2nd INTERNATIONAL CONFERENCE ON
Artificial Intelligence
31 May - 1 June 2026

IICAI
Second International Interdisciplinary Conference on Artificial Intelligence

An Extended Table Detection Framework for Enhanced Recognition of Persian Tables in OCR Systems

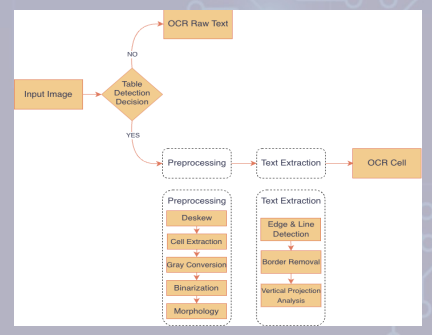
Aras Valizadeh, Zahra Mohammadi, Bita Baroutian, Atefe Aghaei, Fateme Taraghi, Mohsen Ebrahimi Moghaddam*
Department of Computer Science and Engineering Shahid Beheshti University, Tehran, Iran
Corresponding Author: Mohsen Ebrahimi Moghaddam
Professor of Computer Engineering
m_moghadam@sbu.ac.ir

Abstract

Accurate OCR for Persian documents remains challenging, particularly in table-rich layouts with irregular structures and noisy scans. Existing OCR and vision-language models often suffer from structural hallucinations, incorrect cell detection, and unstable table reconstruction. In this work, we propose a hybrid OCR-agnostic framework that combines classical image-processing techniques with modern OCR systems to improve Persian table detection and structural consistency. The framework integrates preprocessing, projection-based refinement, and cell-level analysis to enhance table segmentation before text recognition. A dedicated Persian table dataset containing synthetic and real-world samples was also developed for evaluation. Experimental results demonstrate that the proposed approach significantly improves layout detection accuracy and reduces structural errors compared to recent OCR and VLM-based methods.

Research method

1. Table regions are first detected using the Surya model.
2. Detected tables are preprocessed using:
   • deskewing
   • grayscale conversion
   • Otsu thresholding
   • morphological operations
3. Table cells are extracted through row–column boundary analysis.
4. Canny edge detection and Hough Transform are applied to refine table structures and remove ruling-line artifacts.
5. Projection-based refinement is used to merge incorrectly separated cells.
6. The refined cells are finally forwarded to the OCR engine (Hezar) for text recognition.



conclusion

This work presented a hybrid OCR-agnostic framework for improving Persian table detection and structural reconstruction in OCR systems. By combining classical image-processing techniques with modern OCR models, the proposed approach significantly reduced structural hallucination and segmentation errors commonly observed in VLM-based methods. Experimental results demonstrated superior layout detection performance and strong robustness against noisy and complex Persian table layouts. Furthermore, the introduced Persian table dataset provides a useful benchmark for future research in Persian document analysis and table recognition.

The purpose of the research

This research aims to improve Persian table detection in OCR systems and address the structural limitations of recent VLM-based OCR approaches. The proposed framework focuses on reducing hallucination and segmentation errors that commonly occur in complex and noisy Persian documents. In addition, the study introduces an OCR-agnostic framework that can be integrated with existing OCR systems to enhance table-cell segmentation and structural consistency. To support evaluation and future benchmarking, a dedicated Persian table dataset containing synthetic and real-world samples was also developed.

Practical or simulation results

The proposed framework was evaluated using a dedicated Persian table dataset consisting of synthetic and real-world samples with diverse layouts, noisy scans, merged cells, and irregular structures. Experimental comparisons were conducted against Dots.OCR and DeepSeek-OCR models using layout-wise precision, layout-wise recall, WER, and CER metrics. The results demonstrated that the proposed method significantly improved structural consistency and table reconstruction accuracy. In particular, the framework achieved 0.998 layout-wise precision and 0.9904 layout-wise recall, outperforming existing VLM-based OCR approaches in reliable table detection and reducing hallucination and structural segmentation errors.

References

[1] Z. Khosrobeigi, H. Veisi, E. Hoseinzade, and H. Shabanian, “Persian Optical Character Recognition Using Deep Bidirectional Long Short-Term Memory,” Applied Sciences, vol. 12, no. 22, p. 11760, 2022.

[2] M. Kasem et al., “Deep Learning for Table Detection and Structure Recognition: A Survey,” ACM Computing Surveys, vol. 56, no. 12, 2024.

[3] rednote-hilab, “dots.ocr: Multilingual document layout parsing in a single vision-language model,” GitHub repository, 2025.

[4] G. Kim et al., “OCR-free Document Understanding Transformer (Donut),” in ECCV, 2022.

[5] X. Chen et al., “PaLI: A Jointly-Scaled Multilingual Language-Image Model,” in ICLR, 2023.

[6] H. Wei, Y. Sun, and Y. Li, “DeepSeek-OCR: Contexts Optical Compression,” arXiv preprint arXiv:2510.18234, 2025.

[7] H. Liu et al., “A Survey on Hallucination in Large Vision-Language Models,” arXiv preprint arXiv:2402.00253, 2024.

....