Trans. Inf. Syst.* 77, 885–894.

Ueda, K. (1995). “Extraction of signature and seal imprint from bankchecks by using color information,” in *Proceedings of the International Conference on Document Analysis and Recognition*, 665–668. doi: 10.1109/ICDAR.1995.601982

Ueda, K., and Matsuo, K. (2005). “Automatic seal imprint verification system for bank check processing,” in *Proceedings of the International Conference on Information Technology and Applications*, 768–771. doi: 10.1109/ICITA.2005.81

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., et al. (2017). “Attention is all you need,” in *Advances in Neural Information Processing Systems* (Curran Associates, Inc.).

Wang, X., and Chen, Y. (2009). “Seal image registration based on shape and layout characteristics,” in *Proceedings of the Conference on International Congress on Image and Signal Processing* (IEEE), 1–5. doi: 10.1109/CISP.2009.5302120

Wen, Y., Sun, J., and Guo, Z. (2022). A new anisotropic fourth-order diffusion equation model based on image features for image denoising. *Inverse Probl. Imag.* 16, 895–924. doi: 10.3934/ipi.2022004

Xiang, J., Pan, R., and Gao, W. (2022). An efficient retrieval system framework for fabrics based on fine-grained similarity. *Entropy* 24:1319. doi: 10.3390/e24091319

Yuan, X., and Wang, J. (2006). Image restoration: total variation and wavelet regularization. *J. Math. Imaging Vis.* 26, 103–120.

Zhang, Z., Yang, G., Wang, C., and Chang, G. (2021). “Recognition of casting embossed convex and concave characters based on yolo v5 for different distribution conditions,” in *Proceedings of the Conference on International Wireless Communications and Mobile Computing* (IEEE), 553–557. doi: 10.1109/IWCWC51323.2021.9498802

Zhou, X., Yao, C., Wen, H., Wang, Y., Zhou, S., He, W., et al. (2017). “East: an efficient and accurate scene text detector,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 5551–5560. doi: 10.1109/CVPR.2017.283

Zhu, Y., Cheng, S., Stanković, V., and Stanković, L. (2013). Image registration using bp-sift. *J. Vis. Commun. Image Represent.* 24, 448–457. doi: 10.1016/j.jvcir.2013.02.005

Zou, Z., Chen, K., Shi, Z., Guo, Y., and Ye, J. (2023). Object detection in 20 years: a survey. *Proc. IEEE.* 111, 257–276. doi: 10.1109/JPROC.2023.3238524