Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org
© 2024 SCPE. Volume 25, Issues 4, pp. 3126–3138, DOI 10.12694/scpe.v25i4.2997

RESEARCH ON AUTOMATIC UNATTENDED BILL COLLECTION, PASTE AND VERIFICTION INTEGRATED ROBOT EQUIPMENT AND CONTROL PLATFORM BASED ON DEEP CONVOLUTIONAL NEURAL NETWORK

CHAO WANG,* XI CHEN, AND YING WANG

Abstract. A new solution for fully automated and unmanned ticket pasting verification based on deep convolutional neural networks is designed to address the issues of low efficiency, error-proneness, and wastage of manpower in the supplier service hall. The technology makes full use of machine vision and image processing, AI precise positioning correction algorithm and other methods to build an automatic unattended bill collection, paste and verification platform. Through the technologies of high-speed identification of invoice information, 3D vision-guidance planning, control of the path of robotic arm , detection of invoice pasting and repeating based on ultrasonic sensors, and tidal temporary storage of paper invoices, and so on, the automatic high-speed identification and inspection of bills in the supplier service hall are realized, and the efficiency and accuracy of bill processing in the supplier hall are improved. Experiments show that this research method reinforces ability of identification calibration and order correlation, and improves the efficiency of Invoice filing.

Key words: integrated robot,AI,3D vision,Automatic ticket collection and paste, Tidal operations,Document recognition.

1. Introduction and examples. OCR short for Optical Character Recognition , the main role of OCR technology is to convert the text in the image into an editable text format, allowing for subsequent text processing and analysis. It can convert scanned paper documents or pictures into an editable text format. NLP short for Natural Language Processing , the main role of NLP techniques is to transform natural language into a form that computers can recognize. It enables parsing and understanding of textual content, facilitating more advanced natural language processing tasks. The basic principle of NLP technology lies in converting natural language into a computer-recognizable form, such as logical expressions or vector space models. AI Vision techniques, also known as computer vision, are a branch of computer science that training computers to replicate the human visual system. This allows digital devices, such as face detectors and QR code scanners, to recognize and process objects in images and videos just like humans do.

With the rapid advancement of artificial intelligence technology, AI visual inspection technology is progressively emerging as a novel "game-changing tool" across diverse industries. The general trend is that AI machines are replacing human work in all directions. AI machines have advantages such as automation technology, high efficiency, high precision, and non-contact capabilities. They are widely used in various fields including industrial production, agriculture, animal husbandry, medicine, smart cities, and the military. Overseas AI machine vision testing has entered a period of significant growth. Currently, the share of AI machine vision testing in China remains relatively low. However, with a substantial increase in demand, AI machine vision testing is exhibiting an accelerated growth trend.

2. Problem formulation. Some business processes, such as the collection, verification, and creation of certificates for paper documents utilized in power grid engineering projects both domestically and internationally, can be effectively executed through information systems. However, manual tasks involving the separation, flipping, certificate creation, ticket issuance, and physical document filing are still in the preliminary research phase.

In recent years, in China, OCR and NLP technology have been commonly used for the identification and verification of various types of paper documents with unstructured information. The OCR intelligent conversion of electronic text has undergone several generations of technological upgrades, making it a mature

*Beijing GuoDentsu Network Technology Co.,LTD, Beijing, Hebei 100000, China (362165419@qq.com))

technology. Based on this, there are related products available that have intelligent acquisition for paper documents and bill verification functions. However, these products only offer functions such as acquisition and checking, which cannot automatically complete the full process operations required for verifying various bills and other documents, matching and pairing them, creating certificates, pasting them into books, conducting multi-link consistency verification, and performing automatic archiving. The main pain point of this business is not only the manual searching and checking, but also the technical challenges such as correlating bills with orders, verifying invoices, printing vouchers, and automating sticker. The related research on this technology is still lacking in our country. Many experts and scholars in our country are still puzzled by practical problems such as pasting in large quantities, correcting errors, and preventing loss of bills.

3. Related work. In order to efficiently address the business pain points, such as a large number of invoice vouchers and repetitive manual labor in the company's supplier service hall, the company is focusing on various aspects of its invoicing process, which includes invoice collection, separation, verification, settlement, and certification. The company is also exploring and researching complex machine learning algorithms to innovate advanced technologies. As a result of these efforts, they have independently developed a fully automatic intelligent ticket collection and labeling machine that enables fully automated operations for receipt handling, division, inspection, review, posting, and archiving.

3.1. Visual positioning guidance. This paper presents a path planning and control method for manipulators based on 2.5D vision guidance. By utilizing a combination of 2D visual analysis and laser measurement technology, multiple invoices can be efficiently collected with just one key press, achieving an accuracy rate close to 100% in fast identification and automatic matching. The proposed system integrates advanced vision technology with image processing, artificial intelligence, and pattern recognition techniques. We apply three techniques, CCD, OCR and NLP, to achieve visual localization and recognition. CCD(Charged Coupled Device) is a technology that uses image sensor to capture and process images in real time to realize target positioning. OCR (Optical Character Recognition) is a software technology that can automatically identify the text into the computer. That is, the printed characters in the paper document is converted into a black-and-white dot matrix image file by optical means and the converted into text format by recognition software. NLP (Natural Language Processing) refers to the technology of interacting with machines using the natural language used by human communication. Natural language can be read and understood by computer through processing .The CCD intelligence is effectively employed to gather multiple invoice images, integrating AI perception analysis and self-evolution learning of big data. This comprehensive approach ensures the delivery of accurate invoice data through the utilization of AI visual intelligence and OCR+NLP review double insurance. Consequently, it enables an initial review of invoices with a remarkably high accuracy rate nearing 100% [1, 2, 3].

The real world is three-dimensional, and the image projected on the camera lens (CCD/CMOS) is two-dimensional, the ultimate purpose of visual processing is to extract the relevant three-dimensional world information from the perceived two-dimensional image [4, 5]. To put it simply, optical processing is carried out on the surrounding environment of the robot. First, the camera is used to collect image information, compress the collected information, and then feed it back to a learning subsystem composed of neural network and statistical methods. Finally, the learning subsystem will connect the collected image information with the actual position of the robot to accomplish autonomous navigation and positioning function [6, 7, 8].

3.1.1. Camera calibration algorithm.

Calibration parameters. The objective of camera calibration is to rectify lens distortion, establish a linear correlation between image position and spatial coordinates, and convert camera coordinates into pixel coordinates. 2D-3D mapping parameters: Camera calibration entails employing the direct linear transformation method along with a calibration board for point calibration in order to compute the camera parameters (including correction for camera distortion).

$$Z\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & 0 & C_x \\ 0 & f_x & C_x \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = KP \tag{3.1}$$

In the above formula, K is the Intrinsics of the camera.

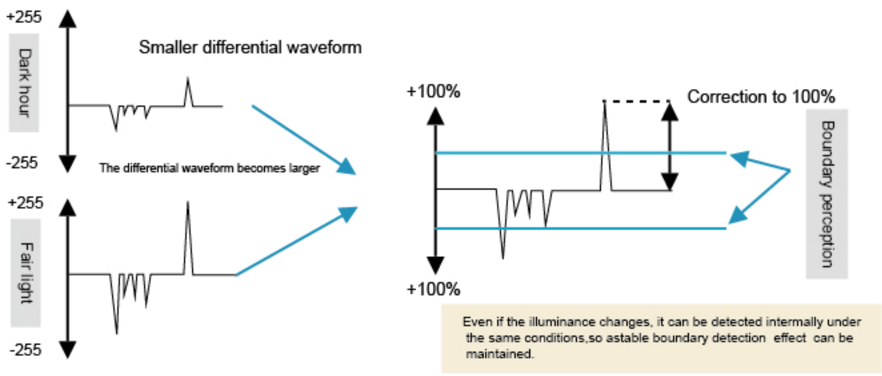

Fig. 3.1: Subpixel processing graph

Calibrating external parameters. The purpose of external parameters is to establish the correspondence between the camera coordinate system and the external world coordinate system (manipulator coordinate system / 3D space) and convert the coordinates of image elements into world coordinates. In this study, we utilize P's coordinates in the camera coordinate system (i.e., with the camera as the origin). Therefore, it is necessary to initially transform Pw from the world coordinate system to P in the camera coordinate system. The pose of the camera can be described by a rotation matrix R and a translation vector t, thus:

$$ZP_{uv} = K(RP_w + t) = KTP_W \tag{3.2}$$

where R and t represent the external parameters of the camera.

3.1.2. Machine vision and image processing. First of all, perform preprocessing which includes hashing, noise reduction, filtering, binarization, and edge detection. The second step involves extracting features by mapping the feature space to parameter space. The third step focuses on image segmentation and template recognition. In the fourth step, adopt the edge grabbing algorithm: it utilizes the Canny edge detection algorithm to apply Gaussian filtering on the image for noise elimination. Then, it calculates the gradient of each pixel to determine the direction of brightness change and finally detects edges through non-maximum suppression and double threshold processing. The fifth step involves utilizing two perpendicular edges to calculate the intersection points, while also determining the coordinate deviation and angle offset in relation to the template [9].

(1) Gaussian filtering (noise reduction). Gaussian blur is an image blurring filter that applies a normal distribution to calculate the transformation of each pixel in the image.

$$G(u, v) = \frac{1}{2\pi\sigma} * e^{\frac{-(u^2+v^2)}{2\sigma^2}} \tag{3.3}$$

In two dimensions, the contour line which surfaces generated by this formula is concentric circles that are normally distributed from the center. The convolution matrix, which a non-zero distribution and consisted of pixels is subsequently transformed using the original image, whereby each pixel's value is determined through a weighted average of its neighboring pixels' values. Notably, the original pixel's value carries the highest weight due to its largest Gaussian distribution value, while weights decrease for neighboring pixels as they move farther away from the original pixel. This method effectively preserves edge effects better than other equalizing blur filters and thus represents an optimal approach for enhancing image quality in various applications.

(2) Boundary detection (grabbing algorithm: using Canny edge detection algorithm). First of all, the information is differentiated. This is shown in Figure 3.1.

Then the information is subpixel processed, the following is shown in Figure 3.2.

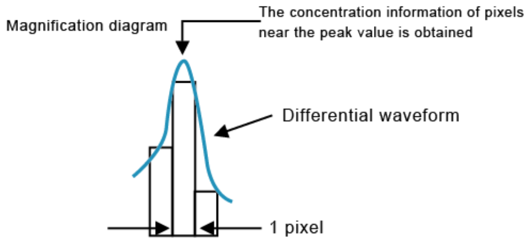

Fig. 3.2: Subpixel processing graph

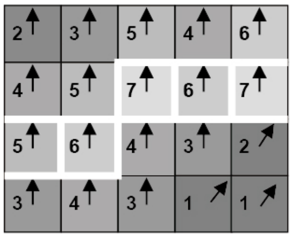

Fig. 3.3: Gradient scan image

(3) Compute image gradient. Employ the Sobel operator to compute the initial derivatives (inflection points) in both horizontal and vertical orientations for the smoothed image, subsequently determining the gradient magnitude and direction of the boundary based on the derived gradient map.

$$\begin{aligned} &Edge\_\ Gradient(G) - \sqrt{{G_x}^2 + {G_y}^2} \\ &Angle(0) = \tan \frac{G_x}{G_y} \end{aligned} \tag{3.4}$$

(4) Suppress non-maximum values. After obtaining the magnitude and direction of the gradient, scan the entire image, exclude points on the boundary, examine the first pixel point, and verify if it is indeed the point with the highest gradient in that same direction. Retain only the largest gradient value [10]. This process is illustrated in Figure 3.3.

(5) Threshold value confirmation. When the gradient of the image surpasses the maximum value, it is deemed as the authentic boundary. If it falls below the minimum value, it is disregarded. If it lies within these two values, its linkage to a confirmed true boundary point is verified. If connected, it is retained, otherwise, omitted. This process is exemplified in Figure 3.4.

(6) Calculate offsets and angles. Configure the edge position mode across multiple components and ascertain the X or Y coordinates of the specimen. The verification procedure is depicted in Figure 3.5.

3.1.3. Height measurement algorithm. When the laser pulse ranging method is in operation, the laser transmitting tube is initially oriented towards the target to emit a focused laser pulse. Upon reflection from the target surface, the laser scatters omnidirectionally. Subsequently, a portion of this scattered light returns to the sensor receiver and is efficiently captured by an advanced optical system before being projected onto an avalanche photodiode for further analysis. An avalanche photodiode is an optical sensor with an amplification function inside, so it detects extremely weak light signals and converts them into corresponding electrical signals.

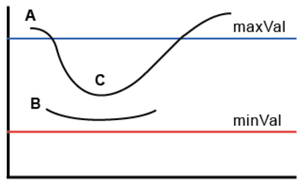

Fig. 3.4: Threshold confirmation image

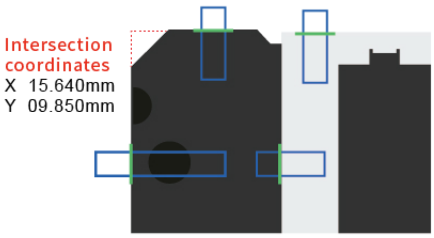

Fig. 3.5: Calculus diagram

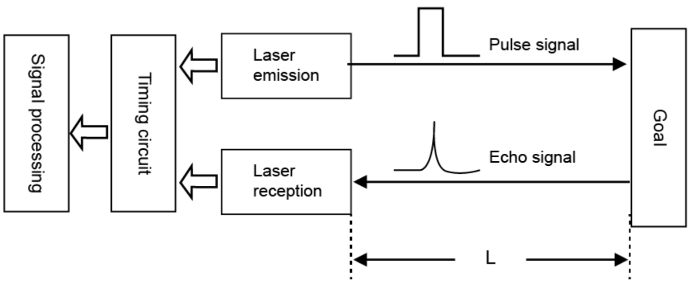

Fig. 3.6: Flow chart of laser pulse ranging

By measuring the time it takes the light pulse to travel to and from the point to be measured, multiplying by the speed of light and dividing by 2, the distance to the target to be measured is calculated by the following formula:

$$d = ct/2$$

where d: Measure the distance between points A and B, c: the speed of light, t: The duration of the light pulse's travel between points A and B once. This is shown in Figure 3.6.

After measurement and calculation, the distance d is obtained. This is shown in Figure 3.7.

When the specified distance is detected, the robot transitions into a precise positioning lock mode, decelerates gradually, and ultimately halts. The travel distance can be computed $d = \int(time * speed)$.This is shown in Figure 3.8.

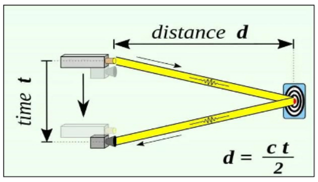

Fig. 3.7: Schematic diagram of laser pulse ranging method

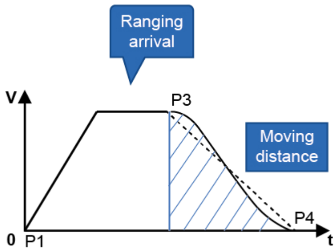

Fig. 3.8: Diagram of the z-direction localization algorithm

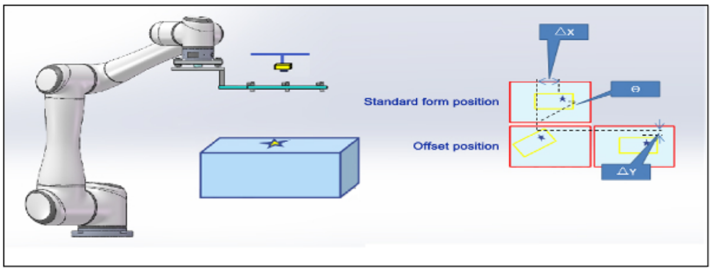

Fig. 3.9: Diagram of the robot correction algorithm

3.1.4. Robot correction algorithm. When invoices are separated, vouchers are printed, and multiple sheets are pasted, there may be a slight horizontal offset between the actual single ticket's position and the target coordinate. The manipulator needs to real-time correct the deviation $\Delta X, \Delta Y$, and offset Angleθ based on the difference between the actual bill's pose and target coordinate in order to ensure accurate positioning [11]. This is shown in Figure 3.9.

Computing center of rotation offset: compensation value Offset

$$X = \cos(\theta) * (x - \mathrm{rx0} - \sin(\theta) * (y - \mathrm{ry0}) + \mathrm{rx0} + \Delta x$$

$$\mathrm{Offset}Y = cos(\theta) * (Y - \ RY0) - sin(\theta) * (x - \mathrm{rx0}) + \ \mathrm{ry0} + \Delta Y$$

Finally, Offset$X$, Offset$Y$, Offsetθ is calculated.($\Delta X, \Delta Y$, are the horizontal deviation, rx0 and ry0 are the vertices of the manipulator, x and y are the coordinates of the current actual bill, and x0 and y0 are the coordinates of the target vertices.)

In this study, the manipulator effectively rectifies real-time deviation displacement and precisely aligns the target coordinates to ensure precise adherence of invoices and vouchers [12, 13, 14].

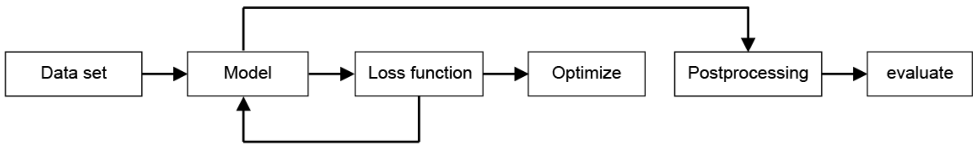

Fig. 3.10: Flowchart of text detection and recognition algorithms

3.2. Text recognition of bills. The text recognition of bills includes three parts: text detection, text recognition and information extraction. In the text detection part, the influence of different feature extraction networks on network performance improvement is analyzed based on the DBNet network. The core of DBNET is segmentation-based text detection, that is, each text block is semantically segmented, and then the segmentation probability map is simply binarized, and finally the detection results are converted into box or poly format. In the text recognition part, text recognition is studied using the network structure of CRNN(Convolutional Recurrent Neural Network) and CTC.For information extraction, a method based on text pattern and keyword matching is employed to obtain keyword key-value pairs [15, 16, 17].

The flow of text detection and recognition algorithms can typically be categorized into the following steps, as illustrated in Figure 3.10:

1) Data set. To gather a substantial amount of text data, which is then labeled to identify any undesirable content. his dataset serves is the foundation for algorithm training.

2) Model. Model is a commonly used feature extraction method. Each word or phrase within the text can be considered a feature. By extracting these features, this allows the transformation of the text into a digital format, which suitables for computer processing.

3) Loss function and Optimize. The extracted features and labeled datas are fed into a machine learning model, which undergoes training using a loss function. Throughout this process, adjustments can be made to enhance the algorithm's performance by modifying its parameters and structure.

4) Postprocessing and evaluate. After the training completed , the model is applied to real text detection tasks. When the new textual input, the algorithm will predict, assesses, and filter based on its learned model.

The text on the paper invoice presents the characteristics of dense and multi-scale. In the selection of the text detection network, the integrity of the detection text area should be taken into account as much as possible, and the adhesion between different texts should be avoided as far as possible. Therefore, this research chooses DBNet as the research object of this project. DBNet is a text detection network based on pixel segmentation [18]. It uses FPN to integrate multi-scale features, optimizes the training process through differentiable binarization to simplify post-processing, and reduces text sticking by shrinking text boxes. The DBNet network structure is shown in the following figure 3.11.

The DBNet text detection model can be divided into three parts. The first part is the backbone network, which has a feature extraction function. After inputting the image to be processed into the backbone network, multi-scale features are obtained through down-sampling operations. Section two, the FPN feature fusion structure is used to perform up-sampling and fusion operations from deep layers to shallow layers on these features. Concatenation (con-cat) is applied on the fused features to obtain feature maps. Both concatenation and addition require that the feature maps have the same resolution for processing [19]. The difference is that con-cat increases the number of channels by superimposing features while the single feature remains unchanged. Adding the corresponding elements of the feature, and the number of channels also remains unchanged. The third part of the network is responsible for obtaining the output content. After acquiring the feature graph, it further predicts the probability graph and threshold graph, and obtains an approximate binary graph through differentiable binarization. Finally, post-processing is performed to obtain the final text enveloping curve.

After the DBNet network detects the text on the ticket image, it is displayed as four-point mark boxes to

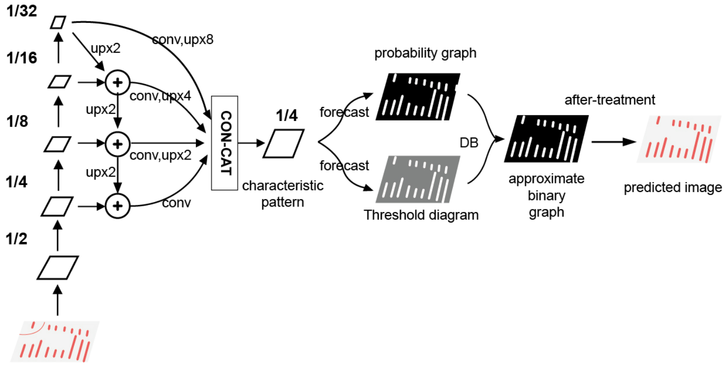

Fig. 3.11: DBNet network structure diagram

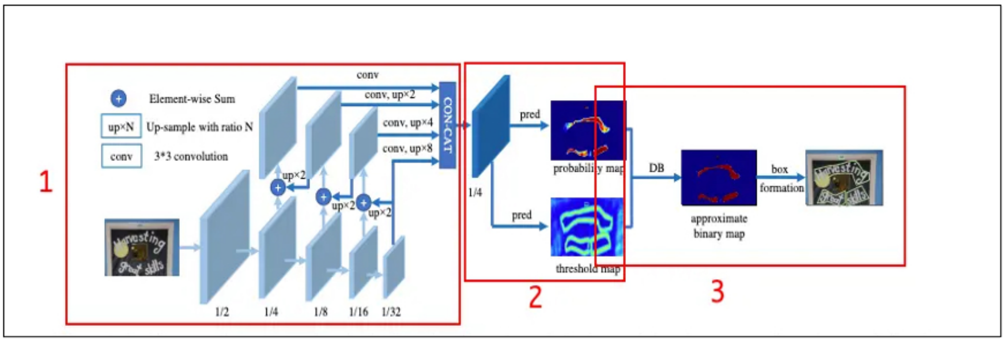

Fig. 3.12: The CRNN identifies the network logic graph

indicate its position. Then, the text content within these mark boxes is recognized. The CRNN recognition network used in this project consists of three main parts, as illustrated in Figure 3.12.

The first module: the red box on the left uses an FPN architecture consisting of a bottom-up convolution operation and a top-down upsampling operation to obtain multi-scale features. The lower part of the figure is a 3x3 convolution operation. According to the convolution formula, the feature maps of 1/2, 1/4, 1/8, 1/16, and 1/32 of the original image size are obtained respectively. Then, the upsampling x2 is performed from top to down, and then fused with the feature map of the same size generated from bottom to up. After fusion, 3x3 convolution was used to eliminate the aliasing effect of upsampling. Finally, the output of each layer is upsampled and unified into a feature map of 1/4 size.

The second module: the 1/4 size feature map is transformed through a series of convolution and transpose convolution mechanisms to the probability map P and threshold map T, which can refer to the FCN network structure . And the purpose is to generate the feature map P and T with the same size as the original map.

The third module: the DB method is used to obtain the approximate binary map through the feature map P and T.

For the recognition of the text content on the ticket, it is important to note that the extracted text lines from text detection may vary. CRNN, as the current mainstream text recognition network, transforms the problem of recognizing text into a sequence recognition problem. Its advantage lies in its ability to detect texts of any length. Additionally, CTC in transcription layer can reduce the cost of marking text content and fulfill the requirements for bill text recognition.

The text image size of the input CRNN network is variable. In this project, the constructed CRNN network requires maintaining the feature size of the input loop layer at (channel , height , width)=[512,1,24], so preprocessing of the image is necessary before performing convolution operations. Taking the ResNet series feature extraction network as an example, it requires carrying out double of five times down-sampling operations in the network. The feature map size after the last subsampling operation is 1/32 of the original image. Therefore, it is necessary for the image height to be fixed as a multiple of 32. The preprocessing strategy adopted in this project involves first calculating the aspect ratio of text recognition images in the dataset. According to the statistics, the average aspect ratio of images used for text recognition in the dataset is approximately 1/3. Subsequently, the image is resized to a height of 32, and any portion with a width less than 100 is padded with zeros. If the width exceeds 100, it is scaled down to 100. The image size remains fixed at [3,32,100] after preprocessing.The image was down-sampled by convolutional neural network, and the input image was changed from gray image (single channel) to color image (512 channels). The height is 1/32 of the original, so it's 1, which is important; Through this step, we get 24 feature quantities, so the width is 24. In summary, we finally get the feature map of size [512,1,24]. During prediction for a single instance, only the width needs to be adjusted to a multiple of 32. However, when predicting multiple invoice pictures in batches, each batch's longest picture sets the width standard for that particular batch.

The pre-processed images are fed into the convolutional network to generate the feature sequence $(X_1, X_2, ..., X_{t-1}, X_t)$, which is then passed through the loop layer. In RNN-based sequence data processing, both LSTM and BiLSTM structures can be used to model the information of the context. Longer distance dependencies are better captured using LSTM models, but cannot encode information from backward to forward. BILSTM can better capture bidirectional semantic dependencies in the case of finer classification granularity, such as the five classification tasks of strong positive, weak positive, neutral, weak negative and strong negative, that is, when attention needs to be paid to the interaction between sentiment words, degree words and negative words. However, the all-in-one machine is mainly used for the identification and processing of invoices, and the semantics and its emotional color not need be judged, so LSTM is chosen. In natural language processing tasks, a bidirectional language model is used, that is, two unidirectional LSTMS are considered to accomplish it.

After the recurrent layer and argmax operation processing, the largest of character category of each frame in the sequence is output as the result of this frame. Because the sequence characteristics of adjacent are similar, there are duplicate adjacent characters in the output string.For example, for the ticket text string 'CNY 18.6', possible output strings include 'CNY 188.6600', 'CNY 118...6600', 'CNY 1888.60', etc. At this time, the final recognition result cannot be obtained through the string with repeated characters. The role of CTC is to predict the sequence label with the highest probability combined with each frame of the sequence.

For invoice recognition tasks, after obtaining text content through text detection and recognition, it is necessary to further extract the corresponding relationship between keywords and text. This relationship can be expressed as key-value pairs through certain processing methods. Such tasks fall under Natural Language Processing (NLP) problems. The text on invoices usually appears in pairs, starting with a descriptive word followed by its corresponding specific value [20].

3.3. Tidal temporary storage, give full play to the ability of multithreading parallel work. This study proposes a method that utilizes tide temporary storage technology to achieve cache operation for large-volume paper invoice submission. Under the business scenario of submitting a large number of paper invoices, the double-thread parallel operation of document submission and ticket pasting can be realized, while applying the working principle of tide temporary storage to implement peak cutting and valley filling. The diagram illustrating tidal temporary storage technique is shown in Figure 3.13.

1. The tide points are automatically monitored and managed, with ticket collection at the front desk and background verification work being conducted simultaneously.
2. When the temporary storage capacity exceeds the upper limit, it will stop the ticket collection task, when the attached work of the background is completed and the working-storage section is below the tide point level, and resume the ticket collection work.
3. The batch operation is primarily associated with the task to prevent temporary gridlock. When the order certificate is not generated, switch to temporary priority and continue requesting certification.
4. After completing the information binding of the order, please check if the ticket data matches the order

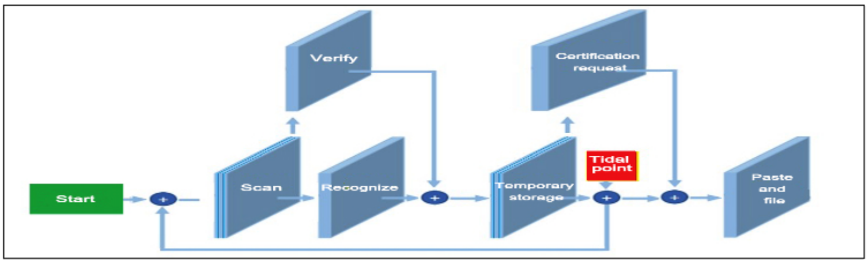

Fig. 3.13: Logic diagram of tidal temporary storage technology

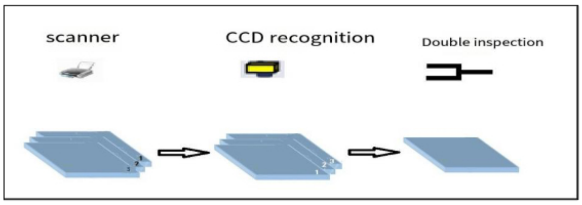

Fig. 3.14: Logic diagram of tidal temporary storage technology

information. Then, initiate the temporary storage task and concurrently submit a certificate request.

5. When the invoice is stored in the temporary storage process, it will switch to ticketing mode after receiving successful information about the previous temporary order certificate. The timing of this switch is controlled by the tidal point.
6. The dynamic distribution of tidal points is determined by the temporary occupancy level and time-weighted comprehensive evaluation, and it is adjusted based on previous statistical results.

3.4. The duplicate and adhesion of invoice detection. This study proposes a research method that utilizes AI intelligent detection to accurately identify and locate multiple invoices by detecting their adhesion. The design includes a double-sheet feed detection mechanism, which employs high-performance ultrasonic sensors to test the paper's weight (ranging from 27-413g/m2) and thickness (0.08mm-0.2mm). If multiple invoices are detected, the operation will be halted. Figure 3.14 illustrates the schematic diagram of detection of the duplicate and adhesion of invoice.

1. The application of antistatic and brushing techniques greatly reduces adhesion occurrences, thereby improving equipment stability.
2. The visual positioning CCD utilizes template matching, edge detection, and analysis of edge shapes to identify the presence of border adhesion. Once such adhesion is detected, an alarm is triggered and the operation is halted.
3. The first step in invoice entry is scanning. The primary function of the scanning operation is to separate a stack of invoices into individual ones. However, if there is adhesion between the invoices, it may cause a paper jam and trigger an alarm, leading to the suspension of further operations.
4. Positioning CCD as the function for correcting invoices, while also performing a secondary verification of invoice information. If the serial number on the invoice and the scanning position of the serial number are opposite, While CCD detection cannot scan the identification information in reverse order, an alarm will be triggered indicating either adhesive or missing tickets. In both cases, business operations will be suspended.
5. Place two detection sensors on the robot arm to detect when an invoice is removed. When a duplication phenomenon is detected, an alarm will be triggered and operations will be suspended.

The application of automatic unattended all-in-one machine solves the problem of manual operation in the

Table 4.1: Analysis of application results for an integrated robot capable of automatically handling bill receipts, stickers, and inspections without human intervention.

Verification item	result	unit
Visual localization	<0.5	mm
Recognition speed	<60	ms
Recognition accuracy	>99.93	%
Efficiency of automatic receipt and Posting of documents	2043	copies

Table 4.2: Analysis of application results for an integrated robot capable of automatically handling bill receipts, stickers, and inspections without human intervention.

Measurement position	Visual positioning deviation	Recognition speed	Recognition accuracy
1#	0.435mm	58ms	>99.93 %
2#	0.461mm	55ms	>99.93 %
3#	0.398mm	57ms	>99.93 %
4#	0.457mm	57ms	>99.93 %
5#	0.493mm	49ms	>99.93 %
6#	0.480mm	56ms	>99.93 %
7#	0.311mm	58ms	>99.93 %
8#	0.274mm	47ms	>99.93 %
9#	0.472mm	55ms	>99.93 %
10#	0.489mm	58ms	>99.93 %
11#	0.463mm	56ms	>99.93 %
12#	0.466mm	52ms	>99.93 %

financial process. The application scenarios of the automatic unattended invoice verification machine are as follows:

1. Supplier service platform: For the platform, whether it is delivering the supplier's invoice to the customer or paying the supplier through the invoice, it will receive a large number of invoices that need to be sorted, identified and archived. The application of all-in-one computer can reduce the four links of invoice review, entry, verification and filing, which greatly saves manpower and material resources;
2. Financial management of electric power enterprises: the realization of functions such as intelligent receipt of paper documents, bill verification, pasting into books and automatic archiving can help group enterprises optimize the process of invoice tax declaration, verification, archiving, inventory management and other operations, more efficient docking with ERP management system, and optimize enterprise resource management data;
3. Duplicate check of electronic invoices: due to the duplication of invoices, financial personnel must be extra careful when dealing with electronic invoices, for fear of repeated reimbursement. All-in-one machine can greatly solve this problem.

4. Results and discussion. Based on the AI vision of an automated unattended bill collection and inspection integrated robot system and control platform, we investigate the application of visual positioning guidance, bill text recognition, temporary storage during peak periods, re-tensioning, and adhesion detection in the process of bill collection and inspection. Through the conducted experiment, it was observed that the visual positioning deviation is below 0.5mm, the bill recognition speed is less than 60ms, and the recognition accuracy exceeds 99.93%. Additionally, the daily average efficiency of automatic receipt and pathing of documents surpasses 2000. As depicted in Table 4.1 and Table 4.2, successful implementation of intelligent automatic processing for invoice documents within the settlement hall has been achieved, leading to a significant enhancement in on-site work efficiency (refer to Figure 4.1).

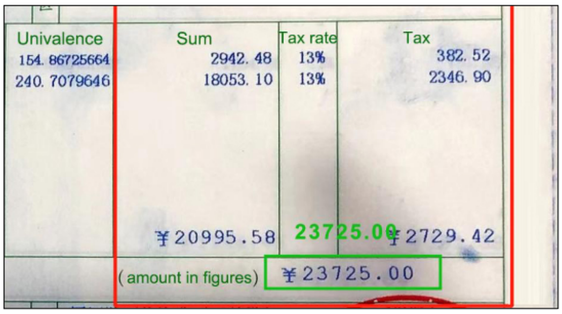

Fig. 4.1: Invoice information extraction results

1. Through the utilization of visual analysis technology, this study efficiently gathers multiple invoices with a single click, swiftly identifies and automatically matches them, resulting in an accuracy rate that is nearly 100%. The system employs advanced vision technology, integrating cutting-edge image processing techniques with state-of-the-art artificial intelligence algorithms and pattern recognition methodologies.
2. Collect multiple invoice images intelligently through CCD, combine AI visual analysis with self-evolution learning of big data, deliver accurate invoice data in the form of intelligent and manual double insurance, and achieve nearly 100% accuracy in the first review of invoices.
3. In this study, the DBNet model leverages ResNet18 as the feature extraction architecture for precise detection of text regions on the ticket. Moreover, to ensure accurate identification of textual content, the CRNN and CTC models employ MobileNetV3 as their feature extraction framework.
4. This study proposes a method that utilizes tidal temporary storage technology to achieve mass paper invoice delivery buffer operation. In the business scenario of mass paper invoice submission, the delivery operation and sticker operation are carried out in parallel as dual-thread processes. The order of document delivery and subsequent sticker application is executed simultaneously, employing the working principle of temporary storage to facilitate peak scene application.
5. This study proposes an AI-based research method for the intelligent detection of invoice adhesion, with the aim of accurately identifying and locating multiple positions where invoices are affixed. The proposed method incorporates a dual others-into-paper detection design, employing high-performance ultrasonic sensors capable of detecting paper weights ranging from 27-413 g/m2. Once more than one sheet of paper is found, the operation will be halted.

5. Conclusion. Through the utilization of an all-in-one machine for bill receipt, pasting, and verification, the functionalities of invoice verification and voucher management are effectively realized. This innovative solution offers the power company a sophisticated, user-friendly, and integrated platform for "verification and automatic certificate creation," thereby facilitating intelligent automated processing of invoice vouchers within the settlement hall. Create a fully automatic intelligent invoice voucher integrated machine to solve the operational problems of a large number of manual receipt, sticker of contract hall. It promoted the upgrading and transformation of the special work of supplier settlement, improved the satisfaction of the business environment, and truly promoted the construction of intelligent robots for invoice production from the theoretical system to the application and practice.

6. Limitations and further work. Notwithstanding, there exist certain limitations in this paper. Firstly, the ambit of bill recognition solely concentrates on the compilation, identification and authentication of specific invoices for power enterprise suppliers, however, it does not encompass receipts such as train tickets. Secondly, the AI vision-based automatic unattended ticket collection and inspection integrated robot equipment and control platform is currently primarily designed to cater to the specific business characteristics of the power industry. However, further adaptations are required to suit diverse business application scenarios such as in telecommunications and government enterprises. Future research can also be carried out from the following aspects:(1) Combining the latest industrial and information technology to further enhance the accuracy and efficiency of automating certificate making and ticket sticking. (2)With the aim of establishing a modern smart supply chain, exploring additional robots to replace complex and repetitive manual tasks, developing various intelligent and automated industrial robots, promoting customization for multiple scenarios in smart supply chains, applying automated robots, and enhancing the overall operational intelligence and automation level of supply chain scenes have become crucial steps for power enterprises to optimize their business environment.

In our second set of experiments, we took $c = d = 100$ and carried out trials analogous to those in the first set above. No preconditioning was used in these experiments, both because we wanted to compare the methods without preconditioning and because the fast Poisson preconditioning used in the first set of experiments is not cost effective for these large values of c and d. We first allowed each method to run for 600 iterations, starting with zero as the initial approximate solution, after which the limit of residual norm reduction had been reached.

REFERENCES

[1] Xiuzhi L, Lijian X, Xin G, et al., *Determination of beef tenderness based on airflow pressure combined with structural light three-dimensional vision technology*, Meat Science , 2023 , 202.

[2] Mingzhi L, JuCui W, Malik J, et al., *Research on nonlinear tracking and evaluation of sports 3D vision action*, Nonlinear Engineering , 2023 , 12(1).

[3] Kusumam K, Krajník T, Pearson S, et al., *3D-vision based detection, localization, and sizing of broccoli heads in the field*, Journal of Field Robotics , 2017 , 34(8).

[4] H. C B, Anders C,C.Ø. H L, et al., *Measuring the interior of in-use sewage pipes using 3D vision*, Automation in Construction,2023,151.

[5] Weichselbaum J, Zinner C, Gebauer O, et al., *Accurate 3D-vision-based obstacle detection for an autonomous train*, Computers in Industry , 2013 , 64(9).

[6] Zannatha I M J, Tamayo M J A, Sánchez G D Á, et al., *Development of a system based on 3D vision, interactive virtual environments, ergonometric signals and a humanoid for stroke rehabilitation*, Computer Methods and Programs in Biomedicine , 2013 , 112(2).

[7] Weizhen Dong, Yan Chen, Hailing Liang, *Improve DBNet and CRNN face logo don't method*, computer engineering and design , 2023 , 44(01):116-124 , The DOI: 10.16208/j.i ssn1000-7024 , 2023 , 01 , 016.

[8] Jose C M, Irazu D F H, Oyuki L P R, et al., *Language meets YOLOv8 for metric monocular SLAM*, Journal of Real-Time Image Processing , 2023 , 20(4).

[9] Xiuli D, Kai L, Yana L, et al., *Motor Imaging EEG Signal Recognition of ResNet18 Network Based on Deformable Convolution*, Electronics , 2022, 11(22).

[10] Yu W, Ibrayim M, Hamdulla A., *Scene Text Recognition Based on Improved CRNN*, Information , 2023 , 14(7).

[11] Krzysztof O, Piotr L., *A method supporting fault-tolerant optical text recognition from video sequences recorded with handheld cameras*, Engineering Applications of Artificial Intelligence , 2023 , 123(PB).

[12] Luis H, Ahmed K, Chen-Hua R Y., *A bioinspired modular soft robotic arm*, Engineering Research Express , 2023 , 5(1).

[13] Mateusz P, Marcin R, Robert P., *Prototype and Design of Six-Axis Robotic Manipulator*, Journal of Automation, Mobile Robotics and Intelligent Systems , 2023 , 16(1).

[14] Fenggang Zhu, *A Validation Study of a Deep Learning-Based Doping Drug Text Recognition System to Ensure Safe Drug Use among Athletes*, Machinery & Electronics , 2023 , 41(03):44-49.

[15] Sangyong L, Jae Hyeon P, Jiwun Y, et al., *Space Motion Trajectory Planning Method for a Six-Axis Manipulator with Joint Velocity Constraints*, Healthcare (Basel, Switzerland) , 2023 , 11(12).

[16] Enrique V, H. A T, Antonio R, et al., *End-to-End page-Level assessment of handwritten text recognition*, Pattern Recognition , 2023 , 142.

[17] Jianrong Cao, Fatong Cao, Ming Wang, et al., *Multi-objective semantic segmentation algorithm for underwater images based on YOLOv5 and FCN-DenseNet*, Computer system application, 2022 , 31(12):309-315 , The DOI: 10.15888 / j , carol carroll nki , Csa , 008850.

[18] Yijie Liu, *Invoice recognition based on deep learning research*, Xi 'an petroleum university, 2022. The DOI: 10.27400 , dcnki , Gxasc , 2022 , 000869.

[19] Rui Yang, Jie Yang, *Invoice identification and management system based on image analysis*,Journal of computer age, 2020(10):4 to 8 , The DOI: 10.16644 / j. carol carroll nki cn33-1094 / tp. 2020.10.002.

[20] Seghiri T, Ladaci S, Haddad S, *Fractional order adaptive MRAC controller design for high-accuracy position control of an industrial robot arm*, International Journal of Advanced Mechatronic Systems , 2023 , 10(1).

Edited by: Jingsha He
Special issue on: Efficient Scalable Computing based on IoT and Cloud Computing
Received: Dec 7, 2023
Accepted: Feb 2, 2024