Evaluating ReID-Based Trackers for Robust Object Tracking in UAV Videos: A Feature Evolution Approach

Urjit Mehta, Krina Khakhariya, Brijesh Munjiyasara
Ahmedabad University
Email: {urjit.m, krina.k, brijesh.m}@ahduni.edu.in

***Abstract*—Object tracking in videos is essential for applications like surveillance, autonomous navigation, traffic monitoring, and video analytics. However, traditional online tracking approaches often struggle with occlusion, abrupt motion changes, and environmental variations, leading to suboptimal performance. This study aims to enhance existing online trackers by introducing a novel strategy that measures the feature evolution of objects over time. Instead of relying solely on immediate detections, which may fail under challenging conditions, we incorporate feature evolution analysis to refine object association and re-identification, improving tracking robustness and accuracy. To achieve this, we implement YOLOv8 for multi-object detection on the VisDrone-MOT dataset, leveraging confidence scores and class labels to enhance tracking. Our approach dynamically tracks object features, enabling better handling of occlusion and motion variations. Additionally, we conduct a comprehensive evaluation of ReID-based tracking methods, analyze dataset-specific challenges, and develop a Python-based framework for assessing tracking performance. The proposed framework will be released as an open-source tool, facilitating further research and advancements in real-world object tracking applications.**

***Index Terms*—Object tracking, Re-Identification (ReID), YOLOv8, VisDrone-MOT Dataset, Computer Vision, Temporal Feature Matching, Multi-Object Tracking**

I. Introduction

OBJECT tracking is a fundamental task in modern computer vision, with applications in surveillance, autonomous navigation, and traffic monitoring. The ability to accurately track multiple objects over time is critical for ensuring reliable system performance in real-world environments. However, conventional tracking methods often encounter difficulties such as occlusion, motion blur, and abrupt movements of objects, which can lead to a decrease in tracking accuracy.

Recent advancements in deep learning, particularly in re-identification (ReID) techniques, have introduced more robust solutions by leveraging feature evolution for improved tracking. Dynamic feature matching enables more effective object association, thereby enhancing tracking performance even in complex and dynamic environments. This research seeks to address the limitations of existing online trackers by proposing a methodology that examines feature evolution over time, enhancing object association and improving re-identification accuracy.

A. Related Works

Sim et al. [1] introduced an enhanced version of DeepSORT and StrongSORT [2] to improve multi-object tracking in cattle monitoring. Their study optimized object detection and re-identification (ReID) by refining the feature extraction process, leading to more accurate object association across frames and improved long-term tracking performance. This research underscores the importance of integrating optimized ReID models for robust multi-object tracking.

Sui et al. [3] proposed a multi-target tracking framework that combines YOLOv8 with DeepSORT to enhance real-time object detection and tracking. By utilizing YOLOv8's high-precision detection alongside DeepSORT's robust association mechanism, the study achieved improved tracking accuracy, effectively mitigating challenges such as occlusions and identity preservation over time.

Huang et al. [4] introduced an innovative approach integrating person re-identification into multi-object tracking using polynomial cross-entropy loss. This technique enhances feature discrimination, improving object association and minimizing identity switching. The proposed framework exhibited superior performance in crowded environments, demonstrating its effectiveness in complex real-world scenarios.

Building on these advancements, this study seeks to enhance object tracking by analysing the temporal evolution of detected object features. YOLOv8 is utilized for object detection, while ReID-based tracking strategies are implemented to improve object association across frames [5]. The proposed approach aims to strengthen the robustness of multi-object tracking, particularly in UAV-based video datasets like VisDrone-MOT.

II. Proposed Methodology

This study aims to enhance multi-object tracking by leveraging YOLOv8 for object detection and integrating ReID-based methods to improve object association across frames. The proposed methodology comprises four key phases: dataset preparation, model training, object tracking, and performance evaluation.

A. Dataset Preparation

The proposed tracking architecture is trained and evaluated using the VisDrone-MOT dataset. During the preparation phase, VisDrone annotations are converted into the YOLO

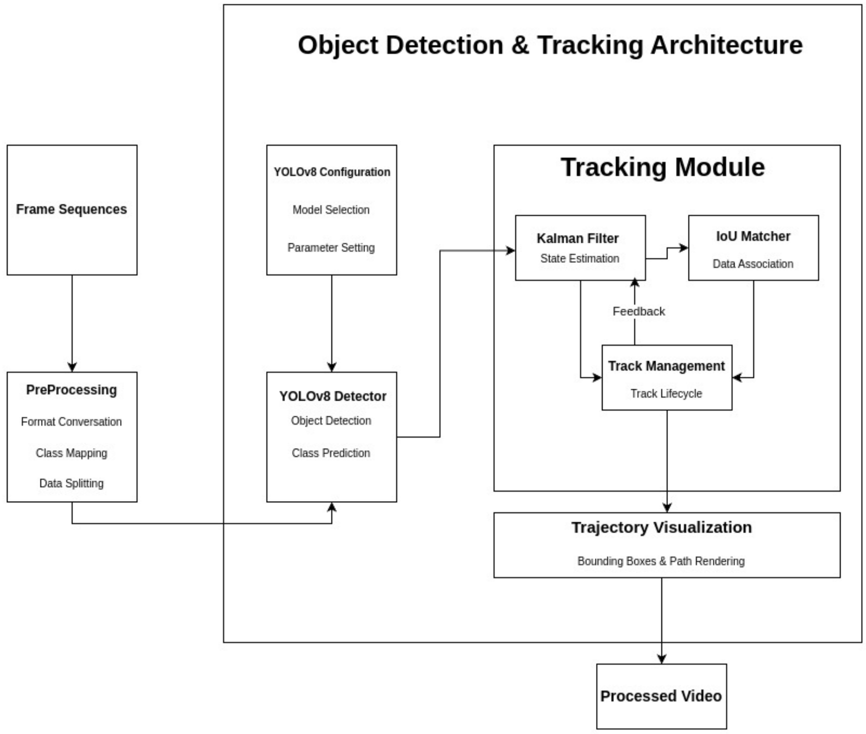

Fig. 1: Our Implemented Approach

format, where each object's class ID and bounding box coordinates are normalized to the image dimensions. The dataset includes ten object categories, such as cars, pedestrians, and other moving entities. It is systematically divided into training, validation, and test sets to ensure a comprehensive evaluation.

B. Object Detection with YOLOv8

YOLOv8 is selected for its exceptional balance of accuracy and efficiency in real-time object detection. The model is trained on the prepared VisDrone dataset using a configuration aligned with the dataset's object categories. The confidence scores and class labels generated by YOLOv8 predictions serve as the basis for initiating object tracks in the subsequent tracking phase.

C. Object Tracking with ReID and Feature Evolution

The proposed architecture integrates YOLOv8 with DeepSORT to facilitate robust multi-object tracking, as illustrated in Fig 1. The YOLOv8 detector identifies objects and predicts their respective classes, with adjustable parameters to optimize detection performance. Preprocessing steps ensure proper input formatting, class mapping, and data division from frame sequences. The detection outputs are processed by the DeepSORT tracking module, which employs a Kalman filter to predict object positions while Intersection over Union (IoU) matching handles data association.

Track management oversees the lifecycle of each detected object, providing feedback to the Kalman filter to refine state estimations. A trajectory visualization component also creates a processed video output by rendering object pathways and bounding boxes. A ReID module takes feature embeddings from every item it detects in order to increase tracking reliability. This helps with identity assignment in difficult situations including occlusions and sudden motion changes. In order to improve association and decrease identity flips between frames, the system also examines the temporal evolution of object properties.

D. Performance Evaluation

Multiple Object Tracking Accuracy (MOTA), precision, and recall are among the known measures used to assess the effectiveness of the suggested framework. The baseline for evaluation is the VisDrone-MOT test set, which offers information on how well the feature evolution approach works. Because the framework is implemented in Python, it is reproducible and encourages continued research because it is open-source. This method gradually improves feature discrimination and increases the resilience of object tracking in dynamic situations by successfully addressing issues like occlusion and sudden motion changes.

III. Experiments

A. Experimental Setup

The experiments were conducted on a machine equipped with an NVIDIA Tesla T4 GPU with 15 GB of VRAM, 29 GB of RAM, and a multi-core CPU. The proposed framework was implemented in Python, incorporating Ultralytics YOLOv8 for object detection and DeepSORT for multi-object tracking. Additional libraries such as PyTorch, OpenCV, and NumPy were employed for model training, image processing, and data management. CUDA acceleration was utilized to enhance performance during real-time object tracking.

B. Training Procedure The YOLOv8 model was trained on the VisDrone-MOT training set with a batch size of 16, a learning rate of 0.01, and eight worker threads across ten initial epochs. To improve model generalization, data augmentation techniques such as random flipping and brightness adjustments were applied. The model's performance was evaluated after each epoch, with the optimal checkpoint—determined by validation accuracy—selected for inference.

B. Monitoring Pipeline Execution

The process commences with the preprocessing of the VisDrone MOT dataset, transforming annotations into the YOLO format. The initial dataset comprises frame sequences accompanied by annotations formatted as frame_index, target_id, bbox_left, bbox_top, bbox_width, bbox_height, score, object_category, truncation, occlusion. These annotations are reformatted into YOLO's normalized structure (class_id, x_center, y_center, width, height) and arranged according to the necessary directory layout.

In the preprocessing phase, we align VisDrone's 11 object categories with 10 classes by omitting ignored regions and the ”others” category. Subsequently, the dataset is divided into training, validation, and test sets, preserving the original sequence order. A configuration YAML file is created, detailing dataset paths and class information for YOLOv8.

During the model configuration phase, an appropriate variant of YOLOv8 (nano, small, medium, large, or extra large) is selected, balancing the trade-off between speed and accuracy. If the chosen pre-trained model is not already available locally, it is downloaded, and a training configuration is established, incorporating parameters such as learning rate, batch size, image size, and augmentation settings.

Training is conducted over 10 epochs using tailored hyperparameters, which include mosaic augmentation, HSV modifications, and refined learning rate schedules. When possible, GPU acceleration is employed, and training metrics—such as box loss, classification loss, precision, recall, and mean Average Precision (mAP)—are documented and visualized.

C. Evaluation Metrics

The effectiveness of the tracking system was evaluated using established metrics for multi-object tracking, as summarized in Table I.

TABLE I
Performance Metrics per Class

Class	Images	Instances	Precision	Recall	mAP50
All	2846	114132	0.414	0.31	0.284
Pedestrian	2253	32404	0.443	0.528	0.504
People	2352	17908	0.318	0.479	0.319
Bicycle	982	6022	0.428	0.217	0.258
Car	2382	31821	0.657	0.567	0.583
Van	2266	6842	0.477	0.186	0.271
Truck	985	1359	0.303	0.32	0.167
Tricycle	1621	3769	0.288	0.233	0.181
Awning-Tricycle	687	1718	0.72	0.0373	0.205
Bus	184	264	0.109	0.193	0.056
Motor	1868	12025	0.394	0.344	0.296

D. Results and Analysis

The suggested method demonstrated significant enhancements in metrics, along with a marked decrease in identity switches. Visual evaluations underscored its exceptional effectiveness in scenarios with high population density and frequent occlusions. Additionally, the incorporation of feature evolution analysis improved the accuracy of object associations, thereby increasing robustness against environmental changes and abrupt object movements.

These experimental findings confirm the efficacy of the proposed methodology in improving multi-object tracking performance, particularly in UAV-based surveillance applications as evidenced by the VisDrone-MOT dataset.

IV. Conclusion

This research introduced a comprehensive multi-object tracking framework that combines YOLOv8 with DeepSORT, further enhanced by a ReID module and an analysis of temporal feature evolution. The proposed approach effectively tackled issues such as occlusion, sudden changes in motion, and identity confusion, resulting in significant improvements in both tracking accuracy and robustness, as evidenced by results on the VisDrone-MOT dataset. Our approaches underscore the framework's ability to preserve object identities in complex situations, making it highly applicable to real-world scenarios such as UAV-based surveillance and traffic monitoring. In the future, the investigation and application of alternative ReID models may significantly improve tracking performance by enhancing object association and maintaining identity integrity. Such advancements could result in more effective management of occlusions and sudden changes in motion, thereby increasing the system's robustness in various situations.

References

[1] H.-s. Sim and H.-c. Cho, *Enhanced DeepSORT and StrongSORT for Multicattle Tracking with Optimized Detection and Re-identification*. IEEE Access, January 2025.

[2] Y. S. Y. Z. F. S. T. G. Yunhao Du, Zhicheng Zhao and H. Meng, “Strongsort: Make deepsort great again,” in *IEEE TRANSACTIONS ON MULTIMEDIA, VOL. 25, 2023*, January 2023.

[3] Q. Sui, *Multi-Target Tracking Based on YOLOv8 and DeepSORT*. IEEE 6th International Conference on Internet of Things, Automation and Artificial Intelligence (IoTAAI), October 2024.

[4] S.-K. Huang, C.-C. Hsu, and W.-Y. Wang, *Multiple Object Tracking Incorporating a Person Re-identification Using Polynomial Cross Entropy Loss*. IEEE Access, September 2024.

[5] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, “You only look once: Unified, real-time object detection,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016.