

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

Probabilistic Hough Transform for Rectifying Industrial Nameplate Images: A Novel Strategy for Improved Text Detection and Precision in Difficult Environments.

Han Li ^{1,†}, Yan Ma ^{1,†} , Hong Bao ^{1,*} and Yuhao Zhang ²

¹ Beijing Key Laboratory of Information Service Engineering, Beijing Union University, Beijing 100101, China; baohong@buu.edu.cn
² China Mining Products Safety Approval and Certification Center, Beijing, 100013, China; 59209122@qq.com
* Correspondence: baohong@buu.edu.cn; Tel.:18810208075
† These authors contributed equally to this work.

Abstract: Industrial nameplates serve as a means of conveying critical information and parameters. In this work, we propose a novel approach for rectifying industrial nameplate pictures utilizing a probabilistic Hough transform. Our method effectively corrects for distortions and clipping, and features a collection of challenging nameplate pictures for analysis. To determine the corners of the nameplate, we employ a progressive probability Hough transform, which not only enhances detection accuracy but also possesses the ability to handle complex industrial scenarios. The results of our approach are clear and readable nameplate text, as demonstrated through experiments that show improved accuracy in model identification compared to other methods.

Keywords: industrial image processing; feature amplification; image transformation strategy; text detection; Probabilistic Hough Transform

1. Introduction

The effective operation of industrial systems relies heavily on the accurate collection of equipment information during maintenance. Industrial nameplates are a vital form of identification, displaying critical data and parameters. The use of optical character recognition (OCR) algorithms to extract information from nameplates has become a key technology in the field of intelligent industrial production, providing valuable support to operators. However, ageing, corrosion, rust, and other defects can render nameplates difficult to read using OCR algorithms, leading to errors in equipment identification. Therefore, the development of effective methods for safely, quickly, and efficiently capturing equipment information has become a prominent area of research in the fields of computer vision and industrial scenarios. The ability to accurately extract information from damaged nameplates has significant scientific and practical value, enabling efficient maintenance and ensuring the safe and reliable operation of industrial systems.

Artificial intelligence development is currently at the stage referred to as weak artificial intelligence[1,2]. Object detection, as a crucial component of OCR, leverages predefined data sets to locate objects or their key features for recognition. There are established examples of computer vision systems in complex conditions, such as the detection of traffic signs in complex scenarios is a significant challenge in the field of autonomous driving, presented by Liang et al. [3]. The author of the paper suggests a new information acquisition method that could open up new research possibilities for computer vision techniques in industry. The photographs of equipment nameplates may present limitations such as varied shooting conditions and a high degree of irrelevant information. To accurately identify the nameplate location, we created the Equipment Nameplate Dataset (MEND) and built a machine learning algorithm that is established upon the use of Probabilistic Hough Transform. Elimination of external regions of the image is performed by perspective

modification. Comparison experiments using DBNet [4] were conducted on a subset of nameplate images from the MEND set, yielding effective detection results with high computational efficiency.

The environmental conditions and image information quality can greatly impact the accuracy of text detection in industrial nameplates [5]. The accuracy of text recognition for industrial nameplates depends heavily on the image quality and the conditions in which it is taken. Any minor alterations to the image, like fuzziness or static, or any environmental changes, such as brightness, can significantly hamper the precision of the text recognition procedure[6].

There are a range of strategies that can be implemented to mitigate the negative consequences of text detection.

If one wants to prevent any undesirable results, they can modify the visual realism of the image. By taking a close look at the image and making the necessary changes to brightness, hue, or contrast, one can achieve the desired result[7]. Taking this approach will guarantee that the image appears as natural as it can and any undesirable results are avoided.

Filtering techniques can be employed to heighten the quality of digital images, thus diminishing the amount of image noise[7–9]. Various strategies like blurring, smoothing, and sharpening are utilized in these approaches to decrease the quantity of unnecessary noise in the picture. Methods of sifting can be utilized to maintain the vital aspects of an image while discarding undesirable components, such as static. However, filtering techniques can only process images with a certain texture pattern, and once the noise has changed due to the industrial production environment, the algorithm cannot adjust.

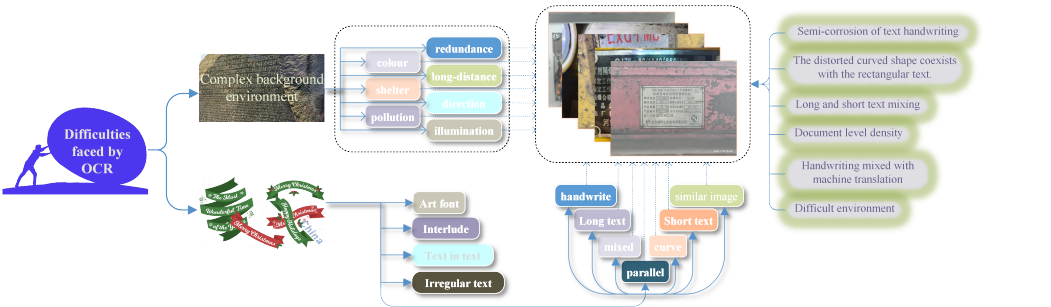


Figure 1. Challenges of OCR in industrial environments.Optical Character Recognition (OCR) faces several challenges in industrial settings. It includes the ability to recognize text in a variety of sizes, fonts, and orientations, as well as the capability to identify text in images that have low resolution and are affected by noise., such as those with multiple light sources or bright lighting. These difficulties must be overcome to enable accurate and reliable text recognition in industrial environments.

Frequency domain analysis and threshold segmentation are effective strategies for emphasizing key elements in images[5,9]. These techniques are employed to discern and separate entities or forms that are not readily perceptible in the original picture. The threshold segmentation technique uses a predetermined level to determine the edges of an item or shape in the image. Alternatively, frequency domain analysis techniques leverage frequency-based filters to recognize a certain trait in an image. By using threshold segmentation or frequency domain analysis techniques to standardize the image, important details can be made more visible and easier to detect.

While frequency domain analysis and threshold segmentation are useful techniques in image analysis, they also have certain limitations when applied to industrial nameplate recognition. One of the main drawbacks of threshold segmentation is that it requires a predetermined threshold level, which can vary depending on the specific image being analyzed. This can lead to inaccuracies in detecting and separating key elements in the image, particularly in cases where there is a lot of noise or distortion. Similarly, frequency domain analysis can be limited by its dependence on frequency-based filters. In certain scenarios,

these filters may not effectively capture the necessary features of the image, leading to inaccurate or incomplete recognition of the nameplate. As a result, these techniques may need to be supplemented with other methods in order to achieve accurate and reliable nameplate recognition in industrial scenarios.

Implementing the aforementioned techniques either alone or in tandem can lead to improved image data extraction. However, in industrial settings, due to their often intricate and unpredictable nature, nameplate recognition technology is somewhat limited in its capabilities[10].

Our research introduces a novel probability-based Hough transform method for corner detection in industrial nameplates. The main research objective is to improve the accuracy and speed of corner detection in complex scenarios, by reducing the negative sample impact. Furthermore, our method offers proper image pre-processing, which can contribute to the improvement of end-to-end text detection algorithms. To enhance the stability of detection models, we also provide challenging industrial nameplate datasets for training.

In the second section, we provide a literature review of the Hough transform and related research in the field of computer vision. This includes a discussion of its applications, strengths, and limitations.

The third section describes our proposed method, which involves corner recognition, projection transformation, and associated information derived from the progressive probability Hough transform.

The fourth section presents the experimental evaluation of our method, which includes a data collection process and a comparison of text detection results obtained from the experimental data. The evaluation demonstrates the effectiveness and superiority of our proposed method in terms of recognition accuracy and efficiency.

Finally, we summarize our findings and contributions in the conclusion section, highlighting the potential implications of our research in the field of computer vision.

2. Related Works

The research on image processing technologies amongst the scope of machine vision concentrates on identifying and processing information from equipment nameplates. It is widely acknowledged that the images of nameplates contain significant amounts of straight lines, with the frame of the nameplate being comprised of the four longest lines.

Research have explored the use of machine learning o detect the corners of industrial nameplates. For example, the RTLD proposed by Lai et al.[24] is capable of identifying and recognizing bold text lines in equipment nameplate images.However, some potential drawbacks that may arise in the use of this method include its accuracy and effectiveness in identifying and recognizing bold text lines across a variety of nameplate images. Chen et al .[11] created a correction module within a deep learning framework for handling text that is curved or oriented in multiple directions, as part of their development of the Natural Scene Recognizer. Li et al.[25] proposed a nameplate identification and recognition application using a seed approach to eliminate irrelevant factors. To enhance text identification, Panhwar et al.[14] included form detection of signboards. They gave priority to the similarity in shape between square signboards and industrial nameplates.. The issue of long-range micro detection was successfully resolved by WideSegNeXt[26].

Several image processing applications use the Hough transform and its variants for different tasks. For example, Zhao et al. [16] combined the Hough transform with neural networks to transform the challenge of finding the midpoint in parameter space into a semantic line recognition problem. Kagawa et al. [17] proposed a robust method for aging socket IC seat detection based on the Hough transform to ensure the correct placement of IC chips. Aslani et al. [18] proposed a new method of counting incomplete or clipped red blood cells through iterative Hough transform, which also involves image preprocessing. Additionally, Marzougui et al. [19] proposed a visual-based lane tracking method and Kumar et al. [20] proposed an efficient method for highway lane detection. Ahmad et al. [21] used probabilistic Hough transform for clustering and effectively corrected skewed

Table 1. Contributions and improvements from related works.

Article	Contribution	Shortcomings and future improvements
Chen et al.[11]	They put forward a model for multi-directional nameplate text recognition, which converts curved text into nearly horizontal text.	Unable to process extremely long text, lack of long text data set verification.
Ning et al.[12]	This study proposes a method of using the improved Faster-RCNN as a label positioning algorithm.	Typical nameplates have visible locations, so basic target detection algorithms are adequate and faster.
Zhang et al.[13]	The study proposed a similar edge detection method to locate and transform the position of nameplates.	CTPN is a powerful text detection algorithm, but now there are stronger text detection algorithms available.
Panhwar et al.[14]	This research suggests a framework for signboard text detection and recognition in natural environment.	The accuracy of model identification needs to be improved now, and the speed of experimental results is not improved.
Wu et al.[15]	The study proposes a region-based method for providing good coverage of the corners of plaques.	Despite the algorithm's impressive performance, its limited background fails to prove its applicability in complex industrial settings.
Zhao et al.[16]	Hough transform is incorporated into the deep learning framework, and a new semantic line detection method in natural scenes is proposed.	Hough transform is common in line segment and simple figure prediction, and this improved method can be used to predict other shapes in complex background.
Kagawa et al.[17]	A robust method for detecting aging sockets to ensure normal use of IC chips in sockets was proposed in the article.	More images need examining, applying this method to different types of diffuse reflection, and analyzing the properties of diffused reflected light in detail is necessary.
Aslani et al.[18]	A method for counting incomplete or fragmented red blood cells was proposed in the article, which was preceded by image preprocessing.	Despite saving researchers a lot of time, this method still has significant errors in medicine and requires more precise performance.
Marzougui et al.[19]	A visual-based lane tracking method was proposed in this study.	Errors in traffic scenarios can be deadly, making this method challenging to use in practical applications.
Kumar et al.[20]	An efficient method for detecting highway lanes was proposed in this study.	Method detects both curved and straight lane lines but needs enhanced accuracy for real road conditions.
Ahmad et al.[21]	The study used the probabilistic Hough transform for clustering, effectively correcting skewed documents.	Detecting densely packed text can be challenging, requiring additional processing. Although it may be effective in clean environments, it has not been demonstrated to be robust across varied backgrounds.
Ma et al.[22]	This paper suggests using deep compression learning to calculate edges in high-resolution images of deep-sea mining.	Data compression is inevitably accompanied by the loss of accuracy, so it is necessary to extract and retain key information modules.
Yang et al.[23]	A multi-feature fusion network is proposed and a new end-to-end 3D object detection framework is designed.	The accuracy of single-stage detection algorithm is often lower than other methods, and the detection speed of this method has no obvious advantage.

documents. Moreover, the Hough transform is generally used for processing planar lines or curves, but Kiyara et al. also applied this technique to 3D technology for better results. In nameplate recognition scenarios, however, the Hough transform is sensitive to image noise and these methods may encounter difficulties in detecting lines and symbols in images that are blurred or contain occlusions.

The field of deep learning is constantly striving to optimise both speed and accuracy. The integration of the Hough Transform algorithm into various applications is evidence of this, as seen in studies such as the one on the deep fuzzy hash network for image retrieval efficiency [22,23]. The method suggested in this research uses the Progressive Probabilistic Hough Transform for corner detection in industrial nameplates. The results of this approach show higher recognition accuracy and faster recognition time compared to traditional methods. Image pre-processing using this technique can improve the overall performance of end-to-end text recognition algorithms. The Hough transform has proven to be a valuable tool for detecting and identifying shapes in images, making it a widely used technique in machine learning applications.

However, these methods may not work well in industrial nameplates, as the negative samples in such images can significantly affect the accuracy of corner detection.

To address this issue, our study proposes a novel probability-based Hough transform method for corner detection in industrial nameplates. Our method reduces the impact of negative samples by using a modified scoring function, which improves the accuracy of corner detection in industrial nameplates. Our technique also allows for proper image pre-processing, which contributes to the improvement of end-to-end text detection algorithms.

Overall, while previous works have proposed effective methods for corner detection, they may not be suitable for industrial nameplates due to the presence of negative samples. Our proposed method addresses this issue and achieves high accuracy in corner detection.

3. The Proposed method

3.1. Hough transform and its variants

The Hough transform is a powerful instrument in computer vision and image processing for recognizing shapes in images. This approach is founded on the duality between points and lines in images. The Hough transform maps each point in image space to a corresponding point in a parametric space called Hough space. In Hough space, each line in the image can be represented as an individual point.

The Hough Transformation has been widely used in many applications, including lane detection, edge detection, and text recognition. It has proven to be a robust and effective technique, particularly when dealing with images containing noisy or cluttered data. The Hough Transformation has also been extended to recognize other shapes, such as circles and ellipses, by modifying the parameter space and the corresponding algorithms. The formula for a straight line is as follows:

$$y = k \times x + b$$
(1)

In the Cartesian coordinate space, the variable k is used to represent the slope of a line, while the variable b is used to represent the intercept of the line. For any line passing through a point (x_0, y_0) in the image space fulfill $y_0 = k * x_0 + b$. The point (x_0, y_0) produces a Hough space K and B straight line. It's written as $b = -k * x_0 + y_0$. The point (k_0, b_0) in Hough space can also be used to generate a line in the Cartesian coordinate system. The following is a representative expression:

$$b_0 = k_0 \times x + y$$
(2)

The Hough space represents all possible lines in the image space as a parameter space. In Hough space, each line is denoted by a point with two parameters, k and b . These parameters describe the line's slope and intercept of the line in the Cartesian space. A

mathematical expression can be used to describe the relationship between an image and Hough space:

$$(x_0, y_0) \rightarrow (k_0, b_0) = (-\frac{x_0}{y_0}, \frac{y_0}{k_0}) \tag{3}$$

where (x_0, y_0) is a point in the image space and (k_0, b_0) is the corresponding point in the Hough space. The inverse mapping, from Hough space back to image space, is given by the equation:

$$(k_0, b_0) \rightarrow (x_0, y_0) = \left(\frac{b_0}{k_0}, -\frac{b_0}{k_0^2 + 1} \right) \tag{4}$$

The Hough space is typically discretized into a grid of cells, each representing a range of k and b values. The value of each cell in the Hough space is set to the number of image points that are mapped to that cell. This is known as the Hough accumulator. The highest value in the accumulator represents the most likely line in the image. The line in the image space can be obtained by mapping back the corresponding point in the Hough space using the inverse transformation.

Utilizing the Hough transform, one can convert the task of detecting lines in an image area to finding corresponding points in the Hough space. Each line in image space corresponds to a specific k_0 and b_0 in Hough space, and finding these parameters in Hough space can be used to identify lines in image space. The Probabilistic Hough Transform is an advanced version of the Hough transform that can more accurately detect shapes and line segment lengths.

The basic idea of the Probabilistic Hough Transform is to randomly select a feature point in the image and assign it to a line in parameter space. The algorithm then iteratively examines additional feature points and updates the weights of the corresponding lines in parameter space. The lines with the highest weights are selected as the most likely line segments in the image.

Mathematically, the Progressive Probabilistic Hough Transform can be characterized as follows:

Initialisation: Given an picture with N pixels and a set of M lines in the parameter space, the algorithm initialises the line weights to zero.

Random selection: A feature point is randomly selected from the image and designated as a point on a particular line in the parameter space.

Weight Update: The algorithm examines the feature point and updates the weight of the corresponding line in parameter space. If the feature point is on the line, the weight is increased. If the feature point is not on the line, the weight is decreased.

Line selection: After iteratively examining all the feature points, the algorithm selects the lines with the highest weights as the most likely line segments in the image.

Line segment estimation: The algorithm estimates the endpoints of the line segments based on the selected lines in the space of parameters.

In this way, the Progressive Probabilistic Hough Transform provides a fast and accurate method for detecting and estimating line segments in images.

3.2. Two-sided detection utilizing probabilistic Hough transform

The Probabilistic Hough Transform is an advanced version of the Hough Transform that is employed to spot line segments in an image. Unlike the Hough transform which is mainly used for detecting straight lines, the Probabilistic Hough Transform is a more advanced algorithm that can identify both the start and end points of line segments. This allows for more accurate and efficient detection of complex shapes in images.

In this method, the two largest points in the Hough space are selected to represent the two longest line in the image. These line segments can then be used to determine the corner points of an object in the image, such as the four corners of a nameplate (as

shown in Figure 2(a)). The Progressive Probabilistic Hough Transform thus enables a more sophisticated form of shape detection and recognition in image processing.

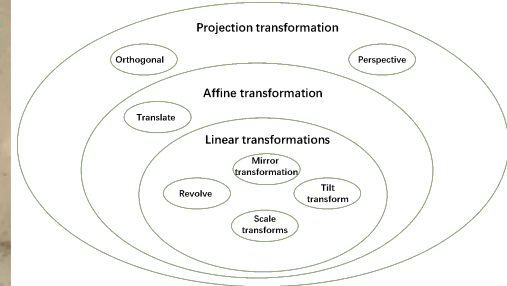
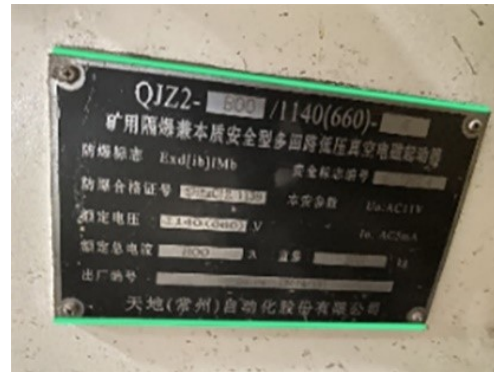


Figure 2. Process for correcting the uneven shape of industrial nameplates using the Probabilistic Hough Transform. (a) Find the edge in the image with the greatest length. (b) The relationship between inclusion and inclusion between projection transformation and its subsets. The double line segment obtained from the transform is used to extract four corners of the nameplate, which are then projected onto a rectangular shape based on their coordinates.

The figure 2 illustrates that four points can be derived from two lines. By analyzing the coordinates of the four corners, it is possible to precisely determine the positions of the figure's upper-left, upper-right, lower-left, and lower-right corners.

3.3. Relevant Principles of Image Correction Based on Probabilistic Hough Transform

The projection of the four corner coordinates obtained from the Progressive Probabilistic Hough Transform is presented. The projection process involves mapping a graphical object from one plane to another using a projection matrix. Projection transformations can be divided into two categories: affine transformations and linear transformations. The relationship between the two subsets of projection transformations is shown in Figure 2(b).

Linear transformation includes various transformations, including image rotation, skew transformation, scaling, and their combinations. After linear transformation, the origin of the image coordinates remains unchanged, and straight lines retain their straightness and parallel lines their parallelism.

Affine transformation, also known as affine projection, is a combination of linear transformation and translation. In geometry, it refers to the process of transforming one vector space into another by applying a combination of translation and linear transformations.

An equation can be used to illustrate this concept.

Assume K and J are two vector spaces. Defined as follows:

$$\begin{cases} K = (x, y) \\ J = (x', y') \end{cases} \quad (5)$$

The transformation of the vector space from K to J is given by the following formula:

$$J = K \times \omega + b \quad (6)$$

The equation will be divided to produce:

$$\begin{cases} x' = \omega_{00} \times x + \omega_{01} \times y + b_0 \\ y' = \omega_{10} \times x + \omega_{11} \times y + b_1 \end{cases} \quad (7)$$

Matrix multiplication conversion:

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} \omega_{00} & \omega_{01} & b_0 \\ \omega_{10} & \omega_{11} & b_1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = M \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \quad (8)$$

The Probabilistic Hough Transform is used to locate the four corners of an industrial nameplate in the image. A parameter matrix "M" is used to implement geometric transformations such as translation, scaling, rotation and mirroring. M acts as an intermediary between two vector spaces that undergo an affine transformation. Affine transformation is a combination of translation and linear transformation where the vector space is transformed and translated from one space to another.

Our approach study applies a perspective transformation, also known as projection transformation, to turn the irregular square in the 2D image into a rectangular shape. Using the Probabilistic Hough Transform, the four vertices of the nameplate of the equipment are determined. The projection matrix then projects the irregular quadrilateral onto the chosen plane. The projection transformation is completed through three stages.

1. General perspective transformation formula:

$$\begin{bmatrix} u' \\ v' \\ \omega \end{bmatrix} = \begin{bmatrix} u \\ v \\ \omega \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ \omega_{21} & \omega_{22} & a_{23} \\ \omega_{31} & \omega_{32} & a_{33} \end{bmatrix} \quad (9)$$

The parameters u and v , which define $\omega = 1$ and $a_{33} = 1$ and the image coordinates (x, y) generated via projection transformation, are taken from the original image. When x and y are satisfied:

$$\begin{cases} x = \frac{x'}{\omega'} \\ y = \frac{y'}{\omega'} \end{cases} \quad (10)$$

2. The eight parameters of the perspective transformation matrix need four coordinates to match to the eight equations that must be resolved. Two particular types of perspective transformation are linear transformation and affine transformation. The converted forms of x_n and y_m satisfy:

$$\begin{cases} x_n = \frac{x'}{\omega'} = \frac{a_{11} \times u + a_{21} \times v + a_{31} \times 1}{a_{13} \times u + a_{23} \times v + 1} \\ y_m = \frac{y'}{\omega'} = \frac{a_{12} \times u + a_{22} \times v + a_{32} \times 1}{a_{13} \times u + a_{23} \times v + 1} \end{cases} \quad (11)$$



Figure 3. Text instances of nameplates and labels are depicted in this image, which represents a typical example from our industrial nameplate dataset. The instances of text are particularly challenging to predict due to their complexity, which is characterized by variations in color, shape, background, and dense with special distributions. These challenging aspects of the dataset require advanced text detection and recognition techniques to achieve accurate results.

3. The solution to the equations provides the necessary eight parameters for the perspective transformation. By multiplication with a projection transformation matrix, the 2D image is converted into a new planar image. It is assumed that the projection center, target point, image point and align in a straight line, and the surface that casts a shadow is rotated around the perspective axis by a specified amount. The target matrix is obtained by

preserving the original projection geometry while eliminating the original projection light beam.

4. Experiments & Analysis

4.1. Dataset

In this study, a dataset (Mining Equipment Nameplates Dataset, MEND) was created for detecting and capturing text on industrial equipment nameplates. The images in the dataset were collected from photographs of industrial equipment nameplates from various mining companies in China, which contain more than 700 photographs. The images are split into 600 training set items and 142 test set items, but the images in the dataset are inconsistent and present challenges such as excessive brightness or poor lighting, partial occlusion, distorted shapes, contamination, rotation, and uneven shapes. This makes the data set more challenging than others.

The purpose of MEND is to track the text content and recognise all text occurrences in the images. It serves as a greatful dataset for testing text recognition models and character recognition for industrial nameplates, offering opportunities for end-to-end research in text detection and recognition. Moreover, this dataset can be leveraged for further research in this domain.

Several high quality datasets were used for model pre-training. The use of large, publicly available datasets improves the accuracy and quality of deep learning algorithms. The synthetic dataset SynthText, which consists of 800,000 images generated by combining 8k background images, was used with the ResNet18 and ResNet50 pre-training models [27]. The bilingual MSRA-TD500 dataset [28] also used the ResNet18 and ResNet50 pre-training models and consisted of 200 test images and 300 training images annotated with line-level text instances in both English and Chinese. The ICDAR2015 dataset [29], with 1,000 training images and 500 test images each with a resolution of 720 x 1280, was annotated with word-level text instances and used the ResNet50-D pre-training model. Finally, the TOTALTEXT dataset [30], consisting of 300 test images and 1255 training images annotated with word-level text occurrences, was used with the Res-Net50-D pre-training model and contained text in various forms, such as horizontal, multidirectional and curvilinear.

4.2. Metrics

When evaluating text recognition models, accuracy, precision, recall and F-measure are commonly used metrics. To fully understand these metrics, it is important to examine the confusion matrix. The confusion matrix is a graphical illustration that shows the association between predicted and actual categories, and it is also referred to as an error matrix or contingency matrix.

The confusion matrix typically has columns representing the predicted categories and rows representing the actual categories. The sum of each column represents the total number of predictions made for each category, while the sum of each row represents the total number of occurrences of each actual category. The numbers in each cell indicate the count of actual data points that were correctly classified as belonging to the corresponding predicted category.

Table 2. Confusion matrix used to form evaluation metrics.

/		Actual class	
		Positive	Negative
Predicted class	Positive	True Positive(TP)	False Positive(FP)
	Negative	False Negative(FN)	True Negtive(TN)

Accuracy: the proportion of correctly predicted samples in all samples; nevertheless, this assessment indicator suffers greatly when the positive and negative samples are out of proportion.

$$Accuracy = \frac{TP + NP}{TP + TN + FP + FN} \quad (12)$$

Precision: the measure of how well the model identifies true positive cases out of all positive predictions.

$$Precision = \frac{TP}{TP + FP} \quad (13)$$

Recall: consider the ratio of positively anticipated samples to all positive samples.

$$Recall = \frac{TP}{TP + FN} \quad (14)$$

A statistic called F-measure is sometimes referred to as F-ScoreHmean. The weighted harmonic average of Precision and Recall is known as F-Measure. It is referred to as the F1 value when $\beta = 1$. A higher F1 score indicates improved performance of the experimental approach.

$$F_{measure} = \frac{(1 + \beta^2) \times Precision \times Recall}{(\beta^2 \times Precision + Recall)} \quad (15)$$

4.3. Perspective Transform Based on Probabilistic Hough Transform

The performance of text detection can be impaired by the slanted orientation and abundant surrounding spaces of typical equipment nameplate images. It is crucial to isolate text instances and improve accuracy by removing the cluttered text samples from distorted areas of the nameplate. This study employs a six-stage process to rectify the irregular quadrilateral shape of the nameplate in the image into a regular rectangle. The streamlined procedure is depicted in Fig.4.



Figure 4. (a) Pictures to be processed; (b) Locate the image position to be transformed by detecting the two longest edges of the industrial image; (c) All the text information features are enlarged as the final result of projection transformation.

This method requires a nameplate that is easy to extract and has clear characteristics. To minimize the time spent on the experiment, the image is first resized. The size of the picture is decreased by a factor of 10 using the Pillow.shape() method. In order to identify the shape of the edge of the grayscale image (see Fig.5(a)), one must locate the boundaries of the picture and convert the low resolution thumbnail into a grayscale format.

After the image is binarized, Gaussian smoothing is applied to enhance the capability of the Probabilistic Hough Transform to identify the longer line segments. This is achieved by connecting short line segments to form a longer straight line (see Fig.6). In this research, we employ the Probabilistic Hough Transform method to identify the double lines of the preprocessed image. By utilizing these line segments, it is possible to calculate the corner coordinates of the Cartesian coordinate space for the thumbnail image. Then, the desired corner point of the original image is obtained by multiplying the value of the corner coordinates by ten (see fig.5(b)). Once the corner coordinates of the nameplate have been identified, a perspective transformation matrix is applied to the region to convert it into a standardized rectangular shape(see fig.7(a)).



Figure 5. (a)Display of gray scale diagram. (b)The position of the original picture after enlargement and positioning. The figures describe the process of extracting the four coordinates by using the two longest line segments for edge detection. To reduce irrelevant features, only grayscale images are suitable for this process.

In this study, we have compared the efficacy of two distinct techniques, namely two-line detection and perspective transformation, in the context of nameplate image processing. To validate the performance of these approaches, we have employed the differential binarization model and conducted experiments on nameplate photos. The outcome of the text detection procedure are presented in Figure 7.

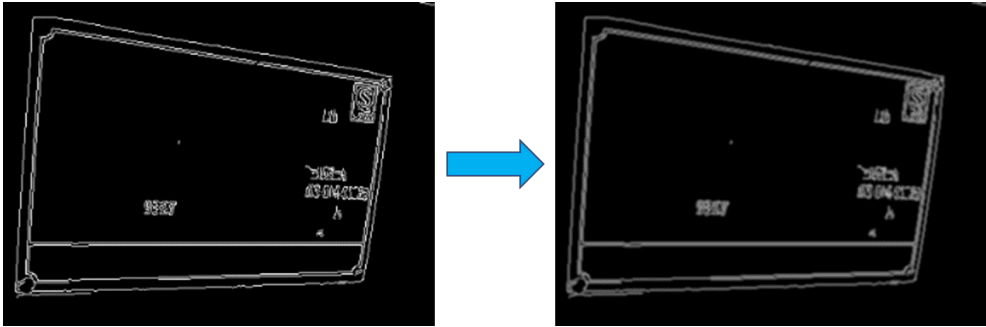


Figure 6. After edge detection, a Gaussian smoothing filter is applied to improve the accuracy of the resulting line segments. The left line segment appears straighter after the smoothing filter is applied, which helps to achieve more accurate results.



Figure 7. (a)Test results after correction progress using DBNet. (b)The pictures after corrected.

Specifically, our research focuses on the detection and correction of duplicate line segments in conventional images. We have utilized a ResNet-18 lightweight backbone in the DB model, which improves the accuracy and recognition performance and simplifies the binarization process of the segmented network. The outcome of our investigation, pertaining to the repair and recognition of double-line segments in conventional images, is demonstrated in Figure 7(b).

4.4. Comparative experiment

The effectiveness of conventional text detection techniques is limited to basic text distribution and lacks robustness, as the spacing between characters must not be too close for traditional text recognition. Two popular text detection algorithms, CTPN[31] and EAST[32], possess limitations, as CTPN can only detect horizontally dispersed text while the EAST method produces incorrect predictions for vertical text instances. Moreover, the unique visual characteristics of warped text remain unresolved by both approaches despite their widespread use. To ensure reliable text detection, we leverage DBNet and SAST[33], both of which exhibit high detection accuracy. Deep learning models often incorporate multiple underlying structures, and ResNet was proposed in 2015 to resolve the bottleneck problem associated with the network depth of classic deep learning networks. While it was generally believed that the effectiveness of a network increased with its depth, research showed that excessively deep networks could worsen the network effect by causing the gradient to vanish. He Kaiming’s ResNet[34,35] effectively addressed this issue. As a result, we have chosen to employ several iterations of ResNet as the underlying network structure for the detection model.

Table 3. The results of a comparative test using DBNet with ResNet18 as the backbone. The performance of models built from different scenarios is not directly transferable to complex industrial situations. Nevertheless, our training with new images has significantly improved model accuracy.

Dataset	Precision	Recall	F-measure	FPS
TD500	0.86	0.79	0.82	66
TD500 trained model in MEND	0.34	0.14	0.84	58
SynthText pre-trained in MEND	0.14	0.01	0.02	61
SynthText pre-trained, and trained in new pictures in MEND	0.84	0.81	0.82	52

Table 4. Comparison test: DBNet takes ResNet50 as the backbone. Using a deeper neural network as the backbone ResNet50 with slightly reduced prediction speed still achieved significantly better performance than ResNet18.

Dataset	Precision	Recall	F-measure	FPS
TD500	0.91	0.80	0.85	40
TD500 trained model in MEND	0.36	0.26	0.30	28
SynthText pre-trained in MEND	0.06	0.16	0.86	25
SynthText pre-trained, and trained in new pictures in MEND	0.86	0.84	0.85	36

Table 5. Comparative test results of the SAST approach using the ResNet50-D network as a backbone for text recognition in complex industrial scenarios. To optimize its performance, the approach is initially trained on the TOTALTEXT dataset and then fine-tuned on new images. The results show that the SAST approach achieves high accuracy while maintaining appropriate recognition speed.

Dataset	Precision	Recall	F-measure	FPS
ICDAR2015	0.91	0.83	0.87	12
TD500 pre-trained model in MEND	0.09	0.083	0.08	10
ICDAR2015 pre-trained, and trained in new pictures in MEND	0.72	0.45	0.55	11
TOTALTEXT	0.86	0.84	0.85	25
TOTALTEXT pre-trained model in MEND	0.17	0.08	0.10	25
TOTALTEXT pre-trained and trained in new pictures in MEND	0.91	0.86	0.89	25

Table 6. Comparative experiment: Faster RCNN and YOLOv3. As a comparison for text detection in complex environments, experiments were conducted using Faster-RCNN and Yolov3. First, training was performed on the extraction of features of industrial nameplates separately, and then fine-tuning was carried out on the generated data according to the strategy. The results showed that both methods achieved a slight improvement.

Dataset	Precision	Recall	F-measure	FPS
Faster-RCNN trained in MEND	0.71	0.73	0.72	14
Faster-RCNN trained in new pictures in MEND(ours)	0.85	0.75	0.80	14
Yolov3 trained model in MEND	0.72	0.45	0.55	25
Yolov3 trained in new pictures in MEND (ours)	0.80	0.82	0.81	25

ResNet18 has fast convergence time and high accuracy performance compared to other residual network models. A deeper model may yield better results due to its ability to extract more complex features by increasing the network depth. ResNet50, which has a more complex networked structure than ResNet18, offers improved accuracy without showing any noticeable degradation. For the differentiable binarization approach, we used both ResNet18 and ResNet50 as the basic networks. In addition, we used the enhanced 50-layer residual network ResNet50-D[35] as the basis for the SAST detector. This was done to speed up the subsequent training phase.

We conducted experiments on three different datasets, namely MSRA-TD500, ICDAR2015, and TOTALTEXT, with the aim of training the models immediately and collecting evaluation data. To increase the model’s resilience, a novel dataset was generated by utilizing the gathered data and the rectified training images. During the pre-training phase, three separate datasets and two different detectors were used to achieve better weighting and detection speed. We performed over 5,000 iterations on all pre-trained models, with about 200 images transformed into different perspectives for the new dataset.

We evaluated the performance of DBNet using ResNet18 and ResNet50 as backbones on the MSAR-TD500 and MEND datasets. As shown in the tabular data, it is not feasible to directly apply the models to complex industrial scenarios. As part of our evaluation of the SAST approach, we conducted a comparative experiment using the ResNet50-D enhanced residual network as the backbone for our system, with the same dataset to ensure a fair comparison. Specifically, we evaluated our method’s efficacy and the difficulty level of the dataset using the ICDAR2015 and TOTALTEXT datasets. The results of the experiment, including precision, recall, and F-measure, are presented in Table 4. Our pre-trained and re-trained models demonstrated excellent performance on these datasets. However, it should be noted that text identification algorithms trained on other datasets may not provide optimal results when used in complex industrial environments. This is because, in addition to the correctness of the character set, line-level text samples must also be collected to address the challenges presented by such settings.

The method proposed in this paper is a pre-processing technique for text detection, which does not affect the efficiency of the model’s use, but can enhance the inference accuracy. In practical applications, this pre-processing algorithm has a relatively small computational cost and high operating efficiency on the experimental host, with a speed of up to 20fps. The delay caused by this method during barcode scanning or data collection is negligible for practical applications. Our method has also effectively improved the detection accuracy in experiments with Faster RCNN and YOLOv36. However, since these two detection algorithms are general-purpose and not specifically optimized for text detection, their accuracy is lower than that of other algorithms

5. Conclusion

We propose a correction module to solve the challenge of unevenly shaped images of industrial equipment nameplates that reduce the accuracy of text identification. The

correction module is implemented using the Probabilistic Hough Transform and includes two tasks: perspective image correction and double long line segment recognition. Our method transforms the unusual shapes of nameplates into approximately rectangular shapes, benefiting both character recognition and text detection. It is a fast and simple process that outperforms deep learning approaches in terms of accuracy, speed, and practicality. The proposed method significantly increases the reliability of the detection model by introducing new data labeling training. It is highly suitable for industrial applications and could be integrated into full text recognition in the foreseeable future.

While the proposed correction module using Probabilistic Hough Transform shows promise in addressing the challenge of unevenly shaped industrial equipment nameplate images, there are still some potential limitations to consider. Firstly, the performance of the module may be affected by the quality of the input images. Images that are too blurry or contain significant noise may not be effectively transformed into rectangular shapes, thus reducing the accuracy of text identification. Secondly, the correction module may not be effective for nameplates with extremely irregular shapes or unusual orientations. In such cases, additional preprocessing steps or more sophisticated techniques may be required to achieve satisfactory results.

Funding: This research was funded by the National Natural Science Foundation of China (Nos. 61932012, 61802019, 61871039, 61906017, and 62006020), the Academic Research Projects of Beijing Union University (No. ZK80202003 and ZK10202202), the Beijing Municipal Education Commission Science and Technology Program (Nos. KM202111417007, KM201911417003, KM201911417009, and KM201911417001), the Beijing Union University Research and Innovation Projects for Postgraduates (No. YZ2020K001), and the premium Funding Project for Academic Human Resources Development in Beijing Union University under Grant BPHR2020DZ02.

References

1. Yu, X.; Ye, X.; Zhang, S. Floating pollutant image target extraction algorithm based on immune extremum region. *Digital Signal Processing* **2022**, *123*, 103442. <https://doi.org/https://doi.org/10.1016/j.dsp.2022.103442>.

2. Yu, X.; Tian, X. A fault detection algorithm for pipeline insulation layer based on immune neural network. *International Journal of Pressure Vessels and Piping* **2022**, *196*, 104611. <https://doi.org/https://doi.org/10.1016/j.ijpvp.2022.104611>.

3. Liang, T.; Bao, H.; Pan, W.; Pan, F. Traffic sign detection via improved sparse R-CNN for autonomous vehicles. *Journal of Advanced Transportation* **2022**, *2022*, 1–16.

4. Liao, M.; Wan, Z.; Yao, C.; Chen, K.; Bai, X. Real-time scene text detection with differentiable binarization. In Proceedings of the Proceedings of the AAAI conference on artificial intelligence, 2020, Vol. 34, pp. 11474–11481.

5. Sabu, A.M.; Das, A.S. A Survey on various Optical Character Recognition Techniques. In Proceedings of the 2018 conference on emerging devices and smart systems (ICEDSS). IEEE, 2018, pp. 152–155.

6. Wu, W.; Xing, J.; Yang, C.; Wang, Y.; Zhou, H. Texts as Lines: Text Detection with Weak Supervision. *Mathematical Problems in Engineering* **2020**, *2020*, 1–12.

7. Li, J.; Huang, T.; Yang, Y.; Xu, Q. Detection and Recognition of Characters on the Surface of Metal Workpieces with Complex Background. In Proceedings of the 2020 IEEE 4th Information Technology, Networking, Electronic and Automation Control Conference (ITNEC), 2020.

8. Yuan, J.; Guo, M.; Huang, B.; Hu, R.; Dian, S. Processing and Recognition of Characters Image in Complex Environment. In Proceedings of the 2022 International Conference on Innovations and Development of Information Technologies and Robotics (IDITR), 2022, pp. 100–104.

9. Khan, T.; Sarkar, R.; Mollah, A.F. Deep learning approaches to scene text detection: a comprehensive review. *Artificial Intelligence Review* **2021**, *54*, 3239–3298.

10. Long, S.; He, X.; Yao, C. Scene text detection and recognition: The deep learning era. *International Journal of Computer Vision* **2021**, *129*, 161–184.

11. Chen, X.; Zhang, Z.; Qiao, Y.; Lai, J.; Jiang, J.; Zhang, Z.; Fu, B. Orientation robust scene text recognition in natural scene. In Proceedings of the 2019 IEEE International Conference on Robotics and Biomimetics (ROBIO). IEEE, 2019, pp. 901–906.

12. Baifeng, N.; Ganzi, H.; Yu, Y. Research on nameplate image recognition algorithm based on R-CNN and SSD deep learning detection methods. In Proceedings of the 2022 IEEE 2nd International Conference on Power, Electronics and Computer Applications (ICPECA). IEEE, 2022, pp. 580–584.

13. Shuliang, Z.; Xin, H.; Haochen, Z.; Jianyu, W. Design of text position detection method for electrical equipment nameplate. In Proceedings of the 2021 International Conference on Cyber-Physical Social Intelligence (ICCSI). IEEE, 2021, pp. 1–4.

14. Panhwar, M.A.; Memon, K.A.; Abro, A.; Zhongliang, D.; Khuhro, S.A.; Memon, S. Signboard detection and text recognition using artificial neural networks. In Proceedings of the 2019 IEEE 9th international conference on electronics information and emergency communication (ICEIEC). IEEE, 2019, pp. 16–19.

15. Wu, Y.; Li, Z.; Wang, Y.; Huang, Z.; Zheng, Z. Application Research of Feature Extraction Method of Power Equipment Nameplate. *Computer Science and Application* **2019**, *09*, 2084–2097.

16. Zhao, K.; Han, Q.; Zhang, C.B.; Xu, J.; Cheng, M.M. Deep hough transform for semantic line detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **2021**, *44*, 4793–4806.

17. Kagawa, T.; Ikemoto, M.; Ohtake, S. A robust method of IC seating inspection in burn-in sockets using Hough transform. In Proceedings of the 2022 IEEE International Conference on Consumer Electronics-Taiwan. IEEE, 2022, pp. 1–2.

18. Aslani, A.A.; Zolfaghari, M.; Sajedi, H. Automatic Counting Red Blood Cells in the Microscopic Images by EndPoints Method and Circular Hough Transform. In Proceedings of the 2022 16th International Conference on Ubiquitous Information Management and Communication (IMCOM). IEEE, 2022, pp. 1–5.

19. Marzougui, M.; Alasiry, A.; Kortli, Y.; Baili, J. A lane tracking method based on progressive probabilistic Hough transform. *IEEE access* **2020**, *8*, 84893–84905.

20. Kumar, S.; Jailia, M.; Varshney, S. An efficient approach for highway lane detection based on the Hough transform and Kalman filter. *Innovative infrastructure solutions* **2022**, *7*, 290.

21. Ahmad, R.; Naz, S.; Razzak, I. Efficient skew detection and correction in scanned document images through clustering of probabilistic hough transforms. *Pattern Recognition Letters* **2021**, *152*, 93–99.

22. Ma, C.; Li, X.; Li, Y.; Tian, X.; Wang, Y.; Kim, H.; Serikawa, S. Visual information processing for deep-sea visual monitoring system. *Cognitive Robotics* **2021**, *1*, 3–11.

23. Yang, S.; Lu, H.; Li, J. Multifeature fusion-based object detection for intelligent transportation systems. *IEEE Transactions on Intelligent Transportation Systems* **2022**.

24. Lai, J.; Guo, L.; Qiao, Y.; Chen, X.; Zhang, Z.; Liu, C.; Li, Y.; Fu, B. Robust text line detection in equipment nameplate images. In Proceedings of the 2019 IEEE International Conference on Robotics and Biomimetics (ROBIO). IEEE, 2019, pp. 889–894.

25. Li, J.; Zhang, W.; Han, R. Application of machine vision in defects inspection and character recognition of nameplate surface. In Proceedings of the 2014 13th International Symposium on Distributed Computing and Applications to Business, Engineering and Science. IEEE, 2014, pp. 295–298.

26. Nakayama, Y.; Lu, H.; Li, Y.; Kamiya, T. WideSegNeXt: semantic image segmentation using wide residual network and NeXt dilated unit. *IEEE Sensors Journal* **2020**, *21*, 11427–11434.

27. Gupta, A.; Vedaldi, A.; Zisserman, A. Synthetic data for text localisation in natural images. In Proceedings of the Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 2315–2324.

28. Yao, C.; Bai, X.; Liu, W.; Ma, Y.; Tu, Z. Detecting texts of arbitrary orientations in natural images. In Proceedings of the 2012 IEEE conference on computer vision and pattern recognition. IEEE, 2012, pp. 1083–1090.

29. Karatzas, D.; Gomez-Bigorda, L.; Nicolaou, A.; Ghosh, S.; Bagdanov, A.; Iwamura, M.; Matas, J.; Neumann, L.; Chandrasekhar, V.R.; Lu, S.; et al. ICDAR 2015 competition on robust reading. In Proceedings of the 2015 13th international conference on document analysis and recognition (ICDAR). IEEE, 2015, pp. 1156–1160.

30. Ch'ng, C.K.; Chan, C.S. Total-text: A comprehensive dataset for scene text detection and recognition. In Proceedings of the 2017 14th IAPR international conference on document analysis and recognition (ICDAR). IEEE, 2017, Vol. 1, pp. 935–942.

31. Tian, Z.; Huang, W.; He, T.; He, P.; Qiao, Y. Detecting text in natural image with connectionist text proposal network. In Proceedings of the Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part VIII 14. Springer, 2016, pp. 56–72.

32. Zhou, X.; Yao, C.; Wen, H.; Wang, Y.; Zhou, S.; He, W.; Liang, J. East: an efficient and accurate scene text detector. In Proceedings of the Proceedings of the IEEE conference on Computer Vision and Pattern Recognition, 2017, pp. 5551–5560.

33. Wang, P.; Zhang, C.; Qi, F.; Huang, Z.; En, M.; Han, J.; Liu, J.; Ding, E.; Shi, G. A single-shot arbitrarily-shaped text detector based on context attended multi-task learning. In Proceedings of the Proceedings of the 27th ACM international conference on multimedia, 2019, pp. 1277–1285.

34. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 770–778.

35. He, T.; Zhang, Z.; Zhang, H.; Zhang, Z.; Xie, J.; Li, M. Bag of tricks for image classification with convolutional neural networks. In Proceedings of the Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2019, pp. 558–567.