VAGAN: Vehicle-Aware Generative Adversarial Networks for Vehicle Detection in Rain*

Jian-Gang Wang, Senior Member, IEEE, Kong-Wah Wan, Wei-Yun Yau, Chun Ho Pang, Fon Lin Lai

Abstract—Vision-based vehicle detection under bad weather conditions is still a challenging problem. Adhesive raindrops on windshield have been known to diffract light and distort parts of the scene behind them. In this paper, we propose a Vehicle-Aware Generative Adversarial Networks (VAGAN) to improve vehicle detection from rain images. We train a Generative Adversarial Network (GAN) on image pairs, each comprising of an original rain image and one that is manually labelled with colored bounding boxes of the vehicles therein. The latter represents a fake version of the original image emphasizing the regions of interest. To further enhance vehicle awareness, we exploit the fact that the vehicle rear lights are usually turned on during rainy conditions to compute a saliency map of image, and use it formulate a background preserving constrain on the learning vehicle loss function. We show that this novel adversarial framework is able to generate new images with colored regions overlap over vehicles, hence effectively learning to differentiate image background from vehicles. The final vehicle detection in the generated images is not affected by image translation noise because we can simply use color segmentation to localize the vehicles. Experimental results on a large dataset show that our approach is an effective way to solve vehicle detection in rain images, achieving state-of-the-art performance.

I. INTRODUCTION

There has been impressive progress recently in deep learning based vehicle detection. However, relatively little attention was paid to the vehicle detection under bad weather conditions. In general, Autonomous Vehicle (AV) performs vehicle detection with an in-car camera and the raindrops that remain on the windshield pose as a nuisance since they hamper the detectability of the vehicles in a scene. As a result, much more effort has been placed on image enhancement and de-raining algorithms in the recent years. However, the help from de-raining is quite limited because the underlying image models to represent rain streaks or raindrop are very different from the images captured by an in-car camera. Furthermore, most existing de-rain processes are slow and compute intensive. Radar is relative insensitive to rain compared with camera. However, we have found that radar detections are affected by rain reflection on the road and often produce non-negligible level of false positives. In addition, the vehicle information content obtainable from radar sensor is limited because the returned point cloud is very noisy. For example, commercial radar devices such as Delphi [1] and mmWave TI [2] only provide a single point as output for each detected vehicle. A more comprehensive survey on on-road vehicle detection can be found in [3].

In this paper, we deviate from the conventional de-rain approach of aiming at better visual quality. Instead, we aim at improving vehicle detectability under rainy conditions by exploiting incorporated vehicle contours as well as the the fact that the vehicle rear lights are usually turned on during rainy conditions. We call it “vehicle-aware” and formulate it as an image-to-image translation problem. The key idea is that a new image could be generated from the raw rain image in which the vehicles are highlighted while the background is preserved well.

Unsupervised image-to-image translation [4–8] has received much attention due to the recent progress in generative adversarial networks (GAN). The advantage of the GAN [9] is the ability to learn mathematical functions that can map one domain of data to another, this ability has been adopted in image enhancement and domain adaptations applications with great success. However, when an image has several target instances, the translation process involves considerable shape changes. In this paper, we propose a novel GAN-based image-to-image translation to incorporate vehicle appearance and status in rain images. An image-to-image translation is learned automatically from a training set containing rain images and their counterparts which are manually labelled with colored bounding boxes of the vehicles therein. To further enhance vehicle awareness, we exploit the fact that the vehicle rear lights are usually turned on during rainy conditions, and compute a saliency map of the image. The resulting binarized saliency regions, corresponding to rear lights, are then used to formulate a background preserving constrain on the vehicle loss function. Once the translation is trained, given a rain image, the color bounding boxes of the vehicles therein can be predicted in the target image and the vehicles can be easily detected with the help of the predicted color bounding boxes.

To the best of our knowledge, we are the first to propose and implement a vision system to detect vehicles in heavy rainy conditions. Compared with the existing perception in rain approaches, the advantageous of our approach include: (1) No need to use extra hardware, e.g. radar; (2) The domain knowledge about the vehicles as well as the fact that the vehicle rear lights are usually turned on during rainy conditions are exploited to emphasize the vehicle awareness. This makes the image-to-image translation focus on the region of interest. (3) The detection accuracy under rainy condition is

*This research is supported by the Agency for Science, Technology and Research (A*STAR) under its <GAP(ETPL/18-GAP055-R20A)>. The authors are with the Institute for Infocomm Research (I2R), 1 Fusionopolis Way, #21-01 Connexis (South Tower), Singapore 138632 (phone: 65-6408-2481; fax: 65-6776-1378; e-mail: {jjwang, kongwah, wyyau, pang_chun_ho, flail}@i2r.a-star.edu.sg).
much higher than the de-rain assistant vehicle detectors because our vehicle-aware approach significant improves the vehicle detectability compared with the existing de-rain approaches which consider only visual quality.

The rest of the paper is organized as follows. Section II discusses the related work on image-to-image translation and de-raining. Section III discusses the proposed vehicle-aware generative adversarial networks. A color-based vehicle detector is discussed in Section IV. The benchmark on a database is discussed in Section V. Finally, the conclusion is given in Section VI.

II. RELATED WORK

Vehicle detection under heavy rain is one of the more challenging and unsolved problems in the field of autonomous vehicle research. There is relatively little literature about vehicle detection from heavy rain images [10] compared with large amount reports on vehicle detection from clear images.

The de-raining concept from single image or video has been extensively investigated. Garg and Nayer [11-12] did pioneering research on modeling, detecting and removing rain from video. Later, some approaches are proposed to de-raining from single image. These methods are not applicable to the conditions discussed in this paper because the images discussed in their approaches are quite different from our images captured form an in-car camera. The effect of rainfall on the windshield has not been considered in most of the previous de-rain approaches. The assumptions they made to build physics models are not true for our images.

Detecting and removing raindrops which adhered to a windscreen have been explored [13-15]. Roser et al. [13-14] models raindrop shape as sphere crown or Bezier curve. It is clear that the models cannot cover all kinds of raindrops in real applications. You et al. [15] propose a method to detect and remove raindrops from video. The authors acknowledged that the method does not work with highly dynamic raindrops.

Recently, conditional generative adversarial networks (GAN) [16] is used to do de-raining from image [17]. As pair of images (rain image, clear image) of the same scene are required in order to train a condition GAN, in their approach, a simulated rain image is generated from a clear image. The model obtained from the simulated images could not be applicable to real images obtained from in-car camera because the rain effects on the windshield is not simulated.

Mo et al. [18] proposed an instance-aware image-to-image translation where the instance is represented as vehicle segmentation mask. It aims to solve the problem when image translation faces multiple instance which have significant changes in shape. The method works well, however, the requirement of segmentation mask could be tough for our images captured under heavy rain conditions. Actually, the mask required by their method is nothing but the results what we want to resolve in this paper. Similar to [18], mask contrast-GAN [19] and Attention-GAN [20] require segmentation mask and not applicable for our applications.

III. VEHICLE-AWARE IMAGE-TO-IMAGE TRANSLATION

In order to detect vehicle from a rainy image, one could think about de-rain technology. However, the degradation caused by rain could be too serious to make the vehicles clear enough to be detected, especial in heavy rain conditions. In addition, the models used to describe the rain streak or raindrop [15] could not be accurate for images captured by an in-car camera. The de-rain algorithm is with high complexity and not suitable for online real-time applications, like autonomous vehicle.

In general, a large training database is required to effectively training a network to achieve good performance. However, it is expensive to collect a large database. Image-to-image translation is a technology to consider this problem and generate more data from the existing data. However, the existing image-to-image translation can produce only limited plausible alternative data.

In this paper, we focus on improving vehicle detection rate rather than visual quality. The idea is that some image-to-image translation technology could generate new images from a rain image such that the vehicle can be detected from the new image more robustly than from the original image. The flowchart of our proposed approach is shown in Fig. 1. Given a rainy image, our goal is to lean mapping function (we call it VAGAN) which translates the image to a new image in which the background is preserved and the vehicle is highlighted with their bounding boxes. In our approach, the vehicle awareness is further improved by exploiting the fact that the vehicle rear lights are usually turned on during rainy conditions, and a saliency map of the image is used to learn the mapping function.

One of the disadvantages of the existing image-to-image translation is that all pixels on the original image are translated with the same criteria and the interesting parts are not emphasized. Noted this, in this paper, domain knowledge about the vehicle, including bounding boxes and rear lights, are incorporated into the image-to-image translation.

![Flowchart of proposed approach](image)

Generative Adversarial Network (GAN) have been of particular interest in the deep learning community, particular for image generation. It includes a generator which is
responsible for generating images which is similar to the data already seen and a discriminator which is responsible for telling apart a real image.

In this paper, we adopt GAN as an image translator to generate images for improving vehicle detection accuracy. In particular, a modified GAN is proposed to translate rainy images to new images in which the targets are prominent by their bounding boxes. We use conditional GAN [21] because it appears that the paired training samples are always available for our approach. Besides the rainy images, the saliency map of the images is also used to construct background preserving loss function for the generator. Although much information about the vehicles are lost in rainy images due to the heavy rain, the rear lights are still observable under most cases and become an important clue to locate vehicles from an image. Actually, they are used to follow vehicles or observe vehicles around the host vehicle by drivers. Furthermore, a color-based bounding box detector is developed in our approach to detect vehicles from the new images obtained from the VAGAN.

A. Vehicle-Aware GAN

Assume \( \{x, S_x, y, S_y\} \) represents a training sample, where \( x \) and \( y \) represent images, \( y \) are the images similar to \( x \) but with vehicle bounding boxes in specific color; \( S_x \) and \( S_y \) are saliency maps of \( x \) and \( y \), respectively.

A conditional GAN learns a mapping from observed image \( x \) and random noise vector \( z \) to \( y \). Define a generator, \( G(x, z) \rightarrow y \). The \( G \) is trained to produce outputs that cannot be distinguished from “real” images by an adversarially trained discriminator, \( D \), which is to do as well as possible at detecting the generator’s “fakes”.

The generative adversarial loss [21] is defined as follows.

\[
L(G, D) = L_{GAN}(G, D) + \lambda E_{x,y,z}[\|y - G(x, z)\|_1] \quad (1)
\]

where

\[
L_{GAN}(G, D) = E_{x,y}[\text{log}(D(x, y))] + E_{x,z}[\text{log}(1 - D(x, G(x, z)))] \quad (2)
\]

In this paper, a new background preserving loss is proposed which uses saliency map image to represent loss. The new loss is defined as a pixel-wise weighted \( l_1 \)-loss where the background is with weight 1 and vehicle are with 0. Only the pixel in background in both original and translated ones are considered. For original \((x, S_x)\) and translated \((y, S_y)\), where \( S_x, S_y \) are binary represented saliency map,

\[
L_o = \|w(S_x, S_y) \odot (x - y)\|_1 \quad (3)
\]

where \( L_o \) represents background preserving loss enforces to vehicle while keeping background. \( \odot \) is the element-wise product.

The weight \( w(S_x, S_y) \) in (3) is defined as follows.

\[
w(S_x, S_y) = \begin{cases} 
1 & \text{if } S_x \neq S_y \\
0 & \text{else}
\end{cases} \quad (4)
\]

Finally, the learning procedure aims at:

\[
G^* = \arg \min_G \max_D L_{GAN}(G, D) + \lambda E_{x,y,z}[\|y - G(x, z)\|_1] + \lambda_o L_o \quad (5)
\]

where \( \lambda_o \) controls the relative importance of our newly proposed background preserving loss \( L_o \).

B. Salience region

As discussed in the last section, one of contributions of this paper is that the saliency map of image is exploited to introduce a background preserving loss function for the proposed vehicle-aware GAN. Although much vehicle information is lost under heavy rain conditions, it is a fact that the rear lights are turned on in rain and could be an important clue for a driver to follow vehicles ahead. For autonomous vehicle, we believe that the rear lights information is helpful to design a vision-based vehicle detector.

As for rear lights, it is naturally to find the region of interest on images based on color information. Different color space saliency map has been proposed for extracting saliency map from an image. Without loss of generality, in this paper, a non-parameter model is proposed to extract saliency map of rainy image. We use RGB color space rather than HSV [22] because we want to show that our approach is relatively insensitive to the quality of the saliency region.

We assume that there are three kinds of rear lights: red, white and amber. The saliency score of an image consists of three components: red score, \( S_r \), white score, \( S_w \), and amber score, \( S_a \). If \( N_d(i, j) \) are the pixels around \((i, j)\) and with a maximal distance, \( d \), from \((i, j)\), then

\[
S_r(i, j) = \sum_{(i', j') \in N_d(i, j)} H_r(i', j') \quad (6)
\]

where \( H_r \) is the normalized histogram of the red light. The overall histogram of \( r \) is as follows.

\[
H = \max(H_r, H_w, H_a) \quad (7)
\]

Then overall saliency map, \( S \), can be computed as,

\[
S(i, j) = \sum_{(i', j') \in N_d(i, j)} \max(H(i', j')) \quad (8)
\]

The saliency scores are computed recursively with the histogram models until it is less than a threshold, \( T \). The type of the pixel is set to be the color corresponding to the maximum saliency score. In our experiments, \( d \) is set to be 2, \( T \) is set to be 0.2. It should be noticed that these settings are selected without special tuning.

![Example Images](image_url)

Figure 2. Some examples of the proposed saliency regions. Left: original images; middle: saliency maps; right: the images that retain the original images corresponding to the binarized saliency regions
By doing so, the majority of pixels could be filtered out by the overall saliency score, and the types of remaining small portion of pixels could be determined by individual saliency models. Some examples of the proposed saliency model are shown in Fig. 2. It can be seen that the rear lights or traffic lights are kept while the background are removed in the binarized saliency regions.

In our approach, the input includes rainy images and their binary represented saliency regions and the output image domain includes rainy images with bounding boxes (manually labelled) and their binary represented saliency region. An example is shown in Fig. 3.

Once the vehicle-aware model is trained, we can translate a test image $X$ to a new image $Y$ with the GAN, an example is shown in Fig. 4. It can be seen that the proposed vehicle-aware conditional GAN can achieve good translated image where the predicted bounding boxes are close to their ground-truth.

Some image-to-image translation results are shown in Fig. 5. The original images and the predicted images are shown on the left and right, respectively. We can see that the vehicles (bounding boxes) are predicted well. Although some artefacts are generated on the predicted images due to the heavy rain, they will not affect the detection of the vehicles because only the color bounding boxes will be used to detect vehicles.

IV. COLOR-BASED VEHICLE DETECTION FROM TRANSLATED IMAGES

On the translated images obtained by our vehicle-aware GAN image-to-image translation presented in Section III, the predicted bounding boxes of vehicles have the same uniform color with the training data. Although a vehicle detector can be trained to detect the vehicles, a simple color-based shape detector should be more robust because the effects of the possible artefact can be prevented by using the known color.

There have been more color shape detection approaches in the literature. In this paper, we have proposed a HSV color based approach to detect rectangles from an image as the bounding boxes of the vehicle are in known color on the new images obtained by our vehicle-aware GAN.

First of all, the image is converted from RGB to HSV. In this paper, green color is adopted to represented bounding boxes of the vehicle. As for green color, the range of the values corresponding to the three channels, $(H_l, H_u)$, $(S_l, S_u)$ and $(V_l, V_u)$ can be estimated from some training images. Once a translated image is given, a pixel $i$, with value $(p_H, p_S, p_V)$, on the bounding box of the detected vehicles can be extracted by using the above ranges. A binarized representation image which contains only vehicles’ bounding boxes can be obtained as follows.

\[
B_i = \begin{cases} 
1 & \text{if } (H_u < p_H < H_l) \cap (S_u < p_S < S_l) \cap (V_l < p_V < V_u) \\
0 & \text{otherwise}
\end{cases}
\]

for all $i$ on the bounding box of the detected vehicles (9)
The resultant pixels on the binarized image with value 1 can be grouped as rectangles by using OpenCV function. As the width of the bounding boxes maybe more than one pixel, the function has three options to group rectangles from an binarized image by using: (1) external pixels only; (2) internal pixels only and (3) using both external and internal pixels. As the contour extraction used external contour pixels, it encounters a problem that the closer vehicles could be merged as a single vehicle. To solve this problem, we are proposing in our approach that the final rectangles shall be obtained by subtracting the internal operation results from the external operation results.

In our experiments, the range of green color in HSV is defined as follows.

\[(H_l, S_l, V_l) = (65, 60, 60)\]
\[(H_u, S_u, V_u) = (80, 255, 255)\]

The translated image is thresholded among upper and lower range of green color in HSV. An example of the proposed color-based vehicle detection algorithm is shown in Fig. 6. The detection results based on all bounding boxes pixels obtained by our vehicle-aware GAN is shown in Fig. 6(a); the results obtained using only internal pixels of the bounding boxes are shown in Fig. 6(b); the final vehicle detection results, obtained by subtracting (b) from (a), are shown in Fig. 6(c).

![Image](image.png)

(a)  (b)  (c)

Figure 6. Color-based vehicle detection

Once vehicles are detected, they are classified into different categories by a CNN classifier. In this paper, a classifier is trained with five class vehicles: car, bus, truck, van and lorry.

V. EXPERIMENTAL RESULTS

Mean average precision (mAP) is a standard metric in evaluate the accuracy of vehicle detectors like YOLO [23] or SSD [24]. Average precision computes the average precision value for recall value over 0 to 1. It sounds complicated but actually pretty simple as we illustrate it with an example. Precision measures how accurate is detector’s predictions. i.e. the percentage of vehicle detector’s predictions are correct. Recall measures how good you find all the positives. They are defined as equations (13) and (14).

\[
\text{Precision} = \frac{TP}{TP + FP} \tag{10}
\]
\[
\text{Recall} = \frac{TP}{TP + FN} \tag{11}
\]

where \(TP\) represents number of the True Positive samples, \(FP\) represents the number of the False Positive samples and \(FN\) represents the number of the False Negative samples.

In this section, the quantitative analysis of our method in terms of precision and recall is conducted. For this purpose, a large database has been collected using our autonomous vehicle. We use IoU (intersection over union) to measure how much our predicted boundary overlaps with the ground truth (the real vehicle boundary). In our evaluation, the prediction is correct if IoU \(\geq 0.5\).

In order to evaluate the performance of the detector, a database includes 3067 rain images is collected. The vehicles on the images are categorized into five classes. The vehicles on the images are manually labeled and used as ground truth. The number of vehicles for each class are: car (7510), bus (465), lorry (47), truck (165) and van (299).

![Image](image.png)

Figure 7. The detection results

![Image](image.png)

Figure 8. Average precision and mAP

![Image](image.png)

Figure 9. Average miss rate
The detection results, include number of false positive (FP) and true positive (TP) are shown in Fig. 7. The number of false negative (FN) can be computed from the number of vehicles per class and the detection results. The precession and recall can be computed as equations (10) and (11). The average precision and average miss rate are shown in Fig. 8 and 9, respectively.

Some detection results are shown in Fig. 10 where the detection results are shown in red color and ground truth are shown in green.

Figure 10. Some detection results. Red: detection results; green: ground truth.

The experimental results have shown that the proposed vehicle-aware GAN has improved the vehicle detection accuracy in rain conditions. The mAP given in Figure 8 is comparable with the state-of-the-art vehicle detectors, e.g. YOLO [23] or SSD [24], which reported their results on clean images.

VI. CONCLUSION AND FUTURE WORK

The vehicle detection under bad weather conditions is one of challenging problems in the field of autonomous vehicle research. Most of the de-raining approaches aim at improving visual quality instead of vehicle detection. In this paper, a vehicle-aware image-to-image translation is proposed to detect vehicle from a single rainy image. By incorporating vehicle information (represented as bounding boxes and saliency map), a predicted image is generated from the image-to-image translation and vehicle can be detected by a color-based vehicle detector. The use of saliency map can improve awareness of the rear lights during image translation. By using the translated image, the vehicle detection accuracy under heavy rain conditions has been improved significantly.

The experimental results on a rain database collected from internet and our autonomous vehicle have shown that the proposed approach can achieve detection accuracy comparable with the state-of-the-art vehicle detector on clean images.

The extension of the vehicle-aware generative adversarial networks presented in this paper to the general object (e.g. pedestrian) detection, lane detection and traffic light recognition [22] could be investigated in the near future because they are needed to operate an autonomous vehicle under rain conditions.

REFERENCES