Deep learning-based vehicle occupancy detection in an open parking lot using thermal camera

Vijay Paidi[1] ✉, Hasan Fleyeh[1], Roger G. Nyberg[1]

[1]*Dalarna University, Faculty of Data & Information Sciences, Borlange, Sweden*
✉ *E-mail: vpp@du.se*

Abstract: Parking has been a common problem over several years in many cities around the globe. The search for parking space leads to congestion, frustration and increased air pollution. Information of vacant parking space would facilitate to reduce congestion and subsequent air pollution. Therefore, the aim of the study is to acquire vehicle occupancy in an open parking lot using deep learning. Thermal camera was used to collect videos during varying environmental conditions and frames from these videos were extracted to prepare the dataset. The frames in the dataset were manually labelled as there were no pre-labelled thermal images available. Vehicle detection with deep learning algorithms was implemented to perform multi-object detection. Multiple deep learning networks such as Yolo, Yolo-conv, GoogleNet, ReNet18 and ResNet50 with varying layers and architectures were evaluated on vehicle detection. ResNet18 performed better than other detectors which had an average precision of 96.16 and log-average miss rate of 19.40. The detected results were compared with a template of parking spaces to identify vehicle occupancy information. Yolo, Yolo-conv, GoogleNet and ResNet18 are computationally efficient detectors which took less processing time and are suitable for real-time detection while Resnet50 was computationally expensive.

1 Introduction

Parking can be a daunting task due to the availability of a limited number of parking spaces compared to the number of vehicles. According to previous studies, it can take up to 14 min to find a parking space [1, 2] which can lead to congestion, increased air pollution and user frustration. Therefore, parking management is an integral component for city planning administrators, and this is one of the research themes in smart city development. Innovative parking has been an important research area as it enables accessibility to commuters and is capable of enhancing business opportunities [3]. The parking problem applies to both major and minor cities due to higher demand and limited resources, though there can be a difference in demand between different places. Lack of parking spaces at places of interest can lead to loss of business opportunities. Therefore, business organisations spend higher expenditure to acquire the sufficient number of parking spaces which cover large areas of scarce land resources. Due to space restrictions, in urban areas, new parking areas are being built in multi-storey buildings or basements. To avoid parking problems, public transport can be utilised. However, using public transport all the time might not be a convenient option for everyone. The other way is to use smart parking applications to efficiently utilise parking spaces and reduce congestion, user frustration and pollution.

There are several web and mobile smart parking applications available. Smart parking applications are available for closed parking lots while there are no applications available for open parking lots [3]. The demand of parking spaces in an open parking lot is higher compared to a closed parking lot. An open parking lot is subjected to different environmental conditions such as snow, rain, darkness and sunny. The expenditure made on a parking guidance system or smart parking application for open parking lots cannot be retained directly as the parking spaces are available freely. There are several tools used to identify parking occupancy such as infrared and ultrasonic sensors, magnetometers, vehicle *ad hoc* networks (VANET) and microwave radars [4, 5]. However, sensors, VANET, microwave radars and magnetometers need higher expenditure in installation and maintenance activities. Therefore, computer vision is a suitable technology to identify parking occupancy which can cover a large number of parking

spaces with a single camera. The number of parking spaces covered by a camera is dependent on the focal length, height and angle of the positioned camera. Since computer vision is one of the feasible smart parking tools for open parking lots, a thermal camera is used to capture data in this paper. A thermal camera is expensive compared to a visual colour camera, but it facilitates in detecting objects in any weather and light conditions and it also has less privacy restrictions. Object detection was implemented to identify vehicles which comprise of image classification and object localisation phases. Multi-object detection was performed since the parking lot consists of multiple vehicles. Emphasis was given to deep learning architectures as it is one of the intelligent contemporary methodologies which do not need custom modifications or morphological operations to detect and classify an object. Since there are already well-defined deep learning architectures, some of them were implemented and evaluated on thermal data in this paper.

There are several smart parking solutions available as mentioned earlier. However, there is no smart parking tool which is suitable for all economic, geographic and environmental conditions. Therefore, the large-scale utilisation of smart parking solutions is still not available. The current available smart parking applications are limited to few parking lots in several countries and none of these applications provide parking occupancy information for an open parking lot. Therefore, this paper aims to identify vehicle occupancy information in an open parking lot with deep learning and thermal camera. The first step is to identify vehicles using object detection algorithms. The second step is to compare the vehicle detection results with a template of parking spaces and acquire vehicle occupancy information.

A thermal camera is commonly used for security applications or to identify overheated equipment [6, 7]. The use of a thermal camera for identifying parking occupancy was not widely discussed in previous literature. According to a study, vehicle detection was performed on moving vehicles using a thermal camera and deep learning [8]. A moving vehicle contains heat at tyres, windshield, engine or lights. Another study used thermal camera and deep learning to detect idle vehicles as longer durations of idleness is prohibited due to increased environmental pollution [9]. The temperature output from the engine and exhaust was used to detect the vehicles. However, these features are dependent on
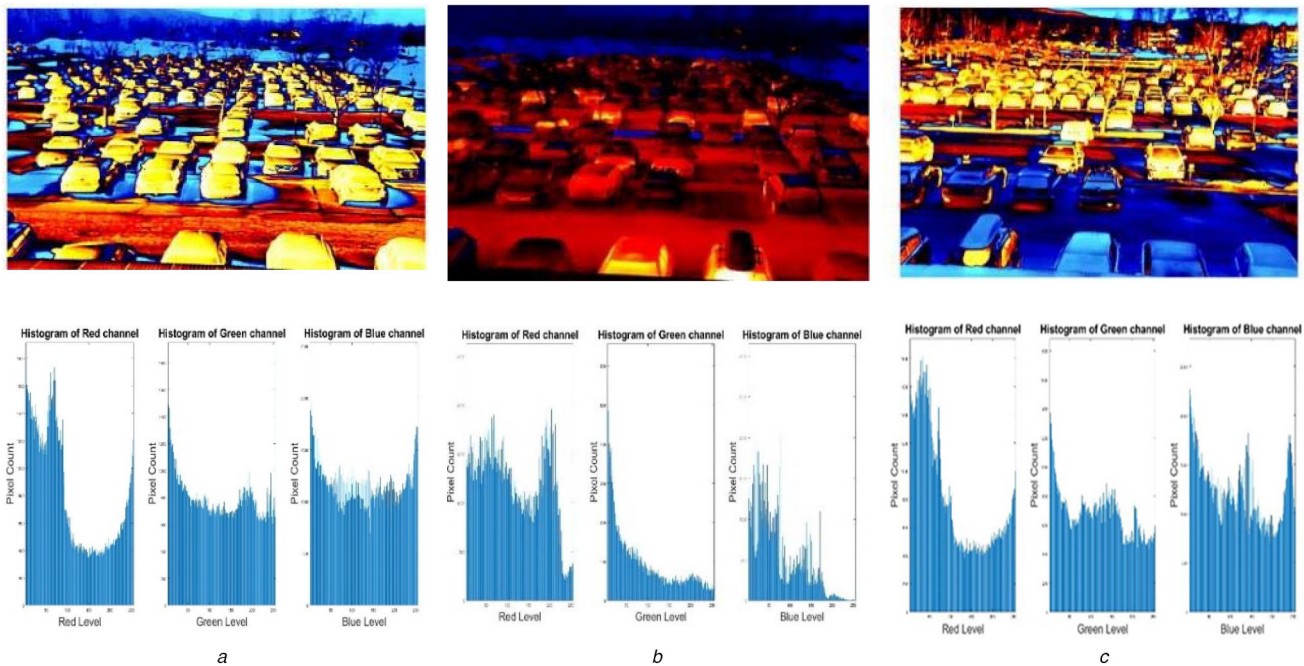
Fig. 1 *Diverse images in the dataset with histogram*
(a) Sunny in winter morning, *(b)* Dark snowy winter morning, *(c)* Dusk or dawn

time in a parking lot as the vehicle cools down completely. A commercial company utilises a combination of visual and thermal cameras to identify parking occupancy [10]. There are several studies utilising the visual camera for vehicle identification. Thermal cameras were largely used on images or videos collected during dusk or night conditions [8]. A combination of visual spectrum images for daylight and thermal images for night conditions were utilised to enable detection in varying illumination conditions [11]. There is scarce of the literature identifying vehicles only using the thermal camera in varying illumination and environmental conditions [12].

The contribution of the paper is to address this research gap by utilising a thermal camera to identify vehicles and generate parking occupancy in varying illumination and environmental conditions. Unlike a visual camera, features of an object change based on temperature using a thermal camera. A vehicle can be parked for a few hours, during which the vehicle can cool down leading to diminished features of the vehicle. However, a smart parking system should be able to detect parking occupancy continuously irrespective of heat emitted from a vehicle. Therefore, a dataset was created with images representing varying illumination and environmental conditions as shown in Fig. 1. Deep learning algorithms were utilised to identify parking vehicles as it is capable of learning complex features. A custom deep learning algorithm based on Yolo, depthwise and residual layers was developed to identify vehicles. Multiple deep learning object detection algorithms such as Yolo, Yolo-conv, Googlenet, ResNet18 and ResNet50 architectures' performances were evaluated and compared. Empty parking space features are hard to distinguish from the background and the lines of the parking space are also not visible in most conditions. Thus, a template is used to generate parking occupancy and this process is discussed more in Section 3.

The remaining sections of the paper are organised in the following way. Section 2 discusses the relevant literature while Section 3 discusses about the dataset and proposed method. Section 4 presents the results obtained in this paper along with analysis. Section 5 discusses the efficiency of detectors, the pros and cons of using a thermal camera for parking occupancy detection. Finally, the paper is ended with the conclusion in Section 6.

2 Literature review

Extensive research on the use of various algorithms or detectors for identifying vehicles or other objects is already available. Object detection capability kept evolving over the years and new detectors

or architectures were generated. Since vehicle occupancy detection was performed using a camera, relevant algorithms or detectors are discussed in this section. Edge detectors such as Canny and Sobel can be used to identify vehicles [13, 14]. Edge detectors perform efficiently when the vehicles can be recognised. However, when using thermal cameras, due to loss of heat in the vehicle over a period, they can become dark and can be challenging to recognise or detect using edges. In such scenarios edge detectors might not be a suitable option for vehicle detection. Histograms of Oriented Gradient (HOG) descriptors and Viola Jones are efficient human or pedestrian detectors [15] which can also be used for vehicle detection. HOG detector trains using positive and negative images using a linear Support Vector Machine (SVM) classifier [16].

In another study, the Kalman filter was used to identify and classify night time traffic surveillance [6]. Headlight and visible vehicle features were used to detect vehicles. Similarly, in another study, [17] vehicles during night time were detected using blob properties which were classified by an SVM classifier and Kalman filters were invoked to track the detected vehicles. In a similar study, the infrared thermal camera was used to study traffic flow using Viola jones detector. The tyres and windshield of the vehicles were used to identify the vehicles [18]. Based on the chosen features, positive and negative images were used to train the detector. However, since the vehicles are moving on the road, heat can be captured by the windshield or by the tyres which might not be applicable to this paper, since the vehicles are stationary in a parking lot. In a similar study, an unmanned aerial vehicle with the help of a thermal camera is deployed for detection of people and cars [19]. However, since vehicles are moving on the road, heat can be captured by the windshield or tyres which might not be applicable to this paper, as the vehicles are stationary in a parking lot. Based on another study, the thermal camera was used to classify vehicles types using features extracted from the front end of the vehicle [11]. Classification of vehicle types using thermal camera performed poorly since a thermal camera shows features in pseudo colours based on generated heat. In another similar study, HOG and SVM classifiers were implemented to detect moving vehicles using thermal camera [20]. Vehicle category classification using a thermal camera was performed by heat distribution analysis in another study [21]. Majority of the studies utilising the thermal camera for object recognition or classification were dependent on the heat signature of the vehicle. However, relying on heat signature alone would not be efficient in-vehicle detection at an open parking lot. Detection of people and cars was performed using multiple cascaded Haar classifiers. Haar uses a set of weak

Fig. 2 *Area of parking spaces*

Fig. 3 *Count of diverse images*

classifiers to form a strong classifier. However, the position and angle of the vehicle affect the classification [17]. Haar also requires good edges and lines for better detection which is a challenge due to varying temperatures of vehicles in an open parking lot. Increased performance normally requires higher computational costs and to maintain accuracy with lower computational costs, a binary sliding window detector also called ACF detector was developed [22, 23]. An image is computed to multiple channels and then sum every block of pixels to generate lower-resolution channels. A multi-scale sliding window is employed and boosting is performed to identify the object in ACF. Multiple additional frame boxes or bounding boxes can be created where non-maximum suppression and threshold can be used to reduce the detected bounding boxes.

Neural networks are an evolving data processing system used for classification purposes. It is inspired by the human brain nervous system where extracted features are passed through layers to generate high dimensional information [24]. Deep learning is capable of handling complex object detections and is one of the popular methods available [25]. Deep residual learning framework improves the detection and accuracy of a convolutional neural network [26]. Pre-trained residual learning networks with varying layers is available which can be used to perform custom detection [27]. Similarly, there are several other frameworks with varying architectures and performances. Googlenet is one such deep learning architecture which consumes less computational costs [28]. Faster Regional Convolutional Neural Network is a deep learning neural network object detector where sliding window detector and region proposal networks were used [29]. This object detector can be combined with existing deep learning architectures to make use of well-developed architectures and improve accuracy. Yolo is another popular model used for classifications which are fast and computationally efficient. It uses a single convolutional neural network to detect the objects of interest. Yolo model was used in to count traffic [30] and people [31] with high accuracy.

Yolo is a computationally efficient and popular classification algorithm and is used as one of the detectors in this paper. Similarly, Googlenet and Resnet18 were also computationally efficient with higher recognition rates, and therefore used in this paper. However, there are not many studies discussing the performance of various deep learning algorithms in recognising vehicles at multiple environment and time durations using a thermal camera.

3 Data and method

This section describes the data collection, processing, implementation of algorithms and parking occupancy generation.

3.1 Data with thermal camera

The data for vehicle occupancy detection was captured using an Axis Q1942-E thermal camera. The thermal camera was installed on a two-storey building and was equipped with 19 mm focal length objective. The viewing angle is 32° and the detection range of a vehicle is 1757 m. Only the first four rows of the parking lot were selected to identify vehicle occupancy as shown in Fig. 1. The green rectangular region is the region of interest. The vehicles outside the region of interest are small and majorly occluded which was the reason for exclusion. The data was collected in different
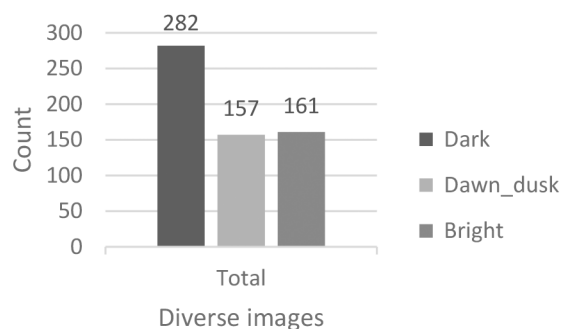
weather conditions such as snow, rain, dark and bright conditions. The installed thermal camera records videos based on motion detection which lead to a collection of several small interval videos in various weather conditions. The videos are accessed and stored on a local disk using an application. The size of the total collected videos is nearly 30 GB. Few frames from these videos were collected and stored which represents diverse conditions. Each image consists of various vehicles and the size of each image is 600×800 pixels.

In Scandinavia or other similar countries, daylight is shorter during winter and longer during summer. Snow can be seen for three to four months during winter. A visual colour camera can face issues with low light levels and snow conditions which are common in Scandinavian or other similar countries. Based on the privacy policy guidelines in Sweden, it is restricted to use video surveillance in open public areas where individuals can be recognised or identified [32]. Therefore, the use of a thermal camera had facilitated to avoid these restrictions where no individuals can be identified or recognised. In this way, 600 frames were collected from several videos representing varying environmental conditions.

The test parking site used in this study is in Sweden where the vehicles arrive between 7:00 and 8:00 in the morning and leave the parking space around 16:00–17:00. There are heaters located in the parking lot which would be used before leaving. The use of heaters generate heat in the vehicles which can be seen in Fig. 1b where one vehicle can be seen bright even in dark evening conditions. In the morning, vehicles are warm and can be easily identified during winter or sunny conditions. Fig. 2 illustrates the diverse images collected from the thermal camera. Fig. 1a illustrates the vehicles in bright conditions, Fig. 1b illustrates vehicles in dark conditions while Fig. 1c illustrates vehicles in dawn and dusk conditions which are subjected to shadowy regions. The histograms of the three images illustrate the variance of pixels in RGB channels during each condition. In bright conditions, RGB pixels were mostly similar. In dark conditions, Red pixels are in higher proportion while in dawn or dusk conditions, Blue and Green pixels are in higher proportion. During dark conditions, heat in the vehicles reduce over a period and they appear dark making it hard to detect as shown in Fig. 1b. The thermal camera uses pseudo colours to display data as shown in Figs. 1 and 2, where bright colour represents warm objects, while the blue or dark colour represents cooler objects. Fig. 3 illustrates the diversity of the images in the dataset collected in various light conditions. Images were divided into three scenes; bright, dark, dawn or dusk. Bright images consist of vehicles with a bright colour like Fig. 1a. The dark images consist of a greater number of red pixels leading to darker images as shown in Fig. 1b. Dawn or dusk images consists of shadows leading to a greater number of green and blue pixels as shown in Fig. 1c.

3.2 Labelling the dataset

The pre-trained vehicle detection algorithms could not identify all the vehicles in the dataset for automatic label creation. Since there was no pre-labelled dataset available, vehicles were labelled manually, and occupied spaces were labelled as cars. The diversity between the images was maintained to facilitate vehicle detection with varying environmental conditions. Each image consists of 30
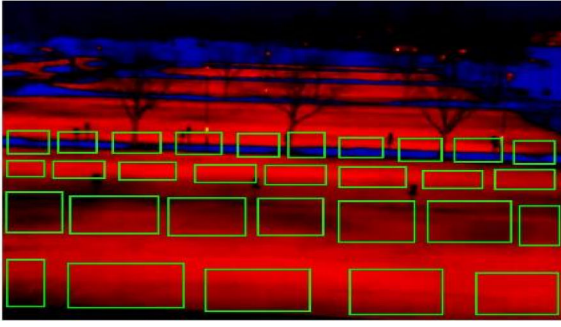
Fig. 4 *Template with bounding boxes*

parking spaces. Images which do not have variations compared to previous images were removed to avoid duplicates and overfitting. A total number of labels in the dataset were 8007. The average size of each label is 90×80 pixels

$$L_t = \left\{ L_1^t, L_2^t, L_3^t, L_4^t, \ldots, L_n^t \right\} \tag{1}$$

where L is the labels, t is the number of images and n represents the total number of labels in one image. The maximum number of labels per image is 30 and it is dependent on the number of vehicles occupying parking spaces.

3.3 Detectors

The detectors implemented in the paper are.

3.3.1 Yolo v2: It is a logistic regression-based multi-object detection algorithm with low computation overhead. It is inspired by the architecture of GoogleNet and aims to reduce the number of parameters by downsampling. It uses a single convolutional neural network which makes it faster compared to other object detection algorithms. It divides the image into 7×7 grid and uses two bounding boxes along with predicted probabilities for object detection [31]. The linear activation function is used for the final layer while a leaky linear activation function as shown in (2) is used on all other layers to enable backpropagation [33]

$$\varnothing(y) = \begin{cases} y, & \text{for} \quad y > 0 \\ 0.1y, & \text{for} \quad y \leq 0 \end{cases}. \tag{2}$$

To improve the computational performance of the detector, multiple downsampling was performed on the input image which leads to coarse features for prediction. It can have problems in detecting smaller objects due to spatial constraints and can also lead to incorrect localisations.

3.3.2 Yolo-conv: It is a custom deep learning algorithm based on Yolo and Xception architectures. Xception is a deep learning network inspired by inception v3 which is a later version of GoogleNet [34]. It consists of depthwise separable convolutional layers and residual connections. Depthwise convolution consists of spatial convolution based on each channel followed by pointwise convolution. Residual connection facilitates in avoiding vanishing gradient and enables deep networks. The number of anchors was increased to 11 which lead to higher Intersection over Union (IoU) value. Combination of depthwise separable convolution layers and residual layers facilitates in reducing parameters and improving accuracy. Since objection detection should be performed based on pseudocolour images, reducing the parameters facilitates in filtering unnecessary features. The final layers consist of Yolo detection and classification layers [31].

3.3.3 GoogleNet: It is a 22-layer pre-trained deep learning network which is computationally efficient and is also called Inception network [35]. It uses several small convolutional neural networks which reduces the number of parameters leading to faster object detection. The last three layers were updated to train with

thermal images dataset. It consists of multiple inception modules to facilitate deep learning and global average pooling layer was used at the end of the network to improve the accuracy. Inception modules also facilitate in improving localisation and object detection. It also consists of a dropout layer, a convolution layer with 128 filters to reduce parameters, a fully connected layer and a linear layer with softmax loss as the classifier [35].

3.3.4 ResNet50 and ResNet18: ResNet50 is a 50-layer pre-trained deep learning network while ResNet18 is an 18 -layer deep network. Resnet uses identity shortcut connections to skip layers and avoid gradient vanishing problem with deep networks [26]. Vanishing gradient problem can occur with deeper networks without the use of skip connections. Residual blocks are assigned for every few blocks of layers. A residual building block is as defined in (3)

$$y = F(x, \{W_l\}) + x \tag{3}$$

x is the input vector and y is the output vector of layers and the function $F(x, \{W_l\})$ is the residual mapping [26]. The use of residual blocks does not increase the complexity or parameters. However, the computational time increases by going deeper. Bottleneck designs are implemented for deeper architectures to reduce the dimensions with the use of convolutions and improving computation time.

3.4 Performance metrics

Average precision and log average miss rates were measured to evaluate the performance of detectors as the detectors perform multi-object detection. The performance of the detectors was also evaluated using the precision-recall curve and log-average miss rate curve [36]. The precision-recall curve gives the measure of relevance and it provides an overview of performance on positive detection. Precision gives the fraction of correctly classified among the total of true positives and false positives. Recall gives the fraction of correctly detected among the total of true positives and false negatives. Miss rate curve gives the error rate of detection per image which gives an overview of false detections. To avoid overfitting, improve performance and validate the entire dataset, *k*-fold cross-validation was performed where *k* is 5. Each *k* consists of equal number of frames. However, the number of labels for each *k* will vary as they are dependent on the occupancy. Approximately 1600 vehicle labels with varying environmental conditions would be present in each *k*. The created dataset consists of more dark images compared to bright or dusk images as vehicle detection is challenging in these conditions and require more training images. The average precision and log-average miss rate for the *k*-fold cross-validation was presented in Section 4. The vehicle occupancy is captured using two steps. The first step is to identify vehicles using detectors with a threshold of 0.5 and the second step is to perform IoU with a template of parking spaces and generate parking occupancy information. The template represents 30 parking spaces as shown in Fig. 4

$$E = l_1, l_2, \ l_3, l_4, \ldots, l_m. \tag{4}$$

As shown in (4), E represents the template, l represents labels with pixel coordinates of parking space while m represents the number of parking spaces.

As shown in (5) the detector generates bounding boxes for identified vehicles where D_i represents detected vehicles per image, i represents the image and p represents number of detected vehicles

$$D_i = D_1^i, \ D_2^i, \ D_3^i, \ D_4^i, \ldots, D_p^i. \tag{5}$$

The IoU measures the overlap between two sets of bounding boxes where a set of bounding boxes is D_i while the other set of bounding boxes is E. If the IoU is >0.2, it assumes the parking space is occupied and if the IoU value is <0.2 it assumes it as a free parking space. With higher IoU value false positives were identified,
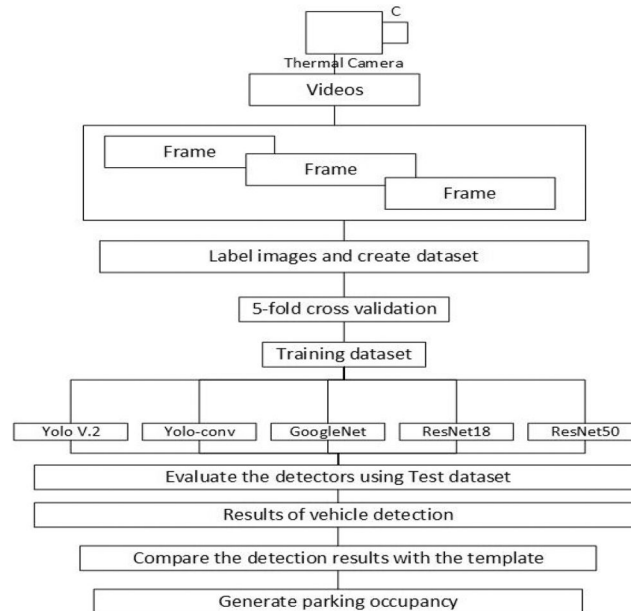
Fig. 5 *Overview of the process*

Table 1 Performance metrics of detectors

DT	AP	MR	ATR	ATE	ET
Yolo v2	59.78	73.57	7.7	0.14	0.5
Yolo-conv	74.79	70	10.5	0.08	0.5
GoogleNet	94.59	27.10	66.2	1.7	1.13
ResNet18	96.16	19.40	62.6	1.1	0.80
ResNet50	94.35	26.53	143.6	2.82	4.3

therefore the IoU value is decreased to 0.2. This is due to overlapping bounding boxes due to occluded images of vehicles. A free parking space is represented by green rectangles and red rectangles represent occupied parking spaces. Yolo, GoogleNet, ResNet18 and Resnet50 used 10 epochs and 0.001 base learning rate for training. Pre-trained GoogleNet, ResNet18 and Resnet50 networks were used as they were experienced in extracting features from several images and provide better detection rates [37]. The training was performed using a Nvidia Quadro P5200 GPU.

3.5 Workflow of the paper

Deep learning algorithms are not widely evaluated on thermal images with varying environmental and heat signatures of the vehicles. Hence, the performance of various deep learning algorithms along with a custom deep learning algorithm is evaluated with this paper. Residual layers enable the algorithm to go deep without reducing too much performance. Therefore, the custom deep learning algorithm utilises Yolo, depthwise and residual layers to balance efficiency and object detection performance. In this paper, deep learning algorithms are only used to identify vehicles and not empty parking spaces. The parking space lines will not be visible during snow conditions or the lines can also be occluded by a vehicle. Deep learning algorithms might require a greater number of empty parking space images to achieve higher detection rates. Thus, to overcome this problem, only vehicle detection was performed using deep learning algorithms. The result from vehicle detection is compared with a template of the parking lot to generate parking occupancy. The template consists of all the parking spaces bounding boxes in the region of interest. The template initially assumes the parking lot to be empty. Detection results consist of bounding boxes of the vehicles which will be compared with the template. Finally, IoU values are generated to classify the parking occupancy. The comparison of detection results with the template is computationally efficient. Hence, the performance of the algorithms is only dependent on the

processing of deep learning algorithm to detect the vehicles. An overview of the process is illustrated in Fig. 5.

4 Results
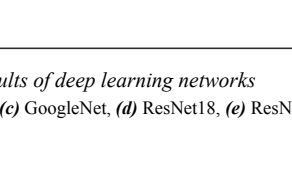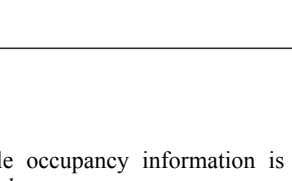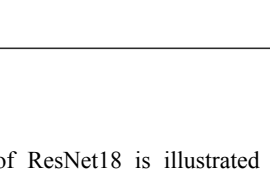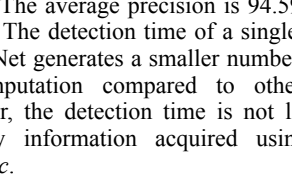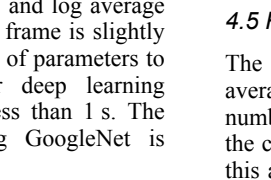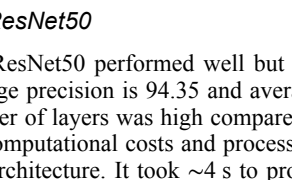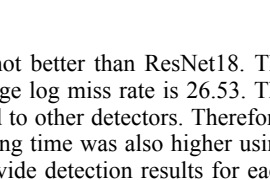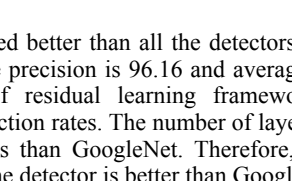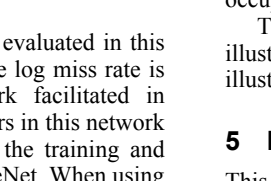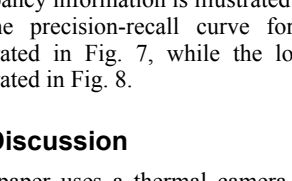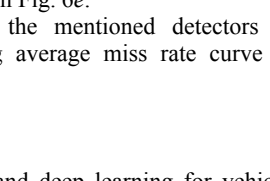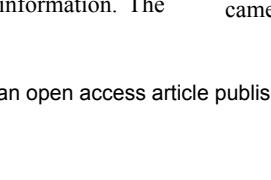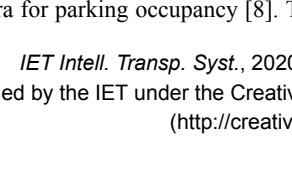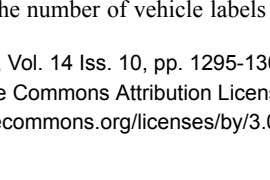
This section discusses and analyses the results obtained from implementing the detectors using 5-fold cross-validation. In Table 1, DT is a detector, AP is average precision, MR is log-average miss rate, ATR is average training time in minutes, ATE is an average testing time in minutes, ET is the execution time of one frame in seconds.

4.1 Yolo

The average precision of Yolo is 59.78 while log average miss rate is 73.57. Yolo uses logistic regression to detect the vehicles in the parking lot and there are a large number of false detections identified using this detector. As mentioned in Table 1, computation and training time is low, efficient and is one of the fastest detectors available. However, the efficiency in detection did not add value in precision and miss rates. More number of training images might be necessary to reduce the miss rate and improve precision. The vehicle occupancy information acquired by Yolo is illustrated in Fig. 6a.

4.2 Yolo-conv

The custom detector Yolo-conv performed better than Yolo where average precision is 74.79 and log average miss rate is 70. As shown in Fig. 6b, the detection of vehicles has improved. However, occlusions such as a group of p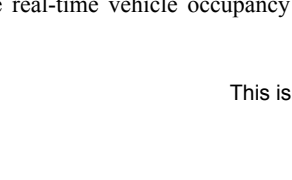eople were considered as vehicles and false detections lead to higher miss rates. Despite additional layers compared to Yolo, the detection rate of the custom detector is the same as Yolo.

Fig. 6 *Detection results of deep learning networks*
(a) Yolo, *(b)* Yolo-conv, *(c)* GoogleNet, *(d)* ResNet18, *(e)* ResNet50

4.3 GoogleNet

GoogleNet deep learning network lead to good detection results as shown in Table 1. The average precision is 94.59 and log average miss rate is 27.10. The detection time of a single frame is slightly above 1 s. GoogleNet generates a smaller number of parameters to improve the computation compared to other deep learning networks. However, the detection time is not less than 1 s. The vehicle occupancy information acquired using GoogleNet is illustrated in Fig. 6*c*.

4.4 ResNet18

ResNet18 performed better than all the detectors evaluated in this paper. The average precision is 96.16 and average log miss rate is 19.40 The use of residual learning framework facilitated in improving the detection rates. The number of layers in this network is 18 which is less than GoogleNet. Therefore, the training and detection time of the detector is better than GoogleNet. When using this network on real-time videos, a number of frames can be reduced to provide real-time vehicle occupancy information. The

vehicle occupancy information is of ResNet18 is illustrated in Fig. 6*d*.

4.5 ResNet50

The ResNet50 performed well but not better than ResNet18. The average precision is 94.35 and average log miss rate is 26.53. The number of layers was high compared to other detectors. Therefore, the computational costs and processing time was also higher using this architecture. It took ~4 s to provide detection results for each image using a single GPU as shown in Table 1. The vehicle occupancy information is illustrated in Fig. 6*e*.

The precision-recall curve for the mentioned detectors is illustrated in Fig. 7, while the log average miss rate curve is illustrated in Fig. 8.

5 Discussion

This paper uses a thermal camera and deep learning for vehicle occupancy detection. There were limited studies using the thermal camera for parking occupancy [8]. The number of vehicle labels in
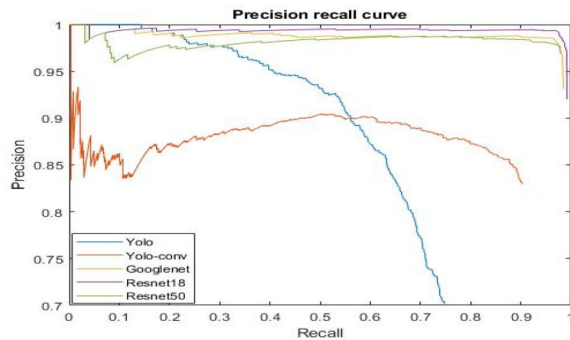
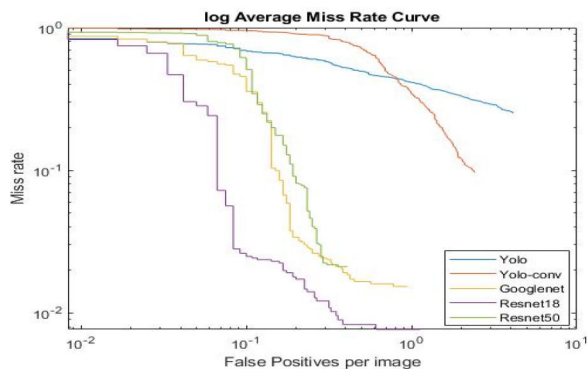**Fig. 7** *Precision-recall curve of the detectors*

Fig. 8 *Log-average miss rate curve of detectors*

the training dataset is ∼6400 for the purpose of object detection. The available pre-trained detectors were trained with millions of images and labels. Therefore, pre-trained network architectures of GoogleNet, ResNet and Yolo were used along with custom network Yolo-conv. Since, they were already pre-trained detectors, fewer number of training set images were enough to perform object detection. Several thousands of images would be needed if pre-trained network architectures were not used. Despite the smaller number of training images, GoogleNet and ResNet networks had higher precision values and lower miss rates. GoogleNet and ResNet networks designed for feature extraction and object detection using visual colour images also performed well in extracting features from thermal images. The use of deep learning facilitated in vehicle detection irrespective of the visual or infrared spectrum. The training time of Yolo is ∼8 min for 6400 labels and its performance is low compared to other detectors. As shown in Fig. 6, a group of people were also recognised as a vehicle which might be due to coarse features used for detection and spatial constraints. Yolo-conv which consists of several convolutional and residual layers facilitated in improving the accuracy compared to Yolo. Detection of smaller objects has improved as shown in Fig. 6*b*. Despite having several convolutional and residual layers, the custom detector still recognised people as vehicles and few vehicles could not be detected. Therefore, softmax, fully connected and classification layers performed better vehicle detection which was present in ResNet and GoogleNet architectures. The vehicles in bright conditions are mostly recognised by all the evaluated algorithms as the features of the vehicle are distinguishable. In dawn and dusk conditions, there were minimal false detections. The combination of vehicle and shadowy features lead to false detections where the shadow regions consist of more red and blue pixels. Majority of the false detections occurred with dark images where vehicles are hard to recognise. A dark image consists of higher number of red pixels where features of the vehicle are blended with the background features. Therefore, all the algorithms consist of false detections with dark images as shown in Fig. 6. However, the number of false detections with dark and dusk images is higher with Yolo compared to GoogleNet and ResNet algorithms. This can also be dependent on the camera resolution and recognition distance. The performance of the algorithms can also be improved by increasing the size of the dataset and

modifying loss function. GoogleNet and ReseNet18 training times were almost similar and the detection time of ResNet18 is slightly faster than GoogleNet. ResNet uses skip functions to facilitate the use of larger number of layers. However, in this case, ResNet18 provided better results compared to ResNet50. More number of layers can be useful if there were millions of labels. The computation time of ResNet50 is also comparatively higher than other detectors. Computation time plays a vital role in providing real-time vehicle occupancy. Therefore, ResNet50 is not suitable for real-time vehicle occupancy.

Higher number of diverse images can be included to improve the detection rates in any environmental or light conditions. If the position of the camera is placed at an increased height or placed on top of the parking lot, unoccluded images of vehicles can be obtained which would lead to better data quality. A better-quality dataset can also lead to a reduction of miss rates. The current position of the camera and viewing angle does not cover all the vehicles in the first four rows. Reducing the focal length would increase the viewing angle of the camera, thereby covering a greater number of vehicles. The IoU value used in this paper is 0.2. for acquiring parking occupancy. A vehicle might park in between two parking spaces which commonly occurs during snow conditions as lines of parking spaces might not be visible. In such cases, the current IoU value would mark the two parking spaces occupied which is true in practical terms but opposite in actual terms.

6 Conclusion and future work

In this paper, real-time vehicle occupancy using a thermal camera and deep learning was proposed. Deep learning object detection algorithms such as; Yolo, Yolo-conv, GoogleNet, ResNet18 and ResNet50 were implemented to identify multiple vehicles as they are dynamic and intelligent contemporary detectors. These detectors facilitate vehicle detection in varying illumination and environmental conditions. However, there were more false detections with dark images due to diminished vehicle features and a greater number of red pixels. The detected results were compared with a template of parking spaces to acquire real-time vehicle occupancy information. ResNet18 performed better compared to other detectors and it can be used to capture real-time vehicle occupancy as it uses less computational time. The future work can be focused on improving the performance of the detectors by increasing vehicle images. There are several other deep learning networks such as Single-shot detector, Faster RCNN, MobileNet and NasNet which can be evaluated in the future. A fast detector such as Yolo architecture was modified using convolutional and residual layers. However, there was a marginal improvement in the performance of the detector. Sharpening and contrast filters can be applied on images to improve the quality of features and thereby improving the performance of the detectors. However, usage of filters can increase the computation time of the detector which is not suitable to generate real-time parking occupancy. This paper addresses the first step in identifying vehicle occupancy information and the next step would be to automate the process of identifying vehicle occupancy suitable for any parking lot and provide navigational directions to probable empty parking spaces. This information will be fed to a smart parking application which can be utilised by the users.

7 References

[1] Shoup, D.C.: 'Cruising for parking', *Transp. Policy*, 2006, **13**, (6), pp. 479–486
[2] Polycarpou, E., Lambrinos, L., Protopapadakis, E.: 'Smart parking solutions for urban areas'. 2013 IEEE 14th Int. Symp. and Workshops on a World of Wireless, Mobile and Multimedia Networks (WoWMoM), Madrid, Spain, 2013
[3] Paidi, V., Fleyeh, H., Håkansson, J*., et al.*: 'Smart parking sensors, technologies and applications for open parking lots: a review', *IET Intell. Transp. Syst.*, 2018, **12**, (8), pp. 735–741
[4] Hassoune, K., Dachry, W., Moutaouakkil, F*., et al.*: 'Smart parking systems: a survey'. 2016 11th Int. Conf. on Intelligent Systems: Theories and Applications (SITA), Mohammedia, Morocco, 2016

[5] Idris, M.Y.I., Leng, Y.Y., Tamil, E.M.*, et al.*: 'Car park system: a review of smart parking system and its technology', *Inf. Technol. J.*, 2009, **8**, (2), pp. 101–113

[6] Robert, K.: 'Night-time traffic surveillance: a robust framework for multi-vehicle detection, classification and tracking'. 2009 Sixth IEEE Int. Conf. on Advanced Video and Signal Based Surveillance, Genova, Italy, 2009

[7] Sosnowski, T., Bieszczad, G., Madura, H.: 'Image processing in thermal cameras', in Nawrat, A, Bereska, D, Jędrasiak, K (Eds.): '*Advanced technologies in practical applications for national security*' (Springer, Cham, Switzerland, 2018) pp. 35–57

[8] Chang, C.-W., Srinivasan, K., Chen, Y.Y.*, et al.*: 'Vehicle detection in thermal images using deep neural network'. 2018 IEEE Visual Communications and Image Processing (VCIP), Taichung, Taiwan, 2018

[9] Bastan, M., Yap, K.-H., Chau, L.-P.: 'Remote detection of idling cars using infrared imaging and deep networks', *Neural Comput. Appl.*, 2018, **32**, pp. 1–11

[10] Eliko.: 'Parking space detection system', 2019 [cited 2019 2019-10-23]. Available at: https://www.eliko.ee/projects/parking-space-detection-system/

[11] Nam, Y., Nam, Y.-C.: 'Vehicle classification based on images from visible light and thermal cameras', *EURASIP J. Image Video Process.*, 2018, **2018**, (1), p. 5

[12] Chien, S.-C., Chang, F.C., Tsai, C.C.*, et al.*: 'Intelligent all-day vehicle detection based on decision-level fusion using color and thermal sensors'. 2017 Int. Conf. on Advanced Robotics and Intelligent Systems (ARIS), Taipei, Taiwan, 2017

[13] Bao, P., Zhang, L., Wu, X.: 'Canny edge detection enhancement by scale multiplication', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2005, **27**, (9), pp. 1485–1490

[14] Maini, R., Aggarwal, H.: 'Study and comparison of various image edge detection techniques', *Int. J. Image Process. (IJIP)*, 2009, **3**, (1), pp. 1–11

[15] Dalal, N., Triggs, B.: 'Histograms of oriented gradients for human detection'. CVPR 2005 IEEE Computer Society Conf. on Computer Vision and Pattern Recognition, San Diego, CA, USA, 2005

[16] Mao, L., Xie, M., Huang, Y.*, et al.*: 'Preceding vehicle detection using histograms of oriented gradients'. 2010 Int. Conf. on Communications, Circuits and Systems (ICCCAS), Chengdu, China, 2010

[17] Fleyeh, H., Mohammed, I.: 'Night time vehicle detection', *J. Intell. Syst.*, 2012, **21**, (2), pp. 143–165

[18] Iwasaki, Y., Misumi, M., Nakamiya, T.: 'Robust vehicle detection under various environmental conditions using an infrared thermal camera and its application to road traffic flow monitoring', *Sensors*, 2013, **13**, (6), pp. 7756–7773

[19] Gaszczak, A., Breckon, T.P., Han, J.: 'Real-time people and vehicle detection from UAV imagery'. Intelligent Robots and Computer Vision XXVIII: Algorithms and Techniques, Int. Society for Optics and Photonics, San Francisco Airport, California, United States, 2011

[20] Irhebhude, M.E., Ali, M.A., Edirisinghe, E.A.: 'Pedestrian detection and vehicle type recognition using centrog features for nighttime thermal images'. 2015 IEEE Int. Conf. on Intelligent Computer Communication and Processing (ICCP), Cluj-Napoca, Romania, 2015

[21] Sangnoree, A., Chamnongthai, K.: 'Thermal-image processing and statistical analysis for vehicle category in nighttime traffic', *J. Vis. Commun. Image Represent.*, 2017, **48**, pp. 88–109

[22] Dollar, P., Appel, R., Belongie, S.*, et al.*: 'Fast feature pyramids for object detection', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2014, **36**, (8), pp. 1532–1545

[23] Liu, K., Mattyus, G.: 'Fast multiclass vehicle detection on aerial images', *IEEE Geosci. Remote Sens. Lett.*, 2015, **12**, (9), pp. 1938–1942

[24] Lan, W., Dang, J., Wang, Y.*, et al.*: 'Pedestrian detection based on YOLO network model'. 2018 IEEE Int. Conf. on Mechatronics and Automation (ICMA), Changchun, China, 2018

[25] Amato, G., Carrara, F., Falchi, F.*, et al.*: 'Deep learning for decentralized parking lot occupancy detection', *Expert Syst. Appl.*, 2017, **72**, pp. 327–334

[26] He, K., Zhang, X., Ren, S.*, et al.*: 'Deep residual learning for image recognition'. Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition, Las Vegas, Nevada, 2016

[27] Sheng, M., Liu, C., Zhang, Q.*, et al.*: 'Vehicle detection and classification using convolutional neural networks'. 2018 IEEE 7th Data Driven Control and Learning Systems Conf. (DDCLS), Enshi, China, 2018

[28] Aswathy, P., Mishra, D.: 'Deep GoogLeNet features for visual object tracking'. 2018 IEEE 13th Int. Conf. on Industrial and Information Systems (ICIIS), Rupnagar, India, 2018

[29] Ren, S., He, K., Girshick, R.*, et al.*: 'Faster R-CNN: towards real-time object detection with region proposal networks'. Advances in Neural Information Processing Systems, Montréal, Canada, 2015

[30] Lin, J.-P., Sun, M.-T.: 'A YOLO-based traffic counting system'. 2018 Conf. on Technologies and Applications of Artificial Intelligence (TAAI), Taichung, Taiwan, 2018

[31] Ren, P., Fang, W., Djahel, S.: 'A novel YOLO-based real-time people counting approach'. 2017 Int. Smart Cities Conf. (ISC2), Wuxi, China, 2017

[32] Datainspektionen.: 'The purposes and scope of the general data protection regulation', 2018. Available at: https://www.datainspektionen.se/other-lang/in-english/the-general-data-protection-regulation-gdpr/the-purposes-and-scope-of-the-general-data-protection-regulation/

[33] Redmon, J., Divvala, S., Girshick, R.*, et al.*: 'You only look once: unified, real-time object detection'. Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition, Las Vegas, Nevada, 2016

[34] Chollet, F.: 'Xception: deep learning with depthwise separable convolutions'. Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition, Honolulu, Hawaii, 2017

[35] Szegedy, C., Liu, W., Jia, Y.*, et al.*: 'Going deeper with convolutions'. Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition, Boston, Massachusetts, 2015

[36] Hossin, M., Sulaiman, M.: 'A review on evaluation metrics for data classification evaluations', *Int. J. Data Mining Knowl. Manage. Process*, 2015, **5**, (2), p. 1

[37] Jo, S.Y., Ahn, N., Lee, Y.*, et al.*: 'Transfer learning-based vehicle classification'. 2018 Int. SoC Design Conf. (ISOCC), Daegu, South Korea, 2018

1302

IET Intell. Transp. Syst., 2020, Vol. 14 Iss. 10, pp. 1295-1302