Spectrum of Engineering Sciences
ISSN (e) 3007-3138 (p) 3007-312X
Volume 4, Issue 5, 2026

AN INTEGRATED YOLOV8N-CRNN-DEEPSORT FRAMEWORK FOR AUTOMATIC NUMBER PLATE RECOGNITION

Qamar Ayyub[*1], **Qamar Farooq**[2], **Haroon Noor**[3], **Asad Ullah gill**[4], **Muhammad hamza Afzal**[5]

[*1,2,3,4,5]*Department of Computer Science, the superior university Lahore (Faisalabad Campus)*

[*1]qfarooq506@gmail.com

DOI: https://doi.org/10.5281/zenodo.20325339

Keywords

automatic number plate recognition; license plate recognition; YOLOv8n; CRNN; CTC loss; EasyOCR; DeepSORT; computer vision; intelligent transportation systems.

Article History

Received: 26 March 2026
Accepted: 06 May 2026
Published: 21 May 2026

Corresponding Author: *
Qamar Ayyub

Abstract

Automatic Number Plate Recognition (ANPR) supports traffic monitoring, access control, parking management and intelligent transportation analytics. A reliable ANPR system must localize plates, transcribe characters, associate repeated detections across frames and store audit-ready records. This paper presents a compact and reproducible ANPR framework that integrates YOLOv8n for plate detection, a Convolutional Recurrent Neural Network with Connectionist Temporal Classification (CRNN-CTC) for sequence recognition, EasyOCR for pseudo-labeling and fallback recognition, and DeepSORT for multi-frame association. YOLOv8n was trained as a one-class detector using 640-pixel images. Detected plate crops were passed to the CRNN recognizer. EasyOCR generated pseudo-labels where manually verified crop text was unavailable and was retained as a fallback recognizer for short CRNN outputs. DeepSORT linked detections across frames for vehicle-level logging. YOLOv8n validation produced precision of 0.980, recall of 0.943, mAP@0.50 of 0.971 and mAP@0.50:0.95 of 0.686 on 2,048 validation images containing 2,195 plate instances. A separate IoU-matched inference evaluation produced 92.32% detection accuracy/ACR, 97.46% precision, 94.58% recall and 96.00% F1 score. The CRNN recognizer achieved 86.61% exact-match training accuracy and 97.58% character-level F1 against EasyOCR pseudo-labels. The proposed workflow demonstrates a practical ANPR pipeline for research-grade intelligent transportation studies. Tracking accuracy is not claimed because temporal identity ground truth was not available.

1. Introduction

Automatic Number Plate Recognition (ANPR) is often described as an optical character recognition problem, but practical systems require more than text reading. The plate must first be localized under viewpoint variation, motion blur, shadows, vehicle occlusion and background clutter. The extracted crop must then be normalized and transcribed despite plate fonts, low contrast, dirt, reflective material and non-standard image acquisition. In video, the same vehicle can appear in multiple frames, so repeated detections must be associated before reliable logging is possible. A publishable ANPR study should therefore separate detection performance, transcription performance and tracking association rather than report a single unsupported number.

This paper reports an integrated pipeline developed from the experimental workflow submitted with this study. The detector is based on YOLOv8n, a lightweight one-stage object detector selected for efficient plate localization [1], [2]. The recognizer is based on a CRNN trained with CTC loss, a sequence-learning approach that avoids manual character segmentation [3], [4]. EasyOCR is used in two roles: it provides pseudo-labels for plate crops where manual text annotations are not available, and it functions as a fallback recognizer when the CRNN output is too short for a plausible plate string [5]. DeepSORT is used for temporal association, not as a classifier, and its role is reported separately [6].

The contribution of this study is not the invention of a new detector or a new tracking algorithm. Its contribution is the construction, evaluation and transparent reporting of a complete ANPR workflow suitable for a short original research article. The paper emphasizes defensible claims: YOLO detector validation metrics, an IoU-matched detection accuracy/ACR calculation, CRNN sequence-recognition results and a clear limitation that DeepSORT ACR cannot be computed without frame-level identity ground truth.

2. Literature Review

Modern ANPR pipelines typically combine object detection and optical character recognition. Earlier methods used hand-crafted features, morphology, edge analysis and connected components. These methods are interpretable but fragile when plates are tilted, partially occluded or captured in uneven illumination. Classical feature-based work, including Haar-like features and sliding-window plate localization, remains important historical context for understanding why learned detectors have become dominant in current traffic-vision applications [7], [8], [9].

One-stage detectors such as YOLO are widely used in real-time traffic applications because they balance detection accuracy and inference speed. YOLO-style models predict bounding boxes and class probabilities in a single forward pass, which makes them suitable for embedded or near-real-time use [1]. For ANPR, the detector is frequently trained as a one-class model because the object of interest is the license plate region rather than the vehicle class. YOLOv8 provides a practical training and deployment framework for this type of detection pipeline [2].

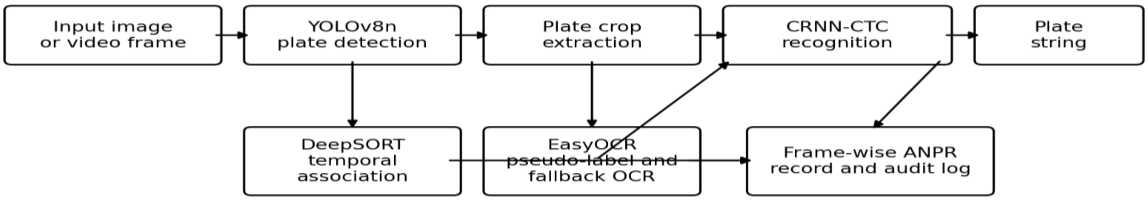
Recognition of the cropped plate region is a sequence problem. Character segmentation is difficult because plate crops may contain touching characters, non-uniform spacing and image degradation. CRNN models address this issue by using convolutional layers for spatial feature extraction, recurrent layers for ordered sequence modeling and CTC loss for alignment-free transcription [3], [4]. This design allows the recognizer to learn plate strings without requiring bounding boxes around individual characters.

EasyOCR and similar OCR systems are valuable for rapid prototyping, but pseudo-labels are not equivalent to human-verified ground truth [5]. When OCR output is used as training supervision, recognition metrics must be interpreted as agreement with pseudo-labels rather than independent real-world transcription accuracy. The present paper follows that conservative interpretation.

Tracking-by-detection methods connect object detections over time. SORT introduced a simple and efficient online tracking framework [10], while DeepSORT extends the approach by combining motion estimation with a deep appearance association metric [6]. In an ANPR pipeline, tracking helps reduce duplicate logs and maintain stable vehicle identities across consecutive frames. Tracking accuracy metrics such as MOTA, IDF1 or identity-switch rate require temporal ground truth. If such annotations are absent, reporting only the tracking configuration is more defensible than claiming a tracking ACR.

3. Research Methodology

The proposed workflow contains four modules: plate detection, crop generation, text recognition and temporal association. The detector receives raw traffic images or video frames and outputs plate bounding boxes. The crop-generation module extracts the predicted plate regions. The recognizer converts each crop into a text string. The tracking module links detections across frames and produces a structured record containing frame index, track identifier, bounding box and recognized plate string.



Pipeline output: track ID, bounding box, frame index, recognized text, and audit-ready log row

Fig. 1. System architecture of the YOLOv8n-CRNN-EasyOCR-DeepSORT ANPR pipeline.

3.1. YOLOv8n Plate Detection

YOLOv8n was trained for one-class plate localization. Images were resized to 640 pixels during detector training and inference. The detector predicts plate bounding boxes and confidence scores. During evaluation, predicted boxes are matched to ground-truth boxes using an Intersection over Union (IoU) threshold. A prediction is counted as a true positive when the predicted box overlaps the target plate region above the selected threshold and the prediction is not a duplicate assignment.

The detector output is the foundation of the whole ANPR system. Poor localization directly harms recognition because the crop can miss characters or include distracting background. For that reason, detector metrics are reported separately from recognition metrics. This separation also prevents the paper from hiding detector errors inside an overall pipeline score.

3.2. CRNN-CTC Plate Recognition

The recognition module uses a CRNN architecture. Convolutional layers extract visual features from the plate crop. The resulting feature sequence is passed to recurrent layers to model left-to-right character dependencies. A final prediction layer outputs per-time-step character probabilities. CTC loss learns the alignment between visual features and target plate strings without requiring individual character boxes.

The recognizer was trained using plate crop transcripts generated or assisted by EasyOCR. During inference, CRNN output is decoded into a candidate plate string. If the decoded text is too short for a plausible plate, EasyOCR is used as a fallback. This hybrid decision rule improves robustness in cases where the custom recognizer returns incomplete text, but it also means that human-verified OCR test labels are still needed before claiming definitive real-world transcription accuracy.

3.3. DeepSORT Temporal Association

DeepSORT stands for Simple Online and Realtime Tracking with a Deep Association Metric. In this pipeline, DeepSORT receives plate detections and associates them across sequential frames. The tracker uses motion prediction and association logic to maintain track identities and reduce repeated logging of the same vehicle across consecutive frames.

DeepSORT is not an OCR model and does not improve the character classifier by itself. Its function is temporal organization. Therefore, this paper reports DeepSORT configuration and system role but does not report DeepSORT ACR, MOTA, MOTP or IDF1. Those values require annotated temporal identity ground truth, which was not present in the experimental workflow. Claiming a DeepSORT accuracy number without identity annotations would be methodologically invalid.

3.4. Workflow Summary

Table 1. Functional stages of the proposed ANPR pipeline.

Stage	Input	Output	Purpose
YOLOv8n detector	Traffic image or video frame	Plate bounding box	Localize number plate region
Crop extraction	Frame and bounding box	Plate crop image	Prepare recognizer input
CRNN-CTC recognizer	Plate crop	Predicted plate string	Primary text recognition
EasyOCR fallback	Plate crop	Alternative text string	Pseudo-labeling and fallback recognition
DeepSORT tracker	Frame-wise detections	Track ID and log rows	Associate repeated observations over time

4. Experimental Setup and Evaluation Metrics

The detector was evaluated on 2,048 validation images containing 2,195 plate instances. Standard detector metrics were calculated, including precision, recall, mAP@0.50 and mAP@0.50:0.95. A separate final inference evaluation used IoU matching to calculate detection accuracy/ACR, precision, recall and F1 score. The CRNN recognizer was evaluated using exact-match accuracy and character-level F1 against the pseudo-label text available in the experiment. Object-detection metrics follow common benchmark reporting conventions used in PASCAL VOC and COCO-style evaluations [11], [12].

Accuracy/ACR in this paper refers to the final detection accuracy calculated from IoU-matched plate detection outcomes. It does not refer to DeepSORT tracking. DeepSORT ACR is deliberately not reported because the experiment does not include identity-labeled trajectories. Recognition accuracy refers to the CRNN exact-match rate at the string level, while character-level F1 measures agreement at the character level and is less strict than full-plate exact matching.

For reproducibility, the final submission should include optimizer choice, learning rate schedule, batch size, epoch count, data split policy and augmentation settings in the methods section or supplementary notebook. If Adam or another adaptive optimizer is used, it should be reported explicitly rather than implied [13].

Table 2. Evaluation metric definitions and reporting scope.

Metric	Definition in this paper	Reported for
Precision	TP / (TP + FP)	YOLOv8n detector
Recall	TP / (TP + FN)	YOLOv8n detector
F1 score	Harmonic mean of precision and recall	YOLOv8n final detection run
mAP@0.50	Mean average precision at IoU 0.50	YOLOv8n validation
mAP@0.50:0.95	Mean average precision averaged over IoU thresholds	YOLOv8n validation
Exact-match accuracy	Full predicted string equals target string	CRNN recognizer
Character-level F1	Character-level sequence agreement	CRNN recognizer
Tracking ACR	Requires identity ground truth	Not reported

5. Results and Analysis

The YOLOv8n detector achieved strong localization performance on the validation split. The validation precision was 0.980 and recall was 0.943, showing that the detector produced few false plate detections while retaining most true plate instances. The mAP@0.50 value of 0.971 indicates high localization quality under the commonly used 0.50 IoU threshold. The mAP@0.50:0.95 value of 0.686 is lower, as expected, because stricter

IoU thresholds penalize small localization errors more heavily.

The final IoU-matched inference evaluation produced a detection accuracy/ACR of 92.32%. In the same run, precision was 97.46%, recall was 94.58% and F1 score was 96.00%. These numbers support the detector as a strong plate localization component for the proposed pipeline. They should not be interpreted as end-to-end legal enforcement accuracy because OCR and tracking evidence are separate components.

Table 3. YOLOv8n detector performance.

Evaluation item	Value
YOLOv8n validation images	2,048
YOLOv8n validation plate instances	2,195
Validation precision	0.980
Validation recall	0.943
Validation mAP@0.50	0.971
Validation mAP@0.50:0.95	0.686
Final detection accuracy/ACR	92.32%
Final detection precision	97.46%
Final detection recall	94.58%
Final detection F1 score	96.00%

The CRNN recognizer achieved 86.61% exact-match training accuracy and 97.58% character-level F1 against the EasyOCR-assisted pseudo-label set. The difference between exact-match accuracy and character-level F1 is meaningful. Exact-match accuracy requires every character in the plate string to be correct, so a single wrong character makes the whole prediction incorrect. Character-level F1 is more tolerant and shows that most characters were correctly recognized even when full-string agreement failed.

Table 4. Recognition performance and interpretation.

Recognition metric	Value	Interpretation
CRNN exact-match accuracy	86.61%	Full plate string matched pseudo-label
CRNN character-level F1	97.58%	Most individual characters matched pseudo-label
EasyOCR role	Pseudo-label + fallback	Assists training and fallback recognition
Independent human-labeled OCR test accuracy	Not available	Required for stronger publication claims

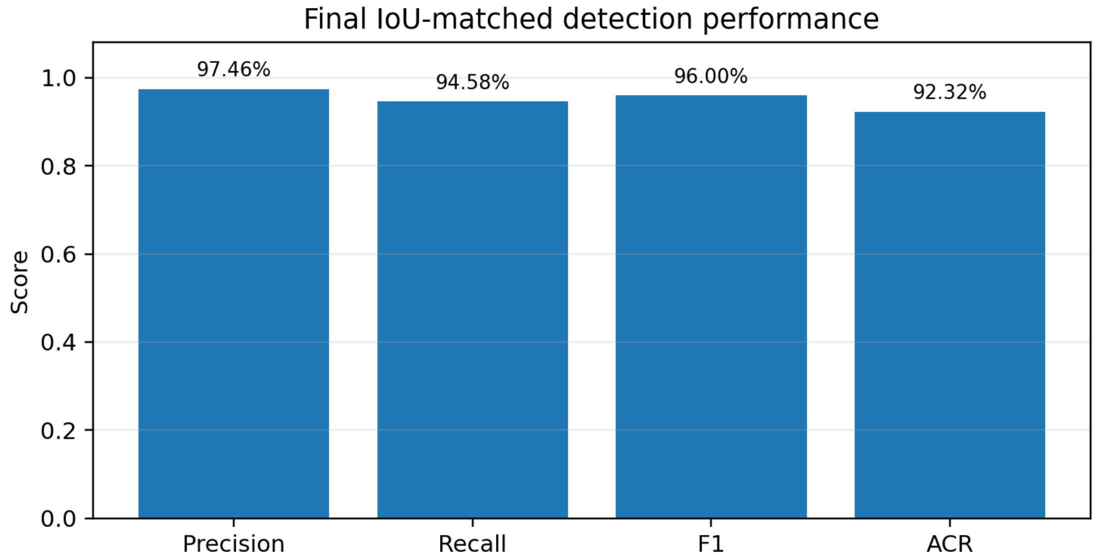


Fig. 2. Final detection metrics from the IoU-matched evaluation run.

DeepSORT provides the structural mechanism for assigning persistent identifiers to detections across frames, but no numerical tracking accuracy is reported. This is a strength of the reporting rather than a weakness of the implementation: tracking metrics require manually annotated identity labels over video sequences. Without that annotation layer, a numerical DeepSORT ACR would be fabricated or misleading.

6. Discussion

The results indicate that the proposed system is strongest at plate localization and character-level recognition. The detector achieved high precision and recall, which is important because false crops increase OCR load and missed plates prevent any downstream recognition. The recognizer results show that the CRNN learned useful sequence patterns, but exact-match accuracy remains lower than character-level F1. For ANPR deployment, exact-match accuracy is the stricter and more operationally relevant metric because a single wrong character can identify a different vehicle.

The use of EasyOCR is practical but must be reported carefully. EasyOCR-assisted labels reduce annotation effort and allow fast model development, but they introduce label noise. Therefore, CRNN metrics in this paper are best understood as agreement with OCR-assisted pseudo-labels. A future version of this study should include a manually verified OCR test set. That addition would strengthen the paper for reviewers because it would separate model performance from pseudo-label quality.

DeepSORT adds value by reducing duplicate logging and giving the pipeline video-awareness. Yet its evaluation remains incomplete until identity ground truth is added. A stronger future experiment would annotate vehicle identities across selected video sequences and report MOTA, IDF1, identity switches and track fragmentation. Until then, the most accurate statement is that DeepSORT was integrated for association but not quantitatively evaluated as an identity tracker.

For Y-category journal submission, the paper should be positioned as an applied computer-vision system evaluation. The novelty is the integrated workflow and transparent component-level measurement, not a claim of architectural originality. The final target journal should be verified in the HEC Journal Recognition System at the date of submission [14], [15].

7. Limitations

1. The OCR labels used for CRNN training and evaluation were assisted by EasyOCR; independent human-labeled OCR test data were not available.
2. DeepSORT tracking accuracy, ACR, IDF1 and identity-switch rate are not reported because temporal identity ground truth was not available.

3. The reported performance should not be generalized to all plate designs, countries, lighting conditions or enforcement settings without additional external validation.
4. The study does not include calibrated speed measurement, violation detection or legal decision automation.
5. Dataset redistribution must be checked against license and privacy constraints before journal upload.

8. Ethical Considerations

ANPR systems process vehicle identifiers and can affect privacy, surveillance and administrative decision-making. The present paper is framed as a research system, not an automated enforcement tool. Before real deployment, the dataset source, consent or legal basis, retention period, access control, anonymization policy and audit procedure must be documented.

Published figures should avoid exposing identifiable plate strings unless the images are synthetic, anonymized or explicitly permitted by the dataset license.

Any operational use should include human review, error appeal procedures and jurisdiction-specific compliance with data-protection rules. The model outputs should not be used as sole evidence for legal enforcement without independent validation, calibrated imaging hardware and approved governance procedures.

9. Conclusion and Future Work

This paper presented a compact ANPR framework integrating YOLOv8n detection, CRNN-CTC recognition, EasyOCR-assisted pseudo-labeling/fallback recognition and DeepSORT temporal association. The detector achieved validation precision of 0.980, recall of 0.943, mAP@0.50 of 0.971 and mAP@0.50:0.95 of 0.686. The final detection evaluation produced 92.32% accuracy/ACR, 97.46% precision, 94.58% recall and 96.00% F1 score. The CRNN recognizer achieved 86.61% exact-match accuracy and 97.58% character-level F1 against OCR-assisted pseudo-labels.

The main conclusion is that the pipeline is suitable for research-grade ANPR experimentation, especially when results are reported at component level. Future work should add manually verified OCR labels, temporal identity annotations for DeepSORT evaluation, cross-dataset testing and ablation studies comparing CRNN-only, EasyOCR-only and hybrid recognition modes.

Funding: [Insert funding statement. If no external funding was received, write: The author(s) received no external funding for this study.]

Conflict of interest: The author(s) declare no conflict of interest.

References

J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, “You Only Look Once: Unified, Real-Time Object Detection,” in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA: IEEE, Jun. 2016, pp. 779–788. doi: 10.1109/CVPR.2016.91.

Ultralytics, ‘YOLOv8 Documentation,’ Ultralytics Docs. [Online]. Available: https://docs.ultralytics.com/models/yolov8/. Accessed: May 19, 2026.”

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, doi: 10.1109/TPAMI.2016.2646371.

A. Graves, S. Fernández, F. Gomez, and J. Schmidhuber, “Connectionist temporal classification: labelling unsegmented sequence data with recurrent neural networks,” in *Proceedings of the 23rd international conference on Machine learning - ICML ’06*, Pittsburgh, Pennsylvania: ACM Press, 2006, pp. 369–376. doi: 10.1145/1143844.1143891.

"JaidedAI, 'EasyOCR: Ready-to-use OCR with multilingual support,' GitHub repository. [Online]. Available: https://github.com/JaidedAI/EasyOCR. Accessed: May 19, 2026."

N. Wojke, A. Bewley, and D. Paulus, "Simple online and realtime tracking with a deep association metric," in *2017 IEEE International Conference on Image Processing (ICIP)*, Beijing: IEEE, Sep. 2017, pp. 3645–3649. doi: 10.1109/ICIP.2017.8296962.

R. Lienhart and J. Maydt, "An extended set of Haar-like features for rapid object detection," in *Proceedings. International Conference on Image Processing*, Rochester, NY, USA: IEEE, 2002, p. I-900-I–903. doi: 10.1109/ICIP.2002.1038171.

K. Deb, H.-U. Chae, and K.-H. Jo, "Vehicle License Plate Detection Method Based on Sliding Concentric Windows and Histogram," *J. Comput.*, vol. 4, no. 8, pp. 771–777, Aug. 2009, doi: 10.4304/jcp.4.8.771-777

S. Du, M. Ibrahim, M. Shehata, and W. Badawy, "Automatic License Plate Recognition (ALPR): A State-of-the-Art Review," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 23, no. 2, pp. 311–325, Feb. 2013, doi: 10.1109/TCSVT.2012.2203741.

A. Bewley, Z. Ge, L. Ott, F. Ramos, and B. Upcroft, "Simple online and realtime tracking," in *2016 IEEE International Conference on Image Processing (ICIP)*, Phoenix, AZ, USA: IEEE, Sep. 2016, pp. 3464–3468. doi: 10.1109/ICIP.2016.7533003.

M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman, "The Pascal Visual Object Classes (VOC) Challenge," *Int. J. Comput. Vis.*, vol. 88, no. 2, pp. 303–338, Jun. 2010, doi: 10.1007/s11263-009-0275-4.

T.-Y. Lin *et al.*, "Microsoft COCO: Common Objects in Context," in *Computer Vision – ECCV 2014*, vol. 8693, D. Fleet, T. Pajdla, B. Schiele, and T. Tuytelaars, Eds., in Lecture Notes in Computer Science, vol. 8693. , Cham: Springer International Publishing, 2014, pp. 740–755. doi: 10.1007/978-3-319-10602-1_48.

D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," 2014, *arXiv*. doi: 10.48550/ARXIV.1412.6980.

"Higher Education Commission Pakistan, 'HEC Journals and Publication Policy 2024.' [Online]. Available: https://www.hec.gov.pk/english/services/faculty/journals/Pages/default.aspx. Accessed: May 19, 2026."

"Higher Education Commission Pakistan, 'HEC Journal Recognition System (HJRS).' [Online]. Available: https://hjrs.hec.gov.pk/. Accessed: May 19, 2026."