IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 9 July 2024, accepted 25 July 2024, date of publication 29 July 2024, date of current version 7 August 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3435386

RESEARCH ARTICLE

Multi-Device Universal Automation Data Acquisition and Integration System

QINGHUA SONG[1], YAJUN LIU[2], HAOYUE SUN[2], YONG CHEN[1], AND ZHENG ZHOU[1]
[1]Information Center, Zhangjiakou Cigarette Factory Company Ltd., Zhangjiakou, Hebei 075000, China
[2]College of Information Engineering, Hebei University of Architecture, Zhangjiakou, Hebei 075000, China

Corresponding author: Yajun Liu (lyj2100@hebiace.edu.cn)

This work was supported in part by the Research on Data Acquisition and Integration Based on Deep Learning under Grant 2221008A, in part by the Research on Data Acquisition and Integration of Tobacco Material Inspection under Grant ZY012022E001, in part by the Non-Invasive Monitoring Research of Office Building Electrical Equipment Based on Machine Learning under Grant 2022CXTD09, in part by the Science Research Project of Hebei Education Department under Grant QN2024148, and in part by the Deep Learning Behavioral Recognition Fall Detection Research under Grant 2022CXTD04.

ABSTRACT With the advent of Industry 4.0 era, smart manufacturing is rapidly developing, which puts higher requirements on data acquisition (DAQ). Currently, enterprises face problems in data acquisition such as a variety of equipment types, equipment heterogeneity, customized development, high costs, and so on. To solve this problem, this paper proposes an innovative approach to address the challenges in data acquisition by combining You Only Look Once version 8 (YOLOv8) and Optical Character Recognition (OCR) technology. First, the data area to be collected is labeled, and then the YOLOv8 model is trained. The model with the shortest recognition time and highest efficiency is selected with comparable detection effects to fully guarantee real-time data collection. Second, to ensure the accuracy of data collection, the OCR model is trained a second time using the dataset, improving performance compared to the most effective PaddleOCR by 8 percentage points. Additionally, considering the aging and weak computational capacity of individual devices, this study adopts a client/server architecture, deploying the computational load required for recognition to the server to achieve generalization, stability, and reliable operation of the system. Lastly, the scheme is tested on 16 types of 30 devices.

INDEX TERMS Data acquisition, OCR, YOLOv8, Industry 4.0.

I. INTRODUCTION

In the era of big data, data is the new gold [1], which profoundly affects people's life, work, and thoughts. Various countries, enterprises, and individuals have realized the important value of data [2], [3]. However, due to the low-value characteristics of the data, data collection is often still mainly manual, even with the time-consuming and laborious development of data collection systems, these systems are often designed for a specific type of equipment, greatly reducing their general value [4], [5], [6].

As shown in Figure 1, DAQ is the process of gathering, analyzing, and recording information concerning specific phenomena. It is typically the initial stage in data analysis and application. Voltage, current, temperature, pressure, sound, and other types of data are mainly obtained conventionally. However, with the development of the Industry 4.0 [7] era and smart manufacturing, there is a growing requirement for data acquisition.

Traditionally, both hardware and software can be used to examine data gathering.

A. HARDWARE-BASED DAQ

Using hardware like sensors for data acquisition, it is possible to acquire common physical signals like the voltage, current, and temperature. This hardware typically has commercial data acquisition software that is simple to use, less expensive, and more dependable [8]. While it is occasionally essential to acquire signals that are not supported by DAQ hardware, this necessitates custom development,

The associate editor coordinating the review of this manuscript and approving it for publication was Abdel-Hamid Soliman.

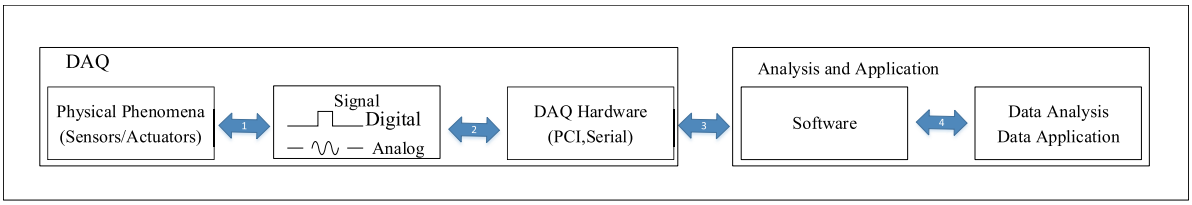


FIGURE 1. Data acquisition and application system diagram.

which is expensive [6], [9] and requires reliability testing [9], [10].

B. SOFTWARE-BASED DAQ

In the context of data acquisition, there are two primary software-based methods: web crawling and Object Linking and Embedding (OLE) for Process Control Unified Architecture (OPC UA). Web crawling techniques typically extract data from web pages [11] using mature frameworks such as Scrapy and Selenium to obtain web page source code by simulating a browser request and parsing it for storage. This method suits software developed with browser/server architecture but faces challenges in data collection for software developed under client/server architecture. To address the difficulties of data acquisition and development of supervisory systems due to the variety of plant equipment and protocols, OPC UA can be utilized. For instance, products from Siemens, ABB, and others already support the OPC protocol. However, some devices do not support the OPC protocol, and even for those that do, it is necessary to understand the driver interface standard, data request parameters, parsing format, etc., which constitutes a significant workload and complicates data acquisition [12], [13].

The OLE for Process Control (OPC) plays a pivotal role in the underlying communication protocols and data exchange mechanisms in industrial automation systems. The OPC standard is divided into OPC Classic and OPC Unified Architecture (UA) versions. OPC Classic is based on Microsoft's DCOM technology (Distributed Component Object Model), mainly for data exchange within local networks. However, due to limitations in DCOM's security and network traversal capabilities, its application is restricted across networks or the internet. In contrast, OPC UA, as a cross-platform data exchange standard, overcomes the limitations of OPC Classic in terms of security and network communication. It introduces more flexible data modeling capabilities, supporting complex data types and structures, thereby enhancing the efficiency and flexibility of data exchange in industrial automation systems. However, this approach still requires the analysis of data protocols such as Message Queuing Telemetry Transport (MQTT) and Advanced Message Queuing Protocol (AMQP).

Both methods require customized development, which cannot be done universally for different devices, websites, instruments, etc. [14], [15]. Moreover, the cost associated with customized development are a significant overhead for companies, which are of concern to researchers [7], [9], [14], [16].

C. CONTRIBUTION OF THIS ARTICLE

Through subsections I-A and I-B, it can be found that traditional hardware and software have difficulties in conducting data collection. Due to differences in equipment procurement batches, manufacturers, development models, data communication interfaces, and protocols, etc., significant difficulties arise in data collection. Even if OCR is used for text recognition when extracting the experimental results data from the testing equipment, OCR text recognition is the full picture, and the text detection and text recognition modules need to spend a lot of time on recognition calculations, while bringing great difficulties to the later data screening, as shown in Figure 2. For example, as shown in Figure 3, a trademark paper sample for testing is shown. When the data is generated, there are three data areas that need to be collected, delineated and marked with three red rectangular boxes. If the whole picture is identified, it is not only time-consuming but also difficult to process the data at a later stage.

Due to the difficulties encountered by companies in data collection, this paper makes the following contributions:this paper proposes a novel data acquisition framework that ignores the problems of heterogeneity of underlying devices, software data encryption, no interface to the software, and different development modes, and truly achieves simple deployment, ease of use, and high versatility. Under the guidance of this framework, first, a labeled dataset was created for training the YOLOv8 model. This dataset mainly focuses on labeling the data that needs to be collected from the detection result page. Secondly, screen recording, camera capture and screenshots are utilized to obtain information from the inspection results. The YOLOv8 model is then applied to perform object detection and select the regions of data that need to be collected. Third, the detected areas are passed to the trained OCR model for recognition and the recognition results are saved. Finally, real-time integration with the company's big data platform is realized through an interface. With the above methods, visible and required data can be collected.

II. RELATED WORK

Data acquisition has a lengthy history [17] and has been used in countless enterprise sectors [8]. It also provides theoretical training to some industries. Data acquisition is the initial step of industrial applications [7], [18], data

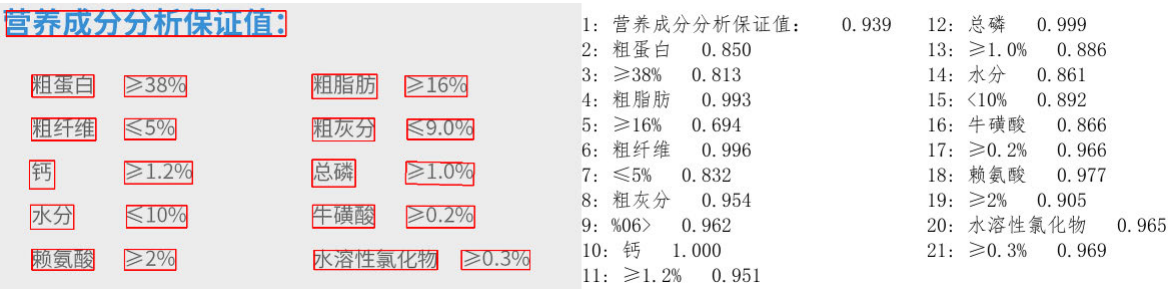


FIGURE 2. OCR text detection and recognition.

FIGURE 3. Trademark paper testing data.

analysis [18], [19], smart manufacturing [20], and other technologies. The duration and accuracy of the acquisition are crucial for the subsequent work. Tasić and Hajro [21] blended sensors with a high-speed data acquisition system to record and analyze real-time modifications that occur during the welding process. These modifications were cautiously analyzed using several signal processing, statistical, and data mining methods to provide valid records for welding result evaluation. Zhu et al. [22] proposed a joint RecurDyn-EDEM-AMESim simulation for describing the working load of a bulldozer underneath linear pushing conditions and proposed a simulation-based strategy to consider the working load of development machinery. Ng et al. [23] introduced a novel approach through the CARE (Content-Aware Resource Efficiency) network series and monitoring system, CARENet, which facilitates the computational aggregation and remote monitoring of patient-specific pulmonary conditions and airflow parameters. The collected data are stored in community-connected storage (Network-Attached Storage, NAS) and analyzed in real-time using web-based software. CARENet's network architecture is intricately designed around three core subsystems: Data Acquisition (DAQ), Data Storage, and Server along with a Data Management Platform (DMP). The DAQ subsystem directly captures Ventilation Waveform Data (VWD) from mechanical ventilators, transmits this data via Wi-Fi to the servers housed in the Data Storage and DMP subsystems, and connects to the wider internet through a 4G router. This setup grants CARENet the flexibility to adapt to changes in network conditions, ensuring uninterrupted data collection and real-time analysis. The CARE network employs a modular and scalable architecture within its system design to accommodate fluctuations in network conditions, user demands, and content attributes, thereby optimizing resource allocation and enhancing the Quality of Experience (QoE). CARENet dynamically adjusts its data transmission and processing strategies by continuously monitoring network status and user interactions. For instance, in the event of detected decreases in network quality, the system can autonomously modify the frequency of data synchronization

or opt for more stable network paths to guarantee the accurate transmission and timely analysis of critical medical data. Additionally, the DMP of CARENet can be tailored to meet the specific needs of physicians and nursing staff, adjusting data display and analysis functionalities to provide the most pertinent information, thus optimizing user experience and supporting decision-making in treatment processes. Sun et al. [24] proposed a multi-peak data acquisition and evaluation machine known as INSMA for gathering data generated by various scientific gadgets in contemporary intensive care units (ICU) to enhance patient care and support doctors in decision-making. Zhang et al. [25] proposed a seismic statistics acquisition. Huang et al. [26] designed a seismic records acquisition gadget primarily based on wi-fi community transmission to enhance the low-frequency response and low sensitivity of current acquisition systems. Samourkasidis et al. [27] designed and applied an environmental statistics acquisition module (EDAM) to facilitate time collection records acquisition and integration, and the generalization of the technique was once established in seven cases.

The above literature reveals that data collection is used in various industries, but each industry needs to develop its own methods, which not only raises the cost of data collection but also limits generalizability.

YOLO is an acronym for You Only Look Once, and its core idea is to transform the target detection task into an end-to-end problem, which has achieved impressive results in target detection [28], [29], [30]. YOLOv8 is one of the YOLO series of models, and YOLOv8 has several significant advantages over traditional target detection methods. Firstly, in terms of speed and efficiency, YOLOv8 adopts a lightweight network structure, which makes it capable of real-time detection while maintaining high accuracy. Secondly, in terms of accuracy and performance, YOLOv8 uses a lightweight network structure while introducing techniques such as adaptive training, multi-degree inference, and data augmentation to further improve the performance of the model. In addition, YOLOv8 is very versatile and flexible. The model can be used for multi-target detection, such as people, cars, plants, vegetables, etc. It is used in multiple scenarios, such as automated driving, traffic flow monitoring, smart security, etc. In addition, YOLOv8 also supports real-time target detection and can track continuous video streams to complete target tracking and recognition.

OCR emerged in the late 1920s and is now developing into one of the traditional challenges in the field of vision research. OCR techniques are often used in document recognition [31], [32], [33], license ticket recognition [34], [35], license plate recognition [36], [37], [38], natural scene text recognition [39], [40], and other fields. With the development of deep learning technology in the field of vision and the improvement of hardware computing power, OCR technology has broken through the bottleneck of traditional technical frameworks and has become a current research hotspot that is widely used. In particular, the neural network architecture proposed by Shi et al. [41], [42] integrates feature extraction, sequence modeling, and transcription into a unified framework, which can achieve end-to-end training, handle arbitrary character length, and perform better in scenarios with and without lexicon text. Moreover, the generated model is small, which is more convenient to apply in practice. The model was tested on IIIT-5K, Street View text, and ICDAR datasets, and the experiments proved the superiority of the algorithm over existing techniques. Since then the research on OCR has focused on text recognition tasks for arbitrary shapes in natural scenes [43], [44].

The references indicate that data acquisition is widely used across various industries, but each industry requires customized development based on its specific characteristics. For example, Kumar et al. recorded and analyzed real-time changes in the welding process using data acquisition devices, while Ng et al. proposed the CARENet system, which optimizes resource allocation and enhances user experience and data processing capabilities through a modular and scalable architecture. Although these studies have achieved significant results in specific applications, their customized nature leads to high costs and limited generalizability.

Building on existing data acquisition technologies, a highly versatile and cost-effective data acquisition solution is proposed. The core of this solution utilizes YOLOv8 and OCR technologies. By employing the YOLOv8 model for detecting data acquisition areas, its advantages in speed, efficiency, and accuracy enable real-time data area detection, providing a solid foundation for data acquisition. Additionally, the application of OCR technology in document and license plate recognition supports data area identification. By combining these two technologies, a new data acquisition system has been developed that allows for what-you-see-is-what-you-get data acquisition. This solution not only offers high accuracy and real-time performance but also enables broad data acquisition in various industrial applications, enhancing the system's versatility and applicability.

In summary, based on existing data acquisition systems, innovative data acquisition methods and processing techniques have been developed to address the high cost and customization issues faced by enterprises. Our developed data acquisition system, with its wide applicability and low cost, effectively solves the challenges enterprises encounter in data acquisition.

III. METHODOLOGY AND DESIGN

This system is developed to address the common data capture challenges faced by enterprises. The system can capture data from the original equipment using screenshots, video recording, cameras, and other methods. The YOLOv8 model is used to detect the data areas that need to be captured, which are then passed to the trained OCR model to complete the detection and recognition tasks. The recognition results can be saved to a file, a database, or integrated with the

enterprise's big data platform for real-time data collection. The realization framework is shown in Figure 4.

A. RESEARCH METHOD OR EXPERIMENTAL DESIGN

The core of this paper involves using YOLOv8 and OCR technology to complete the data acquisition. The specific data acquisition process is shown in Figure 5.

As shown in Figure 5, to complete data acquisition using YOLOv8 and OCR, the original data region that needs to be acquired must first be labeled. Then, the YOLOv8 model is trained to automatically identify this data region. Subsequently, the pictures, videos, and data captured by the camera can be passed to the model to automatically recognize the captured region. Once the area is recognized, OCR can be used to perform text recognition and save the results. Since data collection involves numerous numbers, units, and Chinese characters, the OCR model is retrained to ensure its accuracy meets enterprise-level requirements. If the recognized data needs to be docked with the enterprise's big data platform, it can be analyzed through the interface of the enterprise's big data platform and then the data can be uploaded in real time.

B. DATA COLLECTION

Figure 5 shows that the core part of data acquisition involves dataset construction and labeling, YOLOv8 model training, and OCR detection and recognition. The following is an example of the detection results of trademark paper to demonstrate the entire system using the data acquisition system.

Before training the YOLOv8 model, it is necessary to construct a training set and a validation set. In this study, trademark paper equipment was used to construct the dataset and label it using the labeling tool LabelImg. Figure 6 shows the data acquisition and labeling in the case of maximizing the interface on the computer.

In Figure 6 Area1 and Area2 are the data to be captured, which are labeled using the LabelImg tool, while the data are labeled in a variety of cases because the device may exist not maximized on the display, and the detection results may appear in more than one area. This can be seen in Figure 7.

The above labeling is for test results that can be displayed on the computer desktop. In this situation, screenshots, screen recording, cameras, and other methods can be used to obtain the data. However, the test results of some instruments and equipment are directly displayed on the terminal. In this case, a camera can be used to take a picture of the display. Then, select the area to be collected and label it. Finally, use the camera to complete the data collection.

IV. YOLOV8 MODEL TRAINING AND SELECTION

A. DATASET

Data collection for 16 types of equipment was completed using the approach described in subsection III-B. These devices mainly include tensile testing machines, compression testing machines, high-performance liquid chromatography (HPLC) systems, thickness gauges, and more. Approximately 400 images were collected for each device. To provide a more intuitive display of the results, one of the devices was taken as an example. A total of 400 images were labeled for this device, of which 300 were used to train the model and 100 were used to test the model. The model training process was visualized using wandb, as shown in Figure 8. To mitigate the challenges posed to data collection by fluctuations in image quality, obstructions, and environmental conditions, images were annotated from multiple perspectives and data preprocessing techniques such as contrast adjustment, noise reduction, and image sharpening were employed. These measures significantly alleviate the impact of low lighting, blurring, and other quality issues.

B. YOLOV8 MODEL TRAINING AND EVALUATION

1) EVALUATION METRICS

$$\text{Precision} = \frac{TP}{TP + FP} \tag{1}$$

In formula 1, TP represents the number of true positives, which are the instances correctly predicted as positive. FP represents the number of false positives, which are the instances incorrectly predicted as positive.

$$\text{Recall} = \frac{TP}{TP + FN} \tag{2}$$

In formula 2, TP represents the number of true positives, which are the instances correctly predicted as positive. FN represents the number of false negatives, which are the instances incorrectly predicted as negative.

$$\text{mAP}_{0.5} = \frac{1}{N} \sum_{i=1}^{N} \text{AP}^{i}_{0.5} \tag{3}$$

In formula 3, N is the total number of categories, and $\text{AP}^{i}_{0.5}$ is the average precision for the i-th category at an IoU threshold of 0.5.

$$\text{mAP}_{0.5:0.95} = \frac{1}{N} \sum_{i=1}^{N} \left(\frac{1}{10} \sum_{j=1}^{10} \text{AP}^{i}_{0.5+0.05(j-1)} \right) \tag{4}$$

In formula 4, N is the total number of categories. $\text{AP}^{i}_{0.5+0.05(j-1)}$ is the average precision for the i-th category at different IoU thresholds ranging from 0.5 to 0.95, with increments of 0.05.

$$\text{GFLOPs} = \frac{\text{Total Floating Point Operations}}{10^9} \tag{5}$$

In formula 5, the Total Floating Point Operations represent the total number of floating-point operations performed in a single forward pass of the model.

$$\text{Speed (ms)} = \frac{\text{Total Inference Time (ms)}}{\text{Number of Inferences}} \tag{6}$$

In formula 6, the Total Inference Time (ms) is the total time taken for all inferences, measured in milliseconds.

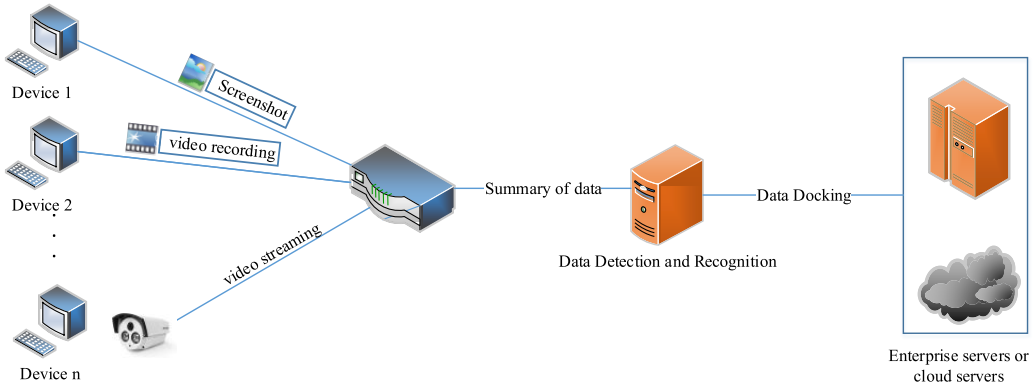


FIGURE 4. Data acquisition system.

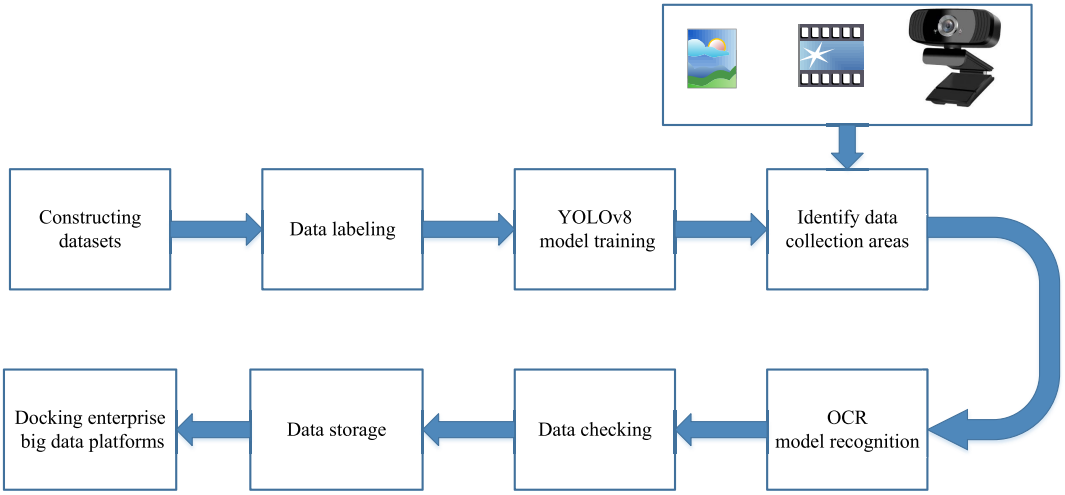


FIGURE 5. Data acquisition process.

The Number of Inferences is the total number of inference operations performed.

$$\text{Parameters} = \sum_{i=1}^{L} \text{Params}_i \tag{7}$$

In formula 7, L represents the number of layers in the model. Params_i denotes the number of parameters in the i-th layer.

C. YOLOV8 MODEL TRAINING AND EVALUATION

To accurately identify data collection areas and efficiently complete data collection, precision, recall, mAP_0.5, and mAP_0.5:0.95 were comprehensively considered when selecting a model. To improve efficiency and achieve effective data collection, computational complexity (GFLOPs), inference speed (Speed, ms), and model parameters (Parameters) were chosen as evaluation metrics for model performance.

In constructing Figure 8, the large number of models caused some areas to overlap, making it difficult to clearly observe the models' performance. Therefore, the training process of the models was smoothed to obtain Figure 8. As seen in Figure 8, after 100 iterations, all models tended to stabilize, and among the four evaluation metrics, the YOLOv8n model performed the best overall, providing a reliable foundation for data area recognition. It should be noted that we set the iteration count for YOLOv8 and YOLOv10 models to 200, but due to the patience parameter, which terminates model training early when performance improvement stalls, some models did not reach 200 iterations.

To better evaluate the models' performance, we visualized the parameters, speed, and other aspects of the trained models, resulting in Figure 9. As shown in subplot a of Figure 9, YOLOv8n has the lowest GFLOPs at 8.195 (similar to YOLOv10n), indicating that YOLOv8n has the lowest computational complexity. This makes YOLOv8n the best choice, especially in resource-constrained scenarios. Subplot b of Figure 9 reveals that YOLOv8n has the fastest inference speed at only 1.491 ms, with YOLOv8s ranking second at 3.022 ms, which is twice that of YOLOv8n. Compared to

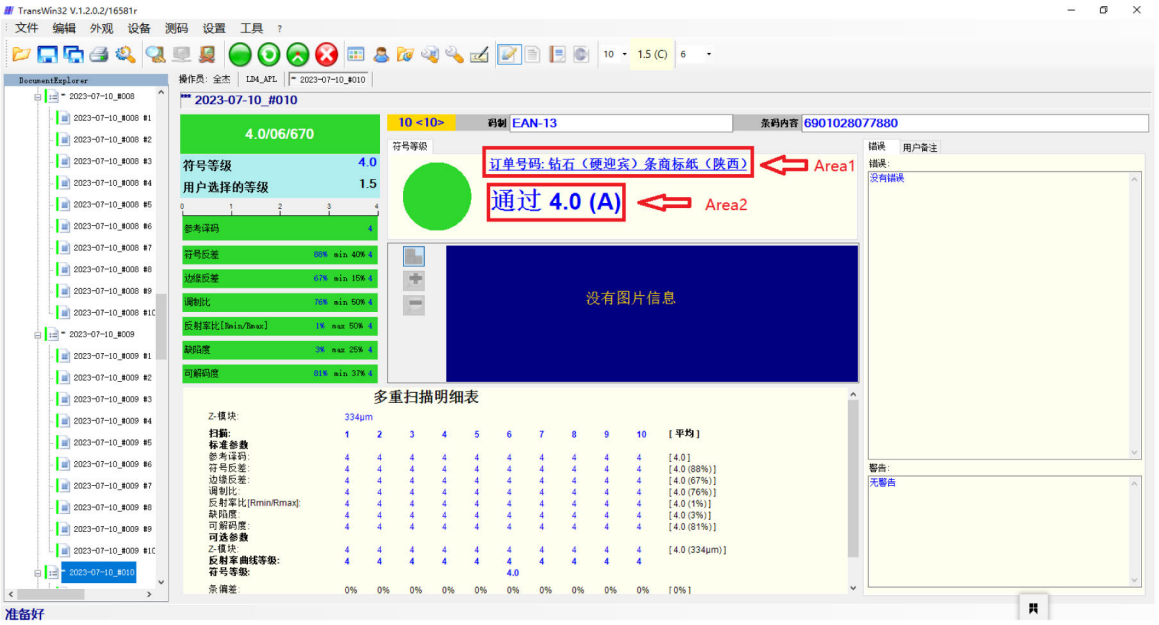


FIGURE 6. Software desktop maximized labeling.

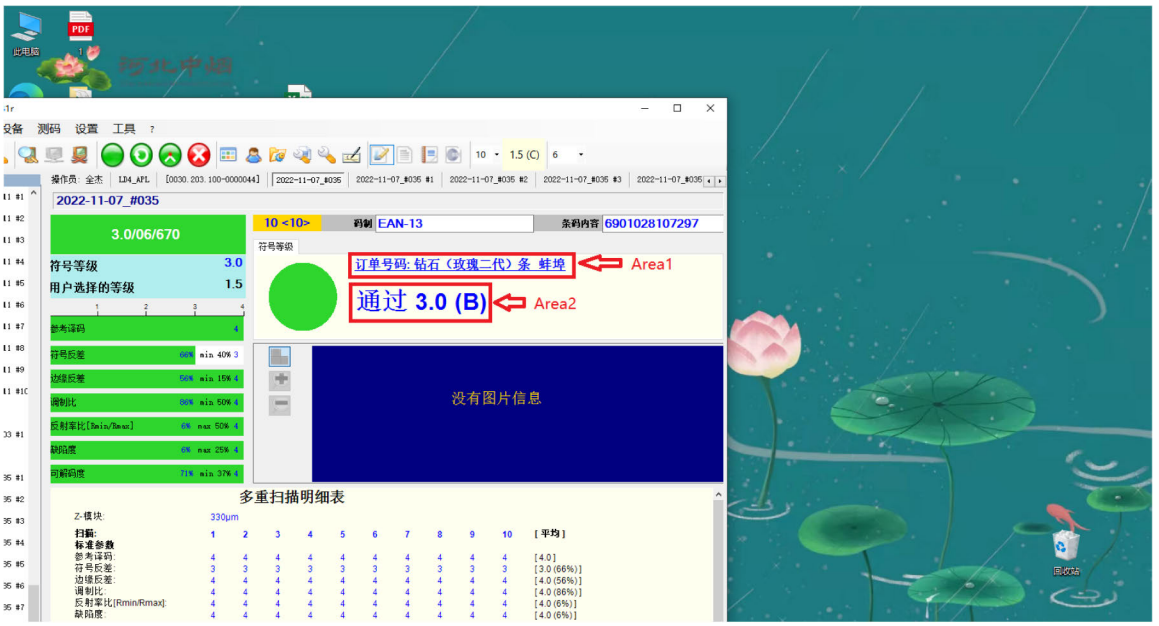


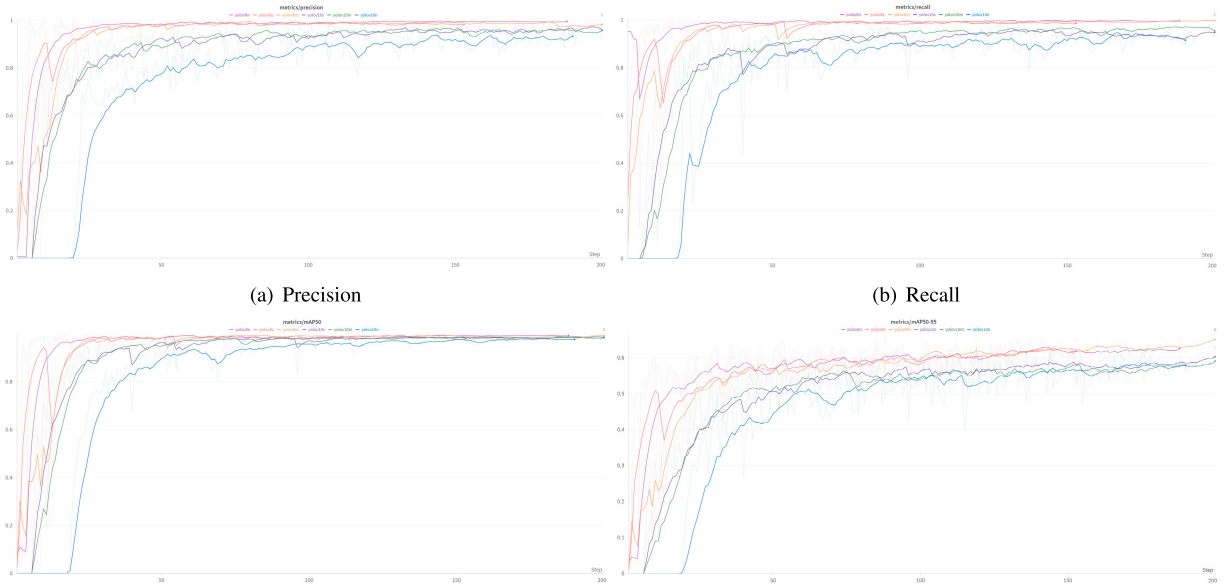
FIGURE 7. Software desktop non-maximized labeling.

other models, YOLOv8n shows a significant advantage in speed. Subplot c of Figure 9 shows that the YOLOv10n model has the fewest parameters, followed by YOLOv8n, with a difference of 303,418 parameters between them.

Combining the metrics of precision, recall, mAP_0.5, mAP_0.5:0.95, GFLOPs, inference speed, and model parameters, it was found that although YOLOv8n has more parameters than YOLOv10n, its overall performance is superior. Therefore, to achieve efficient, fast, and real-time data collection in resource-constrained environments, the YOLOv8n model was chosen for data collection area recognition.

The experiments were conducted on a workstation equipped with an NVIDIA RTX 4090 D GPU and 64GB of RAM. The training time for the YOLOv8n model was 5.26 hours, while testing and validation each took 1.01 hours.

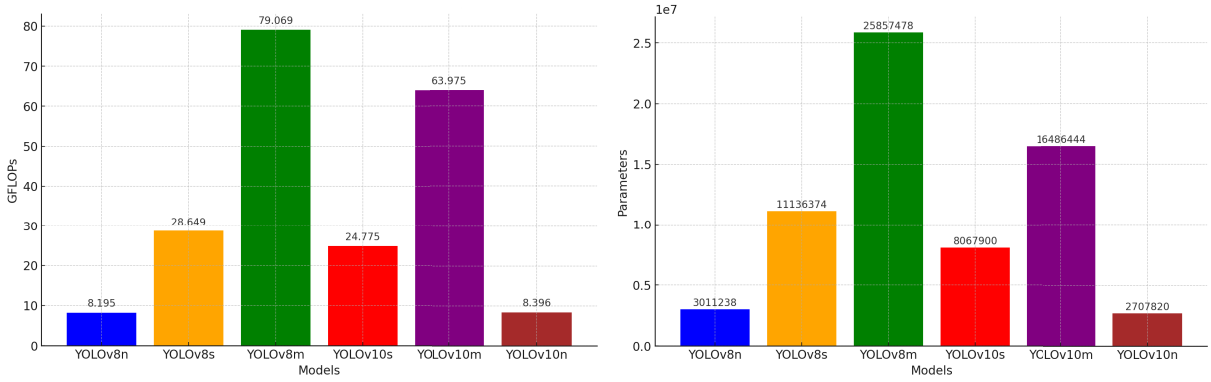
The validation set was tested using the YOLOv8 model and the results are shown in Figure 10. Through the seven



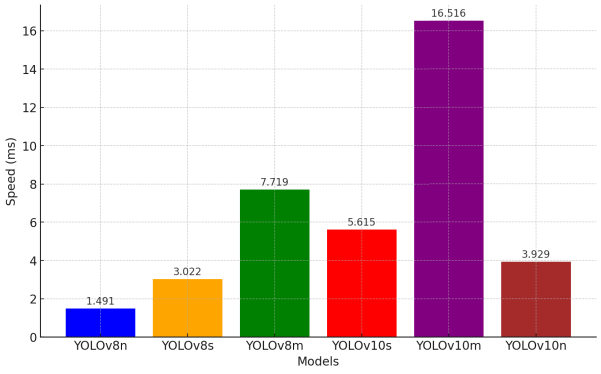
(a) Precision (b) Recall

(c) mAP_0.5 (d) mAP_0.5:0.95

FIGURE 8. Evaluation of model detection performance.



(a) GFLOPs Comparison of YOLO Models (b) Parameters Comparison of YOLO Models



(c) Speed(ms) Comparison of YOLO Models(PyTorch)

FIGURE 9. Evaluation of model speed and complexity.

indicators of checking precision, checking recall, mAP_0.5, mAP_0.5:0.95, computational complexity (GFLOPs), inference speed (Speed, ms), and model parameters (Parameters), it can be found through Figure 8, Figure 9, and Figure 9 that the detection of the current region of the data can be achieved with a small amount of labeling. Combined with

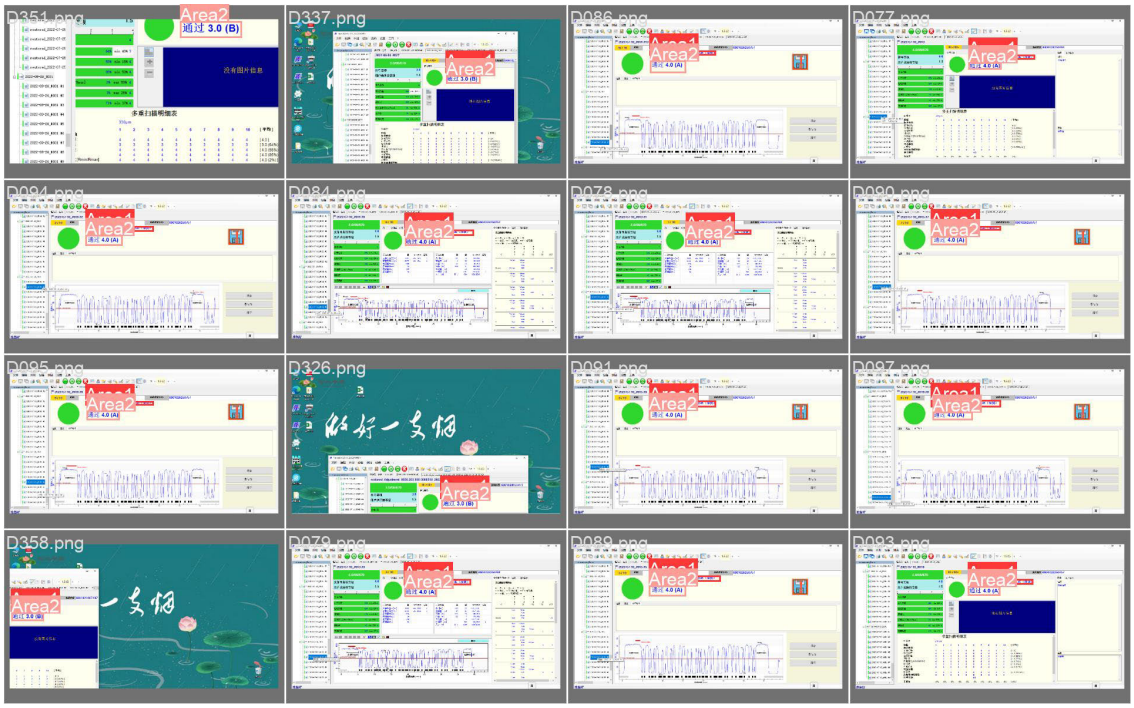


FIGURE 10. Experimental results on the YOLOv8n model validation set.

Figure 8, it can be observed that despite the YOLOv8n model having fewer parameters and lower complexity, its performance is not inferior to the more complex YOLOv8s, YOLOv8m, and YOLOv10 series models, even with fewer iterations. However, it is notable that the YOLOv10n model does not perform well with fewer iterations. As the number of iterations increases, particularly after 100 iterations, the overall performance of the models becomes comparable. Considering the requirement of real-time data collection, the YOLOv8n model, which has fewer parameters and layers, is used as the data region detection model in this study.

Figure 10 shows that the YOLOv8n model can detect Area1 (red area) and Area2 (pink area), which are the data areas to be captured in many cases. Although there is an overlap in the labeling of the two areas, it does not affect the recognition, which lays the foundation for automatic data collection.

V. OCR MODEL TRAINING AND SELECTION

A. ANALYSIS OF TEXT RENDERER'S ALGORITHMS

In order to train an OCR model that meets the requirements of data collection, we use the Text Renderer algorithm. Text Renderer employs a robust set of algorithms and techniques designed to efficiently convert text strings into high-quality graphical representations. This process is crucial for generating diverse datasets needed for training accurate OCR systems. The core components of Text Renderer involve a variety of text rendering techniques, including font variation, text layout adjustment, style simulation, and the application of text effects such as shadows and borders. These techniques collectively ensure the production of rich and varied text images that can simulate real-world text appearance under different conditions.

The choice of rendering engine in Text Renderer significantly affects the rendering quality, performance, and overall compatibility with different operating systems and rendering environments. Text Renderer supports several engines like Pillow, OpenCV, and bespoke rendering algorithms. Pillow is renowned for its comprehensive image processing features, making it ideal for applications requiring intricate text effects. Conversely, OpenCV is preferred for its processing efficiency, particularly beneficial for generating images on a large scale. The selection between these engines depends on the specific requirements of the rendering task at hand, balancing between quality and performance needs.

To optimize rendering outcomes, Text Renderer integrates advanced preprocessing techniques before rendering. These include dynamic contrast adjustments for better visibility under varying lighting conditions, noise reduction to enhance clarity, and geometric transformations for correcting text orientation and ensuring uniformity in text size across different images. Additionally, the system is continuously updated to improve its adaptability to different operating systems, enhancing its usability across a broader range of applications.

In conclusion, the integration of sophisticated algorithms and the strategic choice of rendering engines in Text Renderer contribute significantly to its capability to produce high-quality text images. This flexibility allows for the generation of tailored datasets essential for the development of OCR models that are both accurate and efficient across diverse real-world scenarios.

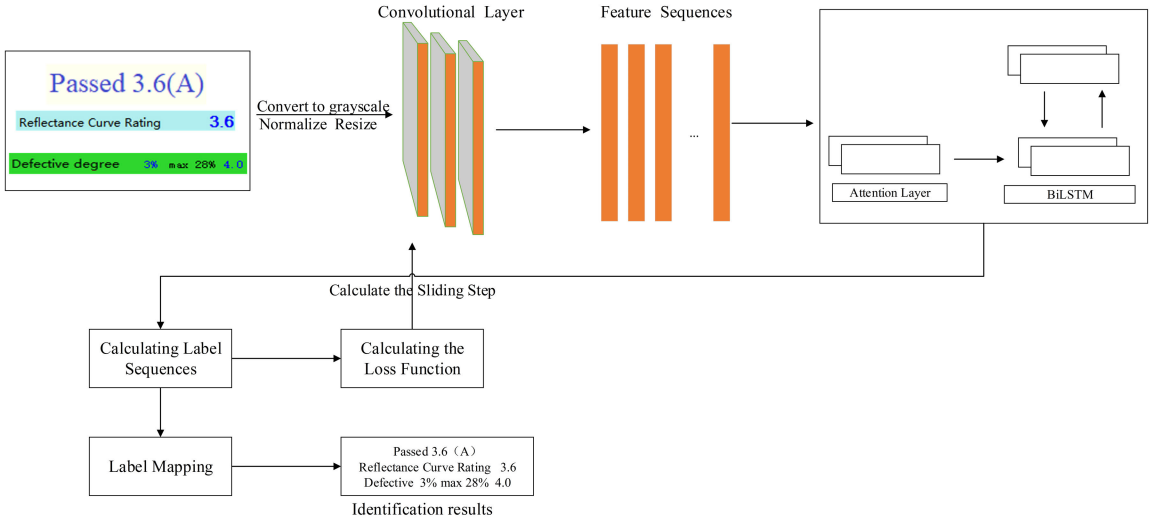


FIGURE 11. OCR model training and recognition.

TABLE 1. Comparison of text recognition models.

Models	Accuracy	Time	Model Size
Cncor	76.90%	706.67 s	2.47 GB
Easyocr	67.98%	2079.04 s	2.46 GB
PaddleOCR	87.30%	1229.29 s	2.87 GB
Pytesseract	36.73%	885.66 s	172 MB
Trained Model	95.70%	1032.76 s	2.15 GB

B. DATASET

Since the contents recognized in this study mainly include Chinese characters, numerical values, English letters and units, numerical features are particularly important. Therefore, the dataset used in the OCR model of this paper consists of two parts. The first part is a dataset of 500,000 news texts with Chinese tones generated by the Text Renderer; the second part is a dataset of 500,000 data containing letters, numbers and units generated by a Text Renderer. In the model training, the batch size is set to 64. Since the data generated by the Text Renderer is generated by serial number, i.e., there are numbers in front of the text in the dataset, and there are also numbers after the text, shuffling operation should be performed to train a better model. The acquisition of the Text Renderer and dataset are given in the data and code availability section of the paper. In order to enhance the robustness of the OCR model, we apply similar operations to the image data here as in subsection IV-A.

C. MODEL CONSTRUCTION AND TRAINING

The OCR model selected in the paper is mainly based on the PP-OCRv2 model built by Du et al. [42]. To speed up the training of the model, the images are converted to grayscale images before training and input to the network.

The core network structure is shown in Figure 11.

Figure 11 shows the structure of the text recognition network algorithm and the core architecture of text recognition in this paper, where the core work includes convolutional layers, feature sequence extraction, bidirectional LSTM neural network, and label transcription mapping and output of end-to-end text recognition results.

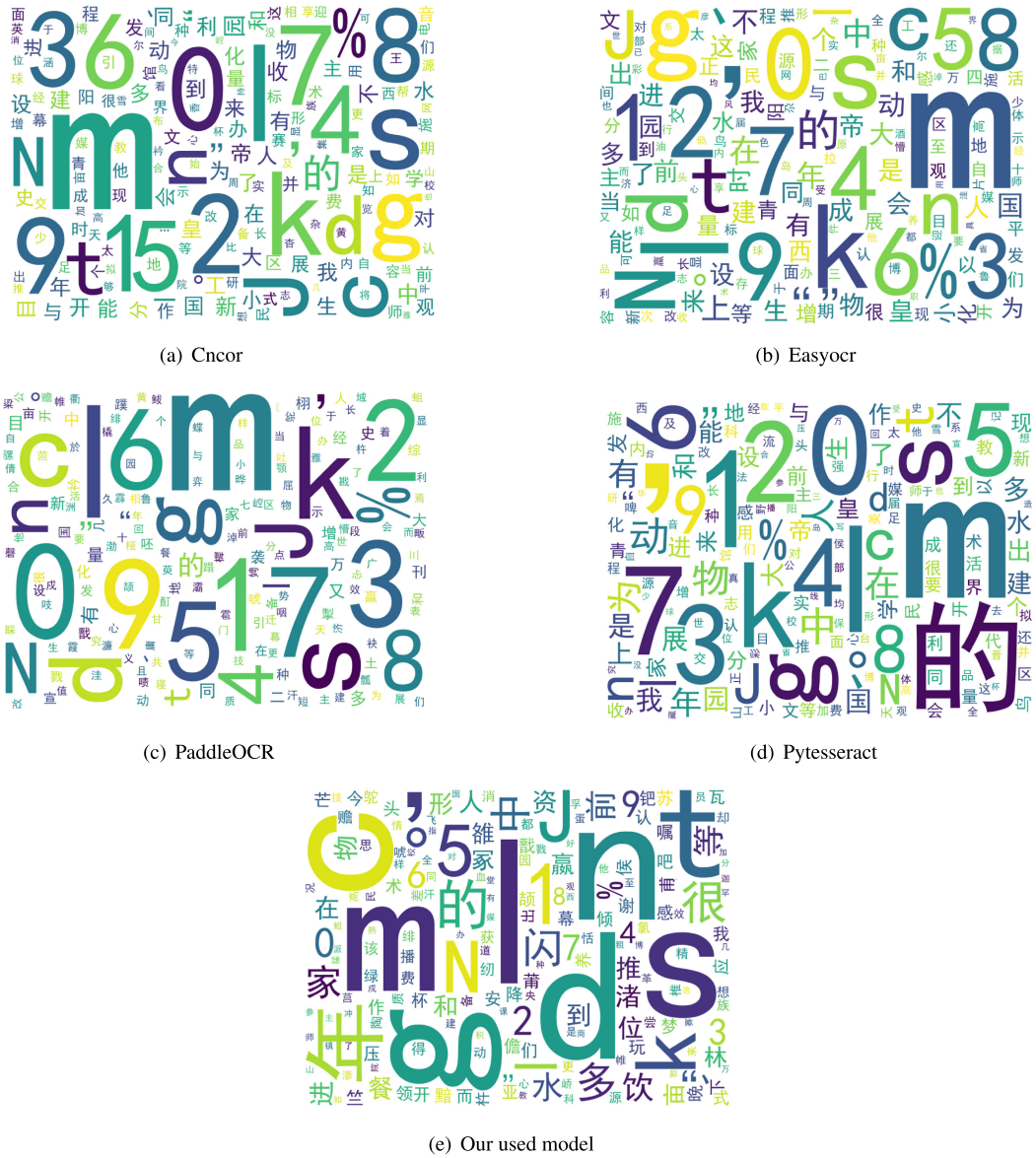
The training process of the OCR model used in this paper based on PP-OCRv2 model training is as follows: (1) train the dataset in subsection V-B using the OCR model; (2) save the model with the best training result after several iterations; (3) load the trained model into the data acquisition system to complete the recognition task.

After the above operation, the core data acquisition and integration tasks have been completed, but to help the enterprise break the data communication barriers, this system analyzes the enterprise's existing big data platform protocols and connects the local data with the remote server through the API interface provided by the enterprise to provide the theoretical basis for data analysis and data decision making.

D. MODEL CONSTRUCTION AND EVALUATION

There are already many OCR open-source frameworks that can be called directly to complete text recognition tasks. To select the optimal model, the mainstream OCR recognition frameworks Pytesseract, Easyocr, Cncor, and PaddleOCR are compared with the model proposed in this paper from multiple dimensions. To prevent single photos, which cause errors in the experiments, the test dataset consists of 500 text data and 500 photos of numbers, units, etc., and the five models were compared in terms of accuracy, time and model size. The experimental results are shown in Table 1.

From Table 1, it can be found that the model trained in this paper has the advantages of the highest recognition accuracy, shorter recognition time, and smaller model size (The model size here is the software size for software distribution using the pyinstaller package.). On the test set of 1000 images, the model has the highest recognition accuracy with smaller size compared to EasyOCR, Cncor and PaddleOCR models, which improved by 8 percentage points compared to PaddleOCR and shortened by 196.53 seconds compared to PaddleOCR.If the detection data similar to



(a) Cncor

(b) Easyocr

(c) PaddleOCR

(d) Pytesseract

(e) Our used model

FIGURE 12. Model recognition error word cloud map.

Figure 3 is processed, it is believed that the time used will be greatly reduced compared to the whole image being recognized.

During testing, it was found that the Cncor model was the fastest in recognition and the Pytesseract model was the smallest. However, both models have a low recognition accuracy, so they cannot be used in industry for practical applications.

Further analysis of the recognition results shows that the accuracy of the model is above 90% for both Chinese and English alone, but when the document is a mixture of English, Chinese, numbers, and units, the accuracy of the generic OCR model decreases because the Chinese characters are very different from the Latin type characters, and the error rate is higher, especially for unit recognition such as u and μ. In terms of data collection, numeric values and units are very important information, and the model used in this paper is trained on a processed dataset, so the results are better than the generic OCR model.

Additionally, it is important to note that due to the diversity of languages, extracting data in different languages requires retraining the OCR model. This not only ensures high accuracy but also helps in keeping the model smaller and reducing hardware resource consumption.

The words that Pytesseract, Easyocr, Cncor, PaddleOCR, and our used model failed to correctly recognize on the test set are transformed into word cloud plots to achieve Figure 12. The erroneously identified words from the Cncor, Easyocr, PaddleOCR, Pytesseract, and Our used model are shown by the a, b, c, d, and e plots, respectively. the text's word cloud graphic. The first four models are shown to have a higher mistake rate in identifying numbers and units, which may be attributed to two factors: 1. low model recognition accuracy and 2. a higher frequency of numbers and units in the data set.

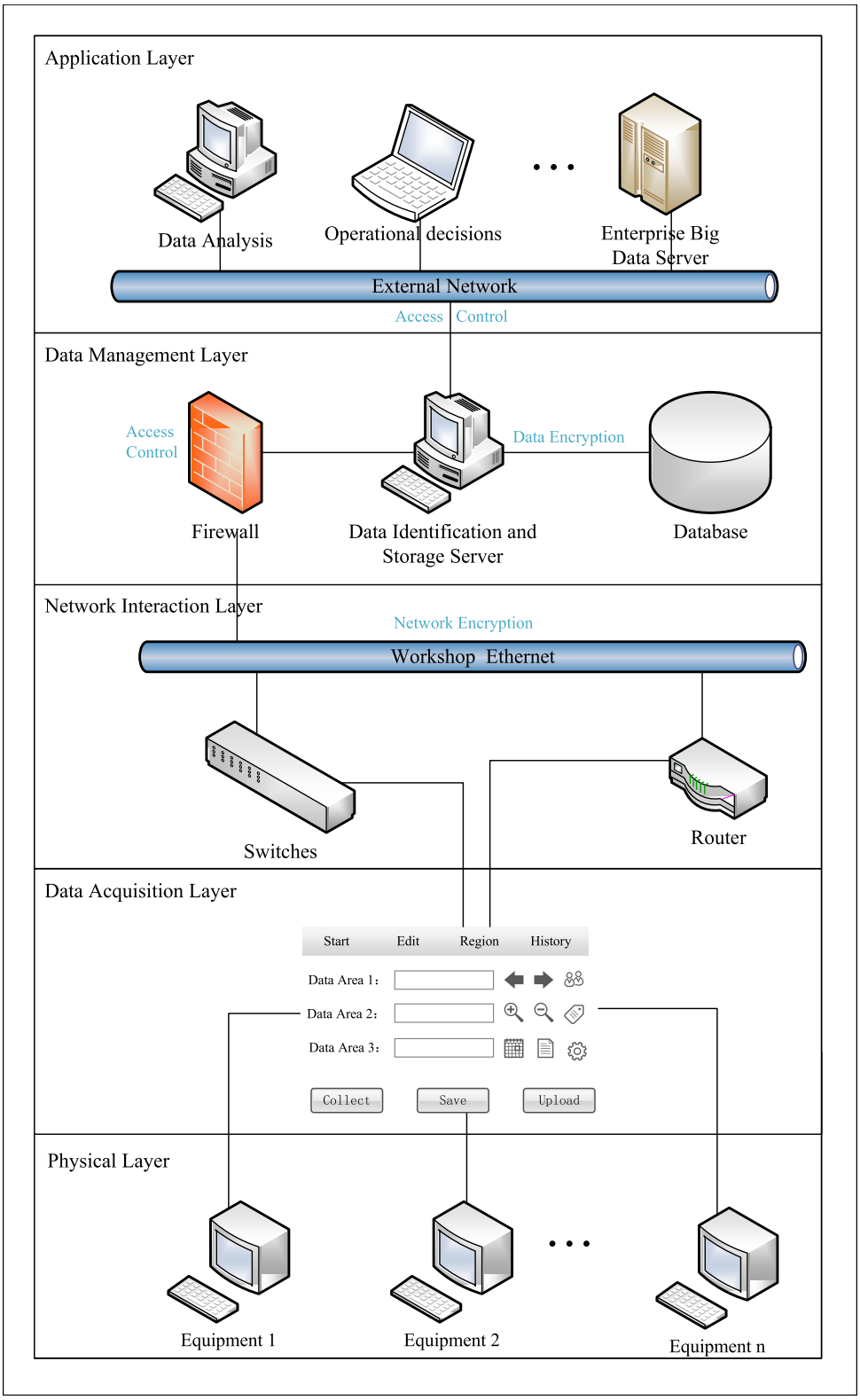


FIGURE 13. Data acquisition system deployment.

The model used in engineering is depicted in the fifth picture. The error rate of numbers and other things was decreased since the model employed a numerical dataset for improved training.

VI. SYSTEM DEPLOYMENT AND TESTING

A. SYSTEM DEPLOYMENT

The company currently uses 16 types of test equipment, totaling 30 units. All test devices are networked using a

FIGURE 14. Data region detection and identification.

TP-LINK 48-port switch that is configured for network communication. The servers are deployed in the same network as the test devices. The client sends the data of the data experiment results to the server in the form of pictures or videos, and the server completes the identification and saves it, and if necessary, it can be directly accessed to the enterprise's big data cloud platform.

The testing and deployment framework is shown in Figure 13. The architecture is mainly divided into five layers, from bottom to top: the physical layer, data acquisition layer, network interaction layer, data management layer, and application layer. The devices in the physical layer are multi-source devices, and the models, manufacturers, communication protocols, and development modes can be different. Through the use of image processing and OCR technology, the differences in the underlying hardware are resolved so as to realize data acquisition and integration of multiple devices and provide a theoretical basis for data analysis, computing decisions, etc.

Given the sensitivity of industrial data, several measures have been taken to ensure data security, as illustrated in Figure 13 with blue annotations indicating the locations of data access controls and encryption measures. Firstly, all data transmitted between devices and the central server is encrypted using the industry-standard AES-256 encryption algorithm. This ensures that even if the data is intercepted, it remains unreadable without the appropriate decryption key.

Secondly, the data acquisition system has implemented strict access controls to protect the data. Only authorized personnel can access the system, and access is granted based on the principle of least privilege. This means that users can only access the data and system functionalities necessary for their roles.

Additionally, continuous intrusion detection systems and regular security audits are in place to identify and mitigate vulnerabilities. In the event of a security breach, an incident response plan is established to quickly address and resolve the issue.

Through these measures, data security is maximized, and industrial data is protected from potential threats.

By using the YOLOv8n model in section IV to detect the data area that wants to be collected, and then combined with the OCR model trained in section V to complete the data collection, of which Figure 14 is the result of a certain experiment, in the process of collection of the results of the experiment at random dragging, it can be found that YOLOv8 can be tracked in real time to collect the data area, and at the same time, the recognition results are displayed and saved to the local Result.txt file.

Currently, the data acquisition system has been tested on 30 devices. However, with industrial development, the number of devices could reach hundreds or even thousands. This significant increase in computing load can be managed by optimizing algorithms and leveraging distributed computing resources to handle the increased data processing requests. Additionally, as the number of devices grows, network bandwidth and latency may affect the system's real-time performance. Utilizing data compression and other techniques can alleviate these issues. Furthermore, to enhance system robustness during scaling, introducing redundancy and failover mechanisms can improve fault tolerance. This ensures continuous operation even if individual devices fail.

VII. CONCLUSION

This research aims to provide users with an efficient and versatile data collection system. The system collects raw data

by taking screenshots and videotaping, detects valuable data areas using the trained YOLOv8 model, and then recognizes and saves them using the optimized OCR model; at the same time, in order to ensure the security of the enterprise data, the OCR model is deployed to a local server for local deployment and operation. This data collection method ignores the heterogeneity of the underlying equipment, data encryption, different development modes, etc. There is no need to capture packets, analyze data protocols, etc., which can truly realize the generalized collection of data, even if the individual device displays the data in the terminal. However, since the OCR model cannot achieve 100% recognition accuracy, it is still necessary to train the model for individual tasks to meet the enterprise data collection needs. It is believed that with the improvement of hardware computing power and OCR recognition accuracy, the system can solve the data collection problems of many enterprises in the near future.

In the future, the functionality and performance of the data acquisition system will be further enhanced. Firstly, the integration of additional types of data, such as audio signals, will be explored to expand the system's scope. Secondly, algorithms will be further optimized to achieve more efficient and accurate data acquisition. Additionally, the system architecture will be extended to integrate with cloud computing, allowing for the handling and analysis of larger-scale data. These changes will not only enhance the system's functionality and performance but also broaden its applicability in various industrial sectors, providing more effective data acquisition solutions for a wider range of enterprises.

VIII. DATA AND CODE AVAILABILITY

You can get the data and code availability through this link:https://gitee.com/jian123654/data_acquation_master.

The dataset is available through this link https://pan.baidu.com/s/1MRBcX3xWl-HuKvK616g8tg. The extraction code is kj02.

Text Renderer can be accessed through this link https://github.com/oh-my-ocr/text_renderer.

Wandb can be accessed through this link:https://wandb.ai/

REFERENCES

[1] V. Mayer-Schönberger and K. Cukier, *Big Data: A Revolution That Will Transform How We Live, Work, and Think*. Boston, MA, USA: Houghton Mifflin Harcourt, 2013.

[2] Y. Liu, J. Fan, J. Zhang, X. Yin, and Z. Song, "Research on telecom customer churn prediction based on ensemble learning," *J. Intell. Inf. Syst.*, vol. 60, no. 3, pp. 759–775, Jun. 2023.

[3] Q. Zhang, J. Lu, and Y. Jin, "Artificial intelligence in recommender systems," *Complex Intell. Syst.*, vol. 7, pp. 439–457, Feb. 2021.

[4] R. Bork, J. Hanks, D. Barker, J. Betzwieser, J. Rollins, K. Thorne, and E. von Reis, "advligorts: The advanced LIGO real-time digital control and data acquisition system," *SoftwareX*, vol. 13, Jan. 2021, Art. no. 100619.

[5] G. Urbikain and L. N. L. De Lacalle, "MoniThor: A complete monitoring tool for machining data acquisition based on FPGA programming," *SoftwareX*, vol. 11, Jan. 2020, Art. no. 100387.

[6] B. Mobaraki, S. Komarizadehasl, F. J. C. Pascual, J. A. Lozano-Galant, and R. P. Soriano, "A novel data acquisition system for obtaining thermal parameters of building envelopes," *Buildings*, vol. 12, no. 5, p. 670, May 2022.

[7] R. P. Pontarolli, J. A. Bigheti, F. O. Domingues, L. B. R. de Sá, and E. P. Godoy, "Distributed I/O as a service: A data acquisition solution to Industry 4.0," *HardwareX*, vol. 12, Oct. 2022, Art. no. e00355.

[8] X. Yang, P. Moore, and S. K. Chong, "Intelligent products: From lifecycle data acquisition to enabling product-related services," *Comput. Ind.*, vol. 60, no. 3, pp. 184–194, Apr. 2009.

[9] J. A. Basson, A. Broekman, and S. W. Jacobsz, "TD-DAQ: A low-cost data acquisition system monitoring the unsaturated pore pressure regime in tailings dams," *HardwareX*, vol. 10, Oct. 2021, Art. no. e00221.

[10] A. Broekman, S. W. Jacobsz, H. Louw, E. Kearsley, T. Gaspar, and T. S. D. S. Burke, "Fly-by-Pi: Open source closed-loop control for geotechnical centrifuge testing applications," *HardwareX*, vol. 8, Oct. 2020, Art. no. e00151.

[11] C. Wang and P. Wei, "A novel web page text information extraction method," in *Proc. IEEE 3rd Inf. Technol., Netw., Electron. Autom. Control Conf. (ITNEC)*, Mar. 2019, pp. 2213–2218.

[12] S. Khalil and M. Fakir, "RCrawler: An R package for parallel web crawling and scraping," *SoftwareX*, vol. 6, pp. 98–106, Jan. 2017.

[13] B. Goldschmidt, D. Schug, C. W. Lerche, A. Salomon, P. Gebhardt, B. Weissler, J. Wehner, P. M. Dueppenbecker, F. Kiessling, and V. Schulz, "Software-based real-time acquisition and processing of PET detector raw data," *IEEE Trans. Biomed. Eng.*, vol. 63, no. 2, pp. 316–327, Feb. 2016.

[14] T. Nieva and A. Wegmann, "A conceptual model for remote data acquisition systems," *Comput. Ind.*, vol. 47, no. 2, pp. 215–237, Feb. 2002.

[15] T. Kos, T. Kosar, and M. Mernik, "Development of data acquisition systems by using a domain-specific modeling language," *Comput. Ind.*, vol. 63, no. 3, pp. 181–192, Apr. 2012.

[16] M. Haizad, R. Ibrahim, A. Adnan, T. D. Chung, and S. M. Hassan, "Development of low-cost real-time data acquisition system for process automation and control," in *Proc. 2nd IEEE Int. Symp. Robot. Manuf. Autom. (ROMA)*, Sep. 2016, pp. 1–5.

[17] H. Poon, "Applications of data acquisition systems," *Comput. Ind.*, vol. 13, no. 1, pp. 49–59, 1989.

[18] A. Corradi, G. Di Modica, L. Foschini, L. Patera, and M. Solimando, "SIRDAM4.0: A support infrastructure for reliable data acquisition and management in Industry 4.0," *IEEE Trans. Emerg. Topics Comput.*, vol. 10, no. 3, pp. 1605–1620, Jul. 2022.

[19] A. R. Lupini, A. Y. Borisevich, and J. C. Idrobo, "Acquisition and fast analysis of multi-dimensional STEM data," *Microsc. Microanalysis*, vol. 23, no. S1, pp. 168–169, Jul. 2017.

[20] L. Wang, "From intelligence science to intelligent manufacturing," *Engineering*, vol. 5, no. 4, pp. 615–618, Aug. 2019.

[21] P. Tasić and I. Hajro, "Influence of heat input on geometry of GMAW fillet welds of unalloyed steel," *Zavarivanje I Zavarene Konstrukcije*, vol. 66, no. 4, pp. 161–167, 2021.

[22] X. Zhu, L. Pan, Z. Sun, Y. Wan, Y. Huang, and J.-H. Choi, "Simulation tool for dozer data acquisition," *Autom. Construct.*, vol. 142, Oct. 2022, Art. no. 104522.

[23] Q. A. Ng, Y. S. Chiew, X. Wang, C. P. Tan, M. B. M. Nor, N. S. Damanhuri, and J. G. Chase, "Network data acquisition and monitoring system for intensive care mechanical ventilation treatment," *IEEE Access*, vol. 9, pp. 91859–91873, 2021.

[24] Y. Sun, F. Guo, F. Kaffashi, F. J. Jacono, M. DeGeorgia, and K. A. Loparo, "INSMA: An integrated system for multimodal data acquisition and analysis in the intensive care unit," *J. Biomed. Informat.*, vol. 106, Jun. 2020, Art. no. 103434.

[25] X. Zhang, S. Zhang, J. Lin, F. Sun, X. Zhu, Y. Yang, X. Tong, and H. Yang, "An efficient seismic data acquisition based on compressed sensing architecture with generative adversarial networks," *IEEE Access*, vol. 7, pp. 105948–105961, 2019.

[26] Y. Huang, J. Song, W. Mo, K. Dong, X. Wu, J. Peng, and F. Jin, "A seismic data acquisition system based on wireless network transmission," *Sensors*, vol. 21, no. 13, Jun. 2021, Art. no. 4308.

[27] A. Samourkasidis, E. Papoutsoglou, and I. N. Athanasiadis, "A template framework for environmental timeseries data acquisition," *Environ. Model. Softw.*, vol. 117, pp. 237–249, Jul. 2019.

[28] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2016, pp. 779–788.

[29] P. Alves-Oliveira, S. Gomes, A. Chandak, P. Arriaga, G. Hoffman, and A. Paiva, "Software architecture for YOLO, a creativity-stimulating robot," *SoftwareX*, vol. 11, Jan. 2020, Art. no. 100461.

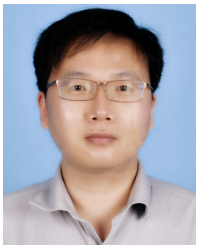
[30] Y. Guo and M. Zhang, "Blood cell detection method based on improved YOLOv5," *IEEE Access*, vol. 11, pp. 67987–67995, 2023.
[31] M. Liao, Z. Wan, C. Yao, K. Chen, and X. Bai, "Real-time scene text detection with differentiable binarization," in *Proc. AAAI Conf. Artif. Intell.*, vol. 34, no. 7, 2020, pp. 11474–11481.
[32] A. Chaudhury, P. S. Mukherjee, S. Das, C. Biswas, and U. Bhattacharya, "A deep OCR for degraded Bangla documents," *ACM Trans. Asian Low-Resource Lang. Inf. Process.*, vol. 21, no. 5, pp. 1–20, Sep. 2022.
[33] J. Martínek, L. Lenc, and P. Král, "Building an efficient OCR system for historical documents with little training data," *Neural Comput. Appl.*, vol. 32, no. 23, pp. 17209–17227, Dec. 2020.
[34] H. Zhang, B. Dong, Q. Zheng, B. Feng, B. Xu, and H. Wu, "All-content text recognition method for financial ticket images," *Multimedia Tools Appl.*, vol. 81, no. 20, pp. 28327–28346, Aug. 2022.
[35] B. Gatos, D. Danatsas, I. Pratikakis, and S. J. Perantonis, "Automatic table detection in document images," in *Proc. Int. Conf. Pattern Recognit. Image Anal.*, 2005, pp. 609–618.
[36] Salma, M. Saeed, R. ur Rahim, M. Gufran Khan, A. Zulfiqar, and M. T. Bhatti, "Development of ANPR framework for Pakistani vehicle number plates using object detection and OCR," *Complexity*, vol. 2021, pp. 1–14, Oct. 2021.
[37] V. Gnanaprakash, N. Kanthimathi, and N. Saranya, "Automatic number plate recognition using deep learning," *IOP Conf. Ser., Mater. Sci. Eng.*, vol. 1084, no. 1, 2021, Art. no. 012027.
[38] A. Mulyanto, E. Susanti, F. Rossi, W. Wajiran, and R. I. Borman, "Penerapan convolutional neural network (CNN) pada pengenalan aksara Lampung berbasis optical character recognition (OCR)," *J. Edukasi Penelitian Informatika*, vol. 7, no. 1, p. 52, Apr. 2021.
[39] R. Kapoor, M. Sushama, B. R. Andem, and A. S. P. S. Sindhura, "Natural scene text to voice signal conversion for visually impaired using deep neural network," in *Proc. Innov. Power Adv. Comput. Technol. (i-PACT)*, 2021, pp. 1–5.
[40] V. L. Cu, X. V. Truong, T. D. Luu, and H. V. Nguyen, "Region awareness for identifying and extracting text in the natural scene," in *Proc. 6th Int. Congr. Inf. Commun. Technol.*, 2022, pp. 501–510.
[41] B. Shi, X. Bai, and C. Yao, "An end-to-end trainable neural network for image-based sequence recognition and its application to scene text recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 11, pp. 2298–2304, Nov. 2017.
[42] Y. Du, C. Li, R. Guo, C. Cui, W. Liu, J. Zhou, B. Lu, Y. Yang, Q. Liu, X. Hu, D. Yu, and Y. Ma, "PP-OCRv2: Bag of tricks for ultra lightweight OCR system," 2021, *arXiv:2109.03144*.
[43] M. Liao, B. Shi, and X. Bai, "TextBoxes++: A single-shot oriented scene text detector," *IEEE Trans. Image Process.*, vol. 27, no. 8, pp. 3676–3690, Aug. 2018.
[44] B. Shi, M. Yang, X. Wang, P. Lyu, C. Yao, and X. Bai, "ASTER: An attentional scene text recognizer with flexible rectification," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 9, pp. 2035–2048, Sep. 2019.

QINGHUA SONG was born in Hefei, Anhui, China, in August 1972. He received the degree in computer technology and application, in July 1997, and the bachelor's degree in computer science and technology, in 2016. He is currently a Senior Engineer with the Information Center, Zhangjiakou Cigarette Factory Company Ltd., engaging in data acquisition, data integration, and data visualization. On this basis machine learning, deep learning, and image processing. His main research interest includes artificial intelligence, and his specialty is the cross-research of data processing and product development.

YAJUN LIU was born in Zhangjiakou, Hebei, China, in August 1992. He received the bachelor's and master's degrees in computer science, in 2016 and 2019, respectively. He is currently a Lecturer with Hebei University of Architecture, where he has been teaching and researching in computer science, since 2019. He also leads students to participate in a variety of big data and data modeling codes and win awards. He has published several academic papers and chaired or participated in several projects. His current main research interests include machine learning, deep learning, image processing, and artificial intelligence, and his expertise lies in data acquisition, modeling, and analysis.

HAOYUE SUN was born in Zhangjiakou, Hebei, China, in 1980. He received the master's degree from Hebei University of Technology. He is currently an Associate Professor with the College of Information Engineering, Hebei University of Architecture; and the Vice Dean, responsible for teaching and research in computer science. During the working period, he has published many high-level papers, organized many research projects, and achieved many results in teaching and research. Some of the results have been translated into high-level. His core research interests include computer networks, big data analysis, and image recognition technology.

YONG CHEN was born in Yining, Xinjiang, in January 1969. He received the degree in industrial electrical automation, in July 1991, and the bachelor's degree in electrical engineering and automation, in March 2010. He is currently a Senior Engineer with Zhangjiakou Cigarette Factory Company Ltd. He has presided over a number of scientific research projects and received several invention patents and awards. He engaged in new product development, software development, data-driven, and application innovation. His main research interests include data analysis and product development.



ZHENG ZHOU was born in Dandong, Liaoning, China, in December 1982. He received the bachelor's degree in computer science, in June 2006, and the master's degree in business administration, in 2013. He is currently with Zhangjiakou Cigarette Factory Company Ltd. He is mainly responsible for data collection, analysis, and decision-making in the field of big data in the company. He has chaired and participated in several projects, published several papers, and applied for several patents and soft writings. His research interests include machine learning, deep learning, and image processing.