applied sciences

MDPI

Article

A Novel Method for Unexpected Obstacle Detection in the Traffic Environment Based on Computer Vision

Wenyan Ci [1], Tianxiang Xu [1], Runze Lin [2] and Shan Lu [3,*]

[1] School of Engineering, Huzhou University, Huzhou 313000, China
[2] State Key Laboratory of Industrial Control Technology, Zhejiang University, Hangzhou 310027, China
[3] Institute of Intelligence Science and Engineering, Shenzhen Polytechnic, Shenzhen 518055, China
* Correspondence: lushan@szpt.edu.cn

Abstract: Obstacle detection is the basis for the Advanced Driving Assistance System (ADAS) to take obstacle avoidance measures. However, it is a very essential and challenging task to detect unexpected obstacles on the road. To this end, an unexpected obstacle detection method based on computer vision is proposed. We first present two independent methods for the detection of unexpected obstacles: a semantic segmentation method that can highlight the contextual information of unexpected obstacles on the road and an open-set recognition algorithm that can distinguish known and unknown classes according to the uncertainty degree. Then, the detection results of the two methods are input into the Bayesian framework in the form of probabilities for the final decision. Since there is a big difference between semantic and uncertainty information, the fusion results reflect the respective advantages of the two methods. The proposed method is tested on the Lost and Found dataset and evaluated by comparing it with the various obstacle detection methods and fusion strategies. The results show that our method improves the detection rate while maintaining a relatively low false-positive rate. Especially when detecting unexpected long-distance obstacles, the fusion method outperforms the independent methods and keeps a high detection rate.

Keywords: unexpected obstacle detection; computer vision; semantic segmentation; open-set recognition algorithm; uncertainty degree; Bayesian fusion

Citation: Ci, W.; Xu, T.; Lin, R.; Lu, S. A Novel Method for Unexpected Obstacle Detection in the Traffic Environment Based on Computer Vision. *Appl. Sci.* **2022**, *12*, 8937. https://doi.org/10.3390/app12188937

Academic Editors: Jose Santamaria and Zhengjun Liu

Received: 16 July 2022
Accepted: 4 September 2022
Published: 6 September 2022

1. Introduction

Road traffic safety hazards are currently a common social safety issue all over the world. In terms of vehicle type, car-related accidents account for more than two-thirds of all and are the most harmful. To improve traffic safety, the automotive industry developed airbags, anti-lock braking systems, and electronic stability systems were used in the early days. In recent years, research in the field of automotive safety has moved toward a more intelligent Advanced Driving Assistance System (ADAS). ADAS can perceive the surrounding environment and alert the driver or control the vehicle automatically when danger is present, thus reducing or avoiding the hazards of traffic accidents. It is a new generation of active safety systems [1]. Since most accidents originate from collisions between vehicles and obstacles, obstacle avoidance is the primary task of ADAS, and accurate detection of obstacles is the basis for ADAS to take obstacle avoidance measures.

Unexpected hazards on the road (e.g., lost goods, loose stones, etc.) are an easily overlooked problem for traffic safety, and they can easily cause traffic accidents, which have a great impact on people's property and even their lives. US traffic reports show that approximately 150 people die each year in the US due to lost goods on the road [2]. As such, obstacles usually have small sizes and complex shapes, and achieving accurate and effective detection is a hard task for both drivers and general detection systems. Therefore, it is greatly significant to include the detection of unexpected obstacles in vehicle environment perception systems and develop an effective detection method.

There are currently two solutions in the field of automotive environment perception, namely computer vision solution and radar solution. At the current research stage, both solutions have certain advantages and disadvantages, but the computer vision solution is closer to the human driving form and has great potential for development. Vision is the most important way for humans to obtain information, and studies have shown that about 90% of the information obtained by drivers comes from vision [3]. Computer vision takes a camera as a sensor, and the images acquired by the camera are most similar to the real world perceived by the human eyes, which contains a wealth of information that can be used for a variety of tasks in the field of environment perception. In addition, computer vision solutions are low-cost and more easily marketable.

Stereo vision and deep learning are the two main detection methods used in vision solutions. Stereo vision methods [4,5] mainly rely on the stereo information contained in stereo image pairs captured by binocular cameras and use the differences in the geometric structure of objects to recognize the image content. However, stereo vision techniques are strongly influenced by distance and are not effective in detecting obstacles at long distances. Deep learning methods can handle complex feature information thanks to the powerful feature extraction capabilities of convolutional neural networks. In addition, they have certain advantages over stereo vision methods in terms of detection distance and accuracy, and semantic information can be obtained. Consequently, this paper studies an unexpected obstacle detection method based on deep learning, and it mainly has the following three contributions:

1. A semantic segmentation method suitable for detecting unexpected obstacles on the road is proposed. Traditional supervised machine learning systems can only identify the classes of obstacles present in the training set but are powerless for unexpected obstacles. A powerful advantage of DeepLabV3+ is its ability to use the learned contextual features of the images to generalize information far beyond their training data. This paper highlights the contextual features of unexpected obstacles on the road by reasonably defining the pixel classes of the training dataset and uses DeepLabV3+ to perform semantic segmentation of images to achieve effective detection of unexpected obstacles.
2. An open-set recognition algorithm for unexpected road obstacles is designed. Depending on the difference in uncertainty between the known and unknown classes, it segments the unexpected obstacles by adaptive threshold. The algorithm is very different from semantic segmentation methods in terms of the principle and the image information used, so it can complement semantic segmentation methods very well.
3. A probabilistic fusion detection method based on the Bayesian model is proposed. The Bayesian model is a classical probabilistic prediction model, which makes the prediction results more consistent with the human brain's judgment and decision-making process by using multivariate data for joint inference. We incorporate the detection results of the two aforementioned methods into the Bayesian model in a probabilistic form to improve the system's ability to detect unexpected obstacles.

The rest of the paper is organized as follows. Related work is introduced in Section 2. Section 3 presents the core algorithms, including unexpected obstacle detection based on semantic segmentation, unexpected obstacle detection based on uncertainty, and probabilistic fusion. Section 4 is the experiment and result analysis. The conclusions are made in Section 5.

2. Related Work

The early machine learning methods divided the obstacle detection task into two steps: feature extraction and feature classification. The artificially designed features were input into some shallow classifiers for detection and recognition, and the detection accuracy and classification ability are limited. In recent years, deep learning techniques using deep neural networks as tools have greatly advanced the development of artificial intelligence. Among them, deep convolutional neural networks (CNNs) [6–8] have a powerful feature

extraction capability, which can learn the features of the object better and achieve higher accuracy detection. Therefore, CNNs are widely used in obstacle detection tasks [9–11]. For example, Qi et al. [9] first targeted obstacle regions by the maximum difference and morphological Region Of Interest (ROI) extraction method and then input the resulting ROI into CNN for obstacle recognition, thus improving the accuracy of obstacle detection and recognition. Jia et al. [10] added the global information of a single image to the classifier while using CNN combined with Deep Belief Network (DBN) to build an obstacle detection model, which experimentally proved to have good detection capability. Levi et al. [11] proposed an obstacle detection and road segmentation method called StixelNet, which simplified the detection task to a stixel regression problem and introduced a loss function based on a semi-discrete representation of the obstacle position probability to train the network, and achieved good results on the KITTI dataset [12]. However, these traditional CNN-based obstacle detection methods are less flexible and cannot provide dense and accurate obstacle locations.

An effective way to use CNNs for obstacle detection is semantic segmentation. Semantic segmentation is the classification of each pixel in an image to achieve region segmentation, which is suitable for application in pixel-level scene representation and obstacle detection. In 2015, Long et al. proposed the Fully Convolutional Network (FCN) [13], replacing fully connected layers in the traditional CNN with convolutional layers. FCN is widely used in semantic segmentation because the convolutional layer can retain a higher image resolution, which is beneficial for pixel-level task such as semantic segmentation. The evaluation results of the PASCAL VOC dataset [14] showed that FCN-based semantic segmentation methods [15,16] take the lead. Badrinarayanan et al. proposed SegNet [17] based on FCN, whose encoder part used the first 13 layers of the convolutional network of VGG-16 [18], and each encoder layer corresponded to a decoder layer. The final output of the decoder was fed to a softmax classifier to generate a class probability for each pixel independently. To address the problem of a single scale of FCN feature extraction, the Unet [19] proposed by Ronneberger et al. adopted the feature channel concatenation method, thus retaining more image information. In addition, its unique U-shaped network structure enabled it to have better access to the semantic information of small objects. The DeeplabV3+ model [20] proposed by Chen et al. in 2018 incorporated an Encoder-Decoder structure and combined it with Atrous Spatial Pyramid Pooling (ASPP), which enlarged the perceptual field of the model by introducing atrous convolution and further improved the segmentation capability of the model for objects of different sizes. In recent years, some researchers have applied semantic segmentation models to the field of obstacle detection and achieved good results. For example, Shin et al. [21] proposed a deep residual network EAR-Net for road scene segmentation on the basis of ASPP, which improved the accuracy of segmentation by utilizing depthwise separable convolution and interpolation for feature extraction and recovery. Valdez-Rodríguez et al. [22] combined depth estimation with a semantic segmentation model and proposed a hybrid CNN network. This model can separate the foreground and background in images and achieve good segmentation results on the SYNTHIA-AL dataset [23].

The above-mentioned semantic segmentation methods are only used to detect the types of obstacles that already exist in the training set, while unexpected obstacles (which do not exist in the training set) are not detected. Creusot et al. [24] first explored this particular problem, and the paper used a restricted Boltzmann machine neural network to detect anomalous regions on the road. There are currently some methods for the detection of unexpected classes, such as open-set recognition, out-of-distribution detection, anomaly detection, and novelty detection, etc. However, the difference between the latter three methods and open-set recognition is that they cannot distinguish known classes in the training set [25]. The methods to solve the open-set recognition problem are referred to as open-set recognition algorithms [26]. Scheirer et al. [27] conducted pioneering research on the open-set recognition problem and defined it as a constrained minimization problem. The article proposes a novel method to deal with the task of open-set recognition, which

sculpts a decision space from the marginal distances of a 1-class or binary SVM with a linear kernel. Bendale et al. [28] investigated the use of deep neural networks for open-set recognition to provide deep-learning-based recognition systems with the ability to reject unexpected class. However, the above two approaches have not carried out targeted research in the area of unexpected obstacle detection.

As a matter of fact, using a single method for obstacle detection in practical problems is very limited. If different detection methods can be integrated, the purpose of complementing each other's advantages can be achieved, thereby improving the detection quality. In recent years, many researchers have adopted this idea. For example, Schneider et al. [29] defined the model as an energy minimization problem and used semantics and geometry as unary data terms in it. This model can jointly infer semantic and geometric clues to achieve better detection results. Zhang et al. [30] proposed an obstacle detection method using fusion of radar and visual data, and its data fusion methods are divided into two categories: spatial fusion and time fusion. The former can realize the unification between coordinate systems through a series of coordinate transformations, and the latter can ensure the synchronization of data fusion in time. It can be seen that the fusion method can consider a variety of data information, which effectively improves the detection quality. For fusion strategies, Bayesian is a probability-based fusion framework, which is very classic and widely used. Therefore, this paper adopts Bayesian as the fusion framework.

In fact, semantic segmentation models have inherently powerful generalization capabilities and are capable of acquiring much more information than the types of data provided in the training dataset. For example, semantic segmentation models can learn the contextual information of an image from a large amount of training data [17,19]. Pixels in an image are usually not isolated, and contextual information reflects some connections between a pixel and its neighboring pixels. The information generalization capability of semantic segmentation models provides powerful technical support for the detection of unexpected obstacles. In this paper, a semantic segmentation method that can highlight the contextual information of unexpected obstacles is designed for such a complex problem of unexpected obstacle detection. In addition, as the semantic method only utilizes the semantic information of the image, the detection accuracy is limited, so this paper also designs an open-set recognition algorithm based on the uncertainty. Unexpected obstacles are detected by incorporating the two methods into the Bayesian framework for joint inference.

3. Proposed Methods

Figure 1 shows the flow chart of our method. First, the input image is fed into Unexpected Obstacles Detection Based on Semantic Segmentation (UOD-SS) and general semantic segmentation. The Unexpected Obstacles Detection Based on Uncertainty (UOD-Un) based on uncertainty degree uses the results of the above two methods. Then, the detection probabilities provided by the above two methods are fused through a Bayesian framework (Probabilistic Fusion Based on UOD-SS and UOD-Un, PF-SSUn). Finally, the output image is the superposition of general semantic segmentation and probabilistic fusion results.

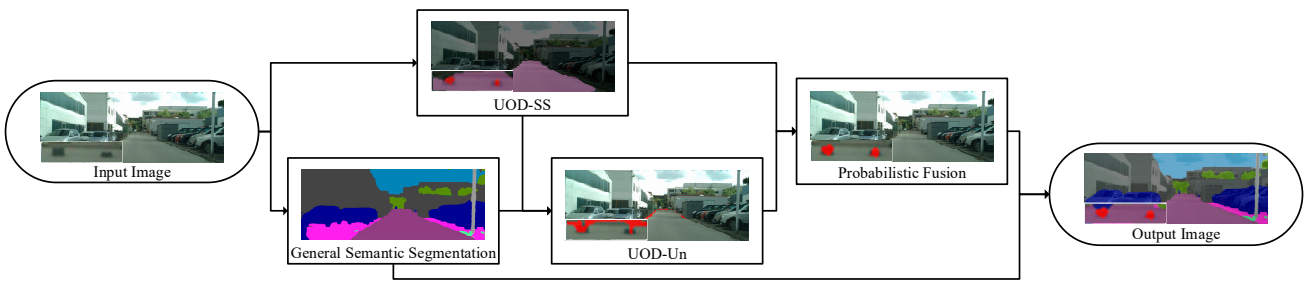
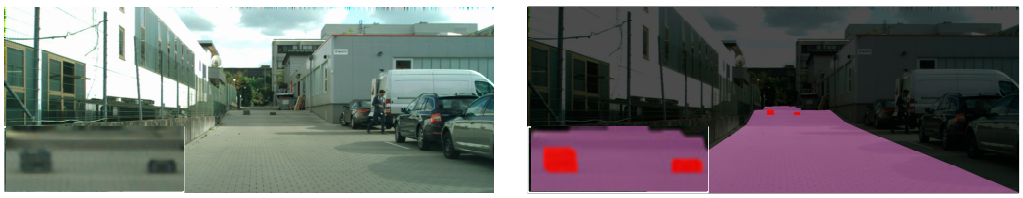


Figure 1. Flowchart of the method.

3.1. Unexpected Obstacle Detection Based on Semantic Segmentation (UOD-SS)

In this paper, by reasonably defining the classes of the training dataset, the contextual features of unexpected obstacles on the road are highlighted. The powerful information generalization ability of the semantic segmentation model is used to realize the detection of unexpected obstacles; at the same time, the free space where the vehicle can travel is output. To highlight the contextual features of unexpected obstacles, traffic scenes are divided into three classes: free space, unexpected obstacles, and background, where the background class is defined as any image region except free space and unexpected obstacles on the road. Figure 2 shows this classification method. There are two dropped cargoes in Figure 2a, shown in the image as unexpected obstacles on the road. Figure 2b is the scene image after segmentation according to the above classification method. The purple part represents free space, the red part represents unexpected obstacles, and the uncolored part represents the background. It can be seen from Figure 2b that these two unexpected obstacles are particularly prominent in the free space. Due to the small number of scene classes, the network ignores the shape and other features of the obstacles and focuses on the common contextual properties of unexpected obstacles on the road. Therefore, this kind of semantic information can be learned by virtue of the semantic segmentation model to segment unexpected obstacles at different distances and appearances.



(**a**) (**b**)

Figure 2. Classification of road scenes: (**a**) original image; (**b**) scene classification.

The advanced DeepLabV3+ is used as the semantic segmentation model, and ResNet-50 is adopted as the backbone feature extraction network. The model combines the encoder-decoder structure and ASPP, which can well restore the edge information of the image and learn multi-scale features. The architecture of DeepLabV3+ for the UOD-SS method is shown in Figure 3. To solve the problem of semantic segmentation of road scenes, the model introduces atrous convolution. Compared with standard convolution, atrous convolution can increase the receptive field of the model and retain more spatial information about small objects. For unexpected obstacles of small size, atrous convolution can effectively reduce the interference of large background regions on feature extraction. At the same time, it can make the feature map in the model retain the boundary information of the object as much as possible without adding too many calculation parameters. For a two-dimensional signal such as an image, the mapping formula for the input and output of the atrous convolution is as follows:

$$y[i] = \sum_{k} x[i + d \cdot k]\omega[k]. \tag{1}$$

Among them, i is the index value of a single element of the feature map; d is the dilation rate of the atrous convolution; $\omega[k]$ is the convolution kernel of size k.

Aiming at unexpected obstacles on the road, the original atrous convolution with the dilation rate of 6, 12, and 18 in DeepLabV3+ is changed to 4, 8, and 12, respectively. As the resolution of the feature maps continues to decrease, the smaller dilation rate can efficiently extract low-resolution feature maps. At the same time, the ASPP structure adopted by the model integrates multi-scale feature information and uses a convolution kernel with a larger dilation rate to segment large objects and smaller one to segment small objects, thereby enhancing the model's ability to segment objects of different sizes.

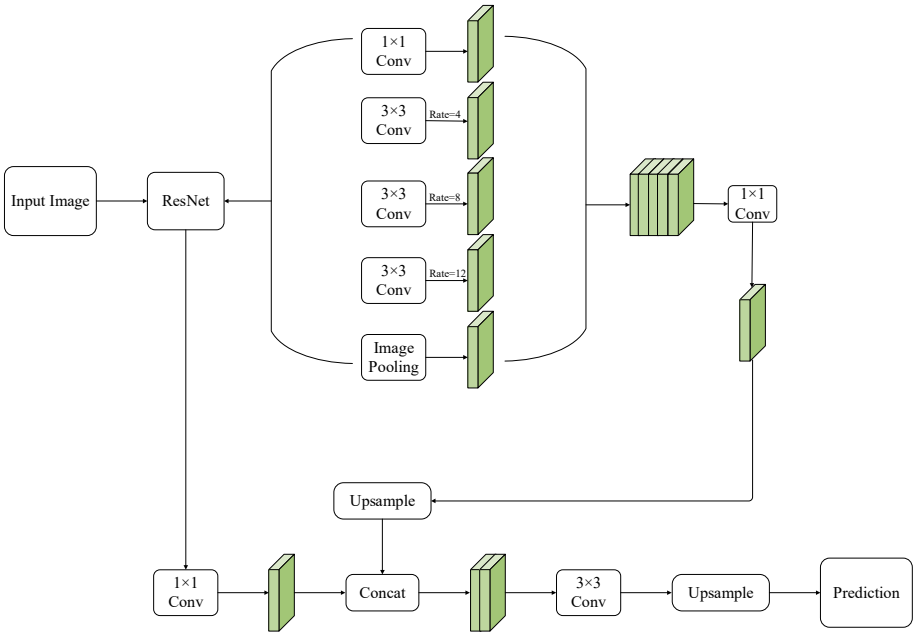
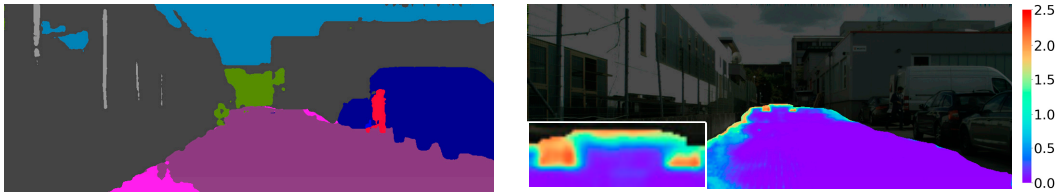


Figure 3. The architecture of DeepLabV3+ for UOD-SS method.

Finally, this semantic segmentation network is trained using the Lost and Found dataset [5], which contains road scenes with high complexity and various unexpected obstacles.

3.2. Unexpected Obstacle Detection Based on Uncertainty (UOD-Un)

In order to further improve the detection effect of unexpected obstacles, an open-set recognition algorithm UOD-Un based on uncertainty is designed in this paper. We compute the uncertainty degree from the known class probabilities obtained from general semantic segmentation. The general semantic segmentation is different from the three-class semantic segmentation in the UOD-SS method in Section 3.1; its training set contains common objects in the traffic scene. In general, semantic segmentation, the predicted probabilities of unexpected obstacle (unknown class) are usually more dispersed among known classes, with higher uncertainty relative to known classes. Figure 4 shows the uncertainty degree distribution of general semantic segmentation. Figure 4a shows the result of general semantic segmentation for the scene in Figure 2a. Figure 4b shows the uncertainty degree distribution map of Figure 4a, and the ROI in Figure 4b is the free space class and unexpected obstacle class region in Figure 2b. From Figure 4b, it can be seen that the area of the unexpected obstacle class is red, indicating that this area has a higher uncertainty degree. Therefore, the uncertainty degree can be used to distinguish known and unknown classes so that unexpected obstacles can be detected.



(**a**) (**b**)

Figure 4. Uncertainty degree distribution of general semantic segmentation: (**a**) general semantic segmentation; (**b**) uncertainty degree distribution.

The flow of UOD-Un is shown in Figure 5. Firstly, the general semantic segmentation is performed on the original image, and then the ROI is determined according to the detection

result of UOD-SS. Finally, the uncertainty degree of the pixels in this region is calculated, and the threshold of the uncertainty degree is used to segment unexpected obstacles. When performing general semantic segmentation, we also choose the DeepLabV3+ model and use the Cityscapes dataset [31] as the training set, which can complement the Lost and Found dataset to a certain extent.

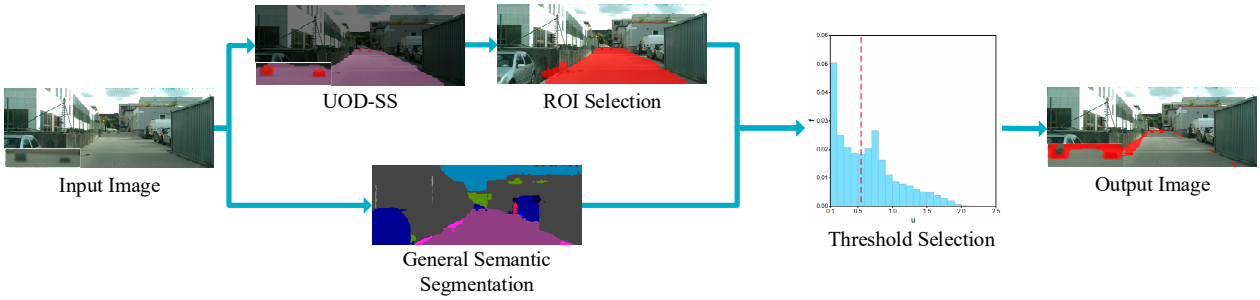


Figure 5. The flow chart of UOD-Un.

Entropy can describe the uncertainty of each possible event of the information source, so it is used as the measurement criterion of pixel uncertainty in this paper. The uncertainty degree of the pixel i in the input image is defined as follows:

$$u_i = -\sum_{c=1}^{N} p_{i,c} \log p_{i,c}. \quad (2)$$

Among them, N is the total number of classes; c is the class index; $p_{i,c}$ represents the prediction probability of the class c of the pixel i in the image, which can be obtained by the Softmax layer of the semantic segmentation model.

There is also a high uncertainty at the class boundary of the image, and this method may mark these pixels as an unknown class, resulting in a high false-positive rate of the segmentation results. Therefore, it is necessary to filter out these points that interfere with the detection results as much as possible. Given that unexpected obstacles are generally scattered in free space, the free space class and unexpected obstacle class regions detected by UOD-SS are first selected, and then the region is expanded by 3 pixels in each of the length and width dimensions of the image. The expanded region is used as the ROI of UOD-Un, so that the selected ROI can avoid the interference of the background, thereby minimizing the occurrence of false positives.

According to the previous analysis, unexpected obstacles can be segmented by the uncertainty degree threshold, so it is very important to choose this threshold reasonably. An adaptive threshold selection method is proposed in this paper. Figure 6 shows the uncertainty degree distribution histogram of an image. The horizontal axis is the uncertainty degree (u) represented by entropy, and the vertical axis is the frequency (f) of pixels with a certain uncertainty degree. The figure reflects the general distribution of uncertainty degree. Considering that the number of pixels belonging to the known classes is large and mainly distributed on the left, while the number of pixels belonging to the unknown class is small and mainly distributed on the right, if the known classes and the unknown class are regarded as two different sets, then the point with the loosest coupling between the two sets should be selected as the threshold point for segmentation. According to this threshold selection principle, first find the leftmost and rightmost frequency peak intervals in the uncertainty degree distribution histogram (to avoid chance, the count of pixels in the peak interval is required to be greater than 100), and then take the midpoint of the minimum interval between these two peaks as the threshold, as shown by the red dotted line in the figure. Finally, all the pixels whose uncertainty degree is higher than the threshold are marked as an unknown class. The selection of the threshold is shown in Algorithm 1.

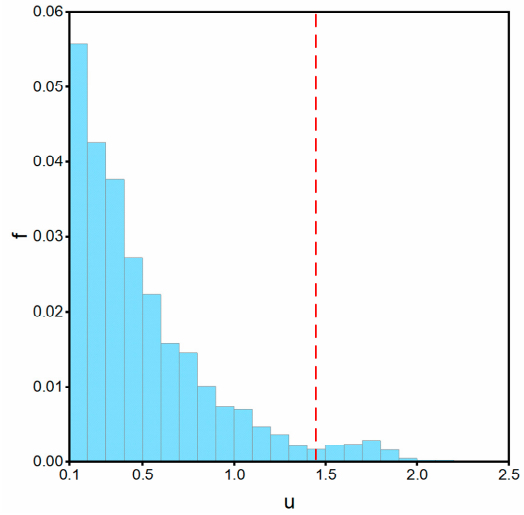


Figure 6. Uncertainty degree distribution histogram (due to the high proportion of known pixels, the leftmost part of the histogram is omitted to achieve a better display effect).

Algorithm 1. The selection of the threshold algorithm.

Input: Un_f: Uncertainty frequency distribution; N: Intervals quantity of uncertainty frequency distribution; $M(\cdot)$: The function to find local maximum of frequency; $C(\cdot)$: The count of pixels in the interval; $min(\cdot)$: The function to find the left endpoint of minimum interval
Output: P_l: Leftmost peak; P_r: Rightmost peak; T: Selected threshold
$i = 1, j = N, l, r$
1. Find the leftmost peak
while $i \leq N$ **do**
  if $Un_f[i] = M(Un_f)$ **and** $C\left(Un_f[i]\right) > 100$
  $P_l = Un_f[i]$
  $l = i$
  end if
$i = i + 1$
end while
2. Find the rightmost peak
while $j > 0$ **do**
  if $Un_f[j] = M(Un_f)$ **and** $C\left(Un_f[j]\right) > 100$
  $P_r = Un_f[j]$
  $r = j$
  end if
$j = j - 1$
end while
3. Determine the threshold
$T = min(Un_f[l : r]) + 0.05$

3.3. Probabilistic Fusion (PF-SSUn)

Psychological research shows that human judgment and decision making is a combination of a variety of information and gives results in a probabilistic way. For intelligent vehicles, obstacle detection is also a judgment and decision-making process. The UOD-SS and UOD-Un methods proposed in this paper detect unexpected obstacles from the perspectives of semantic information and uncertainty information, respectively. The two methods are independent of each other, so the fusion of the obtained results can bring many new advantages. The Bayesian framework is suitable for decision inference based on multi-factor probabilistic methods, which is more in line with the human brain's judgment and decision-making process than the methods of obstacle detection using single informa-

tion. In this paper, it is used to fuse the results obtained by the above two independent methods, and the probability that a pixel is an unexpected obstacle can be expressed as:

$$p(o) = \frac{p_{pr} \cdot p_{SS} \cdot p_{Un}}{(p_{pr} \cdot p_{SS} \cdot p_{Un}) + (1 - p_{pr})(1 - p_{SS})(1 - p_{Un})}. \quad (3)$$

Among them, p_{pr} represents the prior probability; p_{SS} and p_{Un} represent the probability that the pixel is judged as an unexpected obstacle by UOD-SS and UOD-Un, respectively. Finally, the judgment of whether the pixel is an unexpected obstacle depends on the threshold of $p(o)$.

p_{SS} can be obtained directly from the Softmax layer of the network, and the Softmax function can be expressed as:

$$Softmax(z_j) = \frac{e^{z_j}}{\sum_{c=1}^{N} e^{z_c}}. \quad (4)$$

Among them, z_j represents the output of the previous stage unit of the class j; c is the class index; N is the total number of classes. Therefore, p_{SS} can be expressed as:

$$p_{SS} = \frac{e^{z_3}}{e^{z_1} + e^{z_2} + e^{z_3}}, \quad (5)$$

where z_1, z_2, and z_3 represent the output of the previous stage unit of background class, free space class and unexpected obstacles class, respectively.

p_{Un} can be obtained according to the degree of deviation between the uncertainty degree of each pixel and the threshold. Assuming that the probability at the threshold is 50%, it can be expressed as:

$$p_{Un} = \frac{1}{1 + exp(u_t - u_i)}. \quad (6)$$

Among them, u_t represents the value of the uncertainty degree at the threshold; u_i represents the uncertainty degree of pixel i.

4. Experiment and Result Analysis

4.1. Experimental Environment and Parameter Settings

The experiment is based on the deep learning framework Pytorch, and the programming language is Python 3.6. In terms of experimental hardware, the CPU is Intel(R) Core(TM) i7-10750H CPU @ 2.60 GHz, the memory is 16 GB, the GPU is Nvidia Geforce RTX 2060, the video memory is 6 GB, and the CUDA version is 10.2. The prior probability p_{pr} was set to 0.5, and the final decision threshold of $p(o)$ was set to 0.3. During training, the batch size was 8, the initial learning rate of the network was set to 0.0001, and the total iteration was 50. Adam and the cross entropy were used as the network optimization method and the loss function, respectively. The network was trained using a pretrained model.

4.2. Datasets and Preprocessing

The semantic segmentation parts of UOD-SS and UOD-Un were trained using the Lost and Found and the Cityscapes datasets, respectively, and the test set in the Lost and Found dataset was used as the test images. The Lost and Found dataset contains about 2100 images taken from various street scenes with pixel-level semantic annotations. The Cityscapes dataset has 5000 images of urban driving scenes, including many classes, and provides corresponding semantic annotations. Examples of the two datasets are shown in Figure 7. Before the algorithm was implemented, the image was resized from 2048 × 1024 to 1024 × 512, and pre-processing operations such as random cropping, color dithering, and random horizontal flipping were performed on the image.

(**a**) (**b**)

Figure 7. Image examples of datasets: (**a**) the Lost and Found dataset; (**b**) the Cityscapes dataset.

4.3. Evaluation Criteria

For the detection task, the detection rate and the false-positive rate are two important indicators of the detection effect, which reflect the detection method's ability to cover and judge the objects, respectively. In this paper, pixel-level criteria were used to evaluate the detection performance of the algorithm for unexpected obstacles. Pixel-wise Detection Rate (PDR) can be expressed as:

$$PDR = \frac{CDP_{uo}}{AP_{uo}}. \tag{7}$$

Among them, CDP_{uo} is the number of pixels that are correctly detected as unexpected obstacles; AP_{uo} is the number of pixels of real unexpected obstacles. Pixel-wise False-Positive Rate (PFPR) can be expressed as:

$$PFPR = \frac{IDP_{uo}}{AP_{nuo}}. \tag{8}$$

Among them, IDP_{uo} represents the number of pixels that are incorrectly detected as unexpected obstacles; AP_{nuo} is the number of pixels of real non-unexpected obstacles.

4.4. Analysis of Results

To evaluate the detection effect of our fusion method (PF-SSUn) for unexpected obstacles, PF-SSUn was compared with the following three methods: (1) Stixels [4]: It is a classic obstacle detection method based on stereo vision, which can describe the objects with multi-layer stixels; (2) UOD-SS (Unet): The semantic segmentation model DeepLabV3+ of UOD-SS was replaced by Unet to compare the ability of these two semantic segmentation models to detect unexpected obstacles; (3) UOD-SS.

Table 1 lists the performance indicators of PF-SSUn and the above three methods. As can be seen from the table, the Stixels method has the lowest detection rate and the highest false-positive rate, and the detection effect is far worse than the other three methods. This shows that the Stixels method based on stereo vision has obvious disadvantages in the detection of unexpected obstacles compared to the semantic segmentation method. This is mainly due to the fact that stereo vision methods are much more affected by the sizes and shapes of obstacles than semantic segmentation methods. UOD-SS (Unet) is worse than UOD-SS in both detection rate and false-positive rate indicators, and it is difficult to meet traffic safety requirements. UOD-SS has better performance indicators, which is mainly due to the ASPP structure of DeepLabV3+ and the introduction of atrous convolution to improve its detection ability. Compared with UOD-SS, PF-SSUn significantly improves the detection rate of unexpected obstacles. Although the false-positive rate also increased to a certain extent, considering the improvement in detection rate, the cost is acceptable.

Table 1. Comparison of performance indicators of the four detection methods.

Method	PDR (%)	PFPR (%)
Stixels	52.41	4.11
UOD-SS(Unet)	63.23	0.45
UOD-SS	83.58	0.30
PF-SSUn	92.33	0.37

From a safety point of view, to a certain extent, we pay more attention to the improvement of the detection rate. Therefore, two scenes in the Lost and Found dataset are selected to test the frame-by-frame detection rate of the above methods, and the results are shown in Figure 8 (the distance of obstacles is from far to near). Since the Stixels method has significantly the lowest detection performance, only three other semantically relevant methods are shown here. It can be seen that the detection rates of the three methods are relatively high at close range, and the performance is stable. However, the detection capability of PF-SSUn has been significantly improved at long distances. Especially in scene 2, which is more difficult to detect, UOD-SS (Unet) can hardly detect long-distances unexpected obstacles, and the detection rate of UOD-SS is also significantly lower, but PF-SSUn has always maintained a stable and relatively high detection rate. This is mainly because it is very difficult to detect unexpected obstacles at a long distance. PF-SSUn makes up for the detection omission of UOD-SS and improves the detection rate by probability fusion of UOD-SS and UOD-Un.

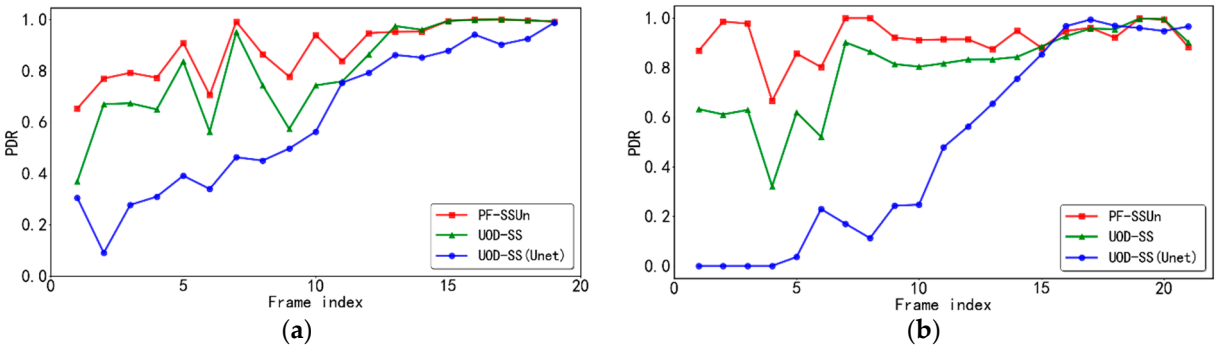


Figure 8. Comparison of frame-by-frame detection rate of the three semantically relevant methods: (**a**) Scene 1; (**b**) Scene 2.

Figure 9 shows the PDR over PFPR scatter plots for the three methods in the two scenes. A high detection rate and low false-positive rate (upper left corner of the figure) can reflect the best detection performance. As shown in the figure, the marked points of PF-SSUn are closer to the upper part of the figure than the other two methods, and only a few points are located on the right side of the figure due to the increase in the false-positive rate, but their values are always very small (less than 1%). Due to the low detection rate of UOD-SS (Unet), many points appear at the bottom of the figure. The marked points of UOD-SS are closer to the upper part of the figure than UOD-SS (Unet), but there are still many points scattered in the middle of the figure.

To evaluate the detection effect of different fusion strategies, PF-SSUn was compared with the following four methods: (1) UOD-SS; (2) UOD-Un; (3) Fusion-And: Both methods detect a pixel as an unexpected obstacle, then it is determined that the pixel is an unexpected obstacle; (4) Fusion-Or: At least one method detects the pixel as an unexpected obstacle, and the pixel is determined to be an unexpected obstacle.

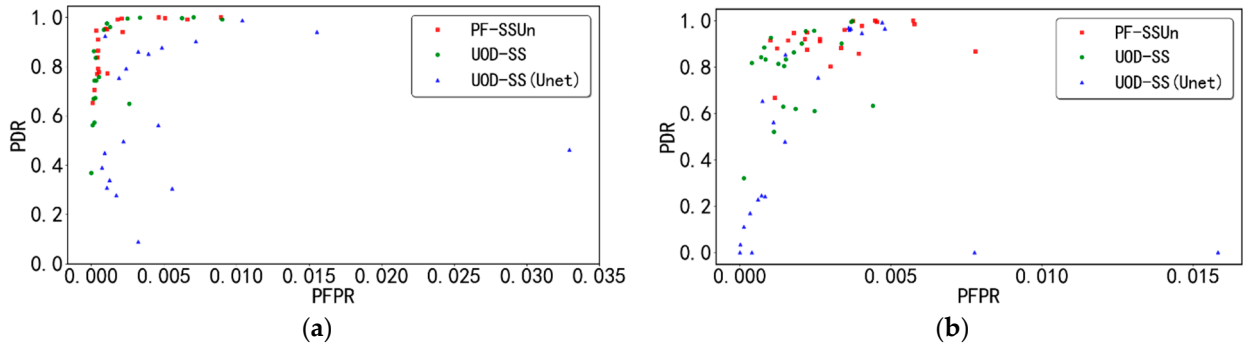


Figure 9. PDR over PFPR scatter plot of the three semantically relevant methods: (**a**) Scene 1; (**b**) Scene 2.

Table 2 lists the performance indicators of PF-SSUn and the above four methods. It can be seen from the table that the detection rate of both independent methods is lower than that of PF-SSUn. When used as an independent method to detect unexpected obstacles, UOD-Un performs unsatisfactorily in detection rate and false-positive rate. However, because UOD-Un uses a completely different detection method from UOD-SS, and the uncertainty information it uses is quite different from semantic information, it plays a good supplementary role to the latter. In terms of fusion strategy, Fusion-And focuses on reducing the false-positive rate, so it achieves the lowest false-positive rate, but the detection rate indicator is very poor. Fusion-Or achieves the highest detection rate, but it also has the highest false-positive rate. The detection rate of PF-SSUn also achieves very good results (only slightly lower than Fusion-Or), but the false-positive rate is only one-ninth of Fusion-Or. This shows that our fusion strategy maintains a relatively low false-positive rate while focusing on improving the detection rate. Figure 10 shows the scatter plots of PDR over PFPR for different fusion strategies in two scenes. As can be seen from the figure, the points of PF-SSUn are concentrated in the upper left area, while the points of Fusion-And and Fusion-Or produce downward and right deviations, respectively.

Table 2. Comparison of performance indicators between independent method before fusion and different fusion strategies.

Method	PDR (%)	PFPR (%)
UOD-SS	83.58	0.30
UOD-Un	62.43	3.17
Fusion-And	52.69	0.08
Fusion-Or	93.32	3.39
PF-SSUn	92.33	0.37

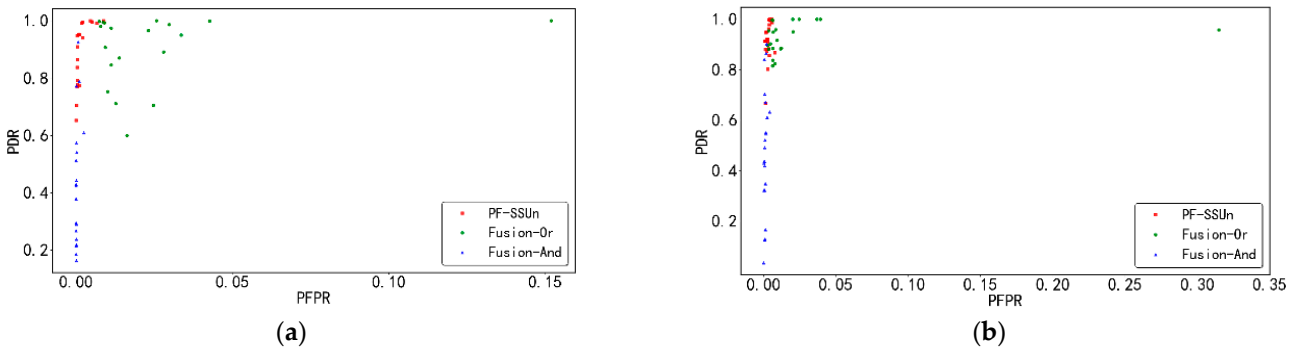


Figure 10. PDR over PFPR scatter plot of the three fusion strategies: (**a**) Scene 1; (**b**) Scene 2.

Figure 11 shows the comparison between the detection results of the three methods proposed in this paper with the ground truth on three images in the test set. These three images are selected because they have great differences in scene, lighting, and appearance of unexpected obstacles, and the distance is very long. The detection results of the last row are obtained by the general semantic segmentation in Section 3.2 and the superposition of PF-SSUn results. The unexpected obstacle in the image on the left column is a toy car. It can be seen from the enlarged image that UOD-SS has detected a part of the toy car, while UOD-Un performs better. The fusion method (PF-SSUn) embodies the advantages of UOD-Un in this unexpected obstacle detection. The middle column image contains three unexpected obstacles. UOD-SS misses many areas due to the long distance of obstacles. Although there are more false-positive pixels in the detection results of UOD-Un, it detects most of the unexpected obstacle areas. The detection results after fusion cover most areas of unexpected obstacles, and the false-positive pixels are significantly reduced compared to UOD-Un. Although there are still many false-positive pixels in the image, considering that the unexpected obstacles in this column of images are far away and small in size, and the detection task is very difficult, it can still be said that PF-SSUn achieves the desired detection effect. The image in the right column contains two unexpected obstacles. Both UOD-SS and UOD-Un have many missed pixels, and PF-SSUn fills these missed areas through Bayesian probability fusion.

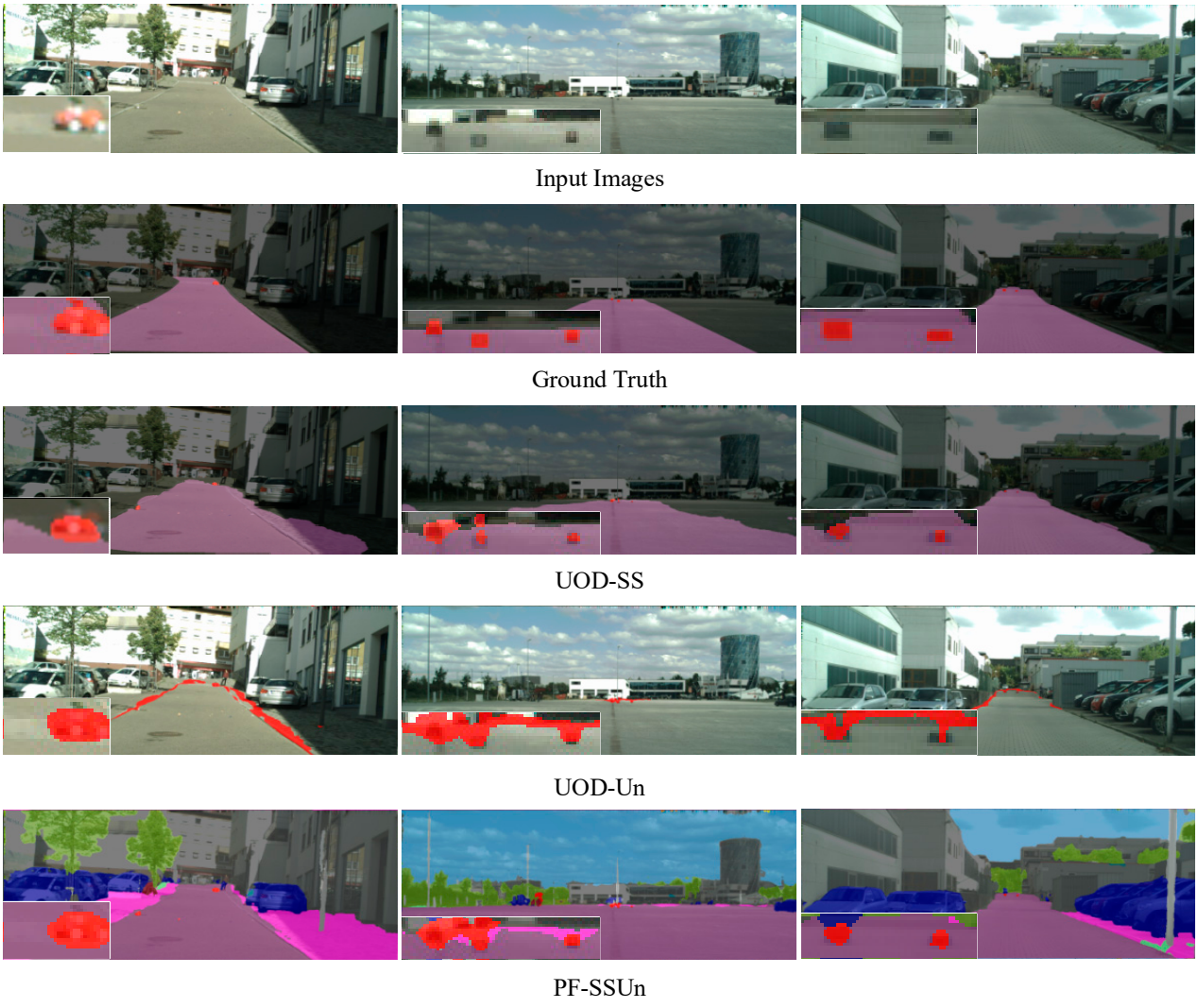


Figure 11. Display of test results.

5. Conclusions

In order to improve the environment perception ability of ADAS, this paper proposes an unexpected obstacle detection method based on Bayesian probabilistic fusion. Firstly, an unexpected obstacle detection method based on semantic segmentation (UOD-SS) is proposed. This method highlights the contextual information of unexpected obstacles by reasonably dividing the scene classes so that the semantic segmentation model can more easily identify unexpected obstacles. Furthermore, we propose an uncertainty-based unexpected obstacle detection method (UOD-Un), which uses an adaptive threshold to segment known and unknown classes. Finally, the detection probabilities provided by the above two methods are fused through the Bayesian framework, and the obtained result reflects the joint decision making based on various information. Experiments implemented on the public dataset Lost and Found show that the pixel-level detection rate and false-positive rate of our method are 92.33 and 0.37%, respectively. Even in the detection of long-distance and small-size unexpected obstacles, our method has achieved good detection results. It can be seen from the experimental results that there are many false-positive pixels in the detection results of UOD-Un. Although the fusion method can remove some of them, the remaining pixels still have a certain effect on the final false-positive rate. Therefore, our future work is to study how to reduce the false-positive rate of UOD-Un. In addition, since the core of this paper is two unexpected obstacle detection methods and their fusion, the existing DeepLabV3+ network is selected for semantic segmentation. If a more powerful semantic segmentation network can be designed, it may improve the detection effect. This is also our future work.

Author Contributions: Conceptualization, W.C., R.L., and S.L.; Funding acquisition, W.C.; Methodology, W.C., R.L. and S.L.; Project administration, W.C.; Software, T.X. and R.L.; Validation, T.X. and R.L.; Writing—original draft, W.C. and T.X.; Writing—review and editing, W.C. and S.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research has been supported by the Public Welfare Technology Application Research Program of Zhejiang Province, China (grant No. LGG22F030016); by Innovation Team by Department of Education of Guangdong Province, P.R. China (grant No. 2020KCXTD041).

Acknowledgments: The authors would like to thank Editor-in-Chief, Editor, and anonymous Reviewers for their valuable reviews.

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

References

1. He, B.; Wen, Y.; Liu, S. Development of Advanced Driving Assistance System (Level 2). *Autom. Dig.* **2020**, *2020*, 48–53.
2. Ramos, S.; Gehrig, S.; Pinggera, P.; Franke, U.; Rother, C. Detecting unexpected obstacles for self-driving cars: Fusing deep learning and geometric modeling. In Proceedings of the 2017 IEEE Intelligent Vehicles Symposium (IV), Redondo Beach, CA, USA, 11–14 June 2017; pp. 1025–1032.
3. Li, P.; He, C.; Li, Z. Comparsion and Analysis of Eye Orientation Method in Driver Fatigue Detection. *For. Eng.* **2008**, *24*, 35–38.
4. Pfeiffer, D.; Franke, U. Towards a Global Optimal Multi-Layer Stixel Representation of Dense 3D Data. In Proceedings of the 22nd British Machine Vision Conference(BMVC), Dundee, UK, 29 August-2 September 2011; pp. 1–12.
5. Pinggera, P.; Ramos, S.; Gehrig, S.; Franke, U.; Rother, C.; Mester, R. Lost and found: Detecting small road hazards for self-driving vehicles. In Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Daejeon, Korea, 9–14 October 2016; pp. 1099–1106.
6. LeCun, Y.; Boser, B.; Denker, J.S.; Henderson, D.; Howard, R.E.; Hubbard, W.; Jackel, L.D. Backpropagation Applied to Handwritten Zip Code Recognition. *Neural Comput.* **1989**, *1*, 541–551. [CrossRef]
7. Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. Going Deeper with Convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition(CVPR), Boston, MA, USA, 7–12 June 2015; pp. 1–9.
8. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
9. Qi, G.; Wang, H.; Haner, M.; Weng, C.; Chen, S.; Zhu, Z. Convolutional neural network based detection and judgement of environmental obstacle in vehicle operation. *CAAI Trans. Intel. Tech.* **2019**, *4*, 80–91. [CrossRef]

10. Jia, B.; Feng, W.; Zhu, M. Obstacle detection in single images with deep neural networks. *Signal Image Video Process* **2016**, *10*, 1033–1040. [CrossRef]
11. Levi, D.; Garnett, N.; Fetaya, E.; Herzlyia, I. StixelNet: A Deep Convolutional Network for Obstacle Detection and Road Segmentation. In Proceedings of the 26th British Machine Vision Conference (BMVC), Swansea, UK, 1 January 2015; pp. 109.1–109.12.
12. 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 (CVPR), Providence, RI, USA, 16–21 June 2012; pp. 3354–3361.
13. Long, J.; Shelhamer, E.; Darrell, T. Fully Convolutional Networks for Semantic Segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition(CVPR), Boston, MA, USA, 7–12 June 2015; pp. 3431–3440.
14. Everingham, M.; Eslami, S.; Van Gool, L.; Williams, C.K.; Winn, J.; Zisserman, A. The Pascal Visual Object Classes Challenge: A Retrospective. *Int. J. Comput. Vis.* **2015**, *111*, 98–136. [CrossRef]
15. Monteiro, M.; Figueiredo, M.A.; Oliveira, A.L. Conditional Random Fields as Recurrent Neural Networks for 3D Medical Imaging Segmentation. *arXiv* **2018**, arXiv:1807.07464.
16. Sharma, A.; Tuzel, O.; Jacobs, D.W. Deep Hierarchical Parsing for Semantic Segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition(CVPR), Boston, MA, USA, 7–12 June 2015; pp. 530–538.
17. Badrinarayanan, V.; Kendall, A.; Cipolla, R. SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 2481–2495. [CrossRef] [PubMed]
18. Simonyan, K.; Zisserman, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. *arXiv* **2014**, arXiv:1409.1556.
19. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention(MICCAI), Munich, Germany, 5–9 October 2015; pp. 234–241.
20. Chen, L.-C.; Zhu, Y.; Papandreou, G.; Schroff, F.; Adam, H. Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 801–818.
21. Shin, S.; Lee, S.; Han, H. EAR-Net: Efficient Atrous Residual Network for Semantic Segmentation of Street Scenes Based on Deep Learning. *Appl. Sci.* **2021**, *11*, 9119. [CrossRef]
22. Valdez-Rodríguez, J.E.; Calvo, H.; Felipe-Riverón, E.; Moreno-Armendáriz, M.A. Improving Depth Estimation by Embedding Semantic Segmentation: A Hybrid CNN Model. *Sensors* **2022**, *22*, 1669. [CrossRef] [PubMed]
23. Bengar, J.Z.; Gonzalez-Garcia, A.; Villalonga, G.; Raducanu, B.; Aghdam, H.H.; Mozerov, M.; Lopez, A.M.; van de Weijer, J. Temporal Coherence for Active Learning in Videos. In Proceedings of the IEEE/CVF International Conference on Computer Vision Workshop (ICCVW), Seoul, Korea, 27–28 October 2019; pp. 914–923.
24. Creusot, C.; Munawar, A. Real-time small obstacle detection on highways using compressive RBM road reconstruction. In Proceedings of the IEEE Intelligent Vehicles Symposium (IV), Seoul, Korea, 27 August 2015; pp. 162–167.
25. Jafarzadeh, M.; Dhamija, A.R.; Cruz, S.; Li, C.; Ahmad, T.; Boult, T.E. A Review of Open-World Learning and Steps Toward Open-World Learning Without Labels. *arXiv e-prints* **2020**, arXiv:2011.12906.
26. Geng, C.; Huang, S.-j.; Chen, S. Recent Advances in Open Set Recognition: A Survey. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**, *43*, 3614–3631. [CrossRef] [PubMed]
27. Scheirer, W.J.; de Rezende Rocha, A.; Sapkota, A.; Boult, T.E. Toward open set recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **2012**, *35*, 1757–1772. [CrossRef] [PubMed]
28. Bendale, A.; Boult, T.E. Towards Open Set Deep Networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 1563–1572.
29. Schneider, L.; Cordts, M.; Rehfeld, T.; Pfeiffer, D.; Enzweiler, M.; Franke, U.; Pollefeys, M.; Roth, S. Semantic stixels: Depth is not enough. In Proceedings of the 2016 IEEE Intelligent Vehicles Symposium (IV), Gothenburg, Sweden, 19–22 June 2016; pp. 110–117.
30. Zhang, X.; Zhou, M.; Qiu, P.; Huang, Y.; Li, J. Radar and vision fusion for the real-time obstacle detection and identification. *Ind. Robot.* **2019**, *46*, 391–395. [CrossRef]
31. 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.