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*—Strong multi-object tracking (MOT) is essential in tasks like UAV surveillance, traffic monitoring, and autonomous systems. Traditional tracking methods fail to handle occlusions, sudden object movements, and appearance changes. This work introduces a dual-pipeline framework to enhance object association and identity retention through the use of detection-based tracking and Re-Identification (ReID)-based feature embeddings. The initial pipeline combines YOLOv8 with DeepSORT for real-time tracking and detection, and the second utilizes a ReID model based on ResNet50 for guaranteeing identity consistency between frames. Both frameworks are tested on the difficult VisDrone-MOT dataset. An improved, Python-based evaluation platform is created to consistently compare performance metrics. The system is extensible, modular, and open-source and is meant to allow for future research in object tracking based on videos.**

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

I. Introduction

Object tracking in videos is a central function in many fields like surveillance, autonomous systems, traffic analysis, and video intelligence. Most of the current online tracking techniques suffer from difficulties in dealing with occlusions, sudden object movement, and appearance changes, which tend to compromise tracking performance. This paper introduces a dual-pipeline approach to improve multi-object tracking by separately investigating both detection-based and Re-Identification (ReID)-based tracking methods. Rather than depending only on frame-to-frame detection, the method focuses on temporal evolution of features to enhance object association and identity continuity. The initial pipeline combines YOLOv8 with DeepSORT for real-time tracking based on object detection, motion modeling, and appearance matching. The second pipeline utilizes a ReID model with ResNet50 as its base to obtain strong feature embeddings for identity tracking between frames. Both pipelines are tested on the VisDrone-MOT dataset with consideration for challenging aspects like dense scenes and camera movements. A Python-based dedicated evaluation platform is constructed to enable systematic performance evaluation and comparison. The derived framework is extensible and will be made publicly available as an open-source tool to facilitate future research in video-based object tracking.

A. Related Works

Sim et al. [1] presented an improved version of DeepSORT and StrongSORT [2] for better multi-object tracking in cattle monitoring. Their research maximized object detection and reidentification (ReID) by optimizing the feature extraction process, resulting in better object association between frames and enhanced long-term tracking performance. This work emphasizes the significance of incorporating optimized ReID models for reliable multi-object tracking.

Sui et al. [3] presented a multi-target tracking framework which is a combination of YOLOv8 and DeepSORT for improved real-time object detection and tracking. Through the integration of YOLOv8's high-precision detection with DeepSORT's strong association mechanism, the research attained higher tracking accuracy, successfully overcoming problems like occlusions and identity persistence over time. Huang et al. [4] presented a novel solution incorporating person re-identification into multi-object tracking with polynomial cross-entropy loss. This method improves feature discrimination, enhancing object association and reducing identity switching. The proposed framework showed better performance in dense scenes, showcasing its robustness in real-world complicated situations. With these developments as a basis, this research aims to further improve object tracking by examining the temporal development of detected object features. YOLOv8 is used for object detection, while ReID-based tracking techniques are employed to enhance object association between frames [5]. The suggested method aims to further enhance the robustness of multi-object tracking, especially in UAV-based video datasets such as VisDrone-MOT.

II. Proposed Methodology

This work proposes a two-pipeline method to improve multi-object tracking (MOT), decoupling object detection with tracking (YOLOv8 + DeepSORT) from person Re-Identification (ReID) with a ResNet50 backbone. The two separate pipelines enable separate evaluation and combination of detection-based tracking and ReID-based tracking. Fig. 1 and Fig. 2 illustrate the architectural flow of the two pipelines.

A. Dataset Preparation

The VisDrone2019-MOT dataset is shared by both pipelines. The dataset is made YOLO-compatible by rescaling

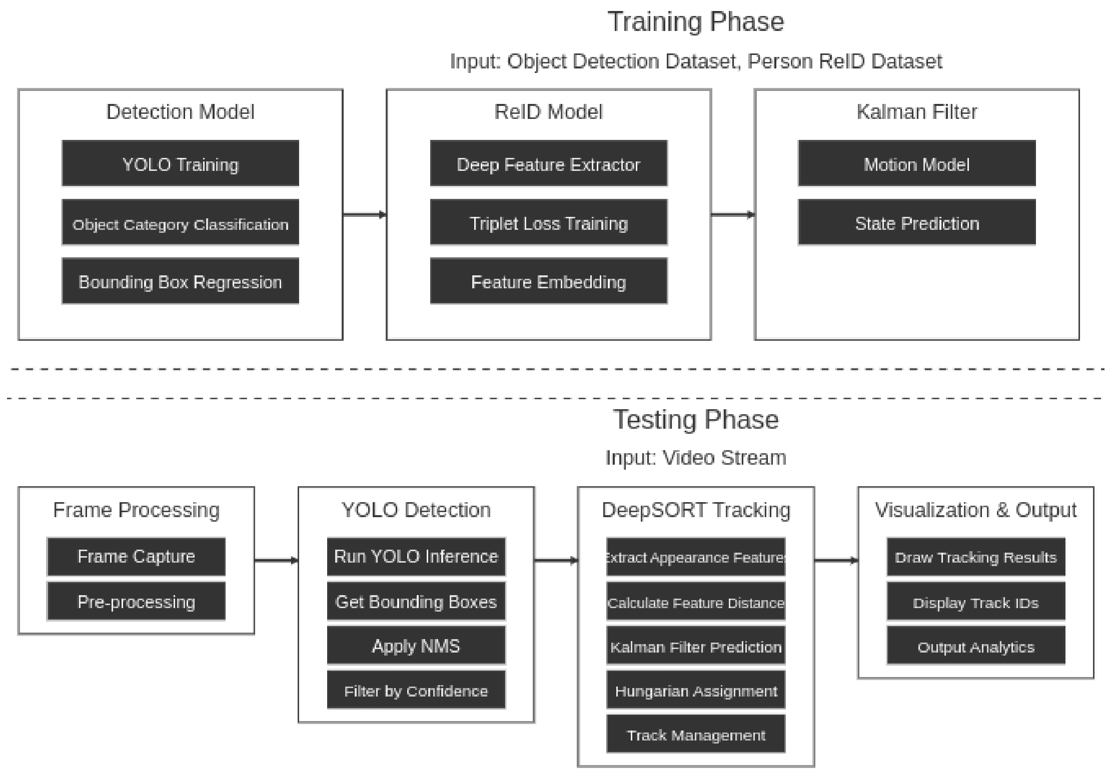

Fig. 1: YOLO + DeepSOrt

bounding box coordinates in terms of image sizes. Object categories are pedestrians, vehicles, bicycles, etc. It is split into training, validation, and testing sets for consistent evaluation between pipelines.

B. Pipeline 1: YOLOv8 + DeepSORT Tracking

Object detection and tracking in this pipeline are closely integrated show in Figure 1:

- **Detection Model:** YOLOv8 is trained on VisDrone dataset with data augmentation of mosaic augmentation and HSV transformation. Models of type YOLOv8n or YOLOv8m are utilized based on system constraints.
- **Tracking:** DeepSORT receives detected bounding boxes, which combine a Kalman Filter and Hungarian assignment for motion forecasting and object linking. The pipeline manages occlusions and sudden motions successfully.
- **Testing Phase:** Inference in YOLOv8 involves object detection per frame, with subsequent DeepSORT tracking, which preserves identities by motion modeling and appearance matching.

C. Pipeline 2: ResNet50-Based ReID Tracking

This standalone pipeline shown in Figure 2 is dedicated exclusively to identity preservation and linking through deep features

- **Preprocessing:** Ground truth cropped object images are normalized with OpenCV and PIL.
- **Feature Engineering:** ResNet50 learns discriminative deep features which are learned with triplet loss in a ReID training architecture built on top of PyTorch.
- **Model Training:** The ReID model is learned from labeled object tracks and verified using cross-validation techniques to learn good quality embedding.
- **Testing Phase:** Object crops on unseen videos are fed through the trained ResNet50 to get embeddings. These are then temporally matched to produce track IDs.

D. Performance Evaluation

Individual pipeline evaluation is done on the VisDrone2019-MOT test set for each pipeline. For Pipeline 1, measures like MOTA (Multi-Object Tracking Accuracy), MOTP (Precision), and ID switches are computed. Pipeline 2 deals with ReID-specific evaluation, cosine similarity-based ID association accuracy, and switch rates. From the results, one can see that YOLOv8 + DeepSORT performs well for real-time cases, whereas the ResNet50-based ReID pipeline performs better in preserving identity consistency, especially under frequent occlusion or appearance change.

Both pipelines are executed based on Python libraries such as OpenCV, PyTorch, and Ultralytics' YOLOv8 framework. The modular design accommodates future fusion or comparison of hybrid tracking methods.

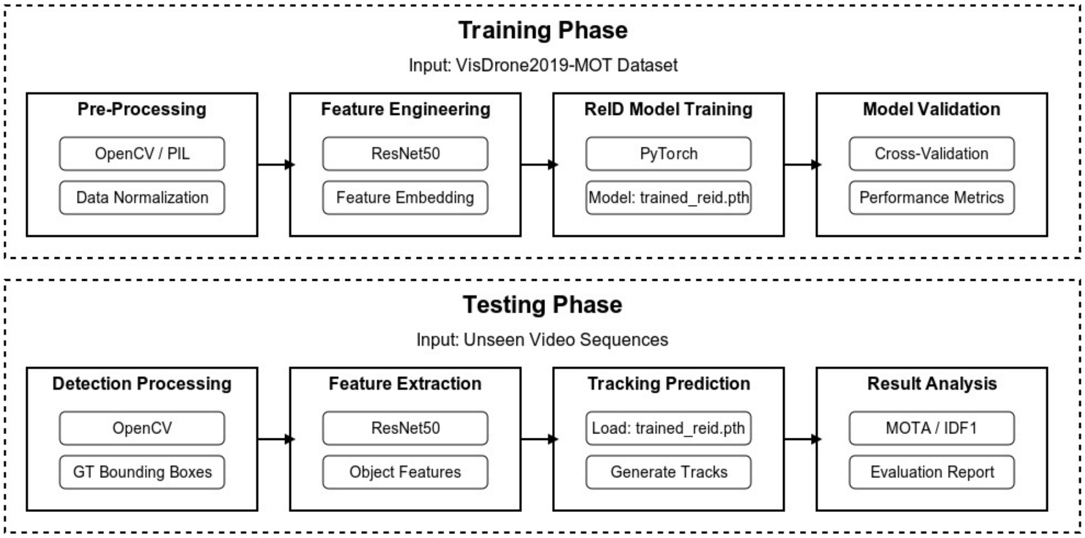

Fig. 2: ResNet50-based ReID tracking

E. Evaluation Metrics

The performance of the tracking system was measured in terms of established multi-object tracking metrics, as listed 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

F. Results and Analysis

Fig. 3 illustrates the qualitative tracking results of the proposed approach on the VisDrone-MOT benchmark. Each object is labeled with a unique ID, and its trajectory is properly maintained across frames. The visualization clearly presents the potency of the tracking pipeline under complex and varied conditions like high object density, scale changes, and partial occlusion.

The suggested method showed a great decline in identity flips, even where there are numerous pedestrians and cars traveling close to each other. As indicated in the picture, the system correctly has unique IDs for nearby people and moving cars, which proves its sustainability where there are crowds.

Integration with appearance-based embeddings using the ResNet50 model for ReID further improved the tracker to handle consistent identities irrespective of visual resemblance or occlusion. The analysis of temporal evolution of features enhanced object association with time, further increasing resistance to changes in the environment and sudden object movements.

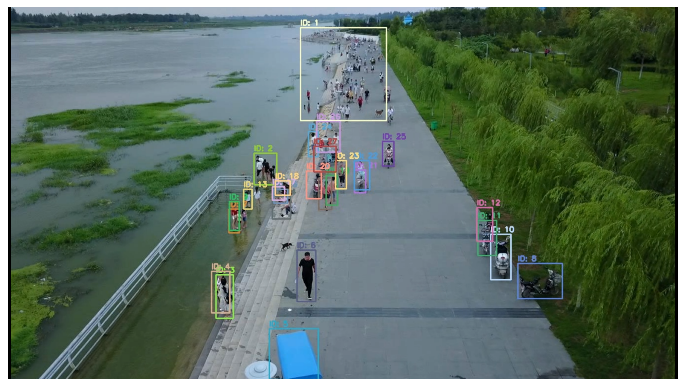

Fig. 3: Example of ReID

III. Conclusion

This paper introduced a comparative study of two autonomous object tracking pipelines in the context of UAV videos. The initial pipeline, which featured the integration of YOLOv8 and DeepSORT, attained effective real-time object tracking with excellent performance on detection-based metrics. The second pipeline, founded on a ResNet50-based ReID model, had superior identity association performance against occlusion and appearance variation. With comprehensive analysis on the VisDrone2019-MOT dataset, the work emphasizes the advantages and limitations of motion modeling and feature-based identity association. In the future, research will aim to construct a hybrid tracking framework that integrates detection and ReID-based information more harmoniously, investigate enhanced backbone architectures, and involve attention mechanisms for better identity memory in dense and dynamic scenes.

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] 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.

[3] 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.

[4] 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.

[5] 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.