Towards Trustworthy Perception Information Sharing on Connected and Autonomous Vehicles

Jingda Guo
Computer Science and Engineering
University of North Texas
Denton, TX, USA
jingdaguo@my.unt.edu

Qing Yang
Computer Science and Engineering
University of North Texas
Denton, TX, USA
qing.yang@unt.edu

Song Fu
Computer Science and Engineering
University of North Texas
Denton, TX, USA
song.fu@unt.edu

Ryan Boyles
Computer Science
California State University - Sacramento
Roseville, CA, USA
ryanboyles14@gmail.com

Shavon Turner
Computer Science
Grambling State University
Grambling, LA, USA
turnershavon14@gmail.com

Kenzie Clarke
Computer Science
Texas Christian University
Fort Worth, TX, USA
kenzie.clarke@tcu.edu

***Abstract*—Sharing perception data among autonomous vehicles is extremely useful to extending the line of sight and field of view of autonomous vehicles, which otherwise suffer from blind spots and occlusions. However, the security of using data from a random other car in making driving decisions is an issue. Without the ability of assessing the trustworthiness of received information, it will be too risky to use them for any purposes. On the other hand, when information is exchanged between vehicles, it provides a golden opportunity to quantitatively study a vehicle's trust. In this paper, we propose a trustworthy information sharing framework for connected and autonomous vehicles in which vehicles measure each other's trust using the Dirichlet-Categorical (DC) model. To increase a vehicle's capability of assessing received data's trust, we leverage the Enhanced Super-Resolution Generative Adversarial Networks (ESRGAN) model to increase the resolution of blurry images. As a result, a vehicle is able to evaluate the trustworthiness of received data that contain distant objects. Based on the KITTI dataset, we evaluate the proposed solution and discover that vehicle's trust assessment capability can be increased by $11 - 37\%$, using the ESRGAN model.**

***Index Terms*—Connect vehicles, trustworthy information sharing, object detection, autonomous vehicles, image super-resolution**

I. Introduction

A fully autonomous vehicle (AV) is a vehicle that can guide itself without human interaction [**?**], which is commonly referred to as driverless car, robot car, or self-driving car. Although the reality of autonomous vehicles being deployed on our roads in a massive scale is a ways off, many connected vehicles already exist [**?**]. Connected vehicle (CV) technologies allow vehicles to communicate with each other and the world around them [**?**]. AV and CV are two key enabling techniques for future transportation system, which aims at reducing traffic accidents, enhancing quality of life, and improving efficiency. The differences between the CV and AV are often blurred, however, it is clear that these two techniques are complementary to each other. CV allows autonomous vehicles to exchange real-time sensor information to each other, which is extremely useful to extending autonomous vehicle's field of view [**?**]. The extended field of view on autonomous vehicles is beneficial at times where there are occlusions preventing a complete perception of the environment. On the other hand, AV is a better platform to manifest the benefits of CV technology as the massive amount of information exchanged among vehicles can only be processed by computer (not human) in real time [**?**].

One killer application of CAV technology is called precise cooperative perception [**?**], which enables real-time perception information to be shared amongst vehicles. Perception for autonomous vehicle is defined as the ability of a vehicle to collect information and extract relevant knowledge from the environment, developing a contextual understanding of environment, such as where obstacles are located, detection of road signs/marking, and categorizing data by their semantic meaning [**?**]. As object detection is a fundamental function of autonomous vehicle's perception functionality, in this paper, we focus on achieving trustworthy information sharing among CAVs where vehicles can share their object detection information to each other. Based on various object detection algorithms, a vehicle is able to detect in real-time the locations and types of objects around it. Such information will be transmitted to nearby vehicles. When a vehicle receives the object detection results from others, it will evaluate the trustworthiness of the received information, before the data is fused with the local ones.

A. Problem Statement

Unlike traditional server-oriented systems, e.g., Internet or cloud computing, in future CAV systems, an autonomous vehicle plays both the roles of information provider and information consumer. The sensors on autonomous vehicles (or roadside units), however, may be unreliable and vulnerable to physical attacks. Therefore, establishing trust in

978-1-7281-6059-7/20/$31.00 ©2020 IEEE
DOI 10.1109/MetroCAD48866.2020.00021

a broad range of vehicles, across dispersed settings and at massive scale, is an extremely important but also a technically challenging problem. Traditional security techniques generally protect vehicles from malicious attacks, by restricting access to only authorized ones, e.g., via the public key infrastructure (PKI) [**?**]. In future CAV systems, however, an autonomous vehicle will need to frequently protect itself from those that offer data/information, so the problem in fact is reversed. It is possible that an authorized vehicle acts deceitfully by providing false/misleading information, due to its defect hardware, software bugs, or even for selfish purposes. A trust system, on the other hand, can provide protections against these threats. Therefore, in this paper, we concentrate our research on how to achieve a trustworthy perception information sharing among connected and autonomous vehicles.

B. Limitation of Prior Art

There has been a flurry of research on trust modelling and trust management for vehicular ad hoc networks (VANET) [**?**]. The fundamental idea of these works is to let vehicles evaluate the trustworthiness of data shared from other vehicles, and based on the evaluation results, determine the trust of the corresponding vehicles. Obviously, the context of trust here is defined as the ability of a vehicle providing trustworthy or reliable data to others. If two vehicles have no data shared previously, they can leverage other vehicles, whose trust values are known, to infer other's trustworthiness. Several accurate trust model and trust computation algorithms are proposed in the literature; however, it is not clear how a vehicle can effectively evaluate the trust of another vehicle. Theoretically, a vehicle can assess the trustworthiness of received data if and only if it can directly measure/sense the same information. Taking object detection as an example, to verify the information shared from others, a vehicle must be able to successfully detect the objects (shared from others). This is a challenging problem, particularly, when the verifying vehicle cannot clearly view the objects, e.g., when they are far away from the vehicle.

C. Proposed Solution

To achieve trustworthy information sharing among CAVs, the first issue is to properly define a vehicle's trustworthiness. Aiming at the cooperative perception application, a vehicle's trust can be defined as its ability of correctly detecting (and classifying) objects captured by its sensors, e.g., cameras. Note that a vehicle does not have to be malicious to lie, i.e., its sensors could simply be faulty. In addition, untrustworthy object detection results may be sent by a vehicle simply because it employs a low-precision object detection model. As a result, a vehicle may not intentionally lie about their perception results, however, it still sends untrustworthy information, e.g., due to blocked field-of-view of its cameras.

With the trust context in place, the second challenge is to devise a mechanism that allows vehicles to effectively and efficiently assess the trust of other vehicles. To address this issue, we propose to use the Dirichlet-Categorical (DC) trust model to measure vehicles' trust, based on the quality and quantity of data sharing among them. Specifically, the object detection results shared from a vehicle are evaluated by a receiving vehicle, leveraging its own perception capability. As a result, the shared information is classified into three categories: trustworthy, untrustworthy, and uncertain. Based on the evaluation results, information provider's trustworthiness is quantified using the DC model.

To assess the trustworthiness of sharing information, a vehicle must be able to clearly observe and detect the objects shared from others. In many circumstances, the vehicle may have a hard time to detect and classify objects that are far away from it. To successfully detect these distant objects, we propose to use the Enhanced Super-Resolution Generative Adversarial Networks (ESRGAN) to increase the resolution of images that contain distant objects. With the enhanced images, a vehicle could detect more objects that are claimed to be undetectable before. As such, more accurate trust evaluation can be achieved on received data, therefore, precise trust assessment of the information sender can be accomplished.

II. Towards Trustworthy Perception Information Sharing for CAVs

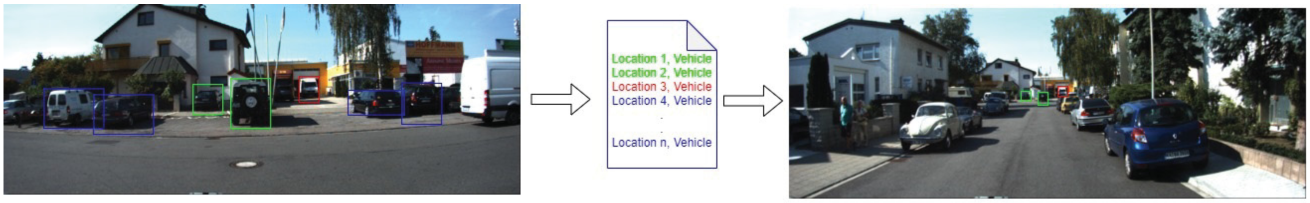
To enable trust assessment on CAVs, we leverage the DC trust model to quantify a vehicle's trust, based on the nature of the data it shares with others (section II.A). Shared information is evaluated by receiving vehicles and grouped into three categories: trustworthy, untrustworthy and uncertain (section II.B). This process is called trust evidence collections, which provides necessary data to conduct vehicle's trust assessment (section II.C). When direct trust assessment is deemed impossible, trust inference is employed to estimate a vehicle's indirect trust (section II.D). With both measured and inferred trust, a trustworthy perception information sharing system becomes possible on CAVs where only data shared from trustworthy peers (vehicles) are processed by receiving vehicles.

A. Vehicle Trust Model

Due to its simplicity, the Dirichlet-Categorical (DC) trust model [**?**] is adopted to quantify and compute the trust value of a vehicle. According to the DC model, a vehicle only needs to count the number of positive, negative, and uncertain evidences collected from other vehicles to assess their trustworthiness. From a trustor vehicle i's perspective, a trustee vehicle j's trust can be modeled as a DC distribution that is represented as an opinion.

$$\omega_{ij} = \langle \alpha_{ij}, \beta_{ij}, \gamma_{ij} \rangle \, | a_{ij}.$$

Here, ω_{ij} denotes i's opinion on j's trust, or i's trust in j behaving as expected in the future. The parameters $\alpha_{ij}, \beta_{ij}, \gamma_{ij}$ refer to the amounts of observed positive, negative and uncertain evidence, respectively. a_{ij} is a constant formed from an existing impression without solid evidences, e.g. prejudice, preference and general opinion obtained from hearsay. For example, if i always distrusts/trusts a vehicle j from a certain



(a) Sender (b) Message (c) Receiver

Figure 1: Illustration of perception information sharing among connected and autonomous vehicles. (a) Sender processes its camera data to determine the locations and types of nearby objects. (b) Locations and types of objects detected by the sender are encapsulated into a message which is then sent to the receiver via DSRC (dedicate short range communication). (c) Based on the object detection results, receiver is able to verify the trustworthiness of the received information.

automobile manufacturer, then a_{ij} will be smaller/greater than 0.5. As vehicles are interacting with each other on the road, e.g., sharing information for achieving cooperative perception, more evidence will be accumulated to derive a more accurate trust assessment of vehicles.

B. Trust Evidence Collection

When a vehicle receives information from others, it could potentially verify the trustworthiness of the information, such as the information provider. As shown in Fig. 1, to realize cooperative perception, a vehicle (sender) sends a message of its object detection results to another vehicle (receiver). In the transmitted message, a list of detected objects along with their locations are shared to the receiver(s). If there is an overlapping area between the views of cameras on both the sender and receiver, the receiver can use its own object detection results to determine if the sender is transmitting trustworthy information. Be checking the received information, in the form of $< L, T >$, the receiver can obtain three possible outputs: (1) received information is trustworthy, (2) not trustworthy, (3) or uncertain. In the message, $L = (\{x_1, y_1, z_1\}, \{x_2, y_2, z_2\})$ provides the location of an object, in a 3D coordinate system; $T = (V|P|C)$ indicates whether a vehicle (V), pedestrian (P), or cyclist (C) is detected. As shown in Fig. 1, for objects that can be detected by both the sender and receiver, we mark them using green boxes. Because the receiver can observe and detect these objects, it is able to check whether the received information is trustworthy or not. If there is any error in the location or type of the objects, the received information (of these objects) is treated untrustworthy. If there exist objects that are out of the receiver camera's scope, in this case, the receiver cannot verify the received information (of these objects), which is considered uncertain. It is worth mentioning that although the object marked in a red box is observable by both the sender and receiver, because it is far away from the receiver, the receiver fails to detect it. As such, the received information of this object will be considered uncertain as well.

C. Vehicle Trust Assessment

Based on collected (trustworthy, untrustworthy, and uncertain) evidence, a vehicle makes use of the DC model to assess the trust of other vehicles. For a certain vehicle j, e.g., the sender in Fig. 1, let's assume it currently detects n objects and the detection results are shared with vehicle i. Of out the n objects, we assume vehicle i can successfully verify r_{ij} of them, i.e., they are considered trustworthy. Similarly, we use s_{ij} to denote the number of objects that are decided as untrustworthy by vehicle i. Then, we can use $t_{ij} = n - r_{ij} - s_{ij}$ to denote the number of uncertain objects, i.e., those not verifiable by vehicle i. Based on the DC model, vehicle i could form an opinion about vehicle j's trust, i.e. $\omega_{ij} = \langle \alpha_{ij}, \beta_{ij}, \gamma_{ij} \rangle$ where $\alpha_{ij} = r_{ij}$, $\beta_{ij} = s_{ij}$, and $\gamma_{ij} = t_{ij}$. As the DC trust model is established upon historical evidence, if vehicle i receives more information from j, it could aggregate these new evidences with old ones to derive a more accurate estimate of vehicle j's trust. For example, if vehicle i collects another set of objects $r'_{ij}, s'_{ij}, t'_{ij}$, then ω_{ij} is updated as follows: $\alpha_{ij} = r_{ij} + r'_{ij}$, $\beta_{ij} = s_{ij} + s'_{ij}$, and $\gamma_{ij} = t_{ij} + t'_{ij}$. For a given trust opinion, e.g., $\omega_{ij} =< \alpha_{ij}, \beta_{ij}, \gamma_{ij} >$, vehicle i can assess vehicle j's trust by computing the expected belief of this opinion [**?**]. With the trust assessment procedure, a vehicle is able to detect the faultiness of another vehicle, e.g., noticing repeated "blind spots" in its shared data.

D. Vehicle Trust Inference

If a vehicle receives information from a "stranger" whom it did not have interactions with previously, the vehicle will start the trust assessment procedure by verifying the trustworthiness of shared data. If the vehicle is unable to verify the received data, e.g., due to sensor faults or limited field of view, it will infer the sender's trust, based on others' recommendation. Based on the DC model, if vehicle i trusts vehicle j and j trusts vehicle k, then i can derive an indirect trust of k, even if i did not interact with j before. This process is called trust propagation, which enable trust assessment of vehicles that are not encountered with previously. Details about trust propagation and trust fusion can be found in [**?**]. As a vehicle may have multiple "friends" suggesting another vehicle's trust, these recommendations need to be fused into a consensus one by aggregating the evidence from each suggested trust opinion. This process is named trust fusion, which combines multiple trust opinions to derive a single one. Leveraging trust propagation and trust fusion, a vehicle is able to assess the trust of every piece of received information, thus realizing a trustworthy cooperative perception system.

To achieve cooperative perception, when received data is fused into local ones, the receiver could consider historic trust information from specific vehicles to discount their data. This is different from existing solutions which treats all inputs as equally valid and true. The trust value of a sender could be used as the weight to reflect its trust and the confidence of its data. For vehicles whose trust value is lower than a certain threshold, all their data would be treated untrustworthy and omitted in cooperative perception.

III. Verification of Distant Objects

When an object is far away from a vehicle, the distant object may be undetectable by the vehicle, thus affecting its trust assessment of the sender vehicle(s). To tackle this issue, we propose to use the Enhanced Super-Resolution Generative Adversarial Networks (ESRGAN) [?] to improve the resolution of distant objects' images, thus achieving a better object detection and classification performance.

A. Enhance Distant Object's Resolution

As shown in Fig. 2, a vehicle is able to detect most of the objects in this image. The numbers shown on top of each box indicates the detection score of the corresponding objects. It is worth mentioning, however, there are objects (marked in green box) that are not detected by the vehicle. The reason of the misdetection is that these objects are too far away from the vehicle, i.e., these objects' resolution is too low to be detected. For example, the size of these objects is typically less than 100×100; however, most state-of-the-art image classification models requires an input size of 224×224. Simply classifying these objects with pre-trained models will not offer meaningful results. To enhance the object detection performance, ESRGAN is employed to increase the resolution and enhance the details of distant objects. ESRGAN is an enhanced version of SRGAN [?], leveraging the discriminative network to discriminate the reality of generated images from generative network. By removing all batch normalization layer and introdcue the dense connection between Residual Blocks, ESRGAN is able to generate more realistic images after super-resolution.

B. Verification of Distant Objects

After receiving the object detection results from other vehicles, a receiving vehicle is able to crop out from its local image several regions that are reported to contain objects. Within these regions, if there are objects detected by the vehicle, then the trust assessment process is carried out. Otherwise, the vehicle employs ESRGAN to enhance the resolution of the cropped regions, i.e., adding enough details/features to improve object detection performance. Then, the enhanced image will be fed into existing object detection modules to check if any objects are detected in these regions. Depending on detailed implementations, it is possible to combine the object detection and classification into one module. Nevertheless, the vehicle is able to produce a new object detection results for these regions. Together with its original object detection results, the vehicles are able to verify whether the information provider is telling the truth. The entire verification process is illustrated in Fig. 3, where ESRGAN can be replaced by other resolution-enhancement solutions, which will be investigated in our future works.

As shown in Fig. 4, we can clearly see the difference between the original image and the enhanced image. As more details are added into the enhance image, more features in the image can be extracted by the object detection/classification module. In our experiments, the size of most enhanced images is increased from 10KB to 150 KB. The actual resolution of enhanced images can vary, due to size of the original input image. In order to achieve a better super-resolution performance, we customize the settings of ESRGAN and retrain our own model on the KITTI dataset [?]. Details of the training and testing process can be found in the next section.

IV. Experiments and Result Analysis

To evaluate the proposed solution's performance, we designed our experiments by assuming two vehicles exchange object detection results to each other. Due to space limitation, we only present how ESRGAN can enhance object detection performance. In order to run our experiments, we need the ESRGAN, prepared for generating super-resolution images, two pre-trained neural networks (one for image classification and one for object-detection), and a dataset to test on. The image classification model we used was the ResNeXt-101 32s48d model, which was pre-trained on ImageNet. The object detection model we used was YOLOv3, which was pre-trained on the COCO dataset. We trained our ESRGAN on images from the KITTI 2D Object Detection Dataset, because it incorporated a lot of various vehicles in the wild.

A. Training Dataset and Training Details

The data pre-processing for training our super-resolution model can be mainly summarized into two steps. Firstly, images from the KITTI dataset are very large (1382*512), so directly using them in training is not efficient and may result in a degraded performance of our model. To address this issue, we divided each image into 30 small images, with the size each sub-image being 100×100. This particular size is chosen because it reflects the size of image that contains distant objects. As such, we expect to achieve a similar super-resolution result on our blurry images containing distant objects. Secondly, to obtain low-resolution images for training our GAN model, we follow the instruction of the original ESRGAN [?] to generate 1/4 upsample low-resolution sub-images for training.

The processor of our workstation is Intel(R) Xeon(R) E3-1270v6 and the graphic processing unit (GPU) is 8 NVIDIA Tesla K80. We implement this work on Python 3.7 and Pytorch 1.1. We trained our model over 225000 iterations with an approximate 50-hour training time.

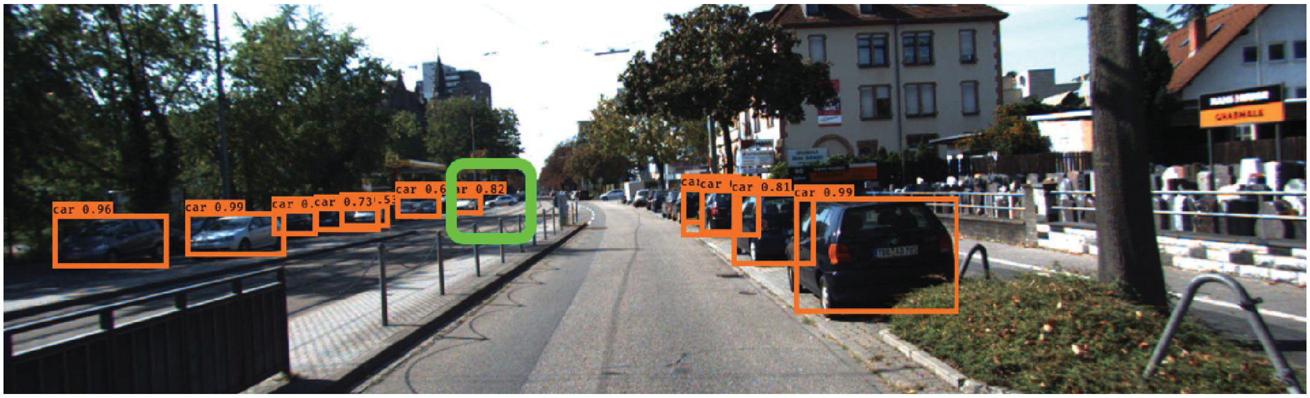


Figure 2: An illustration of the existence of blurred objects in a vehicle's camera data.

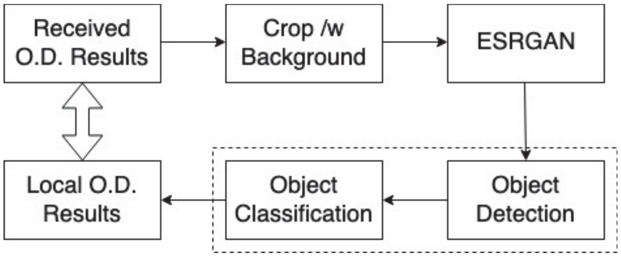


Figure 3: Flow chart of distant object verification using ESRGAN. Based on the received object detection (O.D.) results for other vehicles, distant objects from the receiving vehicle are cropped from the original image. The cropped image contains enough background information to ensure the ESRGAN model works effectively.

B. *Object Detection Improvement*

We evaluate our super-resolution results obtained from ESRGAN, compared with the blurry images detected by the ResNeXt101 model pretrained on Imagenet. In our experiments, all classes in Imagenet related to vehicles (e.g., cab, minivan or sport car) are considered as correct labels. The evaluation results are shown on Table I. Distant objects (vehicles) are difficult to be classified correctly on ResNeXt101, which only has a TOP-1 accuracy of 45.71% and Top-5 accuracy of 82.86%. After super-resolution, the classification accuracy improved significantly, i.e., a Top-1 accuracy of 62.01% and Top-5 accuracy of 91.57%. Our model shows a remarkable improvement for finding undetected or distant vehicles on CAVs.

Table I: Accuracy Comparison on ResNeXt101

Classification	TOP-1	TOP-5
Low-Resolution	45.71%	82.86%
High-Resolution	62.01%	91.57%

V. Discussions

In our experiments, we also try to use our model to improve the detection precision by image super-resolution. Considering that simply applying super-resolution on a whole image is resource consuming and unrealistic in reality, we only apply on part of image which is distant from the sensing vehicle. Figure 4 shows the original detection result on a KITTI image by YOLOv3. The region in the green box contains several vehicles that cannot be detected. After we applied our model on the blurry region, Figure 4 shows the successful improvement on vehicle detection. Multiple vehicles can now be detected; however, extensive experiments shows that this improvement on detection is unstable. Our model does not work well for every case. Image super-resolution can help for detection tasks, but more suitable models should be designed for this specific task.

ESRGAN is the state-of-the-art super-resolution architecture and it works well on different tasks; however, ESRGAN is mainly designed to improve the resolution and the reality of output images, with less consideration about the running time or computation resource consumption. Current detection or classification models on CAVs are required to be real-time and ESRGAN is too heavy for real time tasks. A light-weight super-resolution model is needed for a real-time implementation, especially for autonomous driving or face recognition systems.

VI. Related Work

We summarize the existing works related to trust modelling and trust management in vehicular network and image super-resolution techniques in this section.

A. *Trustworthy Vehicular Networks*

Most works of trustworthy vehicular network studied in the literature are focusing on securing or authenticating vehicles based on the public key infrastructure (PKI). While PKI builds the first line of defense [?], it only provides the identification of legitimate vehicles but not the trustworthiness of data being shared. To understand trustworthiness in computer networks, there are intensive studies on trust in multi-agent systems [?] and mobile ad hoc networks (MANET) [?]. In addition, trust management was intensively studied in distributed systems [?]. Only a few works are proposed for trustworthy vehicular

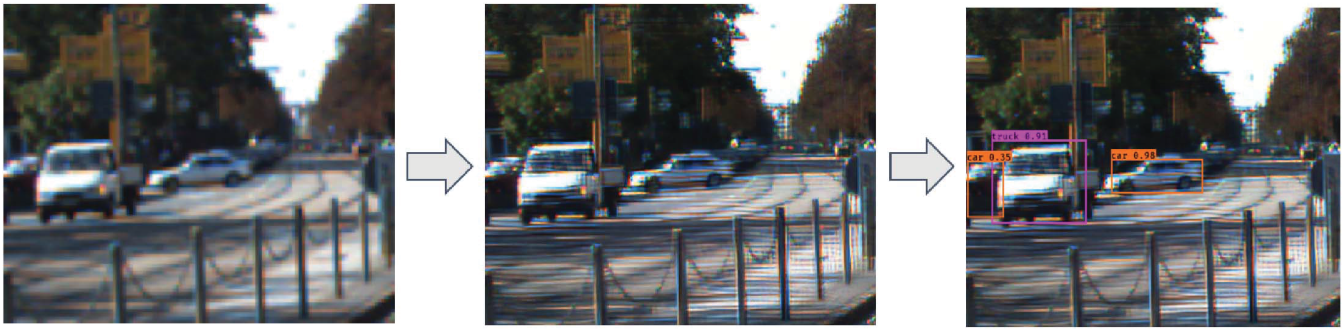


Figure 4: Detection of distant objects through ESRGAN. The left subfigure shows the region that is reported to contain objects. Enhancing the resolution of this region by ESRGAN, a higher-resolution image is obtained in the middle subfigure. Feeding the enhanced image into an object detection model, three vehicles are successfully detected in the right subfigure.

ad hoc networks (VANETs), which can be classified into two categories: information-oriented [?] and entity-oriented trust models [?]. Inspired by securing information integrity, researchers developed several approaches [?] allowing vehicles to decide how to trust received messages. On the other hand, there are works on studying the trustworthiness of vehicles [?]. Although several mathematical models are proposed to quantify vehicle's trust, it is not clear how they are applied in a real-world vehicular system to facilitate trustworthy information sharing among vehicles.

B. Image Super-Resolution

Image super-resolution (SR) techniques reconstruct a higher-resolution image or sequence from the observed lower-resolution images. Plenty of well-known traditional methods have been proposed already [?], [?], [?]. Recently, super resolution methods based on convolutional neural networks (CNNs) show a significant performance improvement. In [?], a CNN is combined with sparse coding to offer an image super-resolution solution. [?] used very deep recursive layers to improve performance without introducing more parameters. Several Generative Adversarial Networks [?] based super-resolution methods can also achieve a good performance on image super-resolution task. For example, [?] improves the super-resolution performance by defining a novel perceptual loss based on high-level feature maps. [?] improves the SR-GAN by replacing all batch normalization layer with a residual dense block.

VII. Conclusions

For the first time, we apply the trust modelling and trust management techniques, designed for vehicular networks, onto autonomous vehicles to realize a trustworthy perception information sharing on CAVs. To enable a vehicle to assess the trustworthiness of more data shared from others, ESRGAN is adopted to enhance the resolution of images that contain distant objects. Based on shared object detection results, we first crop out all distant objects from captured images and apply our super-resolution model on these blurry images to detect possible objects. We evaluate our results on a pre-trained ResNeXt101 model and the results shows our framework could significantly improve car classification accuracy: top-1 accuracy improves from 45.71% to 62.01% and top-5 accuracy improves from 82.86% to 91.57%.

Acknowledgment

The work is supported by National Science Foundation (NSF) grants NSF CNS-1761641 and CNS-1852134.