Research article

Multi-objective pedestrian tracking method based on YOLOv8 and improved DeepSORT

Wenshun Sheng[*], **Jiahui Shen, Qiming Huang, Zhixuan Liu and Zihao Ding**

Pujiang Institute, Nanjing Tech University, Nanjing 211200, China

* **Correspondence:** Email: sws@njpji.edu.cn.

Abstract: A multi-objective pedestrian tracking method based on you only look once-v8 (YOLOv8) and the improved simple online and real time tracking with a deep association metric (DeepSORT) was proposed with the purpose of coping with the issues of local occlusion and ID dynamic transformation that frequently arise when tracking target pedestrians in real complex traffic scenarios. To begin with, in order to enhance the feature extraction network's capacity to learn target feature information in busy traffic situations, the detector implemented the YOLOv8 method with a high level of small-scale feature expression. In addition, the omni-scale network (OSNet) feature extraction network was then put on top of DeepSORT in order to accomplish real-time synchronized target tracking. This increases the effectiveness of picture edge recognition by dynamically fusing the collected feature information at various scales. Furthermore, a new adaptive forgetting smoothing Kalman filtering algorithm (FSA) was created to adapt to the nonlinear condition of the pedestrian trajectory in the traffic scene in order to address the issue of poor prediction attributed to the linear state equation of Kalman filtering once more. Afterward, the original intersection over union (IOU) association matching algorithm of DeepSORT was replaced by the complete-intersection over union (CIOU) association matching algorithm to fundamentally reduce the target pedestrians' omission and misdetection situation and to improve the accuracy of data matching. Eventually, the generalized trajectory feature extractor model (GFModel) was developed to tightly merge the local and global information through the average pooling operation in order to get precise tracking results and further decrease the impact of numerous disturbances on target tracking. The fusion algorithm of YOLOv8 and improved DeepSORT method based on OSNet, FSA and GFModel was named YOFGD. According to the experimental findings, YOFGD's ultimate accuracy can reach 77.9% and its speed can reach 55.8 frames per second (FPS), which is more than enough to fulfill the demands of real-world scenarios.

Keywords: pedestrian tracking; YOLOv8; DeepSORT; multi-objective tracking; association matching

1. Introduction

Since the start of the twenty-first century, urbanization has grown more intense and difficulties with city traffic have spread to include both motor cars and pedestrians. The majority of pedestrian-motor vehicle collisions are brought on by pedestrians' improper adherence to traffic regulations. The impact of a traffic accident on pedestrians and the flow of traffic along the entire route is very serious. Real-time traffic systems have replaced other methods as the primary way for police to identify the cause of traffic accidents in order to address them swiftly and return to regular road traffic. Due to the rise in automobile traffic in recent years, urban road traffic frequently presents complicated scenarios, which is quite likely to make it difficult for the traffic system to identify partially concealed pedestrian targets. Currently, deep learning-based target detection algorithms and classical detection algorithms are the two main types of target identification algorithms for complicated traffic scenarios [1]. Traditional detection methods typically first extract image candidate frames using sliding windows, followed by feature extraction for the local information of each window, then classification processing for the features extracted [2]. Accordingly, the typical flaws of conventional detection algorithms can be summed up as poor target recognition precision, sluggish computing performance and insufficiently classified derived image categories. Utilizing deep learning for multi-objective recognition and tracking has gained widespread acceptance in the current research community as a result of the growing development of deep learning. As a result, a innovative YOFGD model is suggested in this research to enhance all elements of the target identification and tracking model's performance. The detector in the model is you only look once-v8 (YOLOv8) [3], which can extract more detailed the feature information. The model uses the enhanced deep association metric (DeepSORT) [4] network OFGD as a tracker, which significantly boosts tracking efficiency and accuracy.

The remainder of this essay is structured as follows: The related work is introduced in Section 2. The design and implementation of the model are described in Section 3. The tracking effect of the YOFGD model is compared to that of existing models for target identification and tracking in Section 4. The study's summary and future prospects are presented in Section 5.

2. Related works

Numerous academics have made tremendous progress in the field of multi-objective tracking since the integration of deep learning and pedestrian tracking.

G. Yang et al. in [5] proposed to incorporate the Kalman filter into Kalman on K-KCF visual tracking framework based on deep learning to solve the tracking failure problem caused by pedestrian occlusion in densely populated situations.

In [6], M. I. H. Azhar et al. proposed to use YOLO in combination with DeepSORT to build a system for real-time pedestrian tracking in rows, which was able to successfully detect and track the movement paths of people at an average rate of 2.59 frames per second (FPS).

D. Stadler et al. in [7] suggested a cluster-aware non-maximum suppression (CA-NMS) to solve the issue of missed detections that frequently arise in multi-objective pedestrian tracking in order to decrease the frequency of missed detections of the method. In order to enhance the association performance in cluttered settings, they also presented a new tracking pipeline that blends

detection-by-detection tracking and regression-based tracking patterns. Last but not least, the findings also showed that the study made a lot of progress in tracking effectiveness.

A multi-pedestrian tracking algorithm based on the attention mechanism and double data association is proposed in [8]. It adds a feature pyramid network and a high-resolution feature map at the neck layer of the network to further improve the network's capacity to extract epistatic information. Additionally, it improves the spatial attention mechanism module, which increases the model's accuracy in pedestrian spatial localization. The algorithm's effective tracking performance is lastly demonstrated empirically. In small-scale situations, the pedestrian tracking effect needs to be increased.

In [9], Q. Gao et al. established a target tracking approach based on DeepSORT and an optimized version of YOLOv5, which employs complete-intersection over union(CIOU) to compute the loss function and includes the attention mechanism into YOLOv5 to increase the model's tracking accuracy. Through rigorous testing, the model eventually reached a greater tracking accuracy of 54.3%.

The aforementioned study shows that, while the combination of deep learning and multi-objective tracking enhances the accuracy and speed of pedestrian tracking to some extent, the algorithm is frequently subject to omission and false detection issues when performing target tracking. This research suggests a pedestrian tracking method based on YOLOv8 and the enhanced DeepSORT to reduce such issues.

These are the primary contributions of this work:

1) In this study, a novel adaptive oblivious Kalman filtering [10] technique, the adaptive forgetting smoothing (FSA) Kalman filter, is proposed to improve the extracted features and association matching component in DeepSORT.

2) This study presents a generalized trajectory feature extractor model (GFModel) to more fully extract contextual information.

3) To quantify the match between the detection frame and the prediction frame and to increase the accuracy of target matching, this work utilizes the CIOU [11] correlation matching metric.

3. Model design and implementation

This study proposes a multi-objective pedestrian tracking model (YOFGD) based on YOLOv8 and improved DeepSORT, which addresses the issue of low accuracy and efficiency during target identification tracking caused by occlusion and too small targets.

3.1. YOLOv8 network

In order to further enhance the accuracy and sensitivity of the network, YOLOv8 boosts the backbone network, the detection head, and the loss function from the previous YOLO series' basic framework. This algorithm is currently recognized as an advanced target detection algorithm. The overall network structure of YOLOv8 is depicted in Figure 1. While the backbone network's general structure is similar to that of YOLOv5, YOLOv8 does not adopt the backbone's C3 module (Conv1, Conv2 and Conv3) from YOLOv5. Instead, it combines the idea of efficient layer aggregation networks (ELAN) [12] from YOLOv7 to combine C3 and ELAN to form the CSPDarknet53 to 2-stage feature pyramid network (C2F) module [13]. This enhancement enables the network to acquire richer gradient flow information, increasing the YOLOv8 network's accuracy in image

recognition.
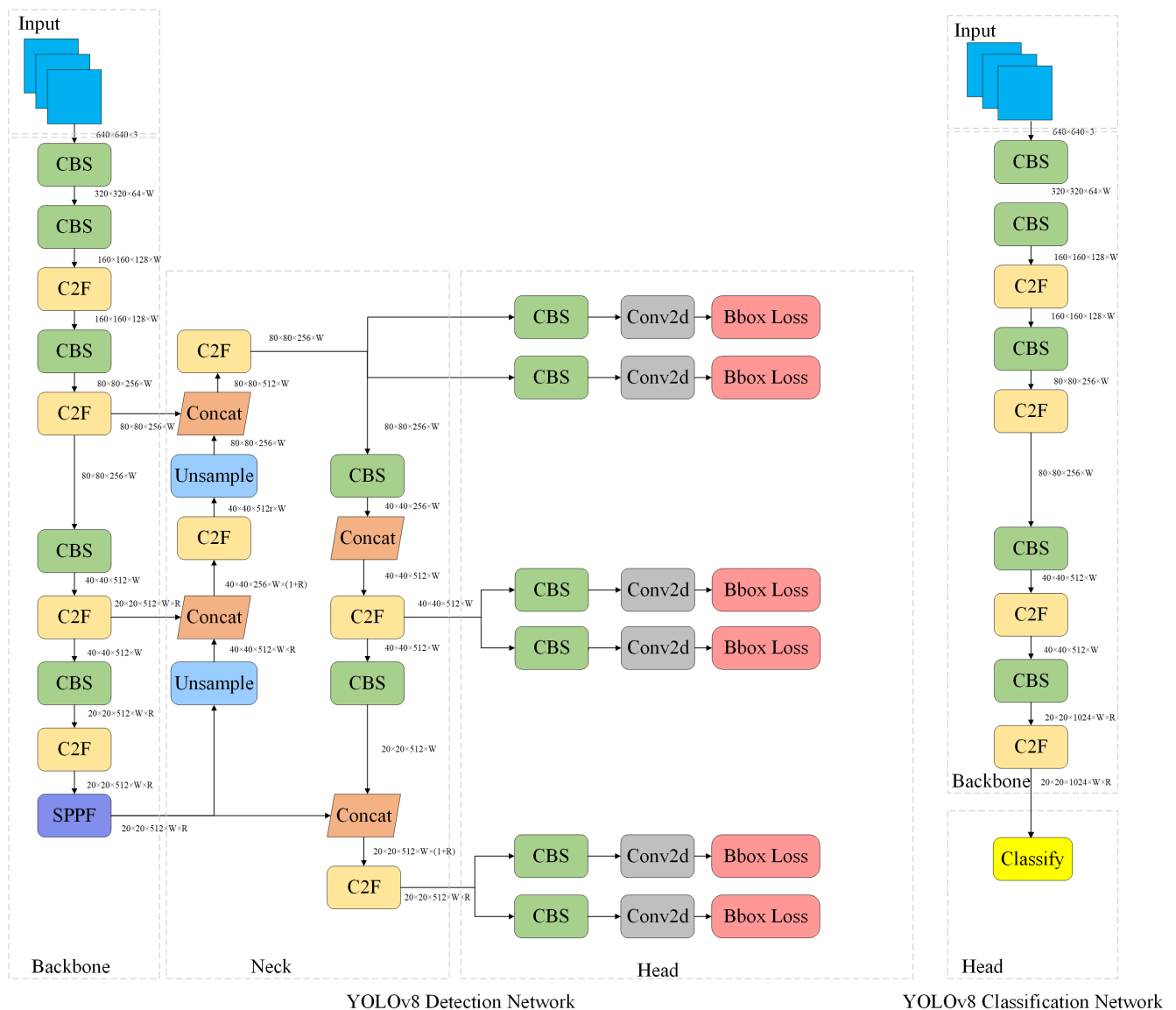
Figure 1. YOLOv8 overall framework.

In contrast to the coupled head utilized in YOLO's previous series, the detecting head of YOLOv8 is a decoupled head. The decoupled head is capable of extracting all target location and classification data and learning each individually by using a classification and detection network, then fusing the data. This clear branch learning concept successfully lowers the network's computational cost, preventing the overfitting phenomena. Additionally, it optimizes the model's performance in terms of generalization and resilience.

Classification and regression are two of the branches that make up the YOLOv8 loss function calculation. In contrast to earlier versions, the classification branch continues to employ the binary cross entropy (BCE) loss [14] while the regression branch uses the distribution focal loss (DFL) [15]

and CIOU loss [16]. Target identification time is significantly increased thanks to the combination of the two loss functions, which makes it possible to gather frame regression information about targets with more accuracy.

3.2. Improved DeepSORT network

The appearance and mobility of tracked targets can both be extracted using DeepSORT, an end-to-end tracking method [17]. Nonetheless, the initial appearance model and re-recognition, motion model, and data association elements of DeepSORT are no longer able to match the current real-time and efficient target tracking requirements due to the growing complexity of the tracking scenarios. For this reason, the feature extraction and association matching components of DeepSORT are optimized in this study. The enhanced DeepSORT algorithm is referred to as the OFGD tracking algorithm, and its algorithm flowchart is displayed in Figure 2.
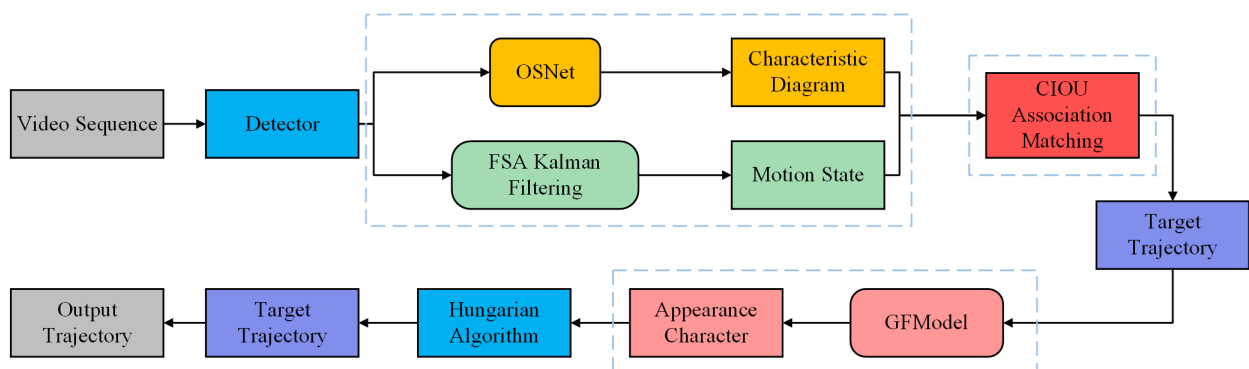

Figure 2. Algorithm flowchart of OFGD.

The three major components of the OFGD tracking method are detection, feature extraction, and association matching. In order to extract the feature information, the input video sequence must first be identified using the omni-scale network (OSNet) [18] network. The goal trajectory is then obtained by matching it with the FSA Kalman filter suggested in this research and correlating the matching's findings with CIOU. The Hungarian algorithm [19] is subsequently combined with the GFModel to complete the feature extraction of global contextual information, which reduces the issue of competing algorithms for accurate tracking, such as occlusion and target number transformation due to changes in the scene. This improves tracking accuracy.

3.2.1. OSNet feature extraction network

Considering the Jetson [20] computing platform is utilized for network deployment in this article, the feature extraction network frequently fails to synchronize in real time while operating. DeepSORT's feature extraction network uses a straightforward convolutional neural network (CNN). The size of the model is decreased from the original 45 to 2.5M while maintaining the tracking accuracy in this paper's solution, and the network topology is depicted in Figure 3. OSNet is utilized as the feature extractor.
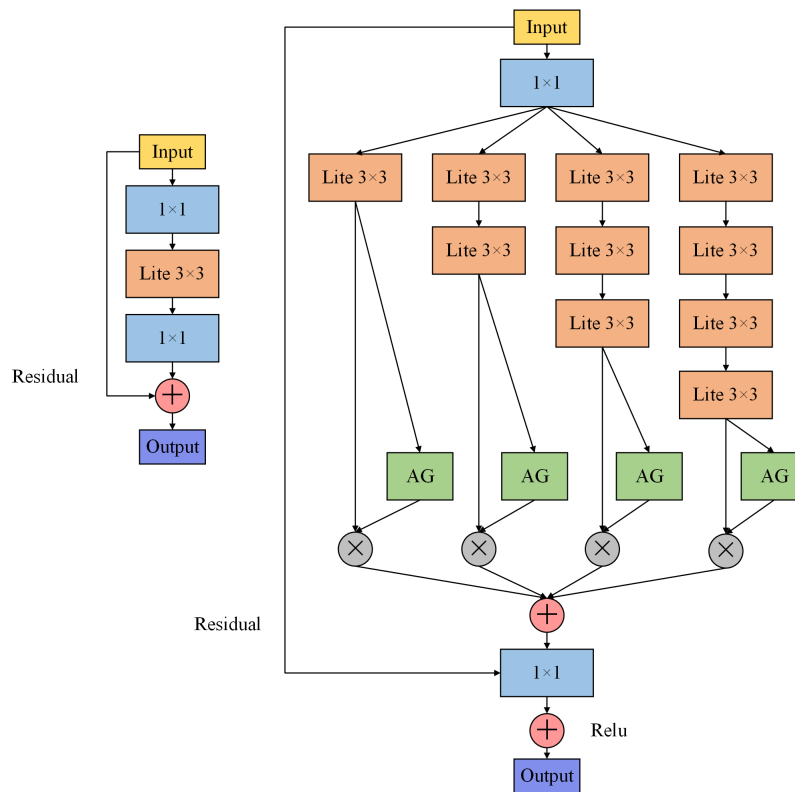
Figure 3. OSNet network structure.

As demonstrated in Figure 3, OSNet is composed of up of several residual blocks with convolutional feature streams. It also incorporates the unified aggregation gate (AG) [21] for dynamic scale fusion, a structure that makes it simpler for the network to extract global features. Moreover, OSNet's network architecture is extensively lightweight, which increases OSNet's efficiency and ease of device deployment.

3.2.2. FSA Kalman filter algorithm

The two fundamental components of the Kalman filter method, which is used to characterize uniform linear motion [22], are prediction and update. To make a prediction is to estimate the present instant's state based on the posterior estimate of the previous moment and to determine the current moment's prior estimate. The update can be further subdivided into measurement updates and time updates. Equation (3.1) displays the time update equation for the Kalman filter method, while Eq (3.2) displays the measurement update equation.

$$\begin{cases} \hat{x}_{\bar{t}} = A\hat{x}_{t-1} + Bu_{t-1} \\ P_{\bar{t}} = AP_{t-1}A^T + Q \end{cases} \tag{3.1}$$

Here, $\hat{x}_{\bar{t}}$ represents the prior state estimate of the t moment, $\hat{x}_{t-1}$ represents the posterior state estimate of the $t-1$ moment, A represents the state transition matrix, B represents the matrix that converts the input to the state, u_{t-1} represents the input of the $t-1$ moment, $P_{\bar{t}}$ represents the prior estimation covariance of the t moment, P_{t-1} represents the posterior estimation covariance of the $t-1$

moment and Q represents the process excitation noise covariance.

$$\begin{cases} K_t = \frac{P_t - H^T}{HP_t - H^T + R} \\ \hat{x}_t = \hat{x}_{\bar{t}} + K_t (z_t - H\hat{x}_{\bar{t}}) \ , \\ P_t = (I - K_t H) \, P_{\bar{t}} \end{cases} \tag{3.2}$$

where K_t represents the measurement noise covariance, H is for the observation matrix, R is for the Kalman gain matrix and z_t is for the observation value.

Subsequently updating the system parameters, the new data on parameter updating changes faster than the old data due to the covariance, and other metrics cannot distinguish between the old data from earlier observations and the new data from more recent observations [23]. This study proposes a novel adaptive oblivious Kalman filter method, the FSA Kalman filter, the precise stages of which are displayed in Algorithm 1, in order to enhance the sensitivity of the algorithm when responding to parameter modifications.

Algorithm 1: Adaptive forgetting Kalman filter algorithm FSA

Input: Initialize the measurement matrix z_t, the confidence of the measured value c_t, the predicted mean value $\hat{x}_{\overline{t-1}}$, the variance $P_{\overline{t-1}}$, observation matrix H_t, noise covariance R_t, kalman gain K_t, adaptive factor μ

Output: Output the final predicted mean $\hat{x}_{\bar{t}}$ and variance $P_{\bar{t}}$

1 Predict target status;

2 The results of tracking and detection are matched;

3 The matching and detection results are updated according to $y_t = z_t - H_t \hat{x}_{\overline{t-1}}'$;

4 The measurement noise ' covariance $\tilde{R}_t = \left[(1 - c_t) \times \frac{1}{k} \sum_{t=1}^{k} c_t \right] \times R_t$ is used to represent the measurement noise scale;

5 Calculating the kalman gain K_t;

6 Update parameters, $\begin{cases} \hat{x}_{\bar{t}} = \hat{x}_{\overline{t-1}}' + K_t y_t \\ P_{\bar{t}} = \left(\frac{1}{\mu} \right) (I - K_t H_t) \, P_{\overline{t-1}}', \mu = 0.95 \end{cases}$.

3.2.3. CIOU association matching

Target matching between anticipated and detected locations by Kalman filtering technique is carried out by the DeepSORT algorithm using the cost matrix and Hungarian algorithm. DeepSORT employs intersection over union (IOU) for correlation matching to improve better tracking by further determining the match between the actual detection frame and the projected detection frame. Equation (3.3) illustrates the IOU calculation formula.

$$IOU = \frac{|A \cap B|}{|A \cup B|}, \tag{3.3}$$

where the numbers A and B represent the expected and real target bounding boxes, respectively.

IOU retains some restrictions, nevertheless, when coping with unique situations. For instance, when there is no overlap between the two bounding boxes, IOU is zero, which leads the gradient to also be zero, making it impossible to do secondary data optimization [24]. CIOU is chosen to substitute IOU for association matching in this study in order to solve this issue, and its formula is shown in Eq (3.4).

$$CIOU = 1 - IOU + \frac{\rho^2 (b, b^{gt})}{c^2} + \alpha v, \tag{3.4}$$

where α represents the weight function, and v is used to gauge how consistently the detection frame to target frame ratio is maintained.

Figure 4 displays the matching impact of CIOU when IOU is zero.
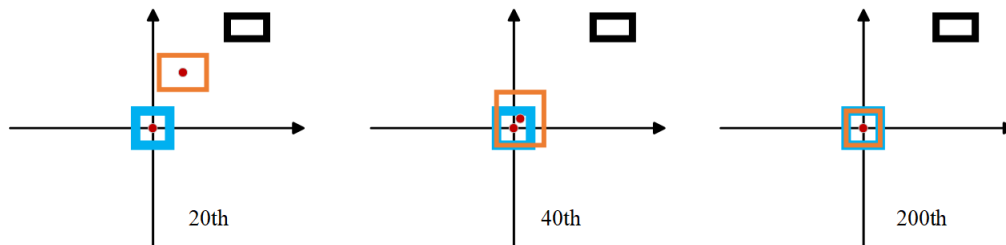
Figure 4. The matching process of CIOU.

As shown in Figure 4, the CIOU can still direct the movement of the target frame regardless of if there is no overlap between the target frame and the prediction frame. The association matching of the OFGD algorithm is substantially accelerated and refined by the CIOU's capability to swiftly return the target frame to the origin position without shifting the position of the prediction frame.

3.2.4. GFModel trajectory feature extractor

The generalized trajectory feature extractor GFModel with strong generalization is suggested in this study as a means of lowering noise resulting from occlusion and scene changes. The unique network structure is depicted in Figure 5. Its input is a collection of frames, which can more thoroughly extract the global contextual information and spatial properties.
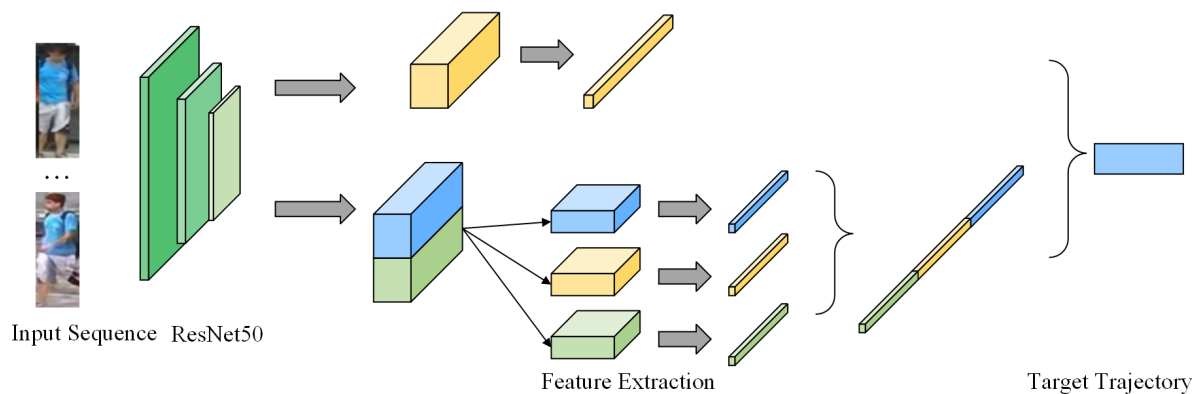
Figure 5. Network structure of GFModel.

4. Experimental results and analysis

This experiment is based on YOLOv8 and OFGD network with Windows environment, Python 3.6.13 as development language, NVIDIA GeForce RTX 2070 SUPER (8G) as GPU and Intel(R)

Core(1TM) i5-10500 CPU@3.10GHz as CPU configuration.

4.1. Experimental data sets

Market-1501 [25] and CUHK03 (The Chinese University of Hong Kong) [26] are two large-scale pedestrian re-identification datasets that were appropriate for the analysis in this paper. Additionally, because Market-1501 and CUHK03 datasets have the same structure, they can be trained together to increase the tracker's accuracy while also generating more data. Bounding_box_train, Bounding_box_test, Gt_query and Gt_galley, which represent the training set, test set, real to-be-queried picture, and real-queried image, respectively, are separated into four pieces for each dataset.

4.2. Evaluation index of experiment

In this paper, a total of four common evaluation metrics for multi-objective tracking are chosen, including multiple object tracking accuracy (MOTA), identification F-score (IDF1), mostly tracked (MT) and FPS, in order to scientifically evaluate the performance of the YOFGD model proposed in this paper from a holistic viewpoint.

1) MOTA

MOTA index integrates four factors: false postive (FP), false negetive (FN), ID switch (IDSW) and ground truth (GT). FP represents the number of targets falsely detected, FN represents the number of real targets not detected, IDSW represents the number of ID switches for the same target and GT represents the number of real objects. The specific formula is shown in Eq (4.1).

$$MOTA = 1 - \frac{FN + FP + IDSW}{GT} \tag{4.1}$$

The better the MOTA represents the model's overall performance, the closer it is to one.

2) IDF1

According to Eq (4.2), IDF1 refers to the F-value recognized by each pedestrian ID and is the reconciled mean of identification precision (IDP) and recall (IDR).

$$IDF1 = \frac{2IDTP}{2IDTP + IDFN + IDFP} \tag{4.2}$$

3) MT

As schematically depicted in Figure 6, MT stands for the number of GT trajectories where the percentage of successfully tracked frames exceeds 80% of the total number of frames.
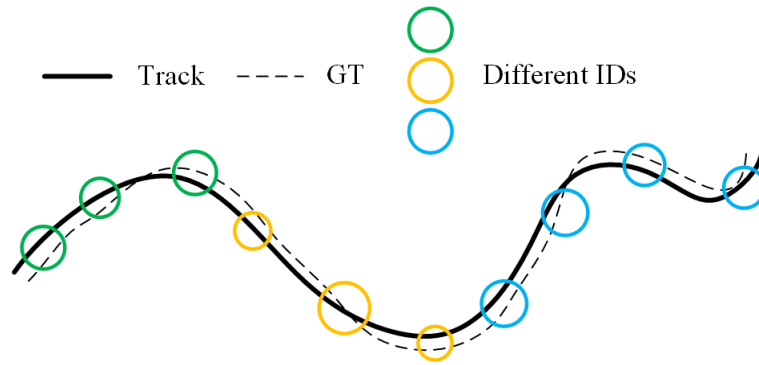
Figure 6. Trajectory tracking diagram.

4) FPS

Target detection speed is frequently assessed using FPS, which is the number of images that can be processed in a second. The model's detecting speed increases with increasing FPS.

4.3. Analysis of experimental results of YOFGD model

This paper initially adopts the OSNet network for pedestrian tracking training, and the initial learning rate is set at 0.00005. After 180,000 iterations, an excellent distinction between pedestrians and pedestrians can be produced. This allows for a reasonable evaluation of the final tracking effect. The model is further trained in this study employing the Torch framework [27], and the resulting training loss curve and classification loss accuracy are displayed in Figure 7 in accordance.
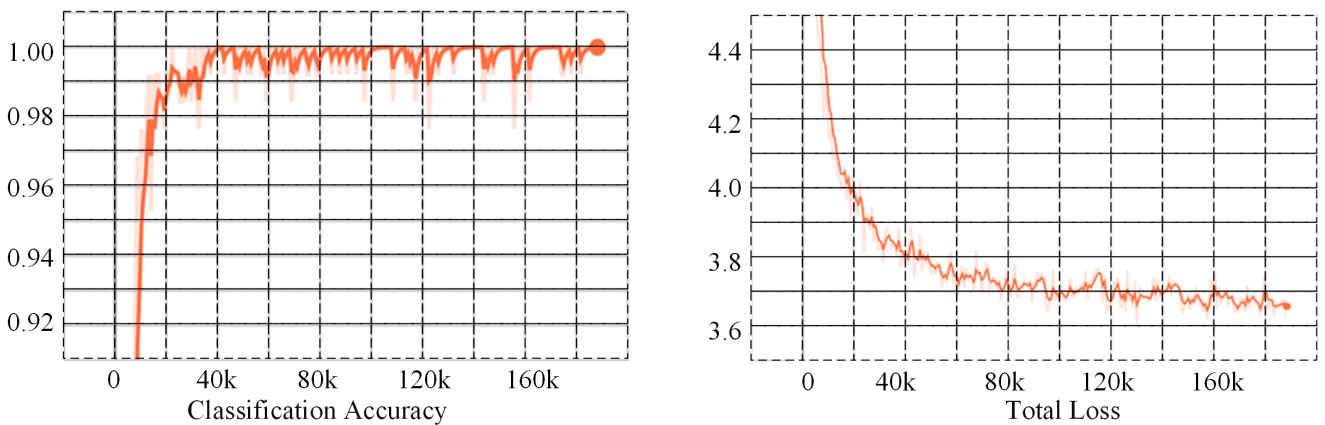
**Figure 7.** The training curve of the tracking model.

Some sequences from multiple object tracking 17 (MOT17) were selected for this paper's demonstration in order to illustrate the implications of the YOFGD model, as seen in Figure 8.
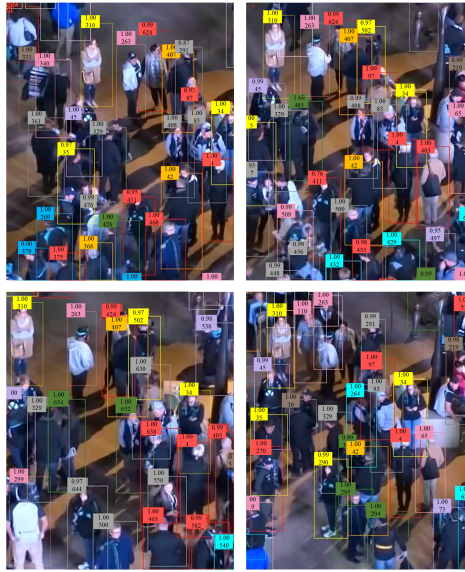
Figure 8. Partial tracking results of MOT17 dataset.

According to the results shown in Figure 8, every pedestrian was accurately tracked. The YOFGD model didn't exhibit any missed detection or misdetection during tracking.

The model of YOLOv8 paired with DeepSORT was chosen for experimental comparison with the YOFGD model described in this paper in order to more effectively illustrate the distinctions between alternative tracking algorithms, as seen in Figure 9.
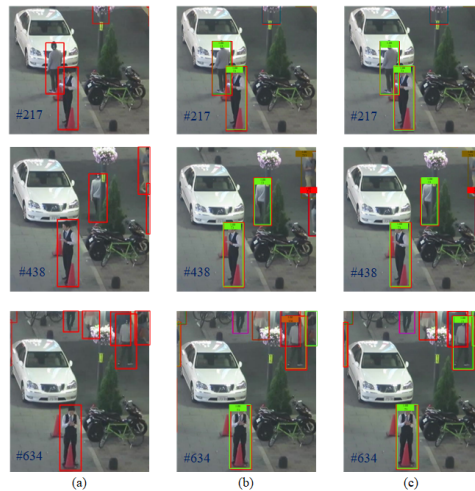
Figure 9. Comparison of partial tracking results.

The real frame is (a), the predicted frame generated by the YOLOv8-Deepsort model is (b) and the predicted frame generated by the YOFGD model is (c). In (b), the tracked target is the pedestrian in the green frame of frame 217 with ID 13. This target can be tracked normally at frame 438; however, in frame 634, the device occlusion and the target's distance cause the ID to be misaligned. The person's ID remains unaltered in (c), which successfully tracks the individual in the distance. It is clear that the

YOFGD model can continue to perform well at tracking irrespective of the presence of complicated scenarios.

In this study, the MOT16 and MOT17 datasets are used to test the YOFGD, and the test results are displayed in Tables 1 and 2.

Table 1. The performance of YOFGD model on MOT16.

Algorithm	MOTA↑	IDF1↑	MT↑	FPS↑
YOFGD	**69.7**	**71.3**	**41.5**	55.8
YOLOv8+CNN+Deepsort	61.6	62.4	34.9	57.5
JDE-864	62.1	56.9	34.4	24.1
JDE-1088	64.4	55.8	35.4	18.9
FairMOT	68.7	70.4	39.5	25.9
Tracktor	54.5	52.5	19.0	5.0
CenterTrackPub*	67.6	57.2	32.9	6.8

Table 2. The performance of YOFGD model on MOT17.

Algorithm	MOTA↑	IDF1↑	MT↑	FPS↑
YOFGD	**77.9**	**76.0**	**45.7**	48.7
YOLOv8+CNN+Deepsort	67.5	55.9	36.4	56.3
JDE-864	66.6	57.2	32.9	19.4
JDE-1088	67.6	57.4	32.2	18.5
FairMOT	67.5	69.8	37.7	18.9
Tracktor	56.4	52.3	19.5	25.0
CenterTrackPub*	61.5	53.3	26.4	18.0

The experimental results demonstrate that the MOTA of YOFGD on the MOT16 dataset reaches 69.7%, which is competitive in the same type of algorithms and indicates that its tracking accuracy has reached a high level when compared to other models. The model in this work has made significant progress in the continuity and accuracy of tracking, as demonstrated by the fact that IDF1 reaches 76.0% on the MOT17 dataset, suggesting that a significant percentage of detection targets among detected and tracked targets receive proper ids. The model in this work has a high tracking accuracy in real nonlinear motion situations, as evidenced by the MT value of 45.7%, which shows that trajectories accurately tracked in 80% frames account for a higher fraction of all trajectories. While the tracking speed may not be the fastest, the model can fully address practical concerns because pedestrian tracking in the real world requires a speed requirement of 30 frames per second.

In comparison to similar research, the MOTA and IDF1 of the model proposed by [28] on the MOT17 dataset were found to be 32.587 and 43.793%, respectively, whereas the MOTA and IDF1 of the model proposed by [29] on the MOT17 dataset were found to be 56.215 and 62.823%, respectively. In these two measures, the model presented in this study performs more superiorly. Furthermore, the trials did not show a balanced connection between tracking accuracy and computational complexity [28, 29]. This paper's tracking efficiency isn't the best, but it is still quite

competitive.

5. Conclusions

This research proposed an innovative multi-objective pedestrian tracking technique based on YOLOv8 and the enhanced DeepSORT. The Kalman filter, feature extraction network and IOU were each improved in this paper starting with feature extraction and association matching in the multi-objective tracking problem. In order to increase the tracker's accuracy, this research first suggested a new adaptive oblivious Kalman filter technique called the FSA Kalman filter. A GFModel was then suggested to enhance the tracker's capacity to extract global data. In order to further increase the tracker's accuracy, the match between the detection and prediction frames was measured using the CIOU correlation matching metric. The YOFGD model put forward in this paper significantly improved tracking accuracy in the experiments. The research in this paper still needs to be expanded upon in certain areas. For example, the model's accuracy has to be further enhanced because it occasionally switches pedestrian IDs incorrectly in scenes with a high density of pedestrians. Meanwhile, YOFGD compares favorably to other models in terms of tracking accuracy and its tracking speed is not at a high level; hence, future studies on the model's speed are anticipated.

Acknowledgments

Conflict of interest

The authors declare there is no conflict of interest.

References

1. H. Liu, F. Dong, Multi object detection algorithm under complex traffic conditions based on YOLOv4, *Foreign Electron. Meas. Technol.*, **41** (2022), 41–47. https://doi.org/10.19652/j.cnki.femt.2204351

2. Z. X. Zou, K. Y. Chen, Z. W. Wei, Y. H. Gou, J. P. Ye, Object Detection in 20 Years: A Survey, *Proc. IEEE*, **111** (2023), 257–276. https://doi.org/10.1109/JPROC.2023.3238524

3. F. M. Talaat, H. ZainEldin, An improved fire detection approach based on YOLO-v8 for smart cities, *Neural Comput. Appl.*, **35** (2023), 20939—20954. https://doi.org/10.1007/s00521-023-08809-1

4. M. I. H. Azhar, F. H. K. Zaman, N. M. Tahir, H. Hashim, People tracking system using DeepSORT, in *2020 10th IEEE International Conference on Control System, Computing and Engineering (ICCSCE)*, Penang, Malaysia, (2022), 137–141. https://doi.org/10.1109/ICCSCE50387.2020.9204956

5. G. Yang, Z. Chen, Pedestrian tracking algorithm for dense crowd based on deep learning, in *Proceedings of 2019 6th International Conference on Systems and Informatics (ICSAI)*, (2019), 568–572. https://doi.org/https://doi.org/10.1109/ICSAI48974.2019.9010144

6. M. I. H. Azhar, F. H. K. Zaman, N. M. Tahir, H. Hashim, People tracking system using DeepSORT, in *Proceedings of 2020 10th IEEE International Conference on Control System, Computing and Engineering (ICCSCE)*, (2020), 137–141. https://doi.org/https://doi.org/10.1109/ICCSCE50387.2020.9204956

7. D. Stadler, J. Beyerer, Multi-Pedestrian tracking with clusters, in *Proceedings of 2021 17th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, (2021), 1–10. https://doi.org/https://doi.org/10.1109/AVSS52988.2021.9663829

8. C. Li, Y. Wang, X. Liu, A Multi-Pedestrian tracking algorithm for dense scenes based on an attention mechanism and dual data association, *Appl. Sci.*, **12** (2022), 9597. https://doi.org/10.3390/app12199597

9. Q. Gao, Z. He, X. Jia, Y. Xie, X. Han, Lightweight high-precision pedestrian tracking algorithm in complex occlusion scenarios, *KSII Trans. Int. Inform. Syst.*, **17** (2023), 840–860. https://doi.org/10.3837/tiis.2023.03.009

10. R. A. Zitar, A. Mohsen, A. E. Seghrouchni, F. Barbaresco, N. A. Al-Dmour, Intensive review of drones detection and tracking: Linear Kalman filter versus nonlinear regression, an analysis case, *Arch Comput. Methods Eng.*, **14** (2023), 2811-–2830. https://doi.org/10.1007/s11831-023-09894-0

11. X. B. Liu, X. Z. Yang, Y. Chen, S. T. Zhao, Object detection method based on CIoU improved bounding box loss function, *Chinese J. Liquid Cryst. Displ.*, **38** (2023), 656–665. https://doi.org/10. 37188/CJLCD. 2022-0282

12. C. Y. Wang, H. Y. M. Liao, I. H. Yeh, Designing network design strategies through gradient path analysis, preprint, arXiv: 2211.04800.

13. H. T. Liu, X. H. Duan, J. M. Guo, H. Y. Liu, J. Gu, H. Chen, DC-YOLOv8: Small-Size object detection algorithm based on camera sensor, *Electronics*, **12** (2023), 2323. https://doi.org/10.3390/electronics12102323

14. H. Z. Xu, H. J. He, Y. Zhang, L. F. Ma, J. T. Li, A comparative study of loss functions for road segmentation in remotely sensed road datasets, *Int. J. Appl. Earth Observ. Geoinform.*, **116** (2023), 1569–8432. https://doi.org/10.1016/j.jag.2022.103159

15. M. S. Hossain, J. M. Betts, A. P. Paplinski, Dual Focal Loss to address class imbalance in semantic segmentation, *Neurocomputing*, **462** (2021), 69–87. https://doi.org/10.1016/j.neucom.2021.07.055

16. I. Pacal, D. Karaboga, A robust real-time deep learning based automatic polyp detection system, *Comput. Biol. Med.*, **134** (2021), 104519. https://doi.org/10.1016/j.compbiomed.2021.104519

17. A. X. Zhao, J. Q. Yang, H. B. Yang, X. G. Shi, W. X. Fu, S. Liu, et al., Indoor multi-object personnel recognition and tracking across camera based on optimized DeepSORT and FastReID, *J. Xi'an Univ. Sci. Technol.*, **43** (2023), 620–630. https://doi.org/10.13800/j.cnki.xakjdxxb.2023.0320

18. T. Jin, X. Ye, Z. Li, Z. Huo, Identification and tracking of vehicles between multiple cameras on bridges using a YOLOv4 and OSNet-Based method, *Sensors*, **23** (2023), 5510. https://doi.org/10.1109/ICSPIS56952.2022.10043932

19. C. Nie, Z. Ju, Z. Sun, H. Zhang, 3D object detection and tracking based on Lidar-Camera fusion and IMM-UKF algorithm towards highway driving, in *IEEE Trans. Emerging Topics Comput. Intell.*, **7** (2023), 1242–1252. https://doi.org/10.1109/TETCI.2023.3259441

20. S. Mittal, A Survey on optimized implementation of deep learning models on the NVIDIA Jetson platform, *J. Syst. Arch.*, **97** (2019), 428–442. https://doi.org/10.1016/j.sysarc.2019.01.011

21. S. Uladzislaum, X. Feng, Modified omni-scale net architecture for cattle identification on their muzzle point image pattern characteristics, in *International Conference on Computer, Artificial Intelligence, and Control Engineering (CAICE 2023)*, **12645** (2023), 489–494. https://doi.org/10.1117/12.2681201

22. H. Xie, Z. Xiao, W. Liu, Z. Ye, PVNet: A Used Vehicle Pedestrian Detection Tracking and Counting Method, *Sustainability*, **15** (2023), 14326. https://doi.org/10.3390/su151914326

23. R. P. Tripathi, A. K. Singh, P. Gangwar, Fractional order adaptive Kalman filter for sensorless speed control of DC motor, *Int. J. Electron.*, **110** (2023), 373–390. http://dx.doi.org/10.5081/jgps.2.1.42

24. N. Wanchaitanawong, M. Tanaka, T. Shibata, M. Okutomi, Multi-modal pedestrian detection with misalignment based on modal-wise regression and multi-modal IoU, *J. Electron. Imaging*, **32** (2023), 013025. https://doi.org/10.1117/1.JEI.32.1.013025

25. N. K. S. Behera, P. K. Sa, S. Bakshi, U. Bilotti, Explainable graph-attention based person re-identification in outdoor conditions, *Multimed Tools Appl.*, **2023** (2023), 99–108. https://doi.org/10.1007/s11042-023-16986-3

26. M. K. Vidhyalakshmi, E. Poovammal, V. Bhaskar, J. Sathyanarayanan, Novel similarity metric learning using deep learning and root SIFT for person re-identification, *Wireless Personal Commun.*, **117** (2021), 1835–1851. https://doi.org/10.1007/s11277-020-07948-1

27. O. Tomarchio, D. Calcaterra, G. D. Modica, P. Mazzaglia, Torch: a tosca-based orchestrator of multi-cloud containerised applications, *J. Grid Comput.*, **19** (2021), 1–25. https://doi.org/10.1007/s10723-021-09549-z

28. M. Razzok, A. Badri, I. E. Mourabit, Y. Ruichek, A. Sahel, Pedestrian detection and tracking system based on Deep-SORT, YOLOv5, and new data association metrics, *Information*, **14** (2023), 218. https://doi.org/10.3390/info14040218

29. X. Xiao, X. Feng, Multi-Object pedestrian tracking using improved YOLOv8 and OC-SORT, *Sensors*, **23** (2023), 8439. https://doi.org/10.3390/s23208439