Research Article

Tracking Vehicle Cruising in an Open Parking Lot Using Deep Learning and Kalman Filter

Vijay Paidi , **Hasan Fleyeh** , **Johan Håkansson** , and **Roger G. Nyberg**

Dalarna University, Faculty of Data & Information Sciences, Borlänge, Sweden

Correspondence should be addressed to Vijay Paidi; vpp@du.se

Received 31 May 2021; Accepted 17 August 2021; Published 24 August 2021

Academic Editor: Alain Lambert

Copyright © 2021 Vijay Paidi et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Due to the lack of wide availability of parking assisting applications, vehicles tend to cruise more than necessary to find an empty parking space. This problem is evident globally and the intensity of the problem varies based on the demand of parking spaces. It is a well-known hypothesis that the amount of cruising by a vehicle is dependent on the availability of parking spaces. However, the amount of cruising that takes place in search of parking spaces within a parking lot is not researched. This lack of research can be due to privacy and illumination concerns with suitable sensors like visual cameras. The use of thermal cameras offers an alternative to avoid privacy and illumination problems. Therefore, this paper aims to develop and demonstrate a methodology to detect and track the cruising patterns of multiple moving vehicles in an open parking lot. The vehicle is detected using YOLOv3, modified YOLO, and custom YOLO deep learning architectures. The detected vehicles are tracked using Kalman filter and the trajectory of multiple vehicles is calculated on an image. The accuracy of modified YOLO achieved a positive detection rate of 91% while custom YOLO and YOLOv3 achieved 83% and 75%, respectively. The performance of Kalman filter is dependent on the efficiency of the detector and the utilized Kalman filter facilitates maintaining data association during moving, stationary, and missed detection. Therefore, the use of deep learning algorithms and Kalman filter facilitates detecting and tracking multiple vehicles in an open parking lot.

1. Introduction

Congestion and pollution from traffic are major problems in many urban areas. Congestion is commonly observed due to high traffic density at peak hours in or close to popular destinations [1, 2]. The majority of the high traffic flows are also related to the low supply of parking spaces. It has been estimated that up to 50% of the traffic in traffic dense environments are trying to find an empty parking space [3]. This indicates that congestion and excess driving that occur during the search of vacant parking spaces lead to increased pollution from the traffic. To understand the magnitude of this problem, the cruising of vehicles should be captured to comprehend how people drive in a parking lot. Previous studies used Geographic Processing Systems (GPS) and visual cameras to generate cruising patterns. According to [4], the time taken to occupy an empty parking space is approximately 1.18 minutes and the data were collected using GPS. In another study [5], a visual camera was placed

on traffic signal poles to capture the number of vehicles cruising for parking in an on-street parking lot. However, there are limited empirical cruising data available within parking lots.

There are basically two types of common parking lots, closed and open. Open parking lots are places where vehicles can be parked without any fee for a limited duration. Therefore, they are in higher demand compared to closed parking lots where parking fees need to be paid. Popular retail stores or destinations often provide large open parking lots, and they are often located in a sparse locality, unlike closed parking lots that are commonly placed in the densest parts of the urban areas. Due to the higher demand of open parking lots, there is a greater need to utilize the space efficiently. Allocating higher parking fees to all parking lots can be one way to reduce the demand. However, it would also impact the business opportunities. The other way is to efficiently utilize the space on parking lots using parking assisting systems [6]. Parking assisting systems utilize

sensors either to get a count of the number of vehicles in the parking lot or to get individual parking occupancy information. Since open parking lots provide free parking spaces, it might not be affordable in many cases to install and maintain parking assisting systems. Since parking assisting systems are not available for open parking lots, it is imperative to quantify the magnitude of additional cruising occurring at open parking lots. Therefore, it is essential to capture the trajectories of vehicles moving in the open parking lot. Being able to detect and identify trajectories of moving vehicles would facilitate understanding the choice of a parking location and counting moving vehicles in a parking lot at a time. Thus, we need to find ways to detect and track moving vehicles in a parking lot.

Multiobject detection and tracking should be performed as there would be multiple vehicles moving in the parking lot. Object tracking using a camera facilitates video surveillance, self-driving vehicles, and robotic systems [7, 8]. Object tracking is complex due to changes in object scale, illumination, occlusion, and rotation [9]. Data association should be performed on detected objects in every frame which might lead to errors due to the mentioned complexities. There is a previous research performed on improving multipedestrian tracking in cluttered environment where object detection was mostly performed using segmentation, foreground detection, and optical flow. However, not much noise was present in those studies compared to the data captured by thermal camera. Noise in this paper is referred to nonfocus objects like pedestrians, bicycles, strollers, baskets, trees, or interference like electronic noise. Over the years, deep learning architectures improved the efficiency of tracking by detection method which makes it suitable for multiobject detection [9, 10]. In the tracking by detection method, objects are detected in each frame continuously and are suitable for videos or real-time detection. Deep learning algorithms are well suited to identify objects of interest with varying heat signatures [11]. Deep learning algorithms such as Yolo are suitable for object detection as they are fast, accurate, and computationally efficient. A modified Yolo was utilized where anchor boxes and size were updated. Anchor boxes are predefined bounding boxes that facilitate performing object detection. A customized Yolov3 was also developed for this study to improve the stability and performance of multiobject tracking. The detected vehicles are assigned IDs using the Kalman filter. Kalman filter is a linear estimation method which is utilized to track objects [12].

In previous literature, vehicle trajectories were generated at roundabouts, lane changes, and other similar road segments to understand driving behavior using a visual camera [13, 14]. However, there is a scarcity of research to detect and track moving vehicles in a parking lot. The thermal camera was also not used for data collection to perform vehicle tracking. This paper intends to address this research gap by generating cruising patterns of vehicles in the parking lot. Therefore, the aim of the paper is to use the thermal camera to detect and track moving vehicles in an open parking lot and evaluate the performance of algorithms in varying environmental and illumination conditions.

The contribution of the paper is to develop a methodology to detect and track vehicles in an open parking lot using data collected from thermal camera. To the best of our knowledge, thermal camera was not utilized in previous studies to detect and track cruising vehicles. The paper also contributes to evaluating the performance of deep learning algorithms like Yolov3, modified Yolo, and customized Yolo on detecting only moving vehicles in a parking lot. Trajectories of each cruising vehicle in an open parking lot were generated using Kalman filters.

2. Literature Review

This section focuses on previous research where multiple vehicles or objects detection and tracking was performed. The suitability of proposed methods for multivehicle detection and tracking using a thermal camera is discussed. Literature search was performed using keywords such as vehicle detection, tracking, object detection, multiobject detection and tracking, parking lots, thermal camera, and deep learning. Google Scholar was the primary database utilized for data collection.

In [15], vehicle detection in various illumination and weather conditions was performed using the first-order difference of Gaussian and multiscale edge fusion detection method. However, this method would not be suitable for an open parking lot where objects such as bicycle and pedestrians cannot be removed completely without compromising quality of vehicle detection. In [16], multiple pedestrians and vehicles were detected and tracked using the optical flow algorithm. Segmentation and motion vector estimation was performed to detect multiple objects. However, optical flow is not suitable for open parking lots as it would also detect pedestrians and other nonfocus objects like bicycles, strollers, and so on. In another study, traffic congestion was identified using videos and images, where locality constraint metric learning was used to capture spatial and temporal information of vehicles simultaneously with textual features and distance metrics [17]. There were no noise or nonfocus objects included in the dataset. Further, grouping of vehicles was performed to capture congestion but individual detection and tracking of vehicles was not achieved. Therefore, the locality constraint metric learning approach is not suitable for multivehicle detection in an open parking lot.

According to [18], vehicles are detected based on color transformation, background subtraction, and edge features. Color transformation and background subtraction using a thermal camera would not facilitate detecting moving vehicles in an open parking lot as the heat signatures of the vehicles will be similar to those of pedestrians and background in bright conditions. Therefore, vehicle detection based on color and edge features is not suitable for open parking lots. In another study, detection and tracking were coupled to reduce detection errors [19]. Background subtraction was performed to identify and track objects. However, this approach is not suitable for noisy data where other nonfocus objects are also visible. In another similar study, segmentation and blob analysis was performed to detect vehicles and Kalman filter was used to track vehicles at

railway crossing [20]. Segmentation and blob analysis is not suitable for noisy data like an open parking lot. According to [21], real-time vision-based vehicle detection and tracking was performed using vehicle shadow features, ROI entropy, and edge features [21]. Extracting edge and shadow features is not suitable for thermal camera data as the shadow of the vehicle is also not visible consistently and is based on varying illumination conditions. It would also lead to missed detection when there is poor or no illumination. In another study [22], histogram of oriented gradients (HOG) was used to detect the rear end of the vehicle and tracked using the Kalman filter. Cleaning of images was done using morphological operations to improve the performance of HOG. However, usage of morphological operations on videos would be computationally expensive and the HOG detector requires positive images of different orientations of vehicle to detect the vehicle. Therefore, this approach is not suitable for detecting moving vehicles using thermal camera. In another similar study [23], surveillance of vehicles was performed using detection and tracking. Detection was performed using morphological operations and decision tree, while tracking was performed using Kalman filter. Detection using headlights is not suitable in open parking lot as the vehicles move in multiple directions and headlights are not visible in different orientations. In [24], pedestrians and vehicles were detected using Haar cascade features and Unmanned Aerial Vehicle (UAV) imagery. The UAV is enabled with a visual and thermal camera. Visual camera data were used for vehicle detection due to varying heat signatures of the vehicles in thermal imagery. The negative images or regions were needed to enable faster detection. However, in this paper, the negative regions also consist of vehicle images that are parked. Since positive and negative images consist of vehicles, this approach is not suitable for open parking lot.

Deep learning has improved the detection process and a wide variety of architectures were introduced. Few such architectures which are fast and computationally efficient are Yolo, GoogLeNet, ResNet18, and single shot multibox detector (SSMD). According to [25], SSMD detected vehicles that are illegally parked on the side of a road. Vehicles that stopped in the region of interest were detected using SSMD and tracked using template matching. Tracking using template matching is not suitable to track moving vehicles as the orientation of vehicles changes in the parking lot. In another study [26], a multibox deep learning algorithm was used to detect vehicles and tracked using optical flow. Vehicles were detected and tracked using an onboard camera where the rear end of the vehicles was visible. The use of optical flow is not suitable for tracking vehicles in a parking lot as stationary vehicles would not be tracked. In another similar study [27], vehicles were detected using rear end lights of the vehicle using AlexNet deep learning detector in combination with lidar data. This approach is not suitable using thermal camera as rear lights are only represented through heat emitted and are represented by pseudocolors. When the vehicle is warm, rear lights cannot be distinguished from the vehicle, leading to missed detection. Therefore, this approach is not suitable for thermal camera.

According to [28], tires and windshield of the vehicles were used to detect vehicles to estimate traffic flow using thermal camera. However, as the vehicles move in different directions, the windshield is not visible to enable detection in an open parking lot. In [29], multivehicle detection was performed using deep learning algorithms and thermal camera. A customized deep learning algorithm was used to detect vehicles using heat signatures. The usage of deep learning algorithms is suitable for thermal camera data and is also used for detecting vehicles in this paper.

Yolo is one of the fast and efficient deep learning algorithms which is utilized in this paper. Yolo utilizes a convolutional neural network to detect multiple objects. It utilizes the entire image for object detection purposes and is comparatively faster than sliding-based methods [30]. In [31], Yolov2, long short-term memory (LSTM), and deep reinforcement learning methods were used for multiobject detection and tracking of pedestrians. A visual camera was used for data collection, and object detection was performed using pretrained Yolov2. The detected object was tracked using LSTM and reinforcement learning. The pretrained detectors are trained on color images and therefore cannot be implemented for thermal camera data. Custom training is needed to detect vehicle in different orientations, positions, and occlusions. There are multiple versions of Yolo available and Yolov3 is capable of detecting smaller objects unlike Yolov2 [32]. In another study, tracking by detection was performed using recurrent neural network and LSTM [33]. However, the detection was not performed in a noisy environment like an open parking lot. According to [34], parked vehicles were detected and tracked in varying illumination conditions using videos. Background modeling and subtraction was used to detect static objects and vehicles were detected using fast corners and template matching. However, background modeling and subtraction is ideal to detect parked vehicles and is not suitable for tracking multiple moving vehicles in an open parking lot.

Based on the relevant literature discussed above, multiobject detection and tracking is a complex problem, and it is performed using several methods. However, multivehicle detection and tracking using thermal camera in an open parking lot has not been performed in previous literature. Open parking lot is a noisy environment which consists of vehicles, pedestrians, trees, strollers, baskets, etc. Further, there are just a limited number of studies conducting detection and tracking in noisy environments such as an open parking lot. In this paper, multiobject detection and tracking is only performed on vehicles and other nonfocus objects like pedestrians, obstacles, and bicycle are not detected. Due to varying illumination conditions, heat signature of vehicles, obstacles, and other nonfocus objects, deep learning algorithms are selected to perform vehicle detection. Deep learning algorithms like Yolov3 are fast and efficient in performing object detection on videos in varying illumination conditions. Similarly, Kalman filter is a reliable and computationally efficient tracking estimator. Therefore, Yolov3 and Kalman filter were utilized to detect and track moving vehicles in a parking lot.

3. The Proposed Methodology

This section gives a description and motivation of the choice and setup of the observational test site where the camera is installed. It includes description of the dataset used for training and testing purposes along with utilized detection algorithms, evaluation metrics, and tracking estimator.

3.1. Description of the Test Site. The investigated parking lot is in a mid-sized city in Sweden with a population size of approximately fifty thousand inhabitants. The parking lot is located at a large shopping center as depicted in Figure 1(a). The setup is illustrated in Figure 1(a) which depicts four entrances/exits which are marked as E1, E2, E3, and E4 and the thermal camera is mounted on the roof of the shopping center. The highlighted region marked by green line in Figure 1(a) and Figure 1(b) is the region of interest (ROI). The selected parking lot is one of the main parking lots of the shopping center and reasonable traffic is expected during the opening hours. The maximum time to park a vehicle in the parking lot is 3 hours. So, there is movement of vehicles intermittently throughout the day. Vehicle traffic is expected to surge during lunchtime and evenings during the weekdays and weekends. The total number of parking spaces is 542, while the parking spaces covered in the ROI are 65. The ROI is selected due to the limitation of vehicle visibility outside ROI. However, the use of ROI serves the purpose of this study. The ideal position of a camera to detect vehicles would be to look at either the front or rear end of the vehicles. However, there was no tall structure placed in that location of the parking lot. Thus, the camera was placed on the shopping center which is circular in structure as illustrated in Figure 1(a), which signifies real-world problems. The installed thermal camera is Axis Q1942-E which is equipped with a 19 mm focal length and a viewing angle of 32 degrees.

3.2. Data Collection. Since open parking lots provide free parking spaces, the affordable system must be utilized to collect data. Few such affordable systems comprise GPS and visual camera which can be utilized to track moving objects. The usability of visual camera is limited due to harsh environmental and illumination conditions [35]. Persons or vehicles that can be identified using visual camera is not suitable to utilize due to privacy restrictions [36]. A GPS device needs to be carried by the driver to track the movement of the vehicle and the obtained data can be limited by the number of users or volunteers. The accuracy of the GPS device might also vary due to the position of satellite, weather conditions, or tall structures obstructing the signal [37]. Another affordable solution to overcome accuracy and privacy problems is the usage of thermal camera. A thermal camera identifies objects by emitted heat, and they are represented by pseudocolors. However, it can be deployed in any environmental and illumination conditions. It also avoids privacy-based restrictions as vehicles and pedestrians cannot be recognized [36]. Therefore, in this study, thermal camera is installed in an open parking lot to capture movement of vehicles. A thermal camera is used to

collect videos of the parking lot with varying weather and illumination conditions between January and August 2020. Videos representing varying illumination conditions, weather conditions, obstacles, and pedestrians were used for training and test purposes. As illustrated in Figure 2, the vehicle movement sequence is labeled to create training and test dataset.

Videos were converted to individual frames and objects of interest were labeled manually and partially with point tracker automation algorithm. The training dataset consists of 3000 images while the test dataset consists of 600 images. The training dataset is created using videos from different days and the image augmentation is invoked. Videos representing varying illumination conditions and traffic flows were included in the training dataset. Traffic flow was high during lunch and evening times, and varying illumination conditions were captured during winter, summer, and spring seasons. To improve detection in the parking lot, the training dataset includes noisier and winter condition images compared to other illumination conditions. Winter condition images represent dark conditions with varying vehicle heat signatures while noisy images consist of multiple vehicles, pedestrians, and occluding objects. Similarly, the test dataset is created using videos from different days representing varying illumination conditions and traffic flows which are not included in the training dataset. Training and test datasets are created using different videos without any overlap to avoid overfitting. Augmented images as illustrated in Figure 3 were used during training to improve the performance of the algorithm and avoid overfitting. Brightness, hue, contrast, and rotation were randomly invoked to create the augmented images.

Sample images from the test dataset are illustrated in Figure 4. Vehicles that are moving in the path of parking lot are the only ones detected and tracked. Parked vehicles are not detected or tracked as illustrated in Figure 4. Pedestrians and other objects like bicycles, strollers, and baskets are also visible in the test and train datasets as they are commonly visible in a parking lot. In Figure 4(b), the vehicle is occluded by a flag while pedestrians and strollers are visible in Figure 4(c). Groups of pedestrians with stroller like objects can easily be misidentified as vehicle. The vehicles in the parking lot have varying heat signatures as shown in Figures 4(a) and 4(d), where some are bright while others are mildly bright. The brightness or heat signature of the vehicle is mostly associated with the idleness. If the vehicle is parked for long duration such as 6 hours, the heat of the vehicle dissipates, and it appears dark in a thermal camera. In this case, the vehicle is not parked more than 3 hours. Therefore, the brightness of the vehicle is maintained. However, deep learning algorithms can detect vehicles even when the vehicle appears dark [11].

3.3. Object Detectors and Tracking Estimators. This section discusses the proposed algorithms used to detect and track vehicles. The objects should be located, in order to be tracked. So, the initial step was to estimate the locations of cruising vehicles which is also referred to as localization

FIGURE 1: The parking lot on which parking behavior is observed: (a) area of shopping center parking lot (courtesy of Google Maps) and (b) region of interest in which vehicles are detected and tracked.

FIGURE 2: Major steps in the data collection and processing stages to create final datasets.

FIGURE 3: Augmented images: (a) brightness and rotation modified, (b) hue and contrast modified, (c) saturation, brightness, and rotation modified, and (d) saturation and rotation modified.

FIGURE 4: Sample images of test dataset: (a) dark conditions with multiple vehicle orientations, (b) dark conditions with occluded vehicle, (c) dark conditions with groups of pedestrians, and (d) dark conditions with multiple vehicle positions and orientations.

problem [38, 39]. The next step was to generate trajectories by located object at regular intervals of time which is referred to as tracking. Since tracking by detection is implemented in this paper, detection of vehicles was performed using the proposed deep learning algorithms while the second step was to track the detected vehicles using Kalman filters.

Yolo is a logistic regression-based object detection algorithm which uses single convolutional neural network. It is one of the deep learning algorithms with higher accuracy and computational efficiency. Different versions of Yolo were developed over the years and YoloV3 is used in this paper which consists of residual, convolutional, and upsample layers which facilitates extracting complex abstract features of objects. Convolutional layer extracts feature of objects and assigns values. The residual layer applies threshold operation to each input, while upsample layer increases the input by replicating neighboring pixels. There are other deep networks like Resnet50, Resnet101, and Inceptionv3 with higher accuracy. However, these networks are computationally expensive compared to Yolo. Similarly, other computationally efficient algorithms like Resnet18 and GoogLeNet are not at par with YoloV3 based on initial evaluations. Despite being a deep network, YoloV3 still manages to be computationally efficient. Therefore, YoloV3 is chosen as the detector in this paper. Since the algorithms are deployed on videos or used for near real-time detection, it is imperative to utilize fast and accurate algorithms like YoloV3. YoloV3 used in this study is based on the original architecture. Modified Yolo has similar architecture to YoloV3. However, the anchor boxes' size and number were updated. Modified Yolo is also referred to as mod Yolo in this paper. Custom Yolo uses the size and number of anchor boxes like mod Yolo. However, the architecture is updated

by adding leaky rectified linear unit (ReLU) and batch normalization layers. Kalman filter is utilized to track the detected vehicle as it does not require historical data and is computationally efficient.

3.4. YoloV3. Yolo class object detectors are widely used detection networks due to their speed and accuracy [40]. YoloV3 is updated using YoloV2 and Darknet-19 networks. YoloV2 had the problem of detecting small objects and it does not contain shortcut connections. However, YoloV3 contains upsample layer, concatenation layer, and shortcut connections and contains more layers compared to YoloV2 [32]. It divides the image into grid cells which facilitate detecting objects. It uses K-means clustering to estimate the bounding boxes. Confidence scores are determined as illustrated in

$$C = P \times IoU_{pr}, \quad (1)$$

where C is the confidence score in the interval $[0, 1]$. P is assigned 1 if the object is present in the grid or it is assigned 0. IoU_{pr} assigns a score based on the overlap ratio between predicted and ground truth box. The confidence score determines the accuracy of the predicted bounding box. Therefore, the results presented in this paper display the confidence scores to illustrate the accuracy of prediction.

3.5. Modified Yolo. In this study, multivehicle detection is performed where vehicles vary in their size based on the position in the frame. Multiple vehicles move at multiple positions in the ROI. Therefore, increasing the anchor boxes would facilitate detecting multiple vehicles at different

locations. Since the target objects are smaller in size, the number of anchor boxes utilized is increased to 8 from 6. The sizes of anchor boxes are estimated using the training data. The architecture utilized in mod Yolo is similar to Yolov3.

3.6. Custom Yolo. Custom Yolo is created using modified Yolov3. The layers were updated with batch normalization and leaky ReLU layers. Batch normalization is added to enable more stable predictive behavior and faster learning. Leaky residual layer is added in the network to avoid vanishing gradient problems and better adjustment of weights [41]. The leaky ReLU function has a nonzero gradient unlike the standard ReLU function [42].

$$f(x) = \alpha y + y \begin{cases} y, & \text{if } y > 0, \\ \alpha y, & \text{if } y \leq 0. \end{cases} \quad (2)$$

The α parameter is introduced to have nonzero gradients as illustrated in (2).

3.7. Kalman Filter. Kalman filter is a motion estimation algorithm which facilitates prediction and tracking of objects [43]. It is used to track objects and predict the movement of objects when they are not visible or occluded. It facilitates single and multiobject tracking. The constant velocity motion model is selected for Kalman filter. As mentioned in [39], the target motion model and measurement model for the Kalman filters are given in the following equations:

$$X_t = AX_{t-1} + BU_t + w_t, \quad (3)$$

where X_t is the predicted state at time t , while w_t is the error estimated. A is the matrix that updates X from the previous step. B is another matrix that is used to update acceleration U_t .

$$z_t = H(X_t) + v_t, \quad (4)$$

where z_t is the vector of measurements like position. H is a transformation matrix. v_t is the measure noise.

However, in this study, tracking of vehicle based on prediction is avoided as the vehicle does not have constant velocity all the time. A vehicle in the parking lot can stop for other vehicles and pedestrians or pause before they park. In this paper, centroid positions of the bounding boxes are generated from utilized object detectors and they are assigned ID and age based on the visibility of the object. Assigning ID for individual vehicles or objects is referred to as the process of data association. Since missing detection is common with object detection algorithms, the parameters in Kalman filter are modified to facilitate tracking of vehicle even with missed detection. Tracking is maintained even when the vehicle is invisible for 50 frames. However, this can be reduced when the performance of detection is improved.

3.8. Vehicle Detection. Yolov3, modified Yolo, and custom Yolo are utilized to detect moving vehicles in the parking lot. As illustrated in Figure 5, the proposed algorithms are trained using the training dataset which includes augmented

FIGURE 5: Vehicle detection process.

images. The learning rate of object detectors is 0.001 and the number of epochs used for training is 40. The training time for each object detector is approximately 6 hours. The trained detectors are then evaluated using the test dataset. The use of augmented images in training datasets improves the generalization of the algorithms. A workstation with i7 processor and Nvidia Quadro P5200 GPU was used in this study. MATLAB platform was used to train, detect, and track the cruising vehicles. Overlap threshold of 0.9 and confidence threshold of 0.7 were used for detection purposes. The overlap threshold was increased to avoid duplicate detection and the confidence threshold was given 0.7 to reduce false positives with low confidence scores.

Evaluation of algorithms is represented using precision-recall curve and log average miss rate curve. Precision is also referred to as positive predictive value and is the fraction of relevant retrieved labeled detection.

$$\text{Precision} = \frac{\text{true positive}}{(\text{true positive} + \text{false positive})}. \quad (5)$$

Recall is referred to as sensitivity and is the fraction of items positively detected among all the items [44].

$$\text{Recall} = \frac{\text{true positive}}{(\text{true positive} + \text{false negative})}. \quad (6)$$

Average precision is defined as the area under the interpolated precision-recall curve where r_n is the recall levels and P_{intp} is interpolated precision.

$$\text{Average precision} = \sum_{i=1}^{n-1} (r_{i+1} - r_i) P_{\text{intp}}(r_{i+1}). \quad (7)$$

The miss rate curve illustrates the amount of overall false detection. Log average miss rate is computed by averaging miss rate at nine false positive per image (FPPI) reference points which are equally spaced in log space [45].

3.9. Vehicle Tracking Using Detection. Tracking of vehicle movement is performed using detected vehicles obtained from object detectors. The process of vehicle tracking is illustrated in Figure 6. Videos were collected using thermal camera and the videos used to calculate trajectory are not part of the training dataset. Videos were divided into frames

FIGURE 6: Vehicle tracking process.

and the proposed algorithms were utilized for detecting vehicles in each frame. The detected vehicles are then given an ID using Kalman filter which is also referred to as data association. An ID is a sequential number assigned to the detected vehicle. The tracked vehicle is given an individual color based on ID and calculated on the image. IDs are assigned to the detected vehicles and the ID is removed when the vehicle is not available in the frame anymore. Single and multivehicle tracking is performed using the Kalman filter. For convenience of understanding, the trajectories are calculated on a video frame obtained using thermal camera. Trajectories can be used to identify the movement of vehicles in the region of interest.

4. Results and Analysis

The evaluation of the detector is illustrated using precision-recall and log average miss rate curves. The average precision, log average miss rate values, and computation time are presented in Table 1. The table presents the average precision and miss rate values of the detectors achieved on the complete test dataset. Average precision and miss rate are scaled between 0 and 1 where 0 is the lowest and 1 is the highest.

As shown in Table 1, mod Yolo performed with higher precision and less miss rate compared to the other algorithms. Description of evaluation metrics can be found in Section 3.4. The miss rate of Yolov3 is 0.38 and the average precision is 0.75. However, when the number of anchor boxes and sizes are updated, there is an improvement in average precision by 0.16 and the miss rates were reduced by 0.21. Custom Yolo achieved an average precision of 0.83, while miss rates were 0.26. Even though custom Yolo performed better than Yolov3, mod Yolo performed better than custom Yolo. The processing times of all the three detectors were optimal. However, custom Yolo was faster by 0.01 seconds due to the inclusion of batch normalization layers.

4.1. Evaluation of Multivehicle Detection. The precision-recall curve is illustrated in Figure 7(a), and the overall performance of mod Yolo was better compared to other algorithms. The positive detection rate or precision of Yolov3 was the lowest, while custom Yolo performed slightly better than mod Yolo. However, mod Yolo recall rate was better than custom Yolo leading to better overall performance.

As illustrated in Figure 7(b), the modified Yolo has the lowest false positives and miss rates compared to other algorithms. The log average miss rate of Yolov3 was again the least performer with a high number of miss rates and false positives per image. Custom Yolo and mod Yolo false positives per image had similar performance. However, mod Yolo had less miss rate compared to custom Yolo leading to better performance.

Detection of Yolov3, mod Yolo, and custom Yolo from test dataset is illustrated in Figure 8 with their respective confidence scores. As visible in Figure 8(a), the detection of all the three detectors performed similarly with higher confidence scores. Mod Yolo and Yolov3 detected vehicle with a 0.99 confidence score despite the presence of pedestrians. In Figure 8(b), Yolov3 detected all moving vehicles while mod Yolo had better confidence scores compared to Yolov3 and custom Yolo. Custom Yolo and mod Yolo have a missed detection of a moving vehicle. The moving vehicle beside the tree was also successfully detected despite the occlusion by all the detectors. However, Yolov3 and custom Yolo have duplicate detection with varying confidence score values in Figure 8(b). In Figure 8(c), the occluded vehicle is not detected in all the three detectors. One vehicle was detected, and the occluded vehicle was missed. The detected vehicle in mod Yolo had nearly a 0.1 confidence score while custom Yolo detected with a 0.99 confidence score. Even though the size of the vehicle is small, the confidence score is maintained high with mod Yolo and custom Yolo.

4.2. Trajectory of Vehicles Using Data Association. Data association of detected vehicles is performed using Kalman filter. As illustrated in Figure 9, object detector and Kalman filter are used to calculate trajectories of multiple vehicles. The actual vehicle trajectories calculated using gTruth are also illustrated for reference in Figure 9. The gTruth is the annotated set of images used to evaluate performance. Kalman filter was able to maintain data association even during missed detection and was also able to track stationary vehicle located in the path of the parking lot as illustrated in Figure 9(a). Custom Yolo and mod Yolo tracked the vehicle accurately during its movement. However, Yolov3 did not maintain continuous detection and Kalman filter produced errors in data association. The object detector should be continuous in its detection to achieve better tracking results. Yolov3 and custom Yolo produced errors shown in Figure 9(b). However, mod Yolo tracked the exiting vehicle and vehicle which was parked. Custom Yolo falsely detected stationary vehicle as a moving vehicle. Yolov3 produced false positives over parked vehicles. In Figure 9(c), mod Yolo detects the exiting vehicle along with moving vehicles in the

TABLE 1: Object detectors performance.

Detector	Processing time second/frame	Average precision	Log average miss rate
Yolov3	0.02	0.75	0.38
Custom Yolo	0.01	0.83	0.26
Mod Yolo	0.02	0.91	0.17

FIGURE 7: Evaluation of proposed algorithms. (a) Precision-recall curve and (b) log average miss rate curve.

FIGURE 8: Confidence scores of three detectors on sample test images at different times with varying illuminations: (a) image with groups of pedestrians; (b) multiple moving vehicles in dark conditions; (c) occluded vehicle and pedestrians in bright conditions.

FIGURE 9: Vehicle tracking using object detector and Kalman filter varying illumination conditions: (a) multiple moving and stationary vehicle; (b) vehicles exiting and parking; (c) occluded moving vehicle.

TABLE 2: Average precision and miss rates of Figure 9 where A is the average precision and M is the miss rate.

Algorithm	Figure 9(a)	Figure 9(b)	Figure 9(c)	Processing time second/frame
Yolov3	$A = 0.73, M = 0.44$	$A = 0.70, M = 0.42$	$A = 0.52, M = 0.54$	0.16
Custom Yolo	$A = 0.73, M = 0.35$	$A = 0.78, M = 0.22$	$A = 0.81, M = 0.19$	0.1
Mod Yolo	$A = 0.84, M = 0.22$	$A = 0.88, M = 0.18$	$A = 0.82, M = 0.22$	0.17

parking lot. Custom Yolo missed the exiting vehicle, while Yolov3 produced false positives by detecting stationary vehicles, and the detection of moving vehicles was incomplete. The data association performed poorly when there were false positives which can be observed in Figure 9(a) and Figure 9(b) of Yolov3. Therefore, data association was better when the detection results were accurate.

Table 2 describes the average precision and log average miss rate values of the detection performed in Figure 9. Mod

Yolo has better average precision and log average miss rate values compared to other algorithms. Yolov3 has higher missed detection and the average precision is the lowest for Figure 9(c). Custom Yolo has a higher miss rate in Figure 9(a) while Figure 9(c) has the lowest miss rate. Mod Yolo has higher average precision values and the miss rate is considerably lower. The processing time of one frame with a combination of detection algorithms and Kalman filter is optimal as mentioned in Table 2.

5. Concluding Discussion

The usage of deep learning algorithms and Kalman filter performed with good accuracy in detection and tracking of the movement of multiple vehicles in an open parking lot with varying environmental and illumination conditions. A combination of thermal camera and deep learning algorithms enabled the detection of objects in varying illumination and environmental conditions. The mod Yolo algorithm performed better compared to YOLOv3 and customized Yolo. Modifying the size and number of anchor boxes improved the performance of the mod Yolo. Deep learning object detectors are originally trained on color images. However, they performed with high accuracy even with thermal camera data after making modifications such as updating anchor boxes and adding new layers. The tracking performance is dependent on the accuracy of object detection. Kalman filter maintained data association during missed detection and moving and stationary vehicles. The distinct color was assigned to each vehicle based on the assigned ID and calculated on the image.

The performance of mod Yolo was high even though the training dataset is rather small considering the complex setup with varying illumination conditions, noise, multiple orientations, and size of vehicles. The inclusion of augmented images improved the performance of algorithms. In order to maintain data association during missed detection, an assigned ID was maintained for 50 frames. Missed detection mostly occurs during occlusions and different orientations which can be prevented by additional training images. However, with additional training images, the time required to train the detector will also be increased. The computation time to run the detector might also be increased if the architecture is too deep. A balance between the training images and the number of layers should be achieved for higher computational efficiency. Mod Yolo performed better when the anchor boxes are increased. However, the use of anchor boxes would also increase the computation capacity. Therefore, adding too many anchor boxes would also impact the computational efficiency of the algorithm.

The algorithms were trained to detect and track vehicles moving on the parking lot. This enables the use of deep learning algorithms to detect and track specific objects, tasks, or actions. In the future, additional algorithms such as LSTM and recurrent neural networks would be utilized for tracking the vehicle. This study also proves that trajectories of moving vehicles can be utilized to study parking behavior. Parking behavior would facilitate understanding the choice of parking space and the amount of cruising undertaken to reach that space. Furthermore, the detection and tracking performance is fast, so this methodology could be used in near real-time parking assisting systems.

Data Availability

The data are collected using a thermal camera at an open parking lot in Sweden.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

- [1] N. Geroliminis and C. F. Daganzo, "Existence of urban-scale macroscopic fundamental diagrams: some experimental findings," *Transportation Research Part B: Methodological*, vol. 42, no. 9, pp. 759–770, 2008.
- [2] N. Geroliminis and J. Sun, "Properties of a well-defined macroscopic fundamental diagram for urban traffic," *Transportation Research Part B: Methodological*, vol. 45, no. 3, pp. 605–617, 2011.
- [3] S. Gallivan, *IBM Global Parking Survey: Drivers Share Worldwide Parking Woes Technical Report*, IBM, Endicott, NY, USA, 2011.
- [4] P. Van Der Waerden, H. Timmermans, and L. Van Hove, "GPS data and car drivers' parking search behavior in the city of turnhout, Belgium," in *Geoinformatics for Intelligent Transportation*, pp. 247–256, Springer, Berlin, Germany, 2015.
- [5] R. C. Hampshire and D. Shoup, "What share of traffic is cruising for parking?" *Journal of Transport Economics and Policy*, vol. 52, no. 3, pp. 184–201, 2018.
- [6] V. Paidi, H. Fleyeh, J. Håkansson, and R. G. Nyberg, "Smart parking sensors, technologies and applications for open parking lots: a review," *IET Intelligent Transport Systems*, vol. 12, no. 8, pp. 735–741, 2018.
- [7] A. Doulamis, "Dynamic tracking re-adjustment: a method for automatic tracking recovery in complex visual environments," *Multimedia Tools and Applications*, vol. 50, no. 1, pp. 49–73, 2010.
- [8] D. Coppi, S. Calderara, and R. Cucchiara, "Transductive people tracking in unconstrained surveillance," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 26, no. 4, pp. 762–775, 2015.
- [9] W. Luo, X. Junliang, and M. Anton, Multiple object tracking: a literature review," *Artificial Intelligence*, vol. 1, Article ID 103448, 2020.
- [10] J. Lu, "Multi-manifold deep metric learning for image set classification," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Boston, MA, USA, June 2015.
- [11] V. Paidi, H. Fleyeh, and R. G. Nyberg, "Deep learning-based vehicle occupancy detection in an open parking lot using thermal camera," *IET Intelligent Transport Systems*, vol. 14, no. 10, pp. 1295–1302, 2020.
- [12] Q. Li, "Kalman filter and its application," in *Proceedings of the 2015 8th International Conference on Intelligent Networks and Intelligent Systems (ICINIS)*, November 2015.
- [13] C. L. Azevedo, J. L. Cardoso, and M. Ben-Akiva, "Vehicle tracking using the k-shortest paths algorithm and dual graphs," *Transportation Research Procedia*, vol. 1, no. 1, pp. 3–11, 2014.
- [14] Z. Kim, "Real time object tracking based on dynamic feature grouping with background subtraction," in *Proceedings of the 2008 IEEE Conference on Computer Vision and Pattern Recognition*, June 2008.
- [15] K. Mu, F. Hui, X. Zhao, and C. Prehofer, "Multiscale edge fusion for vehicle detection based on difference of Gaussian," *Optik*, vol. 127, no. 11, pp. 4794–4798, 2016.
- [16] K. Kale, S. Pawar, and P. Dhulekar, "Moving object tracking using optical flow and motion vector estimation," in *Proceedings of the 2015 4th International Conference on*

- Reliability, Infocom Technologies and Optimization (ICRITO)(trends and Future Directions)*, September 2015.
- [17] Q. Wang, J. Wan, and Y. Yuan, "Locality constraint distance metric learning for traffic congestion detection," *Pattern Recognition*, vol. 75, pp. 272–281, 2018.
 - [18] L.-W. Tsai, J.-W. Hsieh, and K.-C. Fan, "Vehicle detection using normalized color and edge map," *IEEE Transactions on Image Processing*, vol. 16, no. 3, pp. 850–864, 2007.
 - [19] Z. Wu, "Coupling detection and data association for multiple object tracking," in *Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition*, June 2012.
 - [20] S. S. Mane and S. A. K. Shah, "Object tracking and analysis for detecting various situations at level crossings of railway-road," in *Proceedings of the 2017 International Conference on Computing, Communication, Control and Automation (ICCUBEA)*, August 2017.
 - [21] Y. Chong, W. Chen, Z. Li, W. H. K. Lam, C. Zheng, and Q. Li, "Integrated real-time vision-based preceding vehicle detection in urban roads," *Neurocomputing*, vol. 116, pp. 144–149, 2013.
 - [22] X. Zhang et al., "Real-time vehicle detection and tracking using improved histogram of gradient features and Kalman filters," *International Journal of Advanced Robotic Systems*, vol. 15, no. 1, 2018.
 - [23] K. Robert, "Night-time traffic surveillance: a robust framework for multi-vehicle detection, classification and tracking," in *Advanced Video and Signal Based Surveillance* IEEE, Piscataway, NJ, USA, 2009.
 - [24] A. Gaszczak, T. P. Breckon, and J. Han, "Real-time people and vehicle detection from UAV imagery," in *Intelligent Robots and Computer Vision XXVIII: Algorithms and Techniques* International Society for Optics and Photonics, Bellingham, WA, USA, 2011.
 - [25] X. Xie, "Real-time illegal parking detection system based on deep learning," in *Proceedings of the 2017 International Conference on Deep Learning Technologies*, Chengdu, China, June 2017.
 - [26] H. Guan et al., "Real-time lane-vehicle detection and tracking system," in *Proceedings of the 2016 Chinese Control and Decision Conference (CCDC)*, May 2016.
 - [27] J.-G. Wang, "Appearance-based brake-lights recognition using deep learning and vehicle detection," in *Proceedings of the 2016 IEEE Intelligent Vehicles Symposium (IV)*, June 2016.
 - [28] Y. Iwasaki, M. Misumi, and T. Nakamiya, "Robust vehicle detection under various environmental conditions using an infrared thermal camera and its application to road traffic flow monitoring," *Sensors*, vol. 13, no. 6, pp. 7756–7773, 2013.
 - [29] C.-W. Chang, "Vehicle detection in thermal images using deep neural network," in *Proceedings of the 2018 IEEE Visual Communications and Image Processing (VCIP)*, December 2018.
 - [30] P. Ren, W. Fang, and S. Djahel, "A novel YOLO-Based real-time people counting approach," in *Proceedings of the 2017 International Smart Cities Conference (ISC2)*, September 2017.
 - [31] M.-X. Jiang, D. Chao, and Z.-G. Pan, "Multiobject tracking in videos based on lstm and deep reinforcement learning," *Complexity*, vol. 2018, Article ID 4695890, 12 pages, 2018.
 - [32] J. Redmon and A. Farhadi, "Yolov3: an incremental improvement," 2018, <https://arxiv.org/abs/1804.02767>.
 - [33] A. Sadeghian, A. Alahi, and S. Savarese, "Tracking the untrackable: learning to track multiple cues with long-term dependencies," in *Proceedings of the IEEE International Conference on Computer Vision*, Venice, Italy, October 2017.
 - [34] Q. Fan, S. Pankanti, and L. Brown, "Long-term object tracking for parked vehicle detection," in *Proceedings of the 2014 11th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, August 2014.
 - [35] H. Ichihashi, "Vacant parking space detector for outdoor parking lot by using surveillance camera and FCM classifier," in *Fuzzy Systems* IEEE, Piscataway, NJ, USA, 2009.
 - [36] M. N. Asghar, N. Kanwal, B. Lee, M. Fleury, M. Herbst, and Y. Qiao, "Visual surveillance within the EU general data protection regulation: a technology perspective," *IEEE Access*, vol. 7, pp. 111709–111726, 2019.
 - [37] U. Dogan, M. Uludag, and D. O. Demir, "Investigation of GPS positioning accuracy during the seasonal variation," *Measurement*, vol. 53, pp. 91–100, 2014.
 - [38] S. R. Jondhale and R. S. Deshpande, "Kalman filtering framework-based real time target tracking in wireless sensor networks using generalized regression neural networks," *IEEE Sensors Journal*, vol. 19, no. 1, pp. 224–233, 2018.
 - [39] S. R. Jondhale and R. S. Deshpande, "GRNN and KF framework based real time target tracking using PSOC BLE and smartphone," *Ad Hoc Networks*, vol. 84, pp. 19–28, 2019.
 - [40] L. Zhao and S. Li, "Object detection algorithm based on improved YOLOv3," *Electronics*, vol. 9, no. 3, 537 pages, 2020.
 - [41] S. Mastromichalakis, "ALReLU: a different approach on leaky RELU activation function to improve neural networks performance," 2020, <https://arxiv.org/abs/2012.07564>.
 - [42] A. L. Maas, A. Y. Hannun, and A. Y. Ng, *Rectifier Nonlinearities Improve Neural Network Acoustic Models*, Citeseer, Piscataway, NJ, USA, 2013.
 - [43] X. Li, "A multiple object tracking method using Kalman filter," in *Proceedings of the 2010 IEEE International Conference on Information and Automation*, June 2010.
 - [44] A. Godil, *Performance Metrics for Evaluating Object and Human Detection and Tracking Systems. No. NIST Interagency/Internal Report (NISTIR)-7972*, NIST, Gaithersburg, MA, USA, 2014.
 - [45] P. Dollar, "Pedestrian detection: an evaluation of the state of the art," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 4, pp. 743–761, 2011.