Accounting-Oriented Research on Note Recognition Model based on Information Extraction Algorithm

ZHIYIN LIU
Claro M. Recto Academy of Advanced Studies,
Lyceum of the Philippines University,
Manila,
PHILIPPINES

Abstract: - Enterprise accountants deal with bill reimbursement mostly relying on the traditional manual way to carry out, and the current bill recognition technology makes it difficult to meet the recognition needs of Chinese bills. And there is a lack of open-source Chinese bill recognition models in the training and validation process of the billing model. Aiming at the above challenges, the study proposes an information extraction algorithm based on the optical character recognition technique of deep learning, and the bill recognition model construction is carried out on this basis. Image detection is performed by utilizing detection and recognition neural networks, and image feature extraction is performed by combining convolutional recurrent neural networks with connectionist temporal classification. The validation shows that the accuracy of the research-proposed information extraction algorithm increases by an average of 9.86% compared with other algorithms in the self-constructed cab invoice dataset, and the F1 value in the International Conference on Integration and Innovation of Digital Archival Resources Toward the Enhancement of Public Service Capability 2015 dataset increases by 5.82% and 0.92% compared with other algorithms, respectively. Compared to other models, the study's proposed model increases the average number of frames per second by 34.47% and the average class-wide accuracy by 10.72% in the cab invoice dataset. The bill recognition model based on the information extraction algorithm proposed in the study can meet the bill recognition requirements, has superior recognition accuracy and efficiency, and has application value in enterprise bill recognition.

Key-Words: - Information extraction algorithm, Bill recognition, Deep learning, CTC, RCNN, DBNet.

Received: July 20, 2024. Revised: November 29, 2024. Accepted: December 12, 2024. Published: December 26, 2024.

1 Introduction

With the social and economic development and the continuous improvement of information technology, various types of bills are frequently and closely connected with people's lives, work, and study. Bills, as transaction vouchers, play an indispensable role in enterprise management, [1]. However, the traditional way of relying on accountants to manually check the information of bills with the information of system reimbursement forms is time-consuming, labor-intensive, and inefficient, [2]. The emergence of deep learning technology promotes the development of Optical Character Recognition (OCR), and also provides new ideas for the intelligence of bill text recognition, [3]. Based on OCR technology and deep learning, the use of text detection and text recognition to locate images and extract key information can realize the automation of text information processing and optimize the text-based business processes at work, [4]. However, compared with the general natural scene text recognition task, the text target of the ticket is relatively small and dense, which requires higher recognition accuracy of key information and semantic matching efficiency of key fields. However, the existing methods mainly target ordinary bills without obvious defects, and it is difficult to meet the recognition requirements of Chinese bills. In addition, there is a lack of open-source Chinese bill datasets, which limits the training and algorithm optimization of bill recognition models. Therefore, based on the above challenges, the research carries out the design of a note recognition model based on the Information Extraction Algorithm (IEA) and establishes a Chinese note dataset.

The novelty of the study is that an IEA algorithm is proposed for the recognition needs of accounting bills, and a Chinese bill recognition model is constructed. Meanwhile, in order to solve the lack of open-source Chinese bill datasets, the research innovatively constructs a Cab Invoice Dataset (CID) for model training and validation. The study introduces regularized expressions as a matching tool for the information of the ticket structure data, which improves the recognition efficiency of key fields. The IEA algorithm shows higher performance than other algorithms in terms of accuracy and F1 value on both the self-built CID dataset and the ICDAR 2015 dataset. The research applies deep learning OCR technology to the field of bill recognition and promotes the intelligent development of bill recognition. Meanwhile, the CID dataset established by the research provides a positive contribution to the intelligence of Chinese bill recognition.

2 Related Works

As one of the important research directions in the field of computer vision, OCR, many experts, and scholars have carried out studies and explorations on bill recognition from the aspect of OCR technology. To alleviate the pressure of accounting for financial bills, [5] proposed a fast convolutional recurrent neural network based on a fast financial bill detection network by improving the loss function, region suggestion network, and non-maximal suppression, thus improving the recognition speed in the public service capability improvement-oriented International Conference on the Integration and Innovation of Digital Archival Resources (The International Conference on the Integration of Digital Archival Resources (ICDAR) 2019 dataset recognition speed is increased by 50%. Aiming at the problem that the standard OCR technology is weak in recognizing financial bills of diverse types and mixed languages, [6] also proposed a deep learning-based technique for detecting and recognizing full-content text information of financial bills. Utilizing the financial note character framework for the recognition of mixed characters, thus significantly improves the efficiency of the financial accounting system given. In order to solve the problem of scarcity of datasets for automatic parsing of invoices and poor performance of the algorithms, [7] proposed to utilize the OCR-based text extraction method to transform the images into machine-coded text. The hyperparameters of the object detection algorithm were adjusted so as to achieve effective recognition of multiple types of invoice images. Aiming at the poor applicability and robustness of computer-aided techniques for feature extraction in specific financial invoices, [8] constructed a lightweight recognition model using deep learning techniques. By preprocessing the invoice image using a sequence transformation model and segmenting a single character using a horizontal-vertical projection method, a high accuracy of invoice recognition was obtained in a real dataset. To address the shortcomings of the information extraction algorithms that extract only machine-readable words, a Vietnamese invoice recognition network was proposed based on Vietnamese OCR characters [9]. By utilizing a Graph Convolutional Network (GCN, GCN) for text detection and recognition and another GCN for image classification, a recognition accuracy of 99.50% was obtained.

The recognition model of bill image is mostly built based on deep learning OCR, while the information extraction method also determines the recognition effect of bill image. For the extraction of image information, scholars at home and abroad have also explored it in various aspects. [10] proposed a pre-training model for text and typography collections to fundamentally address the contextual and spatial semantics of image text in 2D space. The error due to wrong text ordering is minimized by encoding the relative position of the text in 2D space and learning the documents with unmarked positions using a region masking strategy. In order to improve the imperceptibility of the generated image, a strategy of embedding watermarks by using the image feature points is proposed on the basis of feature extraction information embedding, [11]. Through discrete wavelet variation processing, thus realized the embedding and extraction of digital image watermarks, and passed various image attack tests. [12] in order to extract the information related to the population of Quebec, designed a set of machine learning model-based registry book information extraction schemes. Classification, text detection, handwritten text recognition, and named entity recognition were performed on the register information pages, which led to the classification of family and genealogical relationships of the Quebec population. [13] proposed a novel generalized image fusion framework based on cross-domain long-range learning and the Swin variant with a view to achieving full fusion and global interaction of complementary information. The fusion of multi-scene images is carried out through intra-domain fusion units based on self-attention and inter-domain fusion units based on cross-attention, thus realizing the structure maintenance as well as detail preservation of digital images. To address the issue of security of information transmission in digital images, [14] proposed a steganography technique using a neural-based transmission algorithm to generate steganographic images. By using a conditional generative adversarial network for image information de-typing, an area under the curve of 0.53 was achieved in the PASCAL VOC12 dataset.

Combined with the above, it can be seen that scholars at home and abroad have conducted research in various aspects using deep learning OCR, and more results have been achieved in note recognition. However, the detection efficiency and accuracy of current bill recognition models are difficult to meet the bill recognition needs of accountants in large Chinese enterprises, and there is a lack of open-source Chinese bill datasets for research. Therefore, the research proposes an IEA algorithm for enterprise accounting and constructs a note recognition model based on it. And Dueto the lack of open-source Chinese bills, the study innovatively constructs a CID dataset for the training of the recognition model.

3 Methods and Materials

To improve the current efficiency of bill recognition, the study first designed the IEA algorithm. Using Detection and Recognition Neural Network (DBNet) for detecting textual information in bills, Convolutional Recurrent Neural Network (CRNN) is responsible for transforming text recognition problems into sequence recognition problems, and connectionist Temporal Classification (CTC) is used to screen and monitor the output strings of CRNN. Secondly, a bill recognition model was constructed based on the designed IEA algorithm. Finally, the study constructed a CID dataset based on the collected national taxi receipts.

3.1 IEA Design based on Deep Learning OCR

The text of bills is characterized by dense and multi-scale, and the integrity of its information extraction directly determines the performance of the algorithm. Therefore, the study designs an IEA algorithm based on deep learning OCR techniques such as DBNet and CRNN. Firstly, the DBNet network is utilized for the detection of text information of bills. DBNet is a pixel segmentation-based text detection network that utilizes Feature Pyramid Network (FPN) for multi-scale feature fusion, [15]. The training process is optimized by Differentiable Binarization (DB) thus achieving simplified processing, [16]. The specific structure is shown in Figure 1 (Appendix), [17].

As can be seen from Figure 1 (Appendix), the DBNet network first inputs the to-be-processed image into the backbone network ResNet18 for multiscale feature extraction, and the features are up-sampled and fused from deeper to shallower levels using FPN and merged to obtain the feature image. Secondly, based on the obtained feature image, further prediction and output of probability image and threshold image are carried out to obtain an approximate binary image using DB. Finally, the text envelopment curve is processed and output. In the DBNet text detection process, the study utilizes data augmentation in the form of pre-processing the input image and using a scaling algorithm to generate training labels to reduce the probability of text adhesion prediction. In this case, the labels of the probabilistic and binary images are obtained by the scaled-down offset of the original localization frame position, which is calculated as shown in equation (1).

$$A = \frac{B(1-r^2)}{C} \tag{1}$$

In equation (1), A denotes the offset of the position reduction of the localization frame; B denotes the area of the localization region; C denotes the perimeter of the localization region; and r denotes the label contraction coefficient, which is taken as 0.4 in the study. The labels of the thresholded image are then computed by the combined calculation of the contracted frame obtained by the inward contraction offset of the original localization frame and the expanded frame obtained by the outward expansion offset. Since the use of fixed-threshold segmentation in IEA leads to non-trivial gradients in the training phase of the DBNet network, it was investigated to train the network using a function that approximates the standard binarization. An example formula for the generating function of the approximated DB image is shown in equation (2).

$$g(x) = \frac{1}{1+e^{-kx}} \tag{2}$$

In equation (2), $g(x)$ denotes the approximated binary image; k denotes the fixed factor; x denotes the difference between the pixel value of the probability image and the pixel value of the corresponding threshold image; and e denotes the logarithm. Therefore, the loss function of the bill detection network in IEA consists of the loss function of the probability image, the threshold image, and the binary image together, as shown in equation (3).

$$\begin{cases} Loss = L_d + \mu L_s + \nu L_f \\ L_d = 1 - 2\dfrac{intersection\_area}{total\_area} \\ L_s = \sum_{i \in s_i} y_i \log d_i + (1 - y_i)\log(1 - d_i) \\ L_f = \sum_{i \in E_d} \left| o_i^* - q_i^* \right| \end{cases} \quad (3)$$

In equation (3), $Loss$ denotes the DBNet loss function in IEA; L_d denotes the loss function of the binary image; μ and ν denotes the corresponding weights of the probability image and the threshold image; L_s denotes the loss function of the probability image; L_f denotes the loss function of the threshold image; $intersection\_area$ denotes the intersection range between the binary image and the labeled information; $total\_area$ denotes the common range between the binary image and the labeled information; y_i denotes the input information, which is 1 or 0; d_i denotes the chance that the input information is predicted to be 1; denotes the data set; denotes the labeled information of the threshold image; denotes the CRNN as a commonly used OCR text recognition network. the chance that the input information is predicted to be 1; s_i denotes the dataset; o_i^* denotes the labeling information of the threshold image; q_i^* denotes the threshold image. CRNN, as a commonly used OCR text recognition network, is capable of transforming the text recognition problem into a sequence recognition problem, [18]. CRNN has certain advantages in long text detection, but the cyclic layer of CRNN is prone to output repetitive strings in sequence features of neighboring moments that are close to each other, [19]. Therefore, the study introduces the combination of CTC transcription and CRNN to construct the IEA algorithm. The study sets the cyclic layer feature size of the CRNN network input in IEA to be fixed as [512,1,24], and its network structure mainly consists of three parts, as shown in Figure 2 (Appendix).

As can be seen from Figure 2 (Appendix), this CRNN network is mainly composed of convolutional, cyclic, and transcription layers. Among them, the convolutional layer extracts the image convolutional feature sequences in the standard frame through preprocessing and convolutional neural network, and the recurrent layer uses Bidirectional Long and Short-Term Memory networks (BiLSTM) for text sequence feature extraction. In addition, the transcription layer uses CTC connected temporal classifier for character alignment, [20]. In this, the sequence of BLSTM is passed into two Long- and Short-Term Memory networks (LSTM) structures in a forward and reverse manner, and the merged output of the two LSTMs is passed as a new sequence into the ever-stacking BLSTM structures, [21]. Before passing into the CTC, the sequences need to be normalized by the Softmax function to get their probability matrix. The Softmax function expression is shown in equation (4).

$$Softmax = \frac{e^{x_i}}{\sum_{i=1}^{n} e^{x_i}} \quad (4)$$

In equation (4), x_i denotes the input sample; n denotes the number of samples. According to the law of probability distribution, the column vector of the matrix is obtained by the Argmax function, which can return the maximum value of the characters in the column vector after calculation, and its expression formula is shown in equation (5).

$$Argmax = \{ x \mid x \in S \wedge \forall y \in S : f(y) \le f(x) \} \quad (5)$$

In equation (5), x and y denote two unequal input sequences; S denotes the function definition domain; $f(y)$ and $f(x)$ both denote functions. After the cyclic layer and Argmax function processing, the maximum probability character category of each frame of the sequence is output as the corresponding result of that frame. At this point, CTC combines each frame of the sequence to predict the highest probability sequence label. The string binding path of CTC in the IEA algorithm is shown in Figure 3 (Appendix).

In Figure 3 (Appendix), the study assumes that the CTC transcription time step is 17, and when its transcription defines the blank character 'ε' , it indicates that there is no corresponding character object in the current frame. For each character string, the study added blanks between each character and at the beginning and end of the character. For the legal paths, all of them can be jointly represented by their corresponding strings and time steps. If the CTC predictions for each frame are more independent, the conditional probability of each path is generated as shown in equation (6).

$$p(h \mid x) = \prod_{t=1}^{T} y_{ht}^{t} \quad (6)$$

In equation (6), p denotes the probability; h denotes the legal path; y_{ht}^{t} denotes the current frame output in h at t. The probability of the output string can be obtained by summarizing the conditional probabilities of all the legal paths, and the specific formula is shown in equation (7).

$$p(w|x) = \sum_{V(h)=w} p(h|x) \tag{7}$$

In equation (7), w denotes the output string; V denotes the de-duplication operation. Meanwhile, the study sets the CTC in IEA to combine sequence frames in the form of dynamic programming and the node value at the current moment consists of the sum of the node values of the previous moment that are about to pass through the node. Therefore, after determining the input and output labels, the conditional probability of generating paths in the IEA algorithm can be transformed into a negative log-likelihood function, which is expressed as shown in equation (8).

$$Loss_{CTC} = \sum_{i=1}^{n} -wnp(w_i|x) \tag{8}$$

In equation (8), $Loss_{CTC}$ denotes the loss function of CTC. Combining the above, the proposed IEA of the study mainly consists of two parts: text detection and text recognition, in which DBNet network is mainly responsible for text detection of the input ticket image, while RCNN combines with CTC to extract information from the text after labeling the location box.

3.2 Bill Recognition Model Construction based on the IEA Algorithm

Based on the proposed IEA algorithm, the study further proceeds with the construction of the note recognition model and the creation of the dataset. Since in the recognition task of bills, the text usually appears in pairs and the bill keywords have a fixed format. Therefore, for the textual performance characteristics of tickets, the study introduces Regular Expression (RE) as the matching of structured data information of tickets.RE, as a basic process of natural language processing, performs the acquisition of textual content that meets the logical conditions using logical formulas composed of pre-defined special characters and their combinations. Definition of information extraction rules based on fields such as province, invoice code and number, license plate number, time (including date of ride, time of boarding and alighting), mileage, and amount of money in the cab ticket. Among them, the invoice code information extraction rule is shown in equation (9).

$$ic = '[0-9]\{12\}(?!\backslash d)' \tag{9}$$

In equation (9), ic means the code of cab invoice; $[0-9]$ means any character between 0-9 in the specified range; $\{12\}$ means match 12 times; $\backslash d$ means match any number between 0-9; $(?!\backslash d)$ means match the string found at the beginning of any string that does not match d. The invoice number information extraction rule is shown in equation (10).

$$in = '[0-9]\{8\}(?!\backslash d)' \tag{10}$$

In equation (10), in denotes the number information of cab invoice; $\{8\}$ denotes matching 8 times. The license plate number information extraction rule is shown in equation (11).

$$cn = '[pa]\{1\}[A-Z]\{1\}[A-Z0-9]\{5\}' \tag{11}$$

In equation (11), cn denotes the license plate number information shown on the cab invoice; $[pa]$ denotes the specified abbreviation of any province in China; $\{1\}$ denotes matching 1 time; $[A-Z]$ denotes any character between the specified ranges A-Z; $[A-Z0-9]$ denotes any character between the specified ranges A-Z, 0-9; $\{5\}$ denotes matching 5 times. For other fields, the study utilizes keywords for field information matching. Combined with the above, the flow of the IEA-based note recognition model proposed by the study is shown in Figure 4 (Appendix).

As can be seen in Figure 4 (Appendix), the model first preprocesses the input ticket image, and second utilizes the DBNet network for text detection. According to the labeled boxes added to the text image by DBNet network, the convolutional layer of CRNN is utilized for image information extraction, and after the cyclic layer and Argmax function, the prediction is performed by CTC transcription combined with sequence features. Finally, RE is utilized to match the information of the structured data of the bill and output the recognition results. At the same time, the research carried out the establishment of the CID dataset. Collect 541 taxi invoice images from different provinces in China through web crawling technology and autonomous forms, and use a crowdsourcing platform to collect 300 taxi invoice images, totaling 841 images. Due to the influence of shooting background, preservation method, etc., the collected ticket images have recognition difficulty, so the study introduces Image Enhancement Technology (IET) and Text Enhancement Technology (TET) to process the CID dataset. IET is based on random cropping, color interference, noise processing, and image blurring to perform bill image enhancement and filter out illegible images. The specific form is shown in Figure 5 (Appendix).

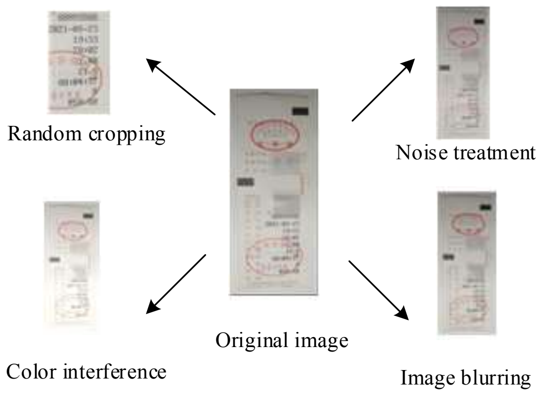


Fig. 5: Schematic of image enhancement of dataset

In Figure 5 (Appendix), random cropping mainly intercepts a certain part of the image arbitrarily from the original image, while color interference randomly adjusts the parameter information such as brightness and contrast of the ticket image within a certain range of values. Noise processing refers to adding various types of noise to the original image, while image blurring refers to blurring the image using different types of filtering techniques. TET refers to replacing the text information of the original image using another word based on text editing algorithm. This is achieved by migrating the foreground style of the text, extracting the background, and fusing the text and background. Based on the above processing, the final CID dataset of 15000 text images is obtained, of which 1000 are test sets.

4 Results

The recognition result of the bills determines the value of the recognition model in the intelligent application of bills. Therefore, the study firstly validates the performance of the IEA algorithm in terms of accuracy and frames Per Second (FPS), and secondly validates and analyzes the proposed bill recognition model.

4.1 Performance Validation of IEA based on Deep Learning OCR

To validate the efficiency of the proposed IEA algorithm for extracting the information from bill images, the study conducted algorithm validation using the self-constructed CID dataset and the ICDAR 2015 dataset. The ICDAR 2015 dataset is a natural scene text dataset released by the International Conference on Document Analysis and Recognition in 2015, which contains 1000 training sets and 500 test sets. The study set the number of iterations of the IEA algorithm to be 1000 and the initial learning rate to be 0.001. The F1 scores and FPS metrics of the IEA algorithm in the CID dataset under different DBNet backbone networks are shown in Figure 6 (Appendix).

A comparison of the F1 scores of the IEA algorithms in the CID dataset under the three backbone networks is shown in Figure 6(a) (Appendix). It can be seen that the F1 scores of all three backbone networks increase and converge with the increase of iterations. After 1000 iterations, the F1 score of the ResNet18 network is 93.68%, which is 7.01% and 8.40% higher than the F1 scores of the other two networks, respectively. This indicates the superiority of the study utilizing ResNet18 as the backbone network. Comparing the FPS of the three networks in Figure 6(b) (Appendix), it can be seen that the FPS of ResNet50 is the lowest among the three networks and its variation is the largest, which indicates that the FPS of ResNet50 is unstable and prone to lagging. While comparing ResNet18 and MobileNetV3, ResNet18 exhibits a more stable FPS. The ResNet18 network is shallower, has fewer parameters, and offers higher computational efficiency, which may be the reason for its superior F1 score. In contrast, the ResNet50 network is deeper, has a higher computational cost, and requires more parameters and computational resources. MobileNetV3, designed for lightweight implementation, significantly reduces its parameter and computational demands, leading to a lower accuracy on some complex tasks compared to the ResNet series. Meanwhile, the study further compares the performance of the three networks in the ICDAR 2015 dataset, as shown in Figure 7 (Appendix).

From the F1 scores of the three backbone networks in the ICDAR 2015 dataset in Figure 7(a) (Appendix), it can be seen that the study proposes that the IEA algorithm constructed by utilizing ResNet18 as the backbone network is more advantageous. A comparison of the FPS of the three backbone networks is shown in Figure 7(b) (Appendix), MobileNetV3 has the highest FPS among the three backbone networks, which is 6.87 after 1000 iterations compared to ResNet18's 6.54. However, a comparison of the magnitude of fluctuations in the network's FPS shows that ResNet18 is more stable. The above results indicate that it is reasonable for the study to utilize ResNet18 as the backbone network for DBNet detection of text in the IEA algorithm. On this basis, the study compares the performance of the IEA algorithm with the currently popular Local Binary Pattern (LBP) algorithm and Scale-Invariant Feature Transform (SIFT). The results of the loss profile comparison of the three algorithms in the two datasets are shown in Figure 8 (Appendix).

Comparing the changes in the loss curves of the three algorithms in the CID dataset in Figure 8(a) (Appendix), it can be seen that the LBP algorithm has the highest loss value and has no tendency to converge after 1000 iterations. While the SIFT algorithm has a loss value of 0.59 after 1000 iterations, which is 79.66% more than IEA. From Figure 8(b) (Appendix), it can be seen that all three algorithms have lower loss values in the ICDAR 2015 dataset than the CID dataset. Among them, the IEA algorithm proposed in the study is more advantageous in terms of loss profile change. As a computer vision algorithm for texture classification, the LBP algorithm is sensitive to changes in lighting conditions, and the stability and reliability of its features can be compromised in the presence of high-frequency noise. This may be the reason for its higher loss curves in both datasets. The SIFT algorithm, on the other hand, experiences a significant performance decline in low image quality or high-noise environments and is prone to getting trapped in local optima. The performance comparison of the three algorithms in the two datasets is shown in Table 1.

As can be seen from Table 1, the accuracy of the proposed IEA algorithm in the CID dataset is 92.12%, which is 15.71% and 4.00% higher than the other two algorithms, respectively. Comparing the F1 values of the three algorithms, the LBP algorithm has the lowest F1 value, which may be because the LBP algorithm is more significantly affected by environmental interference and its multiple sampling of image pixel points and sampling radius leads to a decrease in the ability to characterize the image. Combining the performance validation results of the three algorithms in the ICDAR 2015 dataset, it can be seen that all three algorithms have achieved more than 90% accuracy, but the IEA algorithm proposed in the study still has obvious advantages.

Table 1. Performance comparison of three algorithms

Algorithm	CID datasets		ICDAR 2015 datasets	
	Accuracy (%)	F1-score (%)	Accuracy (%)	F1-score (%)
IEA	92.12	93.68	95.46	94.74
LBP	79.61	78.46	90.78	89.53
SIFT	88.58	89.24	92.21	93.88

4.2 Validation of the IEA-based Note Recognition Model

To further confirm the application value of the IEA algorithm in bill image recognition, the study introduces the previous bill recognition model for performance comparison. The changes in training indexes of different models in the CID dataset are shown in Figure 9 (Appendix).

Figure 9(a) (Appendix) and Figure 9(b) (Appendix) show the training accuracy and FPS comparison results of four models on the CID dataset training set. It can be seen that the ticket recognition model based on the IEA algorithm proposed in the study has a relatively small difference in recognition accuracy in the training set compared to the reference [6], but the model proposed in the study converges faster and has a more stable recognition effect. From the FPS changes of the four models in Figure 9(b) (Appendix), it can be seen that there are significant fluctuations in the FPS of all four models. As the number of iterations increases, the identification model proposed in the study has a higher FPS, while reference [8] has a more stable FPS. The validation results of the four models on the CID dataset test set are shown in Table 2.

Table 2. Comparison of validation results of different models in the test set of CID dataset

Target	Method			
	IEA	Reference [6]	Reference [7]	Reference [8]
Accuracy (%)	94.31	87.35	84.81	78.97
F1-score (%)	93.28	90.05	83.22	75.66
mAP (%)	90.64	89.74	82.55	74.68
FPS	38.97	36.59	32.45	22.04

As can be seen from Table 2, the recognition model based on the IEA algorithm has a high recognition accuracy of 94.31% in the CID dataset, which is an average increase of 12.87% compared to the other methods. Comparing the whole class average precision (mAP) of the four models in the CID dataset, the recognition model proposed in the study increased by 1.00%, 9.80%, and 21.37%, respectively. Compared to the other three models, the bill recognition model based on the IEA algorithm has a more desirable FPS. This indicates that the IEA algorithm designed by the study using deep learning OCR technology has a superior recognition performance in bill recognition, and the recognition model based on IEA is more advantageous than other methods. Finally, the study further analyzes the validation results of the IEA algorithm-based note recognition model, as shown in Figure 10 (Appendix).

As can be seen in Figure 10(a) (Appendix), all the key fields in the bill image are effectively detected by the research-proposed recognition model. Combined with the detection visualization results, it can be illustrated that the research-proposed DBNet with ResNet18 as the backbone network can accurately detect the bill images. The visualization results of the recognition of bill images by the research proposed model are shown in Figure 10(b) (Appendix). It can be seen that the model can recognize the invoice code of the ticket image completely. It also shows that the bill recognition model based on the IEA algorithm proposed in the study has positive application value in the development of bill recognition intelligence.

5 Conclusion

To improve the bill recognition efficiency in the accounting industry and promote the intelligent development of bill recognition, the study proposes an IEA algorithm based on deep learning OCR technology and introduces RE technology to construct a bill recognition model. Firstly, the DBNet network is utilized as the image text detection of the IEA algorithm, and secondly, CRNN is combined with CTC for the recognition module of the algorithm. Finally, the model construction is carried out and a kind of Chinese bill dataset is established. Experimental validation shows that the accuracy of the IEA algorithm in the research self-constructed CID dataset is 92.12%, which is 15.71% and 4.00% higher than the other two algorithms, respectively. The F1 value of the IEA algorithm in the ICDAR 2015 dataset is increased by an average of 3.37% compared to the other two algorithms. The validation of different models shows that the mAP of the proposed model of the study in the CID dataset is increased by 1.00%, 9.80%, and 21.37%, and the FPS is increased by 6.50%, 20.09%, and 76.81%, respectively, over the other models. The results show that the IEA algorithm designed by the study using DBNet network, RCNN, and CTC has superior application in bill recognition, and the bill recognition model based on the IEA algorithm can effectively enhance the accuracy and precision of bill recognition and improve the effect of bill recognition.

The IEA algorithm proposed in our study has made significant technological advancements in the field of bill recognition. Compared with existing methods, the IEA algorithm, by integrating DBNet, CRNN, and CTC technologies, not only improves the accuracy and efficiency of recognition but also enhances the model's generalization ability through the self-built CID dataset. Especially in processing Chinese bills, the IEA has demonstrated outstanding performance, which has been verified by experimental results on both the CID and ICDAR 2015 datasets. Furthermore, the high FPS of the IEA algorithm indicates its potential application value in real-time bill processing. However, the shortcoming of the study is that only the design of the bill recognition model is carried out, and the visualization platform of the recognition system is not explored in depth. In the future, we will further design and develop the automatic processing system of bill images based on the proposed bill recognition model, to realize the intelligent recognition of bills.

References:

[1] Balamurugan A, Krishna M V, Bhattacharya R, Mohammed S, Haralayya B, Kaushik P. Robotic Process Automation (RPA) in accounting and auditing of business and financial information. *The British Journal of Administrative Management,* Vol.58, No.157, 2022, pp. 127-142.

[2] Cartwright D J. Abstracts from the 132nd meeting of the Tennessee academy of science November 18, 2022. *Journal of the Tennessee Academy of Science*, Vol.97, No.1, 2023, pp. 18-54. doi: 10.47226/97.1.18.

[3] Pal S, Roy A, Shivakumara P, Pal U. Adapting a Swin transformer for license plate number and text detection in drone images. *Artificial Intelligence and Applications*, Vol.1, No.3, pp. 145-154. doi: 10.47852/bonviewAIA3202549.

[4] Singh N, Kumar M, Singh B, Singh J. DeepSpacy-NER: an efficient deep learning model for named entity recognition for Punjabi language. *Evolving Systems*, Vol.14, No.4, 2023, pp. 673-683. doi: 10.1007/s12530-022-09453-1.

[5] Zhang H, Dong B, Zheng Q, Feng B. Research on fast text recognition method for financial ticket image. *Applied Intelligence*, Vol.52, No.15, 2022, pp. 18156- 18166. doi: 10.1007/s10489-022-03467-7.

[6] Zhang H, Dong B, Zheng Q, Feng B, Xu B, Wu H. All-content text recognition method for financial ticket images. *Multimedia Tools and Applications*, Vol.81, No.20, 2022, pp. 28327-28346. doi:10.1007/s11042-022-12741-2.

[7] Chazhoor A, Sarobin V R. Intelligent automation of invoice parsing using computer

vision techniques. *Multimedia Tools and Applications*, Vol.81, No.20, 2022, pp. 29383-29403. doi: 10.1007/s11042-022-12916-x.

[8] Wang T, Qiu M. A visual transformer-based smart textual extraction method for financial invoices. *Mathematical Biosciences and Engineering*, Vol.20, No.10, 2023, pp. 18630-18649. doi: 10.3934/mbe.2023826.

[9] Tran A C, Ho L T, Nguyen H T. Information extraction from invoices by using a graph convolutional neural network: a case study of Vietnamese stores. *IEIE Transactions on Smart Processing & Computing*, Vol.11, No.5, 2022, pp. 316-323. doi: 10.5573/ieiespc.2022.11.5.316.

[10] Hong T, Kim D, Ji M, Hwang W, Nam D, Park S. Bros: A pre-trained language model focusing on text and layout for better key information extraction from documents. *In Proceedings of the AAAI Conference on Artificial Intelligence,* Vol.36, No.10, 2022, pp.10767-10775. doi:10.1609/aaai.v36i10.21322.

[11] Shan C, Zhou S, Zhang Z, Li M. Digital watermarking method for image feature point extraction and analysis. *International Journal of Intelligent Systems*, Vol.37, No.10, 2022, pp. 7281-7299. doi: 10.1002/int.22881.

[12] Tarride S, Maarand M, Boillet M, McGrath J, Capel E, Vézina H, Kermorvant C. Large-scale genealogical information extraction from handwritten Quebec parish records. *International Journal on Document Analysis and Recognition (IJDAR)*, Vol.26, No.3, 2023, pp.255-272. doi: 10.1007/s10032-023-00427-w.

[13] Ma J, Tang L, Fan F, Huang J, Mei X, Ma Y. SwinFusion: cross-domain long-range learning for general image fusion via swin transformer. *IEEE/CAA Journal of Automatica Sinica*, Vol.9, No.7, 2022, pp.1200-1217. doi: 10.1109/JAS.2022.105686.

[14] Garg M, Ubhi J S, Aggarwal A K. Neural style transfer for image steganography and destylization with supervised image to image translation. *Multimedia Tools and Applications*, Vol.82, No.4, 2023, pp. 6271-6288. doi: 10.1007/s11042-022-13596-3.

[15] He X, Liu X. Improved angle steel character detection algorithm based on DBNet. *Academic Journal of Science and Technology*, Vol.3, No.3, 2022, pp. 64-69. doi: 10.54097/ajst.v3i3.2549.

[16] Zheng J, Li J, Ding Z, Kong L, Chen Q. Recognition of expiry data on food packages based on improved DBNet. *Connection Science*, Vol.35, No.1, 2023, pp. 1-16. doi: 10.1080/09540091.2023.2202363.

[17] Zhou X, Tang C, Huang P, Tian S, Mercaldo F, Santone A. ASI-DBNet: an adaptive sparse interactive resnet-vision transformer dual-branch network for the grading of brain cancer histopathological images. *Interdisciplinary Sciences: Computational Life Sciences*, Vol.15, No.1, 2023, pp. 15-31. doi: 10.1007/s12539-022-00532-0.

[18] Al-Nabhi H, Krishna K L, Shareef D. Efficient CRNN recognition approaches for defective characters in images. *International Journal of Computing and Digital Systems*, Vol.12, No.1, 2022, pp. 1417-1427. doi: 10.12785/ijcds/1201114.

[19] Yasin S A, Prasada R P. Enhanced CRNN-based optimal web page classification and improved tunicate swarm algorithm-based re-ranking. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, Vol.30, No.5, 2022, pp. 813-846. doi: 10.1142/S0218488522500246.

[20] Noreldeen H A, Huang K Y, Wu G W, Zhang Q, Peng H P, Deng H H, Chen W. Feature selection assists BLSTM for the ultrasensitive detection of bioflavonoids in different biological matrices based on the 3D fluorescence spectra of gold nanoclusters. *Analytical Chemistry*, Vol.94, No.50, 2022, pp. 17533-17540. doi: 10.1021/acs.analchem.2c03814.

[21] He S, Sang X, Yin J, Zheng Y, Chen H. Short-term runoff prediction optimization method based on bgru-bp and blstm-bp neural networks. *Water Resources Management*, Vol.37, No.2, 2023, pp. 747-768. doi: 10.1007/s11269-022-03401-z.

APPENDIX

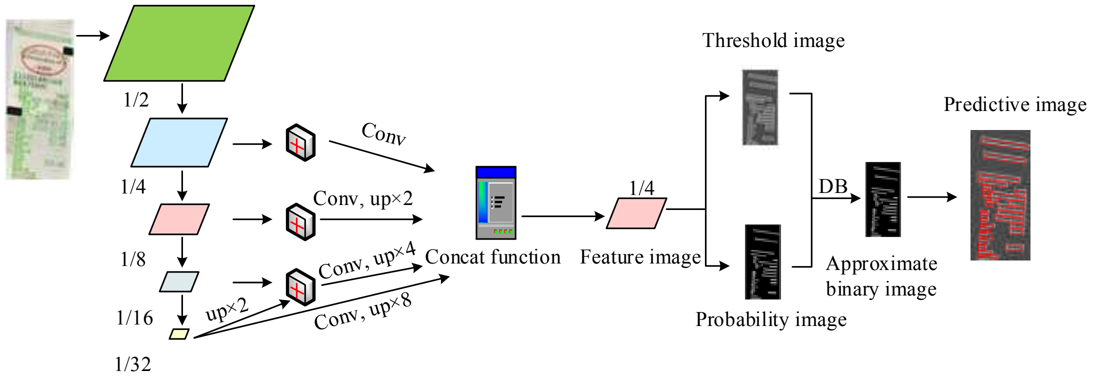


Fig. 1: DBNet structure in IEA

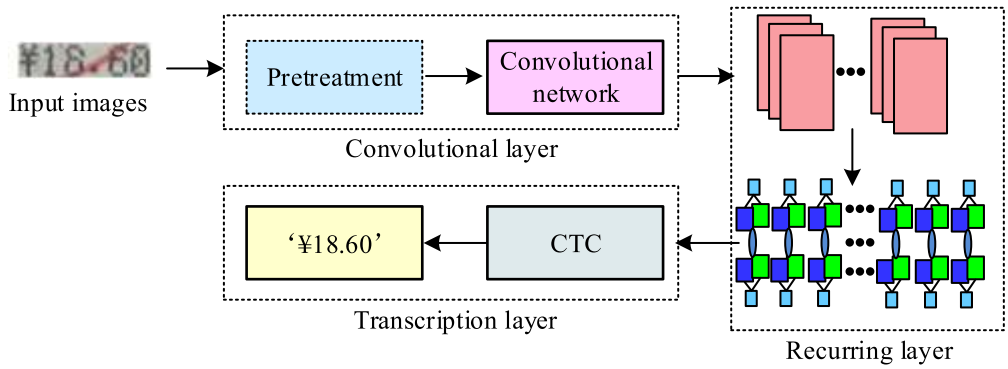


Fig. 2: IEA's CRNN network structure

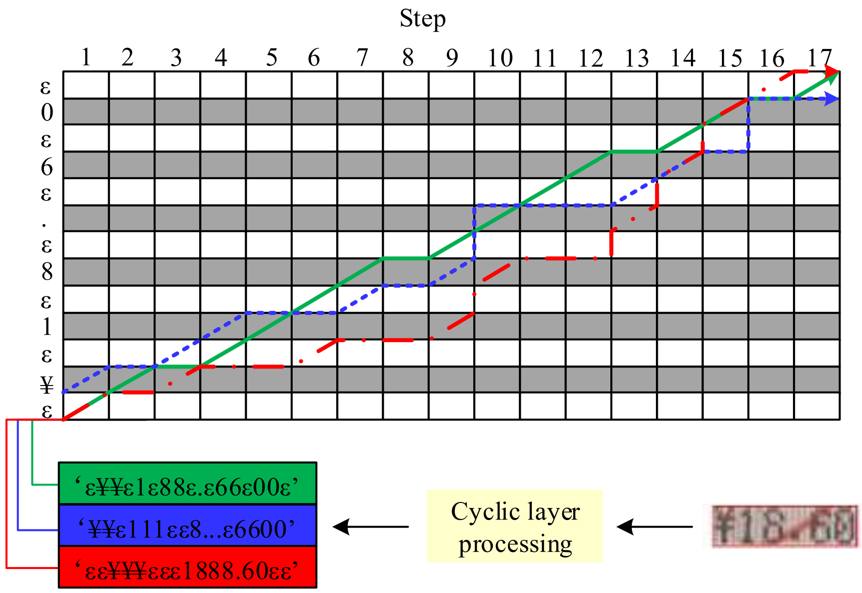


Fig. 3: Schematic of the CTC string path in IEA

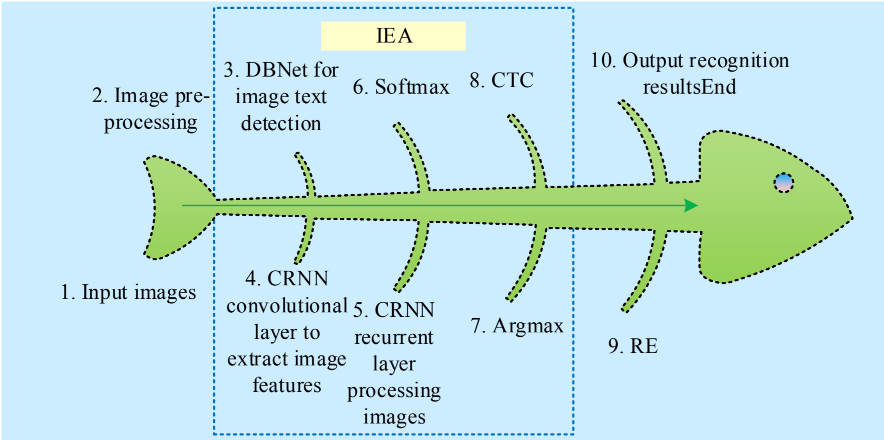
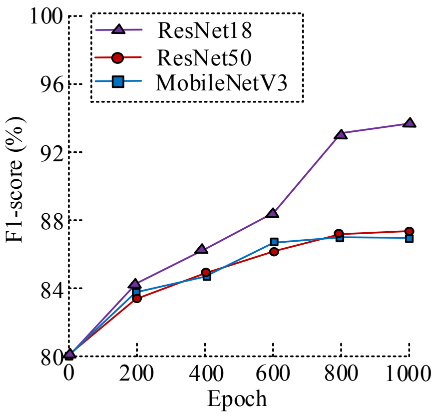
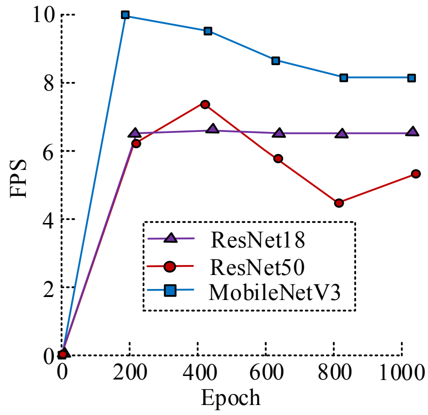


Fig. 4: IEA-based bill recognition modeling process

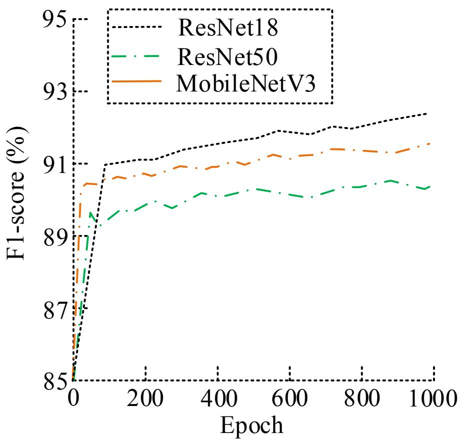


(a) F1-score of IEA algorithms for three backbone networks in CID dataset

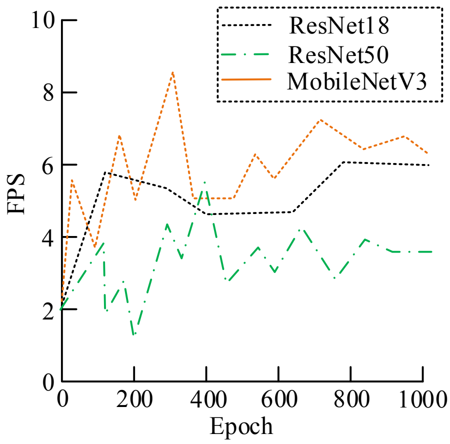


(b) FPS of IEA algorithms for three backbone networks in CID dataset

Fig. 6: Performance comparison of IEA algorithms in CID datasets under different DBNet backbone networks

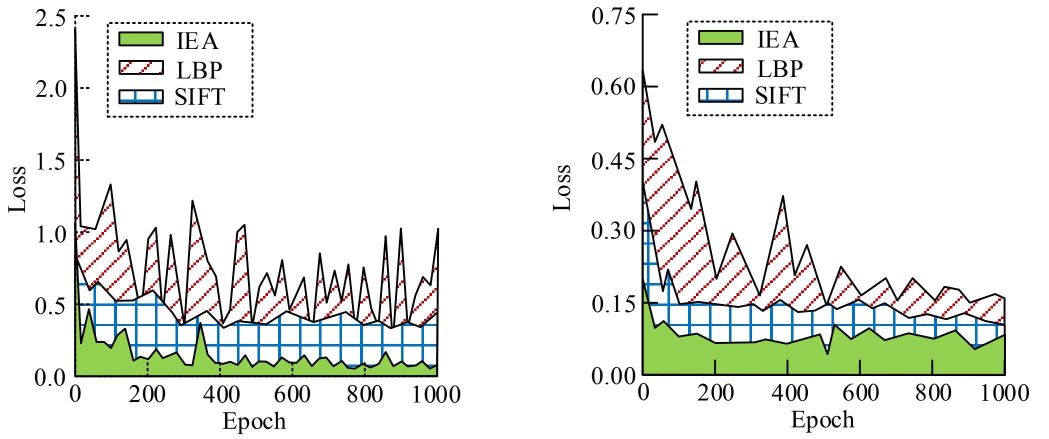


(a) F1-score of IEA algorithms for three backbone networks in ICDAR 2015 dataset



(b) FPS of IEA algorithms for three backbone networks in ICDAR 2015 dataset

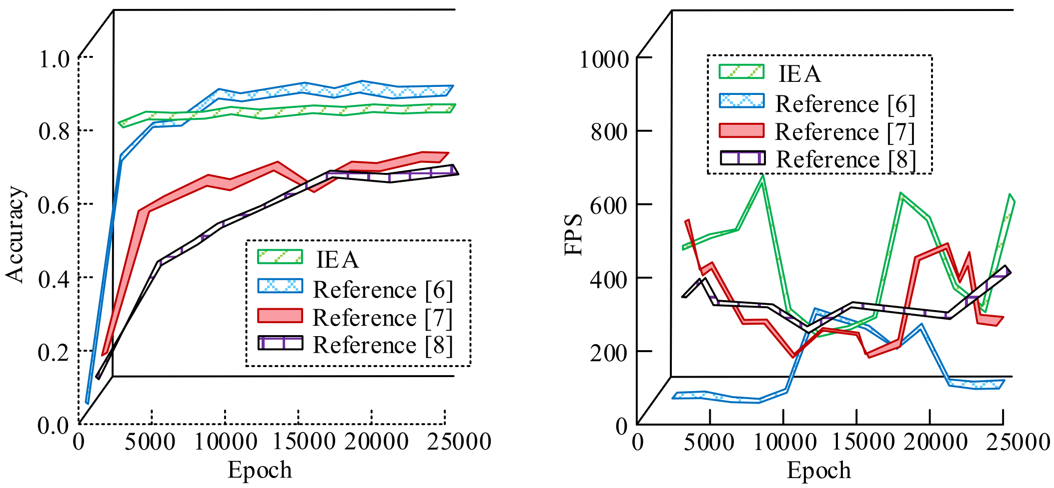
Fig. 7: Performance comparison of IEA algorithms in ICDAR 2015 datasets under different DBNet backbone networks



(a) Comparison of loss curves of three algorithms in CID dataset

(b) Comparison of loss curves of three algorithms in ICDAR 2015 dataset

Fig. 8: Comparison of loss curves of the three algorithms



(a) Variation of training accuracy of different recognition models in CID dataset

(b) Variation of training FPS of different recognition models in CID dataset

Fig. 9: Comparison of training metrics of different methods in the CID dataset

(a) Ticket image inspection results

(b) Ticket image recognition results

Fig. 10: Visualization results of note recognition model based on IEA algorithm

Contribution of Individual Authors to the Creation of a Scientific Article (Ghostwriting Policy)
The author Zhiyin Liu contributed in the present research, at all stages from the formulation of the problem to the final findings and solution.

Sources of Funding for Research Presented in a Scientific Article or Scientific Article Itself
No funding was received for conducting this study.

Conflict of Interest
The authors have no conflicts of interest to declare.