Multi-Camera Vehicle Detection for Parking Management

Ch. Chandra Sekhar[1], Naresh Tangudu[2],Vineesh Pydisetti[3], Pramila Payli[4] ,Amitha Narindi[5], Chinni Krishna Enni[6]
[1, 2, 3, 4, 5, 6] Dept. of IT, Aditya Institute of technology and Management, Tekkali, India.

dlaxmics@gmail.com[1]*,itsajs@gmail.com[2], Vineeshpydisetti@gmail.com[3], paylipramila@gmail.com[4], amithanarindi315@gmail.com[5], chinnikrishnaenni@gmail.com[6]

***Abstract*- Because of increasing urbanization and an increasing number of automobiles on the road, efficient parking space management has become a key concern in modern urban areas. This research proposes an improved approach for an automatic parking system with multi-camera vehicle detection to alleviate the problem. The system's plan of action, which considers many realistic lighting and weather scenarios. The recommended solution comprises a variety of computer vision and Deep Learning models, including R-CNN, YOLO, and AlexNet. The key approach here is the use of pre-trained models to expedite the training process and optimize performance, allowing for automatic detection and real-time monitoring of available parking spaces inside a designated parking lot. This assures accurate vehicle recognition and data fusion techniques will be used to combine vehicle identification results from various cameras. This method to identify the objects on a specific training dataset, then modifies hyperparameters to avoid overfitting using the validation set. This proposed system assessment yields remarkable results in important performance areas such as real-time monitoring capabilities, occlusion elimination, and higher model quality.**

***Keywords:* Spaces detection, Deep Learning, CNN, Mask R-CNN, Object detection, Computer Vision, Automatic spot mapping**

I. INTRODUCTION

With the speed of modern urban lives, parking has become a huge challenge. Because there aren't enough parking spaces and more cars on the road, traditional parking management techniques have proven ineffective. Many annoyances have resulted from this, such as increased pollution, traffic jams, and delays. The introduction of intelligent parking systems in recent times has provided a flicker of cheerfulness. These devices simplify parking procedures by utilizing advanced methods like computer vision and machine learning.

In addition to detecting cars, they also recognize open parking spaces and direct drivers to them in real time. These systems do, however, have several drawbacks, chiefly because they rely on single-camera arrangements that may overlook several cars at once. Most parking lots have surveillance systems easily available, therefore in many areas the solution is just to analyse the data composed by cameras correctly or add extra cameras to provide full surveillance for the network to work well. Current systems primarily employ deep learning or image segmentation methods rather than spot overlays.. On the other hand, current advances in object detection algorithms open the possibility of leveraging these detections to improve automated parking management system performance. This paper presents a new method for addressing this problem. It talks about parking management and detection using multiple cameras. This method partially mitigated the negative aspects of. The following elements are essential to the multi-camera system's efficient operation. Many sensors are needed to cover an extensive area for parking. The rooms for the plan so that it would be expensive, In this research strictly focus on overview of the potential work of the thesis, This Discussion follows as:

A. Potential of multi-camera parking system

Substantial parking management could be entirely altered by this technology. Conventional techniques, such sensor use or human monitoring, tend to be costly inefficient, and inaccurate. On the other hand, multi-camera vehicles recognition can get around those limitations by offering a clear, affordable, and real-time method for tracking and regulating parking availability. The following are some particular benefits of multi-camera for parking management:

- Many variables, including camera quality, lighting, and occlusions, can influence how effective multi-camera vehicle detection is.
- Ensure that that multi-camera systems are being used in an approach that upholds individual privacy is crucial.
- According to the size and complexity of the parking lot, different multi-camera systems can have different costs for implementation.

B. Utilizing object detection for parking space management

Object detection systems that detect cars and assign them to parking spaces have grown more possible in recent years, it advances in object detection technology. The only referred to method in this category is a Convolutional Neural Networks (CNN)-based car detection method. Once trained, the CNN uses a window that moves method to search the entire lots image for cars. But the system isn't complete because the results aren't linked to particular parking spaces. With our work, we hope to make this better by developing an all-inclusive automated parking management system. Among the techniques we use are vehicle detection, automatic spot mapping, multi-camera combination, homographic transformation, and perspective correction to make use of already installed cameras. Additionally, we have created a practical dataset that consists of our thorough and diligent evaluation of the proposed approach addresses the drawbacks of present approaches such as make identification and the division of images, which are altered by changes in light, shadows, and challenges. Proposed system of its kind, which uses object detection to identify

***Corresponding author:** dlaxmics@gmail.com

vehicles and assign them to specific parking spaces. This approach enormously cuts off the quantity of time and effort needed for arrange by doing away with the requirement for individual annotations of every parking spot.

C. Classifying the congruency among Traditional and Modern Approach

Due to limitations in publicly available datasets, a direct quantitative evaluation of our put forward object recognition system and existing methods is not possible. However, because of the system's creative nature, we can demonstrate its theoreticalbenefits. Unlike many existing methods, which require time-consuming human marking of every parking place; in comparison, our solution just requires the total number of places and the corners of the parking lot. This automatic vehicle mapping feature illuminates in larg e parking lots where labelling by hand becomes difficult. Furt hermore, when compared to image segmentation-based systems, our approach is more flexible to varying background due to climate or lighting changes. This separates it from previous efforts while rendering it actually robust in everyday situations.

The key aids of this article work are follows as:

- This approach addresses the limitations of existing methods prone to varying lighting, shadows, and occlusions.
- Automatic mapping detected vehicles to specify parking spots, providing a clear picture of parking availability.
- This System demonstrates resilience to changes in lighting which can impact image segmentation methods.
- Integrate data from multiple cameras, fill in blind spots and further it aid to improve accuracy.

This study suggests a modern object detection based autom ated vacant parking space management system. It overcomes present drawbacks by automatically mapping detected vehicle s to specific parking spaces and illustrating reliability against variable background and object closures. The system undergoes evaluation on a realistic multi-camera dataset with varying lighting and adverse conditions, indicating its relevance to real-world parking management.

II. Literature Review

An overview of the work done on the suggested autonomous parking management method, which looks for parking lots that are occupied or vacant, is provided in this section. Everything was put up by it. Based on how occupied/free approach locations are classified, these studies are grouped into three categories: image division, machine learning approaches (SVMs, NN, etc.) done to identify patches, and vehicle identification techniques based on object recognition systems. Various smart car parks systems that use cameras to find vacant spots have been developed over the last ten years. Several vehicles to detect vacancy, while others model the surface or the backdrop to eliminate vehicles that are parked. Methods for vacancy detection include tracking, background subtraction, and classification, but all have limitations in real-world parking scenarios. However, these can still be incorrect. Previously used systems rely on tracking, subtracting background information, or classification, but all have limitations in practical parking scenarios.

R.M. Nieto et al.(2018) utilized variety of of Deep Learning techniques, including CNN (Convolutional Neural Networks) and Faster R-CNN (Regional with Convolutional Neural Network Features) detectors, which are more effective versions of earlier R-CNN and Fast R-CNN finders, mainly in terms of calculating cost but also in terms of performance [1]. All three variants combine rich characteristics calculated by a convolutional neural network with bottom-up region suggestions. With an accuracy of 90.09%, Faster R-CNN was found to be the approach with the greatest effectiveness.

K. Madhavi et al.(2023) has worked on the different object detection model like Mask R-CNN way has already been presented to use a drone camera to identify the availability of parking places. District-based R-CNN is a convolutional neural network that products class marks and skipping cases per every input along with to a conviction score. The Mask R-CNN Classifier has been created to be confirmed using the previous processed and labelled photos. Our preliminary findings indicate damage may be really linked with 95.13% accuracy on an adapted dataset and 96.87% accuracy on randomly chosen photos. [22].

F. Falchi et al.(2016) proposed a new approach for object detection model like CNN and AlexNet. Both of these models have linked layers, one of which is the output layer. Locally response normalization (LRN), linear rectification (ReLU), and max pooling come after the initial and second convolution layers (conv1-2) [5]. LRN is not utilized by the third layer of convolution (conv-3). This model's accuracy was ninety percent.

B Sairam et al.(2020) has implemented a model that they Parallel to Faster R-CNN, but through the addition of a film of FCN for bounding box segment and object mask proof of identity, Mask R-CNN is used to categorize both booked and parking slots. Here, the face mask serves to pinpoint the precise area that the vehicle covers; if this is less than fifty percent, the area is set aside as partially allocated for light-duty vehicles [23]. With a mask identification rate of more than 92.33% and a boundary recognition percentage of 98.4%, the suggested system shows increased robustness.

Mr. S.S Kumar etal.(2022) demonstrates deep learning(DL) algorithms with outstanding performing methods like Mask R-CNN, also known as Mask RCNN, that represents the cutting edge of image and instance segmentation. On top of Faster RCNN, a Region-Based Convolutional Neural Network, Mask R-CNN was created[12].The performance of YOLO was marginally superior to

Mask R-CNN, exposed by 98.96% and 96.73% precision and 80.93% and75.43% recall, individually.

C. Fatichah et al.(2020) has created an intelligent parking system that uses deep learning algorithms to analyze camera footage and detects when parking spaces are available. This system is divided in two phases. The first thing to do is to indicate the space location on the camera-captured frame of a parking area[8]. A starting picture of a full parking lot is marked with the parking location using a Pre-processed Region-based Convolutional Neural Network (Mask R-CNN), according to this research. Upon the marking of the space spot, AlexNet is used to test if parking spaces are available on video footage, and the results show an accuracy of 73.73%.A new car recognition technique based upon a CNN (convolutional neural network) has been presented by Q. Xiang Wu et al. (2016). Its purpose is to identify and confine vehicles in an open park.

III. MATERIALS & METHODS

3.1 Materials

A. Camera placement and acquisition system

The parking lot data has been collected from large vacant areas in the open space. A contribution of this paper work is fusion with a dataset contain 2000 pictures of marking places that are both empty and occupied, situated on a 250-space parking lot. This dataset was gathered via vecteezy using 4K aerial car departures from an empty parking lot (1290685). The testing data size of 854 x 450 pixels covers all parking area problems. This layer's goal is to locate significant picture patches and compile potent attributes. For our system to identify the picture patches of interest automatically, there has to be a camera calibration step.

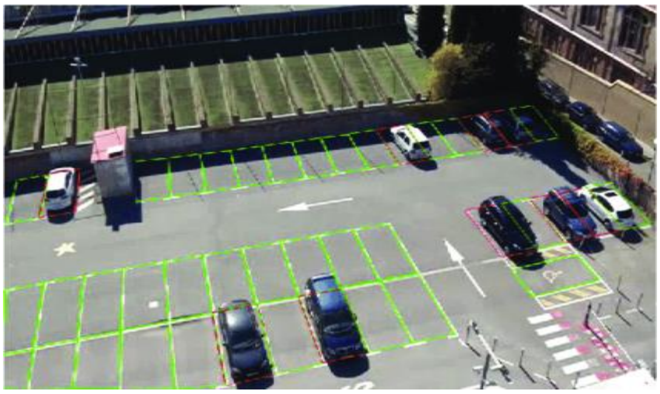

Fig.1 Image acquisition for data collection

The answer basically consists of processing Information acquired from multiple cameras is chosen or adjusted by adding more cameras to offer an entire set that allows the system to function properly. since most parking lots have easily available security cameras.The approach that is being presented primarily aims to address difficult regions such as underlies and occlusions. This is particularly apparent in heavily transferred large parking lots.

Fig.2 Parking lot images captured as (a), (b), (c) in different angles under changing in weather and lighting conditions

3.2 Methods

A Model Approach Overview

This research tackles real-world parking management through object-detection system. Images from a multi-camera setup with varying lighting and weather are captured and carefully labelled. This data fuels the training of the object detection model, which then generates a real-time parking lot occupancy map from live camera feeds. Importantly, the system leverages multiple cameras to handle occlusions and adapts continuously by integrating identified vehicle/vacancy information, making it robust and efficient for managing real-world parking scenarios. The model approach commenced with data caoture, entailing the collection of images from the parking lot under diverse conditions. Subsequently, data annotations was performed to meticulously label the captures images, pinpointing the locations vehicles and vacant parking sapces. The annotated dataset constituted the foundation for training model, a process that progressively refined its capability to detect vehicles and vacant spaces within images. Upon completion of training, the model was deployed for parking lot mapping, generating a visual representation of the occupancy status of parking lot. The model operated by continuously processing real time images of the parking lot, identifying occupied and vacant spaces, and storing this information within the final model. This comprehensive model, encompassing both detection outcomes and parking lot mapping, serves as the cornerstone for facilitating efficient parking management and decision-making. The real-time image processing system utilizes advanced algorithms to sense available and unoccupied parking spaces, contributing to a dynamic in addition up-to-date parking lot mapping. The continuous refinement of this comprehensive model enhances its accuracy over time, providing a solid foundation for effective parking management strategies and informed decision-making. This integrated approach optimizes resource allocation, improves user experience, and supports sustainable urban mobility.

This workflow is based upon the flow of the process in model building and generation of the outcome and finally the parking lot object detection data will be stored in a database and it will continuously monitor the lot and it can tolerate the weather and lighting conditions and it's beneficial for large vacant areas to occupy the parking areas.

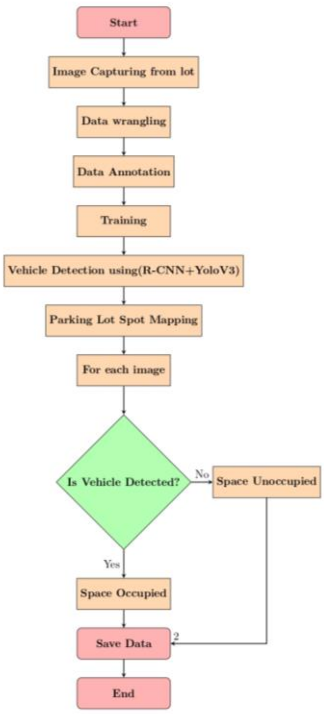

Fig. 3 Flowchart for developed system

Some of the main key scenarios in the flowchart:

- **Data Capturing:** Images have been taken in a multi-camera setup with varying lighting and weather situations. This assures the system's resilience to variances in the actual environment.
- **Data Annotation:** The images are taken and annotated with the open parking spots. This annotated data is used as the object detection mode's training set.
- **Model Training:** The object detection model is trained using the annotated dataset. Through the model's capacity to recognize cars and empty spaces.
- **Parking lot spot mapping:** This model analyses footage, occupied and vacant spaces in each frame to build parking map.
- **Data Storage:** After each vehicle detection, the model turns into an evolving map that knows from its comprehension of parking patterns.

B. Methodology

To locate parking spots, this study employs a variety of techniques. The contrast is increased, coordinate points are automatically marked, and this procedure identifies each vehicle's box binding as the observed item. In this instance, it is best to employ the search process throughout the design and building phases to ensure complete control over the parking lot.: multicamera fusion (assuming multicamera setups are generally available), perspective correction (to allow the reuse of existing camera installations), automatic spot mapping, and vehicle detection.).Preprocessing is required after obtaining the raw image from a camera affixed to a pole or another tall structure, such as a building. This stage of the process is crucial because it determines how accurate the suggested method will ultimately turn out. This part contains different levels of implementation steps for deploying the model and it helps out to reduce the problem time consumption and the detailed way of implementing will pertain all the necessary data and methods for the overall process of the model. In which the model includes some machine learning tools and algorithms to deploy the entire model on a user interface. The model included here are more part deep learning methods [5]. The flow of the methodology looks like as below:

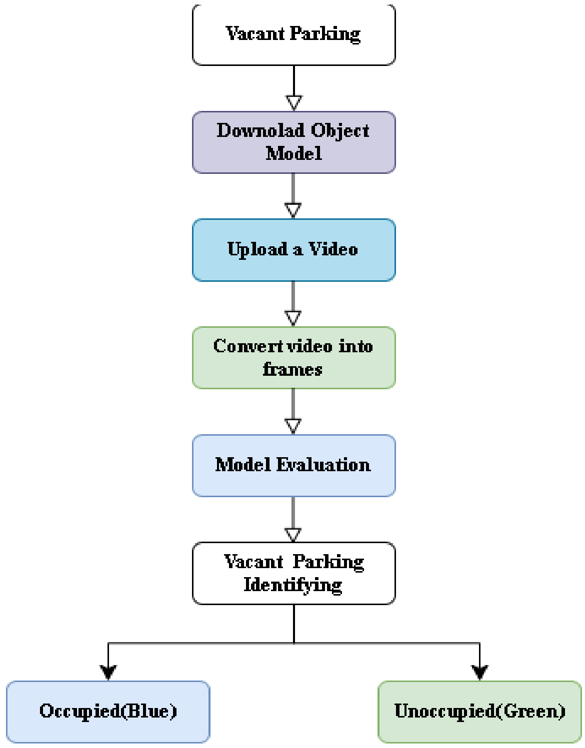

Fig. 4 Block diagram

IV. Proposed Model

A. System Overview

Since security cameras are usually readily available in parking lots, the setup consists primarily of processing the data obtained from the cameras that have been placed throughout the lot or adding more cameras to complete the transmission in order to provide full inclusion that enables the system to function. Applying their findings for the proper operation of independent halting the administrator's systems is presently feasible since object location computations have recently evolved. Without requiring the deployment of brand-new cameras, the system is made to work with already-existing parking lot security cameras with just a small amount of configuration (Fig 4.) The system's performance is limited or impacted by more complex scenarios than those addressed by advanced systems, such as nearly complete occlusions and weather variations (cloudy, rainy, snowy, etc.). Due to the very changing background in this instance, it is not feasible to precisely extract the foreground or define the region of each location since some parked cars entirely obscure the locations behind them. In addition, a situation is taken into consideration that hasn't, as we know, been standard before for these kinds of systems.

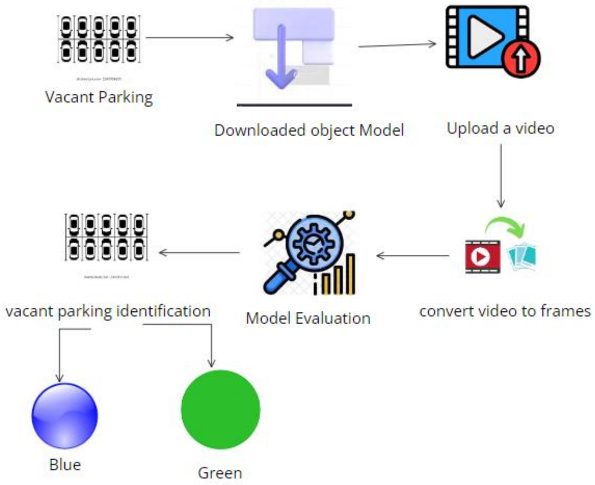

Fig 5. Model Connection layout

The idea behind the proposed multicamera system is to process each camera in parallel before combining (or fusing) the separate outputs. The system's model diagram is shown in Fig. 5.The terms shown in the above image are:

- **Image Data:** A collection of 1000 labelled images capturing parking lot scenes, used to get the vehicle detection model.
- **Training:** Using the collection to teach the model to accurately detect vehicles in new images or videos. Feeding labelled and train images: The model processes labelled images to learn patterns and features that distinguish vehicles from other objects.
- **Model**: The trained algorithm that identifies vehicles in new, unseen images or videos. Some models covered in: Regional space Convolutional Neural Networks (R-CNNs) and YOLO (You Only Look Once): Efficient model for object discovery.
- **Identification**: The processed video stream generated by the model, with visually marked vehicle detections and parking space status.
- **User Interface**: The interactive interface that presents parking information which helps out every user to know about the vacancy in parking area. It includes some features:

In the era of Object recognition and automated vision, a comprehensive approach is employed for developing a vehicle detection model. Initially, a dataset comprising 1000 labelled images capturing various parking lot scenes is assembled. This dataset serves as the foundation for training the model, with an emphasis on teaching it to accurately identify vehicles in new images or videos. The training process involves feeding labelled images into the model, allowing it to discern patterns and features crucial for distinguishing vehicles from other objects. Noteworthy models such as Convolutional Neural Networks (CNNs) and YOLO (You Only Look Once) are explored for their efficiency in real-time object detection. Once trained, the algorithm transforms into a model capable of identifying vehicles in new, unseen images or videos. The model's output manifests as a processed video stream, marked with visually highlighted vehicle detections and parking space status. To enhance user interaction, an intuitive user interface is designed, presenting parking information that includes features aiding users in understanding parking vacancy status within the area. This integrated system harmonizes cutting-edge models, training methodologies, and user-friendly interfaces to deliver a robust solution for vehicle detection and parking space management.

- Real-time parking map indicate available spaces.
- Parking guidance arrows to available spots.

In the domain of computer vision, a vehicle detection model is crafted through the curation of a labelled dataset containing 1000 images capturing parking lot scenes. This dataset becomes the training ground for the model, leveraging techniques from Convolutional Neural Networks (CNNs) and YOLO for efficient real-time object detection. The model, once trained, processes new images or videos, marking vehicles in a generated video stream. The culmination is an interactive user interface that provides parking information, showcasing visually marked vehicle detections and parking space statuses, thus offering users a streamlined experience in monitoring parking availability. The user interface for parking information can be significantly enhanced by integrating various features that capitalize on the capabilities of the trained vehicle detection model. To begin with, real-time occupancy updates can be incorporated to provide users with the latest information about parking space availability. This could include visually marked indicators on a dynamic map, allowing users to quickly identify areas with vacant parking spaces.

Moreover, the user interface could employ a user-friendly dashboard that not only displays real-time information but also offers additional functionalities. For instance, implementing parking space availability indicators could visually represent the status of each parking spot, making it easier for users to navigate and choose parking locations. A dynamic map that actively showcases detected vehicles and their corresponding parking spaces adds a layer of interactivity, enabling users to efficiently locate available spots.

4.1 Image Acquisition

In this image acquisition system, the images are mainly collected from camera's which are allocated in the different viewpoints like top view, side view and front view of the parking area. Mainly, the images are collected from 4K Aerial Car Departs from Empty Parking Lot 1290685 from vecteezy. Our system mainly focuses on the camera calibration and capturing the wide range images from long view. One of the views captured from top angle and make it something different and it helps to capture all the parking spaces that can be observed in fig. 6 which is also performed with frame extraction along with detailed mapping.

Three separate locations inside the parking lot should have been designated for the placement of the long and wide vision cameras that were used to capture the image. Additionally, the camera locates things from a distance and primarily concentrates on those within the area. We will need to look at 2000 distinct zones that this camera will provide. Although this

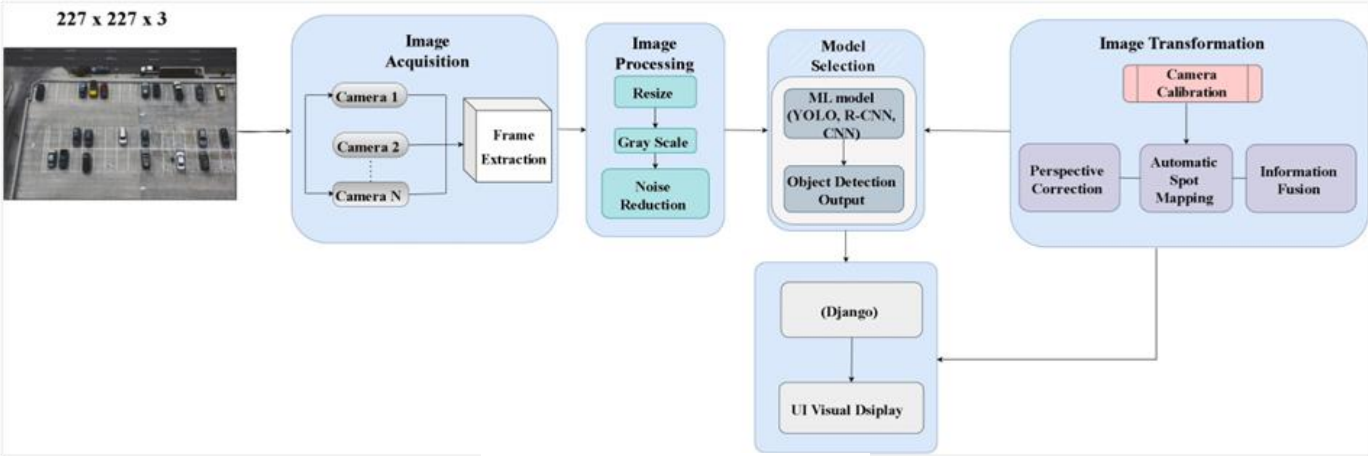

Fig. 6 Layer configuration architecture

seems like a large amount, it's really rather little in comparison to the brute-force sliding window approach. This means that our features are 4096 dimensional. It is also responsible for the frame extraction with exact pinpointing location on the parking lot and it obtain the frames it will remove 40% of the unwanted areas in the image.

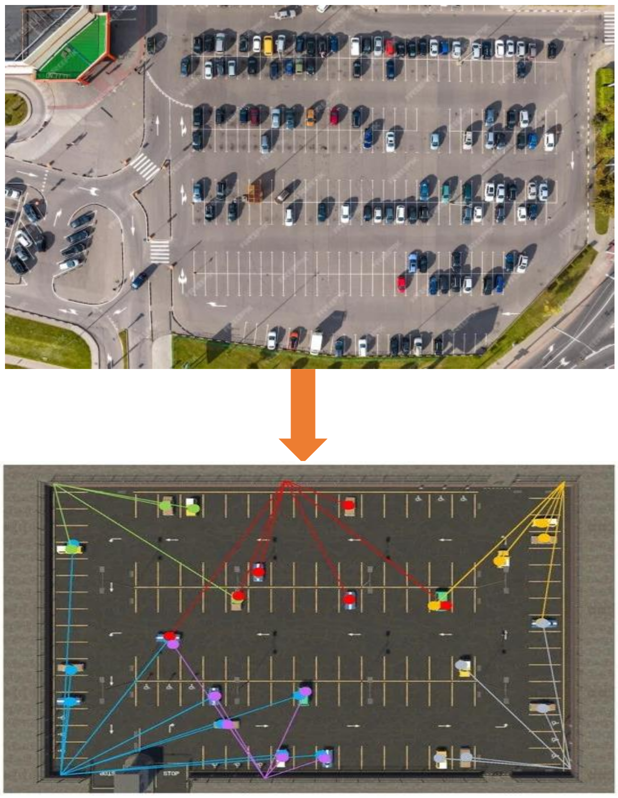

Fig. 7 Parking lot from top view

4.2 Image Processing

In the image processing layer, it handles some other layers in it for making clear and concise view of point in image with putting all the necessary image enhancement qualities. Those are mainly named as:

Resize: It simplifies image size to enable effective model processing. While larger images may capture more detail, smaller ones require less processing.

Grayscale Conversion: Compared to colour images, pictures in grayscale are less prone to lighting changes. It eliminates colour variations to simplify image information. In grayscale, some

features, such as shapes and edges, stand out more, which helps with identifying vehicles.

Noise reduction: Eliminates unwanted noise introduced during camera capture or transmission. Less noisy data V allows the model to learn generalizable features, improving performance on unseen data.

4.3 Methodology

4.3.1 R-CNN

The R-CNN system looks for solutions to the object detection issue, which is the location of things in an image. The region suggestion approach has no bearing on the R-CNN. In the past, item positions and classes were found by employing big and tiny window frames to scan each grid place throughout a picture. Using traditional method i.e. HOG (Histogram of Oriented Gradients), researchers collected features and fed them into SVM to carry out classification tasks. On the other hand, it would take too long to apply CNN to every window frame. To extract regions of interest, RCNN employs a region proposal method, which is essentially bounding box or selective search (greedy algorithms that partition ROI then combine them) By applying CNN to each region proposal and employing Selective Search to identify candidate regions, R-CNN mitigated this issue by lowering the amount of convolutions. It is depicted in (Fig. 7).Additionally, some essential components of this approach are: Bounding box refinement, categorization, feature extraction, and Region Proposal Network (RPN). locations that are represented in the model in (Fig. 8). We combine region proposals into fixed size region proposals as they are not all the same size.

Regional Proposal Network: Recommended areas are the best subsets of the unique image that we agree with desire to have the issues we're observing for.. There was a clear statement of the various region proposal methods available in (Fig. 8). These are "normal" algorithms that function immediately. They don't need any training from us at all. In this study, region ideas are generated through the application of the selective search approach.

4.3.2 Convolutional Neural Networks

The CNN in the article receives instruction to differentiate between two distinct groups. The deep learning algorithms, such as CNN, are comprised of several neural networks and can automatically extract information. Although they may also be utilized for CNN, convolutional neural networks (CNNs) are an artificial neural network type that is similar to BP neural networks and are mostly used for image recognition and computer vision tasks. other forms of data processing, such natural language processing. A collection of two-dimensional neuron arrays, each consisting of a collection of individual neurons, make up each layer. An array of neurons represents a certain type of feature.. In CNN, it mainly involves in types of deep neural network excelling in image recognition and feature extraction [11]. CNN has more layers while in case of model training. And those layers are represented in the below figure (Fig. 9) which is proposed architecture for CNN. And the process of the algorithm provides key features: Efficient feature extraction, adaptable to various parking lot conditions which are responsible for parking management. Each layer is precisely processed in given below (Fig. 9).From there we can now the importance of each layer in this algorithm.

A. The Convolution Layer: The convolution layers entail the automated learning of convolution kernels to convolve the input feature maps. The resulting output is then passed via activation functions to produce output feature maps that the network may use.

B. The Sub-Sampling Layer: If there are N input maps for the subsampling layers, then there will also be N output maps. Furthermore, the output map's dimensions are the same as those of the input map, but its size is smaller than the input map's.

C. The Fully Connected Layer: The CNN's last layer, known as the fully connected layer, is identical to the BP neural network..

4.3.3 Mask R-CNN

Mask R-CNN, also known as Mask RCNN, is a Convolutional Neural Network (CNN) that represents recent advances of image and case portion. On top of Faster R-CNN, a region-based CNN, Mask R-CNN was created. By adding branches to mask objects in combination to identify square limit, Mask R-CNN is a method derived from Faster R-CNN. To build a system that can accurately assess whether a parking spot is occupied or vacant, as well as how much space a car occupies, the classification problem must be separated from the pixel-level mask prediction. Fast R-CNN model still employs a selective search technique to get the proposal region, but it now includes the Region of Interest (ROI) Pooling module. If the vehicle percentage area occupied is higher than 50%, the model suggests that slot is completely occupied.; if not, it detects the slot as partially parked .That training process shows in below (Fig.10).

branches to mask objects in combination to identify square limit, Mask R-CNN is a method based on Faster R-CNN. The goal is to separate the classification issue from the pixel-level mask prediction in order to build a system based on Mask R-CNN that accurately determines whether a parking space is full or vacant, as well as the area that a car covers. The Fast R-CNN model continues to employ a selective search technique to get the proposal region, but it now includes the Region of Interest (ROI) Pooling module. If the vehicle percentage area occupied is greater than 50%, the model shows that the slot is fully occupied. if not, it detects the slot as partially parked .That training process shows in below (Fig.10).

Mask R-CNN, an allowance of the Faster R-CNN model, stands as a pivotal advancement in image and instance segmentation within the realm of Convolutional Neural Networks (CNNs). Built upon the foundation of Faster R-CNN, a region-based CNN, Mask R-CNN introduces additional branches for mask prediction, enabling the precise identification and delineation of object boundaries at the pixel level. This innovative technique goes beyond mere object detection, allowing for detailed segmentation and recognition of distinct instances within an image. The fundamental breakthrough lies in separating the classification task from the mask prediction process, facilitating a more nuanced understanding of visual scenes.

The architecture of Mask R-CNN is characterized by its ability to simultaneously address object detection and instance segmentation. By extending Faster R-CNN with dedicated branches for mask prediction, this model not only discerns objects within an image but also provides a pixel-wise representation of their boundaries. This proves particularly useful in applications such as determining the tenancy of parking spaces. For instance, in context of parking

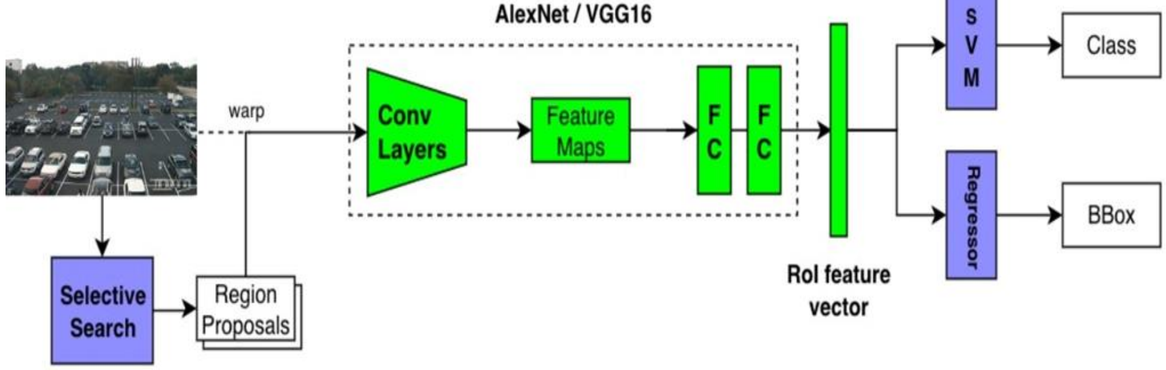

Fig. 8 R-CNN Architecture

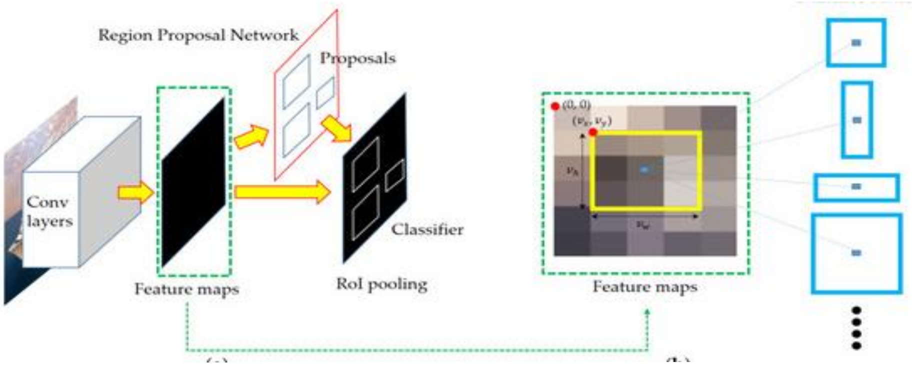

Fig. 9 RPN regional space for parking

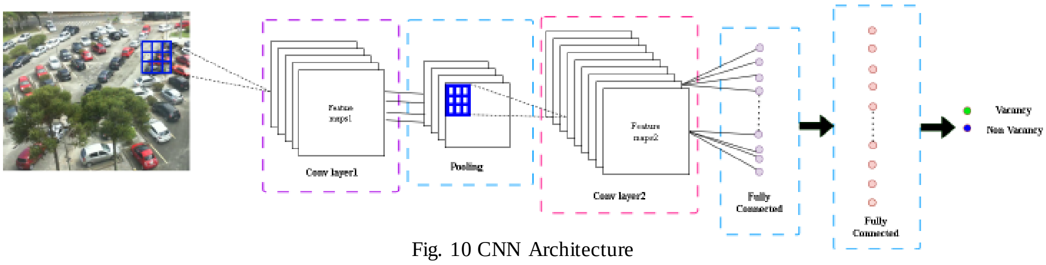

Fig. 10 CNN Architecture

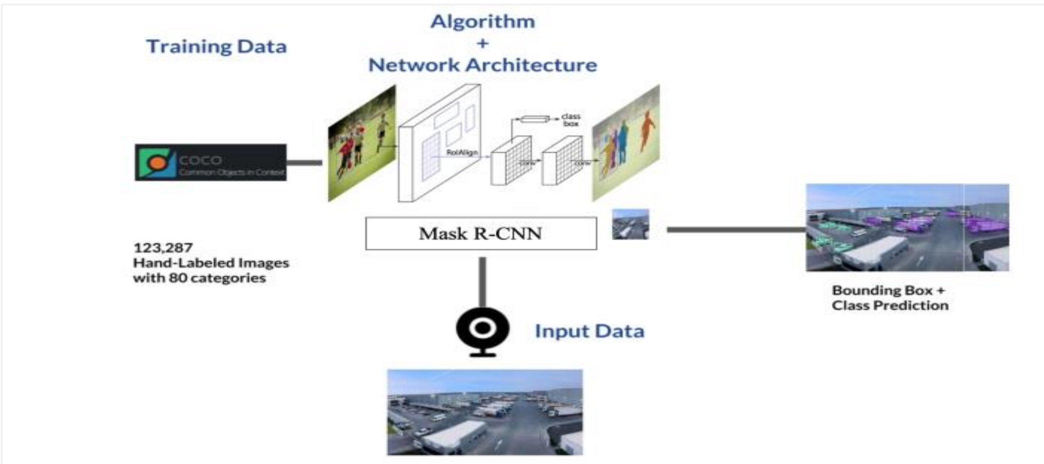

Fig. 11 R-CNN

When it comes to parking spot management systems, Mask R-CNN can reliably determine if a particular space is filled or empty. It also goes a step further by accurately delineating the space occupied by a vehicle, enabling more detailed analysis and decision-making.

Two phases are included in the Mask R-CNN technique. Its early phase is same to Faster R-CNN's. The binary mask for the next step ROI was created using the Mask R-CNN outputs in conjunction with the box's balance and base class. Based on the predictions of the mask, regression and bounding box procedures are performed in parallel with the classification process. The multitask loss during training is determined for each sample..

As seen in fig. 10, Mask R-CNN, an expanded variant of Faster R-CNN organizes. This mask component adds little calculating cost, allowing for rapid system creation and testing.

$$L = L(box) + L(mask) + L(class)$$

In practical terms, this means that Mask R-CNN can contribute significantly to smart parking solutions and traffic management systems. Its pixel-level mask prediction ensures a fine-grained understanding of the spatial distribution of objects within an image, making it a powerful tool for applications where precise localization and segmentation are crucial.

where L(mask) is the middling binary entropy loss over K classes, L(box) is the box's boundaries Report Phrase loss, and L(class) is the level of class_ loss.

4.3.4 AlexNet

AlexNet is a development of the algorithm used for binary classification. Alexnet is well-known for its deep convolutional neural network design and is often used in academic studies. To find automobiles in the camera frames, AlexNet is used. It facilitates determining the existence

of automobiles or vehicles in the parking spots that the cameras have recorded. It may be seen in figure 12 below. The first stages of analyzing video data, identifying cars, and classifying parking spots according to their occupancy state all make use of AlexNet. Effective parking space management can then be achieved by the use of this data in higher level decision-making processes. network architecture with AlexNet as inspiration. Two FC layers, with the output layer, and three convolutional layers are employed in AlexNet. First and second layers (conv1-2), local response normalization (LRN), max pooling, and and rectification linear (ReLU). The third convolutional layer (conv3) doesn't use local response normalization (LRN). The total number of conv1-2 filters and neurons in the layer (fc4) were significantly reduced to match the issue dimensions, resulting in a structure with about 4096-dimensional plane than AlexNet. There were no regularization of dropouts in fc4 and fc5 (the final outcome layers).

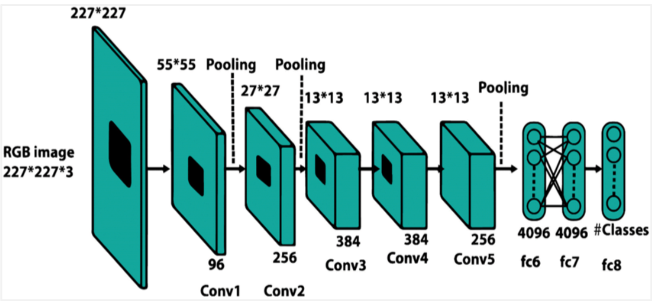

Fig. 12 Architecture for AlexNet

4.3.5 YOLO-V3

Yolo's quick identification speed makes it appropriate for real-time item detection scenarios. This article uses the YOLO v3 network for car detection in parking lots and parking slots. A residual structure for YOLO v3 is included in this publication. The extraction approach uses four different scale feature maps for item recognition and automobile parking space characteristics, which enables deep neural networks to extract more incredibly small data. This work uses four different size feature maps for object identification and residual blocks to extract deep car parking space characteristics based on YOLO v3.

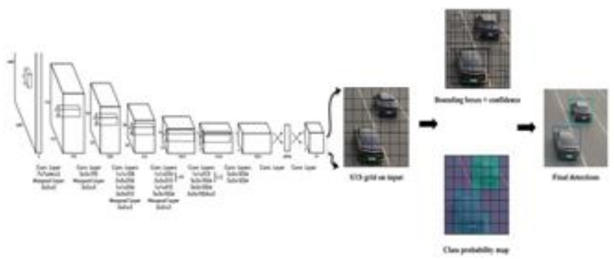

Fig. 13 Yolo v3 model Schematic diagram

As a result, finer characteristics may be extracted by deeper networks. The topic of this work is parking space detection. The item examined in this research and which was shown in (fig. 13) is not suitable for the anchor box created by the original network.

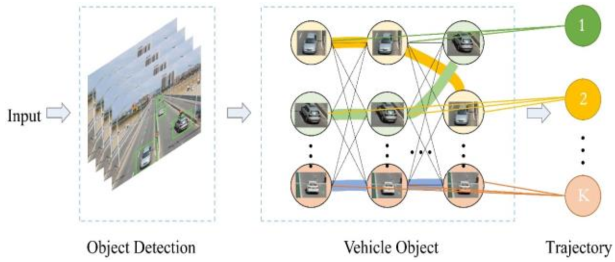

Fig. 14 Representation of object from trajectory

Therefore, to group and categorize the automobile parking spots in the dataset, the K-means clustering technique is applied. Anchor boxes are introduced in YOLO v3 and are utilized in Mask RCNN.

4.3.6 Object Detector

The object vehicle finder is initialized using a car model, and each camera receives its region of interest (ROI) mask to filter out detections from other areas of the parking lot not covered by the cameras. (Figure 14 illustrates an example of one of these masks). This component reads frames from each camera and generates bounding boxes (rectangles) for detected items (cars) using an object detection method. As per the experiment sequences, the existing vehicle models were trained specifically on new vehicle models to address issues such as high perspective, scaling variability, occlusions, and multiple vehicle types.

The primary detection technique employed to assess the upcoming system is the Faster R-CNN (Regions with Convolutional Neural Network Features) indicator. It offers improved performance and computational efficiency compared to its predecessors, namely R-CNN and Fast R-CNN detectors. These advancements are achieved complete the utilization of a convolutional neural network for feature extraction coupled with bottom-up region proposals. A fundamental component of Faster R-CNN is the incorporation of a Region Proposal Network (RPN), which generates region proposals at almost no additional cost.

4.4 Image Transformation

A. Camera Calibration

The bounding box of each identified vehicle is determined by the object detector from the preceding block from the perspective of every camera. This area modifies the location of the detected objects by utilizing the calibration settings. to a common plane from the plane of all camera. First, the block is initialized by using the homograph matrix for each camera, which is created by taking an image from the top view and using four points from each perspective. Obtaining an overhead perspective can be facilitated using tools like Google Earth. While it's not necessary to select identical points for every camera viewpoint, each point chosen must correspond to an image feature. By definition, camera calibrations have 3x3 dimensions for the matrices. Consequently, homograph is simply applied to each bounding box's base midpoint, producing an optimized computation. One point is produced for each vehicle that is detected by this block.

B. Perspective Correction

The displayed places where the results from the previous phases are mapped must have the positions due to the size of the detected items.

Position adjustments are made based on the angle formed by the corresponding points of the space lots and the camera perspective. This enables proper similarity between the vehicle detections and places. The midpoint of the base and the center of the enclosing box are not the same. The midpoint of the base, which is a part of the ground plane, makes it possible to use the attributes of the homograph. Conversely, it is possible that the lens distortion affects the holography's accuracy, leading to errors and roughness in the spot plotting, as seen in (fig. 14).

Fig.15 (a) Initial spot map (b) Corrected grid

It is usual to remedy this issue with intrinsic camera parameters. A switch strategy, which is to use a basic linear adjustment function to rectify the mapping of the grid of spots in the event that these values are unavailable, is shown in fig. 15(b) displays the outcome of this adjustment. To remove the lens's influence, we first establish an even grid and then apply an adjustment factor to the predicted spot..

C. Automatic spot mapping

This section emphasizes the features of homographs. The stated destination places are intended to produce discrete spot numbers on its own, avoiding the need for human registration every place. The source markings are the four corners of the space lots, and the end points are the angles of mock end points shown in fig. 15.

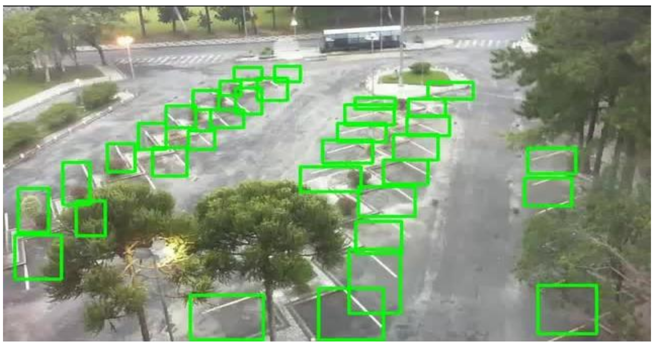

Fig.16 Parking space spot mapping

D. Information Fusion

In multi-camera vehicle detection for parking management, you often encounter problems where features extracted from images don't linearly map to the desired output (e.g., vehicle presence or absence). Non-linear activation functions are useful in this situation. They give the network non-linearity, which enables it to figure out intricate connections between the input and output. Looking into the problem, the convolution layer uses the kernel for feature mapping with some activation function layer i.e., ReLU (Rectified Linear Unit) which is more computational efficient in the non-linearity function to quickly remove the unwanted object from the image and it helps the layer to flatten the image.

$$p(x) = max\ (0, x) \tag{1}$$

ReLU simply apply any negative input values to zero and keeps correct values unchanged. This makes it solving efficient and avoids the vanishing gradient problem that can occur with sigmoid or tanh functions.

To address this issue, another standardized function is sigmoid function, it will be used to combine information from all cameras. This will allow for an easy evaluation of various combination techniques using a different argument. A normalized sigmoid function, P(x), will be used.

$$p(x) = \frac{x}{k\left(1 - \frac{x}{k}\right) + 1}$$

The distance between every camera and the parking center of the lot is represented by x and k is the variable that allows the sigmoid shape to be adjusted. The method offered works well for. The parking lot centre and the camera are standardized to be $0 < x \leqslant 1$. To get the range from -1 to +1 for negative values, the process has to be repeated. This is made possible by providing the absolute value of x to that function.

In this instance, data from many cameras—each of which was trained on an area containing the closest spots—is combined. The situation (mostly the locations of each camera) and the detection technique of choice must be considered while choosing parameter k. if the detected camera has poor resolution, making detection more difficult, or if the detection algorithm's effectiveness drastically decreases with distance. Automatic spot mapping Occupancy matrix Each camera yields ($O_{k,i}$)one occupancy matrix defined as: $O_{k,i}$ (x_d, y_d).

$$O_{k,i}(x_d,y_d) = \Sigma_0^1 \begin{matrix} if\ spot(x_d,y_d) occupied \\ Otherwise \end{matrix}$$

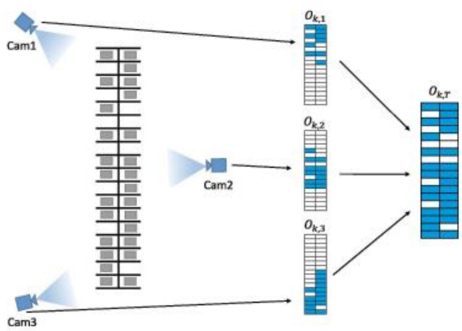

Fig. 17 Multi-Camera calibration with different monitoring views (left, right, top)

V. Evaluation metrics

In this section, we extant every procedure of the experimental results of data with the assessment of each metrics. After training the R-CNN, is it required to search whole image of parking lot to identify the existence of car in the window. The results that are quantified based on the performance to evaluate proposed solution. The results are compared between output and ground truth bounding boxes. The estimation of each metrics follows:

A. Box Loss(Bounding box Regression Loss)

The Box Loss, which measures the accuracy of predicted bounding boxes, is typically calculated using Mean Squared Error (MSE). The MSE formula is given by

$$MSE = 1/n \sum_{i=1}^{n} (Yi - \hat{Y}i)$$

The Mean Squared Error (MSE), utilized in machine learning and statistics, is represented by this formula. Here, n represents the number of data values, (y_i) denotes the actual value of the dependent variable at the i^{th} data point, whereas ($Y^\wedge i$) represents the forecast value. The formula computes the squared difference between the actual and predicted values for each data point. The mean is then calculated by adding the squared differences and dividing by the number of data points (n). This measure evaluates how well the model's predictions align with the actual data.

B. Object Loss(Objectness Loss)

The model's capacity to predict whether an object will be within a bounding box is typically evaluated using Binary Cross-Entropy (BCE) loss. The formula for BCE is:

$$\mathbf{BCE} = -\frac{1}{N} \sum_{i=1}^{n} (Y.\log \widehat{Y}_i + (1 - Y_i) \cdot \log(1 - \widehat{Y}i))^{2}$$

Here, n is the number of models in dataset and $\sum_{i=1}^{n}$ the summation over all individual samples *(i)* from 1 to n,y_i is the true label (ground truth) for the i^{th} sample, where y_i can be 0 or 1 , ($Y^\wedge i$) is the predicted probability that the i^{th} sample belongs to class 1.The loss function penalizes the model based on the difference among the forecast probability $Y^\wedge i$ and the true label (Y_i). The goal is to minimize this cross-entropy loss during the training process, which helps the model learn better representations for binary classification tasks.

C. Class Loss (Classification Loss)

For multi-class classification, the Class Loss is typically calculated using Cross-Entropy Loss:

$$\boldsymbol{Cross - Entroy} = -\sum_{C=1}^{M} y_i, C \log(P_i, C)$$

Here, (M) indicates the number of classes, (y) is a binary indication (0 or 1) indicating if the class label (c) is the correct categorization for the observation (o), and (p) is the predicted class. The formula calculates the cross-entropy by summing over all classes (from 1 to M) the product of the true probability (y_o,c) and the logarithm of the predicted probability (p_o,c) for each class. In essence, it quantifies the difference between the true distribution and the estimate conveyance. A lower cross-entropy indicates better alignment between predictions and actual outcomes. It's a common choice as a loss function for training classification models.

D. Precision (P) and Recall (R)

The Average Precision (AP) mean across all classes and/or IOU (Intersection Over Union) criteria is used to compute MAP for object detection. The precision-recall curve for each class is used to calculate the AP, and the MAP calculation is as follows:

$$\mathbf{P}=TP/(FP+TP)$$
$$\mathbf{R}=TP/(FP+TP) \quad (4)$$

E. MAP (Mean Average Precision)

The Average Precision (AP) mean across all classes and/or IOU (Intersection over Union) criteria is used to compute MAP for object detection. The precision-recall curve for each class is used to calculate the AP, and the MAP calculation is as follows:

$$\boldsymbol{AP} = \int_{0}^{1} P(r)\, dr$$

VI. Experiments and Results

Parking lot data frames are supposed to be have the views in all directions from the top angles(a^o) in wide area that used to design the parking lot with specific lot by keep on view on parking spots in diverse locations. The sequence of the data recorded in the vecteezy parking area in 4K serial lot.

TABLE 1
EACH IMAGE SETS FROM PARKING LOT

Classification name		Edges	Vehicles
Training		640	1784
Test	Sync_ data	1000	482
	All_cam1	500	246
	All_cam2	500	246

After extract the regions of the image based upon the process through RPN network the model can easily recognize the vacant and non-vacant places status with the detailed spot evaluation and ground truth with manually evaluation which previously fall in the camera layers with high resolution and two cameras which are lean on third camera which capture the synchronized views from longer distances.

Later on, when the data from numerous cameras comes together, this will be addressed. In this experiment, the model studies the process entirely and produces the results of the object, class values for object detection concisely. Moreover, the places are captured along with some background area of every image and also it will supposed to detect the object, mask and class for particular location in parking lot The result analyzed and predicted for every epoch in the network and the final analysis should be like:

TABLE 2
OCCUPANCY MATRIX STATUS TABLE

Detective Spot Status	Ground Truth status	Spot evolution
Empty	Empty	TN
Empty	Unavailable	FN
Unavailable	Empty	FP
Empty	Unavailable	TP

In this experiment, the model studies the process entirely and produces the results of the object, class values for object detection concisely. Moreover, the places are captured along with some background area of every image and also it will suppose to detect the object, mask, and class for particular location in parking lot The result analysed and predicted for every epoch in the network. Additionally, it offers critical observations regarding any challenges encountered during the detection process, such as occlusions or complex environmental conditions.

TABLE 3
ANALYSIS ON MASK-RCNN

Epoch	Box loss	Obj loss	Cls_loss
85/99	0.06016	0.192	0.01352
86/99	0.057	0.1843	0.01343
87/99	0.05734	0.2053	0.01313
88/99	0.06049	0.2029	0.01344
89/99	0.05764	0.1987	0.01383
90/99	0.06172	0.2033	0.01288
91/99	0.05814	0.2068	0.01332
91/99	0.05695	0.2106	0.01422
93/99	0.0578	0.1912	0.01328
94/99	0.06397	0.2232	0.01382
95/99	0.05825	0.1990	0.0131
96/99	0.05945	0.1978	0.01415
97/99	0.05778	0.1956	0.01352
98/99	0.05999	0.1968	0.01286
99/99	0.05678	0.2018	0.01301

After all the clear data specified results upon the very class on the object detection with R-CNN with ROI (Region of Interest) layer and for every iteration in the epoch specified on the object loss, class loss and Box loss of the image and generated the data on each curve with all the classes obtained with a MAP-95 with the classification metrics.

TABLE 4
COMPARISON RESULTS OF TWO MODELS

Algorithm	precision	Recall	F1 score
R-CNN	95.02%	94.08%	94.85%
YOLOv3	98.76%	97.65%	98.15%

In this experiment, the corresponding results are carried out by system under parking lot test set, using precision, recall and F1 score as indicator values. False negative indicates samples that are incorrectly marked as positive samples and negative samples too. Which is incorporated in formula (1) and (2) in below as mentioned.

R= True Positive/(True Positive +False Negative) (1)

P=True Positive/(True Positive+ False Positive) (2)

This paper result analysis done with R-CNN and YOLO V3 model for test set of parking lot for rate the model performance and evaluation measure to increase the model backup with more precision and recall values in test dataset

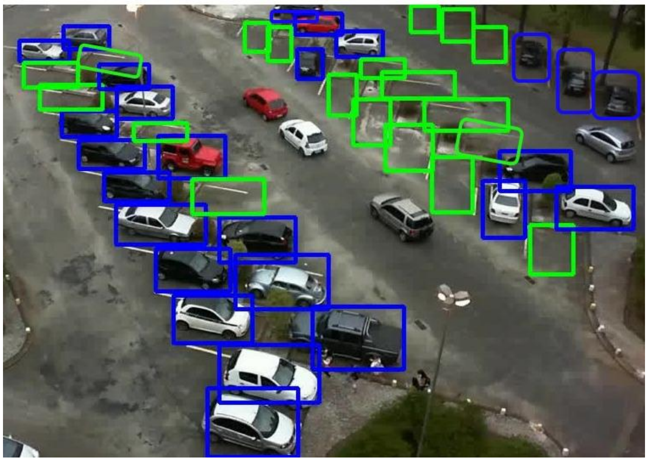

Fig. 18 Experimental result in output window

TABLE 5
RESULTS OF PROPOSED SYSTEM ON IMAGE FRAMES

Image file	Precision	Recall
Frame 1-20 .jpg	95.64	95.64
Frame 21-47 .jpg	97.66	97.66
Frame 48-79 .jpg	98.61	98.61
Frame 79-100 .jpg	96.51	96.51
Frame 101-135 .jpg	97.65	97.65
Frame 136-180 .jpg	98.61	98.61
Frame 181-230 .jpg	97.71	97.71
Frame 231-290 .jpg	95.73	95.73
Frame 310-345 .jpg	96.42	96.42
Frame 346-420 .jpg	97.36	97.36
Frame 421-470 .jpg	96.76	96.76
Frame 471-530 .jpg	97.27	97.27
Frame 531-570 .jpg	96.57	96.57
Frame 571-590 .jpg	98.57	98.57
Frame 591-610 .jpg	97.66	97.66
Frame 611-640 .jpg	95.64	95.64
Average	**97.332**	**97.02**

This result set is classified with the data that presented in the below of the figure of experimental results showed in the final outcome of the model. In this proposed work, we used the pretrained model for Object detection in vehicle parking area. This pretrained Was already been trained with huge number of images of parking area in metropolitan cities, companies, open areas and more vacant space area with vehicles and empty parking area markings that mapped in the entire parking lot dataset for space detection. Our model clearly classifies that parking occupancy matrix for both parkin and vacancy areas widely in spacious area. It has been carried out well and fit for model best outfit with parking images. Finally, our model has given the outcome in terms of precision with yolov3 model is 98.76% and for the RCNN model is 95.02%.That can observe in Table 4.

Simulation and output screens

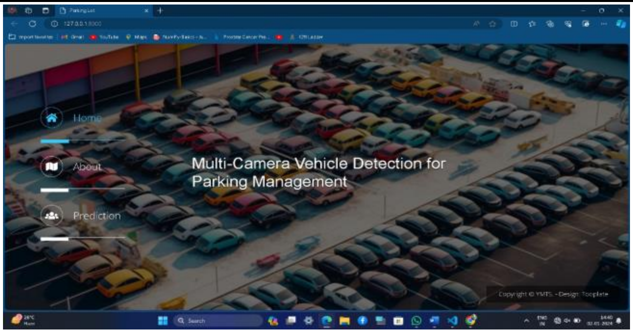

Fig.19 Open Web Browser

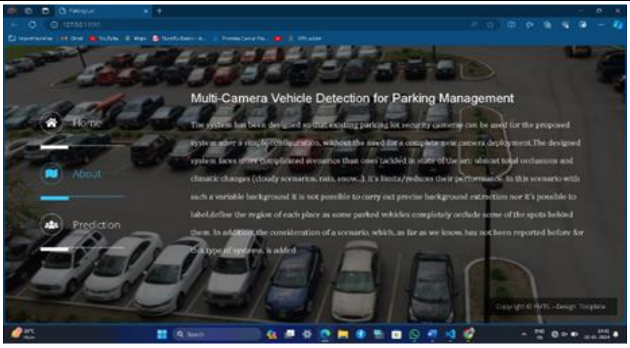

Fig. 20 Select About page for objective

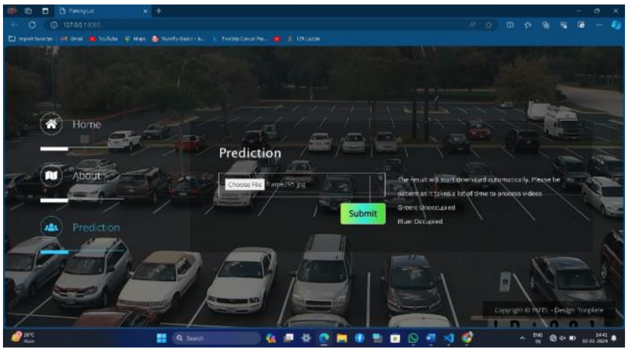

Fig. 21 Choose the input (mp4/img)

Fig. 22 Output Screen

VII. Conclusion

The proposed method relies on a basic principle that performs well under diverse conditions, making it adaptable and reliable. The use of a weight map significantly enhances the method's accuracy and confidence in detecting occupied parking spaces. According to experiments, the proposed approach demonstrates satisfactory performance in parking lot occupation detection [18]. The system utilizes existing security cameras, eliminating the need for additional installations and reducing costs. Unlike methods requiring precise background extraction, this system accounts for variable backgrounds caused by parked vehicles obscuring other spaces. Accurately extracting the background in a dynamic parking lot setting with parked vehicles and changing weather conditions remains a hurdle. Completely hidden parking spaces behind other vehicles pose a challenge for vehicle detection and space mapping.

No new cameras are required to use the system with the parking lot security cameras that are already in place. Compared to earlier systems, this one can manage more complex conditions including nearly complete occlusions and climate shifts. A newly synced and recorded dataset is accessible to the general public. Using video data analysis, the researchers created a system that uses autonomous parking spot detection in outdoor parking lots..

The system uses a two-stage approach:

- The initial step involves marking the space spots on the lot image using a Pre-processed Mask R-CNN, which integrates contrast enhancement with the Mask R-CNN technique.
- Next, feature extraction is performed on the region of interest (ROI) of each marking, utilizing them to classify the ROI as vacant or occupied. This task is accomplished using the Alex Net model.

Future study will focus on a number of areas to enhance the suggested system. Since we have selected a straightforward method for the combination normalized sigmoid functions different functions might be investigated to see how best to fuse or combine the various information sources. To observe how the system behaves under certain circumstances, a new dataset with usual cameras and other layouts may be recorded. A tracker can be added to the series findings to bring together the data got across the series images and provide brief consistency to the vehicle identifications. Furthermore, as the detector is the first step of the system, present and future directions for object detection research may be utilized here.

VIII. References

[1] Nieto, Rafael Martín, et al. "Automatic vacant parking places management system using multicamera vehicle detection." IEEE Transactions on Intelligent Transportation Systems 20.3 (2018): 1069-1080.

[2] [Chen, Y.; Qin, G. Video-based vehicle detection and classification in challange scenarios. Int. J. Smart Sens. Intel. Syst. 2014, 7, 1077–1094.

[3] Yongming, T. Design of Parking Management Information System. Compute. Inf. Technol. 2006, 10, 102–105.

[4] Nurullayev, S.; Lee, S.-W. Generalized Parking Occupancy Analysis Based on Dilated Convolutional Neural Network. Sensors 2019, 19, 277.

[5] Amato, G.; Carrara, F.; Falchi, F.; Gennaro, C.; Machine, C.; Vairo, C. Deep Learning for Decentralized Parking Lot Occupancy Detection. Expert Syst. Appl. 2017, 72, 327–334.

[6] Suhr, J.K.; Jung, H.G. Automatic Parking Space Detection and Tracking for Underground and Indoor Environments. IEEE Trans. Ind. Electron. 2016, 63, 5687–5698.

[7] Jang, C.; Sunwoo, M. Semantic segmentation-based parking space detection with standalone around view monitoring system. Mach. Vis. Appl. 2019, 30, 309–319.

[8] Sliwa, B.; Piatkowski, N.; Westfield, C. The Channel as a Traffic Sensor: Vehicle Detection and Classification based on Radio Fingerprinting. IEEE Internet Things J. 2020.

[9] Wu, L.; Coifman, B. Vehicle length measurement and length-based vehicle classification in congested freeway traffic. Transp. Res. Rec. 2014, 2443, 1–11.

[10] Buch, N.; Velastin, S.A.; Orwell, J. A review of computer vision techniques for the analysis of urban traffic. IEEE Trans. Intel. Transp. Syst. 2011, 12, 920–939.

[11] Daigavane, P.M.; Bajaj, P.R.; Daigavane, M.B. Vehicle detection and neural network application for vehicle classification. In Proceedings of the 2011 International Conference Computational Intelligence and Communication Networks, Gwalior, India, 7–9 October 2011; IEEE: Piscataway, NJ, USA, 2011; pp. 758–762

[12] Mokha, B.S.; Kumar, S. A review of computer vision system for the vehicle identification and classification from online and offline videos. Int. J. Signal Image Process. 2015, 6, 63–76.

[13] Sun, W.; Zhang, X.; Shi, S.; He, X. Vehicle classification approach based on the combined texture and shape features with a compressive, D.L. IET Intel. Transp. Syst. 2019, 13, 1069–1077.

[14] Alex, K.; Ilya, S.; Hinton, G.E. ImageNet Classification with Deep Convolutional Neural Networks. In Proceedings of the Advances in Neural Information Processing Systems 25 (NIPS 2012), Lake Tahoe, NV, USA, 3–8 December 2012; pp. 1097–1105

[15] Wang, H.; Zhang, F.; Cui, P. A parking lot induction method based on Dijkstra algorithm. In Proceedings of the 2017 Chinese automation congress (CAC), Jinan, China, 20–22 October 2017; pp. 5247–5251.

[16] Gantlets, E., Lefauconnier, and A.: The Time Looking for A Parking Space: Strategies, Associated Nuisances and Stakes of Parking Management in France. In: European Transport Conference, pp. 6 (2006)

[17] Wah, C.: Catherine. Parking Space Vacancy Monitoring. Technical Report, University of California (2009)

[18] T. Fabian, “An algorithm for parking lot occupation detection,” in Proc. Compute. Inf. Syst. Ind. Manage. Appl., Jun. 2008, pp. 165–170.

[19] H. xie, Q. Wu, B. Chen, Y. Chen, and S. Hong, “Vehicle detection in open parks using a convolutional neural network,” in Proc. IEEE Int. Conf. Intel. Syst. Des. Eng. , Aug. 2015, pp. 927–930.

[20] C.-C. Huang, S.-J. Wang, Y.-J. Change, and T. Chen, “A Bayesian hierarchical detection framework for parking space detection

[21] framework for parking,” in Proc. Int. Conf. Multimedia Expo, Mar. 2008, pp. 2097–2100

[22] Karanam, Madhavi, et al. "Automatic Vacant Parking Places Management System Using Multicamera Vehicle Detection." E3S Web of Conferences. Vol. 391. EDP Sciences, 2023.

[23] Sairam, B., & Agrawal, A. (2021). VLDNet: Vision-based lane region detection network for intelligent vehicle system using semantic segmentation. Computing, 103(12), 2867-2892.