ORFD: A Dataset and Benchmark for Off-Road Freespace Detection

Chen Min[1,2] and Weizhong Jiang[2] and Dawei Zhao[2,*] and Jiaolong Xu[2]
and Liang Xiao[2] and Yiming Nie[2] and Bin Dai[2]

[1]Peking University

[2]NIIDT

minchen@stu.pku.edu.cn

Abstract

Freespace detection is an essential component of autonomous driving technology and plays an important role in trajectory planning. In the last decade, deep learning based freespace detection methods have been proved feasible. However, these efforts were focused on urban road environments and few deep learning based methods were specifically designed for off-road freespace detection due to the lack of off-road dataset and benchmark. In this paper, we present the ORFD dataset, which, to our knowledge, is the first off-road freespace detection dataset. The dataset was collected in different scenes (woodland, farmland, grassland and countryside), different weather conditions (sunny, rainy, foggy and snowy) and different light conditions (bright light, daylight, twilight, darkness), which totally contains 12,198 LiDAR point cloud and RGB image pairs with the traversable area, non-traversable area and unreachable area annotated in detail. We propose a novel network named OFF-Net, which unifies Transformer architecture to aggregate local and global information, to meet the requirement of large receptive fields for freespace detection task. We also propose the cross-attention to dynamically fuse LiDAR and RGB image information for accurate off-road freespace detection. Dataset and code are publicly available at https://github.com/chaytonmin/OFF-Net.

1. Introduction

The ultimate goal of the development of autonomous driving is to liberate the driver, that is, no human attention is needed for any road environment. In order to achieve this goal, it is necessary to study various road environments, not limited in on-road environments, while the existing researches are mostly focused on on-road environments. Freespace detection, one of the most important technology of autonomous driving, has different concepts for structured on-road and unstructured off-road scenes, as

(a) On-road environment

(b) Off-road environment

Figure 1. Difference between the on-road environment (*i.e.*, the urban environment) and the off-road environment. The on-road environment provides many structural cues, such as lane markings and roadway signs while there are no well paved and clearly outlined roadways in off-road environment.

shown in Fig. 1. For the former, freespace mainly refers to regular roads, while for off-road scenes, the concept of freespace is relatively vague. The autonomous vehicles need to pass through grassy, sandy or muddy off-road environments. This is a big challenge for autonomous vehicles since the off-road environments are complex and diverse, for example, tall grass and short grass are very different in terms of traversability, because tall grass may hide invisible obstacles or holes. To our knowledge, there are relatively few researches on the detection of freespace in off-road environments [23].

Data-driven based methods have achieved great success in the last decade and the world has entered an era of deep learning. With the help of data-driven based deep learning methods, many problems of autonomous driving have been solved and autonomous vehicle is becoming a reality. To improve the performance of deep learning methods for autonomous driving, a great many of datasets for scene per-

1

ception of autonomous driving have been published, such as KITTI [8], nuScenes [2], Waymo [21] and so on. However, the existing published autonomous driving datasets are mainly collected in the urban cities, since the mainstream self-driving companies focus on autonomous driving technology in urban environments. Very few datasets are collected in off-road environments [14].

Freespace detection, also known as traversable area detection, is an essential component of the autonomous driving technology and plays an important role in path planning both in on-road and off-road environments. However, the existing off-road datasets [22, 15, 23, 14] do not focus on the traversability analysis in off-road environments. As such, there is a need for a dataset focused on freespace detection task in off-road environments. To address this, we present the **OFF-Road Freespace Detection** (**ORFD**) dataset, which was collected from the off-road scenes for promoting deep learning research in off-road environments. Taking into account the complexity and diversity of off-road environments and the influence of different light factors and seasonal factors on autonomous driving, we collected the ORFD dataset in different seasons (spring, summer, autumn and winter), at different time (day, evening and night) and in different scenes (such as woodland, farmland, grassland and countryside). The dataset includes a total of 12, 198 LiDAR point cloud and RGB image pairs. We synchronized the LiDAR and RGB image data and labeled the freespace in the image plane. Referring to the processing method of the KITTI urban road dataset [7], we projected the LiDAR point cloud onto the corresponding RGB image plane, so as to obtain the depth information of the freespace.

We also introduce an off-road freespace detection benchmark, called OFF-Net. A Transformer architecture is utilized to enlarge the receptive field and capture the context information. Similar to on-road freespace detection task, one modality is not enough for accurate freespace detection. We propose a cross-attention mechanism, that outputs the modality weights to help dynamically fuse LiDAR point cloud and RGB image information. Experiments show that our algorithm achieves good performance on the off-road freespace detection task evaluated on the ORFD dataset.

The highlights of our work are as follows:

- We propose the first off-road freespace detection dataset, called ORFD dataset, which covers different off-road scenes (woodland, farmland, grassland and countryside), different weather conditions (sunny, rainy, foggy and snowy) and different light conditions (bright light, daylight, twilight, darkness) for the generalization of off-road freespace detection.

- We propose OFF-Net, a Transformer network architecture, to aggregate the context information for accurate off-road freespace detection. We also introduce the

cross-attention mechanism to dynamically fuse data from both camera and LiDAR to leverage the strengths of each modality.

- Our work with the dataset and benchmark will open new research area of the off-road freespace detection to ultimately enhance the perception ability of the autonomous vehicle in off-road environments.

2. Related Work

2.1. Datasets

Autonomous driving technology has developed rapidly in recent years with the help of deep learning methods. The success of deep learning relies heavily on the large-scale training data. According to different collection environments, road dataset for autonomous driving can be divided into two classes: on-road datasets and off-road datasets. The on-road datasets, such as KITTI [8], SemanticKITTI [1], nuScenes [2] and Waymo [21] have been widely used for autonomous driving. However, there are few datasets for off-road environments, and now, we will review the existing off-road datasets.

To our knowledge, DeepScene [22] is the first public off-road dataset for multispectral segmentation collected in unstructured forest environments. YCOR [15] is a more diverse and challenging dataset than DeepScene. However, both DeepScene and YCOR have small amount of data. RUGD [23] was collected in an off-road environment for RGB image semantic segmentation task but only has one modality data. RELLIS-3D Dataset [14] improves RUGD with multi-modal sensor to enhance autonomous navigation in off-road environments. The above datasets do not focus on the analysis of traversability in off-road environments, while the detection of traversable area, namely freespace, is one of the most important modules for autonomous driving, especially in off-road environments. In our work, we create the first dataset with accurate and complete ground truths for off-road freespace detection.

2.2. Methods

The existing freespace detection methods are mostly designed for on-road navigation and most of them fuse the LiDAR and camera information together for accurate freespace detection. RBNet [3] detects both road and road boundary in a single process and models them with a Bayesian network. TVFNet [10] takes the LiDAR imageries and the camera-perspective maps as inputs and outputs pixel-wise road detection results in both the LiDAR's imagery view and the camera's perspective view simultaneously. LC-CRF [9] detects road with LiDAR-camera fusion in a conditional random field (CRF) framework to exploit both range and color information. RBANet [20] proposes the reverse attention and boundary attention units for road

Figure 2. ORFD dataset contains a variety of scenes for the generalization of off-road freespace detection.

Figure 3. Different weather conditions are considered in ORFD dataset.

Figure 4. ORFD dataset was collected at different time of the day to cover the light conditions affecting the autonomous navigation.
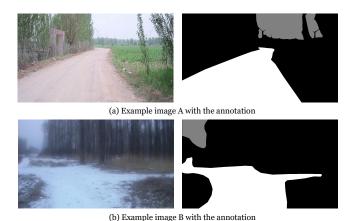

Figure 5. We adopt the concept of 'freespace' to label the ORFD dataset and provide the annotation of three classes, *i.e.*, traversable area (in white color), non-traversable area (in black color) and unreachable area (in gray color).

segmentation. PLARD [4] introduces a progressive LiDAR adaptation-aided road detection approach to adapt LiDAR information into visual image-based road detection and improve detection performance. SNE-RoadSeg [5] infers surface normal information from dense depth images and fuses it with image information to improve the freespace detection performance. The above methods are focused on the on-road freespace detection. In this paper, we introduce a novel Transformer architecture based method for off-road freespace detection and do extensive experiments on our off-road dataset.

3. ORFD Dataset

This section describes the dataset we created for off-road freespace detection called the ORFD dataset. The dataset containing LiDAR point cloud and RGB image information, was collected in a variety of off-road scenes for facilitating deep learning research in off-road environments.

3.1. Data Description

ORFD dataset was collected in off-road environments. Compared with the structured on-road environments, off-road environments vary greatly due to terrain, vegetation, season, weather, time and so on. Therefore, we collected the off-road dataset including a variety of scenes (such as woodland, farmland, grassland and countryside) as shown in Fig. 2, different season and weather conditions (such as sunny, rainy, foggy and snowy weather from spring to winter) as shown in Fig. 3, and different light conditions (bright light, daylight, twilight, darkness) as shown in Fig. 4. The statistics are shown in the Table 1. We collected 30 sequences in various off-road environments in China, and one sequence covers a distance of about 100 meters. We annotated a total of 12, 198 LiDAR point cloud and RGB image pairs. The LiDAR is 40-line, and the size of RGB image is 1280×720.
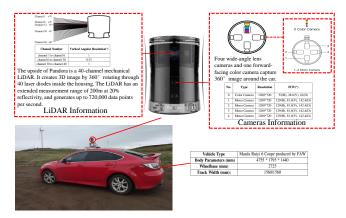
Figure 6. Detail information of vehicle and sensor to collect LiDAR and camera data.

3.2. Data Annotations

We provide pixel-wise image annotations of ORFD dataset for off-road freespace detection. There are three classes: traversable area, non-traversable area and unreachable area as shown in Fig. 5. As mentioned earlier, in off-road environments we use the concept of freespace instead of the road, because the road in the usual on-road environment is clear shown in Fig. 1 (a), while there may not be such a regular road in off-road environments, as shown in Fig. 1 (b), where autonomous vehicles can indeed move in. Therefore, we believe that in the off-road scenes, areas that do not pose a threat to the safety of the autonomous vehicles can be regarded as freespace or traversable area. The non-traversable area is mainly an area composed of objects in the scene that pose a threat to the safe driving of autonomous vehicles. Unreachable area mainly refers to the area composed of objects that are relatively far away, which temporarily does not pose a threat to the safe driving of autonomous, and a typical example is the sky. Therefore, we mainly carry out pixel-level labeling of three types of objects: traversable area, non-traversable area and unreachable area, as shown in Fig. 5.

3.3. Dataset for Deep Learning

Deep learning methods have shown promising results in on-road freespace detection with the public on-road datasets [8, 1]. Only the RGB image lacking the geometry information is insufficient for freespace detection. We provide the ORFD dataset with the RGB image, LiDAR point cloud, calibration, sparse depth, dense depth and ground truth. The sparse depth was obtained by projecting the LiDAR point cloud to the image plane with the calibration information and then the sparse depth was interpolated to get the dense depth. We split the dataset into training, validation and testing set as shown in Table 1. From the table, we can see that the ratio of the three sets is about $7 : 1 : 2$.
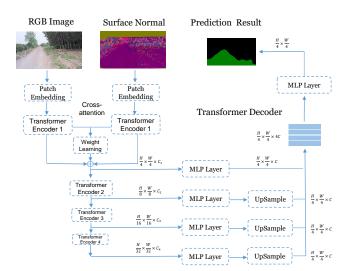

Figure 7. The architecture of our OFF-Net. The Transformer encoder extracts the features from both the RGB image and surface normal and a Transformer decoder predicts the freespace result. A cross-attention is designed to fuse data from both camera and LiDAR to dynamically leverage the strengths of each modality.

As it is hard to define road in grassland, we only choose one sequence of well-annotated data in grassland for training set.

3.4. Sensors

The vehicle used to collect the ORFD dataset is the Mazda Ruiyi 6 Coupe produced by FAW, with the body parameters (mm): $4755 \times 1795 \times 1440$, and the wheelbase (mm): 2725, and track width (mm): $1560/1560$. A sensor fusion kit Pandora produced by Hesai Technology is installed on the top of the vehicle to collect LiDAR point cloud and RGB image data. Pandora is composed of a 40-line mechanical LiDAR at the upper part and 5 cameras distributed around the lower part, including a color camera and 4 wide-angle black-and-white cameras. Detailed parameters of vehicle and sensors are shown in Fig. 6.

3.5. Synchronization and Calibration

Pandora controls the motor rotating and laser firing time of the LiDAR, and at the same time, LiDAR controls the exposure time and frame rate of the cameras. Therefore, Pandora can achieve synchronization of point cloud data from LiDAR and image data from the cameras. We use the method in [18] to calibrate the external parameters of the LiDAR.

4. Method

4.1. Problem Definition

We formulate the off-road freespace detection task as the pixel-wise classification problem on the RGB image plane,

Table 1. Training, validation and testing splits of the ORFD dataset.

Split	Farmland	Woodland	Grassland	Countryside	Sunny	Rainy	Foggy	Snowy	Bright light	Daylight	Twilight	Darkness	Total	%
Train	2718	3180	361	2139	3803	2434	1136	1025	1019	4254	927	2198	8398	68.8
Val	356	302	0	587	356	302	0	587	0	356	302	587	1245	10.2
Test	1129	1064	0	362	1071	405	720	359	361	710	359	1125	2555	20.9
Total	4203	4546	361	3088	5230	3141	1856	1971	1380	5320	1588	4510	12198	100

i.e., whether the pixel belongs to the traversable area. Given the RGB image $\mathbf{I}$ and LiDAR $\mathbf{L}$, the network F predicts the probability map. The goal of off-road freespace detection is to minimize the following loss function:

$$\min_{\theta} \sum_i \mathcal{L}(F(\mathbf{I}_i, \mathbf{L}_i), \hat{\mathbf{Y}}_i), \tag{1}$$

where θ is the parameter of network F and $\hat{\mathbf{Y}}$ is the freespace detection ground truth of the i-th training example.

4.2. OFF-Net

We developed a network called OFF-Net to combine camera and LiDAR information (*i.e.*, surface normal information calculated from the LiDAR point cloud). The reason why we choose the surface normal information as the network input is that the points within the road have similar surface normals, and the surface normals are calculated from dense depth images using the method proposed by Fan [6].

As the freespace detection task needs the network to have large receptive field, however, the CNN has the limited receptive field. Inspired by the success of the Transformer framework in capturing local and global information. We introduce the Transformer network architecture proposed by Xie [24] into the off-road freespace detection task. The structure of the our OFF-Net is illustrated in Fig. 7, and the special modules are described as follows.

4.2.1 Transformer Encoder

In order to obtain multi-level features, we first perform patch embedding on RGB image and surface normal separately with the resolution of $H \times W \times 3$. The outputs of patch emmbedding are hierarchical feature maps and the corresponding spatial resolutions are $\{1/4, 1/8, 1/16, 1/32\}$ of the input size. We then compute the multi-head self-attention function with the heads $\mathbf{Q}$, $\mathbf{K}$, $\mathbf{V}$:

$$Attention(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = softmax(\frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{d_{head}}})\mathbf{V}. \tag{2}$$

Since fixed position encoding will reduce the accuracy as described in [24], we use 3×3 convolution to capture position information for the Transformer encoder. The core formula of position encoding is as follow:

$$\mathbf{x}_{out} = MLP(GELU(Conv_{3\times3}(MLP(\mathbf{x}_{in})))) + \mathbf{x}_{in}, \tag{3}$$

where $\mathbf{x}_{in}$ is the feature from the multi-head self-attention part, *GELU* is the activation function proposed by Dan [13], *MLP* is a fully connected neural network layer and $Conv_{3\times3}$ is a 3×3 convolutional layer.

LiDAR point cloud data contains spatial geometric information but lacks semantic information, while monocular RGB image contains higher-level semantic information of the environment but lacks structual information. And features from RGB image and surface normal contribute differently to the final result of the freespace detection. In order to obtain the best weights for these two types of modalities to improve the detection performance, we design a dynamic fusion module. This module first adds the RGB image features and the surface normal features from LiDAR point cloud, and then sends the superimposed result to the MLP layer with the sigmoid activation function to learn the cross-attention. The cross-attention mechanism is simple but efficient in learning the weight of each modality. In this way, after the processing of the dynamic fusion module, we can get the refined features. The calculation formulas are as follows:

$$Cross\_Attention = \sigma(\mathbf{x}_{img\_in} + \mathbf{x}_{sn\_in}),$$
$$\mathbf{x}_{img\_out} = Cross\_Attention * \mathbf{x}_{img\_in} + \mathbf{x}_{img\_in}, \tag{4}$$
$$\mathbf{x}_{sn\_out} = (1 - Cross\_Attention) * \mathbf{x}_{sn\_in} + \mathbf{x}_{sn\_in},$$

where $\mathbf{x}_{img\_in}$ and $\mathbf{x}_{sn\_in}$ are the learned RGB image and surface normal features after the Transformer block, $\mathbf{x}_{img\_out}$ and $\mathbf{x}_{sn\_out}$ are the refined RGB image and surface normal features, and σ is the sigmoid activation function.

4.2.2 Transformer Decoder

The Transformer decoder is used to fuse local and global information, and it consists of only the MLP layer. The features from the Transformer encoder are first put into the MLP layer to aggregate channel-wise information, and then upsampled to the same size (*i.e.*, 1/4 of the input size). Finally, the features are fused together to obtain the freespace detection result.

4.2.3 Loss Function

We use the binary cross-entropy loss for off-road freespace detection, defined as follow:

$$loss = -\frac{1}{batch} \sum_{i=0}^{batch} \sum_{m=1}^{M} \sum_{n=1}^{N} \hat{\mathbf{Y}}_{mn}^{i} log \mathbf{O}_{mn}^{i}, \quad (5)$$

where $\mathbf{O}_{mn}^{i}$ is the predicted probability of pixel mn of the i-th training sample, and $\hat{\mathbf{Y}}_{mn}^{i}$ is the corresponding ground truth.

5. Experimental Results and Discussion

In this section, a series of experiments are conducted to validate the performance of the proposed dataset and method.

5.1. Experiments

5.1.1 Experimental Setting

We evaluate the proposed OFF-Net method on our ORFD dataset with two classes (*i.e.*, traversable area and non-traversable area, the unreachable area is regarded as non-traversable area for simplicity) and compare it with FuseNet [11] and SNE-RoadSeg [5]. FuseNet extracts features from RGB image and depth image and then fuses depth features into RGB feature maps with VGG16 [19] network architecture, and SNE-RoadSeg extracts and fuses features from RGB image and surface normal information with ResNet-152 [12] network architecture.

We use five common metrics for the performance evaluation of off-road freespace detection: 1) Accuracy=$\frac{TP+TN}{TP+TN+FP+FN}$, 2) Precision=$\frac{TP}{TP+FP}$, 3) Recall=$\frac{TP}{TP+FN}$, 4) F-score=$\frac{2TP^2}{2TP^2+TP(FP+FN)}$ and 5) IOU=$\frac{TP}{TP+FP+FN}$, where TP, TN, FP and FN represent the number of true positive, true negative, false positive, and false negative pixels, respectively.

5.1.2 Implementation Details

We use Pytorch [16] to implement our method, and the model is trained with the momentum (SGDM) [17] optimizer. The initial learning rate is 0.001 and the batch size is set as 8. All experiments are performed using 4 Nvidia RTX 3090 GPU devices. The image size for training and testing is set as 1280×704.

5.2. Results and Discussions

5.2.1 Evaluation on ORFD Dataset

Table 2 shows the quantitative results on the ORFD testing set. Compared to FuseNet with depth and RGB image information as input, SNE-RoadSeg performs much better,

which is because that the freespace can be assumed as a ground plane on which the points have similar surface normals while have different depth, thus the surface normal is more suitable than depth for freespace detection task. Our OFF-Net outperfoms FuseNet by 10.8% on F-score rate and 16.3% on IOU rate. Compared to SNE-RoadSeg, our method obtains 0.7% higher F-score rate and 1.1% higher IOU rate. The results show that our method with the Transformer framework can capture more local and global information for accurate freespace detection performance. Our OFF-Net uses only 25.2M parameters and takes 29.5 ms per input, satisfying real-time requirement. It runs $7 \times$ smaller and $2.7 \times$ faster than SNE-RoadSeg.

5.3. Ablation Studies

In this section, extensive ablation studies are performed to validate several components in the ORFD dataset and the proposed OFF-Net method.

5.3.1 Dataset

To evaluate the influence of different part in the off-road DRFD dataset, we do experiments with different model input on the proposed OFF-Net. This ablation study, presented in Table 3, shows that fusing two modalities can boost the freespace detection performance. The surface normal information is more important than depth information as the freespace has similar surface normals. But only using the surface normal information is not enough as the RGB image can provide the higher-level semantic information. It is needed to fuse the RGB image with the semantic information and LiDAR with the geometric information to leverage the strengths of each modality. Fusing the RGB image with the dense depth performs worse, as the dense depth was obtained by interpolating the sparse depth.

5.3.2 Transformer Encoder

We now analyze the influence of the number of transformer encoder block. In Table 4 we can observe that increasing the number of transformer encoder block can improve the performance with multi-level features. We use four transformer encoder blocks as the spatial resolution of stage 4 is very small.

5.3.3 Attention Mechanism

In this section, we study the effect of the proposed cross-attention mechanism on the fusion of RGB image and surface normals from LiDAR point cloud. It can be seen from Table 4 that the cross-attention increases the IOU rate from 80.9% to 82.3%, and the F-score rate from 89.4% to 90.3%. Therefore, we can draw a conclusion that the proposed cross-attention mechanism can dynamically assign weights

Table 2. Quantitative results on the ORFD testing set

Method	Modality	Accuracy	Precision	Recall	F-score	IOU	Params	Speed
FuseNet [11]	RGB + Sparse Depth	87.4%	74.5%	85.2%	79.5%	66.0%	50.0M	**49 Hz**
SNE-RoadSeg [5]	RGB + Surface Normal	93.8%	**86.7%**	92.7%	89.6%	81.2%	201.3M	12.5 Hz
OFF-Net (ours)	RGB + Surface Normal	**94.5%**	86.6%	**94.3%**	**90.3%**	**82.3%**	**25.2M**	33.9 Hz

(a) RGB Image   (b) Surface Normal   (c) Prediction Result   (d) Ground Truth
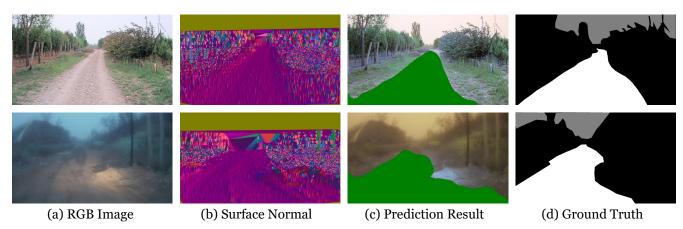
Figure 8. Qualitative results of our OFF-Net on the ORFD dataset. Our OFF-Net can predict off-road freespace accurately. There are also some unsatisfactory results as there are no concept of conventional roads.

Table 3. Impacts of different model input.

Modality	Acc.	Pre.	Recall	F-score	IOU
RGB	88.8%	75.5%	86.4%	80.6%	67.5%
Surface Normal	93.3%	83.7%	93.2%	88.2%	78.9%
RGB + Sparse Depth	90.1%	76.7%	90.9%	83.2%	71.3%
RGB + Dense Depth	86.4%	68.4%	92.3%	78.6%	64.8%
RGB + Surface Normal	**94.5%**	**86.6%**	**94.3%**	**90.3%**	**82.3%**

Table 4. Impacts of Transformer encoder and cross-attention.

Encoder	Cross-attention	Acc.	Pre.	Recall	F-score	IOU
1	✓	92.5%	80.1%	**96.0%**	87.3%	77.5%
2	✓	92.8%	81.1%	95.6%	87.7%	78.2%
3	✓	94.3%	**87.1%**	92.7%	89.8%	81.5%
4	✓	**94.5%**	86.6%	94.3%	**90.3%**	**82.3%**
4	✗	94.1%	86.4%	92.7%	89.4%	80.9%

to the RGB image and surface normal information from Li-DAR point cloud, and these weights are proven to be effective.

5.4. Qualitative Results

It can be seen from Fig. 8 that our OFF-Net can accurately estimate the off-road freespace. Specifically, OFF-Net has a good generalization ability for unseen scenes in the testing set. However, there are some failure cases, especially when the scene is very unstructured. Therefore, de-tecting freespace in an unstructured off-road environment is more challenging than in the on-road environment.

6. Conclusion

In this paper, we introduce an off-road freespace dataset, called the ORFD dataset, collected from a variety of off-road scenes, which, to our knowledge, is the first off-road freespace detection dataset. We believe that the ORFD dataset will help facilitate the study of autonomous navigation in off-road environments. We also introduce a new off-road freespace detection approach called OFF-Net, which unifies the Transformer network architecture to capture the context information. We design the cross-attention to dynamically aggregate information from both camera and LiDAR. In the future, we are planning to collect more freespace detection dataset to cover more off-road scenes.

References

[1] Jens Behley, Martin Garbade, Andres Milioto, Jan Quenzel, Sven Behnke, Cyrill Stachniss, and Jurgen Gall. Semantickitti: A dataset for semantic scene understanding of lidar sequences. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 9297–9307, 2019. 2, 4

[2] Holger Caesar, Varun Bankiti, Alex H Lang, Sourabh Vora, Venice Erin Liong, Qiang Xu, Anush Krishnan, Yu Pan, Giancarlo Baldan, and Oscar Beijbom. nuscenes: A multimodal dataset for autonomous driving. In *Proceedings of*

the IEEE/CVF conference on computer vision and pattern recognition, pages 11621–11631, 2020. 2

[3] Zhe Chen and Zijing Chen. Rbnet: A deep neural network for unified road and road boundary detection. In *International Conference on Neural Information Processing*, pages 677–687. Springer, 2017. 2

[4] Zhe Chen, Jing Zhang, and Dacheng Tao. Progressive lidar adaptation for road detection. *IEEE/CAA Journal of Automatica Sinica*, 6(3):693–702, 2019. 3

[5] Rui Fan, Hengli Wang, Peide Cai, and Ming Liu. Sneroadseg: Incorporating surface normal information into semantic segmentation for accurate freespace detection. In *European Conference on Computer Vision*, pages 340–356. Springer, 2020. 3, 6, 7

[6] Rui Fan, Hengli Wang, Bohuan Xue, Huaiyang Huang, Yuan Wang, Ming Liu, and Ioannis Pitas. Three-filters-to-normal: An accurate and ultrafast surface normal estimator. *IEEE Robotics and Automation Letters*, 6(3):5405–5412, 2021. 5

[7] Jannik Fritsch, Tobias Kuehnl, and Andreas Geiger. A new performance measure and evaluation benchmark for road detection algorithms. In *16th International IEEE Conference on Intelligent Transportation Systems (ITSC 2013)*, pages 1693–1700. IEEE, 2013. 2

[8] Andreas Geiger, Philip Lenz, and Raquel Urtasun. Are we ready for autonomous driving? the kitti vision benchmark suite. In *2012 IEEE conference on computer vision and pattern recognition*, pages 3354–3361. IEEE, 2012. 2, 4

[9] Shuo Gu, Yigong Zhang, Jinhui Tang, Jian Yang, and Hui Kong. Road detection through crf based lidar-camera fusion. In *2019 International Conference on Robotics and Automation (ICRA)*, pages 3832–3838. IEEE, 2019. 2

[10] Shuo Gu, Yigong Zhang, Jian Yang, Jose M Alvarez, and Hui Kong. Two-view fusion based convolutional neural network for urban road detection. In *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 6144–6149. IEEE, 2019. 2

[11] Caner Hazirbas, Lingni Ma, Csaba Domokos, and Daniel Cremers. Fusenet: Incorporating depth into semantic segmentation via fusion-based cnn architecture. In *Asian conference on computer vision*, pages 213–228. Springer, 2016. 6, 7

[12] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016. 6

[13] Dan Hendrycks and Kevin Gimpel. Gaussian error linear units (gelus). *arXiv preprint arXiv:1606.08415*, 2016. 5

[14] Peng Jiang, Philip Osteen, Maggie Wigness, and Srikanth Saripalli. Rellis-3d dataset: Data, benchmarks and analysis. *arXiv preprint arXiv:2011.12954*, 2020. 2

[15] Daniel Maturana, Po-Wei Chou, Masashi Uenoyama, and Sebastian Scherer. Real-time semantic mapping for autonomous off-road navigation. In *Field and Service Robotics*, pages 335–350. Springer, 2018. 2

[16] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, et al. Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems*, 32:8026–8037, 2019. 6

[17] Boris T Polyak. Some methods of speeding up the convergence of iteration methods. *Ussr computational mathematics and mathematical physics*, 4(5):1–17, 1964. 6

[18] Erke Shang, Xiangjing An, Meiping Shi, Deyuan Meng, Jian Li, and Tao Wu. An efficient calibration approach for arbitrary equipped 3-d lidar based on an orthogonal normal vector pair. *Journal of Intelligent & Robotic Systems*, 79(1):21–36, 2015. 4

[19] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014. 6

[20] Jee-Young Sun, Seung-Wook Kim, Sang-Won Lee, Ye-Won Kim, and Sung-Jea Ko. Reverse and boundary attention network for road segmentation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops*, pages 0–0, 2019. 2

[21] Pei Sun, Henrik Kretzschmar, Xerxes Dotiwalla, Aurelien Chouard, Vijaysai Patnaik, Paul Tsui, James Guo, Yin Zhou, Yuning Chai, Benjamin Caine, et al. Scalability in perception for autonomous driving: Waymo open dataset. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2446–2454, 2020. 2

[22] Abhinav Valada, Gabriel L Oliveira, Thomas Brox, and Wolfram Burgard. Deep multispectral semantic scene understanding of forested environments using multimodal fusion. In *International symposium on experimental robotics*, pages 465–477. Springer, 2016. 2

[23] Maggie Wigness, Sungmin Eum, John G Rogers, David Han, and Heesung Kwon. A rugd dataset for autonomous navigation and visual perception in unstructured outdoor environments. In *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 5000–5007. IEEE, 2019. 1, 2

[24] Enze Xie, Wenhai Wang, Zhiding Yu, Anima Anandkumar, Jose M Alvarez, and Ping Luo. Segformer: Simple and efficient design for semantic segmentation with transformers. *arXiv preprint arXiv:2105.15203*, 2021. 5