University Of AL-Qadisiyah

Available online at www.qu.edu.iq/journalcm

JOURNAL OF AL-QADISIYAH FOR COMPUTER SCIENCE AND MATHEMATICS

ISSN:2521-3504(online) ISSN:2074-0204(print)

Optimizing Vehicle Detection and Tracking Efficiency Through YOLO-Based Multi-Objective Approach

Rabab Farhan Abbas[a], Matheel Emaduldeen Abdulmunim[b]*

[a]*Department of Computer Science, University of Technology – IRAQ, Baghdad , IRAQ.Email: Rabab.F.Abbas@uotechnology.edu.iq*

[b]*Department of Computer Science, University of Technology – IRAQ, Baghdad , IRAQ.Email: matheel.e.abdulmunim@uotechnogy.edu.iq*

ARTICLE INFO

Article history:
Received: 23 /03/2024
Rrevised form: 16 /05/2024
Accepted : 17 /05/2024
Available online: 30 /06/2024

Keywords:

Automotive dataset,
Image analysis,
Vehicle recognition,
Vehicle quantification,
Images.

ABSTRACT

In traffic management, a careful detection of cars plays and monitors a vital role in ensuring safety and efficiency. This paper proposes a way to enhance the discovery of vehicles by dividing roads into remote areas and closing. The newly developed approach aims to define a vehicle and calculate vehicles in both fields. Yolov5M network is easy to discover and localize vehicles, exceed the traditional speed tracking methods and enable microorganisms. The proposed form achieves a higher accuracy of 0.833, outperform the Yolov5 Standard Form of 0.67. It also demonstrates the performance enhancement in the results of summons and mean average precision by keeping the training parameters reduction trend the same. Technically speaking, the split of the roads surface is divided into sections using the latest retail techniques and the fine tuning of the Yolov5M network to positively affect the detection and classification of vehicles.

MSC..

https://doi.org/10.29304/jqcsm.2024.16.21543

1. Introduction

Effective traffic management relies on accurate vehicle detection and control to ensure safety, improve transport systems and support smart traffic management. Traditionally, computer visualization methods such as visual flow, continuous differential video frame and rear systems have been used to identify vehicles by distinguishing between motor vehicles and fixed backgrounds [1]. However, these methods may face constraints in complex traffic situations, such as dealing with different vehicle sizes, complex viewing and lighting conditions. [2].

To tackle these issues an improved vehicle detection and tracking method considering the deep learning and convolutional neural networks (CNN) technology [3] is recommended. We concentrate on a near and far tehcnology giving us the ability to count and identify vehicles at different levels both pedestrain and long distance. [4].

∗Corresponding author

Email addresses: *Rabab.F.Abbas@uotechnology.edu.iq*

Communicated by 'sub etitor'

The main part of our method is the YOLOV5M grid, which is a model for modern object identification, but it subgates the speed for accuracy [5]. Yolov5M is an advanced model that is able to not only detect, localize, and classify vehicles on top of images and videos but also when instep of the lighting differences and complex backgrounds [7]. It presents improvements over conventional approaches by way of the multi-scale learning and Ragner of concerning changes, which include another subclass and change of input images [7].

. By incorporating these advanced technologies, our proposed system aims to achieve higher accuracy and accuracy in vehicle detection, superior to previous models [8]. It provides a powerful and effective real-time traffic monitoring solution, offering potential applications in intelligent traffic systems, autonomous vehicles, and safety management [9].

In short, our research advances to the latest in vehicle detection and monitoring by offering a holistic YOLOV5M-based approach. The proposed method demonstrates significant performance improvements, providing a promising path for smart traffic management and safer routes.

In this paper, we use YOLOv5m network to detect vehicle objects.

2. Related Work

2.1 *Detection of vehicles in images or videos*

Intelligent traffic management and highway control depend on vehicle detection and statistics from road surveillance footage. Y. Yuan et al. [10] gathered a sizable database of traffic video data for analysis using the phone. From a lofty vantage point, the farthest road surface can usually be determined. The car object's size varies significantly at this viewing angle, making it difficult to accurately detect small objects that are far from the road. For intricate video scenes, it is critical to appropriately resolve and apply the previously mentioned issues. In order to provide a workable solution based on the vehicle detection results, concentrated on the aforementioned issues. They Funded a small object far from the road requires a lot of sensitivity. For intricate video scenes, it was crucial to apply and handle the aforementioned challenges in a sufficient manner.

X. Li et al. [11] and J. Canny [12] concentrated on the aforementioned difficulties and applied compound detection findings to multi-target detections and vehicle counting observations. Good outcomes in object detection had been attained with Advanced CNN. CNN was susceptible to changes in object detection scaling, though. The spatial constraints of the network prevented advanced perfection through a two-step approach, particularly for small objects. The one-stage system is one stage for object prediction.

The two-step method groups search item regions into blocks according to specific criteria by using ROIs. The search box was filled with parameters of the specified size if it was smaller than the parameters' specified size. This lessens the precision of a small object's detection and destroys its distinctive structure. Object models did not distinguish whether a small or large object was a part of the same order. Inaccurate detections also result from processing objects of the same type using the same system. The following issues can be resolved by using multi-scale image sets or input images, albeit at a considerable computational cost. A. Ayad et al. [13] proposed revised method called the 1Dimention-DenseNet to statement accidents of the car caused by driver fault, which it achieved an unconditional cross-entropy defeat of 0.19 on the validation set.

2.2 Tracking the movement of vehicles

S. Zhang et al. [14] thorough object identification, such multi-object shading, was a crucial task for Intelligent Transportation Systems (ITS) in the automotive industry. J. Solawetz [15] constructed models based on detection were used by most multi-object shaders. For object initialization, densityless tracking (DBT) and density less tracking (DFT) were used. DBT uses backdrop modelling to characterize objects that move in video frames before tracking them. Shader object initialization is necessary for the DFT system; however, it was impossible to control the retention of current items and the insertion of new ones. The problem of objects between images and the similarity of objects in the image must be taken into account by the multi-object tracking algorithm. A useful tool for determining how similar items was to one another in the environment is regular cross-correlation, or NCC. Currently, removing the line position or the trace position will fix this problem. To get past obstacles SIFT points were used to shade objects, albeit slowly, due to variations in the size and illumination of moving objects. The work suggested a field point detection algorithm, in which the ball can travel much faster and reach better birth points than SIFT. Additionally, smaller generic datasets tailored to particular business scenarios exist. The detection of small objects was imprecise due to the perception of changes in convolutional neural networks.

S. Moosbauer et al [16] produce a road vehicle dataset with high resolution and volume. It was capable of providing detailed descriptions for several vehicle objects in various scenes that were captured from motion areas.

The dataset could be used to compare how well various automobile detection algorithms work when... Control our car's balance. T. Liu et al [17] were improved the precision of vehicle recognition in road traffic scenarios; a tiny object detection technique was employed. In order to detect vehicles, the road traffic area was extracted, split into a peripheral area, and the surrounding area is fed into a convolutional network. M. Gao et al. [18] suggested using multiple objects to monitor and analyze traffic scenes on roads. By extracting and comparing the movement direction of cars and the flow of goods, the ORB method located the road detection line to calculate the feature points of the object to be identified.

3. Proposed System

The proposed system defines the basic architecture and methods of car detection and counting systems. It includes a comprehensive approach to corporate video data processing, real-time vehicle detection and tracking, and data collection and analysis for vehicle certification and classification.

3.1 Data Input and Preprocessing

The system begins to enter video data for companies captured from the surveillance cameras along the roads. Raw video clips are prepared to clean the road area, ensuring an accurate and effective discovery of vehicles.

A large portion of pre -processing includes the Road surface fragmentation using advanced object detection techniques and image processing such as gagi mixture modeling. This technology allows the rear overlap from 500 basic video frameworks. Each pixel is analyzed in each frame, and its value is formulated as a Gaous distribution that focuses on a specific value over time. The pixel units that deviate greatly from the medium are an introduction, while those in a specific range of average are considered the background [19].

The resulting background image is more processed using Gaussian 3x3 candidate to calm smaller color differences and neutralize color distribution. This facilitates the surface of the road using an algorithm filling the area, which begins to choose the seed point on the side of the road and gradually fill the areas surrounding similar pixels. This retail allows this accurate estimate to display the road and the number of existing cars, which are found to be 11 cars and one truck in this study.

Fig .1 - Road the surface area of the extraction process.

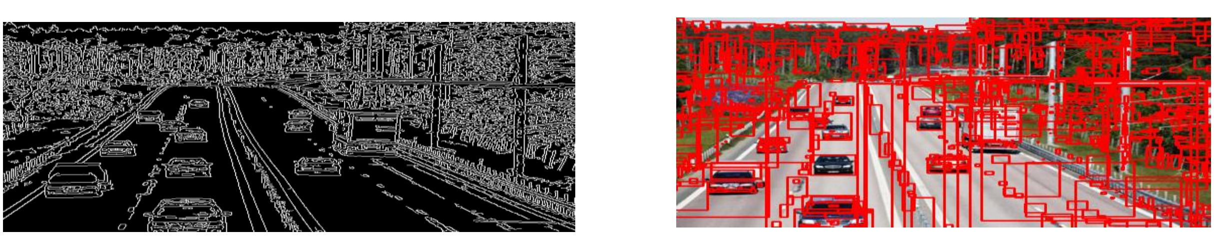

Fig .2 - The process of extracting the number and type of cars.

3.2 Real-Time Vehicle Detection Using YOLOv5

Once the data is already processed, the next step includes the discovery of vehicles in actual time using the Yolov5 network. This network is part of the one -stage family and excels in speed while maintaining a high level of accuracy.

The Yolov5 network consists of three coherent layers. The first section, also known as the spine or CSPDADARKNET, includes joint operations in the censive nerve networks (CNNS) such as notes, activation and maximum assembly. It also includes a simple guidance mechanism that records many features for further treatment.

CSPDARKNET network improves gradient information processing in large -scale images, which reduces the number of parameters and floating-point operations per second. These improvements lead to a faster and more accurate discovery by Yolov5.

In the past, locating those small objects may be tricky. However, the Yolov5 algorithm is powerful enough to solve this by keeping a focus or even exclusively identifying vehicles. Its smaller scale works and all the corrections and additions that he incorporated into the blueprint offer a good example of his fine-tuning of minute details. The model can simultaneously confirm different features in earlier steps while also further developing this model in subsequent stages, ultimately leading to improved detection accuracy.

The primary mechanism in Yolov5 is the so-called focus layer which segments the input image into smaller areas in order to weed them out out and perform the analysis. The network enjoys this network with a higher ability of picking details about the objects on the scene and of depicting things sharply.

The manner of increasing the precision in the detection of smaller objects involves additional insertion of the anchor points. Such anchors help the network configurations that makes smaller network cars unique and distinct thus no loss of data is encountered through the network.

The Yolov5 network has been trained in a specific data collection containing different types of vehicle images. This allowed the network to identify different types of vehicles with high accuracy. The model can classify vehicles into three different categories: buses, trucks and exchange, as shown in the data set. Motorcycles were not within the scope of this study, and therefore they were not included in the detection.

3.3 Data Gathering and Analysis

The final stage involves data collection and analysis of vehicle detection and tracking results. The coordinates of the ocean box obtained from the YOLOV5 network are used to localize vehicles accurately in the original images.

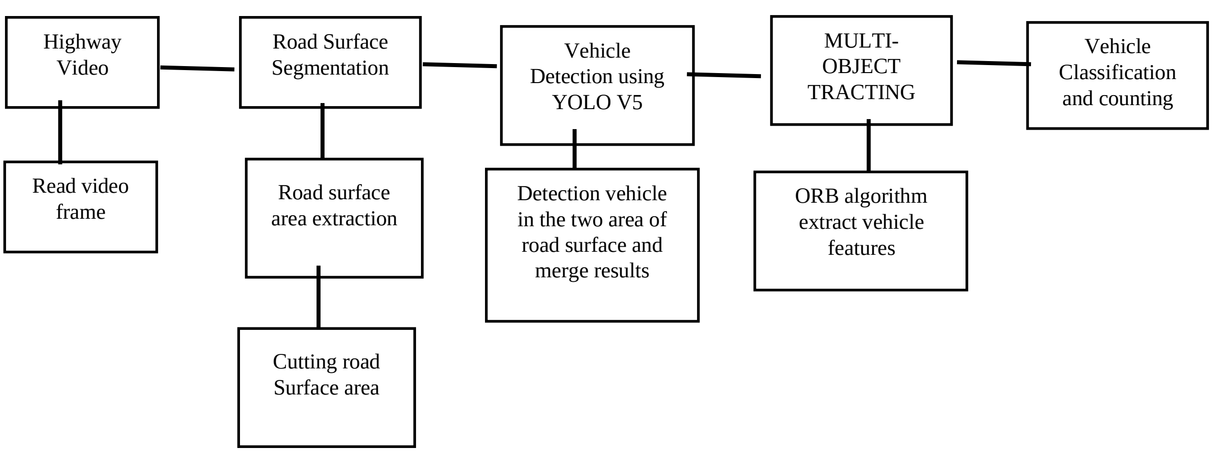

Fig. 3- Illustration of road surface segmentation for highway area estimation

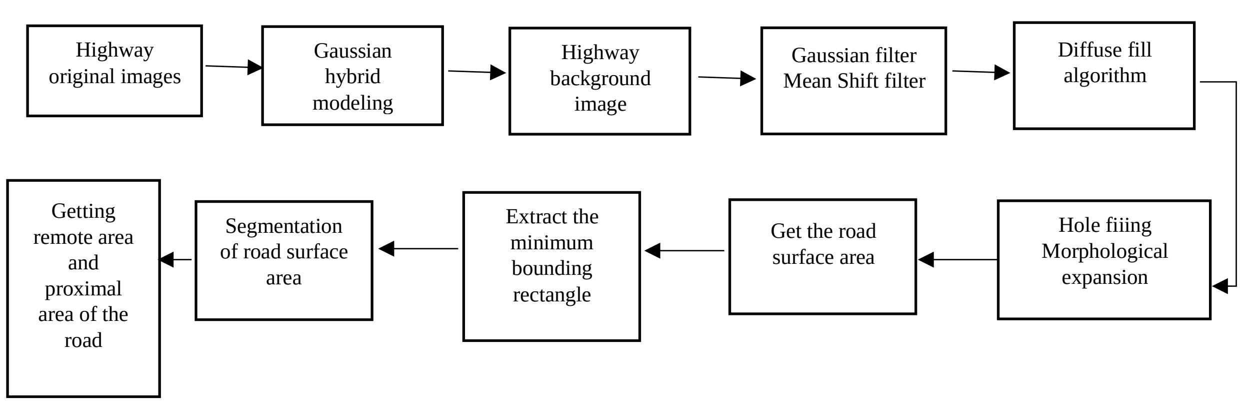

Fig .4 - The Sfera algorithm for multi-object tracking.

After accurately locating the vehicles, the system collects data on the number of vehicles, type and movement. This data is crucial to tracking and analyzing the direction and behavior of vehicles. The field algorithm is used to negotiate mapping between multiple video organisms and frames, which increases the accuracy of organism tracking and data collection.

Data analysis includes the study of pixel concentrations and graphic drawing for decline to improve digital image processing, such as brightness and contrast control, enhance the unification of lighting and details in different regions.

Table 1 - The value of the peaks Grayscale histogram

Rank	Pixel Value	Frequency
1	191	2365
2	192	2364
3	193	2308
4	190	2307
5	189	2213
6	194	2213
7	188	2185
8	187	2155
9	195	2118
10	196	2052
11	186	2048
12	198	2027
13	184	2017
14	197	2012
15	219	1994
16	200	1985
17	199	1949
18	202	1948
19	185	1944
20	201	1926

In conclusion, the proposed system effectively leverages YOLOV5 for real-time vehicle detection and tracking. Through data collection and analysis, the system provides valuable insights for vehicle calculation and monitoring, contributing to improved traffic management and road safety.

4. The Experimental Data

4.1 Data Repository

We may split business image collections into three categories based on how they obtained: images captured by cameras, images captured via surveillance cameras, and images captured by non-surveillance cameras. [21]. With images from both urban and rural scenes, the KITTI standard dataset [22] is an extensive dataset created especially for tasks like autonomous driving and 3D object detection and tracking. Images taken by cameras in a variety of lighting and weather conditions make up the Tsinghua Tencent business sign dataset [23], though no vehicle annotations are present. However, this dataset offers comprehensive details about the cars, including their model, make, and manufacturing date. It, however, do not have many pictures. As an example, there is a collection of 28,300 images showcasing various attributes of vehicles, including their top speed, the number of doors, and the number of seats, class, and type of engine. Additionally, 150,200 photos focus on the overall aesthetic of the cars. Another dataset called the BIT-Vehicle Dataset [24] consists of 10,000 images and classifies vehicles addicted to six groups: SUV, sedan, minivan, truck, motorcycle, and microbus. However, this dataset has certain limitations that make it challenging to use for training convolutional neural networks (CNNs), such as the positive firing angle and the small size of the vehicle objects. In this section, we will introduce a vehicle dataset specifically created from trace surveillance videos.

Similar to the previous dataset, this one also contains a large number of photos providing details on the vehicle's top speed, number of doors, number of seats, and other features, as well as categorization and engine type.

The photographs included in the dataset offer a thorough depiction of the cars' overall looks. Surveillance cameras are used to gather these datasets. The BIT-Vehicle Dataset [25], which is an image collection, is one prominent example. Six categories are used in this dataset to classify vehicles: SUV, sedan, minivan, truck, motorcycle, and microbus. It is crucial to remember that the photos have a positive firing angle and that the car objects are too small to be usefully generalized for CNN training. We will present the vehicle dataset that we created using trace surveillance footage in this section. The dataset includes images taken from roads in Yangshupu and Lin-gang, China.

A 12-meter-tall surveillance camera that mounted on the side of the road observed the traffic. This camera was not stationary; instead, its field of view could be adjusted. The photos taken from this angle show a considerable stretch of the trace as well as vehicles with notable scale variations. The images in the dataset were taken with phone cameras in a variety of settings, at different times of the day, and with different lighting. This dataset's vehicles are divided into three kinds: trucks, cars, and buses (Fig. 11). The associated annotations for every image are kept in a text document and consist of the standard bounding box coordinates and the numerical designation of the vehicle category. Tables 1 gives a summary of the dataset, which has all of the images along with their corresponding bounding boxes. The photos are 1920 x 1080 pixels in resolution and are in RGB format. Notably, the annotations are mainly focused on the objects that are right next to the camera, which leads to vehicles that are different sizes. Features on objects are larger when they are closer to the camera and smaller when they are farther away. Including annotated cases with varying sizes improves the accuracy of detecting small vehicle objects [26]. Our dataset can be applied in multiple domains, including European ones, and provides a comprehensive resource for vehicle detection.

In YOLOv5, the direct calculation of distances between objects is not a primary function of the network. YOLOv5 primarily designed for tasks of object detection and localization, rather than distance estimation. Nevertheless, once the objects, such as cars, have been detected and localized within an image, additional methods can be applied to estimate distances between them.

To estimate distances between objects, one can follow these general steps:

1. Object Detection: Employ this method to locate objects of interest within the input image. The method provides coordinates for each detected object.

2. Perspective Transformation: Correct for any distortions caused by the camera's viewpoint. In this step the crucial for accurate distance calculations when the camera is not calibrated.

3. Bounding Box Overlap: Measure the degree of overlap between bounding boxes of detected objects using metrics like Intersection over Union (IoU).

4. Depth Estimation: Obtain depth information through methods such as stereo vision or depth sensors to determine actual physical distances between objects.

5. Scaling: Convert the measured overlap between bounding boxes into physical distances by use known depth information. This step requires camera calibration and an understanding of camera specifications.

It is important to word that distance estimation accuracy depends on elements together with object detection first-rate, perspective correction, intensity estimation, and digital camera calibration. Additionally, superior functionalities, inclusive of precise form reputation or object counting, might also require additional techniques.

Fig .5 - Vehicle Labelling Categories in the Dataset.

4.2 Tracking multiple objects

This segment demonstrates the utilization of object containers identified inside the previous 'Vehicle Detection with YOLOv5' segment for monitoring a couple of gadgets. The take a look at hired the ORB (Oriented FAST and Rotated BRIEF) technique to extract exceptional features from the recognized motors, which yielded promising results. The ORB set of rules proves to be a appropriate alternative to photo description strategies like SURF (Speeded-Up Robust Features) and SIFT (Scale-Invariant Feature Transform), because it well-knownshows superior performance in terms of matching expenses and computational performance compared to different algorithms.

The ORB approach utilizes the Harris operator to identify corners and the Features from Accelerated Segment Test (FAST) to stumble on feature points. Once the characteristic factors are received, descriptors are computed the use of the BRIEF (Binary Robust Independent Elementary Features) approach. To make sure rotation consistency within the function point descriptors, a coordinate machine is created using the centroid of the point place because the x-axis and the function factor because the middle of the circle. This lets in the coordinate gadget to be rotated along side the photo. Consequently, despite the fact that the image's angle changes, a solid factor can nonetheless be recommended.

The monitoring technique the use of the ORB method involves setting up a threshold for the variety of matching factors obtained. If the whole range of matching points exceeds this threshold, a factor is considered to have a successful fit, and the item's matching box is drawn. The source of the prediction box feature is like this:

For point validation, the Random Sample Consensus (RANSAC) algorithm is employed to remove false noise points from the matching computations. Additionally, an estimation of the homography matrix is made. Subsequently, a perspective transformation is performed using the estimated homography matrix and the original object detection box's location to create an associated prediction box.

The vehicle detection algorithm generates an object detection box, from which point points are selected using the Sphere fashion method. This method ensures that object features are not taken from the entire surface area of the road, as this can significantly impact the results.

The item's vaticination box is drawn inside the following frames with item shadowing taken under consideration. This is because, because the Sphere factor recaptured from the item container indicates, there is a moderate trade in the automobile object among successive frames of the video. It is still viable to track the equal object in spite of these minor changes.

All things considered, the RANSAC algorithm, homography matrix estimation, and Sphere point selection work together to provide precise object box matching and estimation, which makes object tracking in the video sequence easier.

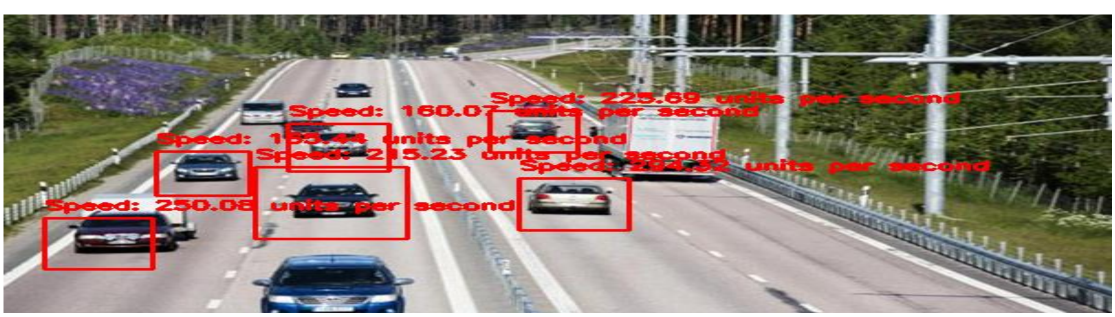

Fig .6 - The Process of multi-object tracking.

We take the shortest distance criterion of the center point into consideration to ascertain whether the vaticination box and discovery box of the upcoming frame match (as illustrated). A match declare when the distance between the center points of the discovery box and the vaticination box is less than a predetermined threshold. The tracking system's particular requirements use to determine this threshold.

The vaticination box exactly matches the item within the subsequent frame while the item bins are matched using the shortest distance criterion. The object's movement or modifications between frames may be detected through the device through comparing the positions of the middle factors.

This matching procedure is essential to preserving the accuracy and flow of object tracking, enabling trustworthy observation and evaluation of the object's motion and behavior during the film clip.

$$T = \frac{boxheight}{2.5} \qquad (1)$$

There is a method used to eliminate lines that haven't changed in ten frames when the scene is captured from a wide angle of a highway camera. This method works especially well in situations where the camera is capturing a steep road face.

A line is deemed redundant and eliminated when tracking a truck or any other object in consecutive videotape frames if it stays the same for ten frames. This is because it is anticipated that the truck will travel farther in these ten frames, and an unaltered line signifies that the line is no longer pertinent to the truck's present location.

The line removal threshold setting does not significantly influence the final counting result. The line removal threshold setting does not significantly influence the final counting result.

This is due to the fact that the vehicle line and the discovery line only ever briefly cross. As a result, the accuracy of the counting process is unaffected even if the threshold is not set precisely.

On the other hand, if the prediction box does not match in a sequence of frames, it is considered that the object (in this case, the truck) is absent from the video scene. Consequently, the prediction box is canceled.

Through the application of these methods, the system attains tracking circles and global object discovery findings from the full trace monitoring videotape perspective. By ensuring precise and dependable object tracking throughout the video sequence, this antedating operation improves the surveillance system's overall efficacy.

5. Results And Disisions

The YOLOV5 model is designed primarily for the detection and localization of objects, and although it does not focus exclusively on the speed of tracking or calculating objects, it offers significant benefits in these areas when combined with additional algorithms and methods. Below, we provide a detailed analysis of YOLOV5's capabilities in various scenarios, including strengths, limitations, and potential applications.

5.1 Object Detection

Yolov5 is particularly good at object detection and classification tasks; in particular, it successfully works with photos and videos which are about vehicles, pedestrians and animals. The examined machine discovered 12 elements of the road, covering an area of about 1044 square meters, and divid ed it into 11 cars or trucks. This demonstrates the effectiveness of YOLOV5 in counting objects; makes it possible to t rack objects as they enter or exit the frame; This is useful for monitoring people for crowd control, traffic analysis an d inventory management.

5.2 Speed Tracking

In addition to the ability to detect and locate moving objects such as cars, people or animals in videos or photos, it can also detect people walking in low light or in the dark. Obtaining this information requires tracking where thes e species follow tires and using the time difference between tires to estimate this. Although Yolov5 is not specifically designed for speed tracking, it can provide accurate results and tracking features, and this information can be used i n conjunction with other algorithms to calculate speed. Technologies such as visual streaming and motion detection can be combined with Yolov5's detection results to obtain accurate predictions of vehicle movements.

The table below shows the speeds of the different vehicles included in this study:

Table 2 - Results of cars speed

Car	Speed
1	250
2	215
3	205
4	150
5	194

5.3 Shape Identification

After training on a diverse dataset, YOLOV5 can recognize and classify most objects based on their shape. YOLOV5's accuracy in distinguishing between cars and trucks based on their unique shapes contributes to its effectiveness in various real-world applications. For example, in cases of legal liability involving traffic accidents, YOLOV5 shape recognition capabilities can help to create the respective vehicle types and situations.

5.4 Potential Applications and Enhancements

To increase YOLOV5's capabilities and achieve specific goals such as speed monitoring, shape recognition, or complex object counting, it can be combined with other computer vision strategies, algorithms, and domain-specific knowledge. For example, integrating YOLOV5 with depth estimation techniques or multi-object tracking algorithms can provide more comprehensive and accurate analysis.

Concrete examples of YOLOv5's applications include:

- Traffic management: Yolov5 detection and classification of Yolov5 can help monitor traffic in actual time, allowing improving improved traffic flow and congestion management.
- Safety and accident analysis: Yolov5 can be used to study road safety, accident scenarios analysis, and to provide visions about the behavior of the car before and during accidents.
- Self -government compounds: The detection of fast and microorganisms in Yolov5 can be used in independent compounds to identify obstacles in actual time, help in mobility and decision -making.

Quantitative results and comparisons with other models, along with references to recent studies, can increase the validity of YOLOV5's effectiveness and highlight areas for improvement. By providing detailed analysis and examples, we can better understand the strengths and limitations of YOLOV5, enabling researchers and practitioners to tap into its full potential in different scenarios.

6. Conclusions

This study created a large collection of data from the camera from the camera perspective, focusing on vehicles and enabling the development of the detection system of mobile phone objects and its shading to effectively monitor video scenes. The establishment of a reference area of the road to be monitored shows a practical and effective approach to analyzing the video scene.

The Yolov5 object detecting the essence of the vehicle tracking data format model on the basis of the vehicle tracking data sets. The search results showed that the proposed method is possible and effective to discover vehicles and monitor scenes that were captured by phone cameras.

The average Yolov5 model uses multi -range attention mechanisms, achieving a higher performance compared to the basic Yolov5 model. Specifically, the modified model has reached 0.833, while the basic model achieved 0.67 resolution. This improvement was achieved in accuracy while maintaining a similar number of training parameters.

In addition, the proposed model showed improvements in the recall and average medium accuracy (MAP). This performance improvement highlights the possibilities of the model for real world applications, such as traffic control, management and self -government navigation.

However, it is still possible to improve the proposed detection and monitoring framework. Future research should focus on improving the speed and performance of the model by developing vehicle detection devices faster specifically designed to use in traffic scenes. Researchers can also explore the integration of modern computer vision methods or take advantage of additional data collections to increase the enhancement of model capabilities.

Moreover, the conclusions extracted from this study are supported by the performance measures and notes obtained during the research. These results are the basis for improvements and future progress in smart traffic control systems.

References

[1] T. Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan and C. L. Zitnick, "Microsoft Coco: Common Objects in Context," in *13th European Conference on Computer Vision*, Zurich, 2014. https://doi.org/10.48550/arXiv.1405.0312

[2] M.A. Kadhim and A. M. Radhi, "Heart disease classification using optimized Machine learning algorithms," *Iraqi Journal for Computer Science and Mathematics*, vol. 4, no. 2, pp. 31- 42, 2023. https://doi.org/10.52866/ijcsm.2023.02.02.004

[3] R.F.Abbas and M.E. Abdulmunim, "Residual Network with Attention to Neural Cells Segmentation," *Iraqi Journal of Science*, vol. 64, no. 4, pp. 2023-2036, 2023. https://doi.org/10.24996/ijs.2023.64.4.37

[4] J. Redmon and A. Farhadi, "Yolov3: An Incremental Improvement," in *IEEE conference on computer vision and pattern recognition*, 2017. https://doi.org/10.48550/arXiv.1804.02767

[5] Y. Bin, F. Pan, L. Xiaoyan, L. Zhijie, and Y. Fuzeng, "A Real-Time Apple Targets Detection Method for Picking Robot Based on Improved YOLOv5," *Remote Sens,* vol. 13, no. 9, pp. 16-19, 2021. https://doi.org/10.3390/rs13091619

[6] S. A. Magalhaes, L. Castro, G. Moreira, F. N. Santos, M. Cunha, J. Dias, and A. P. Moreira, "Evaluating the Single-Shot MultiBox Detector and YOLO Deep Learning Models for the Detection of Tomatoes in a Greenhouse," *Sensors,* vol. 21, no. 10, pp. 35-69, 2021. https://doi.org/10.3390/s21103569

[7] G. Gao, S. Wang, C. Shuai, Z. Zhang, S. Zhang, and Y. Feng, "Recognition and Detection of Greenhouse Tomatoes in Complex Environment," *Traitement du Signal,* vol. 39, no. 1, pp. 291-298, 2022. https://doi.org/ 10.18280/ts.390130

[8] M. E. H. Chowdhury, T. Rahman, A. Khandakar, M. A. Ayari, A. U. Khan, M. S. Khan, N. Al-Emadi, M. B. I. Reaz, M. T. Islam, and S. H. M. Ali, "Automatic and Reliable Leaf Disease Detection using Deep Learning Techniques," *AgriEngineering,* vol. 3, no. 2, p. 294–312, 2021. https://doi.org/10.3390/agriengineering3020020

[9] S. Shim and G. C. Cho, "Lightweight Semantic Segmentation for Road-Surface Damage Recognition Based on Multiscale Learning," *IEEE Access,* vol. 8, no. 15, pp. 102680-102690, 2020. https://doi.org/ 10.1109/ACCESS.2020.2998427

[10] Y. Yuan, M. S. Islam, Y. Yuan, S. Wang, T. Baker and L. M. Kolbe, "EcRD: Edge-Cloud Computing Framework for Smart Road Damage Detection and Warning," *IEEE Internet of Things Journal,* vol. 16, no. 8, pp. 12734-12747, 2020. https://doi.org/10.1109/JIOT.2020.3024885

[11] X. Li, S. Lai, and X. Qian, "DBCFace: Towards Pure Convolutional Neural Network Face Detection," *IEEE Transactions on Circuits and Systems for Video Technology,* vol. 4, no. 32, pp. 1792-1804, 2021. https://doi.org/10.1109/TCSVT.2021.3082635

[12] J. Canny, "A Computational Approach to Edge Detection," *IEEE Transactions on pattern analysis and machine intelligence,* vol. 8, no. 6, p. 679–698, 1986. https://doi.org/10.1109/TPAMI.1986.4767851

[13] A. Ayad and M.E. Abdulmunim, "Detecting Abnormal Driving Behavior Using Modified DenseNet," *Iraqi Journal for Computer Science and Mathematics,* vol. 4, no. 3, pp. 48-64, 2023. https://doi.org/10.52866/ijcsm.2023.02.03.005

[14] S. Zhang, C. Chi, , Z. Lei, and S. Z. Li, "Refineface: Refinement Neural Network for High Performance Face Detection," *IEEE transactions on pattern analysis and machine intelligence,* vol. 43, no. 11, pp. 4008-4020, 2020. https://doi.org/ 10.1109/TPAMI.2020.2997456

[15] J. Solawetz, "YOLOv5 Improvements and Evaluation," https://blog.roboflow.com/yolov5-improvements-and-evaluation , 2020.

[16] S. Moosbauer, D. Konig, J. Jakel and M. Teutsch, "A Benchmark for Deep Learning Based Object Detection in Maritime Environments," in *In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2019. https://doi.org/ 10.1109/CVPRW.2019.00121

[17] T. Liu, B. Pang, L. Zhang, W. Yang and X. Sun, "Sea Surface Object Detection Algorithm Based on YOLO v4 Fused with Reverse Depth wise Separable Convolution (RDSC) for USV," *Journal of Marine Science and Engineering,* vol. 9, no. 7, p. 753, 2021. https://doi.org/ 10.3390/jmse9070753

[18] M. Gao, G. Shi and S. Li, "Online Prediction of Ship Behavior with Automatic Identification System Sensor Data using Bidirectional Long Short-Term Memory Recurrent Neural Network," *Sensors,* vol. 18, no. 2, p. 4211, 2018. https://doi.org/ 10.3390/s18124211

[19] Z. Shao, W. Wu, Z. Wang, W. Du and C. Li, "SeaShips: A Large-Scale Precisely Annotated Dataset for Ship Detection," *IEEE transactions on multimedia*, vol. 20, no. 10, pp. 2593-2604, 2018. https://doi.org/ 10.1109/TMM.2018.2865686

[20] Y. Zhang, Q. Z Li, and F. N. Zang, "Ship Detection for Visual Maritime Surveillance from Non-Stationary Platforms," *Ocean Engineering,* vol. 141, no. 7, pp. 53-63, 2017. https://doi.org/ 10.1016/j.oceaneng.2017.06.022

[21] M. Nalamati, N.Sharma, M. Saqib and M. Blumenstein, "Automated Monitoring in Maritime Video Surveillance System," in *35th International Conference on Image and Vision Computing New Zealand (IVCNZ)*, 2020. https://doi.org/ 10.1109/IVCNZ51579.2020.9290533

[22] H. Zhang, M. Cisse, Y. N. Dauphin and D. Lopez-Paz, "Mixup: Beyond Empirical Risk Minimization," in *6th International Conference on Learning Representations*, Canda, 2018. https://doi.org/10.48550/arXiv.1710.09412

[23] Z. Cai and N. Vasconcelos, "Cascade R-CNN: Delving into High Quality Object Detection," in *IEEE conference on Computer Vision and Pattern Recognition*, 2018. https://doi.org/ 10.1109/CVPR.2018.00644

[24] J. Redmon, S.Divvala, R . Girshick and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2016. https://doi.org/10.1109/CVPR.2016.91

[25] Z. Tian, C. Shen, H. Chen and T. He, "FCOS: Fully Convolutional One-Stage Object Detection," in *IEEE/CVF international Conference on Computer Vision Workshops*, 2019. https://doi.org/ 10.1109/ICCV.2019.00972

[26] Z. Dong, Y. Wu, M. Pei, and Y. Jia, "Vehicle Type Classification using a Semisupervised Convolutional Neural Network," *IEEE transactions on intelligent transportation systems,* vol. 16, no. 4, p. 2247–2256, 2015. https://doi.org/ 10.1109/TITS.2015.2402438