Object Recognition for Organizing the Movement of Self-Driving Car

Wasin AlKishri, Mahmood Al-Bahri

***Abstract*— Today there is a revolution in the automotive industry. Cars are becoming self-driving with advanced sensors, cameras, and recognition algorithms. Algorithms and scenarios are evolving and improving every day, but it is too early for self-driving cars to go out on public roads. The work of recognition algorithms is far from ideal. For the correct and synchronized operation of all elements of an unmanned vehicle, a person needs to transfer all his intellectual experience to the computer systems of the vehicle. The developers are working to ensure that the car can see and understand what is happening around it. Such cars are already driving on the roads in test mode with a pilot, receiving a huge amount of information and learning. Already, tests and trials have proven that unmanned vehicles are safer than vehicles driven by people, and after mass implementation, the death rate on the roads will be reduced several times. Optical vision is an essential component of self-driving car due to the absence of human control. Speed, accurate detection of vehicles, empty parking, pedestrians, traffic signals, streets, and road signs can help self-driving vehicles drive safely and avoid mistakes and road accidents. On the other hand, object identification was challenging because objects in the physical world are affected by strong and low luminance, angularity, and scaling in image capture. The Convolutional Neural Network (CNN) approach was used by several researchers to improve object recognition outcomes in a variety of situations. However, the main disadvantage of this method is being unable to react quickly in real-time scenarios. In this work, the focus was on the YOLO model and the features of its work. Tests and comparisons with other state-of-the-art were given, which show that YOLO is the fastest model for real time objects recognition today. The architecture of the algorithm and the essential components on which the field experiment was conducted are described. Based on the results obtained, it was concluded that the YOLOv4 recognition model shows excellent results in terms of accuracy and speed.**

***Index Terms*— YOLOv3, YOLOv4, Recognition, Unmanned Vehicles, self-driving Car**

I. Introduction

Over the past decade, artificial intelligence and deep learning have become the main technologies behind many breakthroughs in computer vision. Classical sensing, path planning and motion control techniques cannot solve driving scenarios that machine learning and artificial intelligence can easily handle. These technologies are the central components of autonomous driving [1].

Self-driving vehicles provide tracking, computation, and wireless communication technologies. They have on-board units (OBUs) and global positioning system (GPS) receivers, on-board radio modules such as IEEE 802.11p, LTE or 5G modules, and other on-board units. These on-board units map the environment by receiving data from cameras, lidars, radars, ultrasonic sensors, and vehicle motion sensors, and perform calculations and data exchange [2]. Laser Radar (LiDAR) creates a 3D map of the environment by scanning it with a full 360-degree view. This map allows the vehicle to determine distances to various objects. Cameras allow vehicle systems to see signs, objects, road markings and many other elements on the road. The sensors of an unmanned vehicle generate a huge amount of data that is analyzed by machine vision, neural networks, and other software. The generated data must be transmitted without delay and at high speed between other vehicles. This requires the introduction of a new high-speed 5G standard, advanced road transport and telecommunications infrastructure. To ensure the movement of autonomous vehicles, such infrastructure components are required as:

- ITS (Intelligent Transport System).
- V2X-platform as part of the ITS. Vehicle-to-Everything (V2X) is the concept of a vehicle interacting with everything around it.
- Precise positioning system.
- Digital model of roads based on high-precision digital dynamic road maps.
- Stable coverage with high-speed communication channels [3].

The fifth generation of cellular networks (5G) was designed to answer communication challenges by providing a low latency, high reliability, and high bandwidth communication infrastructure. Connected vehicles will be able to work together and organize correct and accident-free traffic [4].

B This paper discusses methods of object recognition by unmanned vehicles. Classical recognition methods and models based on neural networks are considered. The operation of the YOLO (You Only Look Once) model is described in detail, as well as a natural experiment. Based on the results, conclusions were made about the performance of the models and recommendations for use.

Corresponding Author: Wasin Alkishri wkishri@su.edu.om
Department of Computing and IT Sohar University Sohar,
Sultanate of Oman.

Dr. Mahmood Al-Bahri , mbahri@su.edu.om
Department of Computing and IT Sohar University Sohar,
Sultanate of Oman.

II. V2X TECHNOLOGY OVERVIEW

Vehicle-to-everything (V2X) describes the communication network system between the vehicles for the purpose of connecting and exchanging data with other resources and sensors over the internet within high reliability, bandwidth, and low latency. These technical specifications facilitate the operation of fully autonomous driving without the need for human interaction. V2X is composed of various components, firstly, Vehicle-to-Vehicle (V2V) enables the communication between all types of vehicles, such as cars and trucks. Whereas Vehicle-to-Infrastructure (V2I) enables vehicles to recognize empty parking lots and buildings, red light from the green light in a traffic light. Vehicles-to-pedestrian (V2P), this type allows vehicles to recognize pedestrians and distinguish them in the pedestrian crossing lanes and communicate with them through their phones and smartwatches. After collecting data from various sensors and sources, vehicles need to set up a communication network for this information to be exchanged between pedestrians and vehicles through a Vehicle-to-Network (V2N).

The Multidimensional Ecosystem scenario below visualizes the process of self-driving cars interacting with each other, with infrastructures such as traffic lights or parking spaces, pedestrians using smartphones, and data storage through networks. The establishment of requirements depends on use cases. However, the vehicles communication system must handle reliably and cost-effectively. Figure 1. depicts the different types of V2X communications.

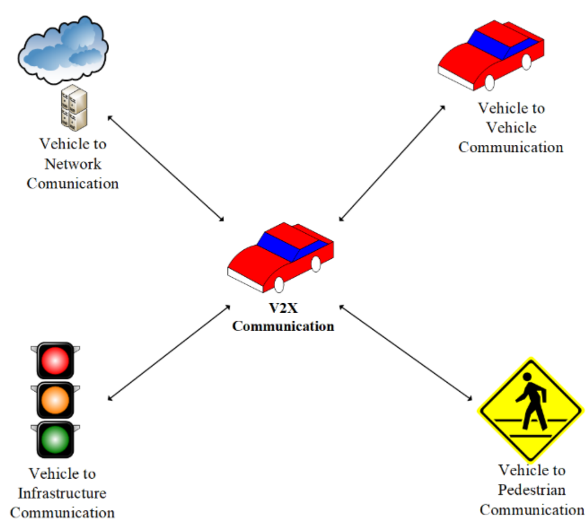

Figure 1. V2X communication types

Although connected and autonomous vehicle technologies are often presented as separate technologies, but they are orthogonal. The effective link between these two technologies enhances their interaction to implement an intelligent transport system better. (V2V) and (V2I) communications allow vehicles to exchange information with each other and with road infrastructure hardware devices such as RSUs to serve various applications. Vehicle-infrastructure collaboration and image processing can be applied to apps [15]. For example, for a collaborative group where cars share their mobility data to maintain group integrity or a collaborative traffic information system, vehicles share their planned traffic projection beacons [16]. Likewise, autonomous cars can also interact with other cars in other situations. For instance, distributing sensory information with road users will make driving safer and help them in different situations such as road congestion or road accidents. Collaboration between autonomous vehicles can have many other primary benefits for instance, a combination health evaluation and the exchange of lane information between self-driving cars will improve navigation efficiency, smoothness, and aid autonomous vehicles if one of their subsystems fails [17]. Other applications of autonomous collaborative communication include collaborative localization through optimized sensor configuration and vehicle-to-vehicle traffic coordination [18].

One of the main criteria standards of V2X is the Wi-Fi fork which is connected over IEEE 802.11p. This is classified as part of the IEEE WAVE or Wireless Access Program (for the automotive environment), and it also operates in the unauthorized frequency band of 5.9 GHz. The IEEE 802.11p protocol, which was completed in 2012, serves as the foundation for DSRC, the short cut of Dedicated Short Range Communications that be in the US and ITS-G5 thru the European Cooperative Intelligent Transport Systems (C-ITS) project in Europe. In addition to V2X that connectivity stretches through line-of-view sensors over 802.11p technologies such as cameras, radar and LIDAR and also their counterparts (V2V/V2I applications) such as electronic toll collection systems, wireless expansion options for V2X include environmental sensing, traffic sign alarms, and crash notification, speed limit notices and automated car parks. The functional properties of 802.11p involve range (which is shorter than 1 kilometer), low latency (2 ms and less), and durability at medium and long distances (1+ kilometers). combats driving needs such as subzero temperatures, rain, on-air suspension, long rough roads, off-road driving, mud, dust, heat, wheel vibration and, etc.”. IEEE 802.11p gives cars a better view of their surroundings and under difficult and harsh weather conditions [14]. IEEE 802.11p is independent of cellular coverage and on-board device (OBU) and roadside unit (RSU) solutions [10].

As with the 5G wireless in-vehicle communication, the leaders (i.e. Cellular V2X or 5G Automotive Association proponents) have expressed a need for a successor to 802.11p, the IEEE 802.2p could be an option. Another big benefit of the C-V2X is it has a number of alternate configurations that allow the vehicle to be properly prepared for almost any potential mission or scenario [15]. The first mode is C-V2X which is considering as a direct connection with s a low-latency around 5.9 GHz bands, which is unlicensed. This mode is optimized for active safety messages including immediate road alerts such as "Immediate Turn Right Here" and other close V2V, V2I, and V2P scenarios. This mode can be similar to the existing IEEE 802.11p technology, which also operates in the 5.9 GHz band. The second mode is communication over a regular cellular network with licensed bandwidth via the Uu interface. It is capable of handling V2N usage cases such as infotainment and delay-tolerant safety alerts about longer-term road dangers or traffic situations. IEEE 802.11p will only cope with this mode by establishing ad hoc connectivity to roadside base stations because it does not use cellular connectivity [10]. When 5GAA, the alliance that supports C-V2X, first started operations in 2017, it had eight founding members. It currently has a membership of 120 people. Audi, BMW, Daimler, Ford Motor, GM, Honda, Hyundai, Nissan, Volkswagen, and Volvo are among those taking part. So do technology giants such as Intel, Samsung, and Qualcomm; automakers such as Alpine, Continental, and Bosch; network equipment suppliers such as Nokia and Ericsson; and networks such as AT&T, T-Mobile, Verizon, and Vodafone. C-V2X is now operational on modern 4G networks. C-V2X infrastructure, developed by 3GPP, an industry body that sets wireless networking protocols, helps cars to relay simple driving information over 4G networks [11,12].

III. OBJECT RECOGNITION BY UNMANNED CARS

A. Object Recognition Architectures

An appropriate model is required for an experiment to recognize different objects, which will provide the necessary speed and accuracy. Thus, the models based on convolutional neural networks were chosen, since they are the most suitable for the goals set in the framework of this work. When recognizing objects, the YOLO (You Only Look Once) models will be used, since they are the most optimal in terms of accuracy and speed [5].

YOLO is the first attempt at developing a rapid, real-time object detector. YOLO is a one-step algorithm for detecting objects. Region-Based Convolutional Neural Network R-CNN model is used to detect the object by extracting features from the regions. It starts by retraining a pre-trained convolutional neural network according to the number of classes to be detected. Then extracts and reshape the Region of Interest for each image. the SVM classifier is used to filters and divide the obtain region into different classes. finally, the R-CNN algorithm will return the boundary boxes for the specified classes concurrently. This algorithm is the fastest and easiest, but it has the potential to slow down performance a bit. That is, the YOLO model does not pass the proposal stage for the region and only expects a limited number of specified frames, and as a result it can quickly make an assumption [6].

There are four versions of the YOLO model. The fourth version was released in 2020. With each version, the architecture changed and the models became more accurate. The work in this paper is limited to the third and fourth versions. So we will not mention the first and second versions.

YOLO version 3 was released in April 2018 and includes some minor enhancements, including the ability to forecast bounding boxes at various scales. YOLO uses the Darknet variant of this version, which has a 53-layer network trained on ImageNet. For the detection mission, another 53 layers are added, giving YOLO v3 with 106-layer fully convolutional base architecture. This is the main cause of YOLO v3's slower performance as compared to YOLO v2. Figure 2 shows the architecture of the YOLO network of the third version [7].

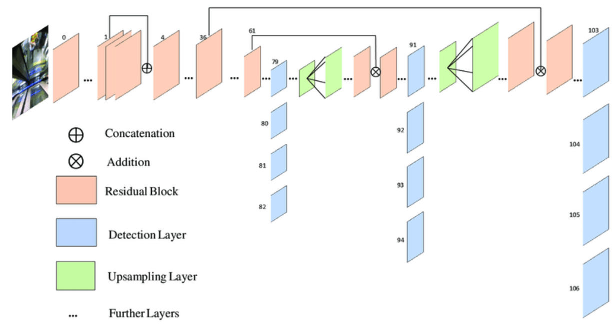

Figure 2. Architecture of YOLO v3

In general, YOLOv3 outperforms and faster than the SSD (Single Shot Detector) model, and worse than RetinaNet, but 3.8 times faster [8]. Figure 3 shows a comparison of different models with YOLO v3.

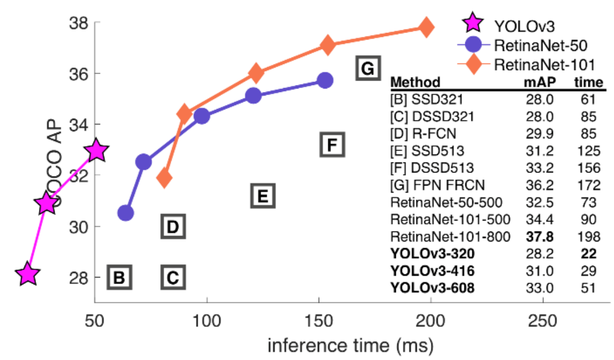

Figure 3. Comparison of recognition architectures

However, in tests with a reduced IOU used to reject detection, YOLO shows accuracy on par with other models, and at the same time with a noticeably slow speed. Benchmark 50 in COCO 50 is a metric that indicates how well forecasted bounding boxes correspond with an object's underlying actual bounding boxes. The number 50 denotes 0.5 IoU in this case. If the difference between the forecast and the truth scale is less than 0.5, the forecast is considered a wrong translation and is labelled as false positive. Figure 4 shows a comparison of models in the COCO mAP-50 benchmark [9].

In April of 2020, version 4 of YOLO was announced, but it was not created by the original YOLO creator. Joseph Redmond announced his departure from the world of computer vision in February 2020, citing his concerns about a possible negative impact on his work. [10].

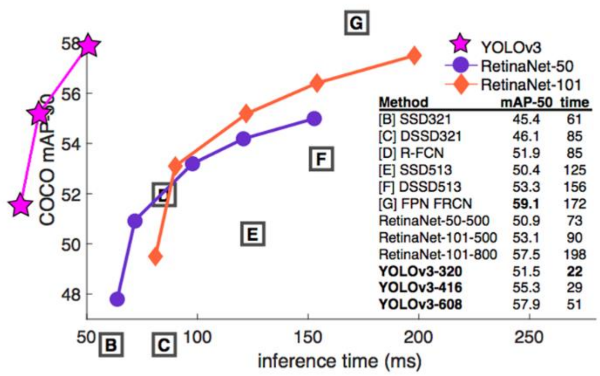

Figure 4. Comparison of models in the COCO mAP-50 benchmark

YOLOv4 is superior to the previous version with several advantages:

- It is a quick, reliable, and effective object detection model that enables anyone with a 1080 Ti or 2080 Ti GPU to train a highly accurate and fast object detector.
- During the detector's training, the effect of current object detection methods such as "Bag-of-Freebies" and "Bag-of-Specials" was checked.
- CBN (cross-iterative batch normalization), PAN (path aggregation network), and other modified modern approaches, have been optimized to be more effective and optimal for training on a single GPU. [11].

YOLO v4 consists of several main components such as:

- Input is where the image is entered.
- Base - related to the network, which takes an image as input and retrieves a feature map - it can be either VGG16, Resnet-50, Darknet-52 or ResNext50;
- Blocks of prediction and additional layers are subsets of the framework, where they serve to increase the recognizability and robustness of functions using the likes of FPN, PAN, RFB and others.
- The prediction block that processes the forecast [10].

The experimental results of YOLOv4 show that the AP value for the MS COCO dataset was 43.5% (65.7% AP50) and that the Tesla v100 was real-time at a speed and precision of approximately 65 FPS, which were the fastest and most precise detectors. YOLOv4 is twice the speed with comparable efficiency of EfficientDet. Moreover, AP and FPS rose respectively by 10 and 12 per cent in comparison with YOLOv3, [10]. Figure 5 compares the speed and accuracy of different architectures with the YOLO v4 model, on an NVIDIA video card with Volta architecture.

B. Model implementation

The experiment used the YOLO models of the third and fourth versions and their variations in the input data size. Windows 10 Pro was chosen as the operating system, as it has more reliable support for drivers and libraries required for video processing. Linux operating systems are notorious for not always accepting the latest driver and library updates in the kernel, which necessitates updating the package manager and installing additional kernel libraries.

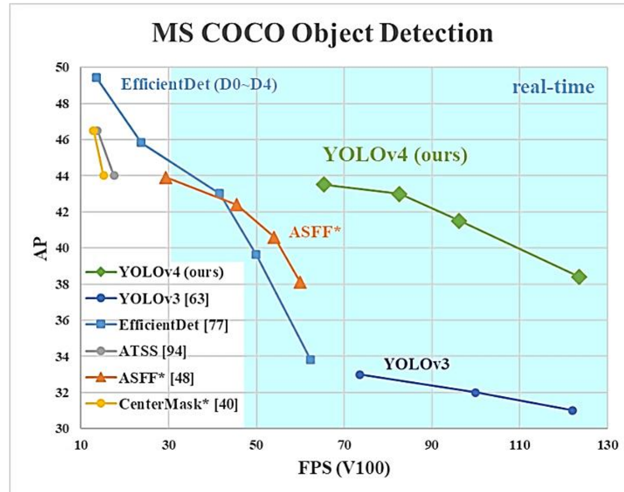

Figure 5. Comparison of different architectures with YOLOv4.

The stationary machine on which the recognition will be carried out has the characteristics indicated in Table 1.

Table 1: Units for magnetic properties

CPU	**AMD FX-8320**
Video card	NVIDIA GeForce GTX 1050
RAM	DDR3, 8GB
HDD	SSD WD BLUE, 250 GB

According to the benchmark results, the video card shows a performance result at the level of 2 teraflops. That is, this stationary machine belongs to the category of an average multimedia computer. NVIDIA video cards are required to fully operate the model with maximum performance since they have CUDA cores.

To implement the model, you need to install components such as:

- OpenCV (version 2.4 and higher) is a library of programming functions primarily intended for real-time computer vision. That is, it is a library used for image processing. It is mainly used to perform all image-related operations.

- CMake (version 3.9.1 and higher) is the cross-platform for handling the software development process using a compiler-independent approach, which is free and open-source.

- MS Visual Studio is an integrated development environment (IDE) for developing software programs and applications designed for Windows and many other platforms.

- Anaconda (with a Python version greater than or equal to 3.7) It is a free and open-source programming environment for the Python and R programming languages used in programming many areas including data science, machine learning applications, large-scale data processing, predictive analytics, etc. which has the objective of simplifying package management and deployment.

- CUDA (version 10.0 and higher) is an NVIDIA toolkit that offers a development platform for creating GPU-accelerated applications with high speed. Accelerated transformations are possible in a variety of fields, including linear algebra, image and video analysis, deep learning, and graph analytics, using GPU-accelerated CUDA libraries.
- cuDNN (version 7.0 and higher) is a deep neural network library that provides primitives for deep neural networks that are GPU-accelerated. cuDNN implements regular protocols such as forward and backward folding, pooling, normalization, and activation thresholds with high precision. Researchers and developers working on frameworks all over the world rely on cuDNN for high-performance GPU acceleration. This enables them to concentrate on neural network training and device development rather than on low-level GPU output tuning.

Next, there is a need to install the Darknet framework that serves as the basis for the YOLO models. This framework is native to the YOLO model and therefore easy to implement. After installing all the necessary components, we can start experimenting.

C. Development of a full-scale experiment

An action camera Digma DiCam-380 was attached to the windshield of a Nissan Maxima car. The main criterion for choosing a camera was the ability to shoot in 1080p and 60 frames per second.

The characteristics of this camera are shown in Table.2

TABLE 2: Units for Magnetic Properties

Maximum video resolution	**3840x2160**
The number of frames at Full HD resolution	60 fps
Viewing angle	160 degrees
Diaphragm	2.5

Further, the footage was passed through the YOLO neural network. For recognition with the Darknet framework, you need to download configuration files and pretrained files for specific YOLO models. These files must be placed in the cfg folder of the Darknet root folder.

To train the files, a large-scale object detection and segmentation dataset, COCO, created by the Google Brain project, was used. This dataset has many different classes, for example: Man, A bike, Car, Bus, Traffic lights, animals, and many other categories.

To start the object recognition with the video stream, it is need to enter in the following command line: darknet detector demo cfg / coco.data cfg / yolov.cfg yolov.weights <video file>

Depending on the version of the model, the configuration file changes, and when the type of models is selected by the size of the input data, the width and height parameters in the configuration files change.

IV. MATHEMATICAL MODEL

In this part, we will discuss a mathematical model that was used to detect objects on the street.

A. Data Preprocessing

In order for the accuracy to be high, some treatments must be performed to change the size of the object. The mathematical model will be based on eight objects that are abundant on various roads in Oman. Such objects are a bike, people, a car, traffic lights, a roundabout, a passenger, an animal, and a truck. We defined the objects as follows:

$$\hat{y} = [p_c \quad b_x \quad b_y \quad b_h \quad b_w \quad c_1 \quad c_2 \quad . . . \quad c_8]^T \tag{1}$$

Where:

- pc: The likelihood of an object in the target area
- bx, by: The coordinates of the target square
- bh: The height of the bounding box with the height of the shape
- bw: View the bounding box with the shape displayed.
- ci: Probability of class i.

The square coordinates need to be converted depending on the target variables that we identified above, so the coordinates in the dataset will be as follows: xmin, ymin, xmax, ymax as in Figure 6.

W: width of the original image.

H: height of the original image.

$$b_x = \frac{(x_{min}+x_{max})}{2*W}$$

$$b_y = \frac{(y_{min}+y_{max})}{2*W}$$

$$b_w = \frac{(x_{max}-x_{min})}{2*W}$$

$$b_y = \frac{(y_{max}+y_{mix})}{2*W}$$

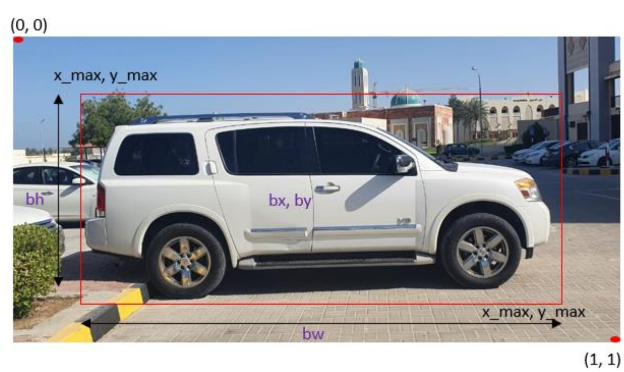

Figure 6. Bounding Box Coordinates in the target variable.

B. Intersection Over Union

The following equation illustrates the matric of Intersection over Union (IoU). It is used to measure the object detection algorithm through the interference between two surrounding boxes.

$$IoU = \frac{\boldsymbol{area\ of\ overlap}}{are\ of\ Union} \tag{6}$$

To obtain this matric, the following are required:

- Square measure on real ground.
- The square measure expected from the model.

The ratio of the intersection region to the union area filled by the ground truth and estimated bounding boxes is called the intersection over union ratio. Fig. 7 depicts the IoU calculation for different bounding box configurations.

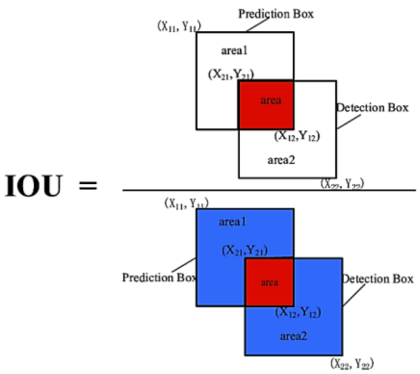

Figure 7. Intersection over Union (IoU) calculation diagram.

C. Defining the Model

In this work, we used a trained network and applied transportation instruction to it until the final prototype appeared. As we mentioned previously, the most recent real-time object recognition method is You only look once (YOLO). YOLO includes 78.6 percent map on 2007 VOC and 48.1 percent map on COCO test-dev. YOLO employs a single neural network to process the whole picture. The captured picture is then divided into regions and the perimeter squares and probabilities for each area are predicted. Planned odds are surrounded by weighted boxes.

V. ANALYSIS OF THE PERFORMED EXPERIMENTS

A. Results of the experiment with the YOLOv3 model

The experiment was carried out on the YOLOv3 model with three variations of the input image resolution: 320x320, 416x416 and 608x608. First, a model with an input image resolution of 320x320 was used. Figure 8 shows the object recognition process.

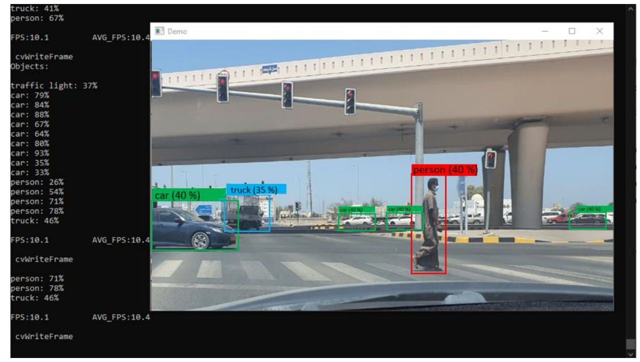

Fig. 8. Object recognition process.

When running the test, the average frame rate was 10.4 frames per second. The model showed satisfactory accuracy as not all small objects such as traffic lights and people were recognized due to range. However, recognition worked well at close range.

The second test was carried out with an input image resolution of 416x416. Figure 9 shows the result of the second test.

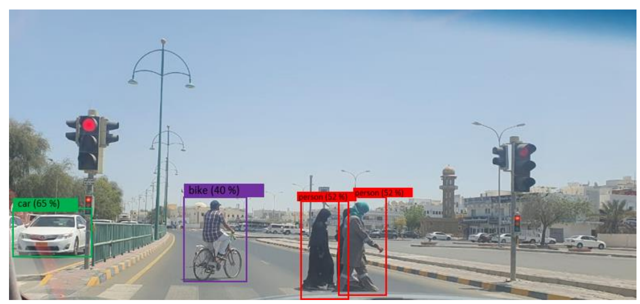

Figure 9. Second test result of YOLOv3.

In this case, the model showed more confident recognition. Passers-by on the sidewalk are better recognized, as are traffic lights. The average frame rate was 10 frames per second. The recognition accuracy of cars, trucks and motorcycles ranges from 50% to 100% depending on the distance.

The final test using YOLOv3 was carried out with an input image resolution of 608x608. This variation is the most demanding on the resources of a stationary machine.

This model showed good results and detected most of the objects with high accuracy. The accuracy of detecting

people and all objects in general have increased compared to other variations. The average frame rate was 7 frames per second.

B. Results of the experiment with the YOLOv4 model

According to the developers, YOLOv4 is the most accurate and fastest "real-time" neural network on the COCO dataset.

As with the third version model, three tests were carried out. With input image resolution 320x320, 416x416 and 512x512.

The first test was carried out on a model with an input image resolution of 320x320. Figure 10 demonstrates the experiment result of the initial test.

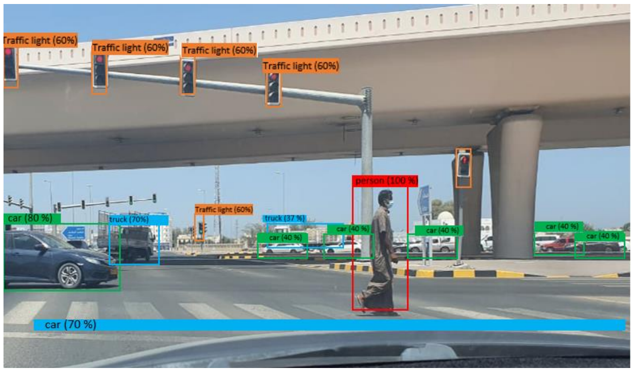

Figure 10. First test result of YOLOv4

The new version of the model showed a very good result. The accuracy was on the level of the most demanding and accurate variation of the YOLOv3 model. Vehicle recognition accuracy ranges from 52% to 100%. Also, the model became more sensitive to small objects such as traffic lights and people. The average frame rate was 8 frames per second.

The second test was carried out with an input image resolution of 416x416. Figure 11 shows the result of the second test.

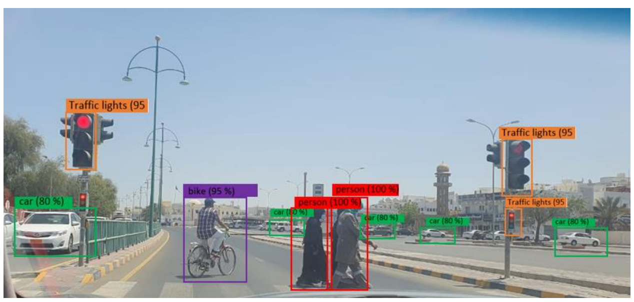

Figure 11. Second test result of YOLOv4

This model shows the best recognition so far. Some objects were not noticed by other variations, while a model with such parameters recognizes them with 50% accuracy. Traffic lights are detected with an accuracy of 30% to 60% at an average distance and with a poor camera. The average frame rate, as in the first test, was 8 frames per second.

So, the last test using YOLOv4 was done with 512x512 input image resolution. This variation is the most demanding and accurate. The result of the third test.

C. *Compare results between YOLOv3 and* YOLOv4

Compared with other variations, this model shows better recognition, since almost all objects were detected. The model recognized all traffic lights and people. In line of sight, vehicles were detected with an accuracy of 60% to 100%. Traffic lights were recognized with an accuracy of 30% to 85%, and people with an accuracy of 30% to 80%. Not the highest accuracy may be due to lighting and poor picture quality. However, the model showed excellent recognition and quick response to objects. The average frame rate, as in the two previous tests, was 8 frames per second.

Table 3 shows the results of the experiments.

TABLE 3: Units for Magnetic Properties

Objects	YOLO v3	YOLO v4
Car	60 %	90 %
Traffic light	0 %	60 %
Bike	66 %	95 %
person	61 %	100

VI. CONCLUSION

Within the framework of this work, classical recognition methods and methods based on the use of deep neural networks were described. The advantages of methods of deep neural networks were identified, and their types and process of work were described. For a full-scale object recognition experiment, the YOLO model was chosen, which is the fastest recognition system to date. Comparisons of different architectures with the YOLO model in benchmarks were given. This model has been implemented and tested with various parameters and has shown excellent results in speed and accuracy.

Thus, it can be concluded that the field of self-driving cars is booming, and companies are offering more and more advanced car architectures. With the advent of 5G technology in general use, there will be a huge leap forward in this area, as self-driving cars are highly dependent on data transmission technologies. Object recognition architectures are also evolving, recognition speed and accuracy are increasing, and very soon self-driving cars will enter public roads.

REFERENCES

[1] Sorin Grigorescu, Bogdan Trasnea, Tiberiu Cocias, Gigel Macesanu. “A Survey of Deep Learning Techniques for Autonomous Driving”. arXiv:1910.07738v2. 2020. pp28.

[2] Ahmad Alhilal, Tristan Braud, Pan Hui. “Distributed Vehicular Computing at the Dawn of 5G: a Survey”. arXiv:2001.07077.2020. 2020. pp34.

[3] Demir, Kadir Alpaslan, Halil Cicibas, and Naiz Arica. "Unmanned Aerial Vehicle Domain: Areas of Research." Defence science journal, 2015, pp65.4.

[4] Ignatova, Lidiia, et al. "Analysis of the Internet of Things devices integration in 5G networks." 2017 Systems of Signal Synchronization, Generating and Processing in Telecommunications (SINKHROINFO). IEEE, 2017.

[5] YOLOv4 — Superior, Faster & More Accurate Object Detection. [Online] / 28.08.2020 // URL: https://medium.com/@riteshkanjee/yolov4-superior-faster-more-accurate-object-detection-7e8194bf1872

[6] Joseph Redmon, Santosh Divvala, Ross Girshick, Ali Farhadi. “You Only Look Once: Unified, Real-Time Object Detection”. arXiv:1506.02640. 2016. pp 10.

[7] A Gentle Introduction to YOLO v4 for Object detection in Ubuntu 20.04. [online] / 01.08.2020 // URL: https://robocademy.com/2020/05/01/a-gentle-introduction-to-yolo-v4-for-object-detection-in-ubuntu-20-04/

[8] YOLO: Real-Time Object Detection. [online] / 23.07.2018 // URL: https://pjreddie.com/darknet/yolo/

[9] Joseph Redmon, Ali Farhadi. “YOLOv3: An Incremental Improvement”. arXiv:1804.02767. 2018, pp 6.

[10] Alexey Bochkovskiy, Chien-Yao Wang, Hong-Yuan Mark Liao. “YOLOv4: Optimal Speed and Accuracy of Object Detection”. arXiv:2004.10934. 2020, pp17.

[11] Recommendation ITU-R M.2083: IMT Vision, “Framework and overall objectives of the future development of IMT for 2020 and beyond,” Sep. 2015.

[12] Dai, Y., Liu, W., Li, H., & Liu, L. (2020). Efficient foreign object detection between PSDs and metro doors via deep neural networks. IEEE Access, 8, 46723-46734.

[13] Bochkovskiy, A., Wang, C. Y., & Liao, H. Y. M. . Yolov4: Optimal speed and accuracy of object detection. arXiv preprint arXiv:2004.10934,2020.

[14] Khakimov, A., Alekseeva, D., Muthanna, A., & Al-Bahri, M. Traffic Offloading Algorithm for VANET Network Based on UAV. In 2020 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering (EIConRus), 2020, pp. 28-32. IEEE.

[15] Hasoon, F.N., Yousif, J.H., Hasson, N.N. and Ramli, A.R. Image Enhancement Using Nonlinear Filtering Based Neural Network. Journal of Computing, 2011, 3(5), pp.171-176.

[16] AL-Balushi AI, Yousif J, Al-Shezawi M. Car accident notification based on Mobile cloud computing. International Journal of Computation and Applied Sciences IJOCAAS, Volume2. 2017 Apr(2).

[17] Yousif JH, AlRababaa MS. Neural Technique for Predicting Traffic Accidents in Jordan. Journal of American Science. 2013 Nov;9(11), pp528-535.

[18] Ateya, A., Al-Bahri, M., Muthanna, A., & Koucheryavy, A. End-to-end system structure for latency sensitive applications of 5G. Электросвязь, 2018, 6, pp 56-61.

Wasin AlKishri, PhD candidate and researcher in computer vision and image processing. lecturer at Sohar University with 2 years of teaching experience in different field of computer science. She received the B.S. in Networking and M.S. in Soft computing from the Sohar University in 2019.

Mahmood Al-Bahri is assistant professor at Sohar University with three years of experience teaching undergraduate and graduate courses in Info- communication and computer science. Extensive knowledge of Forth Industrial Revolution technologies. Author of the Internet of Things book in Arabic Language. Actively participate as speaker in international and local conferences and professional research. Believe in students' abilities to learn and inherent thirst for knowledge with the right environment. Actively work to connect students to their material to transform lives.