Real-Time Human Detection and Tracking using YOLO and DeepSORT on Jetson Nano

Muhammad Ali, Waqas Jahangir, Hassan Akhtar
Department of Electrical Engineering, School of Electrical Engineering and Computer Science (SEECS)
National University of Sciences and Technology (NUST), Islamabad, Pakistan
Email: {muali, wjahangir, hakhtar}.msee24seecs@seecs.edu.pk

I. Introduction

Human detection and tracking are vital tasks in computer vision, enabling intelligent surveillance, crowd monitoring, and autonomous navigation. With the advent of powerful yet compact embedded AI hardware such as the NVIDIA Jetson Nano, it has become possible to perform real-time inference at the edge. The motivation behind this work is to develop a lightweight, accurate, and deployable solution for real-time human tracking that maintains high frame rates and reliability under constrained computational resources. This report presents a focused literature review on YOLO-based detection and DeepSORT-based tracking, evaluates recent advances in the field, and assesses the feasibility of deploying the proposed system on Jetson Nano.

The proposed system pipeline is illustrated in Fig1 1, which shows the block diagram of the real-time human detection and tracking process using YOLOv5 and DeepSORT on Jetson Nano. The process includes frame preprocessing, YOLO-based detection optimized with TensorRT, and DeepSORT tracking using Kalman filtering and the Hungarian algorithm.

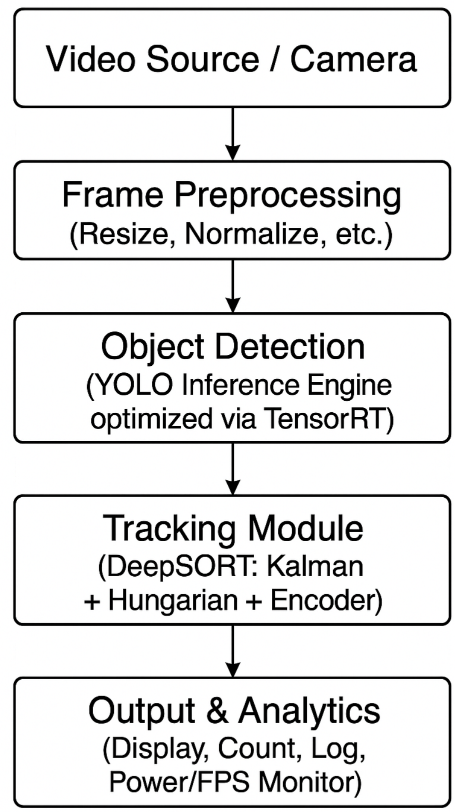

Fig. 1. System block diagram of the proposed real-time human detection and tracking pipeline using YOLOv5 and DeepSORT on Jetson Nano.

II. Literature Review

Human detection and tracking remain core challenges in computer vision, especially when real-time inference is required on resource-constrained edge devices such as the NVIDIA Jetson Nano. The integration of efficient deep learning models with optimized tracking algorithms enables real-world deployment for applications like surveillance, crowd analysis, and autonomous robotics. This literature review synthesizes ten peer-reviewed works, highlighting classical and state-of-the-art methods in detection and tracking, recent advances in edge deployment, and the foundational survey that contextualizes this research domain.

The foundation of the proposed project draws directly from Razzok *et al.* [1], who developed a *Pedestrian Detection and Tracking System* integrating YOLOv5 with Deep-SORT and a novel data association metric. Their approach achieved higher robustness under occlusion and dynamic backgrounds compared to standard SORT-based pipelines. The use of YOLOv5 allowed lightweight yet accurate detection, while their improved metric reduced identity switches in crowded scenes. This paper forms the implementation basis of the present work, as it combines a modern detector with efficient tracking for human motion analysis in real time.

For detection performance at the cutting edge, Wang *et al.* [2] introduced YOLOv7, representing a major leap in real-time detection through its “trainable bag-of-freebies” strategy. It incorporated extended efficient layer aggregation networks (E-ELAN) and auxiliary heads, leading to superior accuracy-speed tradeoffs. YOLOv7 achieved state-of-the-art mean average precision (mAP) across MS COCO benchmarks while maintaining real-time speeds on GPUs. Its modular and scalable design makes it highly adaptable for deployment on embedded systems, aligning closely with Jetson-based human detection applications.

In multi-object tracking (MOT), Zhang *et al.* [3] proposed

ByteTrack, which addressed the limitation of conventional association frameworks that ignored low-confidence detections. By linking every detection box, including those with low confidence, ByteTrack achieved superior tracking accuracy on MOT17 and MOT20 datasets, outperforming DeepSORT and FairMOT. Its simplicity and compatibility with any detector make it an exemplary state-of-the-art approach for robust tracking in dynamic environments.

A broader theoretical overview is provided by Luo *et al.* [4], who compiled a comprehensive survey on multiple object tracking in the age of deep learning. This survey systematically analyzed the evolution from traditional Kalman-filter-based trackers to end-to-end deep learning models. It categorized MOT pipelines into detection-based and joint-detection-tracking frameworks, emphasizing challenges such as occlusion handling, real-time constraints, and re-identification. This paper serves as the principal background reference, situating the proposed project within contemporary MOT research trends.

Recent works further advance this foundation. Shim *et al.* [5] explored attention-based data association in online multi-object tracking, achieving improved temporal consistency by weighting historical track confidence. Gomes *et al.* [6] demonstrated a practical implementation on Jetson Nano using YOLO and V-IOU tracking for counting people and bicycles. Their edge-optimized selective frame downsampling strategy reduced latency and power consumption, confirming the feasibility of deploying detection and tracking models on embedded hardware. Similarly, Humes *et al.* [7] presented Squeezed Edge YOLO, introducing an extremely compact model optimized for onboard inference with minimal accuracy loss, illustrating the trend toward lightweight detectors for constrained devices.

Application-specific studies like Punn *et al.* [8] used fine-tuned YOLOv3 and DeepSORT to monitor social distancing during COVID-19, highlighting the versatility of combined detection–tracking systems in real-world surveillance. Complementing these efforts, Han *et al.* [9] proposed a scalable and fast YOLO optimized for edge computing through structural pruning and layer quantization, reducing computational overhead while maintaining high mAP. Finally, Ganesh *et al.* [10] developed YOLO-ReT, which enhanced feature aggregation and contextual learning to balance accuracy and inference latency on edge GPUs, confirming the potential for high-performance object detection in compact embedded settings.

Collectively, these studies demonstrate rapid progress in efficient detection and tracking for edge deployment. The reviewed works reveal three key research trends: (1) continuous optimization of YOLO architectures for real-time operation; (2) enhanced tracking reliability through improved data association metrics; and (3) increased focus on energy-efficient inference for embedded AI platforms. Building upon these insights, the proposed project aims to implement and optimize the YOLOv5–DeepSORT-based pedestrian tracking system for the Jetson Nano, bridging accuracy and efficiency in real-world human detection scenarios.

TABLE I
SUMMARY OF CORE REVIEWED PAPERS

No.	Category	Paper Title	Venue / Year
1	**Implementation***	Pedestrian Detection and Tracking System Based on Deep-SORT and YOLOv5	MDPI Information, 2023
2	State of the Art in Detection	YOLOv7: Trainable Bag-of-Freebies for Real-Time Object Detectors	arXiv, 2023
3	State of the Art in Tracking	ByteTrack: Multi-Object Tracking by Associating Every Detection Box	ECCV, 2022
4	Survey Paper	A Survey of Deep Learning for Object Detection	Elsevier, 2020

A. *Feasibility for Semester Project*

*1) **Implementation Scope**:*

YOLOv5 + DeepSORT [1]:

- Most balanced and practical technique for real-time tracking.
- Pre-trained YOLOv5s/n weights can be deployed on Jetson Nano via **TensorRT** or **ONNX** optimization.
- DeepSORT's embedding model and Kalman-based tracking require moderate compute and fit comfortably in 4 GB RAM.
- Achieves **15–20 FPS** on optimized setups.

*2) **Project Duration and Resources**:*

Duration: 10 weeks (part-time, 3 students) are allocated for implementation, optimization, and documentation. The detailed week-wise plan is presented in table no II.

TABLE II
PROJECT TIMELINE (10 WEEKS)

Week	Activities / Deliverables
1	Conduct literature review and finalize key research papers. Define objectives and overall architecture of the YOLO-based detection and tracking pipeline.
2	Jetson Nano setup: flash OS image, configure environment, and verify GPU and camera compatibility.
3	Install dependencies (PyTorch, CUDA, cuDNN, TensorRT, OpenCV). Validate YOLOv5 model inference on test images.
4-5	Implement YOLOv5 + DeepSORT pipeline for real-time pedestrian detection and tracking using video streams.
6-7	Optimize model with TensorRT and pruning techniques for real-time edge inference. Test with USB/CSI camera input.
8	Conduct extended validation on dynamic scenes with varying lighting and occlusions. Record and log test performance metrics.
9	Refine the pipeline, prepare visual outputs (bounding boxes, tracking IDs), and perform comparative analysis between methods.
10	Final documentation, report writing, lifecycle analysis, and preparation of presentation for evaluation.

Human Resources and Roles:

- **Muhammad Ali** – MATLAB simulations, algorithm design, and performance analysis.

- **Waqas Jahangir** – Jetson Nano setup, embedded optimization, and camera integration.
- **Hassan Akhtar** – Documentation, report preparation, and system lifecycle coordination.

Hardware / Software Requirements:

- NVIDIA Jetson Nano (4 GB) DevKit (Provided by Sir)
- 5 V / 5 A power supply (Provided by Sir)
- 32-GB microSD card
- USB or CSI camera module (Already available)
- Python 3.8, PyTorch $\geq$ 1.10, OpenCV, TensorRT, Deep-SORT repository

III. Declaration

ChatGPT (Deep Research Mode) was used to search, verify, and summarize research papers and to generate the narrative literature review. All links were opened and confirmed by the team. All reviewed papers were added to the GitHub repository under `/docs/references`. The final narrative was refined and validated by the project authors.

IV. Project Resources

The source code, documentation, and project management details are available on GitHub: **GitHub Repository:** https://github.com/M-A-S1/Real-Time-Human-Detection-and-Tracking-Using-Jetson-Nano
GitHub Project: https://github.com/users/M-A-S1/projects/2

References

[1] M. Razzok, A. Badri, I. El Mourabit, Y. Ruichek, and A. Sahel, "Pedestrian Detection and Tracking System Based on Deep-SORT, YOLOv5, and New Data Association Metrics," *Information*, vol. 14, no. 4, 2023.

[2] C.-Y. Wang, A. Bochkovskiy, and H.-Y. M. Liao, "YOLOv7: Trainable Bag-of-Freebies Sets New State-of-the-Art for Real-Time Object Detectors," *arXiv preprint arXiv:2207.02696*, 2023.

[3] Y. Zhang *et al.*, "ByteTrack: Multi-Object Tracking by Associating Every Detection Box," in *Proc. ECCV*, 2022.

[4] W. Luo *et al.*, "A Survey on Multiple Object Tracking in the Age of Deep Learning," *Artificial Intelligence*, 2021.

[5] C. Shim *et al.*, "Focusing on Tracks for Online Multi-Object Tracking," in *Proc. CVPR*, 2025.

[6] H. Gomes, N. Redinha, N. Lavado, and M. Mendes, "Counting People and Bicycles in Real Time Using YOLO on Jetson Nano," *Energies*, vol. 15, no. 23, 2022.

[7] E. Humes, M. Navardi, and T. Mohsenin, "Squeezed Edge YOLO: Onboard Object Detection on Edge Devices," *arXiv preprint arXiv:2312.11716*, 2023.

[8] N. S. Punn, S. K. Sonbhadra, and S. Agarwal, "Monitoring COVID-19 Social Distancing with Person Detection and Tracking via Fine-Tuned YOLOv3 and DeepSORT," *Applied Intelligence*, 2020.

[9] B.-G. Han, J.-G. Lee, K.-T. Lim, and D.-H. Choi, "Design of a Scalable and Fast YOLO for Edge-Computing Devices," *Electronics*, vol. 9, no. 11, 2020.

[10] P. Ganesh, Y. Chen, Y. Yang, D. Chen, and M. Winslett, "YOLO-ReT: Towards High Accuracy Real-Time Object Detection on Edge GPUs," *arXiv preprint arXiv:2110.13713*, 2021.