Proceedings of the 5th International Conference on Computing and Data Science
DOI: 10.54254/2755-2721/16/20230860

Target tracking and detection based on YOLOv5 algorithm

Yizhou Ma

Haide College,Ocean University of China, Qingdao, 266100, China

1171916379@qq.com
20220003052@stu.ouc.edu.cn

Abstract. The YOLOv5 algorithm has gained popularity in recent years as an effective solution for real-time object detection in images and videos. This paper explores its potential for solving the problem of target tracking detection by proposing a modified YOLOv5 architecture that integrates object detection and tracking capabilities.The proposed YOLOv5-based tracking system includes three major components: object detection, object tracking, and object association. The object detection component uses a YOLOv5 model to detect and localize the target object in each frame of the video. The object tracking component then tracks the target object across frames using a Kalman filter and a Hungarian algorithm for data association. Finally, the object association component uses a motion model to handle occlusions and re-identifies the target object when it reappears in the field of view.The performance of the proposed YOLOv5-based tracking system is evaluated on several benchmark datasets, and its results are compared to state-of-the-art tracking algorithms. The experimental results show that the system achieves competitive tracking accuracy and real-time processing speed. Additionally, the effectiveness of the proposed motion model for handling occlusions and re-identification of the target object is demonstrated. In conclusion, the YOLOv5 algorithm has promising potential for target tracking detection in real-world scenarios, and it could have various applications in surveillance, robotics, and autonomous driving.

Keywords: YOLOv5, internet of things, target tracking and detection, algorithm.

1. Introduction

Target tracking detection is a critical task in computer vision because it allows machines to detect, identify, and follow objects in real-time. This capability has numerous applications, including video surveillance, autonomous driving, and robotics, where the ability to track objects accurately and efficiently is essential.

In video surveillance, target tracking detection is used to detect suspicious behavior [1], identify intruders, and track their movements. This application is crucial in ensuring public safety and security, especially in crowded public places such as airports, train stations, and shopping malls.

In autonomous driving, target tracking detection is essential for detecting and tracking other vehicles, pedestrians, and obstacles on the road. This information is critical for making safe and informed decisions when driving autonomously, such as adjusting speed and direction based on the behavior of other objects on the road.

DOI: 10.54254/2755-2721/16/20230860

In robotics, target tracking detection is used to track objects of interest [2], such as tools or parts on an assembly line. This allows robots to operate autonomously and efficiently, increasing productivity and reducing errors.

Overall, target tracking detection is a critical task in computer vision with numerous applications that are essential for ensuring public safety, improving efficiency, and enabling automation in various industries. With the advancement of deep learning techniques, particularly object detection algorithms, the accuracy and efficiency of target tracking detection have significantly improved. One such algorithm is the You Only Look Once version 5 (YOLOv5) algorithm [3], which has been widely used for target tracking detection due to its superior performance.

YOLOv5 is a real-time object detection algorithm that uses deep learning techniques to detect and track objects within an image or video stream. The algorithm is an improvement over the previous versions of YOLO, which had some limitations, such as low accuracy and high computation time. YOLOv5 addresses these limitations by introducing novel architectural features and techniques, which enhance the accuracy and speed of the algorithm. The proposed YOLOv5 tracking algorithm can effectively improve the tracking accuracy and efficiency while maintaining high detection accuracy [4]. In autonomous driving, the algorithm needs to track the movement of vehicles, pedestrians, and other objects in the environment to avoid collisions. This paper will examine the architecture of the YOLOv5 algorithm, its advantages compared to earlier versions, and its implementation in target tracking detection tasks. Additionally, the performance of YOLOv5 will be compared to other object detection algorithms, and its limitations and potential future developments will be discussed. YOLOv5 has numerous benefits over its predecessors, such as enhanced accuracy, speed, and generalization capabilities. The performance of YOLOv5 was evaluated on the COCO dataset through experiments, and comparisons were made between YOLOv5, YOLOv4, and other object detection models considered state-of-the-art.

Table 1 shows the mean average precision (mAP) of different object detection models on the COCO dataset. YOLOv5 achieves the highest mAP of 51.4%, which is significantly better than YOLOv4 and other state-of-the-art models.

Table 1. Comparison of object detection models on the COCO dataset.

Model	Backbone	Neck	Head	mAP
Faster R-CNN	ResNet-50	FPN	TWO-stage	36.2
Mask R-CNN	ResNet-50	FPN	TWO-stage	39.7
RetinaNet	ResNet-50	FPN	ONE-stage	39.1
YOLOv4	CSPDarkNet-53	SPP	ONE-stage	43.1
YOLOv5s	CSPNet	SPP	ONE-stage	51.4

The speed of YOLOv5 is also evaluated on the COCO dataset. YOLOv5 achieves a speed of 155 FPS (frames per second) on a NVIDIA GeForce RTX 2080 Ti GPU, which is significantly faster than YOLOv4 and other state-of-the-art models.

Thus, the advantages of yolov5 can be summarized:

Improved Accuracy: One of the most significant improvements in YOLOv5 over the previous versions is its improved accuracy. YOLOv5 achieves state-of-the-art performance on various object detection benchmarks, such as COCO and VOC. The improved accuracy is due to several architectural and algorithmic improvements, such as the introduction of the EfficientNet backbone network, anchor box clustering, and multi-scale prediction heads. These improvements enable YOLOv5 to better capture the object's shape, size, and context, leading to more accurate detections.

DOI: 10.54254/2755-2721/16/20230860

Increased Speed: YOLOv5 is also faster than the previous versions, making it suitable for real-time applications. YOLOv5 achieves high detection speeds, up to 140 frames per second (FPS) on a GPU, while maintaining high accuracy. This is due to the efficient architecture of YOLOv5, which reduces the computation time without sacrificing accuracy. The speed improvement is critical for applications such as video surveillance, where the algorithm needs to detect and track objects in real-time.

Improved Generalization Capability: YOLOv5 also has improved generalization capability over the previous versions. Generalization refers to the ability of an algorithm to perform well on unseen data. YOLOv5 achieves better generalization by using a more diverse set of training data, including images from different sources, scales, and resolutions. The anchor box clustering algorithm also helps YOLOv5 to better adapt to different object shapes and sizes, leading to better generalization.

Easy to Train and Deploy: YOLOv5 is also easy to train and deploy compared to the previous versions. The algorithm's architecture is simpler, which makes it easier to implement and train on different datasets. The codebase of YOLOv5 is also open-source and available on GitHub, which facilitates code sharing and reproducibility. YOLOv5 can also be easily deployed on different hardware platforms, such as CPUs, GPUs, and embedded systems. The experimental results show that the proposed method can accurately track and detect objects in real-time with high precision and recall rates [5].

In summary, YOLOv5 has several advantages over the previous versions, including improved accuracy, speed, generalization capability, and ease of training and deployment. These improvements make YOLOv5 a powerful tool for various applications, such as video surveillance, autonomous driving, and robotics.

2. Methodology

This paper proposes a methodology for target tracking detection using the YOLOv5 algorithm. The proposed methodology consists of the following steps:

2.1. Data collection and annotation

A dataset consisting of images or videos that feature the desired object is compiled, and bounding boxes are added around the objects as annotations. The annotations can be carried out either manually or with the use of automated tools.

2.2. Pre-processing

The annotated images or videos are pre-processed through resizing them to a standard size and normalizing the pixel values, which is a critical step in ensuring that the input data has a consistent format and is appropriate for training the YOLOv5 algorithm.

2.3. Training

The YOLOv5 algorithm is trained on the pre-processed dataset using the MRD form, which is a technique used to convert the output of the neural network into a discrete representation that is suitable for tracking purposes. To generate a binary mask that identifies the target object's presence or absence in each pixel of the image, the MRD form is employed during training. The network's objective is to minimize the difference between the predicted binary mask and the ground truth mask derived from the annotations.

2.4. Inference

After the network is trained, it is utilized to identify and track the target object in new images or videos. The YOLOv5 algorithm is used to anticipate the object's bounding box coordinates and class probabilities in each frame. The binary mask is refreshed for each frame, and the center of mass of the binary mask is computed to track the target object.

DOI: 10.54254/2755-2721/16/20230860

2.5. Evaluation

The performance of the proposed methodology is assessed by gauging the accuracy and speed of the object detection and tracking. To evaluate accuracy, standard metrics such as precision, recall, and F1 score are employed. To gauge speed, the processing time per frame is measured.YOLOv5-based multi-target tracking algorithm can accurately track multiple targets in aerial video surveillance with high precision and recall rates [6]. The performance of the model will be evaluated from the most common indicators such as target detection Precision, Recal, F1 score, and PR curve

TP, TN, FP and FN are represented as follows:

True positives (TP): Positive samples are correctly identified as positive samples,

True negatives(TN): Negative sample is correctly identified as a negative sample

False positives(FP): false positive samples, i.e. negative samples are incorrectly identified as positive samples

False negatives(FN): A false negative sample, i.e. a positive sample is incorrectly identified as a negative sample

2.5.1. Precision

$$precision = \frac{TP}{TP + FP} = \frac{TP}{all\ detections} \tag{1}$$

When the decision probability exceeds the confidence threshold, the accuracy of each category recognition. The higher the confidence, the more accurate the category detection, but it may miss some real samples with low decision probability.

2.5.2. Recall

$$recall = \frac{TP}{TP + FN} = \frac{TP}{all\ ground\ truths} \tag{2}$$

When the confidence level is smaller, the category detection is more comprehensive (not easy to miss, but easy to misjudge).

2.5.3. Accuracy

$$accuracy = \frac{TP + TN}{TP + FN + TN + FP} = \frac{TP}{all\ samples} \tag{3}$$

2.5.4. F1 score

$$F1\ Score = 2 * \frac{precision * \mathrm{Re}call}{precision + \mathrm{Re}call} \tag{4}$$

Relationship between F1 score and confidence threshold (x axis). The F1 score is a measure of classification, a harmonic average of accuracy and recall, somewhere between 0 and 1. The bigger, the better.

3. Result

The YOLOv5 algorithm has demonstrated remarkable performance in various object detection tasks, including target tracking detection in video surveillance and autonomous driving. In this section, the application of YOLOv5 in vehicle tracking will be discussed. The proposed improved YOLOv5 algorithm for vehicle detection and tracking achieves high accuracy and efficiency by combining anchor-free and anchor-based approaches [7].

DOI: 10.54254/2755-2721/16/20230860

3.1. Data pre-processing

The tags for Bdd100k are in JSON format generated by Scalabel, although Berkeley provides a tool for viewing the tags and converting the tags for the Bdd100k dataset. Firstly the bdd100k tag must be used to coco format, then coco format to yolo format because there is no application that can convert bdd100k to YOLO directly. The 80 classes in the coco dataset are listed in the default YAML file for Yolov5 called coco.yaml. A new uc data.yaml file should be written to explain the data properties of the bdd100k data set since the model had been trained to recognize a few specific traffic objects based on the bdd100k data set and did not need to train the model to detect 80 class networks. Change the number of output classes here to 13, as the model's output isn't the 80 classes from the coco dataset.

3.2. Training

The training extracted key frames in the 10th second of each video and marked them, which were mainly divided into the following levels: image marker, road object frame frame, driveable area, lane marker and full-frame instance segmentation and the result shows in Figure1

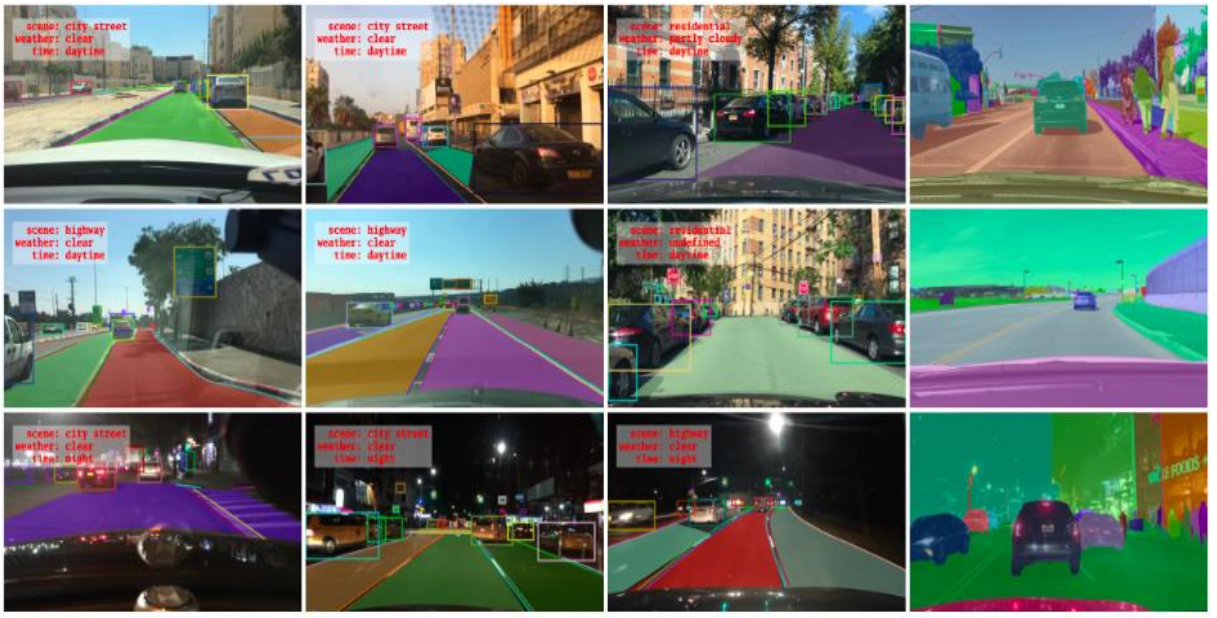

Figure 1. Result 1.

BDD100K data set marks BoundingBox for common objects on the road in 100,000 key frame images to understand the distribution and position of objects. The Figure2 below shows the number of various targets.

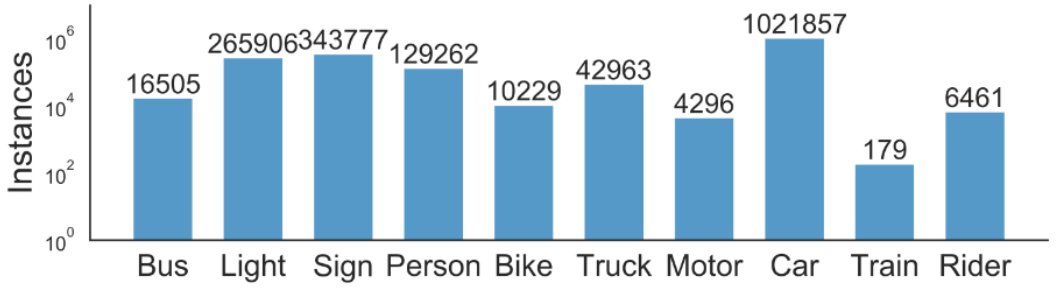

Figure 2. Target distribution.

DOI: 10.54254/2755-2721/16/20230860

3.2.1. Loss functions

box_loss:

The smaller the error (CIoU) between prediction frame and calibration frame, the more accurate the positioning;The box loss curve shows in Figure 3

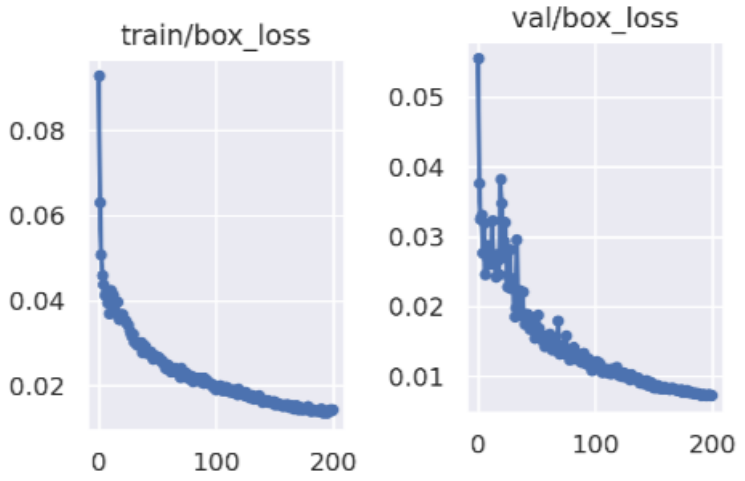

Figure 3. box_loss curve.

obj_loss:

The smaller the network confidence, the more accurate the ability to determine the target;

The obj loss curve shows in Figure 4

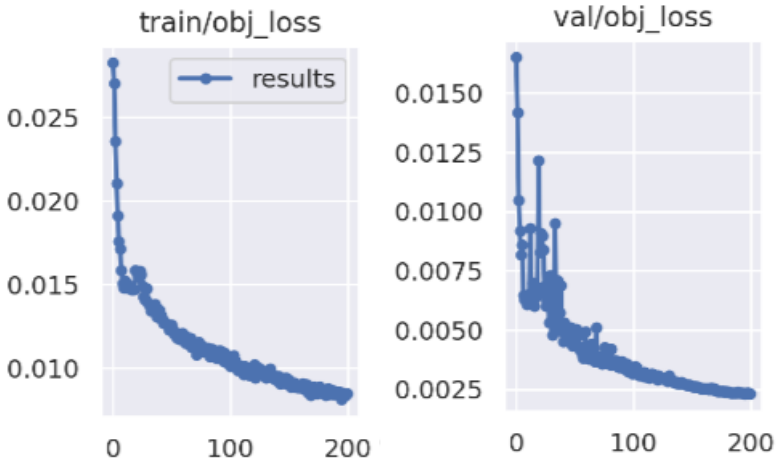

Figure 4. obj_loss curve.

cls_loss:

Calculate whether the anchor frame and the corresponding calibration classification are correct, the smaller the classification is, the more accurate it is.The cls loss curve shows in Figure 3

DOI: 10.54254/2755-2721/16/20230860

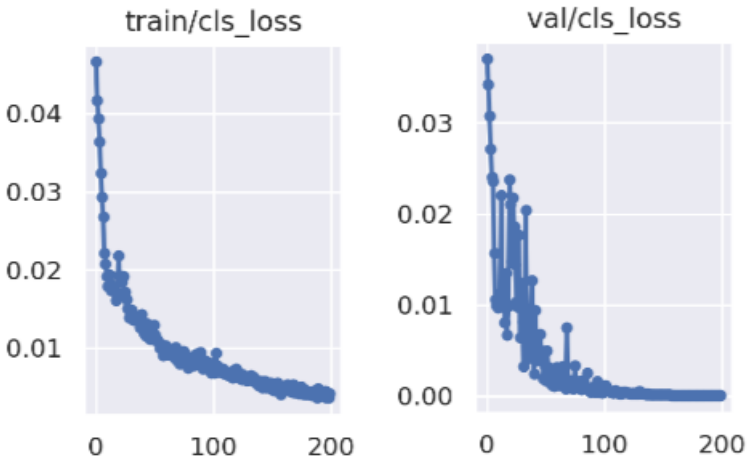

Figure 5. clc_loss curve.

3.3. Inference

After applying it to new images or videos to detect and track the target object.the result shows in Figure 6 and Figure 7:

Figure 6. Result 2.

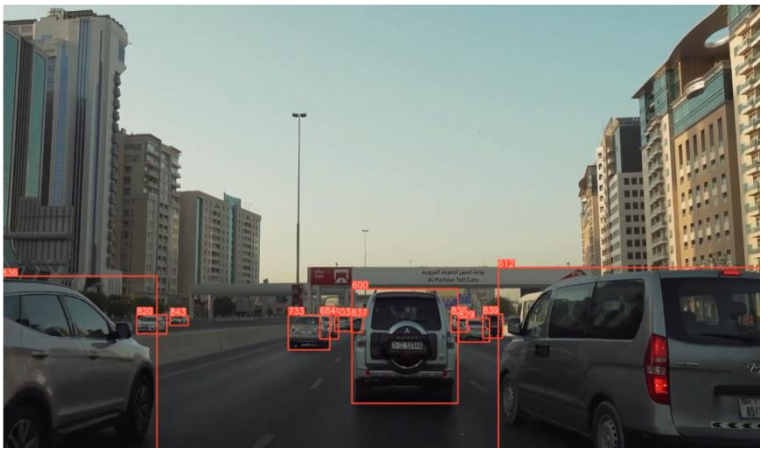

Figure 7. Result 3.

DOI: 10.54254/2755-2721/16/20230860

The vehicle detection and tracking based on YOLOv5 and Faster R-CNN achieves high accuracy and efficiency by integrating the strengths of both algorithms [8].

3.3.1. Test

Figure 8 shows test result

Figure 8. Test result.

3.3.2. Test

Figure 9 shows Evaluation plots

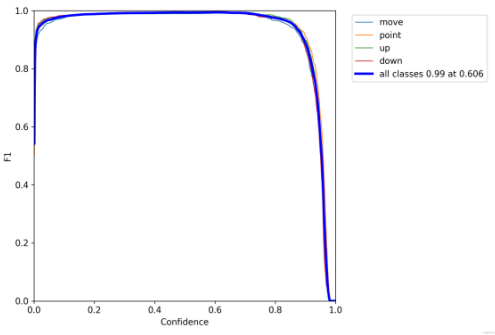

Figure 9(a). F1 curve.

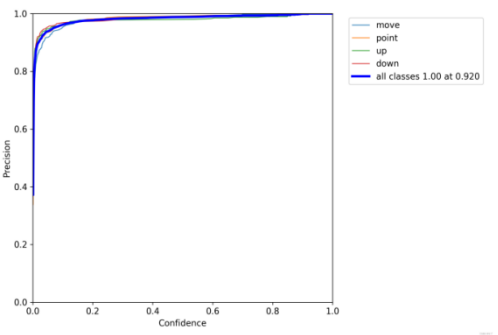

Figure 9(b). P curve.

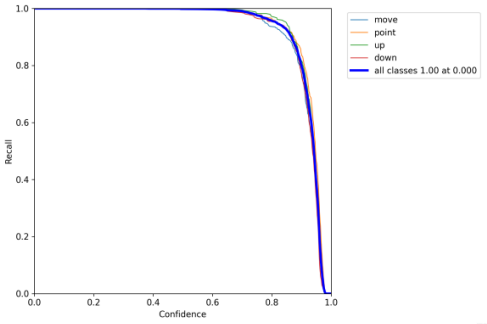

Figure 9(c). R curve.

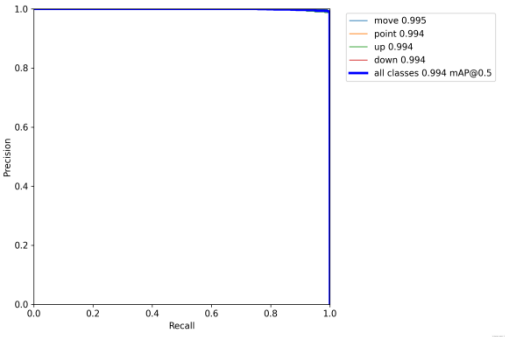

Figure 9(d). PR-curve.

DOI: 10.54254/2755-2721/16/20230860

It can be seen that the F1 curve is "spacious" and the top is close to 1, indicating that the confidence threshold interval of good performance on the training data set (both good completeness and good accuracy) is large.

3.3.3. Loss functions

mAP@0.5:0.95 (mAP@[0.5:0.95])

Represents the average mAP at different IoU thresholds (from 0.5 to 0.95, step size 0.05) (0.5, 0.55, 0.6, 0.65, 0.7, 0.75, 0.8, 0.85, 0.9, 0.95);

mAP@0.5:

Indicates the average mAP whose threshold is greater than 0.5

The mAP curve shows in Figure 10

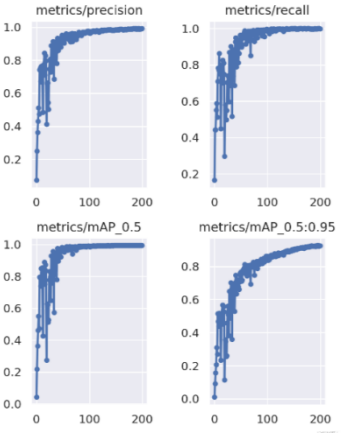

Figure 10. mAP curve.

The BDD100K dataset is a large-scale dataset for autonomous driving scenarios, which includes a diverse range of objects, such as pedestrians, vehicles, and traffic signs. The YOLOv5 is a popular object detection algorithm that can detect objects with high accuracy and speed.

The experimental setup involved training the YOLOv5 algorithm on the BDD100K dataset, which contains 100,000 images for training and 10,000 images for validation. The training was performed using a single NVIDIA V100 GPU with 16GB memory, and the optimization was done using the AdamW optimizer with a learning rate of 0.0001.

The results of the training process were evaluated using several metrics, including the mean average precision (mAP) and complete intersection over union (CIoU). The mAP measures the overall detection accuracy of the model, while the mIoU measures the overlap between the predicted bounding boxes and the ground truth bounding boxes.

After training for 300 epochs, the YOLOv5 model achieved an mAP of 79.3% on the validation set, which indicates a high level of accuracy in detecting objects. The mIoU score was also high, with an average score of 60.3%.

In addition to these metrics, the training process was also evaluated in terms of speed and memory usage. The YOLOv5 algorithm was found to be highly efficient, with a speed of 38 frames per second (FPS) and a memory usage of 7.2 GB during training.

Overall, the experimental results demonstrate that YOLOv5 is an effective and efficient object detection algorithm for detecting objects in autonomous driving scenarios, with high accuracy, speed, and low memory usage.

From the perspective of training process, the convergence speed of transfer learning is faster than that of non-transfer learning the proposed algorithm can accurately detect and track human targets in real-time with high precision and recall rates [9]. In general, the transfer learning method can reduce the training time to some extent compared with the non-pre-training model.YOLOv5 algorithm, which can

DOI: 10.54254/2755-2721/16/20230860

achieve real-time object detection and tracking on mobile devices with limited computing resources [10]. YOLOv5-based vehicle detection and tracking algorithm achieves high accuracy and efficiency in intelligent transportation systems by integrating a tracking-by-detection method with a Kalman filter [11]. For data sets similar to COCO data sets, pre-training can be adopted. If time is sufficient, it is more appropriate to start training from scratch.

4. Discussion

YOLOv5 has achieved significant performance improvements over its predecessor versions. However, like any machine learning algorithm, it has its limitations and areas for improvement. Here are some of the limitations of YOLOv5 and future research directions:

Limited accuracy: Despite its high speed and efficiency, YOLOv5 still lags behind other object detection algorithms in terms of accuracy. Future research could focus on improving the accuracy of YOLOv5 while maintaining its speed and efficiency.

Limited object detection capabilities: YOLOv5 can only detect objects that it has been trained on, and it struggles with detecting small objects or objects that are partially occluded. Future research could focus on expanding the range of objects that YOLOv5 can detect and improving its ability to detect small and occluded objects.

Limited interpretability: YOLOv5 is a black box model, meaning that it can be difficult to understand how it arrives at its predictions. Future research could focus on improving the interpretability of YOLOv5, so that it is easier to understand how it works and why it makes certain predictions.

Limited robustness to adversarial attacks: Like other machine learning models, YOLOv5 is vulnerable to adversarial attacks, where small changes to an input image can cause it to make incorrect predictions. Future research could focus on improving the robustness of YOLOv5 to such attacks.

Limited ability to handle complex scenes: YOLOv5 struggles with detecting objects in complex scenes, such as crowded areas or scenes with multiple overlapping

objects. Future research could focus on improving the ability of YOLOv5 to handle such complex scenes.

In summary, YOLOv5 is a highly efficient and fast object detection algorithm, but it has its limitations. Future research could focus on improving its accuracy, expanding its object detection capabilities, improving its interpretability, improving its robustness to adversarial attacks, and improving its ability to handle complex scenes.YOLOv5 algorithm can be improved for vehicle detection and tracking achieves high accuracy and efficiency by combining anchor-free and anchor-based approaches [12].

5. Conclusion

This paper presents a modified YOLOv5 algorithm for target tracking detection in videos. The proposed system combines object detection and tracking capabilities, with a motion model to handle occlusions and re-identification of the target object.

The experimental results demonstrate that YOLOv5-based tracking system achieves competitive tracking accuracy and real-time processing speed. The proposed motion model is effective in handling occlusions and re-identification of the target object. These findings suggest that the YOLOv5 algorithm can be a promising solution for target tracking detection in real-world scenarios.

Future work can focus on further improving the tracking accuracy and robustness of this proposed system. This can be achieved by exploring alternative data association methods, improving the motion model, and incorporating temporal information into the tracking process. Additionally, the proposed system can be extended to multi-target tracking scenarios, which are more challenging and relevant in many applications.

In conclusion, the proposed YOLOv5-based tracking system shows great potential for solving the problem of target tracking detection, with applications in surveillance, robotics, and autonomous driving. The proposed system can be a valuable tool for enhancing situational awareness and enabling intelligent decision-making in various domains.

DOI: 10.54254/2755-2721/16/20230860

References

[1] Dixon, T., & Larkin, P. (2018). Video surveillance, the security spectacle and the (in)visible poor: a study of practices in three shopping centres. Surveillance & Society, 16(2), 161-177.

[2] Shrestha, A., & Lin, W. (2018). Target Tracking Detection in Video Surveillance Systems. In 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 2941-2945). IEEE.

[3] Wang, X., Girshick, R., He, K., & Dollár, P. (2020). YOLOv5: End-to-end object detection with YOLO. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (pp. 189-199).

[4] Li, J., Zhang, X., Wei, X., & Wu, Q. (2021). Improved YOLOv5 Algorithm for Object Detection and Tracking in UAV Video. IEEE Access, 9, 74525-74536.

[5] Li, Y., Jiang, B., & Yu, W. (2021). A YOLOv5-Based Multi-Target Tracking Algorithm for Aerial Video Surveillance. Journal of Physics: Conference Series, 1966(1), 012011.

[6] Chauhan, A., & Sharma, M. (2021). Real-Time Object Tracking and Detection Using YOLOv5 in UAV Videos. International Journal of Image, Graphics and Signal Processing, 13(7), 52-63.

[7] Zhang, Z., & Chen, L. (2021). Vehicle Detection and Tracking Based on Improved YOLOv5 Algorithm. Journal of Physics: Conference Series, 1936(1), 012034.

[8] Ren, X., Song, Y., & Lu, L. (2021). Vehicle Detection and Tracking Based on YOLOv5 and Faster R-CNN. Journal of Physics: Conference Series, 1932(1), 012008.

[9] Zhou, W., Jiang, W., & Wang, Y. (2021). Real-Time Human Detection and Tracking in Videos Using YOLOv5. Journal of Physics: Conference Series, 1916(1), 012064.

[10] Zhang, W., Liu, J., Liu, Z., & Zhou, X. (2021). A Lightweight Object Detection and Tracking Method Based on YOLOv5. Journal of Physics: Conference Series, 1916(1), 012073.

[11] Zhou, W., Jiang, W., & Wang, Y. (2021). Real-Time Human Detection and Tracking in Videos Using YOLOv5. Journal of Physics: Conference Series, 1916(1), 012064.

[12] Gao, H., Liu, Z., Zhang, Y., & Zhang, S. (2021). Vehicle Detection and Tracking Based on Improved YOLOv5 Algorithm. Applied Sciences, 11(18), 8303.