

Article

# Robust Drivable Road Region Detection for Fixed-Route Autonomous Vehicles Using Map-Fusion Images

Yichao Cai <sup>1,†</sup>, Dachuan Li <sup>2,†</sup>, Xiao Zhou <sup>1,2,\*</sup> and Xingang Mou <sup>1</sup>

<sup>1</sup> School of Mechanical and Electronic Engineering, Wuhan University of Technology, Wuhan 430070, China; easoncai@whut.edu.cn (Y.C.); sunnymou@whut.edu.cn (X.M.)

<sup>2</sup> California PATH, University of California, Berkeley, Richmond, CA 94804-2468, USA; dachuanli@berkeley.edu

\* Correspondence: zhouxiao@whut.edu.cn; Tel.: +1-510-570-9638

† These authors contributed equally to this work.

Received: 30 August 2018; Accepted: 20 November 2018; Published: 27 November 2018



**Abstract:** Environment perception is one of the major issues in autonomous driving systems. In particular, effective and robust drivable road region detection still remains a challenge to be addressed for autonomous vehicles in multi-lane roads, intersections and unstructured road environments. In this paper, a computer vision and neural networks-based drivable road region detection approach is proposed for fixed-route autonomous vehicles (e.g., shuttles, buses and other vehicles operating on fixed routes), using a vehicle-mounted camera, route map and real-time vehicle location. The key idea of the proposed approach is to fuse an image with its corresponding local route map to obtain the map-fusion image (MFI) where the information of the image and route map act as complementary to each other. The information of the image can be utilized in road regions with rich features, while local route map acts as critical heuristics that enable robust drivable road region detection in areas without clear lane marking or borders. A neural network model constructed upon the Convolutional Neural Networks (CNNs), namely FCN-VGG16, is utilized to extract the drivable road region from the fused MFI. The proposed approach is validated using real-world driving scenario videos captured by an industrial camera mounted on a testing vehicle. Experiments demonstrate that the proposed approach outperforms the conventional approach which uses non-fused images in terms of detection accuracy and robustness, and it achieves desirable robustness against undesirable illumination conditions and pavement appearance, as well as projection and map-fusion errors.

**Keywords:** drivable road region detection; autonomous vehicles; map-fusion image; FCNs

## 1. Introduction

Environment perception is a critical technical issue for autonomous vehicles and significant progress has been made during the past decade. In particular, road detection is one of the key aspects of environment perception, since identifying key elements of the road (such as road markings and lane borders) and drivable road regions is essential for providing valid road information to support decision-making and to navigate the vehicle on roads [1].

In the area of autonomous driving systems, various environment perception systems have been designed based on different perceptions sensors (monocular/stereo camera, LiDAR and radar, etc.), among which LiDAR and camera are the most commonly used sensors for environment perception. Although LiDAR has been applied to perception tasks in several autonomous vehicle systems and demonstrated its effectiveness in experimental tests [2,3], its application in drivable road detection is still limited by its disadvantages such as high cost, low resolution and lack of texture information.

Cameras are preferable to LiDAR in road detection due to their low cost, light weight, high resolution and simplicity in installation and configuration.

For camera-based perception systems, the objective of drivable road region detection is to identify a region of pixels that indicates the drivable (or navigation-free) area in a given image captured by the camera, and examples of applications of camera in road/lane detection can be found in [4–9]. The real-world autonomous driving scenarios pose significant challenges to camera-based perception systems, due to the following factors: (1) Unstructured road environments: road markings and lane borders are not always available, and the marking/borders may be too vague to be identified; (2) variable illumination conditions: the images may contain shadows and other undesirable illumination conditions; (3) road curvatures: the camera's field-of-view may not capture the entire region needed due to curved road segments; (4) ununiform pavement appearance and occlusions: the road pavement in the image may contain variable texture and color, and the objects in the camera's field-of-view may cause occlusions in the image. Despite the progress made in the road detection systems, the robust and effective detection of drivable regions in variable and challenging road environments remains a major challenge to be addressed.

To address the above issues for autonomous vehicles, this paper proposes a drivable road region detection approach using a monocular vehicle-mounted camera. This research focuses particularly on autonomous vehicles that have access to route maps (e.g., shuttles, buses and other vehicles operate on predefined fixed routes). Taking advantage of the route map and the vehicle location information, the proposed approach extracts from the full route map a local map which corresponds to the current processed image captured by the camera, which is then fused with the projected image to obtain a map-fusion image (MFI). The MFIs enable robust drivable region detection in the presence of undesirable illumination conditions, ununiformed pavement appearance, occlusions and absence of road/lane marking. In addition, for multi-lane scenarios, the lane navigation information (the lane which the vehicle is supposed to operate in) can also be provided in the MFI. The road detection approach is built upon the Fully Convolutional Network (FCN)-VGG16 framework which is trained using the MFIs. Therefore, the drivable road regions in each MFI are identified using the trained FCN-VGG16. The proposed approach is validated using video data collected from real-world scenarios, and experimental results show that the proposed approach achieves high accuracy and robustness in drivable road region detection in various driving scenarios and outperforms contemporary approach.

The remainder of this paper is organized as follows: Section 2 reviews the existing approaches for road detection. Section 3 formulates the drivable road detection problem. Details of the proposed road detection approach is presented in Section 4. Experimental results based on real-world scenarios are presented and analyzed in Section 5, and the paper is finally concluded in Section 6.

## 2. Related Works

In general, drivable road region detection can be classified as a road detection problem. During the past decade, various computer vision-based approaches have been proposed to address road detection problem for autonomous vehicle systems, and they can be generally divided into three categories. One straightforward paradigm detects the road regions by extracting the lane markings and road boundaries. This category of approaches utilizes conventional image processing techniques to extract the road boundaries by using image features such as edge, position and color [4–9]. After that, the boundaries are fitted (linear models [6,7] and high-order curves/splines are used for straight and curved road segments [8], respectively), and the area between the identified lane markings and/or boundaries is considered as the drivable road region. The conventional approaches have the primary advantage of offering the opportunity to evaluate specific performance criteria and factors that affect the performance, such as descriptors used. However, these approaches are sensitive to illumination variation and occlusion, and their performance may degrade significantly when dealing with roads with vague or discontinuous lane markings. Another category of approaches detects roads by analyzing texture features and attempting to find the inhomogeneity between road area and backgrounds in the

image [10–14]. To specify texture information in the image, Gabor Filter [10,13] and discrete cosine transform [14] are commonly used in this type of road detection approaches and they can achieve desirable performance. Although this type of approaches performs well on rural, single-lane roads, their performance is undesirable when detecting drivable region on multi-lane roads. In general, the conventional approaches described above achieve desirable road detection performance only in specific scenarios and are less effective in more complex driving environments. For instance, the lane on the opposite direction of a two-way road may be incorrectly identified as drivable road region by the marking/boundary-based road detectors. In addition, there are usually multiple driving directions in intersection regions and it is difficult to identify correct drivable road regions only from boundaries and texture information. Therefore, the application of the conventional marking/boundary-based and texture-based approaches to real-world driving scenarios are still very limited due to their limitations.

As alternatives to the conventional approaches, machine-learning based semantic segmentation approaches have emerged as effective tools [15] in computer vision area, and due to their high precision and efficiency, they have been applied to a number of autonomous driving systems [16–19] in recent years. Semantic segmentation approaches can achieve pixel-level precision and demonstrate effectiveness in many road detection applications, and typical semantic segmentation algorithms are built upon CNNs, as proposed in [20–24]. A more recent example of such approaches is the Vanishing Point Guided Network (VPG-Net) [23] which can detect road under complex environmental condition. However, such CNN-based models typically process images using a local block as the input, which requires high computational cost. The performance of machine-learning-based approach are heavily affected by the dataset used to train the network and the size of required dataset may increase dramatically if the network is required to adapt to various use cases. In order to improve the efficiency of CNNs, the Fully Convolutional Networks (FCNs) is proposed in [25], which can recover the class of each pixel from abstracted features in the image. Since it is proposed, a number of research works have focused on making extensions and improvements to FCNs and applying the approach in the field of autonomous driving systems [26–28]. One of the most preferred variants is the MultiNet [28] which is a codec structure model based on FCNs and VGG [27], and it can accomplish joint classification, detection and semantic segmentation via a unified framework. The encoder of MultiNet is designed to extract the features from image, while the decoder is responsible for multiple tasks such as road detection. In the applications of road detection, the training parameters can be initialized with pre-trained VGGs to speed up the training speed, and the road region is detected using FCNs. MultiNet demonstrates appealing performance both in structured and unstructured road with a desirable trade-off between accuracy and training cost. Therefore, taking advantage of the machine-learning based frameworks, the approach proposed in this paper uses an FCN-VGG16 framework to identify drivable road region from map-fusion images.

On the other hand, real-world contains many complex driving scenarios that may be difficult to be handled by systems with single perception data source. As described previously, for intersections containing multiple lanes and multiple driving directions, it can be difficult to determine which direction the vehicle should be driving on by using solely the information of images. Therefore, fusing information from multiple sources is a preferable choice to improve the performance of road detection. [29] proposes a lane marking detection framework that fuses GPS data, vehicle velocity and lane map, and it demonstrates good performance in structured road scenarios. Similarly, GPS data and image are fused for localization and decision making in [30]. Inspired by the previous research, the vehicle location data and route map data are used to generate local route maps in the proposed framework. The vehicle location data can either be obtained from vehicle-mounted GPS or other data feeds such as point cloud map-matching of LiDAR data. We fuse a local route map and its corresponding image captured by vehicle-mounted camera into a map-fusion image (MFI), which provides high level heuristics for the drivable region detection.

### 3. Problem Formulation

Drivable road region detection can be considered as a subtask of road detection which involves all key elements in the road environment. Without considering the pedestrians, vehicles and other obstacles, the drivable road region can be defined as the specific road area that the vehicles are supposed to drive in. For the monocular camera-based systems, the data precepted from the driving environment are the images captured by the vehicle-mounted camera. As shown in Figure 1a, a typical image of the driving environment can be divided into three parts: the background area, the drivable area and the non-drivable road area. Given an image  $I$  (Figure 1a) captured by the camera at certain timestamp  $t$ , the definitions used in drivable road region detection are as follows:

$S$ —the set of all pixels in the image  $I$ ,  $S = \{p|p \in I\}$

$S_b$ —the set of pixels in the background area (or non-road area)

$S_r$ —the set of pixels in the road region

$S_d$ —the set of pixels in the drivable road region

$S_{ud}$ —the set of pixels in the non-drivable road region.

The relationship between the above sets is as follows (Figure 1b):

$$S = S_b \cup S_r, \quad (1)$$

$$S_b \cap S_r = \emptyset, \quad (2)$$

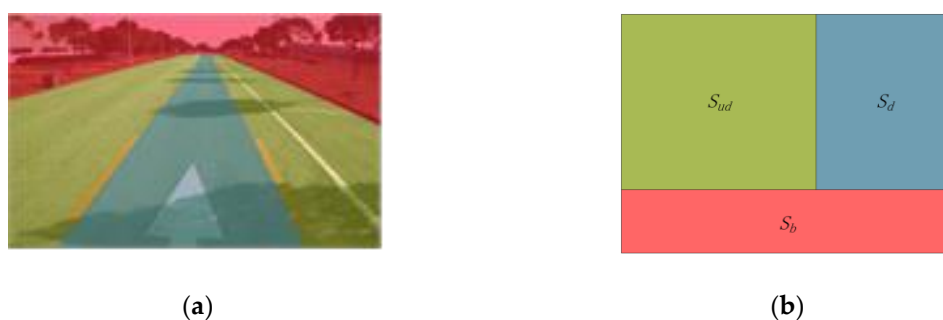
$$S_r = S_d \cup S_{ud}, \quad (3)$$

$$S_d \cap S_{ud} = \emptyset, \quad (4)$$

Given the above definitions, the monocular camera-based road region detection problem can be formulated as follows:

**Definition 1 (Drivable Road Region Detection).** Given a set of pixels  $S$  in an image  $I$ , extract the pixels of  $S_d$  from  $S$ .

In this paper, the objective of drivable road region detection is to extract the drivable regions in the captured images that correspond with the pre-specified route. In addition, for vehicles operating on fixed routes, note that the drivable road regions include regions of both single road segments and intersections.

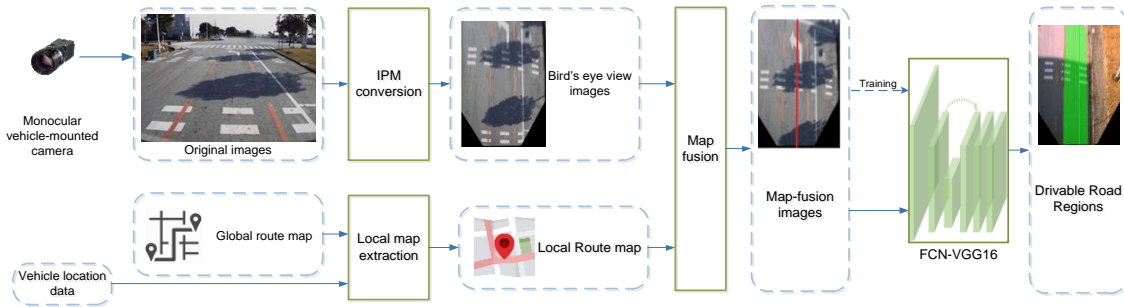


**Figure 1.** Definition of drivable region detection: (a) Regions in an image captured by the camera; (b) Relationship between the regions.

### 4. Drivable Road Region Detection

The overall framework of the proposed approach is depicted in Figure 2. Our approach consists of three primary processes: (1) Image pre-processing: the original images captured by the camera are swapped into bird's-eye view images using Inverse Perspective Mapping approach. (2) Map fusion:

the homogenous local route map of the processed image is extracted from a global route map with the assistance of vehicle location data, and it is then fused with the image to generate a map-fusion image (MFI). (3) Drivable road region detection: a FCNs-based deep neural network (FCN-VGG16) is utilized to detect drivable road region from the MFIs. Note that the FCN-VGG16 are also trained using datasets of MFI images. Details of each process are presented in the following subsections.



**Figure 2.** Overall Framework of the drivable road region detection based on map-fusion and monocular camera.

#### 4.1. Inverse Perspective Mapping

Vehicle-mounted cameras are typically installed in a way that the lens face toward the forward-driving direction of the vehicle. As depicted in Figure 3a, images captured by the vehicle-mounted camera are normally in the front-view, which is more suitable for certain detection applications such as pedestrian detection [31] and traffic-sign detection [32]. However, for drivable road region detection tasks, images in front-view contain a lot of redundant information which may affect the effectiveness of the detection algorithm. Compared with the front view, the bird's-eye view is more preferable in drivable region detection applications as the image contains less redundant information, and the lane markings become roughly parallel [4] (Figure 3b).

Without loss of generality, the road regions can be assumed to be homographic planes in road detection scenarios. Therefore, we can easily map the original images into bird's eye view using Inverse Perspective Mapping (IPM). The IPM can be considered as an image projection process and in our application, the pixels in the original images can be projected to the bird's-eye-view images using a mapping matrix (projection matrix). In our application, a multiple-point correspondence-based method is used to estimate the mapping matrix  $\hat{H}$ :

$$p_i = H p'_i, \quad (i = 1, 2, \dots, n) \quad (5)$$

$$p = \hat{H} p', \quad (6)$$

where  $p'$  and  $p$  represent the homogeneous coordinates of pixels in the original image  $I'$  ( $p' \in I'$ ) and the projected image  $I$  (bird's-eye view) ( $p \in I$ ), respectively. The mapping matrix  $\hat{H}$  can be estimated in two-steps: First, select two sets of pixels  $p'_i$  and  $p_i$  ( $i = 1, 2, \dots, n$ ) from  $I'$  and  $I$ , respectively, and the optimal estimate ( $\hat{H}$ ) of the mapping matrix  $H$  can be calculated based-on Equation (5). Once  $\hat{H}$  is determined, all pixels of the front-view image  $I'$  can be projected into a bird's-eye view image  $I$  using the estimated  $\hat{H}$  (Equation (6)). (Note that the IPM is used in our approach to reduce redundant information instead of trying to ensure the lane markings are parallel.)

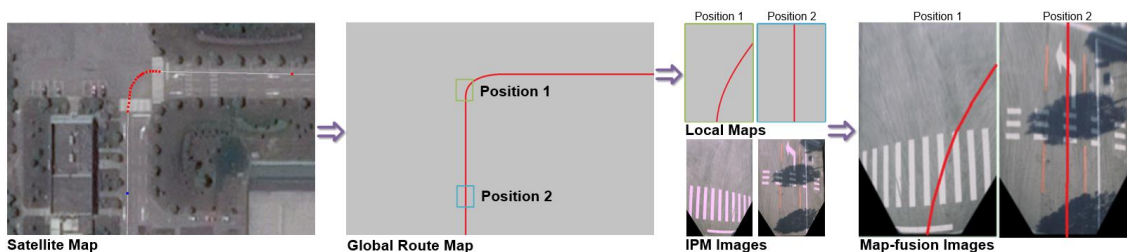


**Figure 3.** Images in front view and bird's-eye view: (a) Front-view image; (b) Bird's-eye-view image.

Note that the above described estimation of projection matrix  $\hat{H}$  is a pre-calibration procedure, and the IPM process in our system uses the fixed  $\hat{H}$  calculated via the pre-calibration. As a matter of fact, the roll and pitch movements of the vehicle can cause variations of  $\hat{H}$ , and therefore, a more accurate way of IPM is to adjust  $\hat{H}$  based on the measurement of the vehicle's roll and pitch in real-time. However, considering that the vehicle performs minor roll and pitch movements at low speeds in regular driving environments, the proposed drivable road detection approach can achieve robustness against such movements and small IPM error (see Section 5.4). Therefore, it is reasonable to use a fixed, pre-calibrated  $\hat{H}$  in our application.

#### 4.2. Map-Fusion Image Generation

Figure 4 depicts the process of map-fusion image generation. The overall process consists of four steps which are described in the following subsections.



**Figure 4.** Generation of map-fusion images.

##### 4.2.1. Map-Fusion Image Generation

For vehicles driving on fixed routes (e.g., shuttles, buses and school buses) or vehicles with pre-established high-precision maps, the route maps provide critical heuristics for the environment perception and navigation of vehicles. A route map is defined as a map that consists of roads/lanes, navigation points and driving directions of the routes that the vehicle is supposed to operate on. In our system, the route map is presented and stored in the form of straight and curved segments (Figure 4), and each point in the route map represents a certain geographical location of the vehicle in the corresponding segment. In addition, for fixed routes, there is only one forward driving direction for the vehicle in a straight road segment or intersection, and this driving direction information is also included in the route map. The global map can be generated using various data sources such as GPS maps and high-precision 2D maps. For the framework proposed in this paper, a satellite map with geo-location information (i.e., GPS coordinates) is used to generate the global route map (Figure 4).

##### 4.2.2. Local Map Extraction

When the vehicle operates on the route which is pre-specified in the global route map, a local route map that corresponds with the road region in the current image from the camera is extracted from the

route map every time an image is captured by the vehicle-mounted camera. In the proposed framework, the matching between the local route map and images is based on GPS data from the vehicle-mounted GPS module. For each frame of image and GPS coordinates, we search its corresponding position and localize the corresponding region in the global route map using GPS coordinates. After that, a local route map with a look-ahead length of 20 m can be extracted from the global route map, using the current position of the vehicle (according to the installation position of the camera, the equivalent height of each bird's-eye-view image is approximately 20 m above the road surface, leading to a look-ahead distance of 20 m). Note that vehicle location from other data feeds (e.g., LiDAR point cloud matching) can also be utilized in the proposed approach.

The synchronization of the images and GPS data is a major issue in local route map extraction, as the update rates of the vehicle-mounted camera and the GPS module may be different. In our approach, it is appropriate to use the GPS data of which the timestamp is closest to that of the image, under the assumption that the vehicle would not move a long distance within milliseconds.

#### 4.2.3. Map-Image Fusion

After the local map is extracted, the image is fused with the corresponding local route map to generate a map-fusion image (MFI), where the road image contains the primary information, and the local route data provide heuristic navigation information of the autonomous vehicle, including location, specified lane and driving direction.

As depicted in Figure 4, in the map-fusion image, the original bird's view image is overlaid with line or curved segment representing the pre-defined route of vehicle specified in the corresponding local route map (Figure 4). The midpoint of the bottom border of the image represent the position of the vehicle at the moment the image is captured. Therefore, the origin of the line/curvature in the local map is aligned with the midpoint of the bottom border of the image. Experiment results indicate that the color and width of the line have slight effect on the detection results. Therefore, in our application, we use green-colored (RGB: 0, 255, 0) lines with a width of 2 pixels to represent the route in the MFI for the subsequent drivable road region detection process. However, for visualization purposes, broad red lines are used to represent the route in Figure 4.

### 4.3. FCN-VGG16 for Drivable Road Region Detection

#### 4.3.1. Neural Network Model

As described in previous subsections, CNNs demonstrate desirable performance in various detection tasks due to its high efficiency and storage-saving characteristics. The neural network model used in this research follows an extension of the CNN model, namely the MultiNet, which is an end-to-end model firstly proposed in [28]. The MultiNet is a multi-task framework and the semantic segmentation model FCN-VGG16 of MultiNet is adopted in the proposed framework for drivable road region detection. The FCN-VGG16 model consists of two major components, a CNN encoder and a segmentation decoder. As an encoder in our framework, a pre-trained VGG16 network which consists of 13 convolutional layers is used to extract features from the MFI images. For the decoder module, the extracted features are processed by a cascade of convolutional layers and up-sampling layers to generate the output. A set of  $1 \times 1$  convolutional layers is responsible for generating a low resolution segmentation. After that, three transposed convolution layers performs up-sampling of the low-resolution segmentation. Finally, skip layers composed by  $1 \times 1$  convolutional layers extract high-resolution features and output the image with detected drivable road regions. The architecture of the FCN-VGG16 model is depicted in Figure 5.

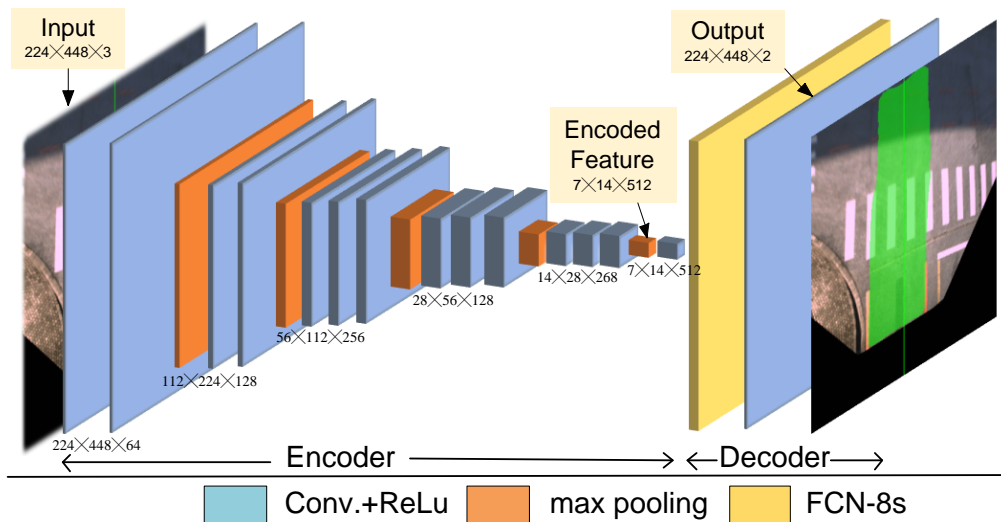


Figure 5. Segmentation model based on FCN-VGG16 [28].

#### 4.3.2. Selection of Loss Function

The Cross-entropy function is widely used as loss function in neural networks as it can effectively avoid the gradient vanishing problem. Therefore, loss function adopted in the training process of the FCN-VGG16 model is a cross-entropy function which is defined as follows:

$$Loss(P, G) = \frac{1}{|I|} \sum_{i \in I} (G_i \log P_i), \quad (7)$$

where  $I$  denotes the MFI inputs,  $|I|$  represents the size of the image,  $i$  represents a pixel in the MFI,  $G_i$  is the ground truth of  $i$ , and  $P_i$  is the predicted binary value of  $i$ . If  $i$  falls in the predicted drivable region, the value of  $P_i$  is set to 1, otherwise  $P_i$  equals to 0.  $G_i$  is calculated in a similar way.

## 5. Experimental Study

### 5.1. Experiment Platform

The proposed map-fusion and FCN-VGG16-based drivable road region detection was tested using real-world driving scenario videos captured by a camera (Figure 6a) mounted on an 18-meter bus (Figure 6b). Parameters of the vehicle-mounted are listed in Table 1. To obtain a desirable field of view, the camera was installed down the center of the upper border of the windshield and below the headliner of the bus. Testing drive was conducted on a 4.3 km testing track which consists of 4 intersections, and videos of the testing drive and the corresponding GPS data were recorded to test the drivable road region detection algorithm.



Figure 6. Experiment Platform: (a) 18 m testing bus, (b) vehicle-mounted camera.



**Table 1.** Parameters of the vehicle-mounted camera.

Parameters	
Model	Daheng MER-125-30-UC
CCD sensor	Sony <sup>®</sup> ICX445
Resolution	1292 × 964
Frame rate	30 fps
Lens	Basler 4 mm fixed focus length

To reduce the training time of the FCN-VGG16 model, we normalize the converted IPM images to a resolution of 224 × 448 (The original resolution is 1292 × 964, see Table 1). The experiments were conducted on a PC with one Intel<sup>®</sup> Core i5-3570 CPU and one NVIDIA<sup>®</sup> 1080Ti GPU. With a learning rate of 0.00001, it took 3.5 h to train the FCN-VGG16 network for 30000 times using the MFI images.

### 5.2. Evaluation Metric

The Pixel Accuracy (PA) [15] is used as the metric for evaluating the proposed approach. Since drivable road region detection can be regarded as a single-class semantic segmentation task, the calculation of PA can be simplified as follows:

$$PA = \frac{p_{tp}}{p_{fp} + p_{tp} + p_{fn}}, \quad (8)$$

where,  $p_{tp}$ ,  $p_{fp}$  and  $p_{fn}$  represent the number pixels of true positive, false positive and false negative cases detected by the algorithm, respectively. Therefore, high Pixel Accuracy value indicate that the algorithm achieves good performance in detecting drivable road regions.

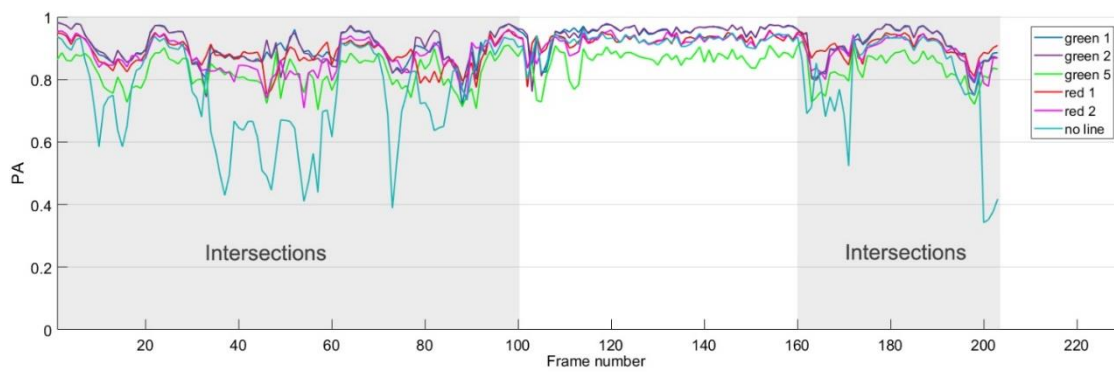
### 5.3. Evaluation of Detection Accuracy

A dataset of 270 frames is selected from the recorded video to evaluate the performance of the proposed approach. The dataset includes straight road segments, curved segments and intersections. For comparison purposes, the algorithm is tested using both map-fusion images (MFI) and raw images (without map-fusion). The PA of detection results are calculated for each frame in both cases.

The route line added to the MFI can be considered as artificial lane-markings. Therefore, the choice of color and width of the route line may affect the drivable road region detection results. The color of the route line should be distinct from the color of the road region in the image, and the width of the line is relevant to the intensity of the high-level intervention. We test different choices of colors and widths of the lines in local maps to demonstrate the effect of MFI in drivable road region detection. The performance (PA) of different choice of line colors and widths for each frame are depicted in Figure 7 (The cases using raw images without map-fusion are marked as ‘no line’).

As shown in Figure 6, the testing site contains two intersections and turning sections without lane markings, and they correspond with the first 100th frames, and the sections from the 160th to 203th frames of the dataset (the other sections correspond with regular straight road segments), respectively. As can be seen from Figure 7, the PA value of detection is quite low at intersections and turning segments (curve marked as ‘no line’) due to the absence of lane markings, and there is little information for the algorithm to determine the drivable road region. Therefore, the route line in the map-fusion image provides important heuristics for drivable road region detection.

As for the route line, experiment results demonstrate mild difference using different choices of line colors and width, as shown in Figure 6. Experiment results also suggest using colors that are distinctive from the color of road surface, such as green and red. In our experiment, we chose to use the green color and the width of 2 pixels for the route line in the MFI (please note that the route lines are rendered in red in the subsequent figures just for clarity of visualization purposes).

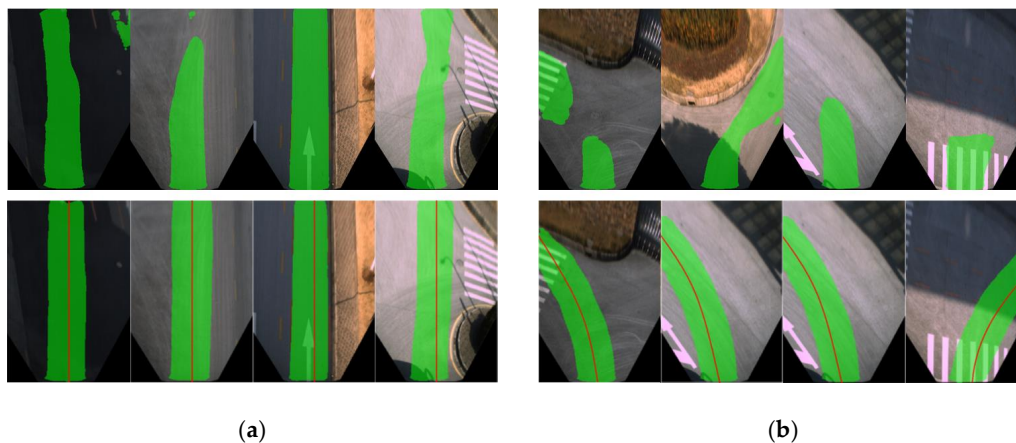


**Figure 7.** PA of detection for each frame.

By calculating the average PA value of all frames in the dataset, the average performance in detecting accuracy of FCN-based detection algorithm using raw images and our map-fusion-based algorithm is evaluated and listed in Table 2. The evaluation results indicate that the proposed approach outperforms conventional FCN-based detection algorithm in terms of detection accuracy (PA). Examples of detection results are depicted in Figure 8, which contains typical scenarios such as intersections, straight roads and turning sections, and the detected drivable road regions are marked in green. In contrast to conventional FCN-based detection algorithm, the proposed approach can still effectively detect drivable road detections even in intersections and turning segments where lane-markings are absent.

**Table 2.** Average detection accuracy.

Approach	PA
FCN without MFI	0.811
MFI and FCN-VGG16	0.917

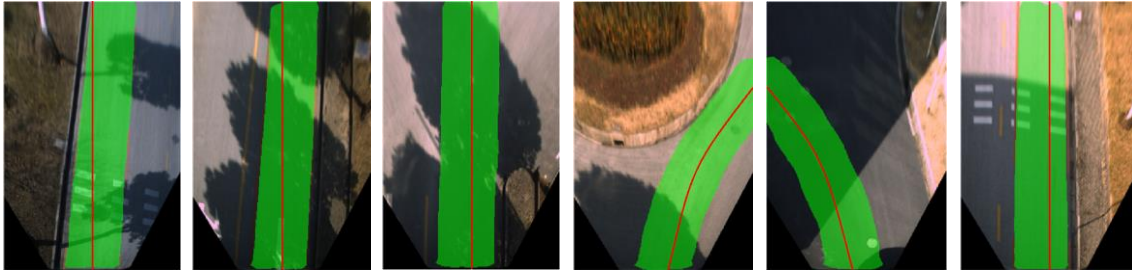


**Figure 8.** Comparison of detection results. The upper row shows the results of conventional FCN without MFI, and the lower row is the detection results of the proposed approach. (a) Detection results of straight road segments; (b) Detection results of intersections and turning segments.

#### 5.4. Evaluation of Robustness

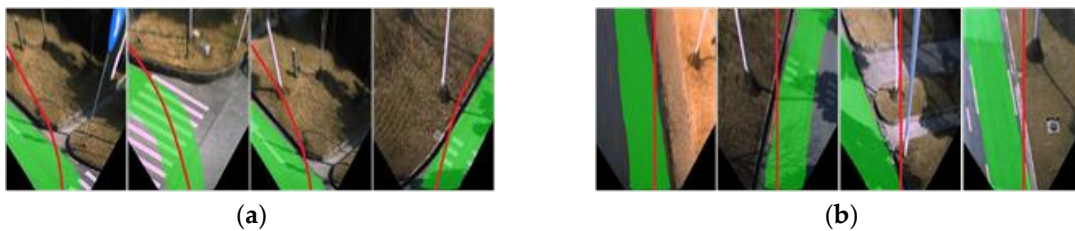
The complex road scenarios of the testing site include absence of lane/road markings, illumination variation, disturbances in road surface appearance and the presence of various road users (other vehicles, pedestrians, bicycles and other road users). In addition, errors introduced by the IPM conversion and local route map extraction process may also lead to detection deviations.

The robustness of our approach was evaluated using the above complex road environments. Examples of the testing results are shown in Figures 9–11. As described in Section 5.3, the proposed approach demonstrates desirable robustness in the absence of road/lane markings (Figure 8). Figure 9 depicts examples of typical illumination variation scenarios including heavy shadows and variations of illumination direction. As can be seen from the figure, the proposed approach achieves desirable robustness in such scenarios, while conventional border or feature-based approaches may fail.



**Figure 9.** Robustness against illumination variation.

Figure 10 demonstrates examples of incorrect local route maps (Figure 10a) and IPM conversions (Figure 10b) due to errors. As can be seen from the figures, the drivable road region can still be detected by our approach even when the route lines are placed in totally incorrect positions of the image. These results indicate that the combination of map-fusion and FCN-VGG16 can improve the robustness of drivable road detection in complex road environments.



**Figure 10.** Robustness against conversion and fusion errors: (a) Detection results in the presence of local map extracting deviation; (b) Detection results in the presence of IPM errors.

Figure 11 shows the detection results of the proposed approach in the presence of undesirable road surface appearance. Still, our approach demonstrates robustness against large pavement cracks (Figure 11a) and inconsistent pavement colors (Figure 11b).



**Figure 11.** (a) Detection results in the presence of large cracks; (b) Detection results in the presence of inconsistent pavement colors.

Figures 9 and 10 present the performance of drivable road region detection in the presence of undesirable factors induced by illumination and road appearance. In most real-world driving scenarios, another aspect of the complexity of road environments is that the road region is usually shared by various road users such as other vehicles, pedestrians, and motorcycles, etc. We tested the propose drivable road region detection approach in various scenarios with other road users, as shown in Figures 12–14 (for visualization purposes, these figures are shown in the front-view). Figure 12 presents the scenario with single, continuously moving pedestrian. The pedestrian walked across the lane the ego vehicle was driving in and away from the testing vehicle. As can be seen from Figure 12, the proposed approach can correctly and continuously extract drivable road regions as the distance between the pedestrian and the ego vehicle increased.



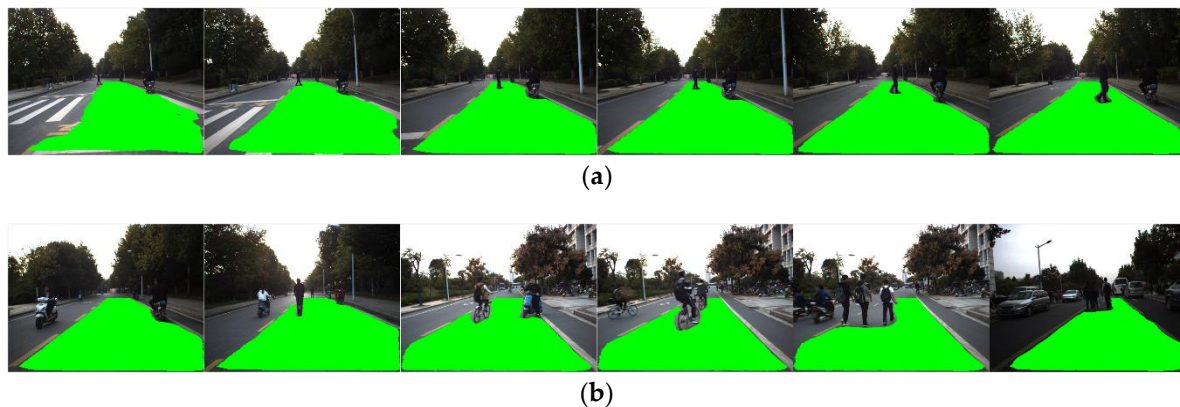
**Figure 12.** Detection results in the presence of single, continuously-moving pedestrian.

Figure 13 shows detection results of various scenarios with different types of road users other than pedestrians (from left to right: other vehicle, motorcycle and bicycle). As can be seen from Figure 13, the road users are of different types and sizes, with different distances and orientation relative to the ego vehicle. As the FCN-VGG16 network was trained using image datasets containing various road users in different scenarios, the proposed approach can effectively identify drivable road regions when different types of road users share the lane the ego vehicle was driving in.



**Figure 13.** Detection results in the presence of various types of other road users.

Testing results in the presence of multiple and various road uses are shown in Figure 14. Figure 14a shows the scenario with a pedestrian and a motorcycle in the lane of the ego vehicle. The pedestrian walked across the lane and toward the ego vehicle, while the motorcycle drove in the same direction as the ego vehicle did with changing distances. Similar to the case shown in Figure 12, the proposed approach can continuously identify drivable road region as the two road users moved in the view of the camera. Figure 14b include more complex, multiple-road users driving scenarios that are common in real-world road environments. In such scenarios, the lane can be shared by multiple road users of different types, and they may drive in different directions. In addition, the opposite lane is usually occupied by other road users at the same time. As can be seen from Figure 14b, our approach can effectively deal with such scenarios and the testing results indicate that the proposed approach has the potential to be applied to autonomous vehicle systems operating in real-world driving scenarios.



**Figure 14.** Detection results in the presence of: (a) Multiple road users; (b) various road users.

## 6. Conclusions

This paper proposes a robust drivable region detection approach based on a monocular vehicle-mounted camera for autonomous vehicles. The proposed approach takes advantage of the heuristics from the route map and vehicle location data and fuses the local route map to generate map-fusion images to improve the robustness of detection. The map-fusion images provide significant heuristics (e.g., heading, location, specified driving direction and lane of the vehicle) for drivable road region detection in complex road environments such as turning segments and intersections in the absence of road/lane markings. In order to extract drivable road regions, the map-fusion images are segmented using an FCN-based neural network which is trained also using a dataset consist of selected map-fusion images. The proposed approach is validated using real world videos captured by a camera mounted on a testing bus. Experiment results show that the proposed approach achieves desirable detection accuracy using an industrial camera, and it demonstrates robustness in complex driving scenarios involving absent road/lane markings, undesirable illumination conditions and pavement appearance, as well as various road users. Evaluation results also indicate that the proposed approach outperforms conventional FCN-based algorithm using raw images in terms of detection accuracy.

The future work will focus on the incorporation of other vehicle location data feeds such as LiDAR point-cloud matching. More comprehensive investigation is also required to further improve robustness of the proposed drivable road region detection approach. The implementation and integration of the drivable-road detection approach to an autonomous driving system will be investigated. In addition, more testing will be conducted to further evaluate the performance of the proposed approach in more real-world driving scenarios.

**Author Contributions:** X.Z. proposed the methodology and designed the overall framework, Y.C. and D.L. accomplished the detailed design of the proposed framework and wrote the manuscript; Y.C. developed and implemented the algorithm; D.L. and X.Z. were in charge of the experiment design and the validation of the proposed approach; X.M. supervised the research.

**Funding:** This research was supported in part by National Natural Science Foundation of China (61701357) and China Scholarship Council under Grants 201606955118.

**Acknowledgments:** The authors would like to thank Kejun Ou for his assistance and insightful suggestions.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Yenikaya, S.; Yenikaya, G.; Düven, E. Keeping the vehicle on the road: A survey on on-road lane detection systems. *ACM Comput. Surv.* **2013**. [[CrossRef](#)]
2. Liu, X.; Deng, Z.; Yang, G. Drivable road detection based on dilated FPN with feature aggregation with feature aggregation. In Proceedings of the IEEE 29th International Conference on Tools with Artificial Intelligence (ICTAI), Boston, MA, USA, 6–8 November 2017; pp. 1128–1134.

3. Gu, S.; Lu, T.; Zhang, Y.; Alvarez, J.M.; Yang, J.; Kong, H. 3D LiDAR + monocular camera: An inverse-depth induced fusion framework for urban road detection. *IEEE Trans. Intell. Veh.* **2018**, *3*, 351–360. [[CrossRef](#)]
4. Shin, B.-S.; Tao, J.; Klette, R. A Superparticle Filter for Lane Detection. *Pattern Recognit.* **2015**, *48*, 3333–3345. [[CrossRef](#)]
5. Wang, Y.; Dahnoun, N.; Achim, A. A novel system for robust lane detection and tracking. *Signal Process.* **2011**, *92*, 319–334. [[CrossRef](#)]
6. Borkar, A.; Hayes, M.; Smith, M.T. Polar randomized Hough transform for lane detection using loose constraints of parallel lines. In Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Prague, Czech Republic, 22–27 May 2011; pp. 1037–1040.
7. Yi, S.-C.; Chen, Y.-C.; Chang, C.-H. A lane detection approach based on intelligent vision. *Comput. Electr. Eng.* **2015**, *42*, 23–29. [[CrossRef](#)]
8. Lim, K.H.; Seng, K.P.; Ang, L.-M. River flow lane detection and Kalman filtering-based B-spline lane tracking. *Int. J. Veh. Technol.* **2012**. [[CrossRef](#)]
9. John, N.; Anusha, B.; Kutty, K. A reliable method for detecting road regions from a single image based on color distribution and vanishing point location. *Procedia Comput. Sci.* **2015**, *58*, 2–9. [[CrossRef](#)]
10. Zhou, S.; Jiang, Y.; Xi, J.; Gong, J.; Xiong, G.; Chen, H. A novel lane detection based on geometrical model and Gabor filter. In Proceedings of the IEEE Intelligent Vehicles Symposium, San Diego, CA, USA, 21–24 June 2010; pp. 59–64.
11. Kong, H.; Audibert, J.-Y.; Ponce, J. General road detection from a single image. *IEEE Trans. Image Process.* **2010**, *19*, 2211–2220. [[CrossRef](#)] [[PubMed](#)]
12. Shang, E.; An, X.; Li, J.; Ye, L.; He, H. Robust unstructured road detection: The importance of contextual information. *Int. J. Adv. Rob. Syst.* **2013**, *10*, 1–7. [[CrossRef](#)]
13. Li, Z.-Q.; Ma, H.-M.; Liu, Z.-Y. Road lane detection with Gabor filters. In Proceedings of the International Conference on Information System and Artificial Intelligence (ISAI), Hong Kong, China, 24–26 June 2016; pp. 436–440.
14. Janda, F.; Pangerl, S.; Schindler, A. A road edge detection approach for marked and unmarked lanes based on video and Radar. In Proceedings of the 16th International Conference on Information Fusion, Istanbul, Turkey, 9–12 July 2013; pp. 871–876.
15. Garcia-Garcia, A.; Orts-Escolano, S.; Oprea, S.; Villena-Martinez, V.; Garcia-Rodriguez, J. A Review on Deep Learning Techniques Applied to Semantic Segmentation. *arXiv* **2016**, arXiv:1704.06857.
16. Zhang, W.-Z.; Wang, Z.-C. Rural road detection of color image in complicated environment. *Int. J. Signal Process. Image Process. Pattern Recognit.* **2013**, *6*, 161–168. [[CrossRef](#)]
17. Geiger, A.; Lenz, P.; Urtasun, R. Are we ready for autonomous driving? The KITTI vision benchmark suite. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Providence, RI, USA, 16–21 June 2012; pp. 3354–3361.
18. Cordts, M.; Omran, M.; Ramos, S.; Rehfeld, T.; Enzweiler, M.; Benenson, R.; Franke, U.; Roth, S.; Schiele, B. The Cityscapes Dataset for Semantic Urban Scene Understanding. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 3213–3223.
19. Oberweger, M.; Wohlhart, P.; Lepetit, V. Hands deep in deep learning for hand pose estimation. In Proceedings of the 20th Computer Vision Winter Workshop (CVWW), Seggau, Austria, 9–11 February 2015; pp. 1–10.
20. Yu, F.; Koltun, V.; Funkhouser, T. Dilated residual networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 636–644.
21. Zhao, H.; Qi, X.; Shen, X.; Shi, J.; Jia, J. ICNet for real-time semantic segmentation on high-resolution images. *arXiv* **2017**, arXiv:1704.08545.
22. Dvornik, N.; Shmelkov, K.; Mairal, J.; Schmid, C. BlitzNet: A Real-Time Deep Network for Scene Understanding. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017; pp. 2380–7504.
23. Lee, S.; Kim, J.; Yoon, J.S.; Shin, S.; Bailo, O.; Kim, N.; Lee, T.-H.; Hong, H.S.; Han, S.-H.; Kweon, I.S. VPGNet: vanishing point guided network for lane and road marking detection and recognition. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017; pp. 1965–1973.

24. Brust, C.-A.; Sickert, S.; Simon, M.; Rodner, E.; Joachim, D. Convolutional Patch Networks with Spatial Prior for Road Detection and Urban Scene Understanding. In Proceedings of the 10th International Conference on Computer Vision Theory and Applications (VISAPP), Berlin, Germany, 11–14 March 2015.
25. Shelhamer, E.; Long, J.; Darrell, T. Fully Convolutional Networks for Semantic Segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 640–651. [[CrossRef](#)] [[PubMed](#)]
26. Pham, V.-Q.; Ito, S.; Kozakaya, T. BiSeg: Simultaneous Instance Segmentation and Semantic Segmentation with Fully Convolutional Networks. In Proceedings of the British Machine Vision Conference (BMVC), London, UK, 4–7 September 2017.
27. Simonyan, K.; Zisserman, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. *arXiv* **2014**, arXiv:1409.1556.
28. Teichmann, M.; Weber, M.; Zoellner, M.; Cipolla, R.; Urtasun, R. MultiNet: Real-time Joint Semantic Reasoning for Autonomous Driving. *arXiv* **2016**, arXiv:1612.07695.
29. Kowsari, T.; Beauchemin, S.S.; Bauer, M.A. Map-based lane and obstacle-free area detection. In Proceedings of the International Conference on Computer Vision Theory and Applications, Lisbon, Portugal, 5–8 January 2014; pp. 523–530.
30. Tao, Z.; Bonnifait, P.; Frémont, V.; Ibañez-Guzman, J. Mapping and localization using GPS, lane markings and proprioceptive sensors. In Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems, Tokyo, Japan, 3–7 November 2013; pp. 406–412.
31. Li, C.; Wang, X.; Liu, W. Neural features for pedestrian detection. *Neurocomputing* **2017**, *238*, 420–432. [[CrossRef](#)]
32. Zhu, Z.; Liang, D.; Zhang, S.; Huang, X.; Li, B.; Hu, S. Traffic-Sign Detection and Classification in the Wild. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 2010–2118.



© 2018 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).