# Real-Time Computational Efficiency Vehicle Detection and Counting Utilizing the Background Subtraction Technique and Non-Maximum Suppression Techniques

Jameelah Kadhim Mezaal
Dept. of Student Activities, University of Basrah, Basrah, Iraq
E-mail: Jkm.alkafaje@gmail.com

**Keywords:** Computer vision, Image processing, Background subtraction, Mixture of Gaussians (MOG), Non-Maximum Suppression (NMS**)**, Vehicle detection.

**Received:** October 24, 2024

*By combining cloud computing, computer vision, and Internet of Things (IoT), it would be able to make the most of both sides. Because the IoT is mostly composed of connected, contained gadgets, it can store and process data gathered through the application of computer vision algorithms. It is able to achieve this by making use of the almost infinite resources provided by cloud organizations, including processing and storage services. The development and execution of a computer vision-based system are examined in this paper. that counts and identifies automobiles using machine learning (ML). The system consists of multiple stages, including initialization, background subtraction, object detection, bounding rectangles, vehicles counting and evaluation criteria. The proposed methodology first separates moving objects from the background and then employs a statistical technique called Mixture of Gaussians (MOG) for background subtraction to identify the automobiles in the image and Non-Maximum Suppression (NMS) to filter out overlapping bounding boxes to enhance the detection operation. The experiment's outcomes show how effectively cars can be found and counted. The result of the experiments using accuracy, precision, f1-score and recall are about 90% for the different types of video and from many corners.*

*Povzetek: Razvit je učinkovit sistem za zaznavanje in štetje vozil v realnem času, ki s tehniko odštevanja ozadja in metodo Non-Maximum Suppression doseže visoko točnost pri računalniškem vidu.*

## 1 Introduction

In computer science, computer vision is the area that focuses on giving computers the ability to analyze, interpret, and comprehend digital images and videos at a high level [1]. Because deep learning and artificial intelligence are developing so quickly, it has grown significantly in the last few years and can be used to automate a wide range of operations. Computer vision technology can be employed to gather specific, location-specific information about ecosystems and their surroundings by using digital images to interpret and recognize the global environment [2]. Over the past ten years, there has been a tremendous advancement in the utilization of machine learning models to multimodal data processing [3]. ML models find useful applications in many domains, including signal processing [7], remote sensing for earth observation and analysis [8], the classification of common human actions [9, 10], biomedical imaging analysis [4-6], and many more.

Systems related to computer vision and image sensing depend on ML models. Advances in machine learning have made it possible to analyze sensor and image data more effectively, which has led to a variety of research projects tackling practical issues in industries like medical services, farming, protection, surveillance, planet perception, and autonomous navigation. Giving computers the ability to see, comprehend, and act on data based on past and present outcomes is the aim of machine learning and computer vision. Both machine learning and computer vision are rapidly developing fields. Brain-human interfaces, the Industrial IoT (IIoT), and the Internet of Things all heavily rely on computer vision.

ML and computer vision are employed to identify and track intricate human behaviors in multimedia streams. Unsupervised learning, semi-supervised learning, and supervised learning are three prominent methods used for prediction and analysis. These techniques utilize ML algorithms, including support vector machines (SVM) and K-Nearest Neighbors (KNN) [11]. Smart sensors and the Internet of Things have grown significantly in a number of industrial applications and research domains. The field of robotics has grown significantly, and its applications can be seen in industries like aerospace, medical care, industry, and learning. It has been found that robots are dependable, effective, and efficient in many of these activities. Robots are limited by a multitude of factors, such as slow internal code execution, tiny storage sizes, latency within the network changeable quality, delay, and insufficient intelligence.

Improved hardware computing and cloud computing could alleviate these limitations. Cloud computing infrastructure has the potential to be highly efficient in the processing and storage of large amounts of data. Integrating IIoT applications across a range of industries requires regulations and smart techniques, such as low-power wireless networking technology and quick sensors for large-scale data analytics. Data from sensors is stored in the cloud for evaluation and forecasting during industrial manufacturing and operation. ML algorithms are used by the IIoT to find and choose pertinent features. The IIoT produces datasets that are utilized for training and evaluating ML models, with the aim of enabling intelligent decision-making. ML has become a crucial component of the (IIoT) to enhance the productivity and effectiveness of industrial outputs [12, 13].

The main objective of this study is to create a system capable of automatically detecting and quantifying cars in real-time video footage. Vehicle detection, also known as computer vision object recognition, refers to the scientific procedures and techniques used by machines to perceive visual information, as opposed to human eyes. A vehicle detection system's primary function is to accurately determine the position of one or more vehicles in input photos.

# 2 Related work

Ref. [14]This paper reviews the research progress in computer vision, focusing on the theoretical basis of Generative Adversarial Networks (GANs). GANs have been rapidly developing in image processing and are increasingly important in fields like medicine, art, and security. Recent research progress includes data enhancement, high quality sample generation, domain transfer, image restoration, and AI security. However, to apply GAN models more rationally, it is necessary to understand their advantages and disadvantages. GANs are trained based on the idea of minimax game, but they are difficult to achieve Nash equilibrium states in actual use and are prone to gradient disappearance and model collapse. To ensure stability, convergence, and efficiency, more suitable objective functions or improved network structures are needed. GANs have broader application prospects when combined with other ML algorithms, such as forecasting problems, target detection, AI medicine, large data modeling, domain transfer, and AI security attack and defense. However, security in high-risk areas like autopilot, smart homes, and AI investments needs to be considered. GANs need to achieve a breakthrough in theory, solve some drawbacks of GAN models, and establish reasonable and accurate generative models through scientific and effective evaluation methods, taking into account the security and robustness when combined with different fields.

Ref. [15] This research examines various ML techniques that utilize diverse data sources for the purpose of detecting plant diseases. The text outlines the main methods used for collecting data, which encompass the IoT, ground imaging, unmanned aerial vehicle (UAV) imaging, and satellite imaging. It also encompasses both conventional and deep learning approaches. Moreover, this study examines the significance of data integration in the context of disease detection for ongoing research. The text highlights the benefits of utilizing intelligent data fusion techniques, which are obtained from diverse data sources, to improve the accuracy of predicting the health condition of plants. Furthermore, it delineates the primary obstacles that this field is presently grappling with. The study finishes by addressing many contemporary concerns and research directions. For this objective, ML methods such as Support Vector Machines (SVM), K-Nearest Neighbors (KNN), and the decision tree-based classifier C5.0 were utilized. Presently, image sensor technologies are burdened by various constraints, including the challenge of implementing the hyperspectral data acquisition technique under field conditions. Moreover, several parameters such as technical characteristics (resolution, brightness, etc.), settings for sample preparation (laboratory or field), and features of the sample (size, texture, humidity, etc.) can also have an impact on the spectrum reflectance. To analyze and extract disease patterns from high-resolution satellite photos, UAVs face challenges related to the climate and logistics. These challenges include dealing with strong winds and rain, managing battery power, and requiring a qualified person to launch and oversee flights. In addition, the presence of clouds and their shadows poses a substantial obstacle.

Ref [16] The objective of this study was to demonstrate the effective utilization of specialized technologies to tackle the most pressing issues in the Fleet Industry. The suggested fleet management system has been constructed utilizing a wide range of technologies, such as the IoT, Computer Vision, ML, Cloud Computing, ML, Deep Learning, and Embedded systems. The application was deployed on the IBM Watson IoT and Heroku platforms to collect data from the vehicle dashboard gadget. The driver's face identification and driving pattern monitoring, as well as the consumption of fuel prediction modules, used OpenCV (Computer Vision), SVN (ML), and CNN (Deep Learning) algorithms. An advanced embedded device, consisting of a gateway, sensor boxes, cameras, and an OBD-II device, was tasked with continuously monitoring the internal environment of trucks. It also detected unauthorized entry into the container, deviations from the designated route, and provided vehicle telematics. Empirical data was gathered to train and evaluate the facial detection and authentication models employed in the system. The simulation results shown that the suggested method can effectively handle and manipulate large volumes of data generated by several vehicles at frequent intervals, fulfilling the practical demand. This is achieved by utilizing IoT technology, a NoSQL CloudantDB database, and Cloud Computing. In addition, the article presented the system's architecture and the results of the experiments carried out on several modules.

Ref [17] This review paper aims to examine the techniques employed in the analysis of fruit images as well as their potential applications. It provides an overview of various methods, such as preprocessing, clustering, feature extraction, and classification, utilizing CNN, YOLO4, and IoT. These methods are utilized to assess the quality of fruits and vegetables based on their size, shape, color, and textures. This study explores a method that is both secure and cost-effective for assessing the freshness of food. It achieves this by examining the produce's size, shape, and color. Given their high susceptibility to damage, fruits should exclusively undergo non-destructive testing methods. The color of the fruit is its most prominent physical characteristic, second only to its size, as it contributes to its visual attractiveness. The results were obtained by evaluating different pre-trained models for fruit photos. The accuracy of CNN is 94.8%, whereas ANN has an accuracy of 91.4% and SVM has an accuracy of 86%. As expected, CNN has got a superior classification result. This survey focuses on the CNNs utilized in the agriculture and agricultural sectors. CNN is commonly utilized for the objective assessment of food's physical characteristics, offering precise, unbiased, and advantageous categorization. The proposed methodology effectively differentiated between unripe and mature product, as well as decaying fruit. The need to assess the freshness of products was examined during the activity. An investigation was carried out to assess a range of ML models capable of differentiating between fresh and decaying fruits. The Convolutional neural network is now the most pragmatic and efficient approach being employed. Empirical evidence has demonstrated that it yields the best accurate outcomes for Image Classification.

Ref. [18] This article proposes a highly effective model for automating the production lines of central processing unit systems in an industry. The model analyzes photos of the production lines and detects any anomalies in their assembly. The system administrator is notified of this information via a cyber-physical cloud system network. Accurate categorization is accomplished by utilizing a system based on ML. Our methods exhibit a 92% precision rate, and this model not only focuses on irregularities but also aids in establishing the perspectives from which production photographs are taken. This paper use multi-instance (MI) learning16 to detect problems. MI learning has previously been utilized for applications like as image retrieval, object recognition, target tracking, and image categorization. This is achieved by incorporating an industrial image processing system into an industry that utilizes CPCS. The image processing system scrutinizes product photos taken during production line activities to detect any faults. Afterwards, the relevant authority is notified of the result. The main challenge in our system lies in extracting the features of the instances. If the required feature extraction mechanisms are not used, the entire visual inspection system is in danger. It is imperative to consider the system's security for next developments. The security concerns are worsened by the incorporation of the cloud in communication. Hence, it is crucial to enforce multiple preventive measures to ensure the security of the entire CPCS system from unauthorized third parties.

Ref. [19] This approach is especially beneficial in circumstances where it is important to ascertain the magnitude of traffic or tally the quantity of distinct vehicle categories that traverse a designated region, such as a street or highway. The paper's data source consists of images of autos that were acquired from various places, ensuring that the camera-to-vehicle distances remained consistent. The image's background is removed, and its morphological traits are extracted and given to four classifiers to aid in the classification process. This article focuses on seven specific kinds of Iranian automobiles. Four machine classifiers, namely support vector, k-nearest neighbor, perceptron neural network, and Bayesian decision theory, were applied to three sets of images: raw, filtered, and noisy. The photos underwent low-pass Gaussian filtering at different frequencies. Afterwards, salt-and-pepper noise and Gaussian noise were added to create noise. The findings demonstrate that our suggested approach, employing the SVM algorithm, has produced favorable results with a precision of 97.1. The algorithm used in this study also yielded positive outcomes when applied to filtered and noisy images. Nevertheless, one could contend that this article has a limited scope as it focuses on the specific autos covered, which represent the bulk of vehicles in Iran. Certainly, it is also possible to attain favorable results by altering the automobiles.

# 3 Computer vision

Attention mechanisms refer to techniques that direct focus towards the significant areas of a picture while ignoring unnecessary portions. The human visual system employs such mechanisms [20, 21] to aid in the efficient and effective analysis and comprehension of complex images. As a result, researchers have been motivated to include attention mechanisms into computer vision systems in order to enhance their performance. Within a vision system, an attention mechanism can be conceptualized as a dynamic selection process that is achieved by assigning variable weights to features based on the significance of the input. [22]

Computer vision emerged in the early 1970s as the visual perception aspect of a larger effort to replicate human intelligence and equip robots with intelligent capabilities. Some early pioneers of artificial intelligence and robotics, including those at MIT, Stanford, and CMU, anticipated that tackling the "visual input" problem would be a relatively simple task compared to more complex challenges like higher-level thinking and planning. In 1966, Marvin Minsky, a prominent figure at MIT, instructed his undergraduate student Gerald Jay “spend the summer linking a camera to a computer and getting the computer to describe what it saw”[23] . It has been determined that the problem is somewhat more challenging than initially thought. What sets computer vision apart from the field of digital image processing is its aim to extract the three-dimensional structure of the environment from images and utilize it as a foundation for comprehensive scene comprehension.

Initial efforts in scene comprehension involved the extraction of edges, followed by the deduction of the three-dimensional configuration of an object or a "blocks world" based on the topological arrangement of the two-dimensional lines. Multiple line labeling algorithms were created during that period. Edge detection was a subject of ongoing research [24].

Attention mechanisms have been beneficial in various visual tasks, such as image classification [25], 3D vision [26, 27], face recognition [28], action recognition [29], object detection, medical image processing, image generation [21, 22], pose estimation [23], person reidentification, few-shot learning, super resolution, multi-modal task and semantic segmentation [22]. The techniques of computer vision are Image Classification The process of categorizing pixels and vector groups in an image based on their characteristics. The act of implementing precise regulations is referred to as image categorization. It is one of the most well-known approaches. Nevertheless, there are other challenges that need to be addressed during its implementation. Suppose we possess a collection of photographs that fall under a specific category, and we have assembled a separate set of test images to evaluate the precision of our predictions. The challenges in this context include the need to alter viewpoints, address deformations, and adjust lighting conditions. Object detection Object detection is a technique employed to precisely identify and determine the exact location of particular items inside an image or video. Object detection enables us to precisely identify the positions or movements of items in a given scene and visually represent them using bounding boxes. The primary distinction between object detection and image recognition lies in the fact that object detection assigns a defined and labeled bounding box to each object in an image or video, specifically identifying the object in question. The model provides predictions for both the label and location of the object. And Object tracking refers to the process of locating and following a certain object in a video or image sequence. Object tracking refers to the process of monitoring and following the movement of an object inside one or several scenes. Initially, object detection is employed, followed by the utilization of deep learning algorithms to track the movement of the detected objects. Tracked items are accompanied by an indicator adjacent to them. The location of the object may be observed in Figure 1 through the bounding box [30].

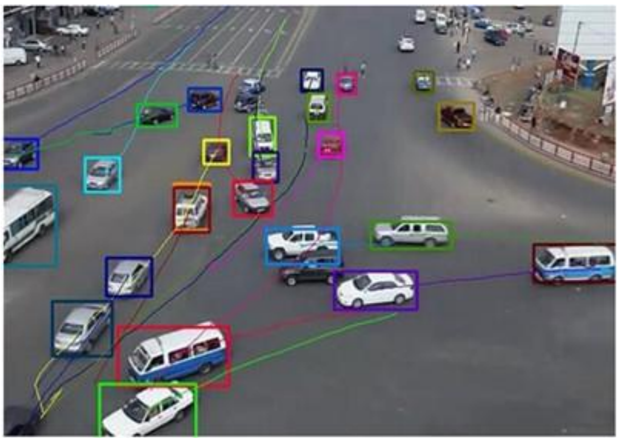


Figure 1: Bounding box in object tracking [30].

# 4 The proposed method

In general, the method used to count vehicles contains many steps such as Background subtraction, object detection, bounding rectangles, center calculation, vehicle counting and display and interaction. This steps can be show in the figure (2) below.

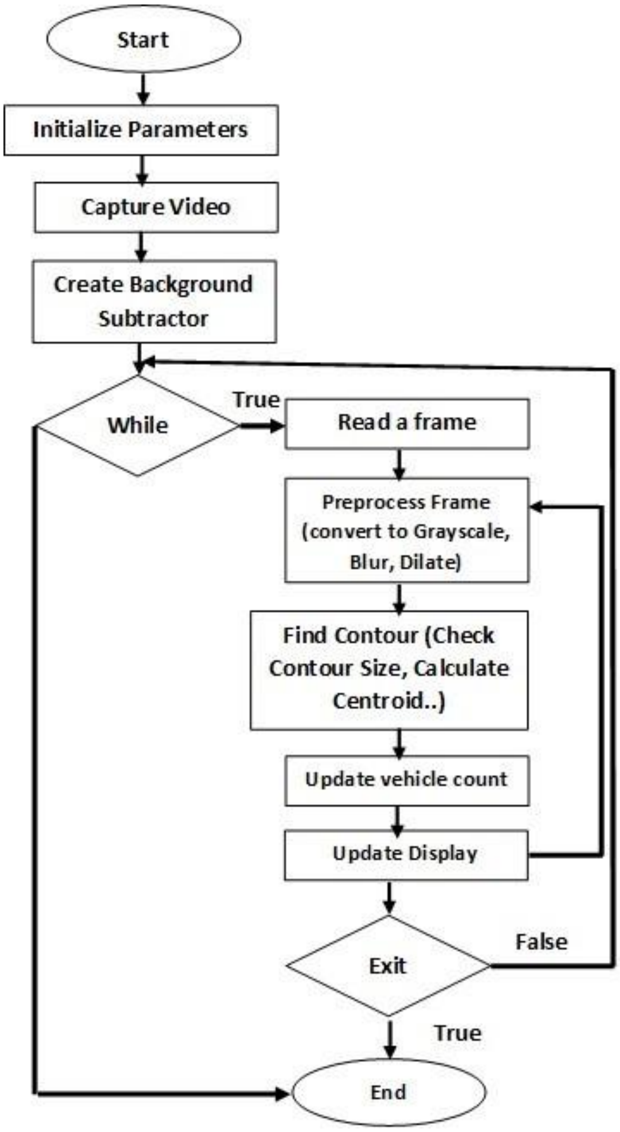



Figure 2: Flow chart of the vehicle counting system.

The following is a summary of the flow chart:

## 4.1 Background subtraction

- To detect moving objects, background removal is applied to the video stream. This technique finds regions of change across consecutive frames to help separate potential things, like cars.
- An important phase in the car detection and counting operation is backdrop subtraction. In order to differentiate moving objects from the background, the research applied the Mixture of Gaussians (MOG) technique.

For background elimination in video modification applications, MOG is an effective a statistical approach. Because it models each background pixel as a blend of Gaussians, it can manage small alterations in the background environment and capture oscillations in illumination. Every frame of the video feed is converted to grayscale to facilitate processing. The next step to help with better detection is to use a Gaussian blur to smooth out and minimize noise in the image. Subsequently, morphological methods are utilized on the obtained image to enhance the recognized motion zones and close any gaps. After that, morphological closure techniques will be employed to help break new boundaries and establish connections with other regions. The contours that contour detection creates from the detected moving objects indicate the boundaries of any possible moving vehicles in the area being detected.

## 4.2 Object detection

Geometric techniques such as dilation are used after background subtraction to improve the detected zones and reduce distortion. The contours of the items are then located, defining the boundaries of vehicles that are allowed.

((Background removal is a prerequisite for the object detection approach. In order to identify moving objects—in this case, vehicles—the background subtraction method first maintains a model of the background scene and then looks for pixels that differ noticeably from the background model. This background subtraction-based object detection technique involves the following steps:

1. Taking a frame of the video and turning it into grayscale.
2. To lessen noise, apply Gaussian blurring to the grayscale image.
3. Making an object for background subtraction and applying it to the grayscale image that is blurry. As a result, a binary image with the moving items highlighted is created.
4. Applying the binary image's morphological operations—dilation and closure—to smooth the areas that were identified.
5. Locating the cars that have been spotted by contours in the binary image that has been processed.
6. Filtering the identified outlines based on size to eliminate any that are too small to be categorized as vehicles.
7. Counting the number of cars that pass the predetermined counting line by drawing bounding boxes around the valid vehicle outlines.

## 4.3 Bounding rectangles

The method used to draw bounding rectangles around the detected vehicles takes a contour as input and returns the coordinates (x,y) of the top-left corner of the bounding rectangle, as well as the width (w) and height (h) of the rectangle. This method can be summarized as follows:

1. For each detected contour (object), a bounding rectangle is determined to help enclose the detected object and aid in further analysis.
2. Iterates through each contour detected in the processed video frame.
3. For each contour, get the coordinates (x, y) of the top-left corner and the width (w) and height (h) of the bounding rectangle.
4. Then checks if the width and height of the bounding rectangle are greater than or equal to the minimum width and height thresholds. If not, the contour is skipped.
5. If the contour passes the size validation, draws a green bounding rectangle around the vehicle. The function takes the following arguments:

• frame1: the original video frame to draw the rectangle on
• (x, y): the top-left corner of the rectangle
• (x+w, y+h): the bottom-right corner of the rectangle
• (0, 255, 0): the color of the rectangle in BGR format (green)
• 2: the thickness of the rectangle border

## 4.4 Center calculation

Each bounding rectangle's center is computed and used as a reference to track the position of the object. The procedure that follows is used to determine the center of the observed vehicles:

1. Gets each detected contour's bounding rectangle coordinates (x, y, w, h).
2. The technique makes use of the following formulas to determine the bounding rectangle's center:

• center_x = x + w // 2 (1)

The x-coordinate of the center can be found by taking the x-coordinate of the top-left corner (x) and adding half the bounding rectangle's width (w // 2).

• center_y = y + h // 2 (2)

By multiplying the axis of of the top-left corner (y) by 1/2 the height of the bounding area (h // 2), the center's y-coordinate is determined.

This simple center computation based on the bounding rectangle coordinates can be used to efficiently and frequently determine the center of identified objects in computer vision applications.

## 4.5 Vehicle counting

- Every object is processed and detected, and while it's at it, its center is examined to see if it crosses a pre-established counting line on the frame.
- When an object's center position intersects with this designated counting line within a specified offset range, the vehicle counts increments.
- Then updates the displayed frame to visually represent the detection process, highlighting the detected objects and counting the number of vehicles that cross the counting line.

Vehicle counting is done by tracking the identified vehicles over a number of frames and incrementing a counter each time a new, distinct vehicle is found.

This is an explanation of the car counting procedure:

1. **Vehicle detection:**

• The method first detects vehicles in each video frame using computer vision techniques, such as background subtraction and contour detection.
• For each detected vehicle, calculates the bounding rectangle coordinates and the center point, as discussed in the previous section.
• Apply Non-Maximum Suppression (NMS) to filter out overlapping bounding boxes to enhance the detection operation.

2. **Vehicle tracking:**

• To keep track of the detected vehicles, maintains a list or dictionary of the vehicle center coordinates from the previous frame.

• When a new vehicle is detected in the current frame, compares its center coordinates to the previously stored vehicle centers.
• If the new vehicle's center is sufficiently far from all the previously stored centers, it is considered a new, unique vehicle.

3. **Vehicle Counting:**

The vehicle counter is increased whenever a new, distinct vehicle is found (that is, when its center coordinates do not correspond to any of the previously stored vehicles).
• Tracks the total number of distinct cars found thus far by using a counter object or a variable.
4. The function keeps track of the prior vehicle centers in this case (previous_centers). The approach first determines whether the center coordinates of a newly detected vehicle differ from all previously stored centers sufficiently. If so, it is recognized as a completely new, unique car and has its vehicle score raised.
5. The exact threshold used to determine if a vehicle is distinctive can be altered based on the needed sensitivity, the characteristics of the video stream, and the cars that are recognized.
6. By using this vehicle monitoring and counting technique, the method may generate a reliable estimate of the total number of distinctive vehicles that have passed through the monitored area. Several applications for traffic analysis and monitoring can benefit from the calculation.

## 4.6 Display and interaction

- The live video feed displays the number of detected vehicles as of right now, which is updated on a regular basis.
-Through visible indicators on the shown video frames, viewers may observe how the detection and counting process works.
The approach uses contour analysis, background subtraction, and object position tracking in relation to a designated counting line to capture cars as they pass across the observed region in the video feed.
This approach determines the layout of the automobile counts systems that are part of the technique. It explains how to analyze video frames and how to identify cars depending on where they are in relation to a particular region of the screen, monitor movement, and recognize elements.

## 4.7 Evaluation criteria

Average accuracy (AP) is the main training parameter used in network performance evaluation, and the validation set is utilized to measure the operational efficiency of the trained network. The following are the equations for P (precision) and R (recall):

$$Accuracy = \frac{TP+TN}{TP+FN+TN+FP} \quad (3)$$

$$\text{precision} = \frac{TP}{TP+FP} \quad (4)$$

$$\text{recall} = \frac{TP}{TP+FN} \quad (5)$$

$$F - Measure = 2 * \frac{precision * recall}{precision + recall} \quad (6)$$

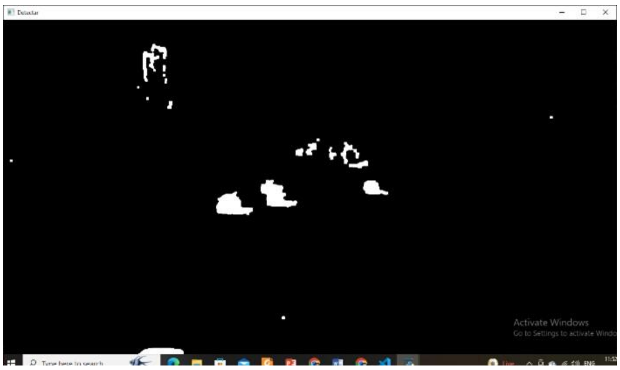
$$MSE = \frac{1}{n}\sum_{i=1}^{n}(yi - \overline{yi})^2 \quad (7)$$

The samples that the classifier correctly classifies as positive and are actually positive are known as True Positives (TP). The samples that the classifier correctly classifies as negative and are actually negative are known as True Negatives (TN). Samples that are truly negative but are mistakenly labeled as positive by the classifier are known as False Positives (FP). Samples that are truly positive but are mistakenly labeled as negative by the classifier are known as False Negatives (FN).

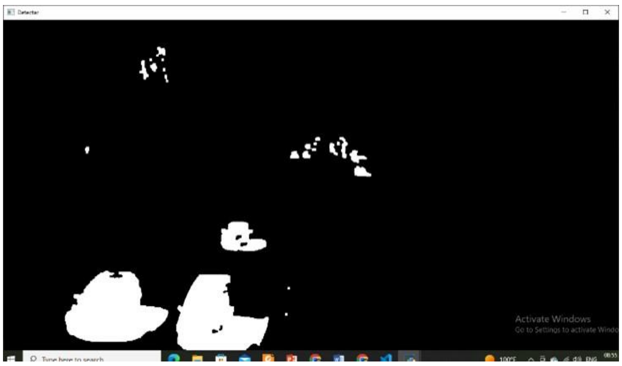
The evaluation criteria are shown in the table (1).

# 5 The experiments and results

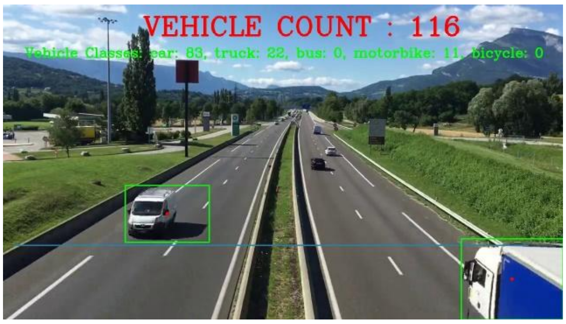
After read the data of video from the CCTV that exist on the road and passing the video in many steps as explained in the previous section the result is as shown in the figure (3):



a. Convert the video frame to B&W

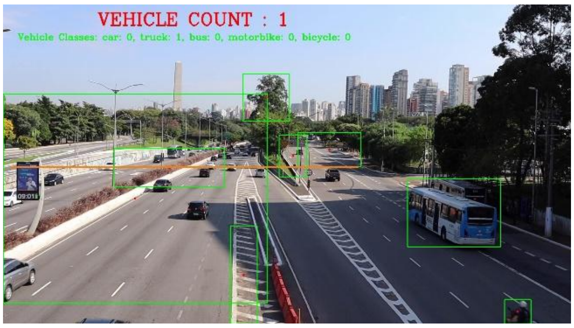


b. Convert the video frame to B&W

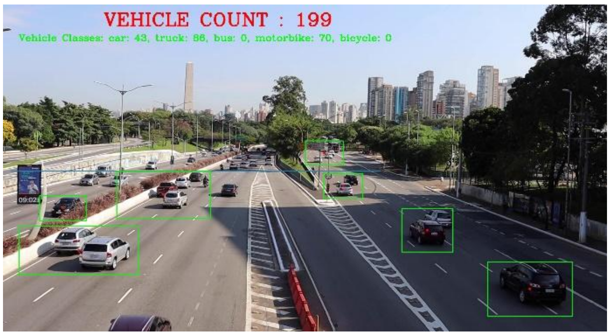



c. C: The first frame in the video-1.


d. The last frame in video-1.




e. E: The first frame in the video-2.




f. The last frame in the video-3.


g. The first frame in the video-3.

h. H: The last frame in video-3.

Table 1: The evaluation criteria for the method.

| Video No. | Accuracy | Precision | F1-Score | Recall |
|---|---|---|---|---|
| 1 | 90.504 | 85.433 | 82.225 | 80.323 |
| 2 | 90.234 | 84.673 | 81.434 | 80.102 |
| 3 | 88.342 | 82.453 | 80.322 | 80.232 |

Table 2: Shows the number of vehicles for each used class.

| Video No. | car | Truck | Bus | Motorbike | bicycle |
|---|---|---|---|---|---|
| 1 | 112 | 54 | 0 | 31 | 0 |
| 2 | 43 | 86 | 0 | 70 | 0 |
| 3 | 32 | 129 | 0 | 57 | 0 |

As shown in the images (c, d, e, f, g, h) that appear the first image and last image for each video, the evaluation criteria and the number of each class in each video is different as a result of the types of vehicles and types of roads and the corner of capturing of video. And the result of accuracy is very good comparing with these reasons.

## 6 Conclusion

In order to precisely identify and count cars in video streams, the vehicle recognition and counting system created in this study makes use of an efficient ML technique. A Gaussian Mixture Model (GMM), a common unsupervised learning method that can mimic the intricate and multimodal distributions found in actual traffic situations, serves as the foundation of the system. This approach's flexibility to adapt to a variety of operational settings is one of its main advantages. The algorithm's flexible structure makes it simple to connect with a wide range of video input sources, including sensors mounted on drones and CCTV cameras along the side of the road. This makes it a versatile option for applications pertaining to transport management, city development, and automated infrastructure for transportation. The algorithm's efficient computing weight and immediate time processing abilities also allow it to be implemented on low-resource devices at the edges, which broadens its possible use to include autonomous cars, traffic signal optimization, and other state-of-the-art smart city projects.

This research can be enhanced by using YOLO framework or adding deep learning algorithms to obtain very good results and enhance the performance and efficiency of the method.

# References


[1] H. Tian, T. Wang, Y. Liu, X. Qiao, and Y. Li, "Computer vision technology in agricultural automation—A review," *Information Processing in Agriculture,* vol. 7, no. 1, pp. 1-19, 2020.

[2] Y. Lu and S. Young, "A survey of public datasets for computer vision tasks in precision agriculture," *Computers and Electronics in Agriculture,* vol. 178, p. 105760, 2020.

[3] K. Sharma and M. Giannakos, "Multimodal data capabilities for learning: What can multimodal data tell us about learning?," *British Journal of Educational Technology,* vol. 51, no. 5, pp. 1450-1484, 2020.

[4] M. L. Giger, "Machine learning in medical imaging," *Journal of the American College of Radiology,* vol. 15, no. 3, pp. 512-520, 2018.

[5] K. Suzuki, "Overview of deep learning in medical imaging," *Radiological physics and technology,* vol. 10, no. 3, pp. 257-273, 2017.

[6] A. Barragán-Montero *et al.*, "Artificial intelligence and machine learning for medical imaging: A technology review," *Physica Medica,* vol. 83, pp. 242-256, 2021.

[7] M.-P. Hosseini, A. Hosseini, and K. Ahi, "A review on machine learning for EEG signal processing in bioengineering," *IEEE reviews in biomedical engineering,* vol. 14, pp. 204-218, 2020.

[8] L. Zhang and L. Zhang, "Artificial intelligence for remote sensing data analysis: A review of challenges and opportunities," *IEEE Geoscience and Remote Sensing Magazine,* vol. 10, no. 2, pp. 270-294, 2022.

[9] L. Chen, J. Hoey, C. D. Nugent, D. J. Cook, and Z. Yu, "Sensor-based activity recognition," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews),* vol. 42, no. 6, pp. 790-808, 2012.

[10] H.-B. Zhang *et al.*, "A comprehensive survey of vision-based human action recognition methods," *Sensors,* vol. 19, no. 5, p. 1005, 2019.

[11] A. I. Khan and S. Al-Habsi, "Machine learning in computer vision," *Procedia Computer Science,* vol. 167, pp. 1444-1451, 2020.

[12] N. F. Kamal, M. Abdelfattah, M. Ferjani, A. Amira, and N. Nawaz, "An IoT Reconfigurable SoC Platform for Computer Vision Applications," in *2019 International Symposium on Systems Engineering (ISSE)*, 2019: IEEE, pp. 1-7.

[13] A. I. Khan and A. Al-Badi, "Open source machine learning frameworks for industrial internet of things," *Procedia Computer Science,* vol. 170, pp. 571-577, 2020.

[14] L. Jin, F. Tan, and S. Jiang, "Generative adversarial network technologies and applications in computer vision," *Computational intelligence and neuroscience,* vol. 2020, no. 1, p. 1459107, 2020.

[15] M. Ouhami, A. Hafiane, Y. Es-Saady, M. El Hajji, and R. Canals, "Computer vision, IoT and data fusion for crop disease detection using machine learning: A survey and ongoing research," *Remote Sensing,* vol. 13, no. 13, p. 2486, 2021.

[16] P. Singh, M. S. Suryawanshi, and D. Tak, "Smart fleet management system using IoT, computer vision, cloud computing and machine learning technologies," in *2019 IEEE 5th International Conference for Convergence in Technology (I2CT)*, 2019: IEEE, pp. 1-8.

[17] R. Karpe, P. Patil, P. Shahane, H. Patki, S. Vispute, and R. Kannan, "A Review of Fruits Image Analysis Using Computer Vision and Deep Learning Techniques," in *International Conference on Soft Computing for Security Applications*, 2023: Springer, pp. 707-723.

[18] N. Rahmatov, A. Paul, F. Saeed, W.-H. Hong, H. Seo, and J. Kim, "Machine learning–based automated image processing for quality management in industrial Internet of Things," *International Journal of Distributed Sensor Networks,* vol. 15, no. 10, p. 1550147719883551, 2019.

[19] M. U. Arif, M. U. Farooq, R. H. Raza, Z. U. A. Lodhi, and M. A. R. Hashmi, "A Comprehensive Review of Vehicle Detection Techniques Under Varying Moving Cast Shadow Conditions Using Computer Vision and Deep Learning," *IEEE Access,* vol. 10, pp. 104863-104886, 2022.

[20] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Transactions on pattern analysis and machine intelligence,* vol. 20, no. 11, pp. 1254-1259, 1998.

[21] M. Hayhoe and D. Ballard, "Eye movements in natural behavior," *Trends in cognitive sciences,* vol. 9, no. 4, pp. 188-194, 2005.

[22] M.-H. Guo *et al.*, "Attention mechanisms in computer vision: A survey," *Computational visual media,* vol. 8, no. 3, pp. 331-368, 2022.

[23] M. A. Boden, *Mind as machine: A history of cognitive science*. Oxford University Press, 2008.

[24] R. Szeliski, *Computer vision: algorithms and applications*. Springer Nature, 2022.

[25] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "Cbam: Convolutional block attention module," in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 3-19.

[26] S. Xie, S. Liu, Z. Chen, and Z. Tu, "Attentional shapecontextnet for point cloud recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 4606-4615.

[27] M.-H. Guo, J.-X. Cai, Z.-N. Liu, T.-J. Mu, R. R. Martin, and S.-M. Hu, "Pct: Point cloud transformer," *Computational Visual Media,* vol. 7, pp. 187-199, 2021.

[28] J. Yang *et al.*, "Neural aggregation network for video face recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 4362-4371.

[29] X. Wang, R. Girshick, A. Gupta, and K. He, "Non-local neural networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 7794-7803.

[30] K. G. Nalbant and Ş. Uyanık, "Computer vision in the metaverse," *Journal of Metaverse,* vol. 1, no. 1, pp. 9-12, 2021.