# On-road vehicle detection in varying weather conditions using faster R-CNN with several region proposal networks

Rajib Ghosh[1] (ORCID)

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2021

## Abstract

Developing automated systems to detect and track on-road vehicles is a demanding research area in Intelligent Transportation System (ITS). This article proposes a method for on-road vehicle detection and tracking in varying weather conditions using several region proposal networks (RPNs) of Faster R-CNN. The use of several RPNs in Faster R-CNN is still unexplored in this area of research. The conventional Faster R-CNN produces regions-of-interest (ROIs) through a single fixed sized RPN and therefore cannot detect varying sized vehicles, whereas the present investigation proposes an end-to-end method of on-road vehicle detection where ROIs are generated using several varying sized RPNs and therefore it is able to detect varying sized vehicles. The novelty of the proposed method lies in proposing several varying sized RPNs in conventional Faster R-CNN. The vehicles have been detected in varying weather conditions. Three different public datasets, namely DAWN, CDNet 2014, and LISA datasets have been used to evaluate the performance of the proposed system and it has provided 89.48%, 91.20%, and 95.16% average precision on DAWN, CDNet 2014, and LISA datasets respectively. The proposed system outperforms the existing methods in this regard.

**Keywords** On-road vehicle detection · Faster R-CNN · Several RPNs · Varying weather conditions

## 1 Introduction

Investigations on developing computer vision based automated systems for detection and tracking of on-road vehicles have gained momentum in the recent past due to the vast applications of these

✉ Rajib Ghosh
  rajib.ghosh@nitp.ac.in

[1] Department of Computer Science and Engineering, National Institute of Technology Patna, Patna, India

systems in this digital era. The main goal of these systems is to detect various on-road vehicles of varying sizes from still images or video data collected from the road. These systems provide innovative support to transport and traffic management in various ways. Vehicle detection systems provide information in vehicle counting and classification at toll plaza [28], vehicle number plate recognition [11], vehicle speed measurement, tracking traffic accidents, etc. These systems also provide assistance to the drivers during driving by tracking front and rear vehicles in the same as well as in different lanes, specially in heavy rain, snowfall or foggy weather. With the continuous development of urban roads, increase in vehicle buying capacity of common people and as a consequence more number of on-road vehicles, the importance of automated vehicle detection systems is gradually increasing. With the advent of more developed computer vision techniques, more developed systems are being enabled in this domain.

Several vision based vehicle detection systems are available in the literature using both Machine Learning (ML) and non-ML methods. In ML base studies, researchers have mostly relied upon three different shallow ML techniques to detect vehicles—Support Vector Machine (SVM) [8, 20, 30, 31], Adaboost classifier [30, 34, 36], and Artificial Neural Network (ANN) [24, 25]. Although shallow Convolutional Neural Network (CNN) has been successfully used in various studies for object detection in natural scene images, a little number of studies [39] have used shallow CNN to locate on-road vehicles in natural scene images. The use of deep CNN has alsobeen reported [16] in the literature for on-road vehicle detection. But, the use of the deep learning method Faster R-CNN with multiple region proposal networks (RPNs) has not been reported yet. The Faster R-CNN structure was first published in NIPS 2015 [27] in order to detect objects from natural scene images. The conventional Faster R-CNN produces regions-of-interest (ROIs) through a single RPN utilizing the feature matrix of the last convolutional layer and therefore cannot always detect vehicles of varying sizes from the images. The present investigation proposes a novel method of generation of ROIs in Faster R-CNN by incorporating several RPNs in the conventional Faster R-CNN structure generated through different intermediate as well as the last convolutional layers. Sizes of these RPNs vary as well and it enables the present system to detect the on-road vehicles of varying sizes in images. The vehicles have been detected from the images collected in good (sunny) as well as challenging weather conditions like blizzard, snowfall, and wet snow weather. Few on-road vehicle images in varying weather conditions from CDNet 2014 [33] and LISA [29] public datasets are shown in Fig. 1.

**The major contributions of the proposed system are as follows:**

1. Images of various challenging weather conditions have been considered in the present work during detecting on-road vehicles.
2. No study has been found in the literature in this problem area on using Faster R-CNN with multiple RPNs, a deep learning approach. The present article proposes a computer vision based on-road vehicle detection method using Faster R-CNN with several RPNs, which is a novel idea. In order to detect the vehicles of varying sizes, several RPNs of varying sizes have been introduced in the conventional Faster R-CNN structure.

The rest of the paper is organized as follows. Related works are discussed in Section 2. Section 3 details the proposed method. Section 4 analyses the vehicle detection results of the present system. Finally, Section 5 concludes the paper with a direction for future research.
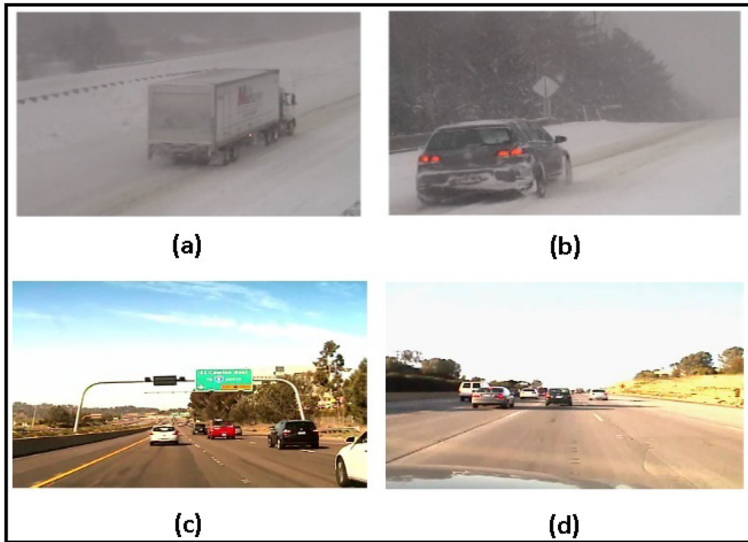
**Fig. 1** Few sample on-road vehicle images in varying weather conditions: (**a**) In blizzard (CDNet 2014 dataset), (**b**) In snowfall (CDNet 2014 dataset), (**c**) In sunny (LISA dataset), and (**d**) In dense (LISA dataset)

## 2 Literature survey

Several studies on computer vision based vehicle detection systems are available in the literature. Most of the studies have relied on different shallow ML techniques to detect vehicles, whereas few studies have used non-ML techniques as well. Some of those related studies are discussed below in brief.

**Studies using ML techniques** Sun et al. [31] presented a computer vision based vehicle detection system utilizing the popular ML technique, SVM. In this study, two different classes were created, one for vehicle and the other for non-vehicle and the classification was performed using SVM after extracting various features from the samples of each class. Cheon et al. [8] proposed another vision based vehicle detection method where the features were extracted through Histogram of Oriented Gradients (HOG) and the classification of vehicle and non-vehicle was performed using SVM. In another study [30] on on-road vehicle detection, two different classifiers were applied, SVM and Adaboost, to detect the vehicles. Feature vectors generated through HOG features were classified using SVM and the vectors generated through Haar-like features were classified using Adaboost classifier. Khairdoost at al. [20] presented an on-road vehicle detection method for both front and rear vehicles. Feature vectors were generated through Pyramid Histogram of Oriented Gradients (PHOG) features and classified using linear SVM. Yan et al. [36] reported one investigation outcome in this regard using Adaboost classifier. Two types of HOG features were extracted from vehicle and non-vehicle samples and the feature vectors were classified using Adaboost classifier in one of the two classes, vehicle or non-vehicle. The use of Adaboost classifier is found in another study [34] as well, where Haar-like features were extracted from vehicle and non-vehicle classes samples and the classification of these feature vectors were performed using Adaboost classifier. Few studies have relied upon ANN as well to detect on-road vehicles. Ming et al. [24] used ANN to detect vehicles based on the information obtained from vehicle tail light. In

another study on the use ANN [25], Haar-like features were extracted from the samples of vehicle and non-vehicle classes and the feature vectors were classified using ANN. Matthews et al. [23] presented a two-stage vehicle recognition method by combining image processing techniques with ANN. ROIs were generated using various image processing techniques which were then fed to ANN for vehicle recognition. Zhou et al. [39] proposed an end-to-end vehicle detection system using shallow CNN. Ghosh et al. [28] presented a system for detection and classification of vehicles in toll plaza using image processing and ML techniques. A two-stage detector using Cascade R-CNN was presented by Cai et al. [2]. Hu et al. [18] proposed SINet, which provided fast vehicle detection using a scale insensitive network. In a recent researcheffort [16], Hassaballah et al. proposed a deep CNN based framework for vehicle detection in adverse weather conditions. In this work, the visibility of the image has been enhanced by enhancing the illumination and reflection component.

**Studies using non-ML techniques** Sivaraman et al. [29] presented an active-learning framework to track on-road vehicles. The system was trained using a supervised learning technique. Chan et al. [5] presented a non-ML based system to detect the preceding vehicles during driving in varying light and weather conditions. Four different vehicular structure related cues were considered and a particle filter, combined with these cues, was used to detect the preceding vehicles. Another non-ML based vehicle tracking system was proposed by Chellappa et al. [6] using acoustic and video sensors. Both video and acoustic sensors were fused to develop the system. Bertozzi et al. [1] proposed another non-ML based vehicle detection system where vehicles were located through detecting its corners as in general vehicles have a rectangular shape with four corners. One template was considered corresponding to each corner to locate all of the four corners of any vehicle. In another non-ML based vehicle detection study [32], a multi-sensor correlation method was proposed utilizing magnetic wireless sensor network. Fossati et al. [10] presented an on-driving vehicle tracking system through detecting the taillights of vehicles, coupled with rule based system. Haselhoff et al. [13] used Haar-like and Triangle features to detect on-road vehicles. Triangle filters were computed for feature extraction using four integral images. Another non-ML based research exploration [38] presented a method to detect partially occluded vehicles using geometric and likelihood information. Vehicle-vehicle occlusions can be detected using these two information. The likelihood information was obtained through generating the connected positive region using the union-find algorithm. Chen et al. [7] proposed one vehicle detection method employing Haar-like features and eigencolours. The occlusion problem was tackled by estimating theprobability of each vehicle using chamfer distance. Recently, another study [9] has addressed the tracking-learning-detection algorithm to track on-road moving vehicles from video data. The Kalman filter has been utilized to predict the vehicleposition during occlusion. The normalized cross-correlation coefficient has been used to get the correct bounding boxes of detected vehicles. Haselhoff et al. [14] presented a learning algorithm for vehicle detection. Vehicles were represented using Haar-like features and a weighted correlation was utilized for the matching purpose. In a recent research effort [15], Hassaballah et al. presented a local binary pattern (LBP) based onâ€'road vehicle detection method in urbanscenarios. In this effort, texture and appearance histograms of vehicles were fed to the clustering forests. You et al. [37] has proposed a word embedding scheme from natural language processing to reflect the contextual relationships between the pixels for image recognition. This word embedding

scheme generates bidirectional real dense vectors. Recently, Cai et al. [3] has proposed a combination of cross-attention mechanism and graph convolution scheme for image classification.

Although several investigations have been reported in the literature in this problem area, the accuracy of detecting the vehicles with varying sizes is not too high. No study has been found in this problem area using Faster R-CNN with multiple RPNs. Due to introducing several varying dimensional RPNs in the traditional Faster R-CNN structure, the proposed system can detect the vehicles with varying sizes more accurately in comparison to the existing studies.

# 3 Proposed method

The present investigation proposes an end-to-end method of on-road vehicle detection and tracking through proposing a modified form of the conventional Faster R-CNN structure. The conventional Faster R-CNN produces ROIs through a single RPN utilizing the feature matrix of the last convolutional layer only. The proposed structure of Faster R-CNN shall include several varying sized RPNs and an additional merger layer to merge these RPNs. The systems employing conventional Faster R-CNN cannot detect varying sized vehicles from on-road images due to the use of a single fixed-sized RPN, whereas the proposed system will be able to overcome this drawback by incorporating multiple RPNs of variable sizes in the conventional Faster R-CNN structure. The proposed modified architecture of Faster R-CNN incorporating several RPNs is shown in Fig. 2.

**The proposed method consists of the following steps**:

1. Input the image frame containing vehicles to the input layer of CNN.
2. Creation of several RPNs in the CNN architecture and setting of different anchors with varying scale and aspect ratio in each RPN.
3. Merging of all the RPNs and pooling of the ROIs.
4. Detection of vehicles based on the trained model generated during training of each RPN.
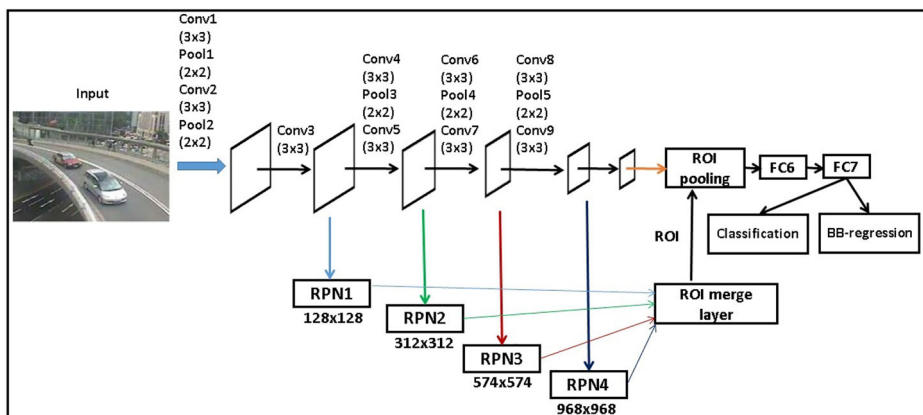


**Fig. 2** The proposed modified architecture of Faster R-CNN incorporating several RPNs

**Table 1** Statistics of different anchors

| Scale | Size(Height×Width) | | | |
|---|---|---|---|---|
| 1:1 | 128×128 | 257×257 | 515×515 | 1031×1031 |
| 1:2 | 105×210 | 211×422 | 423×846 | 847×1694 |
| 2:1 | 200×100 | 370×185 | 670×335 | 1250×625 |
| 1.5:1 | 135×90 | 270×180 | 540×360 | 1080×720 |

## 3.1 Creation of several RPNs

The conventional Faster R-CNN contains a single RPN generated from the feature matrix of the last convolutional layer. In similar-sized object detection problems in images, the receptive field of this single RPN may be capable to contain the features of the objects. However, as far as vehicle detection problem is concerned, it is not capable to detect varying sized vehicles from on-road images because the dimensionality of the receptive field of the single RPN may not be enough to hold the features of thevehicles of all sizes. So, a modified architecture of Faster R-CNN with several RPNs is proposed in the present article to overcome this drawback. The effectiveness of the proposed architecture using several RPNs over conventional Faster R-CNN has been demonstrated in Section 4.

The proposed architecture (shown in Fig. 2) contains four varying sized RPNs. The number "four" has been decided using *Bayesian optimization technique*. RPN1 has been generated from Conv1-Pool1-Conv2-Pool2-Conv3 sequence, whereas RPN2 has been generated from Conv4-Pool3-Conv5 sequence. The dimension of the kernel is $3 \times 3$ for convolution operations, whereas for pooling operations the kernel dimension is $2 \times 2$. The receptive field dimension of RPN1 is $128 \times 128$, whereas it is $312 \times 312$ for RPN2. RPN3 has been created from Conv6-Pool4-Conv7 sequence and RPN4 is created from Conv8-Pool5-Conv9 sequence. The receptive fields of RPN3 and RPN4 are of $574 \times 574$ and $968 \times 968$ respectively. The dimension of the input image is $1200 \times 1200$, so thesevarying dimensional receptive fields of different RPNs can hold the features of the vehicles of almost all sizes. The aforementioned order of receptive field dimensions has provided the best performance, the details are shown in Table 4.

## 3.2 Merging of RPNs and setting anchors

As the outcome of each RPN is more than one different ROI arrays, so these separate arrays are required to be merged into a single array. To achieve this, a separate ROI merger layer has been considered in the proposed modified architecture of Faster R-CNN. This layer merges the separate ROI arrays into a single one. This single merged array has then been fed for ROI pooling. To select the appropriate ROIs for detection of vehicles, the intersection over union (IoU) operation has been performed. IoU measures the area of overlapping between the ROIs

**Table 2** Definition of metrics used in the present work

| Metrics | Definition |
|---|---|
| Precision | (True Positive)/(True Positive+False Positive) |
| Recall | (True Positive)/(True Positive+False Negative) |

**Table 3** Results and the processing speed of the proposed vehicle detection system using Faster R-CNN with several RPNs

| Dataset | Weather condition | Accuracy | Precision | Recall | Processing speed (in fps) |
|---|---|---|---|---|---|
| DAWN | Fog | 90.62% | 90.44% | 90.52% | 20 |
| | Snow | 89.58% | 89.32% | 89.41% | |
| | Rain | 89.21% | 89.04% | 89.12% | |
| | Sandstorms | 89.25% | 89.14% | 89.19% | |
| CDNet 2014 | Sunny | 94.54% | 94.12% | 94.32% | 20 |
| | Blizzard | 92.72% | 92.44% | 92.56% | |
| | Snowfall | 89.46% | 89.14% | 89.28% | |
| | Wet snow | 89.42% | 89.12% | 89.25% | |
| LISA | Sunny evening | 95.28% | 95.12% | 95.21% | 18 |
| | Cloudy morning | 95.18% | 95.08% | 95.14% | |
| | Sunny afternoon | 95.42% | 95.28% | 95.34% | |

and ground-truth bounding boxes. ROIs having the IoU value greater than 0.8 have been chosen as the candidate ROIs to detect vehicles. From the candidate ROIs, 80 ROIs with higher IoU value have been selected finally for the following process. *Bayesian optimization technique* has been used to find the optimal value of number of ROIs.

Different anchors with varying scale and aspect ratio have been set in each RPN. Anchor is a rectangle denoting the region of the target object. The detailed statistics of different anchors in the proposed system are presented in Table 1. To enable the detection of vehicles of various sizes, small anchors (columns 1 and 2 of the Table 1) have been set in RPN1, comparatively larger anchors in RPN2 (columns 2 and 3 of the Table 1) and the largest anchors in RPN3 and RPN4 (columns 3 and 4 of the Table 1).

**Table 4** Precision of the proposed system against different orders of receptive field dimensions of four RPNs

| Dataset | Weather condition | Precision | Order of receptive field dimensions of four RPNs |
|---|---|---|---|
| DAWN | Fog | 90.44% | $128 \times 128$-$312 \times 312$-$574 \times 574$-$968 \times 968$ |
| | | 86.62% | $968 \times 968$-$574 \times 574$-$312 \times 312$-$128 \times 128$ |
| | Snow | 89.32% | $128 \times 128$-$312 \times 312$-$574 \times 574$-$968 \times 968$ |
| | | 85.91% | $968 \times 968$-$574 \times 574$-$312 \times 312$-$128 \times 128$ |
| | Rain | 89.04% | $128 \times 128$-$312 \times 312$-$574 \times 574$-$968 \times 968$ |
| | | 85.42% | $968 \times 968$-$574 \times 574$-$312 \times 312$-$128 \times 128$ |
| | Sandstorms | 89.14% | $128 \times 128$-$312 \times 312$-$574 \times 574$-$968 \times 968$ |
| | | 84.92% | $968 \times 968$-$574 \times 574$-$312 \times 312$-$128 \times 128$ |
| CDNet 2014 | Sunny | 94.12% | $128 \times 128$-$312 \times 312$-$574 \times 574$-$968 \times 968$ |
| | | 90.23% | $968 \times 968$-$574 \times 574$-$312 \times 312$-$128 \times 128$ |
| | Blizzard | 92.44% | $128 \times 128$-$312 \times 312$-$574 \times 574$-$968 \times 968$ |
| | | 87.28% | $968 \times 968$-$574 \times 574$-$312 \times 312$-$128 \times 128$ |
| | Snowfall | 89.14% | $128 \times 128$-$312 \times 312$-$574 \times 574$-$968 \times 968$ |
| | | 84.81% | $968 \times 968$-$574 \times 574$-$312 \times 312$-$128 \times 128$ |
| | Wet snow | 89.12% | $128 \times 128$-$312 \times 312$-$574 \times 574$-$968 \times 968$ |
| | | 83.72% | $968 \times 968$-$574 \times 574$-$312 \times 312$-$128 \times 128$ |
| LISA | Sunny evening | 95.12% | $128 \times 128$-$312 \times 312$-$574 \times 574$-$968 \times 968$ |
| | | 91.32% | $968 \times 968$-$574 \times 574$-$312 \times 312$-$128 \times 128$ |
| | Cloudy morning | 95.08% | $128 \times 128$-$312 \times 312$-$574 \times 574$-$968 \times 968$ |
| | | 90.56% | $968 \times 968$-$574 \times 574$-$312 \times 312$-$128 \times 128$ |
| | Sunny afternoon | 95.28% | $128 \times 128$-$312 \times 312$-$574 \times 574$-$968 \times 968$ |
| | | 91.12% | $968 \times 968$-$574 \times 574$-$312 \times 312$-$128 \times 128$ |

**Table 5**  Optimal set of values of various hyper parameters in Faster R-CNN with several RPNs

| Hyper parameters | Value |
| --- | --- |
| Batch size | 512 |
| Overlap threshold for ROI | 0.8 |
| Number of RPNs | 4 |
| Number of ROIs | 80 |
| Learning Rate | 0.05 |
| Weight decay for regularization | 0.005 |

### 3.3 Training scheme

In the proposed architecture of Faster R-CNN using several RPNs, a separate training data has been prepared corresponding to each RPN due to the existence of variable-sized anchors in the RPNs. Small rectangles in the ground-truth, covering the small-sized vehicles, have been used to train RPN1. The large rectangles, covering the large-sized vehicles, have been used in the ground-truth to train RPN3 and RPN4. In the present work, the larger side of the small rectangles is smaller than 260 unit, whereas the larger side of the large rectangles is larger than 400 unit. These thresholds have been decided empirically. In the proposed architecture, RPN loss function is applied for each RPN. The learning rate has been set as 0.05 to train each RPN. *Bayesian optimization technique* has been used to find the optimal value of learning rate.

## 4 Experimental results and analysis

The experiments have been performed using both conventional Faster R-CNN with single RPN and the proposed architecture of Faster R-CNN with several RPNs to have a performance comparison.

### 4.1 Dataset description

In comparison to other research fields in computer vision, on-road vehicle detection field has fewer publicly available datasets. The present investigation has used DAWN [19], CDNet 2014, and LISA public datasets to train and test the proposed model. The details of the datasets used in the present work are discussed below.



**Fig. 3**  Correct detection of on-road vehicles using the proposed method on few test images from DAWN dataset in (**a**) Fog, (**b**) Snow, and (**c**) Rain conditions. The red bounding box indicates the ground-truth and the green one indicates the detecting box

**Fig. 4** Correct detection of on-road vehicles using the proposed method on few test images from CDNet 2014 dataset in (**a**) Sunny, (**b**) Snow fall, and (**c**) Blizzard conditions. The red bounding box indicates the ground-truth and the green one indicates the detecting box

### 4.1.1 Dawn

This dataset contains various real-world images collected under various adverse weather conditions. The images have been collected from a diverse traffic environments (urban, highway and freeway) as well as a rich variety of traffic flow. This dataset contains 1000 images which are divided into four categories of weather conditions: fog, snow, rain, and sandstorms. These 1000 images have been divided into training and testing sets in 3:1 ratio using holdout method. An uniform number of samples from each weather condition have been considered in both training and testing sets.

### 4.1.2 CDNet 2014

This dataset contains a realistic, camera-captured, diverse set of videos of various on-road vehicles in good (sunny) as well as in challenging weather conditions like blizzard, snowfall, and wet snow under the category "bad weather". Videos were acquiredin both day and night time conditions in CDNet 2014 dataset. The video in sunny condition contains 2500 consecutive frames, whereas the video in blizzard condition contains 7000 consecutive frames. The video sequence in snowfall condition contains 6500 consecutive frames and that of the video sequence in wet snow condition contains 3500 consecutive frames. In each condition, the total number of frames have been divided into training and testing sets in 3:1 ratio using holdout method. An uniform number ofsamples from each weather condition have been considered in both training and testing sets.



**Fig. 5** Correct detection of on-road vehicles using the proposed method on few test images in (**a**) Sunny afternoon (LISA dataset) and (**b**) Wet snow (CDNet 2014 dataset) conditions. The red bounding box indicates the ground-truth and the green one indicatesthe detecting box
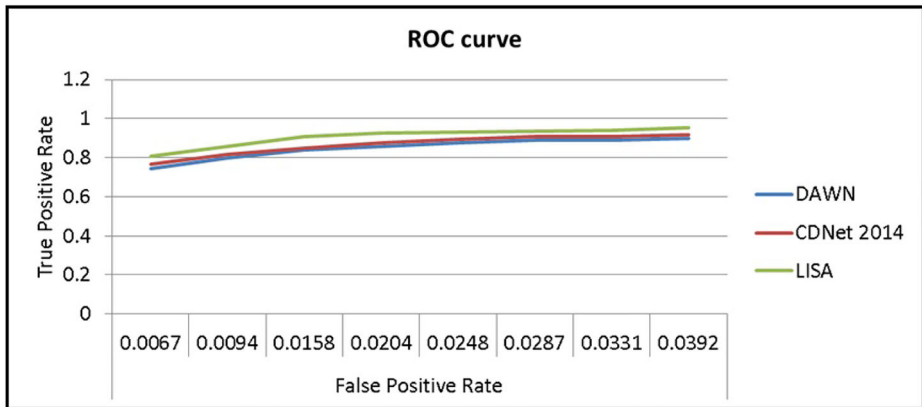
**Fig. 6** ROC curve analysis of the proposed system using Faster R-CNN with several RPNs

### 4.1.3 Lisa

The LISA dataset contains three different videos, collected in three different conditions. The first video contains 1600 consecutive frames, collected during sunny evening rush hour traffic. The second video contains 300 consecutive frames, collected in acloudy morning on urban roads. The third one also contains 300 consecutive frames, captured on the highways in a sunny afternoon. All the video sequences have been captured from an on-board camera. In each condition, the total number of frames have been divided into training and testing sets in 3:1 ratio using holdout method. An uniform number of samples from each weather condition have been considered in both training and testing sets.

### 4.2 Vehicle detection results using faster R-CNN with several RPNs

The performance of the present system using the proposed architecture of Faster R-CNN with several RPNs has been evaluated using various metrics, namely, *accuracy*, *precision*, and *recall*. *Precision* and *recall* are defined in Table 2. The values of these metrics have been found by computing the Intersection over Union (IoU) between the detecting and the ground-truth bounding boxes. If the IoU value is more than 0.8, then the detection has been considered as true positive. The vehicle detection results of the proposed system using Faster R-CNN with

**Table 6** Vehicle detection results using conventional Faster R-CNN with single RPN

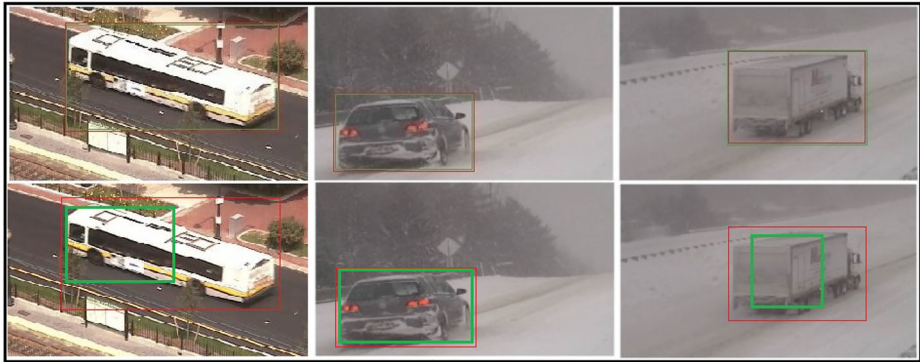| Dataset | Weather condition | Accuracy | Precision | Recall |
|---------|-------------------|----------|-----------|--------|
| DAWN | Fog | 84.56% | 84.38% | 84.48% |
| | Snow | 82.45% | 82.34% | 82.39% |
| | Rain | 82.21% | 82.08% | 82.15% |
| | Sandstorms | 82.25% | 82.13% | 82.19% |
| CDNet 2014 | Sunny | 88.23% | 88.04% | 88.11% |
| | Blizzard | 86.57% | 86.26% | 86.31% |
| | Snowfall | 82.26% | 82.11% | 82.18% |
| | Wet snow | 82.23% | 82.07% | 82.16% |
| LISA | Sunny evening | 88.37% | 88.28% | 88.34% |
| | Cloudy morning | 88.28% | 88.21% | 88.24% |
| | Sunny afternoon | 88.43% | 88.35% | 88.39% |

**Fig. 7** Qualitative result analysis on few test images using Faster R-CNN with several RPNs (top row) and the conventional Faster R-CNN with single RPN (bottom row). The red bounding box indicates the ground-truth and the green one indicates the detecting box

several RPNs in terms of accuracy, precision, and recall are presented in Table 3. Table 3 alsopresents the processing speed of the proposed system in Frames per second (fps) on all the three datasets. The proposed system has been implemented on a single Titan Xp GPU. Table 4 shows the precision of the proposed system against different orders of receptive field dimensions of four RPNs. The optimal set of values of various hyper parameters in Faster R-CNN with several RPNs structure is shown in Table 5. *Bayesian optimization technique* has been used to find this optimal set. Figures 3, 4 and 5 illustrate the correct detection of on-road vehicles using the proposed method on few test images from DAWN, CDNet 2014, and LISA datasets. Detected vehicles have been enclosed within green colored boxes in this figure. The Receiver Operating Characteristic (ROC) curve analysis of the proposed system is shown in Fig. 6.

### 4.3 Vehicle detection results using conventional faster R-CNN with single RPN

The experiment has also been carried out using conventional Faster R-CNN with single RPN to have a performance comparison with the proposed architecture. The vehicle detection results using conventional Faster R-CNN with single RPN are presented in Table 6 in terms

**Table 7** Comparative performance analysis with few existing studies on DAWN dataset

| Dataset | Reference | Method | Backbone | Average precision | Processing speed (in fps) |
|---------|-----------|--------|----------|-------------------|---------------------------|
| DAWN | Cai et al. [2] | Casc-RCNN | ResNet-101-FPN | 81.31% | 16 |
| | Hu et al. [18] | SINet | VGGNet-16 | 79.23% | 17 |
| | Hassaballah et al. [15] | LBP | NA | 82.46% | 18 |
| | Chabot et al. [4] | Deep MANTA | VGGNet-16 | 80.54% | 18 |
| | Xiang et al. [35] | SubCategory | ResNet-101 | 78.34% | 17 |
| | He et al. [17] | Mask RCNN | ResNeXt-101 | 81.74% | 17 |
| | Lin et al. [22] | RetinaNet | ResNet-101-FPN | 71.23% | 18 |
| | Redmon et al. [26] | YOLOv3 | DarkNet-53 | 88.54% | 24 |
| | Li et al. [21] | TridentNet | ResNet-101 | 75.42% | 18 |
| | Proposed method | Faster R-CNN with several RPNs | Faster R-CNN | **89.48%** | 20 |

**Table 8** Comparative performance analysis with few existing studies on CDNet 2014 dataset

| Dataset | Reference | Method | Backbone | Average precision | Processing speed (in fps) |
|---------|-----------|--------|----------|-------------------|---------------------------|
| CDNet 2014 | Cai et al. [2] | Casc-RCNN | ResNet-101-FPN | 83.04% | 16 |
| | Hu et al. [18] | SINet | VGGNet-16 | 80.82% | 17 |
| | Hassaballah et al. [15] | LBP | NA | 83.38% | 18 |
| | Hadi et al. [12] | Adaptive search window | NA | 90.49% | NA |
| | Chabot et al. [4] | Deep MANTA | VGGNet-16 | 81.46% | 18 |
| | Xiang et al. [35] | SubCategory | ResNet-101 | 79.94% | 17 |
| | He et al. [17] | Mask RCNN | ResNeXt-101 | 83.12% | 17 |
| | Lin et al. [22] | RetinaNet | ResNet-101-FPN | 72.54% | 18 |
| | Redmon et al. [26] | YOLOv3 | DarkNet-53 | 89.72% | 24 |
| | Li et al. [21] | TridentNet | ResNet-101 | 77.18% | 18 |
| | Proposed method | Faster R-CNN with several RPNs | Faster R-CNN | **91.20%** | 20 |

of accuracy, precision, and recall. Figure 7 shows qualitative result analysis on few test images using both Faster R-CNN with several RPNs and the conventional Faster R-CNN with single RPN.

The results shown in Tables 3 and 6 clearly state that the proposed architecture of Faster R-CNN with several RPNs performs much better than the conventional Faster R-CNN model with single RPN.

## 4.4 Comparison with the state-of-the-art results

Most of the studies in on-road vehicle detection field have not used publicly available datasets to evaluate the system performance, rather they relied upon self-generated non-public datasets due to the specific designed algorithms. So, the results of theproposed system cannot be compared with the results of the existing studies based on some non-

**Table 9** Comparative performance analysis with few existing studies on LISA dataset

| Dataset | Reference | Method | Backbone | Average precision | Processing speed (in fps) |
|---------|-----------|--------|----------|-------------------|---------------------------|
| LISA | Zhou et al. [39] | Shallow CNN | NA | 79.41% | 4 |
| | Cai et al. [2] | Casc-RCNN | ResNet-101-FPN | 85.21% | 16 |
| | Hu et al. [18] | SINet | VGGNet-16 | 83.45% | 16 |
| | Hassaballah et al. [15] | LBP | NA | 86.72% | 17 |
| | Chabot et al. [4] | Deep MANTA | VGGNet-16 | 84.28% | 17 |
| | Xiang et al. [35] | SubCategory | ResNet-101 | 81.82% | 17 |
| | He et al. [17] | Mask RCNN | ResNeXt-101 | 86.38% | 17 |
| | Lin et al. [22] | RetinaNet | ResNet-101-FPN | 74.62% | 17 |
| | Redmon et al. [26] | YOLOv3 | DarkNet-53 | 93.78% | 24 |
| | Li et al. [21] | TridentNet | ResNet-101 | 79.58% | 16 |
| | Proposed method | Faster R-CNN with several RPNs | Faster R-CNN | **95.16%** | 18 |

**Fig. 8** Few instances of incorrect detection of vehicles in (**a**) Wet snow, (**b**) Normal weather, and (**c**) Blizzard conditions. The red bounding box indicates the ground-truth and the green one indicates the detecting box

public datasets. A few existing similar vehicle detection methods have been evaluated on three public datasets (DAWN, CDNet 2014, and LISA) used in the present study and their results have been compared with that of the proposed system. The comparisons are presented in Tables 7, 8 and 9 for DAWN, CDNet 2014, and LISA datasets respectively. Average precision has been used as comparison metric in these three tables. The average precision has been computed by considering the individual precisions obtained from the test samples under each type of weather conditions.

### 4.5 Strengths of the proposed system

1. The proposed system detects on-road vehicles in good (sunny) as well as challenging weather conditions.
2. The present system can detect varying sized vehicles more successfully in comparison to the existing studies due to the presence of several varying dimensional RPNs in the Faster R-CNN structure.
3. The performance of the proposed system has been evaluated on three widely used public vehicle detection datasets.

### 4.6 Error analysis

For few test images, false positive results have been obtained. The present system could not detect the vehicles in these images properly; the detecting boxes have enclosed some other objects in these images, but not the vehicles. Few erroneous detections have been obtained in bad weather conditions like wet snow or blizzard environments due to the extreme poor visibility of the images and few have occurred in normal weather condition also where some other objects have obstructed the vision of vehicle in the image. Few erroneous outcomes are shown in Fig. 8.

## 5 Conclusion and future scope

Detecting and tracking on-road vehicles from natural scene images in bad weather conditions is a challenging task because of poor visibility of the target object. The present investigation proposes a novel method of on-road vehicle detection by incorporating several RPNs in conventional Faster R-CNN structure. The proposed strategy detects varying sized vehicles

more accurately due to the introduction of several RPNs in the present work. The results show that the proposed strategy outperforms conventionalFaster R-CNN as well as similar other state-of-the-art methods on DAWN, CDNet 2014, and LISA datasets in detecting varying sized vehicles in good as well as challenging weather conditions. The results also demonstrate that the processing speed of the proposed detection strategy is more than other existing systems except one (YOLOv3) on the aforesaid three datasets. There is still room for improvement of the performance of the present system and the future work will be devoted to achieve this goal. The attempt will also be made in future to further increase the processing speed of the vehicle detection system. Classifying the vehicles into different categories (small, medium, and heavy) in order to facilitate the collection of toll at toll plaza is another future direction of the present work. It is also intended to carry out the investigation on vehicle number plate recognition after detecting different varying sized vehicles.

# References

1. Bertozzi M, Broggi A, Castelluccio S (1997) A real-time oriented system for vehicle detection. J Syst Archit 43(1-5):317–325
2. Cai Z, Vasconcelos N (2018) Cascade R-CNN: Delving into high quality object detection, proceedings of the IEEE conference on computer vision and pattern recognition, Salt Lake City, USA, pp. 6154–6162
3. Cai W, Wei Z (2020) Remote Sensing Image Classification Based on a Cross-Attention Mechanism and Graph Convolution, IEEE Geosci Remote Sens Lett, pp. 1–5
4. Chabot F, Chaouch M, Rabarisoa J, Teuliere C, Chateau T (2017) Deep MANTA: A Coarse-to-Fine many-task network for joint 2D and 3D vehicle analysis from monocular image, Proceedings of the IEEE conference on computer vision and pattern recognition, Hawaii, USA, pp. 2040–2049
5. Chan Y, Huang S, Fu L, Hsiao P (2007) Vehicle detection under various lighting conditions by incorporating particle filter, proceedings of the IEEE intelligent transportation systems conference, Seattle, USA, pp. 534–539
6. Chellappa R, Qian G, Zheng Q (2004) Vehicle detection and tracking using acoustic and video sensors, proceedings of the IEEE international conference on acoustics, speech, and signal processing, Montreal, Canada, pp. 793–796
7. Chen D, Chen G, Wang Y (2013) Real-time dynamic vehicle detection on resource-limited mobile platform. IET Comput Vis 7(2):81–89
8. Cheon M, Lee W, Yoon C, Park M (2012) Vision-based vehicle detection system with consideration of the detecting location. IEEE Trans Intell Transp Syst 13(3):1243–1252
9. Dong E, Deng M, Tong J, Jia C, Du S (2019) Moving vehicle tracking based on improved tracking-learning-detection algorithm. IET Comput Vis 13(8):730–741
10. Fossati A, Schnmann P, Fua P (2011) Real-time vehicle tracking for driving assistance. Mach Vis Appl 22:439–448
11. Ghosh R, Thakre S, Kumar P (2018) A vehicle number plate recognition system using region-of-interest based filtering method, proceedings of the 2018 conference on information and communication technology. Jabalpur, pp 1–6
12. Hadi RA, George LE, Mohammed MJ (2017) A computationally economic novel approach for real-time moving multi-vehicle detection and tracking toward efficient traffic surveillance. Arab J Sci Eng 42:817–831
13. Haselhoff A, Kummert A (2009) A vehicle detection system based on Haar and triangle features. Proceedings of the IEEE Intelligent Vehicles Symposium, Xi'an, China, pp 261–266
14. Haselhoff A, Kummert A (2009) An evolutionary optimized vehicle tracker in collaboration with a detection system. Proceedings of the IEEE Intelligent Transportation Systems Conference, St. Louis, USA, pp 1–6
15. Hassaballah M, Kenk MA, Henawy IME (2020) Local binary pattern‐based on‐road vehicle detection in urban traffic scene. Pattern Anal Applic 23:1505–1521
16. Hassaballah M, Kenk MA, Muhammad K, Minaee S (2020) Vehicle Detection and Tracking in Adverse Weather Using a Deep Learning Framework, IEEE Trans Intell Transport Syst, pp. 1–13

17. He K, Gkioxari G, Dollár P, Girshick R (2017) Mask R-CNN, Proceedings of the IEEE conference on computer vision and pattern recognition, Hawaii, USA, pp. 2961–2969
18. Hu X, Xu X, Xiao Y, Chen H, He S, Qin J, Heng PA (2019) SINet: a scale-insensitive convolutional neural network for fast vehicle detection. IEEE Trans Intell Transp Syst 20(3):1010–1019
19. Kenk MA, Hassaballah M (2020) DAWN: Vehicle Detection in Adverse Weather Nature Dataset, arXiv preprint arXiv:2008.05402
20. Khairdoost N, Monadjemi SA, Jamshidi K (2013) Front and rear vehicle detection using hypothesis generation and verification. Sig Image Process 4(4):31–50
21. Li Y, Chen Y, Wang N, Zhang Z (2019) Scale-aware trident networks for object detection, arXiv: 1901.01892. [Online]. Available: http://arxiv.org/abs/1901.01892. Accessed 18 Dec 2020.
22. Lin TY, Goyal P, Girshick R, He K, Dollar P (2017) Focal loss for dense object detection, Proceedings of the IEEE international conference on computer vision, Venice, Italy, pp. 2980–2988
23. Matthews N, An P, Charnley D, Harris C (1996) *Vehicle detection and recognition in greyscale imagery*, control engineering practice, volume 4. Issues 4:473–479
24. Ming Q, Jo KH (2011) Vehicle detection using tail light segmentation, proceedings of the 6th international forum on strategic technology, Harbin, China, pp. 729–732
25. Mohamed A, Issam A, Mohamed B, Abdellatif B (2015) Real-time detection of vehicles using the Haar-like features and artificial neuron networks. Procedia Comput Sci 73:24–31
26. Redmon J, Farhadi A (2018) YOLOv3: An incremental improvement, arXiv:1804.02767. [Online]. Available: http://arxiv.org/abs/1804.02767. Accessed 18 Dec 2020.
27. Ren S, He K, Girshick R, Sun J (2015) Faster R-CNN: towards real-time object detection with region proposal networks. Proceedings of the Neural Information Processing Systems, Montreal, Canada, pp 91–99
28. Singh V, Srivastava A, Kumar S, Ghosh R (2019) A Structural Feature Based Automatic Vehicle Classification System at Toll Plaza, proceedings of the 4th international conference on internet of things and connected technologies. Jaipur, pp 1–10
29. Sivaraman S, Trivedi M (2010) A general active-learning framework for on-road vehicle recognition and tracking. IEEE Trans Intell Transp Syst 11(2):267–276
30. Sivaraman S, Trivedi MM (2014) Active learning for on-road vehicle detection: a comparative study. Mach Vis Appl 25(3):599–611
31. Sun Z, Bebis G, Miller R (2006) Monocular precrash vehicle detection: features and classifiers. IEEE Trans Image Process 15(7):2019–2034
32. Tian Y, Dong H, Jia L, Li S (2014) A vehicle re-identification algorithm based on multi-sensor correlation. J Zhejiang Uni Sci C 15(5):372–382
33. Wang Y, Jodoin PM, Porikli F, Konrad J, Benezeth Y, Ishwar P (2014) CDnet 2014:an expanded change detection benchmark dataset, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 387–394
34. Wen X, Shao L, Fang W, Xue Y (2015) Efficient feature selection and classification for vehicle detection. IEEE Trans Circuits Syst Video Technol 25(3):508–517
35. Xiang Y, Choi W, Lin Y, Savarese S (2017) Subcategory-aware convolutional neural networks for object proposals and detection, Proceedings of the IEEE winter conference on applications of computer vision, Santa Rosa, USA, pp. 924–933
36. Yan G, Yu M, Yu Y, Fan L (2016) Real-time vehicle detection using histograms of oriented gradients and AdaBoost classification. Int J Light Electron Opt 127(19):7941–7951
37. You H, Tian S, Yu L, Lv Y (2020) Pixel-level remote sensing image recognition based on bidirectional word vectors. IEEE Trans Geosci Remote Sens 58(2):1281–1293
38. Yu T, Shin H (2015) Detecting partially occluded vehicles with geometric and likelihood reasoning. IET Comput Vis 9(2):174–183
39. Zhou Y, Liu L, Shao L, Mellor M (2016) DAVE: a unified framework for fast vehicle detection and annotation, proceedings of the European conference on computer vision, Amsterdam, Netherlands, pp. 278–293