2021 IEEE 17th International Conference on Automation Science and Engineering (CASE)
August 23-27, 2021. Lyon, France

2021 IEEE 17th International Conference on Automation Science and Engineering (CASE) | 978-1-6654-1873-7/21/$31.00 ©2021 IEEE | DOI: 10.1109/CASE49439.2021.9551643

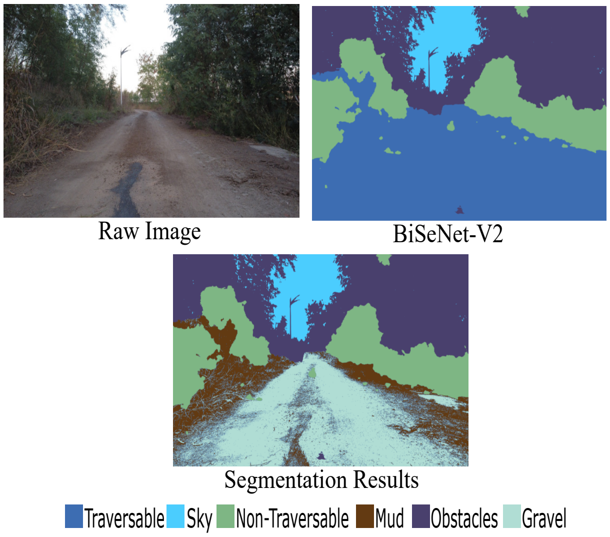
OFFSEG: A Semantic Segmentation Framework For Off-Road Driving

Kasi Viswanath*, Kartikeya Singh*, Peng Jiang, P.B. Sujit and Srikanth Saripalli

***Abstract*— Off-road image semantic segmentation is challenging due to the presence of uneven terrain, unstructured class boundaries, irregular features and strong textures. These aspects affect the vehicle perception. Current off-road datasets exhibit difficulties like class imbalance and understanding of varying environmental topography. To overcome these issues, we propose a framework for off-road semantic segmentation (OFFSEG) that involves (i) a pooled class semantic segmentation with four classes (sky, traversable region, non-traversable region and obstacle) using state-of-the-art deep learning architectures (ii) a color segmentation methodology to segment out specific sub-classes (grass, puddle, dirt, gravel, etc.) from the traversable region for better scene understanding. The evaluation of the framework is carried out on two off-road driving datasets, namely, RELLIS-3D and RUGD. We have also tested the proposed framework on IISERB campus data. The results show that OFFSEG achieves good performance and also provides detailed information on the traversable region.**

I. Introduction

Autonomous off-road driving has a wide range of applications like inspection, exploration, rescue, reconnaissance missions, etc. Off-road environments are often texture rich with indefinite boundaries and less detailed than the urban environments. Non-uniform terrain description makes off-road environments more difficult to understand from the perception perspective for a robust autonomous driving system. On-road driving has received significant attention in the domain of autonomous driving in terms of datasets for segmentation. The datasets [1][2][3] are available for the semantic scene understanding of the on-road environment as well as the state-of-the-art benchmarks for these environments. Compared to urban environments, off-road environments have unstructured class boundaries, uneven terrain, strong textures, and irregular features that hinder the direct transfer of models between the environments. Moreover, there are large differences in class distributions across distinct off-road environments. Thus, there is a need to develop a framework that provides feature-rich semantic information to the vehicle for making better decisions while driving in off-road environments. For example, consider Figure 1 that provides detailed information on the traversable region like mud and gravel which could be used by the path planning module of an autonomous vehicle for better planning as compared to the traversable region without any additional information.


Raw Image

BiSeNet-V2

Segmentation Results

Fig. 1. Segmentation results on IISERB campus frames (top left) Raw image taken from the vehicle (top right) Semantic segmentation of four clases using BiSeNetV2 (below) color segmentation on the traversable region providing additional information about the traversable region

The limited availability of off-road environment based datasets is another challenge that affects the progress in the off-road autonomous driving domain. For the robotic navigation in the off-road environment there are three main datasets (1) RELLIS–3D [4], (2) RUGD [5], and (3) DeepScene [6]. RELLIS–3D [4] is a multimodal dataset collected in an off-road environment, which contains annotations for 13,556 LiDAR scans and 6,235 images where ground truth in terms of annotated labels are provided. RUGD [5] dataset gives a rich ontology and large set of ground truths of 7546 annotations with 24 classes. However, in both RELLIS-3D and RUGD, classes like log, pole, water exhibit low pixel density resulting in class imbalance and hence low mean Intersection over Union (mIoU).

An interesting aspect of off-road driving is that unlike on-road scenes, which have detailed classes like signboard, traffic lights, etc, off-road environments require less features which allows us to pool the classes in RELLIS-3D and RUGD dataset to four classes namely traversable, non-traversable, obstacles and sky. The motive behind clustering classes into four was to group classes according to their semantic contributions in the environment. The sky class explicitly includes the region present in the sky. The traversable class includes all possible traversable regions present in the datasets. The non-traversable class includes all surface regions which are not traversable and do not act as an

Kasi Viswanath, Kartikeya Singh and P.B. Sujit are with the Department of Electrical Engineering and Computer Science, IISER Bhopal, Bhopal - India. e-mail:(kasi18,kartikeyas,sujit)@iiserb.ac.in

Peng Jiang and Srikanth Saripalli are with the Department of Mechanical Engineering, Texas A&M University, College Station, Texas, TX– 77843-3123. e-mail:(maskjp,ssaripalli)@tamu.edu

* Kasi Viswanath and Kartikeya Singh are equal contributors.

obstacle during the off-road navigation. Lastly, the obstacle class explicitly includes all the possible obstacles present in the dataset. By grouping the classes into four, we resolve the class imbalance issue adequately. The traversable class can include additional information like dirt, mud, gravel, etc. These additional details can play an important role in determining the drivable path during autonomous off-road driving.

A transfer learning framework with semantic segmentation for off-road environments was developed in [7]. However, the approach has lower performance for classifying finer features like grass, gravel, bush, etc., which are essential attributes to perform robust robotic navigation in off-road environments. Nefian and Bradski [8] use a hierarchical Bayesian network approach to detect driving regions. However, the approach does not provide detailed features and the environment is not heterogeneous.

The main contributions of this paper are

- Development of a novel, simple and efficient semantic segmentation framework for off-road driving OFFSEG.
- We exploit the off-road driving requirement by pooling the 20 classes in RELLIS-3D and 24 classes in RUGD into 4 classes and mitigate the class imbalance issue.
- We use color layers [9] and K-Means [10] clustering to determine RGB clusters [11] [12] to find finer details on the traversable region.
- We compare OFFSEG results with BiSeNetV2 and HRNETV2+OCR on RELLIS-3D and RUGD. The mIoU of OFFSEG cannot be compared with the results of HRNETV2+OCR and GSCNN from RELLIS-3D because of the 20 pooled to 4 in our proposed framework. Similarly for RUGD the 24 classes were pooled to 4 hence we cannot make a direct inference of the results with other benchmarks.
- We also test OFFSEG on IISERB campus data that was recorded from Indian Institute of Science Education and Research Bhopal (IISERB) campus. The results obtained from all the datasets shows detailed segmentation of all classes.
- We test OFFSEG on NVIDIA Jetson AGX Xavier [13] to record the inference speed of our framework for realtime application.

The rest of the paper is organized further. Section II describes the proposed OFFSEG framework and OFFSEG results are presented in Section III. The conclusions and future work in described in Section IV.

II. Methodology

OFFSEG consists of two stages: semantic segmentation for four classes and color segmentation of traversable region. The input is a raw RGB image and the output is a pixel-wise annotated RGB image. An overview of OFFSEG architecture is represented in Figure 2.

A. Semantic segmentation

Semantic Segmentation is a method of labeling the class of each pixel in an image. Traditionally, image segmentation was carried out by threshold selection, region growing, and other approaches. Recent developments of the Convolutional Neural Network (CNN)[14] yield faster and accurate state-of-the-art segmentation. Most CNN-based architectures consist of an encoder-decoder structure which downsamples the input to extract features and then upsamples with a pooling layer. This results in loss of spatial details. Architectures like HRNET[15] adopts a high resolution multiple branches to recover the spatial information. As shown in Figure 2.a, the first stage of our framework is to perform semantic segmentation on 4 classes. The 20 classes in RELLIS-3D and 24 classes in RUGD datasets were re-categorized into four classes, 1) sky, 2) traversable, 3) non-traversable 4) obstacle. From RELLIS-3D dataset 6 classes were pooled to traversable, 3 were pooled to non-traversable and 10 were pooled to obstacles. As the RELLIS-3D dataset had 94% of the pixels distributed between sky, grass, tree and bushes, pooling them into four different classes solved the problem of class imbalance issue as shown in Figure 3. The pixel-wise annotation of the classes from RELLIS-3D and RUGD were then converted into these four classes for training on the semantic segmentation network.

The traversable class includes sub-classes like puddle, mud, dirt, gravel, etc. as given in Table I. These sub-classes play an important role in determining path during robotic navigation on the off-road environments. Another reason to consider only the traversable class as our region of interest (RoI) is to ignore all other unusable sub-classes present in the environment which are not necessary for determining traversable paths in autonomous driving like pole, bush, etc. These instances do not require fine segmentation to achieve.

B. Color segmentation and sub-class classification

K-Means algorithm has been used to extract the color pools from the output obtained in the previous section. Color pools are used to distinguish between several components present in an off-road environment. Each cluster obtained from the centroid has been transferred into the classification model which gives us the mapping of the required sub-classes in our region of interest as shown in Figure 2.b. The color segmentation algorithm extracts the color masks from the image and inputs these masks into the classification model. The classifier classifies the sub-classes in terms of color clusters and determines the sub-classes like mud, puddle, grass, water, etc. as shown in Figure 2.c. Next, these obtained masks are appended on the segmentation output which was obtained from semantic segmentation resulting in final segmentation as shown in Figure 6.

1) Data pre-processing and data generation for classification: In order to classify the sub-classes of the traversable region, we need to create an image-oriented dataset for the training of a classification model. This dataset comprises of the detailed sub-classes present in the traversable region. The training samples in RELLIS-3D has 6 classes in the traversable region – grass, mud, puddle, dirt, asphalt and

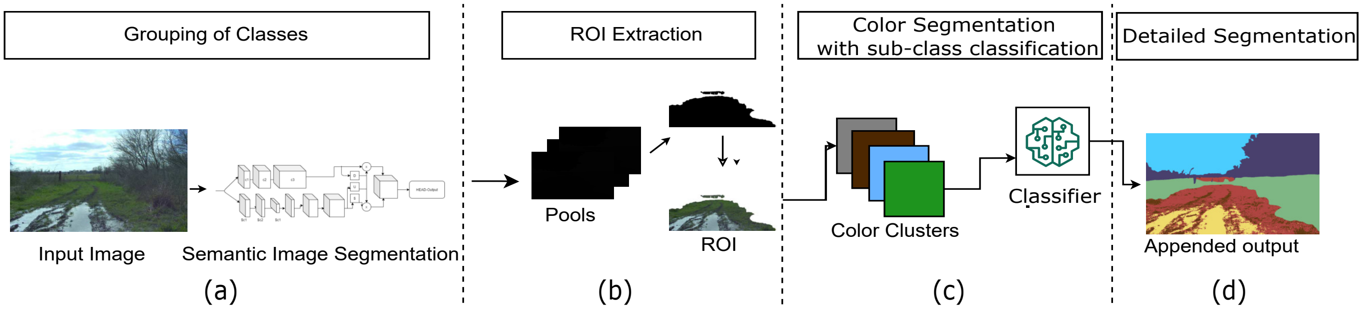


Fig. 2. OFFSEG consists of two stages. (a) Pooling of different classes into four and performing semantic image segmentation (b) The RoI (region of interest) is obtained from the segmentation (c) Color segmentation algorithm is used to segment and classify sub classes like grass, mud, puddle, etc. (d) The append output.

Class Distribution for RELLIS-3D and RUGD			
Sky	**Traversable**	**Non Traversable**	**Obstacles**
Sky[RE,RU]	Grass[RE,RU]	Bush[RE,RU]	Vehicle[RE,RU]
	Dirt[RE,RU]	Void[RE]	Barrier[RE]
	Asphalt[RE,RU]	Water[RE,RU]	Log[RE,RU]
	Concrete[RE,RU]	Deep Water[RE]	Pole[RE]
	Puddle[RE]		Object[RE]
	mud[RE]		Building[RE,RU]
	Sand[RU]		Person[RE,RU]
	Gravel[RU]		Fence[RE,RU]
	Mulch[RU]		Tree[RE,RU]
	Bridge[RU]		Rubble[RE]
	Rockbed[RU]		Pole[RU]
			Container[RU]
			Bicycle[RU]
			Sign[RU]
			Rock[RU]
			Table[RU]

TABLE I
CLASS DISTRIBUTION OF THE POLLED CLASSES FROM RELLIS-3D[RE] AND RUGD[RU]

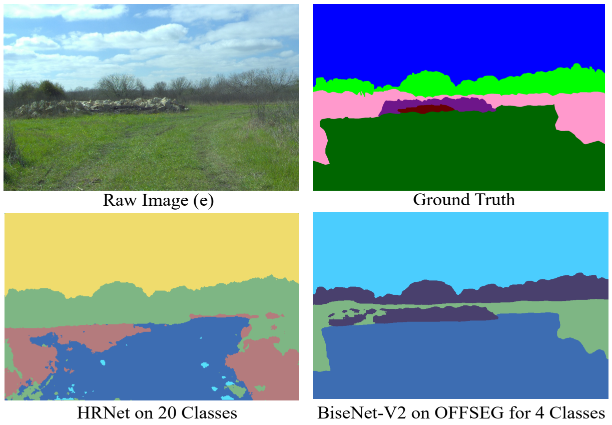

Raw Image (e)
Ground Truth
HRNet on 20 Classes
BiseNet-V2 on OFFSEG for 4 Classes

Fig. 3. Segmentation results from OFFSEG for four class model have been compared with the HRNET 20 class model

concrete, while RUGD has 8 classes in the traversable region – dirt, sand, grass, water, asphalt, gravel, mulch and concrete.

2) Training of classification model: The output from color segmentation needs to be further classified into different subclasses in the traversable region. The classifier differentiates the masks extracted from K-Means clustering and assigns the respective classes to them. Table I shows the different subclasses present in the traversable region of both RELLIS-3D and RUGD datasets.

III. Results

The OFFSEG approach was evaluated on two state-of-the-art off-road datasets RELLIS-3D and RUGD using BiSeNetV2 and HRNETV2+OCR architectures. We have also tested OFFSEG on IISERB campus data.

A. Segmentation

The evaluation of image semantic segmentation of the converted classes of RELLIS-3D and RUGD were done using two state-of-the-art architectures: BiSeNetV2[16] and HRNETV2[15]+OCR[17]. BiSeNetV2 consists of two branches: detail branch and semantic branch. The detail branch extracts spatial details consisting of low-level information and uses shallow layers with wide channels. Meanwhile, semantic branch extracts high-level semantics employing low channel capacity. Then an aggregation layer merges extracted features from the two branches and upsample the output from aggregation layer.

HRNETV2+OCR consists of a High-Resolution Network which acts as a backbone and Object-Contextual Representations (OCR) to enhanced pixel representation of objects. Unlike other segmentation models HRNET maintains high resolution throughout the model avoiding the downsample and upsample process. OCR aggregated the features extracted from HRNET to improve pixel representation. We used 3,302 images for training set, 983 images for validation set and the testing with 1672 images for the RELLIS-3D. For RUGD, we used 4732 images for training set, 932 images for the validation set and 1827 images for the testing set.

1) Quantitative analysis of the architectures used for segmentation: The results obtained from OFFSEG can be seen in Figure 8. The individual IoU breakout of four classes is given in Table II. The mean IoU [18] for the datasets is given by:-

$$mIoU = \frac{1}{n}\sum_{n=1}^{n} Z, \tag{1}$$

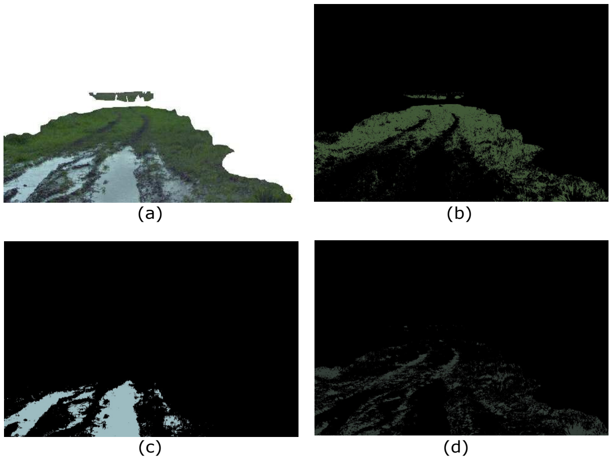


Fig. 4. Color segmentation results on RELLIS-3D (a) Traversable class obtained as RoI from segmentation (b) grass, (c) puddle, (d) mud obtained from color segmentation from RoI.

where, Z is defined as

$$Z = \frac{(TruePositive)_n}{(TruePositive)_n + (FalsePositive)_n + (FalseNegative)_n} \tag{2}$$

where n is the number of classes.

From Table II, the mean IoU obtained for RELLIS-3D on BiSeNetV2 and HRNETV2+OCR were 86.61% and 80.82% respectively. The mean IoU obtained for RUGD on BiSeNetV2 and HRNETV2+OCR were 80.17% and 84.49% respectively. The results obtained in Figure 3 shows the prediction of BiSeNetv2 on a RELLIS-3D frame which contains the class log. The obstacle class in the prediction covers most of the log ground truth labels inferring higher predictions than the prediction of HRNETV2+OCR trained on 20 classes for the log which had 0.0% IoU.

B. Clustering

We obtain color clusters using K-Means algorithm. The set of random k-points has been assigned with the closest centroid from the image which further combines these centroids into separate clusters. By adopting an iterative approach, we obtain a set of all possible color clusters present in the RGB layers. The number of required clusters depends on the incorporation of sub-class properties in a frame with respect to the sub-classes present in the traversable region of the dataset.

C. Color segmentation and sub-class classification

The color masks obtained after applying color segmentation on the RoI is shown in Figure 4 and Figure 5. The accuracy of OFFSEG is determined by how precisely the sub-classes are classified. We trained a classification model using transfer learning with MoblieNetV2[19] as the classifier model pre-trained on ImageNet dataset[20]. The training inputs are classes listed in the traversable region of Table I from both RELLIS-3D and RUGD dataset.The classifier was trained on 23,967 images for 9 classes which achieved a mean accuracy of 97.3% and the outputs obtained from the model are used into the color segmentation algorithm which appends only classified sub-classes into the final result. Note that, quantitative analysis for the color segmentation would lead to inaccurate results as the ground truth for classes in the traversable region is very vague whereas the outputs in our approach are more feature-rich with distinct boundaries. The detailed outputs obtained using OFFSEG expands the application space of the model.

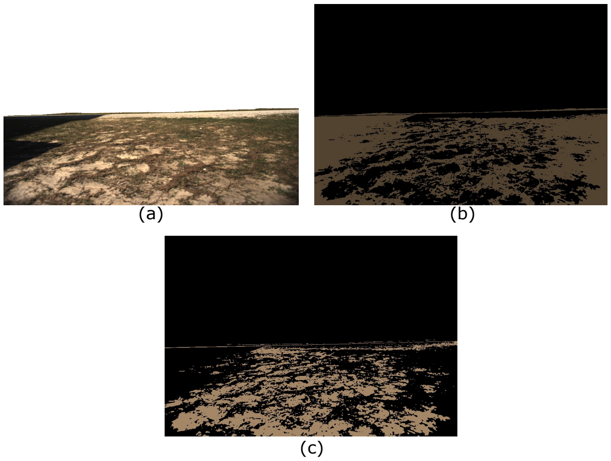


Fig. 5. Color segmentation results on RUGD (a) Traversable class obtained as RoI from segmentation (b) mulch, (c) gravel obtained from color segmentation from RoI.

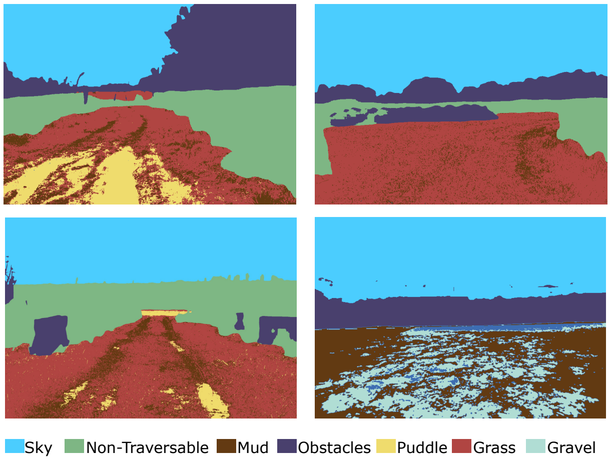


Fig. 6. Final segmentation of classes (sky, traversable, non-traversable, obstacles) and sub-classes (mud, puddle, grass, gravel) in a,e,b,d frames respectively.

D. Inference speed

The inference speed of the whole framework with BiSeNetV2 and HRNETV2+OCR as segmentation model was tested on Jetson AGX Xavier platform [13]. The input resolution of RELLIS-3D is 1024*640 and RUGD is 688*550. The performance in terms of inference speed and MIoU is shown in Figure 7. The figure demonstrates that BiSeNetV2 performs better in terms of fps than HR-NetV2+OCR.

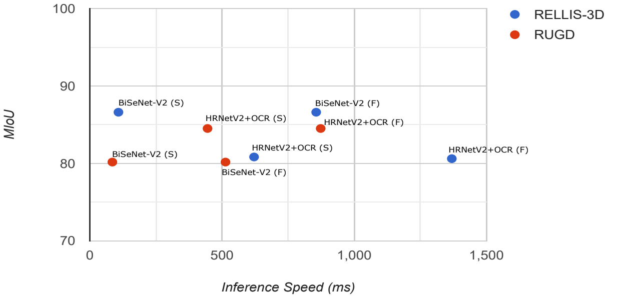


Fig. 7. OFFSEG Performance: Segmentation(S) and OFFSEG(F). The testing was performed on Jetson AGX Xavier platform.

Classes / Models	Sky	Traversable	Non-Traversable	Obstacles	**mIoU**
BiSeNet-V2 [RELLIS-3D]	97.09%	92.30%	77.12%	79.93%	86.61%
HRNETV2 [RELLIS-3D]	96.85%	86.04%	66.22%	74.18%	80.82%
BiSeNet-V2 [RUGD]	90.85%	91.83%	47.81%	90.20%	80.17%
HRNETV2 [RUGD]	92.27%	94.18%	59.92%	91.60%	84.49%
BiSeNet-V2 [IISERB]	95.71%	85.93%	49.31%	66.58%	74.38%

TABLE II

MIOU OF EXPERIMENT RESULTS

E. Testing of framework on IISERB campus data

We further evaluated OFFSEG in a different environment data from IISERB campus. Figure 9 shows the raw images from IISERB campus. The data used for testing includes combination of sub-classes present in RELLIS-3D and RUGD datasets. The data was recorded in a sequential manner using Dji Mavic Mini[21] from an altitude of 1.8 meters. The ground truth is generated using LabelBox[22] for the evaluation. The BiSeNetV2 model trained using RUGD dataset was used for testing and obtained an mIoU of 74.38% and the individual class IoU breakout is shown in Table III. The output obtained from the model followed by our color segmentation algorithm were adequate to provide the detailed sub-class information. From the outputs, we observe that the color segmentation is a very effective mechanism to classify among different sub-classes present in the traversable region.

IV. Conclusion and Future Work

In this work, we have presented an off-road semantic segmentation (OFFSEG) framework for fine semantic segmentation on two off-road datasets. OFFSEG shows affirmative results for achieving good mIoU on off-road environments. The sub-class segmentation within the traversable region from OFFSEG can be used for robust scene understanding and optimized path planning for navigation through off-road environments. Our framework performed well in solving the class imbalance issue which was being accounted in the benchmarks of RELLIS-3D dataset. This framework can be extended to include other sub-classes which are not included in RELLIS-3D and RUGD within the traversable region. Another interesting direction is to study the robustness of the approach under different climatic conditions change as the vegetation and texture of an off-road scene change significantly compared to urban environments. Another interesting study could be the robustness of the approach under different weather conditions.

References

[1] F. Yu, H. Chen, X. Wang, W. Xian, Y. Chen, F. Liu, V. Madhavan, and T. Darrell, "Bdd100k: A diverse driving dataset for heterogeneous multitask learning," 2020.

[2] X. Huang, X. Cheng, Q. Geng, B. Cao, D. Zhou, P. Wang, Y. Lin, and R. Yang, "The apolloscape dataset for autonomous driving," *arXiv: 1803.06184*, 2018.

[3] M. Cordts, M. Omran, S. Ramos, T. Rehfeld, M. Enzweiler, R. Benenson, U. Franke, S. Roth, and B. Schiele, "The cityscapes dataset for semantic urban scene understanding," 2016.

[4] P. Jiang, P. Osteen, M. Wigness, and S. Saripalli, "Rellis-3d dataset: Data, benchmarks and analysis," 2020.

[5] M. Wigness, S. Eum, J. G. Rogers, D. Han, and H. 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)*, 2019, pp. 5000–5007.

[6] A. Valada, G. Oliveira, T. Brox, and W. Burgard, "Deep multispectral semantic scene understanding of forested environments using multimodal fusion," in *International Symposium on Experimental Robotics (ISER)*, 2016.

[7] S. Sharma, J. E. Ball, B. Tang, D. W. Carruth, M. Doude, and M. A. Islam, "Semantic segmentation with transfer learning for off-road autonomous driving," *Sensors*, vol. 19, no. 11, p. 2577, 2019.

[8] A. V. Nefian and G. R. Bradski, "Detection of drivable corridors for off-road autonomous navigation," in *2006 International Conference on Image Processing*, 2006, pp. 3025–3028.

[9] M.-K. Lee, M. R. Golzarian, and I. Kim, "A new color index for vegetation segmentation and classification," *Precision Agriculture*, vol. 22, no. 1, pp. 179–204, 2021.

[10] Y. Ding, Y. Zhao, X. Shen, M. Musuvathi, and T. Mytkowicz, "Yinyang k-means: A drop-in replacement of the classic k-means with consistent speedup," in *International conference on machine learning*. PMLR, 2015, pp. 579–587.

[11] N. A. A. Khairudin, A. S. A. Nasir, L. C. Chin, H. Jaafar, and Z. Mohamed, "A fast and efficient segmentation of soil-transmitted helminths through various color models and k-means clustering," in *Proceedings of the 11th National Technical Seminar on Unmanned System Technology 2019*. Springer, 2021, pp. 555–576.

[12] S. A. Naji, R. Zainuddin, and H. A. Jalab, "Skin segmentation based on multi pixel color clustering models," *Digital Signal Processing*, vol. 22, no. 6, pp. 933–940, 2012.

[13] "Jetson agx xavier," https://developer.nvidia.com/embedded/jetson-agx-xavier-developer-kit/.

[14] A. Garcia-Garcia, S. Orts-Escolano, S. Oprea, V. Villena-Martinez, and J. Garcia-Rodriguez, "A review on deep learning techniques applied to semantic segmentation," 2017.

[15] J. Wang, K. Sun, T. Cheng, B. Jiang, C. Deng, Y. Zhao, D. Liu, Y. Mu, M. Tan, X. Wang, W. Liu, and B. Xiao, "Deep high-resolution representation learning for visual recognition," *TPAMI*, 2019.

[16] C. Yu, C. Gao, J. Wang, G. Yu, C. Shen, and N. Sang, "Bisenet v2: Bilateral network with guided aggregation for real-time semantic segmentation," 2020.

[17] Y. Yuan, X. Chen, and J. Wang, "Object-contextual representations for semantic segmentation," 2020.

[18] M. Everingham, S. A. Eslami, L. Van Gool, C. K. Williams, J. Winn, and A. Zisserman, "The pascal visual object classes challenge: A retrospective," *International journal of computer vision*, vol. 111, no. 1, pp. 98–136, 2015.

[19] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "Mobilenetv2: Inverted residuals and linear bottlenecks," 2019.

[20] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A Large-Scale Hierarchical Image Database," in *CVPR 2009*.

[21] "Dji mavic mini," https://www.dji.com/mavic-mini.

[22] "Labelbox," https://labelbox.com/.

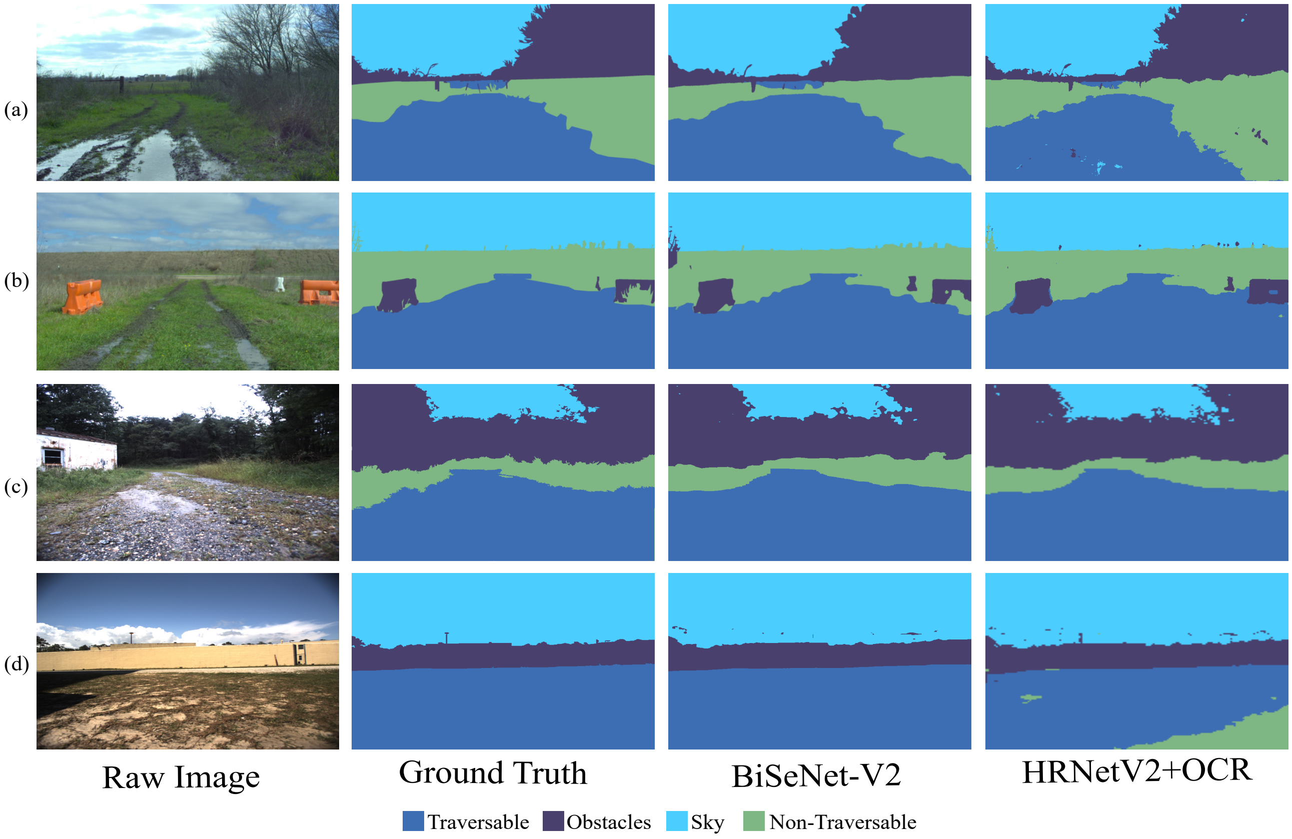


Fig. 8. Segmentation results from BiseNetV2 and HRNETV2+OCR on RELLIS-3D (a,b) and RUGD (c,d) datasets

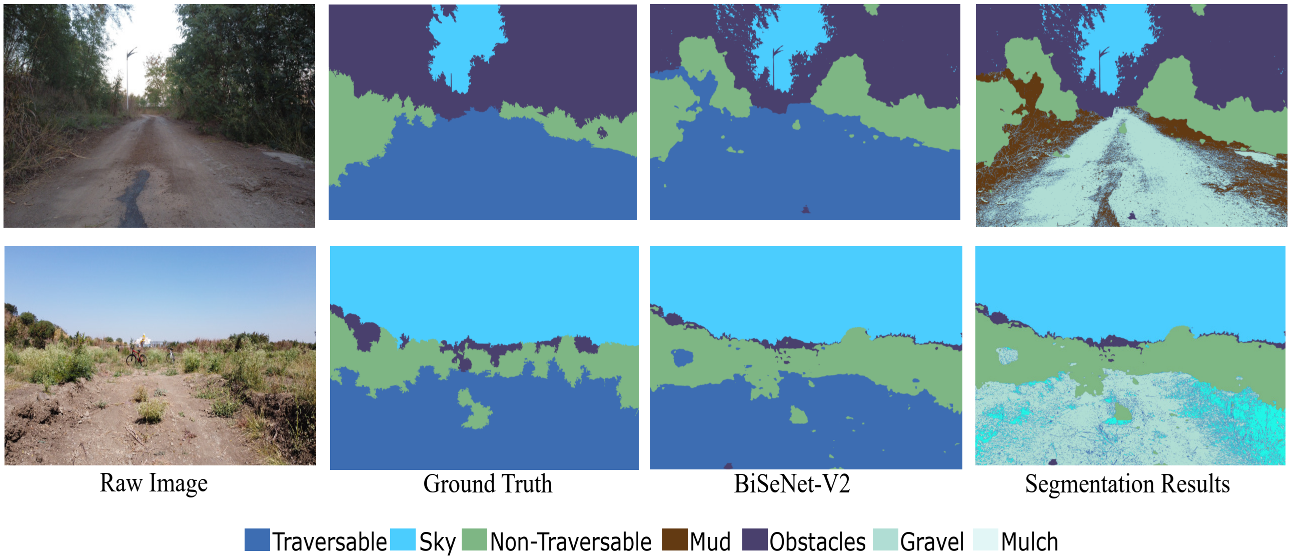


Fig. 9. Segmentation results on IISERB campus frames