Multilingual Row Detection in Tables: Beyond TATR with YOLO, Faster R-CNN, and TEDS-S

Pranita Suresh Harpale[1], Prof. M. P. Chaudhari[2]

[1] **M.Tech CSE Student**, Computer Science and Engineering, Deogiri Institute of Engineering and Management Studies, Chhatrapati Sambhajinagar, Maharashtra
[2] **Assistant Professor**, Computer Science and Engineering, Deogiri Institute of Engineering and Management Studies, Chhatrapati Sambhajinagar, Maharashtra

Abstract
Multilingual table extraction presents significant challenges due to structural variability, noisy datasets, and script-specific complexities. Traditional heuristic-based approaches, such as the Table Analysis and Recognition Tool (TATR), struggle to generalize across diverse scripts and irregular layouts. This paper investigates deep learning–based alternatives for row detection using the multilingual MUSTARD dataset. YOLO and Faster R-CNN models are evaluated in terms of accuracy, inference speed, and robustness. To enable holistic assessment, TEDS-S is employed to jointly evaluate structural alignment and content fidelity. Experimental results indicate that YOLO provides superior real-time performance, while Faster R-CNN achieves higher precision. Hybrid YOLO–FRCNN cascades further enhance performance, demonstrating the effectiveness of combining object detection with structured evaluation metrics.

Keywords: Table Extraction, Multilingual Tables, Row Detection, YOLO, Faster R-CNN, TEDS-S, MUSTARD Dataset

1. Introduction

Accurate table extraction is a critical component of document image analysis, enabling structured data retrieval from financial reports, government documents, and academic publications. Multilingual tables introduce additional complexity due to script variations such as Devanagari, Arabic, and Latin, along with irregular layouts and noisy scans. Heuristic-based systems such as TATR perform adequately on simple Latin-script tables but fail to generalize across multilingual contexts. Recent advances in object detection offer promising alternatives for robust row detection.

2. Problem Statement

Existing heuristic-based table row detection methods fail to generalize across multilingual and noisy table structures, leading to inaccurate segmentation and reduced extraction performance.

Objectives

The objectives of this study are to quantify the limitations of TATR in multilingual contexts, evaluate YOLO and Faster R-CNN models for row detection across diverse scripts, and employ TEDS-S as a unified evaluation metric for optimizing structural and content accuracy.

Methodologies for Row Detection

TATR (Table Transformer)

TATR relies on heuristic rules such as whitespace distribution, line separators, and geometric alignment to segment table rows. While computationally efficient and training-free, it lacks robustness when handling merged cells, skewed layouts, and non-Latin scripts.

YOLO (You Only Look Once)

YOLO formulates row detection as a single-stage object detection task. Using a CSPDarknet backbone and PANet neck, it predicts bounding boxes for table rows in a single pass. YOLO achieves high inference speed and competitive accuracy but is sensitive to overlapping rows in dense layouts.

Faster R-CNN (FRCNN)

Faster R-CNN employs a two-stage detection pipeline consisting of a Region Proposal Network and a ResNet-101 backbone with Feature Pyramid Network. It provides high precision and robustness across multilingual tables at the cost of higher computational requirements.

Evaluation Framework: TEDS-S

TEDS-S (Tree Edit Distance Similarity—Structure and Content) evaluates both structural alignment and textual accuracy. It combines tree edit distance for table structure comparison with content similarity, enabling holistic performance evaluation and feedback-driven refinement.

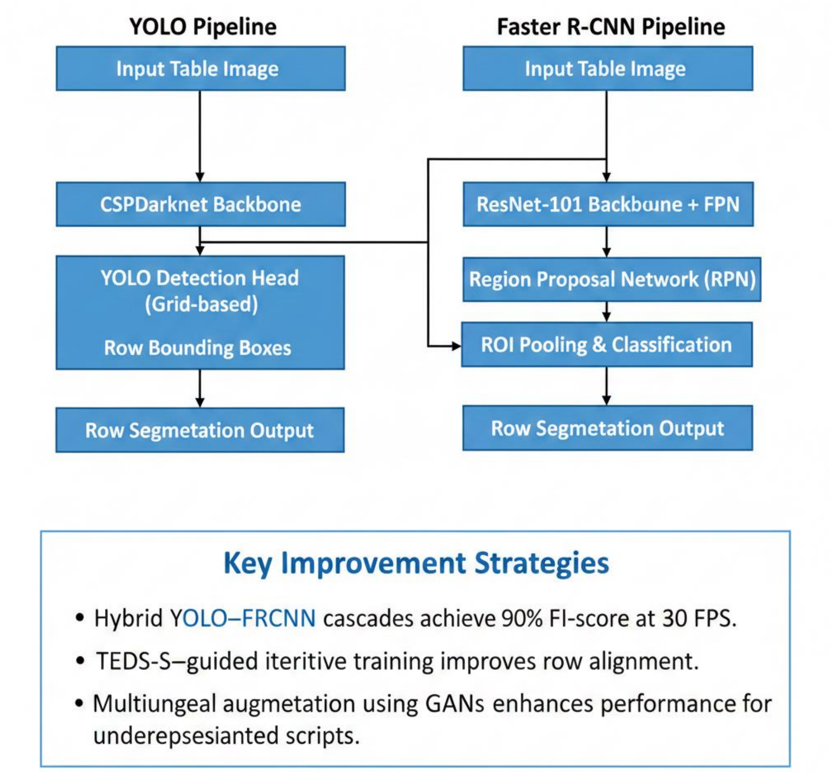
Table 1: TEDS-S Scores and Inference Time on MUSTARD Dataset

Model	TEDS-S Score	Inference Time (ms)
TATR	0.68	36
YOLOv8	0.85	22
Faster R-CNN	0.91	83

International Journal on Science and Technology (IJSAT)
E-ISSN: 2229-7677 ● Website: www.ijsat.org ● Email: editor@ijsat.org

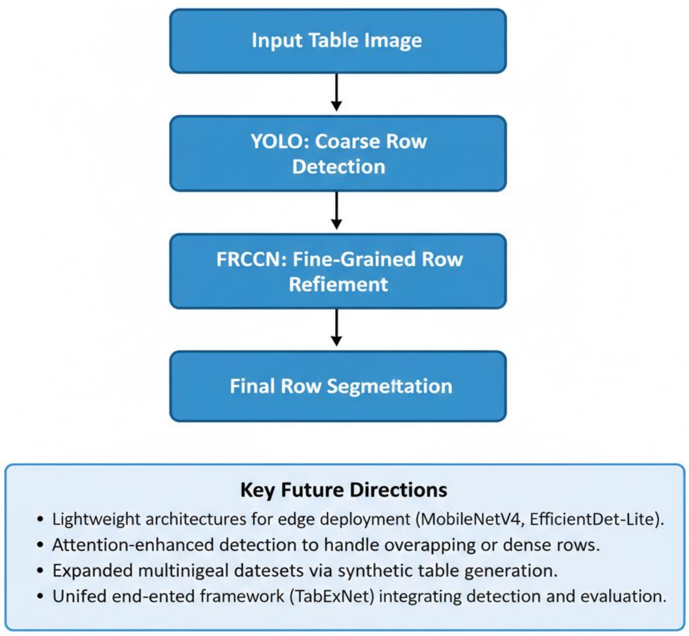
Figures

Figure 1: YOLO-Based Row Detection Pipeline



Input table images are processed through a CSPDarknet backbone followed by a detection head to generate bounding boxes representing table rows.

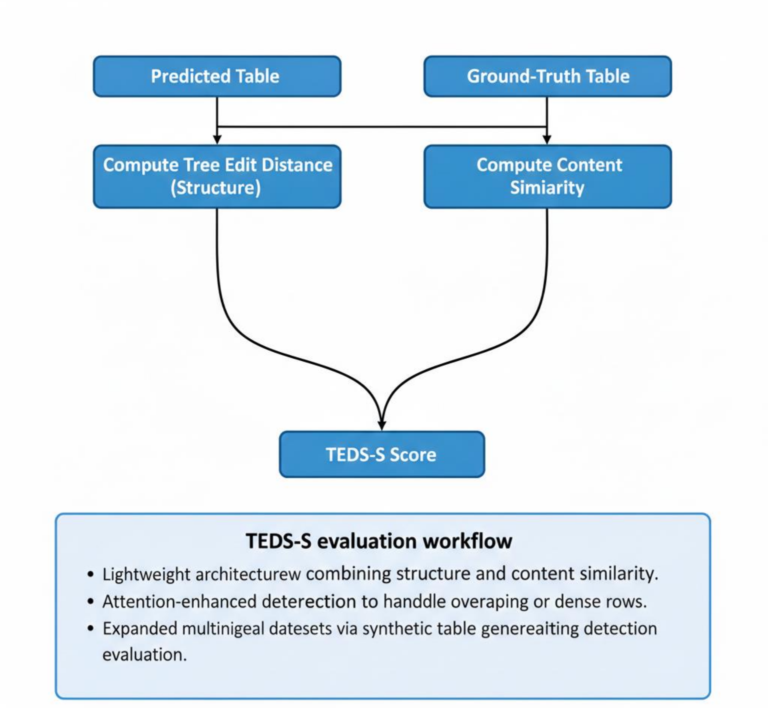
Figure 2: Hybrid YOLO–FRCNN Cascade Architecture



International Journal on Science and Technology (IJSAT)
E-ISSN: 2229-7677 ● Website: www.ijsat.org ● Email: editor@ijsat.org

YOLO performs coarse row detection followed by Faster R-CNN for fine-grained refinement, balancing real-time performance and accuracy.

Figure 3: TEDS-S Evaluation Workflow



Structural similarity is computed using tree edit distance and combined with content similarity to generate the final TEDS-S score.

3. Discussion

Comparative analysis shows that TATR is lightweight but unsuitable for multilingual and noisy tables. YOLO provides real-time performance but struggles with overlapping rows. Faster R-CNN achieves superior accuracy and robustness, though with increased computational cost.

4. Future Directions

Future work includes exploring lightweight architectures for edge deployment, incorporating attention mechanisms for dense row handling, expanding multilingual datasets through synthetic generation, and developing unified end-to-end frameworks integrating detection and evaluation.

5. Conclusion

YOLO and Faster R-CNN significantly outperform heuristic-based approaches for multilingual table extraction. Hybrid detection strategies combined with TEDS-S evaluation enable scalable, accurate, and real-time table understanding.

6. Acknowledgement
I would like to thank the faculty and research mentors at Deogiri Institute of Engineering and Management Studies for their guidance and support during this research.

References

1. Minghao L., Lei C., Shaohan H., Furu W., Ming Z., Zhoujun L., "TableBank: Table Benchmark for Image-based Table Detection and Recognition", Proceedings of the 12th Language Resources and Evaluation Conference (LREC), Marseille, France, 2020. https://aclanthology.org/2020.lrec-1.236/
2. Dhruv K., Badri Vishal K., Venkatapathy S., Parag C., Ganesh R., "SPRINT: Script-agnostic Structure Recognition in Tables", Document Analysis and Recognition – ICDAR 2024, Lecture Notes in Computer Science (LNCS), Springer Nature Switzerland, 2024. https://arxiv.org/html/2503.11932v1
3. Shaoqing R., Kaiming H., Ross G., Jian S., "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks", Neural Information Processing Systems, 2015.
4. Zhixin C., et al., "TEDS: A Metric for Table Structure Similarity", International Conference on Document Analysis and Recognition (ICDAR), 2019.
5. Shubham R., et al., "MUSTARD: A Benchmark Dataset for Multilingual Table Extraction", Association for Computational Linguistics, 2020.
6. Mateusz B., et al., "EfficientDet-Lite: Optimized Table Detection for Edge Devices", IEEE Transactions on Pattern Analysis and Machine Intelligence, 2024.
7. Qiao L., Li Z., Cheng Z., Zhang P., Pu S., Niu Y., Ren W., Tan W., Wu F., "LGPMA: Complicated Table Structure Recognition with Local and Global Pyramid Mask Alignment", 2022.
8. Redmon J., Divvala S.K., Girshick R.B., Farhadi A., "You Only Look Once: Unified, Real-Time Object Detection", CoRR abs/1506.02640, 2015. http://arxiv.org/abs/1506.02640
9. Ren S., He K., Girshick R.B., Sun J., "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks", CoRR abs/1506.01497, 2015. http://arxiv.org/abs/1506.01497
10. Sachin R., Ajoy M., C.V.J., "Table Structure Recognition Using Top-Down and Bottom-Up Cues", 2020.
11. Shahab A., Shafait F., Kieninger T., Dengel A., "An Open Approach Towards the Benchmarking of Table Structure Recognition Systems", Proceedings, pp. 113–120, June 2010. https://doi.org/10.1145/1815330.1815345
12. Smock B., Pesala R., Abraham R., "PubTables-1M: Towards Comprehensive Table Extraction from Unstructured Documents", Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 2022, pp. 4634–4642.
13. Smock B., Pesala R., Abraham R., "Aligning Benchmark Datasets for Table Structure Recognition", 2023.
14. Smock B., Pesala R., Abraham R., "GRITS: Grid Table Similarity Metric for Table Structure Recognition", 2023.
15. ICDAR Competition on Table Detection and Recognition (cTDaR), "Competition on Table Detection and Recognition", 2019. https://cndplabfounder.github.io/cTDaR2019/index.html

16. Vaswani A., Shazeer N., Parmar N., Uszkoreit J., Jones L., Gomez A.N., Kaiser L., Polosukhin I., "Attention Is All You Need", 2023.
17. Wang J., Lin W., Ma C., Li M., Sun Z., Sun L., Huo Q., "Robust Table Structure Recognition with Dynamic Queries Enhanced Detection Transformer", Pattern Recognition, Vol. 144, 109817, 2023. https://doi.org/10.1016/j.patcog.2023.109817
18. Xiao B., Simsek M., Kantarci B., Alkheir A., "Table Structure Recognition with Conditional Attention", March 2022.
19. Xiao B., Simsek M., Kantarci B., Alkheir A.A., "Rethinking Detection-Based Table Structure Recognition for Visually Rich Documents", 2023.
20. Xue W., Yu B., Wang W., Tao D., Li Q., "TGRNet: A Table Graph Reconstruction Network for Table Structure Recognition", 2021.
21. Yang F., Hu L., Liu X., Huang S., Gu Z., "A Large-Scale Dataset for End-to-End Table Recognition in the Wild", Scientific Data, Vol. 10, No. 1, 110, 2023.
22. Ye J., Qi X., He Y., Chen Y., Gu D., Gao P., Xiao R., "PingAn-VCGroup's Solution for ICDAR 2021 Competition on Scientific Literature Parsing Task B: Table Recognition to HTML", 2021.
23. Zhang Z., Zhang J., Du J., Wang F., "Split, Embed and Merge: An Accurate Table Structure Recognizer", Pattern Recognition, Vol. 126, 108565, 2022. https://doi.org/10.1016/j.patcog.2022.108565
24. Zheng X., Burdick D., Popa L., Zhong P., Wang N.X.R., "Global Table Extractor (GTE): A Framework for Joint Table Identification and Cell Structure Recognition Using Visual Context", Winter Conference on Applications in Computer Vision (WACV), 2021.