Computers, Materials & Continua

Tech Science Press

DOI: 10.32604/cmc.2023.046707

ARTICLE

Leveraging Augmented Reality, Semantic-Segmentation, and VANETs for Enhanced Driver's Safety Assistance

Sitara Afzal[1], Imran Ullah Khan[1], Irfan Mehmood[2] and Jong Weon Lee[1,*]

[1]Mixed Reality and Interaction Lab, Department of Software, Sejong University, Seoul, 05006, Korea

[2]Faculty of Engineering and Digital Technologies, School of Computer Science, AI and Electronics, University of Bradford, Bradford, UK

*Corresponding Author: Jong Weon Lee. Email: jwlee@sejong.ac.kr

Received: 11 October 2023 Accepted: 28 November 2023 Published: 30 January 2024

ABSTRACT

Overtaking is a crucial maneuver in road transportation that requires a clear view of the road ahead. However, limited visibility of ahead vehicles can often make it challenging for drivers to assess the safety of overtaking maneuvers, leading to accidents and fatalities. In this paper, we consider atrous convolution, a powerful tool for explicitly adjusting the field-of-view of a filter as well as controlling the resolution of feature responses generated by Deep Convolutional Neural Networks in the context of semantic image segmentation. This article explores the potential of seeing-through vehicles as a solution to enhance overtaking safety. See-through vehicles leverage advanced technologies such as cameras, sensors, and displays to provide drivers with a real-time view of the vehicle ahead, including the areas hidden from their direct line of sight. To address the problems of safe passing and occlusion by huge vehicles, we designed a see-through vehicle system in this study, we employed a windshield display in the back car together with cameras in both cars. The server within the back car was used to segment the car, and the segmented portion of the car displayed the video from the front car. Our see-through system improves the driver's field of vision and helps him change lanes, cross a large car that is blocking their view, and safely overtake other vehicles. Our network was trained and tested on the Cityscape dataset using semantic segmentation. This transparent technique will instruct the driver on the concealed traffic situation that the front vehicle has obscured. For our findings, we have achieved 97.1% F1-score. The article also discusses the challenges and opportunities of implementing see-through vehicles in real-world scenarios, including technical, regulatory, and user acceptance factors.

KEYWORDS

Overtaking safety; augmented reality; VANET; V2V; deep learning

1 Introduction

According to a survey conducted by the World Health Organization (WHO), over 1 million people lose their lives each year due to driver error and dangerous overtaking tactics in road accidents, resulting in tragic and preventable fatalities. When a driver must cross the front car and perceive

impending traffic, especially buses, and trucks, which are hazardous due to their size and ability to block the driver's eyesight, they must perform an overtaking technique.

Overtaking on such vision-blocking vehicles and changing lanes is a challenging task because drivers are unaware of the traffic scenario. Many road tragedies occur at the time when a driver is shifting into another lane for overtaking and they do not know about the oncoming vehicle in another lane.

Many scholars are working to create an effective see-through system that can help and improve the driver's vision to overtake the vision-blocking car safely because there is now no such efficient system that helps the driver decide whether to overtake or not in such a maneuver [1].

Research into intelligent transportation systems has quickly expanded to incorporate a variety of cognitive characteristics into automobiles to create intelligent vehicle systems. The information transfer between the vehicles is ensured by a wireless Vehicle to Vehicle, i.e., V2V communication system. The camera stream sends the view of the leading car to the rear car to improve the vision and assist the driver in overtaking large vehicles such as buses and trucks. Vehicular Ad-hoc Network, i.e., VANET is utilized for establishing a network connection between the front and rear cars due to its advantageous features in high mobility settings, such as quick communication between vehicles in short distances and minimal latency.

Segmentation is one of the challenging tasks in the computer vision area [2–4]. Many researchers are using segmentation for different purposes such as segmentation in the medical field, fire segmentation, and road scenes in urban areas, but in this paper, we performed segmentation for segmenting vehicles in the scene. To achieve this goal, we use transfer learning by using the DeepLabV3 model which is originally trained on the Coco2017 dataset. In the context of deep learning, transfer learning is commonly employed to leverage pre-trained models on large-scale datasets and adapt them to smaller or domain-specific datasets. This study's specific purpose is to determine whether it is possible to enhance the accuracy of semantic segmentation.

We conduct experiments using atrous convolutions from the DeepLab model, using real-world photos from Cityscapes [5], which is presently the default benchmark dataset for semantic segmentation [6]. We used atrous convolution for extracting dense features to improve the accuracy of the model. As we are only interested in vehicle segmentation and the Cityscape datasets [5] have many classes, we ignore other classes and just select the classes related to cars and vehicles. There are 5000 fine-labeled photos and 20,000 coarse-labeled images in Cityscapes [5]. The main issue in training the deep convolutions in semantic segmentation is inadequate semantically labeled data for self-driving vehicles. This issue may probably hinder the development of self-driving cars with a high degree of autonomy.

The degree of user immersion in AR systems is crucial for fusing digitally enhanced objects with the physical environment. With our suggested method, the see-through effect was displayed using cameras in both automobiles and a windshield display in the back car. After the server in the rear car segments the vehicle, the video from the front car will be displayed in the segmented area of the car. This transparent technique will instruct the driver on the concealed traffic situation that the front vehicle has obscured. More precisely, we want to help the driver safely change lanes and pass the vehicle in front of them.

Our system was developed by integrating several modules, such as VANET [7] for facilitating V2V communication, DeepLabV3 for vehicle segmentation, and an augmented reality module, which works

to enhance the front car video on the segmented section of the vehicle, resulting in a see-through effect. Table 1 shows the acronyms used in this manuscript.

Table 1: List of acronyms

Sr. No.	Acronyms	Definitions
1	AR	Augmented reality
2	V2V	Vehicle to vehicle
3	VANET	Vehicular ad-hoc network
4	ADAS	Advanced driver assistance system
5	STS	See-through system
6	VO	Visual odometer
7	V2I	Vehicle-to-infrastructure
8	DSRC	Dedicated short-range communication
9	OBU	On-board unit
10	RSU	Road side units
11	ASPP	Atrous spatial pyramid pooling
12	DCNN	Deep convolutional neural network
13	LCD	Liquid crystal display

This comprehensive approach allowed us to achieve our objective. Three phases make up the suggested system. The front car's video stream is sent to the back car during the first phase via the VANET system installed in both vehicles. The segmentation technique is used in the second step to separate the front car from the rear car video stream. The segmented portion of the preceding automobile is enhanced with the video stream from the front car in the third phase. Our see-through system improves the driver's field of vision and helps him or her change lanes, cross a large car that is blocking their view, and safely overtake other vehicles.

The following are the main contributions of the proposed work:

- We employed the transfer learning approach with DeepLabV3 using atrous convolution for extracting dense features and improving the segmentation accuracy.
- We present the Deep Learning, AR, and ad-hoc network-assisted framework to segment the vehicle in intricate traffic scenarios and communicate the content in real-time among vehicles.
- All the testing and validation is performed on the Cityscapes Dataset, which demonstrates the real-time processing capability of the proposed framework, making it suitable for time-critical application such as traffic and driving.
- We present a lightweight model with high performance while reducing the computational cost. The practical capability of our proposed approach is its lightweight design addressing the real-world deployment in autonomous vehicle systems.

The rest of the sections are organized as follows: In the second section, we discussed some related literature related to see-through car systems. The third section of our paper provides a detailed discussion of each module employed in our proposed methodology. Following this, in the fourth section, we present the results of various experiments carried out to evaluate our system. Lastly, we conclude our paper in the fifth section.

2 Literature Review

This section provides an overview and analysis of the existing literature. The literature review aims to examine and synthesize the relevant studies, theories, and findings in terms of Augmented Reality and Deep learning.

Researchers have recently focused on the Advanced Driver Assistance System, i.e., ADAS to create a comfortable and effective method to see through automobiles. The See-Through System, i.e., STS is used in advance vehicles to prevent the accident ratio during the overtaking maneuver. The system was designed to aid and improve the visual perception of drivers by projecting the view from the front car onto the back of the car, creating a see-through effect to view surrounding vehicles.

In another research [3], they proposed layer-wise training as a novel learning strategy for semantic segmentation and assess it on a light-efficient structure dubbed an efficient neural network (ENet). On two RGB picture datasets on the roadway and off-road pathways, the suggested learning method's outcomes are compared to the classic learning approaches, including mIoU performance, network robustness to noise, and the possibility of reducing the size of the structure. Using this strategy reduces the need for Transfer Learning. When the input is loud, it also enhances the network's efficiency. In self-driving cars, precision is not the only relevant parameter, another important model that can produce a decent Intersection Over Union is also required. In another research [8], they proposed a U-Net architecture that allows and maintains a reasonable balance in terms of accuracy and Intersection Over Union when compared to the FCN architecture tested on the CamVid dataset.

Rameau et al. [9] offered a technique for improving drivers' ability to see past cars in real-time. Two stereo cameras are placed on the front car in their suggested system, and one camera is placed on the next car. The V2V communication system is used for communication and information exchange between the vehicles. Firstly, a fast stereo Visual Odometry i.e., VO computes the real scale 3D map based on stereo image pairs, where the second step is localized and tracked the rear car in the map. Further, a synthetic image is generated by computing the dense depth map of the leading vehicle, the back of the front car is detected through a 3D bounding box, and at last, the synthetic patch is cropped and stitched to the rear of the car.

Researchers [10] explored technology by using semantic networks and dual link prediction. They aim to provide a comprehensive understanding of technological opportunities. By employing dual link prediction techniques within hierarchical semantic networks, the research enhances the identification and assessment of potential technological advancements, offering valuable insights for strategic decision-making and technology innovation. The dual approach uniquely contributes to a nuanced analysis of technology opportunities, making the paper a significant contribution to the field.

Bingjie et al. [11] presented the solution for overtaking maneuvers in driving and see-through cars is another simple and affordable technique. By showing the geometric relationship in 3D space with the aid of two cameras, they first address the formulation problem and its suitable solution in image synthesis: one in the front car and the second in the following car. Patch computation and stitching are performed to augment the video stream coming from the front car camera on its rear by utilizing the Unity and Vuforia marker-based techniques. Another research conducted by Gomes et al. [12] utilized the windshield cameras that transmitted the video stream of the front vehicle through Dedicated Short-Range Communication, i.e., DSRC. To transform the front vehicle into a transparent plane object, numerous techniques are used for STS such as distance sensors, AR, and computer vision. The AR module generates a 3D image of the front vehicle and renders the video stream with computed length on the back of the vehicle that can be displayed on the dashboard of the overtaking vehicle.

The simulation of this system was carried out by the integrating VANET simulator, in-vehicle driving simulator, and simulated 3D road scenarios containing different traffic signs.

To implement the ADAS in a real-world scenario, Gomes et al. [13] proposed a system based on DSRC, augmented reality, and computer vision technology with real-time detection and segmentation of the leading vehicle. The video stream of the leading vehicle is superimposed on the rear of the segmented vehicle, which can be displayed on a transparent LCD for the driver to see through the vehicle. To address the issues of fuzzy vision of contents and the relationship between the static location of the LCD and the driver's eye produces misalignment in the augmentation of digital contents in existing techniques, a solution has been proposed by Michel et al. [14]. Numerous devices are utilized including VUZIX smart glasses for resolving the blurred vision of contents and driver's eyes viewpoint, DSRC radios for communication between vehicles, and GPS sensors for distance calculation between vehicles. Moreover, smart glasses video streams are analyzed through computer vision techniques to detect the rear of the vehicle and attach the front car video streaming on its back for STS. Seong-Woo et al. [15] proposed a technique based on multi-vehicle cooperative perception, to visualize the occluded area for the rear vehicle to see through the front vehicle. The augmented reality term is used for the natural and direct visualization based on a 3D perspective transformation of estimated 2D range visual depth perception of vehicles. The cooperative perception technique can be utilized for the automation of vehicles, while augmented reality can be supportive in driving assistance.

Another research presented by Francois et al. [16], comprised stereo cameras on the front car, and monocular camera on the rear car, and a wireless communication system for transmitting only relevant information for better performance in real-time. Furthermore, a marker-based technique (four markers mounted on each corner of the rear part of the front car) is also developed for inter-car pose estimation and tri-focal tensor-based image synthesis from a disparity map. The coordinates information of the front car is transmitted through the trifocal tensor to the rear car for stitching and disappearing the car from the rear car view.

Table 2: Comparison with state-of-the-art approaches in the field

Reference	Year	Target	Approach	Dataset	Performance measure
[17]	2019	Driver activity recognition	CNN + ResNet 50 Transfer Learning	Custom data collected by Kinect	75% Accuracy
[18]	2019	Urban Autonomous Driving	Reinforcement Learning	Raw bird view images	DDQN: 80% Accuracy TD3: 88% Accuracy
[19]	2021	Autonomous & Manual Transportation System	CNN with 5G ITS	Trajectory dataset natural driving dataset	90.88% Accuracy
[20]	2020	Autonomous Driving	SWGS Algorithm	Fast beam alignment	2.8 Gbs stable rate
[21]	2018	See-through Truck for driver monitoring	CNN	20 truck drivers	0.64 Standard Deviation

(Continued)

Table 2 (continued)

Reference	Year	Target	Approach	Dataset	Performance measure
Our	2023	See-through vehicle	Segmentation + Deep atrous-based CNN	Cityscape	97.1% F1-score

Olaverri et al. [22] conducted a study about STS, a co-ADAS for safely overtaking based on GPS for distance calculation, a webcam, and DSRC radios for transmitting the video stream between vehicles. The system uses the Geo-cast features for the communication between IEEE 802.11p radio and on-board unit, i.e., OBU over an ad-hoc network.

Xhang et al. [17] proposed a study that can understand the behavior of the driver in an intelligent vehicle. For this, they used CNN and attained an accuracy of 75% accuracy using ResNet50. Similarly, other studies also mentioned in Table 2, work for intelligent or autonomous vehicles for the safety of drivers. Synthetic pictures have been effectively used for a variety of self-driving car tasks, including semantic segmentation. Thus, Ros et al. [23] used synthetic images from their SYNTHetic collection of Imagery and Annotations dataset to augment real-world datasets and improve semantic segmentation; Richter et al. [6] extracted synthetic images and data for generating semantic segmentation masks from the video game Grand Theft Auto V and used the acquired synthetic dataset to improve semantic segmentation.

Overall, our study provides valuable insights into the development of deep learning approaches for autonomous vehicles and highlights the potential of these methods for improving the safety and efficiency of autonomous driving technology.

The overall system is evaluated via real-world scenarios and 3D scenarios created in the OpenSceneGraph library with a realistic driving simulator. In our proposed approach, with our method, we merged several modules, including DeepLabV3 for vehicle segmentation, VANET for V2V communication systems, and an AR module for creating see-through effects. Our see-through system improves the driver's field of vision and helps them change lanes, pass a large car that is blocking their view, and safely overtake other vehicles.

3 Proposed Methodology

In this section, we will present each of the components of the research design, participants, data collection methods, data analysis techniques, and ethical considerations that were developed for our algorithm individually. The information provided will enable readers to evaluate the rigor of the study's methods and conclusions.

3.1 V2V Communication Layer

V2V communication layer refers to the wireless communication protocol used for vehicle-to-vehicle communication. It enables vehicles to exchange information with one another, such as speed, location, and direction, in real-time, using DSRC technology. The V2V communication layer is a critical component of connected vehicle technology, which aims to improve road safety, reduce traffic congestion, and enhance the overall driving experience. The visual representation of data cannot

currently be fed through V2V wireless network connectivity by car AR systems. The adoption of DSRC by the automobile sector would enable wireless V2V and vehicle-to-infrastructure, i.e., V2I communication, greatly enhancing the effectiveness of network connectivity in vehicles [9]. Since vehicles are already used in VANET, they are typically just regular vehicular nodes. Vehicles have an On-Board Unit (OBU) that enables them to link with static or mobile Road Side Units (RSUs) or operate as network nodes and interact with one another using V2V or V2I technologies, as appropriate. Additionally, OBU features a GPS sensor to give position data, memory to store and process information, processing capabilities, and compatibility with the IEEE 802.11p communication standard. To keep track of the vehicle's physical condition and related data, it is also connected to wireless sensors placed within the vehicle. These sensors promptly report to the OBU, which then generates an emergency alert, if any abnormal event occurs with the vehicle (such as collisions with an object or another vehicle).

OBU also offers a user interface for facilitating interaction with VANET's security and entertainment apps. Many wireless technologies can be used for V2V communications DSRC is the most used technology for wireless V2V communication, and it is short-range technology that works up to 1000 m. DSRC allocates 75 MHz bandwidth at frequency 5.9 GHz which is handled by the Frequency Communications Commission (FCC). In addition, a bandwidth of 75 MHz is split into seven channels, each of which is 10 MHz wide. One of these channels is designated as a control channel and is only utilized for safety-related purposes, while the remaining six channels are service channels such as 2 spectrum bands preserved for special use, and the rest are used for safety and not safety purposes [24]. DRSC band can be used free because the FCC has no fee charges for using the spectrum, but this spectrum is licensed and different from unlicensed bands in 2.4 GHz, 5 GHz, and 900 MHz which can be also used freely without charges [25].

The performance of communication can be measured by both Third Generation 3G and DSRC technology. The initial form of performance assessment involves a spontaneous method and employs the DSRC protocol, enabling direct communication between the two vehicles. On the other hand, the second assessment can be conducted using an infrastructure-based approach, utilizing 3G technology. In the DSRC-based transmission process, the 802.11p standard, which operates in the 5.9 GHz frequency range, specifies improvements to 802.11 that are necessary to allow wireless local area networks in a vehicle setting. The following equipment is needed for this setup two high-quality antennas that can be attached to the roof of the vehicle, two LinkBirds MX V3 which can provide DSRC connection, wireless drivers for connection establishment of both vehicles, cameras on the vehicle windshield for video streaming and by utilizing the powerful systems in modern vehicles. By using this type of technology, the driver can see the image of another vehicle with a 100 ms delay which can be considered as real-time transmission through VANET.

The second type of vehicle connectivity where transmission occurs through 3G technology is very common in vehicles nowadays and is usually utilized in real-time traffic information systems, emergency messages, and remote diagnostics systems. Utilizing 3G networking technology, which requires two 3G network adapters, a camera positioned on the windshield, and a high-performance system built into contemporary vehicles, the performance of the system may also be evaluated.

However, this infrastructure has a packet delay of almost half a second which cannot be more helpful in overtaking maneuvers. The real position of the front vehicle and the position shown through the video streaming might differ by several meters as a result of this kind of delay. This 3G infrastructure can support several connected vehicle applications, but it is unable to offer the low-latency streaming needed for overtaking assistance. In contrast, DSRC can deliver the required

communication delay and, because of its ad hoc nature, it is also compatible with bandwidth restrictions that lead the DSRC-based technology usage from 3G networking [12].

3.2 Extracting Dense Features Using Atrous Convolution

Extracting dense features using atrous convolution involves applying atrous convolutions in a deep learning model to capture rich and detailed information at multiple scales while preserving spatial resolution. Atrous Convolution is implemented in a fully-convolutional manner, to extract the dense features [26] and this approach proved to be efficient in terms of deep segmentation.

In this network, the combination of max-pool and stride layers reduces the spatial resolution of the resultant feature-map by 32 [27–29]. To regain these pixels, de-convolutional layers [30–34] have been implemented. So, with the utilization of 'atrous convolution', which is initially formed for computations and utilized in DCNNs [35,36]. Consider two-dimensional signals, the atrous convolution is implemented over the input feature map x, for each location I on the output y and a filter w:

$$Y[i] = \sum x[i + r.k]\, w[k] \tag{1}$$

It is comparable to convolving the input x with the up-sampled filters created by inserting r-1 zeros between two successive filter values for each spatial dimension where r represents the stride through which the input signal is sampled (atrous is a French word, and trous means holes in English). For rate $r = 1$, the standard convolution has a special case, and to change this rate value to adaptively enhance the filter's field of view, we can use atrous convolution. Fig. 1 illustrates this.

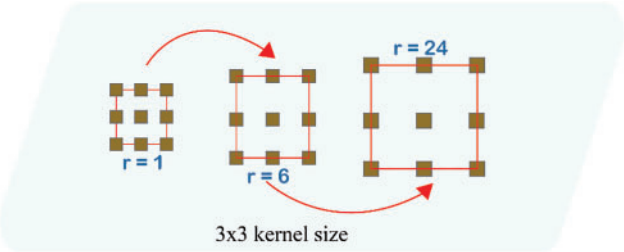

Figure 1: Atrous convolution same kernel size of $3 * 3$, but with different rate

In a fully convolutional network, this atrous convolution allows us to directly regulate the dense feature responses. For the Deep Convolutional Neural Network, i.e., DCNNs [27–29] implemented for classification, the feature responses are reduced by 32 times as compared with the inputted image dimension, so the outputted stride = 32. In the DCNNs, if anyone wants to twice the spatial density of calculated responses of the feature, i.e., 16, the stride to the last convolutional layer or pooling layer is set to 1. Later, all the convolutional layers are substituted with the atrous convo layers and these layers have a rate value of $r = 2$. This allows us to extract denser features and it also not requires extra parameters for learning.

3.3 Enhancing Model Depth with Atrous Convolution

In this proposed approach, we use atrous convolution. Atrous convolution is a method that allows for an increase in the effective area that the model sees of a convolutional neural network (CNN) without significantly increasing the number of parameters or reducing the spatial resolution of the

feature maps. The atrous rate in atrous convolution controls the spacing between the values in the convolutional filters.

By increasing the atrous rate, the effective receptive field expands, allowing the model to capture larger-scale contextual information. This is particularly useful for tasks that require capturing multi-scale information, such as semantic image segmentation or object recognition.

In this proposed approach, we initially look into different atrous rates with atrous convolution. To be concrete, In ResNet, we have duplicated numerous duplicates of block 4 as shown in Fig. 2, and then arrange all these copies in a cascade. There are three convolutions of 3×3 dimensions in these blocks, whereas the last convolutional contains stride 2 except the one in the last block as in the original ResNet. An approach named as multigrid approach, which implements a hierarchy of different sizes grids [37]. We used multigrid to drive our proposed strategy, and we used different atrous rates from block 4 to block 7. For this, we define r1, r2, and r3 as multigrid the same as the unit rate. In the convolutional layer, the multiplication of the unit rate to the corresponding is equivalent to the final atrous rate. In this case, If we consider output rate 16 with multigrid rates, then the respective convolutional layer will be $2 \times [1,2,3] = [2,4,8]$.

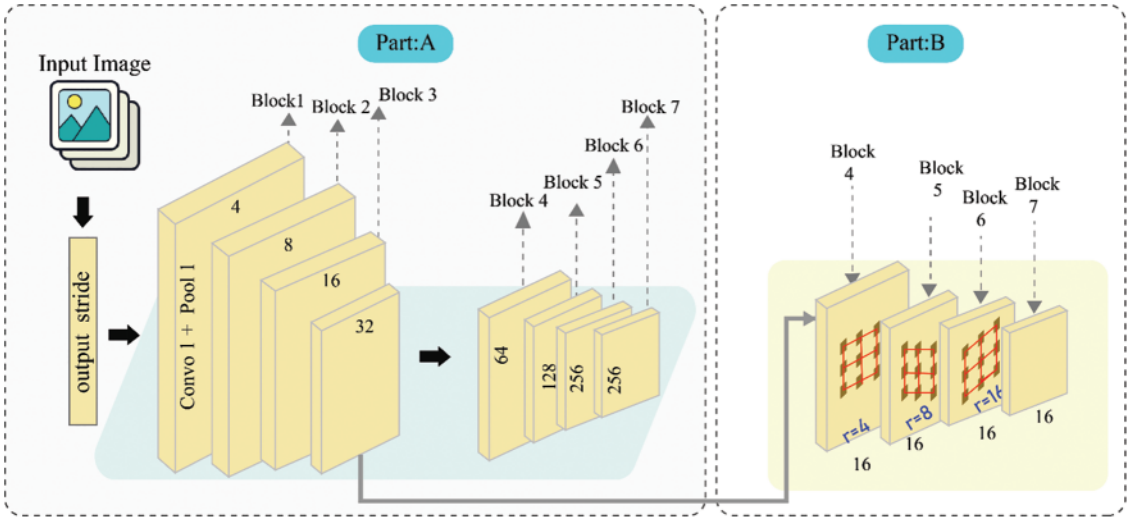

Figure 2: Comparison of modules with and without atrous convolution

The main motivation to implement this module is that this introduced striding can easily capture the long-range info in the deeper blocks. For instance, the features of a complete image can be summed up in the final small feature map, as shown in Fig. 2. But successive striding is damaging for the semantics since the complete detail info is destroyed and that is why we implement atrous convolution with rates set on the anticipated output values as illustrated in part B of Fig. 2, where the stride is 16 in the output. In the model, we do experiments with cascaded ResNet blocks up to the 7th block which means blocks 5, 6, and 7 are duplicates of block 4, which has a stride of 256 in output without atrous convolution.

3.4 Atrous Spatial Pyramid: Incorporating Pooling for Feature Extraction

We conceptualize the pooling in the atrous spatial pyramid that is presented in [4], in which on the top of the feature map, a total of 4 parallel atrous convolutions having distinct rates are applied. The Atrous Spatial Pyramid Pooling, i.e., ASPP is motivated by the success of spatial pyramid-pooling

[38,39], demonstrating that it is effective for sampling data at various sizes for the precise and effective categorization of configurable scale regions. Within ASPP, we have incorporated batch normalization. This ASPP, having distinct atrous rates effectively catches multi-scale info. Though we come across that, the number of valid filter weights, i.e., the rate implemented on a valid region of features rather than zero-padding gets smaller as the rate of sampling gets larger. when a 3×3 filter with a 65×65 feature map and a separate atrous rate is applied to it. In the utmost case, when the rate value gets close to the size of the feature map.

In this extreme case, the filter having 3×3 dimensions degenerates to a simple 1×1 filter rather than capture the whole image context, since just the weight of the center filer is effective. To get better info and to include global context to the framework, we take on image-level features, like to [40,41].

In particular, we implemented global average pooling on the most recent features map, input the generated image-level features to the convolution of 1×1 dimensions with 256 filters, and subsequently bi-linearly up-sampled the feature to the necessary spatial dimension. After all, our enhanced ASPP contains; (i) 1 conv with 1×1 dimensions and 3 conv having 3×3 dimensions with rates = (6,12,18) when the output stride = 16 and (ii) the image-level features as illustrated in Fig. 3. Consider that the rate is double when the output stride = 8. Concatenated features from each branch are then passed through a new conv of 1×1 dimension before the final convolution of 1×1 dimension which generates the final logits.

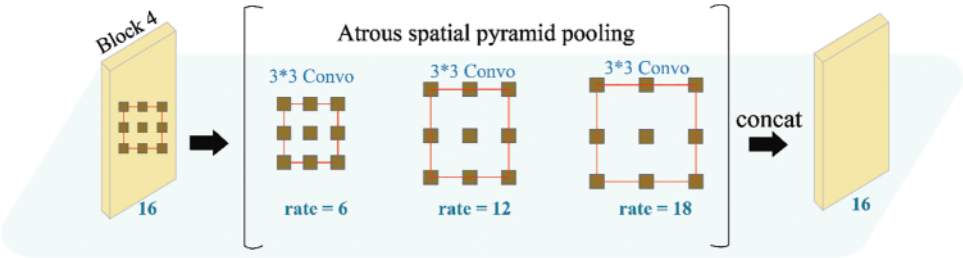

Figure 3: Parallel modules utilizing atrous convolution

3.5 Integrating an Augmented Reality Module

In this section, we discussed the augmented reality technique in ADAS for safe overtaking maneuvers. The term "Augmented Reality Module" typically refers to a software component that integrates computer-generated images or data with the user's real-world environment in real-time. To enhance driver visual perception and ensure safe overtaking, drivers in overtaking vehicles should know about the traffic environment behind the preceding vehicle. For this purpose, both the preceding vehicle and the overtaking vehicle are connected through VANETs for the exchange and transmission of video streams. Both vehicles are equipped with high-resolution cameras to capture the front view of the vehicle. While installing the cameras on the windshield of the vehicles, the angle and position of the camera should be the same in both vehicles to see the area occluded by preceding vehicles. All the processing will be done by the powerful server embedded in advance vehicles which can smoothly run the deep learning models. As both vehicles are wirelessly connected through VANETs, the preceding vehicle sends the video stream S1 to the rear car through the wireless communication channel. VANETs enable novel AR applications due to their qualities designed for high-mobility scenarios, such as zero latency as well as very short associating times.

Our approach uses computer vision techniques to segment the rear of the preceding vehicle where we can superimpose the front vehicle camera view. The see-through system will only be applicable if

there is a short distance between the leading vehicle and the following vehicle. In this module, to create a see-through impact, we want to accept the front car stream S1 into the rear car so that we can add it to the segmented area of the previous vehicle. The forward car is transmitting stream S1 to the rear vehicle through the communication layer, as seen in Fig. 4.

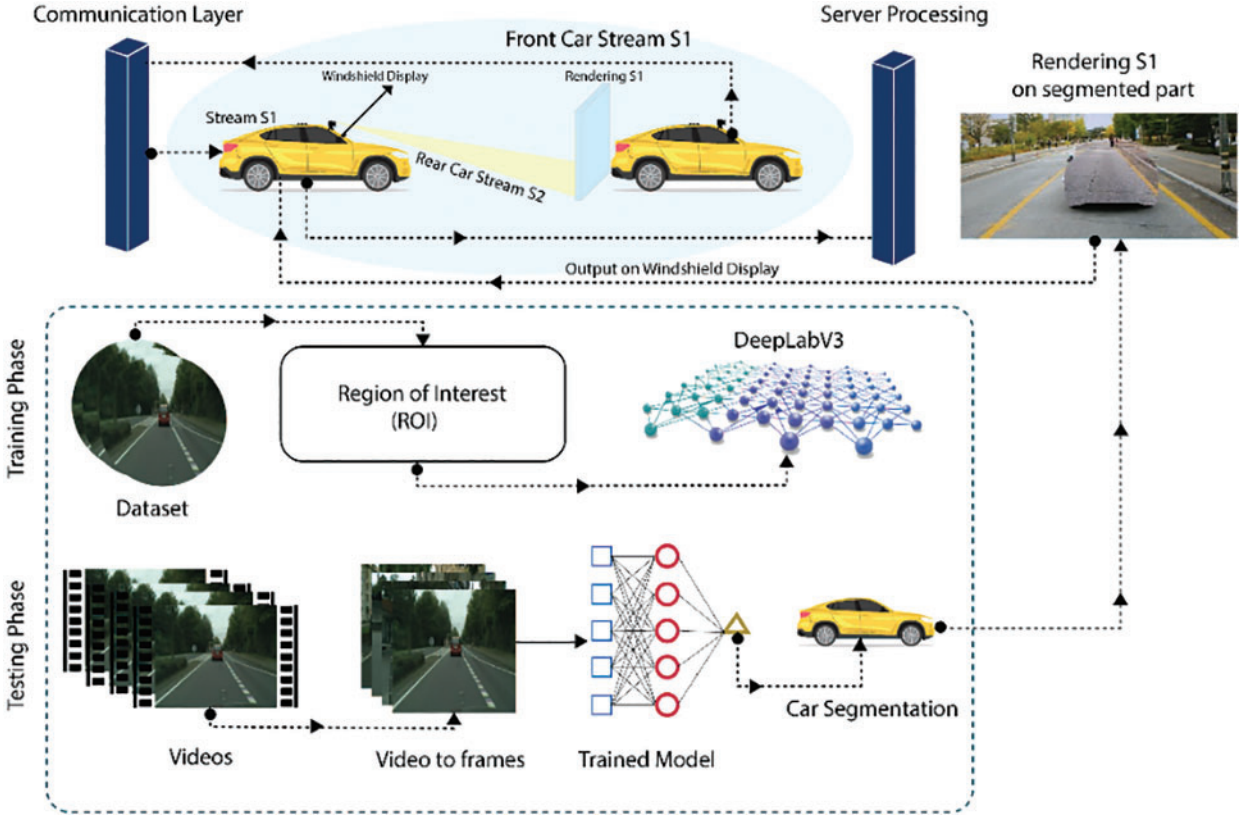

Figure 4: Comprehensive framework of STS

The segmentation method in stream S2 only segments the back end of the preceding vehicle. But even though we require all the surrounding sceneries for a continuous see-through look, we are interested in the segmented part to supplement the S1 and allow drivers to be aware of the traffic conditions for safe overtaking. The whole stream S2 is displayed by inverting each frame from the camera, except the segmented area of the vehicle. To improve the S1 front vehicle camera stream, the rear area of the car is divided, and the beginning and ending points of the segmented region are determined. The driver can then see the area that the front vehicle has obscured by overlaying the front vehicle video on the segmented portion of the car. Only the view obscured by the front car will be shown in a segmented vehicle, thus the leading vehicle's camera angle and location must be comparable to the position of the following vehicle's camera. By utilizing this AR approach, we offer a practical tool that can improve visual perception and help drivers overtake safely.

4 Experimental Design and Outcome Analysis

In this research implementation and testing, we have used Core i9 11th Gen Intel CPU which implants NVIDIA GeForce RTX 3070 GPU system. The implementation is carried out in Pytorch,

which is one of the most popular and open-source deep learning frameworks and is based on the Torch library mostly uses computer vision-related applications. We have tested the model on our own captured video on road scenario and checked the accuracy and complexity of our model. Our model's inference time on the GPU system with the specifications is 66 frames per second (fps), while the entire processing time of the framework is 43 fps with a network delay of 100 ms in the stream transmission from one vehicle to another, which is still real-time. Our suggested solution performs admirably in terms of precision and time complexity, and it may be used in actual traffic situations.

4.1 Overview of the Data Used in the Study

In our proposed approach, we have used the Cityscapes dataset [5]. It contains scenes of random streets collected from 50 different cities and high-quality pixel-level annotations of almost 5000 frames. This dataset is publicly available along with ground truth for research purposes. It contains the annotations for other classes as well instead of cars like bicycles, people, trees, etc. For our proposed approach, we only considered car class, to segment the cars only. Fig. 5 shows the sample images of the cityscape dataset.

Figure 5: Examples of data samples from the cityscape dataset

4.2 Methodology for Car Segmentation Proposed in the Study

In the experimentation, to obtain the dense features, we used the pre-trained ResNet for semantic segmentation. After collecting the cityscape car dataset each pixel is labeled as either car or background.

We used a pre-trained DeepLabV3 model that has been trained on a large-scale dataset like COCO. This model serves as the starting point for transfer learning. We define the ratio of input spatial-resolution to output-resolution similarly to output stride. We used several convolutions based on the TensorFlow framework in our findings. As shown in Fig. 6, we used the cityscape data samples from the pre-trained Coco2017 dataset and the DeepLabV3 model for this study.

In this model, we first initialize the model with pre-trained weights, this initialization allows the model to leverage the learned features and representations from the pre-trained model. After this, we Incorporate atrous convolutions into the modified DeepLabV3 model. Atrous convolutions can be added to the backbone network of DeepLabV3 to capture both fine-grained details and multi-scale contextual information important for car segmentation. We adjust the atrous rates to control the dilation levels and the receptive field size.

Split the car segmentation dataset into training and validation sets. Train the modified DeepLabV3 model on the training set using a suitable loss function, such as pixel-wise cross-entropy loss or dice loss. This large-scale dataset contains 2975 images for training and 500, 1525 for validation and testing respectively. For evaluating our proposed approach, 19 semantic labels were utilized without considering the void labels.

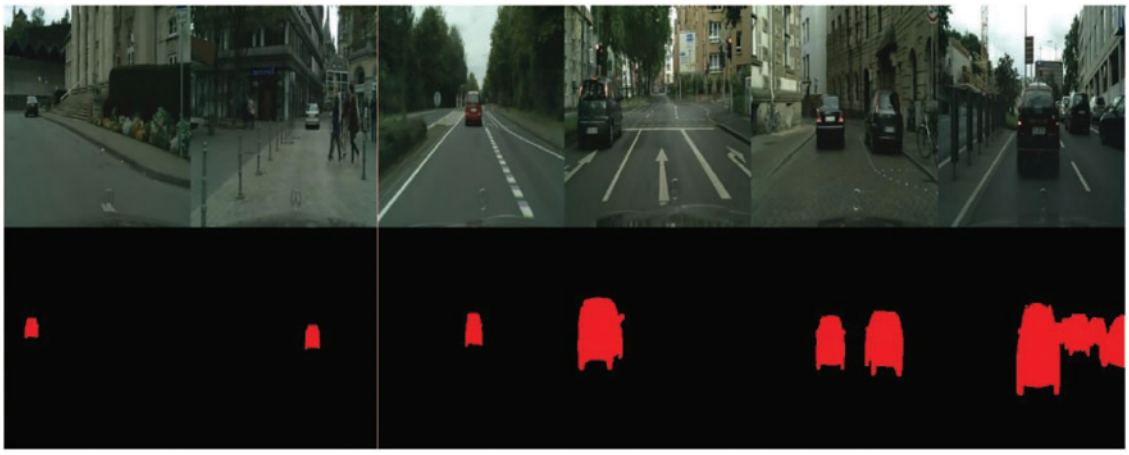

Figure 6: Segmented ROI samples of the dataset

At first, we used our training photos to train the model, and we used a F1-score to assess its performance. Monitor the model's performance on the validation set to prevent overfitting and select the best-performing model. By combining atrous convolution and transfer learning with DeepLabV3, the model can effectively capture fine-grained car details and leverage knowledge from pre-trained models, leading to accurate car segmentations.

Our suggested method has a performance accuracy of 97.1% when using Google Colab with training weights on epoch 25. The below Fig. 7 shows the F1 measure and loss of our proposed model.

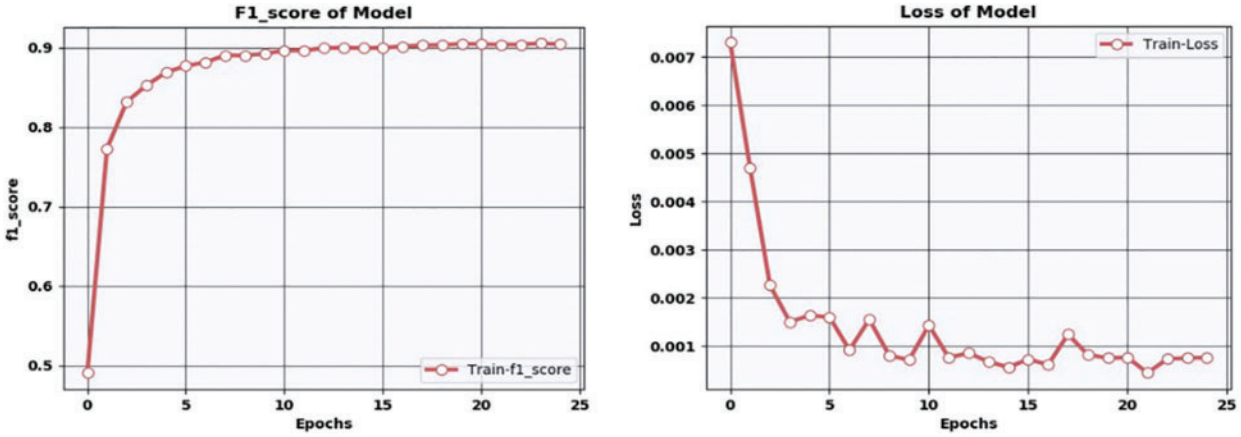

Figure 7: Performance evaluation metrics for our proposed model: F1-measure and loss

4.3 AR Module Results

By fusing virtual and real-world content, augmented reality plays a crucial role in giving virtual data to drivers to improve their sense of sight and enable them to see through the vehicles. As discussed in Section 3, we developed a system through which the driver can safely overtake the vision-blocking vehicle. To locate the zone of interest in which we can add virtual information (front car stream) to assist the driver in this overtaking move, we used the segmentation-based technique. A

hierarchy-blocking vehicle is required for STS implementation, together with a passenger automobile for overtaking. Both cars have windshield recording devices, and the driver can view the results on a small dashboard LCD, i.e., Liquid Crystal Display. The front car's stream is sent over a wireless channel and received by a server in the back car, where the vehicle is segmented using the segmentation method covered in Section 3 above. Finally, the segmented area of the vehicle will display the video feed originating from the front car.

All of the experimental work done for this research was done in a lab, not in a real-world setting. But for recording the augmented reality-based car video, we record it outside the laboratory by wearing HoloLens. As we discussed previously the video is captured through the camera in the same scenario which is considered to be the stream coming front vehicle. The AR module task is superimposing the front car stream on the segmented part which blocks the driver view before applying the segmentation approach to S2, which separates the car, the output stream is multiplied to match the size of the image array.

Fig. 8 shows three different main stages of the AR section, including the car segmentation and then the sample image of the inverted stream. Only the segmented area is then displayed in the stream. Since the other data in the stream is also significant, we inverted the video stream to make the entire stream visible, excluding the segmented portion for the cars. Finally, the c part in Fig. 8 shows the segmented component that displays the front automobile video. However, in the segmented region, we only presented that part that is obstructed by the previous vehicle, not the entire stream, therefore the camera poses of both automobiles should be in the same position for better results. Due to its placement beneath the vehicle, the area beneath the preceding automobile in view is dark and unable to be split; as a result, the stream will not be displayed in this area.

Figure 8: From left to right, (a) Image shows car segmentation in the rear vehicle stream (b) Image shows the inverted stream (c) Image shows the front vehicle video shows the previous car's split portion

5 Discussion

Using augmented reality technology offers a significant advancement in the design and development of a see-through car in resolving the overtaking issues. Our proposed method, which employs a revolutionary deep atrous convolutional learning algorithm for autonomous vehicles, comes out as an efficient solution, outperforming established state-of-the-art methods in terms of performance measures. Our approach significantly enhances the driver's field of view and addresses the challenges associated with overtaking vehicles.

The outstanding F1-score of 97.1% on the challenging Cityscape dataset demonstrates the efficacy of our proposed approach, putting it in the lead in autonomous vehicle recognition.

This prompts further exploration and development, positioning our see-through system as a captivating subject for future studies and advancements in intelligent transportation systems. The evolving landscape of intelligent transportation demands innovative solutions, and our work contributes to the ongoing future of road safety and autonomous vehicle integration. Xing et al. [17] proposed a study that can understand the behavior of drivers in an intelligent vehicle. For this, they used CNN and attained an accuracy of 75% accuracy using ResNet50. Similarly, other studies also mentioned in Table 3, that work for intelligent or autonomous vehicles for the safety of drivers.

Table 3: Comparison with state-of-the-art approaches in the field

Reference	Year	Target	Approach	Dataset	Performance measure
[17]	2019	Driver activity recognition	CNN + ResNet 50 Transfer Learning	Custom data collected by Kinect	75% Accuracy
[18]	2019	Urban Autonomous Driving	Reinforcement Learning	Raw bird view images	DDQN: 80% Accuracy TD3:88% Accuracy
[19]	2021	Autonomous & Manual Transportation System	CNN with 5G ITS	Trajectory dataset natural driving dataset	90.88% Accuracy
[20]	2020	Autonomous Driving	SWGS Algorithm	Fast beam alignment	2.8 Gbs stable rate
[21]	2018	See-through Truck for driver monitoring	CNN	20 truck drivers	0.64 Standard Deviation
Our	2023	See-through vehicle	Segmentation + Deep atrous-based CNN	Cityscape	97.1% F1-score

Overall, our study provides valuable insights into the development of deep learning approaches for autonomous vehicles and highlights the potential of these methods for improving the safety and efficiency of autonomous driving technology.

6 Conclusion

The creation of a see-through vehicle using augmented reality technology has the potential to transform the driving experience and make all roads safer for everyone. To solve the problems with safe overtaking, we created a see-through vehicle technology. In this study, we proposed a novel deep atrous convolutional learning approach for autonomous vehicles using a combination of segmentation and region convolutional neural networks. Our method demonstrated superior performance compared to existing state-of-the-art methods in terms of accuracy and efficiency. Specifically, our model achieved a F1-score measure of 97.1% on the Cityscape dataset, outperforming the previous best result. Our results show that the proposed method is effective in detecting autonomous vehicles in a variety of challenging environments.

Research on intelligent transportation systems has quickly expanded to integrate various intelligent capabilities. Our model is composed of three steps: V2V communication, segmentation, and augmented reality. Initially, the front automobile's video stream is sent to the back car using both cars' V2V models. The back car camera then separates the front car. The segmented portion of the preceding automobile is added to at the end with the video feed of the front car. When segmentation starts, our see-through system will function; however, if the distance increases without segmentation, it will not. Our system will take more time to reply if there is a connectivity problem. A stable and reliable connection for the efficient operation of a see-through system. This delay in response time impacts the system's ability to make quick decisions and real-time info for overtaking and changing lanes.

Our see-through system improves the driver's field of vision and helps him change lanes, cross a large car that is blocking their view, and safely overtake other vehicles. Although there are still some technological and legal issues to be resolved, the advantages of this technology make it an interesting field for future transportation study and development.

Acknowledgement: The authors would like to express their gratitude to XR Research Center, Sejong Univeristy, Seoul, Korea for providing valuable support to carry out this work.

Funding Statement: This work was financially supported by the Ministry of Trade, Industry and Energy (MOTIE) and Korea Institute for Advancement of Technology (KIAT) through the International Cooperative R&D Program (Project No. P0016038) and also supported by the MSIT (Ministry of Science and ICT), Korea, under the ITRC (Information Technology Research Center) support program (IITP-2022-RS-2022-00156354) supervised by the IITP (Institute for Information & Communications Technology Planning & Evaluation).

Author Contributions: The authors confirm contribution to the paper as follows: study conception and design: Jong Weon Lee, Irfan Mehmood; data collection: Sitara Afzal; analysis and interpretation of results: Jong Weon Lee, Sitara Afzal; draft manuscript preparation: Y. Sitara Afzal and Imran Ullah Khan. All authors reviewed the results and approved the final version of the manuscript.

Availability of Data and Materials: Not applicable.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

[1] L. R. Macias, K. Picos and U. Orozco-Rosas, "Driving assistance algorithm for self-driving cars based on semantic segmentation," *Optics and Photonics for Information Processing*, vol. 12225, no. 1, pp. 28–38, 2022.

[2] D. Jurado-Rodríguez, J. M. Jurado, L. Pádua, A. Neto, R. Munoz-Salinas *et al.,* "Semantic segmentation of 3D car parts using UAV-based images," *Computers & Graphics*, vol. 107, no. 1, pp. 93–103, 2022.

[3] S. Shashaani, M. Teshnehlab, A. Khodadadian, M. Parvizi and T. Wick, "Using layer-wise training for road semantic segmentation in autonomous cars," *IEEE Access*, vol. 1, no. 1, pp. 1, 2023.

[4] K. Pasupa, P. Kittiworapanya, N. Hongngern and K. Woraratpanya, "Evaluation of deep learning algorithms for semantic segmentation of car parts," *Complex & Intelligent Systems*, vol. 8, no. 5, pp. 3613–3625, 2022.

[5] M. Cordts, M. Omran, S. Ramos and T. Rehfeld, "The cityscapes dataset for semantic urban scene understanding," in *Proc. of IEEE CVPR*, Las Vegas, New York, pp. 3213–3223, 2016.
[6] S. R. Richter, Z. Hayder and V. Koltun, "Playing for benchmarks," in *Proc. of ICCV*, Venice, Italy, pp. 2213–2222, 2017.
[7] H. C. Jang and B. Y. Li, "VANET-enabled safety and comfort-oriented car-following system," in *Proc. of ICTC*, Jeju-Island, Korea, pp. 877–881, 2021.
[8] A. Lahbas, A. Hadmi and A. Radgui, "Scenes segmentation in self-driving car perception system based U-Net and FCN models," in *Proc. of ICATH*, Rabat, Morocco, pp. 10–18, 2021.
[9] F. Rameau, H. Ha, K. Joo, J. Choi, K. Park *et al.,* "A real-time augmented reality system to see-through cars," *IEEE Transactions on Visualization and Computer Graphics*, vol. 22, no. 11, pp. 2395–2404, 2016.
[10] S. Yu, C. Zhao, L. Song, Y. Li and Y. Du, "Understanding traffic bottlenecks of long freeway tunnels based on a novel location-dependent lighting-related car-following model," *Tunnelling and Underground Space Technology*, vol. 136, no. 1, pp. 105098, 2023.
[11] B. Yuan, Y. A. Chen and S. Z. Ye, "A lightweight augmented reality system to see-through cars," in *Proc. of IIAI-AAI*, Yongpa, Japan, pp. 855–860, 2018.
[12] P. Gomes, C. Olaverri-Monreal and M. Ferreira, "Making vehicles transparent through V2V video streaming," *IEEE Transactions on Intelligent Transportation Systems*, vol. 13, no. 2, pp. 930–938, 2012.
[13] P. Gomes, F. Vieira and M. Ferreira, "The see-through system: From implementation to test-drive," in *Proc. of VNC*, Seoul, Korea, pp. 40–47, 2012.
[14] M. Ferreira, P. Gomes, M. K. Silvéria and F. Vieira, "Augmented reality driving supported by vehicular ad hoc networking," in *Proc. of ISMAR*, Adelaide, Australia, pp. 253–254, 2013.
[15] H. Li and F. Nashashibi, "Multi-vehicle cooperative perception and augmented reality for driver assistance: A possibility to 'see' through front vehicle," in *Proc. of ITSC*, Washignton DC, USA, pp. 242–247, 2011.
[16] F. Rameau, H. Ha, K. Joo, J. Choi, K. Park *et al.,* "A real-time vehicular vision system to seamlessly see-through cars," in *Proc. of ECCV*, Amsterdam, The Netherlands, pp. 209–222, 2016.
[17] C. Olaverri-Monreal, P. Gomes, R. Fernandes, F. Vieira and M. Ferreira, "The See-Through System: A VANET-enabled assistant for overtaking maneuvers," in *2010 IEEE Intelligent Vehicles Symp.*, California, USA, pp. 123–128, 2010.
[18] Y. Xing, C. Lv, H. Wang, D. Cao, E. Velenis *et al.,* "Driver activity recognition for intelligent vehicles: A deep learning approach," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 6, pp. 5379–5390, 2019.
[19] G. Ros, L. Sellart, J. Materzynska, D. Vazquez and A. M. Lopez, "The synthia dataset: A large collection of synthetic images for semantic segmentation of urban scenes," in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, pp. 3234–3243, 2016.
[20] J. Chen, B. Yuan and M. Tomizuka, "Model-free deep reinforcement learning for urban autonomous driving," in *Proc. of ITSC*, Auckland, New Zealand, pp. 2765–2771, 2019.
[21] K. Yu, L. Lin, M. Alazab, L. Tan and B. Gu, "Deep learning-based traffic safety solution for a mixture of autonomous and manual vehicles in a 5G-enabled intelligent transportation system," *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 7, pp. 4337–4347, 2020.
[22] Q. Zhang, H. Sun, Z. Wei and Z. Feng, "Sensing and communication integrated system for autonomous driving vehicles," in *Proc. of INFOCOM*, Toronto, ON, Canada, pp. 1278–1279, 2020.
[23] B. Zhang, E. S. Wilschut, D. M. Willemsen, T. Alkim and M. H. Martens, "The effect of see-through truck on driver monitoring patterns and responses to critical events in truck platooning," in *Advances in Human Aspects of Transportation*, Los Angeles, California, USA, pp. 842–852, 2018.
[24] F. FCC, "Amendment of the commission's rules regarding dedicated short-range communication services in the 5.850–5.925 GHz band," https://www.fcc.gov/document/amendment-commissions-rules-regarding dedicated-short-range-0/ Report and Order 03-324, Tech. Rep, 2003.
[25] D. Jiang and L. Delgrossi, "Towards an international standard for wireless access in vehicular environments," in *Proc. of VTC*, Marina Bay, Singapore, pp. 2036–2040, 2008.
[26] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard *et al.,* "Backpropagation applied to handwritten zip code recognition," *Neural computation*, vol. 1, no. 4, pp. 541–551, 1989.

[27] A. Krizhevsky, I. Sutskever and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems*, vol. 25, no. 1, pp. 1, 2012.
[28] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," arXiv:1409.1556, 2014.
[29] K. He, X. Zhang, S. Ren and J. Sun, "Deep residual learning for image recognition," in *Proc. of CVPR*, NV, USA, pp. 770–778, 2016.
[30] M. D. Zeiler, G. W. Taylor and R. Fergus, "Adaptive deconvolutional networks for mid and high level feature learning," in *Proc. of CVPR*, Boston, MA, USA, pp. 2018–2025, 2011.
[31] J. Long, E. Shelhamer and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. of CVPR*, New York, USA, pp. 3431–3440, 2015.
[32] H. Noh, S. Hong and B. Han, "Learning deconvolution network for semantic segmentation," in *Proc. of ICCV*, NW Washington, USA, pp. 1520–1528, 2015.
[33] O. Ronneberger, P. Fischer and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. of MICCAI*, Munich, Germany, pp. 234–241, 2015.
[34] C. Peng, X. Zhang, G. Yu, G. Luo and J. Sun, "Large kernel matters—improve semantic segmentation by global convolutional network," in *Proc. of CVPR*, Honolullu, USA, pp. 4353–4361, 2017.
[35] P. Sermanet, D. Eigen, X. Zhang, M. Mathieu, R. Fergus *et al.,* "OverFeat: Integrated recognition, localization and detection using convolutional networks," arXiv:1312.6229, 2013.
[36] G. Papandreou, I. Kokkinos and P. A. Savalle, "Modeling local and global deformations in deep learning: Epitomic convolution, multiple instance learning, and sliding window detection," in *Proc. of CVPR*, Boston, USA, pp. 390–399, 2015.
[37] P. Wang, P. Chen, Y. Yuan, D. Liu, Z. Huang *et al.,* "Understanding convolution for semantic segmentation," in *Proc. of WACV*, Lake Tahoe, NV, USA, pp. 1451–1460, 2018.
[38] S. Lazebnik, C. Schmid and J. Ponce, "Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories," in *Proc. of CVPR*, New York, USA, pp. 2169–2178, 2006.
[39] K. Grauman and T. Darrell, "The pyramid match kernel: Discriminative classification with sets of image features," in *Proc. of ICCV*, Las Vegas, USA, pp. 1458–1465, 2005.
[40] W. Liu, A. Rabinovich and A. C. Berg, "ParseNet: Looking wider to see better," arXiv:1506.04579, 2015.
[41] H. Zhao, J. Shi, X. Qi, X. Wang and J. Jia, "Pyramid scene parsing network," in *Proc. of CVPR*, New York, USA, pp. 2881–2890, 2017.