Global Journal of Engineering Innovations & Interdisciplinary Research

Original Article

An OCR-Based Intelligent System for Automated Invoice Data Extraction

J. Swapna Priya[1], Shaik Anwar[2], Shaik Farukh[2], Sayampu Vamsi[2], Shaik Nadeem[2]

[1]Assistant Professor, Department of CSE, KKR & KSR Institute of Technology and Sciences, Guntur, Andhra Pradesh, India
[2]B.Tech Student, Department of CSE, KKR & KSR Institute of Technology and Sciences, Guntur, Andhra Pradesh, India

Correspondence

J. Swapna Priya

Assistant Professor, Department of CSE, KKR & KSR Institute of Technology and Sciences, Guntur, Andhra Pradesh, India

- Received Date: 08 Jan 2026
- Accepted Date: 25 Jan 2026
- Publication Date: 15 Mar 2026

Keywords

Invoice data extraction, OCR, YOLO, Tesseract, document analysis, intelligent automation, structured document processing

Abstract

Accurate and efficient extraction of invoice data is an essential requirement in modern business and financial operations, where organizations process large numbers of in- voices every day. Traditional manual invoice handling is time- consuming, labor-intensive, and highly error-prone. Conventional Optical Character Recognition (OCR) systems offer partial au- tomation by converting document images into machine-readable text, but they often fail to extract critical invoice fields reliably because of variations in layout, font style, image quality, scanning distortions, and background noise. These limitations lead to unstructured outputs and require extensive human verification. This paper presents an OCR-based intelligent system for auto- mated invoice data extraction by integrating YOLO-based object detection with selective OCR. Instead of applying OCR to the entire invoice image, the proposed system first detects important invoice fields such as invoice number, date, vendor details, and total amount. OCR is then applied only to the detected regions of interest, which reduces noise and improves recognition accuracy. The extracted text is further refined using rule-based validation and pattern matching to generate structured outputs in CSV and Excel-compatible form. The reformatted study reports an extraction accuracy of approximately 92–95% and a processing speed improvement of nearly 35–40% over a conventional OCR- only approach, demonstrating the practical value of combining region localization and targeted text recognition for invoice automation.

Introduction

Organizations across finance, logistics, healthcare, retail, and manufacturing rely on invoices to record transactions, validate payments, manage vendors, and support auditing. Although invoice processing is central to day-to-day business operations, many organizations still depend on manual entry or semi-automated pipelines. Manual entry increases turnaround time and operating cost and also introduces avoidable errors in key fields such as invoice number, issue date, tax amount, and final payable amount. OCR-based automation improves doc- ument digitization, but OCR alone does not model document structure, spatial relationships, or field semantics [5,19].

Modern object detectors such as YOLO have made it possible to localize important regions within an image in real time [1]–[4]. In parallel, document understanding research has shown that layout-aware processing is essential for ex- tracting structured information from complex forms, receipts, and visually rich business documents [11]–[13], [15], [16]. These developments motivate an invoice extraction strategy in which field localization and text recognition are performed as coordinated stages rather than as a single OCR pass over the full document.

The uploaded manuscript proposes an OCR-based intelli- gent system that uses preprocessing, YOLO-based region de- tection, selective OCR, and post-processing to improve invoice data extraction. The central idea is straightforward: detect the important fields first, then apply OCR only where needed. This approach reduces irrelevant text extraction, improves robustness to diverse invoice layouts, and produces structured outputs suitable for downstream accounting workflows.

The main contributions of the reformatted paper are as follows. First, it presents a clean IEEE two-column structure for the proposed invoice extraction pipeline. Second, it con- solidates the original methodology, architecture, discussion, and conclusion into a coherent research-paper format. Third, it preserves the reported performance values without inventing missing experimental details.

Problem Statement And Objectives

Problem Statement

In modern business environments, invoices arrive in many formats, templates, resolutions,

Citation: Swapna Priya J, Shaik A, Shaik F, Sayampu V, Shaik N. An OCR-Based Intelligent System for Automated Invoice Data Extraction. GJEIIR. 2026;6(3):215.

and visual styles. Traditional manual processing is expensive and slow, while OCR-only systems often produce unstructured text that does not clearly separate important fields. Performance further degrades when invoices contain skew, blur, scanning artifacts, shadows, low contrast, or complex backgrounds. As a result, critical fields such as invoice number, date, vendor details, subtotal, tax, and total amount may be extracted incorrectly or may require manual correction.

Objectives

The objective of this work is to design an intelligent invoice data extraction system that:

1. detects important invoice fields automatically,
2. applies OCR selectively to relevant regions,
3. improves extraction accuracy and speed over OCR-only processing,
4. reduces noise and human effort, and
5. stores results in structured formats such as CSV and Excel for integration with financial systems.

Literature Review

Traditional OCR engines such as Tesseract convert document images into text effectively when document quality is good and layouts are simple [5]. However, invoice images are often heterogeneous and visually noisy, which makes direct OCR insufficient for reliable field extraction. Classical image processing and document analysis methods improve readability through enhancement, denoising, and skew correction [6], [19], [20], but they still depend heavily on predictable layouts. Deep learning has significantly advanced region localization and document understanding. General object detectors such as Faster R-CNN and SSD established accurate detection baselines [9], [10], while YOLO introduced fast single-stage detection that is suitable for real-time localization [1]–[4]. For text-specific detection, methods such as EAST and CRAFT improved the localization of text regions in complex scenes and documents [7], [8]. These advances support the idea of detecting invoice fields first and recognizing text second.

Research in document intelligence further highlights the importance of layout. Datasets such as FUNSD and CORD demonstrate that structured documents require models that use both text and spatial context [11], [12]. Models such as Chargrid, LayoutLM, LayoutLMv2, DocFormer, and Donut have shown strong performance on visually rich document understanding tasks [13], [15]–[18]. In the invoice domain, CloudScan showed that machine learning can reduce rule- heavy configuration in invoice analysis pipelines [14]. The proposed system in this paper follows the same broad direction of intelligent document extraction, but uses a simpler, practical pipeline based on YOLO plus selective OCR.

Methodology

The proposed methodology consists of five major stages: image acquisition, preprocessing, invoice-field detection, selective OCR, and structured output generation.

Image Acquisition and Preprocessing

Invoice images may originate from scanners, mobile cameras, or digital uploads. Because input quality can vary significantly, preprocessing is used to improve visual clarity before field extraction. Typical operations include denoising, resizing, contrast enhancement, and skew correction. These steps improve the stability of both object detection and OCR [6], [19], [20]. Better input quality reduces false detections and increases text readability.

YOLO-Based Field Detection

After preprocessing, the invoice image is passed to a YOLO-based detector. The detector localizes important fields such as invoice number, date, vendor information, and total amount.

Compared with full-image OCR, this step focuses computation on semantically relevant regions and reduces the effect of irrelevant text blocks, logos, decorative elements, and back- ground artifacts. The choice of YOLO is appropriate for fast and efficient localization in document images [1]–[4].

Region of Interest Extraction

Once the detector predicts bounding boxes, the corresponding regions of interest (ROIs) are cropped from the invoice image. ROI extraction isolates meaningful content and reduces clutter. This stage is important because OCR accuracy decreases when large amounts of unrelated text or image noise are included in the recognition window.

Selective OCR

Tesseract OCR is then applied only to the extracted ROIs rather than to the complete invoice page [5]. This selective OCR strategy improves recognition quality by limiting OCR to the most relevant field regions. It also lowers processing overhead and helps maintain consistent performance across different invoice layouts.

Post-Processing and Data Structuring

The recognized text is refined using pattern matching and rule-based validation. For example, date fields can be checked against valid date patterns, invoice numbers can be normalized, and numeric totals can be validated for currency-like formatting. The final output is converted into structured formats such as CSV and Excel-compatible tables, making the extracted information easier to retrieve, analyze, and integrate with accounting systems.

Proposed Intelligent Invoice Extraction Framework

Figure 1 presents the overall architecture of the proposed system. The pipeline begins with the input module, which receives invoice images from scanning or digital sources. The preprocessing module enhances image quality through resizing, contrast adjustment, noise removal, and skew correction. The YOLO module then detects the relevant invoice fields, after which OCR is applied only to the selected field regions. Finally, post-processing validates the extracted text and stores it in a structured form.

The framework can be understood module by module:

Input Module: Accepts invoice images from scanners, cameras, or uploaded files.

Image Processing Module: Enhances document quality for downstream processing.

YOLO-Based Object Detection Module: Detects and localizes important invoice fields.

OCR Module: Applies Tesseract OCR to the detected field regions.

Post-Processing Module: Performs validation, pattern matching, field labeling, and formatting.

Data Storage Module: Saves the final output in structured formats such as CSV and Excel.

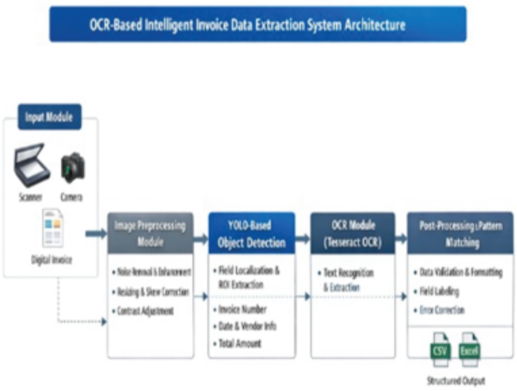

Fig. 1. System architecture for the OCR-based intelligent invoice data extraction pipeline.

Table 1. Reported Performance Comparison

Metric	Existing OCR-Based Method	Proposed YOLO-Integrated Method
Accuracy	85–88%	92–95%
Processing	Standard OCR speed Speed	35–40% faster
Noise Reduction	Limited	Significant reduction
Output Quality	Mostly unstructured text	Structured field-wise extraction

This modular design makes the system practical for business workflows because each stage contributes directly to either accuracy, speed, or output usability.

Results And Performance Analysis

The source manuscript reports that the proposed system was tested on invoice images with different formats and quality levels. According to the paper, the combined YOLO-plus- selective-OCR approach successfully extracted key fields such as invoice number, date, and total amount while reducing un- wanted text. The reported extraction accuracy is approximately 92–95%, and the reported processing speed improvement is nearly 35–40% compared with a conventional OCR-only method.

Because the source manuscript does not specify dataset size, train/test split, annotation protocol, or hardware configuration, the evaluation is best interpreted as a comparative summary rather than a fully reproducible benchmark. To remain faithful to the original document, Table I and Fig. 2 reproduce only the performance information explicitly stated in the paper.

The reported results suggest that region-based OCR improves both precision and efficiency. Instead of processing the entire invoice page, the system performs recognition on a smaller and more relevant set of image regions. This lowers the amount of irrelevant text passed to OCR and reduces the burden of post-correction.

Discussion

The reformatted paper indicates that the proposed system improves invoice extraction by combining localization and recognition. This is consistent with broader document-AI literature, where layout and field position are known to influence extraction quality [11], [12], [15], [16]. For invoices with varied templates, a detection-first strategy is more robust than applying OCR uniformly to the entire page.

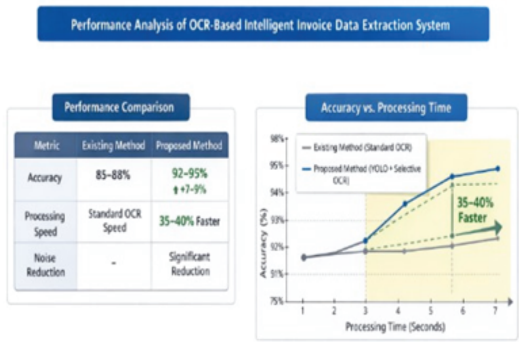

Fig. 1. Accuracy comparison of the existing OCR-based method and the proposed YOLO-integrated invoice data extraction system.

A second advantage is practical deployment. Structured outputs in CSV and Excel-compatible form make the system immediately useful for enterprise workflows such as accounts payable, auditing, and digital archiving. The architecture is also modular, which means that future improvements can be introduced at specific stages without redesigning the whole system. For example, a stronger detector or a layout-aware transformer could replace the current field-localization stage [17,18].

At the same time, the current manuscript leaves some experimental details unspecified. It does not report dataset size, annotation scope, class balance, training settings, or a field-wise breakdown of precision and recall. These omissions do not invalidate the reported gains, but they do limit strict reproducibility and make cross-paper comparison difficult. Future versions of the work should include a larger annotated invoice dataset, ablation studies for each pipeline stage, and separate accuracy scores for fields such as invoice number, date, tax, vendor name, and total amount.

Conclusion and Future Scope

SThis paper presented a reformatted IEEE version of an OCR-based intelligent system for automated invoice data extraction. The proposed approach combines image prepro- cessing, YOLO-based field detection, selective OCR, and rule-based post-processing to extract important invoice information accurately and efficiently. Based on the reported results in the source manuscript, the system achieves approximately 92–95% extraction accuracy and improves processing speed by about 35–40% relative to a traditional OCR-only pipeline.

The work demonstrates that selective OCR over detected invoice regions is a practical and effective strategy for reducing noise, improving field recognition, and generating structured outputs for downstream financial systems. Future work can strengthen the system by expanding the training dataset, adding more invoice field classes, incorporating layout-aware language models, and reporting standardized evaluation met- rics for stronger reproducibility.

References

1. J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, “You Only Look Once: Unified, Real-Time Object Detection,” in Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), 2016, pp. 779–788.
2. J. Redmon and A. Farhadi, “YOLO9000: Better, Faster, Stronger,” in Proc. IEEE Conf. Computer Vision and Pattern Recognition

(CVPR), 2017, pp. 6517–6525.

3. J. Redmon and A. Farhadi, “YOLOv3: An Incremental Improvement,”arXiv preprint arXiv:1804.02767, 2018.
4. A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, “YOLOv4: Optimal Speed and Accuracy of Object Detection,” arXiv preprint arXiv:2004.10934, 2020.
5. R. Smith, “An Overview of the Tesseract OCR Engine,” in Proc. Int. Conf. Document Analysis and Recognition (ICDAR), 2007, pp. 629–633.
6. G. Bradski, “The OpenCV Library,” Dr. Dobb’s Journal of Software Tools, vol. 25, no. 11, pp. 120–126, 2000.
7. X. Zhou, C. Yao, H. Wen, Y. Wang, S. Zhou, W. He, and J. Liang, “EAST: An Efficient and Accurate Scene Text Detector,” in Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), 2017, pp. 5551–5560.
8. Y. Baek, B. Lee, D. Han, S. Yun, and H. Lee, “Character Region Awareness for Text Detection,” in Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), 2019, pp. 9365–9374.
9. S. Ren, K. He, R. Girshick, and J. Sun, “Faster R-CNN: Towards Real- Time Object Detection with Region Proposal Networks,” in Advances in Neural Information Processing Systems (NeurIPS), 2015, pp. 91–99.
10. W. Liu et al., “SSD: Single Shot MultiBox Detector,” in Proc. European Conf. Computer Vision (ECCV), 2016, pp. 21–37.
11. G. Jaume, H. Ekenel, and J.-P. Thiran, “FUNSD: A Dataset for Form Understanding in Noisy Scanned Documents,” in Proc. Int. Conf. Document Analysis and Recognition Workshops (ICDARW), 2019, pp. 1–6.
12. S. Park et al., “CORD: A Consolidated Receipt Dataset for Post- OCR Parsing,” in Proc. Int. Conf. Document Analysis and Recognition Workshops (ICDARW), 2019, pp. 97–102.
13. D. Katti et al., “Chargrid: Towards Understanding 2D Documents,” in Proc. Conf. Empirical Methods in Natural Language Processing (EMNLP), 2018, pp. 4459–4469.
14. R. B. Palm, F. Laws, and O. Winther, “CloudScan—A Configuration- Free Invoice Analysis System Using Recurrent Neural Networks,” in Proc. Int. Conf. Document Analysis and Recognition (ICDAR), 2017,pp. 406–413.
15. Y. Xu, M. Li, L. Cui, S. Huang, F. Wei, and M. Zhou, “LayoutLM: Pre-training of Text and Layout for Document Image Understanding,” in Proc. ACM SIGKDD Int. Conf. Knowledge Discovery & Data Mining (KDD), 2020, pp. 1192–1200.
16. Y. Xu et al., “LayoutLMv2: Multi-modal Pre-training for Visually-Rich Document Understanding,” in Proc. Annual Meeting of the Association for Computational Linguistics (ACL-IJCNLP), 2021, pp. 2579–2591.
17. A. Appalaraju et al., “DocFormer: End-to-End Transformer for Docu- ment Understanding,” in Proc. IEEE/CVF Int. Conf. Computer Vision Workshops (ICCVW), 2021, pp. 993–1003.
18. G. Kim et al., “OCR-Free Document Understanding Transformer,” in Proc. European Conf. Computer Vision (ECCV), 2022, pp. 498–517.
19. J. Liang, D. Doermann, and H. Li, “Camera-Based Analysis of Text and Documents: A Survey,” International Journal on Document Analysis and Recognition, vol. 7, no. 2–3, pp. 84–104, 2005.
20. R. C. Gonzalez and R. E. Woods, Digital Image Processing, 4th ed. New York, NY, USA: Pearson, 2018