

Design and Implementation of an Automatic Vehicle License Recognition System for End-of-Life Vehicle Dismantling Scenarios

Yining Liu *

Shijiazhuang Tiedao University, School of Electrical and Electronic Engineering, China

* Corresponding author: (Email: 2356535464@qq.com)

Abstract: The accuracy and efficiency of vehicle license information entry directly determines whether such operations can truly implement the Standards for Motor Vehicle Recycling and Dismantling Operations and further advance the digitalization of compliant supervision in this sector. To address challenges in this scenario, such as severely damaged documents, complex shooting conditions, and strong field logic, this paper designs and implements an automatic vehicle license information recognition system for end-of-life vehicle dismantling enterprises. The system first employs a YOLOv8 model combined with a multi-frame stability criterion to achieve robust detection of the vehicle license region and capture clear images under complex shooting environments. Subsequently, PaddleOCR is utilized for text extraction. Finally, a multi-layer post-processing engine integrating regular expression matching, semantic analysis, and business rules corrects and structures the raw OCR results, outputting standardized vehicle information. Experiments demonstrate that the designed post-processing rule engine effectively corrects the majority of OCR recognition errors. Ultimately, the system is integrated into a desktop application featuring data visualization and browser automation form-filling capabilities, achieving an end-to-end automated process from physical document scanning to business system entry, thereby enhancing operational efficiency and data accuracy.

Keywords: End-of-life vehicle license recognition, YOLOv8, PaddleOCR, post-processing rule engine, document digitization.

1. Introduction

With the continuous increase in motor vehicles in China and the deepening advancement of the national "Dual Carbon" strategy, the standardized recycling of end-of-life vehicles and the efficient utilization of their embedded resources have become key focal points in building a modern circular economy system. According to the *Administrative Measures for the Recycling of End-of-Life Vehicles* and its implementation rules issued by the Ministry of Commerce, dismantling enterprises must accurately enter vehicle license information into the "National Automobile Circulation Information Management" application service system upon vehicle reception. This process involves dozens of critical fields, including Vehicle Identification Number (VIN), engine number, usage nature, and inspection records. Its accuracy and efficiency directly impact compliant vehicle deregistration, component traceability, and environmental supervision. However, in actual operational scenarios, vehicle licenses of vehicles destined for scrapping often face multiple challenges: poor physical condition (oil stains, wear), complex acquisition environments (uneven lighting, shaking), and strong logical correlations between information fields. These factors lead to low efficiency and high error rates in traditional manual entry methods. General OCR solutions, lacking integration with domain-specific degraded image characteristics and business knowledge, struggle to meet the high-reliability requirements of this business context.

In recent years, deep learning techniques have shown significant promise in the field of document image analysis. However, existing research has predominantly focused on standard documents like ID cards and invoices. Public academic research on the niche scenario of end-of-life vehicle licenses, characterized by both extreme image degradation

and complex business logic, remains scarce. Current solutions either lack robustness against severe image degradation, fail to comprehend the intricate business rules among fields, and rarely consider automated integration with backend government systems.

Through in-depth exploration of the end-of-life vehicle dismantling industry, we naturally conceived the development of a dedicated, tailored intelligent scanning platform for vehicle licenses within this sector. The main contributions of this paper are: First, we design a detection and image capture process integrating a stability criterion to cope with complex shooting environments. Second, we construct a multi-layer post-processing rule engine that deeply integrates domain knowledge, effectively enhancing the robustness of structured extraction for key information under conditions of text damage and blurring. Third, we implement a complete end-to-end application system and validate its effectiveness through real-world scenario testing.

2. Related Work

Driven by the widespread application of document images, the demand for document digitization and intellectualization is increasing, making automated analysis of document images and efficient text recognition crucial foundations for modern documentation. In recent years, China has made significant progress in document image analysis and text recognition technologies. Early work relied on handcrafted features and traditional machine learning methods, such as text localization algorithms based on edge detection and morphological processing. In recent years, deep learning-based end-to-end methods have become mainstream. In the field of general OCR, open-source frameworks like PaddleOCR and EasyOCR integrate advanced text detection (e.g., DBNet, DB++) and recognition (e.g., CRNN, SVTR)

models, demonstrating excellent performance on various public benchmarks and providing powerful foundational capabilities for vertical applications [1-2]. For example, DBNet significantly improves detection performance for curved text through differentiable binarization post-processing, while SVTR introduces a visual Transformer structure to better model long-range dependencies between characters [3].

Combining the traditional "detection-recognition-post-processing" pipeline with the injection of sufficient domain knowledge can significantly enhance the understanding of vertical domain documents. For instance, in ID card, invoice, and business license recognition, structured information extraction is commonly achieved through key field localization, fixed template matching, and rule-based post-processing (e.g., ID number verification) [4-5]. The core idea of these methods is to use the document's prior layout and semantic information to constrain and correct the output of general OCR. Some studies attempt to introduce layout analysis models (e.g., LayoutLM) to understand document structure, but training costs are high in small-sample scenarios [6].

Although previous research has laid a solid foundation for specific document recognition, research on the niche scenario of end-of-life vehicle licenses remains blank. The particularity of this scenario lies in extremely poor image quality and complex business logic. Existing solutions either lack robustness to severe image degradation, cannot understand the complex business logic between vehicle license fields, and rarely consider automated integration with backend government systems [7-8]. Some industrial solutions, while achieving automated form filling, often rely on fixed UI element positioning, lacking visual perception capability and offering poor adaptability [9]. Therefore, the value of this work lies in addressing this specific industrial pain point by constructing a practical system that deeply integrates computer vision technology with domain business rules, filling the gap in automated processing for this field.

3. System Design and Implementation

3.1. Overall System Architecture

Through careful and rational module design, the platform's main functions have been preliminarily realized, primarily comprising five basic modules: image acquisition and preprocessing, vehicle license detection and tracking, text recognition, intelligent post-processing, and user application interaction. The workflow is as follows: The camera video stream, after preprocessing, is fed into the YOLOv8 detection module to locate the vehicle license region. The stability tracking module assesses whether the target is stable; if stable, it triggers automatic capture. The captured image is cropped and then subjected to text detection and recognition by PaddleOCR. The raw recognized text is sent to a custom post-processing engine, which performs field classification, validation, and structuring according to a preset rule set, outputting a standardized vehicle information dictionary. Finally, the results can be saved locally or automatically populated into target business system forms via a browser automation module. The system adopts a layered architecture to ensure loose coupling between algorithm modules and business logic, facilitating subsequent iteration.

3.2. Stability Detection and Capture Based on YOLOv8

Building upon the precise real-time detection capability of YOLOv8s, we further designed a multi-frame tracking-based stability criterion to effectively reduce the impact of shooting shake on image quality. The model was fine-tuned on a self-annotated dataset containing two categories: "license-front" and "license-back". The detection module not only outputs bounding boxes but also maintains a status queue for each target. The system calculates the Root Mean Square Error (RMSE) of the center point movement of the target over N consecutive frames. Capture is triggered only when the RMSE is below a set threshold. This mechanism effectively filters out motion blur caused by manual shaking, ensuring the quality of images fed into the recognition module.

The loss function, as shown in Formula (1), integrates bounding box regression loss (L_{CIoU}), classification loss (L_{BCE}), and distribution focal loss (L_{DFL}).

$$L_{total} = \lambda_{box} \cdot L_{CIoU} + \lambda_{cls} \cdot L_{BCE} + \lambda_{df} \cdot L_{DFL} \quad (1)$$

Where three λ are the weight coefficients for each loss term, and CIoU loss considers overlapping area, center point distance, and aspect ratio.

The stability criterion calculation is shown in Formula (2):

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2 + (y_i - \bar{y})^2} \quad (2)$$

Where (x_i, y_i) is the center point coordinate of the detection box in the i-th frame, and (\bar{x}, \bar{y}) is the average center point coordinate over N frames.

3.3. Text Recognition Based on PaddleOCR

By leveraging the fine-grained detection module based on the DB algorithm and the advanced recognition module based on the SVTR-LC network, our system achieves a higher level of precise detection and recognition. Through a series of preprocessing and feature extraction operations, it ultimately obtains an accurate recognition result for each piece of license information. The detection module is based on the DB algorithm, and the recognition module employs the SVTR-LC network. The system enables document correction functionality to mitigate perspective distortion. The recognition result is a list containing text content and their coordinates. To adapt to the layout of the vehicle license, the system performs initial spatial sorting on the OCR output, ensuring text lines are arranged in order from top to bottom and left to right, providing ordered input for subsequent rule processing.

3.4. Custom Post-Processing Rule Engine

By using regular expression matching and preliminary format validation for key information with distinct fixed patterns such as license plate numbers and VIN codes, illegal data for these fields is preliminarily filtered, playing a crucial role in the subsequent in-depth validation and analysis of information. This information extraction transforms "massive" text into "usable" vehicle information "ore," laying a solid foundation for subsequent vehicle information applications and development. The engine adopts a hierarchical decision framework: it first filters out noise text; then, it performs parallel judgment and extraction on the remaining text through multiple layers of rules. The first layer is the format validation layer, matching fields with fixed

patterns such as license plate numbers and VIN codes using regular expressions. The second layer is the semantic analysis layer, which judges fields without fixed formats but with specific semantic characteristics, such as "owner" and "usage nature." The third layer is the numerical normalization layer, extracting fields containing units, such as "gross mass" and "approved passenger capacity." The fourth layer is the field association and business rule layer, which utilizes logical relationships between fields (e.g., the difference in character composition between the engine number and VIN) to resolve ambiguities, further enhancing robustness in the event of OCR recognition errors.

3.5. System Implementation and Interactive Interface

Leveraging the deep application of PySide6, we present the system's major functions to the user in a desktop format, greatly facilitating the user experience. The graphical user interface primarily consists of a real-time video preview area, a data display area, and a log control area. The browser automation module is responsible for automatically populating the structured data into the web forms of the target business system after recognition is complete, achieving a closed-loop business process.

4. System Testing and Analysis

4.1. Evaluation Metrics

To comprehensively evaluate system performance, metrics are defined:

Field-level extraction accuracy is reflected by the ratio of correctly extracted samples for a single field to the total number of samples. We focus on critical fields where errors are unacceptable, such as VIN and engine number.

4.2. System Performance Test

The test platform configuration is: Intel Core i9-14900H CPU, NVIDIA GeForce RTX 4060 Laptop GPU, 16GB RAM.

4.2.1. Overall Performance

The system achieves an average recognition accuracy exceeding 90% on all 50 images. Field-level accuracy is shown in Table 1. Fields with fixed formats (e.g., VIN, engine number) achieve the highest accuracy, while free-text fields (e.g., owner) show relatively lower accuracy due to the impact of stains and damage.

Table 1. System Overall Field-Level Extraction Accuracy

Field Name	Accuracy	Field Name	Accuracy
License Plate Number	90%	Vehicle Identification Number (VIN)	100%
Engine Number	100%	Owner	86%
Usage Nature	95%	Brand/Model	91%
Registration Date	92%	Vehicle Type	100%

4.2.2. Scenario-Specific Robustness Test

The system performs excellently under conventional working conditions. Preliminary classification (as shown in Table 2) indirectly indicates the preliminary feasibility of text recognition in this domain. The system performs excellently under normal conditions. Under abnormal lighting conditions, using a fill light fixture significantly maintains a high pass rate. In scenarios with skewed printing/oil stain occlusion, the

pass rate drops to 85.7%. The post-processing engine corrects about 70% of the OCR's original field assignment errors but struggles to recover from severe character-level recognition errors.

As shown in Table 2, under different challenging scenarios, our system's performance on various metrics varies but generally meets the requirements of the corresponding challenges.

Table 2. Extraction Accuracy in Different Scenarios

Scenario	Extraction Accuracy
Normal Conditions (good lighting, printing, no damage/occlusion)	99.4%
Abnormal Lighting	93.3%
Skewed Printing/Oil Stain Occlusion...	85.6%

4.2.3. Module Performance Analysis

Beyond achieving relatively high accuracy through further system optimization, the average processing time for a single instance from detection to structured output is approximately 15 seconds. This significantly improves efficiency compared to manual entry (which typically takes 3-5 minutes) and also avoids the increased error rate caused by fatigue from prolonged manual input.

5. Discussion and Limitations

Although its performance on various test metrics is excellent, its practical application still faces several limitations and challenges, such as its applicability in certain special scenarios remains questionable.

First, the generalizability of the rule engine. The system's post-processing heavily relies on the fixed layout and field logic of the current vehicle license. If the license format undergoes significant changes, the rule base requires manual updates, increasing system maintenance costs with policy changes.

Second, the ability to handle extreme damage. When key fields are completely obscured or corroded, the system cannot recover the information. This is a physical information loss issue and requires supplementation through other means, such as querying the vehicle management database.

Third, the complexity of contextual ambiguity. The current rules employ sequential matching and simple contextual judgment. In the future, lightweight machine learning models could be introduced for priority ranking.

Fourth, end-to-end latency. PaddleOCR recognition remains the main time-consuming step. When deploying on lower-end hardware, model optimization needs to be considered to balance speed and accuracy.

6. Conclusion and Future Work

By deeply exploring the pain points of automated vehicle license information entry in end-of-life vehicle recycling operations, we have designed and successfully implemented a targeted intelligent scanning platform, which is now undergoing pilot cooperation with relevant enterprises. This system effectively enhances the robustness and accuracy of document information extraction in complex scenarios by integrating YOLOv8 object detection, PaddleOCR text recognition, and a deeply customized post-processing rule engine. The final software system achieves end-to-end automation from physical document capture to business

system entry, possessing practical value.

Future work will focus on developing recovery techniques for text regions with severe image degradation to further improve recognition rates and accuracy. We will also work on reducing the system's high dependence on fixed layouts by exploring few-shot adaptive methods based on pre-trained document understanding models to enhance the system's generalizability. Furthermore, we will construct larger, more diverse datasets and consider extending the system's capabilities to related business documents such as motor vehicle registration certificates.

References

- [1] Du, Y., Li, C., Guo, R., et al. (2020). PP-OCR: A practical ultra lightweight OCR system. arXiv. <https://doi.org/10.48550/arXiv.2009.09941>
- [2] Huang, Z., Chen, K., He, J., et al. (2020). LayoutLM: Pre-training of text and layout for document image understanding. In Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining.
- [3] Redmon, J., Divvala, S., Girshick, R., et al. (2016). You only look once: Unified, real-time object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.
- [4] Ministry of Commerce of the People's Republic of China. (2020). Implementation rules for the administrative measures on the recycling of end-of-life vehicles. (in Chinese)
- [5] Liu, W., & Zhang, Q. (2021). Research on vehicle license recognition system based on deep learning. *Computer Engineering and Design*, 42(5), 1342–1348. (in Chinese)
- [6] Du, Y., Li, C., Guo, R., et al. (2021). PP-OCRv2: Indispensable strategies for making training better and faster. arXiv. <https://doi.org/10.48550/arXiv.2109.03144>
- [7] Wang, H. (2022). Discussion on informatization construction of end-of-life vehicle recycling and dismantling industry. *Renewable Resources and Circular Economy*, 15(3), 12–15. (in Chinese)
- [8] Chen, G. (2023). *Computer vision: Fundamental algorithms and engineering practices*. Publishing House of Electronics Industry. (in Chinese)
- [9] Ge, Z., Liu, S., Wang, F., et al. (2021). YOLOX: Exceeding YOLO series in 2021. arXiv. <https://doi.org/10.48550/arXiv.2107.08430>