Off-Road Drivable Area Extraction Using 3D LiDAR Data

Biao Gao*, Anran Xu*, Yancheng Pan*, Xijun Zhao†, Wen Yao†, Huijing Zhao*
*Peking University, Beijing, China
†China North Vehicle Research Institute, Beijing, China

***Abstract*— nothing.**

I. INTRODUCTION

nothing.

II. RELATED WORKS

nothing.

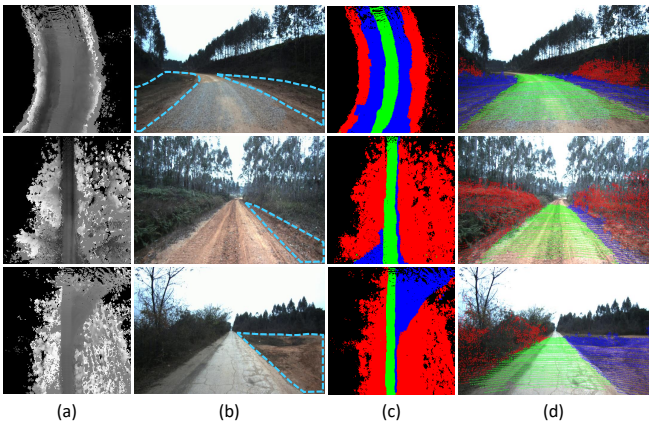

Fig. 1. The ambiguities in off-road drivable area extraction. (a) Input LiDAR data in bird's-eye view. (b) Image reference of input data. (c) Human annotation. (d) Projected point clouds in camera coordinate.

III. METHODOLOGY

A. Problem Definition

Let's denote the origin point clouds from 3D LiDAR sensor as $PC = \{pt_i\}_{0 \le i < N}$, where N is the number of points. We aggregate a few frames' point clouds to get a dense bird's-eye view height map $X = \{x_{j,k}\}_{0 \le j < H, 0 \le k < W}$, which is used as our input data format. The input height map is in the size of $H \times W$ and each pixel $x_{j,k}$ means the physical height of pixel (j, k). The examples of input can be seen in Fig.1(a).

Different from the well-defined road borders in structured urban environment, the main peculiarity in off-road environment is the ambiguous area beside the road margin, which is called grey zone. In order to distinguish it with others, we let $LabelSet = \{unknown, drivable\ zone, obstacle\ zone, grey\ zone\}$, and we use $G = \{g_{j,k}\}_{0 \le j < H, 0 \le k < W}$ to denote human annotated ground truth, where $g_{j,k} \in LabelSet$.

The origin output of our proposed framework is a cost map $C = \{c_{j,k}\}_{0 \le j < H, 0 \le k < W}$, where each $c_{j,k} \in [0, 1]$ evaluates the traversability cost of pixel (j, k). Our proposed framework learns a mapping from input X to cost map C.

$$f_\theta^* : x_{j,k} \rightarrow c_{j,k} \in [0, 1] \tag{1}$$

For the convenience of comparison with human-annotated ground truth G and other baseline methods, we use d_α to remap cost map C to label $Y = y_{j,k} \in LabelSet$.

$$d_\alpha : c_{j,k} \rightarrow y_{j,k} \in LabelSet \tag{2}$$

Therefore, the problem of this work can be formulated as learning a multi-class classifier $f_\theta = d_\alpha(f_\theta^*)$ that maps input $x_{j,k}$ to a label $y_{j,k} \in LabelSet$.

B. Network Architecture

Due to the ambiguity of the grey zone, we hold the view that classifying it as an independent label from the others is not reasonable enough. In some cases, the grey zone is drivable technically but not human-desired, which is very close or even overlapped with the drivable zone in feature space. In other cases, the grey zone may have higher traversability cost than common drivable zone, which is closer or even overlapped with the obstacle zone in feature space. As a result, viewing the ambiguous grey zone as an independent label in training process will cause confusion for the deep learning model, and the experimental results in Sec.V give the evidence of this viewpoint.

Therefore, we can assume that the samples of grey zone are distribute between the drivable samples and the obstacle samples in feature space. The key idea of our proposed method is learning two classification surfaces in feature space and use the margin between them to split the grey zone samples. One classification surface is used to separate the drivable zone samples from others. The other one is used to separate the obstacle zone samples. We can evaluate the traversability cost of a sample by its feature distance to the two surfaces.

As shown in Fig.2, the proposed network has two branches in order to learn the two classification surfaces mentioned above. Both of them are designed according to the common VGG-based fully convolutional network. The difference is that the last layer does not output the discrete labels but the probabilistic predictions from the last softmax layer. We denote them as S_1 and S_2, which are in the range of $[0, 1]$.

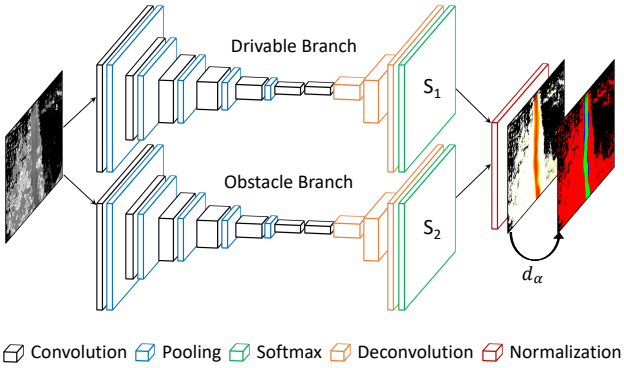

Fig. 2. Network

Each branch is trained end-to-end guided by the following cross entropy loss function:

$$L^{br}(X;\Theta^{br}) = -\sum_{i} y_i^{br} \log P(y_i^{br}|\Theta^{br}),\ br \in \{dri, obs\} \tag{3}$$

where $br \in \{dri, obs\}$ is the name of network's branch. $P(y_i^{br}|\Theta^{br})$ is the probability that pixel i is predicted as label y_i^{br} with the network parameters Θ^{br}. We use dri, obs and gre to represent $drivable, obstacle$ and $grey\ zone$.

When training the network, different label y_i^{br} is used in the two branches.

$$y_i^{br} = \begin{cases} \Psi(\vec{br}), & if\ g_i = gre \\ \vec{g_i} & if\ g_i \in \{dri, obs\} \end{cases} \tag{4}$$

where $\vec{br}$ is the one-hot vector of label $br \in \{dri, obs\}$. $\Psi(\vec{dri}) = \vec{obs}$ and $\Psi(\vec{obs}) = \vec{dri}$. For example, we replace all the pixels with ground truth $g_i = gre$ to the label obs to get y_i^{dri} when training the drivable branch.

We use the following regulation to calculate the traversability cost map C and then get the discrete label:

$$C = \begin{cases} S_1, & if\ S_1 > \alpha_1 \\ 1 - S_2, & if\ S_2 > \alpha_2 \\ \frac{1-S_2}{1-S_1+1-S_2}, & otherwise \end{cases} \tag{5}$$

where the pixels satisfied the first condition will be labeled as $drivable\ zone$, the second is $obstacle\ zone$ and the last is $grey\ zone$.

C. Semi-supervised Learning

In order to reduce the demand of the high-cost human-annotated data, we propose a semi-supervised learning method shown in Fig.3. This framework can use a large number of auto-generated labels and only a small fraction of human-annotated labels to train the network.

For human-annotated data X_h, we use the loss function Equation (3) for training. For data X_w with only auto-generated weakly labels, we define the loss L_{semi}^{br} in branch br as below:

$$L_{semi}^{br}(X_w;\Theta^{br}) = -\lambda \sum_{j} \widetilde{y_j^{br}} \log P(y_j^{br}|\Theta^{br}) \tag{6}$$

where $\widetilde{y_j} \in \widetilde{Y}$ denotes the auto-generated labels and λ is a regularization weight.

When training the network with human-annotated labels and auto-generated labels simultaneously, combine the Equation (3) and Equation (6) together.

$$L^{br}(X_h, X_w;\Theta^{br}) = L^{br}(X_h;\Theta^{br}) + L_{semi}^{br}(X_w;\Theta^{br}) \tag{7}$$

In each training batch, the mean loss of human-annotated and auto-generated data will be used for back propagation.

IV. IMPLEMENTATION DETAILS

A. Training Setup

B. Ground Truth Labeling

C. Evaluation

V. EXPERIMENTAL RESULTS

A. Data set

B. Proposed Method Results

C. Limitations

VI. CONCLUSION

nothing

APPENDIX

ACKNOWLEDGMENT

nothing.

REFERENCES

[1] G. O. Young, Synthetic structure of industrial plastics (Book style with paper title and editor), in Plastics, 2nd ed. vol. 3, J. Peters, Ed. New York: McGraw-Hill, 1964, pp. 1564.

TABLE I

EVALUATION MEASURES

Drivable Zone		Obstacle Zone	
Definition	Explanation	Definition	Explanation
$Q_1 = TP(G_{dri})/\|Y_{dri}\|$	$TP(G_{dri}) = \|G_{dri} \cap Y_{dri}\|$	$Q_1 = TP(G_{obs})/\|Y_{obs}\|$	$TP(G_{obs}) = \|G_{obs} \cap Y_{obs}\|$
$Q_2 = TP(G_{dri})/\|G_{dri}\|$	$TP(G_{dri}) = \|G_{dri} \cap Y_{dri}\|$	$Q_2 = TP(G_{obs})/\|G_{obs}\|$	$TP(G_{obs}) = \|G_{obs} \cap Y_{obs}\|$
$Q_3 = TP(VP_{dri})/\|VP_{dri}\|$	$TP(VP_{dri}) = \|VP_{dri} \cap Y_{dri}\|$	/	/
$F_1 = 2Q_1Q_2/(Q_1 + Q_2)$	F_1 Measure	$F_1 = 2Q_1Q_2/(Q_1 + Q_2)$	F_1 Measure

dri: Drivable zone **obs**: Obstacle zone **G**: Ground truth **Y**: Prediction **VP**: Vehicle Path $\|\mathbf{X}\|$: Pixel number in X

TABLE II

QUANTITATIVE EVALUATION OF DIFFERENT METHODS

	Drivable zone				Obstacle zone		
	Q_1 (PRE)	Q_2 (REC)	Q_3 (ACC)	F_1	Q_1 (PRE)	Q_2 (REC)	F_1
3-class FCN (fully-sup.)	74.93	82.99	**98.92**	78.75	94.36	**98.44**	96.36
Ours (fully-sup.)	**76.01**	**86.72**	98.09	**81.01**	**96.20**	96.75	**96.47**
RG-FCN (weakly-sup.)	59.78	79.15	93.16	68.11	94.46	95.38	94.92
Oxford PP (weakly-sup.)	**97.00**	47.38	83.71	63.66	**98.40**	89.84	93.93
Ours (weakly-sup.)	72.38	78.83	95.21	75.47	96.31	94.84	95.57
Ours (semi-sup.)	81.73	**81.73**	**96.24**	**81.73**	95.60	**97.38**	**96.49**

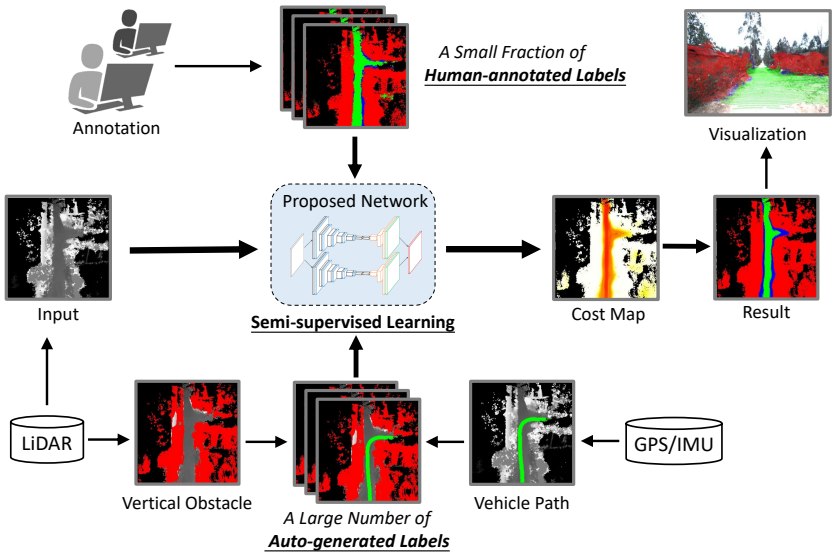

Fig. 3. Overview of the proposed off-road drivable area extraction framework

TABLE III

QUANTITATIVE COMPARISON OF F_1 MEASURE

-	6.25% semi-sup.	12.5% semi-sup.	25% semi-sup.	50% semi-sup.	Fully-sup.
F_1 measure	73.86	75.04	78.96	**81.73**	81.01

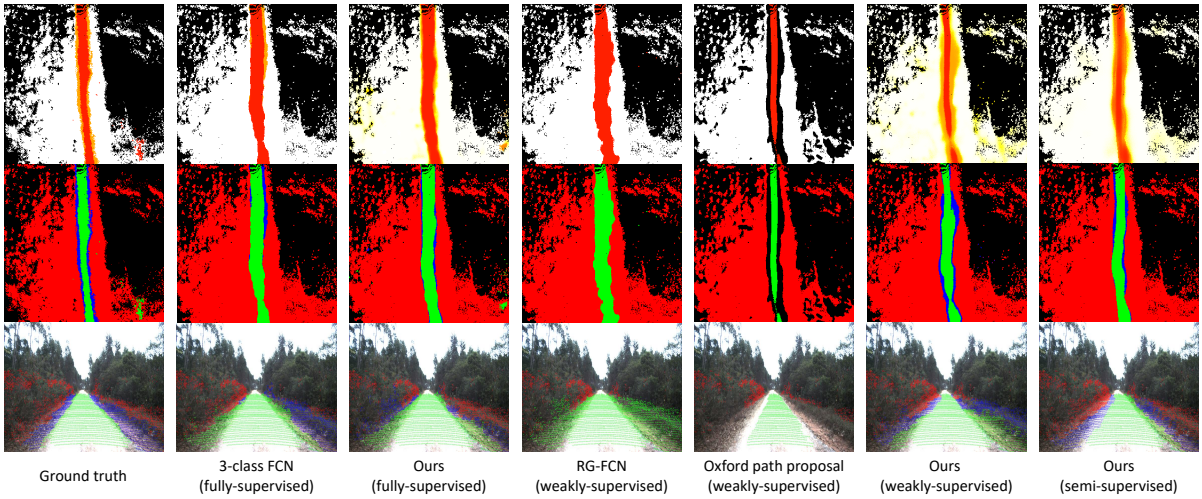

Fig. 4. Qualitative results at straight road scene.

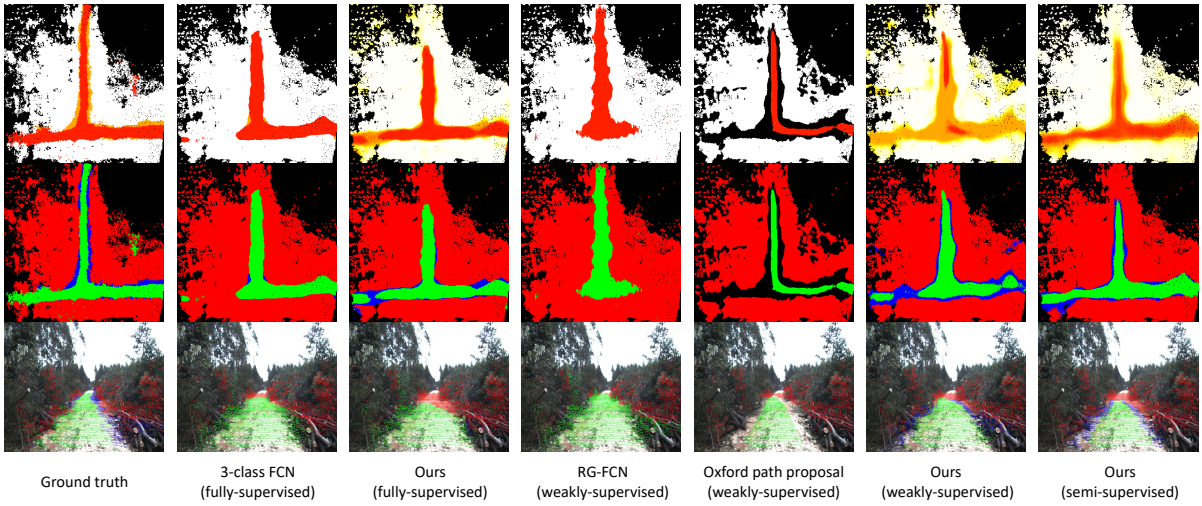

Fig. 5. Qualitative results at cross road scene.

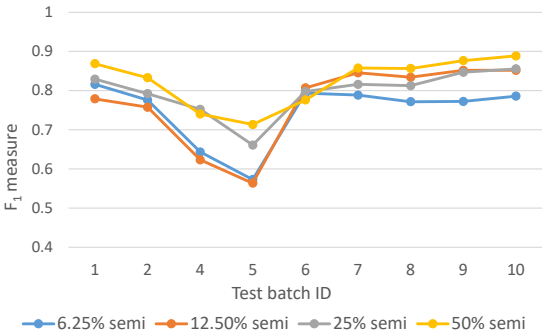

Fig. 6. Quantitative comparison on test set.