TransGSnet: Transformer-embedded Ground Segmentation of Point Cloud for Rough Roads

Dong He and Jong-Hwan Kim
School of Electrical Engineering, Korea Advanced Institute of Science and Technology (KAIST), Daejeon, South Korea

ABSTRACT

Despite comprehensive works on point cloud ground segmentation for flat roads, research on rough roads has rarely been conducted due to datasets' scarcity. To study point cloud ground segmentation on rough roads, in this paper, we provide a synthetic geometric transformation of flat roads motivated by the investigation of real-world rough roads. Our proposed TransGSnet framework consists of two modules: the pillar feature extractor, which turns a raw point cloud into the pseudo image as an intermediate representation, and the transformer-based segmentation network to perform ground segmentation. Specifically, our segmentation network exploits the U-Net architecture and includes three sub-modules: Transformer, mobile block (MB), and convolutional block attention module (CBAM). We thoroughly evaluate our framework in experiments, including several comparisons to the state-of-the-art approaches on semanticKITTI and the synthetic rough road dataset, respectively. As a result, our framework shows an outstanding balance of performance-cost.

Keywords: Semantic Segmentation, Transformers, Point Cloud, LiDAR.

1. INTRODUCTION

Ground segmentation plays a critical role in environment sensing. This knowledge is also helpful in future state prediction, path planning, and collision avoidance in autonomous driving. Researchers have employed diversiform sensors to boost autonomous driving in various driving conditions, particularly the high-definition LiDAR. It can accurately detect the distance to the object in different lighting conditions, thus ensuring the reliability of autonomous driving [1]. Consequently, an increasing number of point cloud datasets have been collected or synthesized for public research. In the Society of Automotive Engineers (SAE), autonomous driving levels 4 and 5 emphasize the adaptive capability for rough road conditions. However, neither real-world nor synthetic point cloud datasets of rough roads are provided because of the troublesome setup. Thus, point cloud ground segmentation has rarely been studied for rough roads so far.

Most approaches of point cloud ground segmentation consider roads to be presumably flat. With this presumption, some previous studies rely on straightforward geometrical clues of ground to determine ground points [2]. Nonetheless, rough road conditions are enormously different from their presumption setup, where most priors fail to handle such conditions. To study point cloud ground segmentation on rough road conditions, we synthesize a rough road dataset based on semanticKITTI [3]. In the international roughness index (IRI), the standard scale for evaluating the roughness of roads, rough roads are often found undulated and uneven [4]. Therefore, we perform jitter, rotation, and translation operations on semanticKITTI to obtain a synthetic rough road dataset. The new synthetic dataset keeps the same data format as semanticKITTI and is available to train or evaluate other ground segmentation approaches.

In the preceding works of point cloud ground segmentation, discretization-based approaches report promising results with less computational cost than other approaches. Following the pipeline of discretization-based approach, we design the TransGSnet framework, which is composed of two major modules: the pillar feature extractor and the transformer-based segmentation network. The pillar feature extractor for representing point clouds to pseudo images is like implementing PointPillars [5]. However, to boost the pillar feature extraction ability, instead of only applying the simplified PointNet [6], a convolutional layer is added to learn the spatial feature of neighbor points within a pillar.

We propose a performance–cost balanced segmentation module that adapts a U-Net [7] architecture with sub-modules. Recently, Transformers emerged with a powerful capacity to extract global features while demanding a high computational cost. For instance, Chen et al. [8] first developed a u-shaped architecture network containing the Transformer layer for medical image segmentation, where their five-layers model demands 98.6 million parameters. Nevertheless, our segmentation module adopts distinct sub-modules for specific layers against their implementation that uses CNNs for all

layers. Although our network is mainly developed for point cloud ground segmentation, it is not confined to this field only and can be adopted in semantic segmentation tasks. The TransGSnet overview is shown in Figure 1.

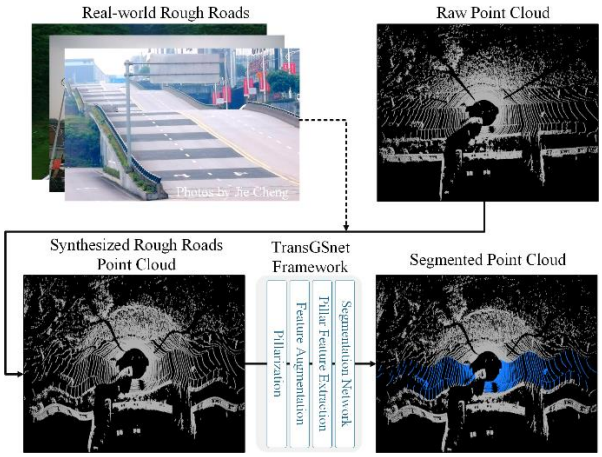

Figure 1. TransGSnet schematic overview. From the investigation of real-world rough roads, we synthesize the rough roads dataset. Input is a synthesized point cloud of the rough road. Output is the classification map with predicted labels. Our TransGSnet framework processes the raw point cloud and estimates the label for each pillar.

Our major contributions are threefold: (1) We synthesize the rough road dataset based on semanticKITTI. We next evaluate geometry-based, and learning-based ground segmentation approaches on the dataset. (2) We propose TransGSnet framework for point cloud ground segmentation. This framework comprises the pillar feature extractor to gain the 2D representation from a 3D point cloud and the transformer-based segmentation network to perform classification. (3) We ablate the Transformer component in segmentation network and normal vector features to verify the improvement.

2. RELATED WORK

2.1 Combination of Transformer and U-Net

U-Net, involving the encoder-decoder bilateral architecture, has gained enormous success in medical image segmentation tasks [9], [10], [11]. It receives images including a few categories and yields the classification maps end-to-end. The Transformer was first developed for sequence modeling as an attention-based block in the natural language process (NLP) [12]. It is remarkable for gaining the long-range dependencies using the attention mechanism, which accesses each item of the entire sequence and reassigns weights of features by aggregating them globally. Furthermore, the vision Transformers (ViT) make it comparable to CNN-based approaches for computer vision tasks [13]. Chen et al. [8] introduced a network combining the Transformer and U-Net for medical image segmentation by replacing the bottom path of the U-Net with the Transformer, yet their network did not consider the computational cost because of their use case. Our segmentation network is based on their work, but the efficiency is additionally considered in our design.

2.2 Efficient Computational Design for U-Net

CNN needs to learn inter-channel and spatial correlations, but Szegedy et al. [14] first argued that the CNN operation could be implemented effectively by separating the channel and spatial correlations. Then, Howard et al. [15] employed MB to reduce parameter size and inference cost. Gadosey et al. [16] introduced a U-Net with MBs and claimed comparable performance with eight times fewer parameters against the original U-Net implementation. Rather than directly using the skip-connection, Bello et al. [17] added an attention module to augment the global representation before each skip-connection. Trebing et al. [18] adopted the above design methodologies, achieving comparable performance to the original U-Net in a weather forecast with fewer parameters.

2.3 Ground Segmentation for Cloud Points

Geometry-based approaches: This approach determines geometrical features such as planes, heights, and lines, and then predict the ground points using (1) plane-fitting model (e.g., RANSAC [19], GPR [2]), (2) elevation-based filter (e.g., elevation maps [20]), or (3) random fields (e.g., MRF [21], CRF [22]). These approaches work well in simple flat road conditions, but they encounter the under-segmentation issue when tested on rough roads under actual driving conditions [1]. Thus, the roughness of roads is normally overlooked in these approaches.

Learning-based approaches: This approach comprises two subcategories: point-based and discretization-based. The former can directly handle raw point clouds. PointNet [6] applied MLPs on points and used a permutation invariant operation to aggregate features of all points. PointNet++ [23] leveraged the local spatial information to enhance the relationship of neighbor points. These approaches are costly because of the point cloud's inherent sparsity. Therefore, the latter appeared as alternatives with higher efficiency by representing a point cloud to the ordered discrete format. VoxelNet [24] discretized a point cloud into the 3D voxel map and then performed the expensive 3D CNN to segment objects. Whereas, PointPillars [5] performed 2D CNN operation after pillarization, yet confirmed a competitive performance with less computational cost. Inspired by them, following discretization-based approaches, we adopt the pillarization process and design the pillar feature extractor to extract features of pillars, and then we perform our proposed segmentation network for ground segmentation.

3. THE SYNTHETIC ROUGH ROAD DATASET

This section investigates real-world rough roads and then defines operations of generating a rough road dataset. According to IRI, the index value higher than 8.0mm/m signifies the rough road condition that adversely affects passengers safety, driving efficiency, and riding quality. The IRI value is calculated using sprung and surface parameters that are highly associated with the longitudinal profile (Z-axis) of the test road [4]. Meanwhile, semanticKITTI dataset was recorded mainly on flat roads. Based on background knowledge, we modulate the Z-axis of point clouds in semanticKITTI dataset to simulate rough roads unevenness. We first execute the jitter and rotation operation to reproduce the uneven ground surface, and translation is also applied to augment the complexity of roads. Jitter, rotation, and translation operations are respectively defined as

$$z_j = z + A\sin\left(B * \sqrt{x^2 + y^2}\right) + C,$$
$$z_r(\theta_z) = \begin{bmatrix} \cos\theta_z & -\sin\theta_z & 0 \\ \sin\theta_z & \cos\theta_z & 0 \\ 0 & 0 & 1 \end{bmatrix},$$
$$x_t = x + offset_x, \quad and$$
$$y_t = y + offest_y,$$

where z_j is the jitter on Z-axis, z_r indicates the rotation of the Z-axis, x_t and y_t are the translation on the x, y coordinates, respectively. Besides, labels from semanticKITTI are affixed to corresponding points. Eventually, the resulting dataset keeps the same data format as semanticKITTI.

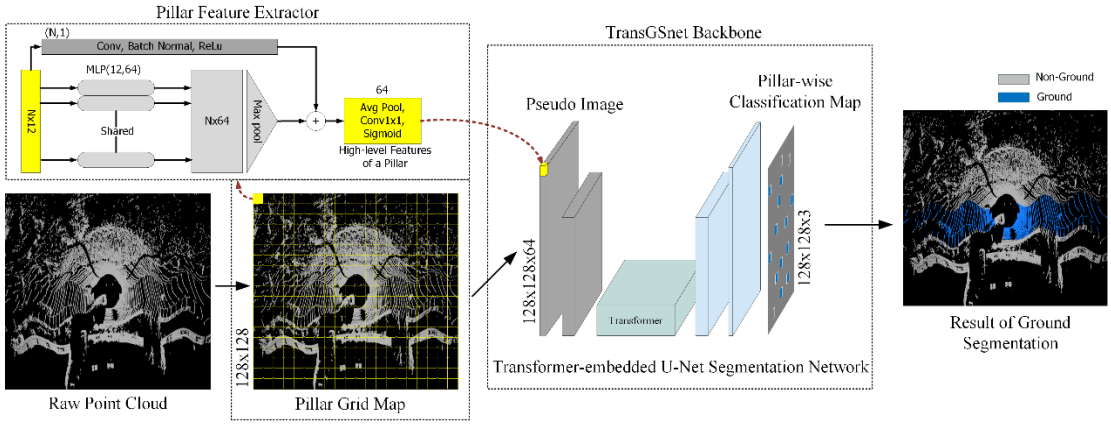

Figure 2. The architecture of TransGSnet framework: 1) 2D representation (pillar feature extractor): 3D point cloud is discretized into a 2D 128x128 pillar grid map. Points with 12-dimensional features (original features of points, pillars coordinates, and normal vectors of pillar) in each pillar go through the pillar feature extractor to be composed the 2D pseudo image. 2) Ground segmentation: Pseudo image goes through a Transformer-embedded segmentation network to estimate the label of each pillar.

4. PROPOSED TRANSGSNET FRAMEWORK

TransGSnet framework involves two essential modules as shown in Figure 2: (1) pillar feature extractor, which converts a scan of 3D points to the 2D pseudo image $R^3 \rightarrow R^2$, and (2) segmentation network, designed as a U-Net shape with three

sub-modules. We describe each stage of the framework in this section. Especially for the segmentation module, we delineate the effective design of the main network and specify the Transformer at the bottom path.

4.1 Pillar Feature Extractor

This paper first discretizes a point cloud into an even pillar grid map by following PointPillars. For more details, the pillar grid map is of size $(128, 128)$ on the x, y coordinates; the size of each pillar is $[0.8, 0.8, 8]$ meters; the size of a point cloud is $([-51.2, 51.2], [-51.2, 51.2], [-5,3])$ meters on x, y, z coordinates and outlines are eliminated. According to the observation of real-world rough roads, we realize that the diversity of normal vectors of rough roads potentially influences classification performance. Therefore, we adopt the surface normal vector algorithm from Open3D [25] to calculate the normal vector for each pillar grid. Moreover, if the point number in a pillar grid is less than three, we search for additional points in neighborhood pillars to deduce the normal vector. We then append normal vector of each pillar grid as features. Eventually, the original features of points are augmented to 12-dimensional features.

We use the pillar feature extractor to generate the pseudo image from the features of pillars. This module extracts two high-level features by using simplified PointNet and CNN separately. Simplified PointNet comprises shared linear layers followed by the batch norm and ReLu, which generates a tensor sized of (C, P, N), where C denotes the feature number, P the pillar number, and N the number of points in a pillar. Then, the max-pooling operation aggregates of features of points in a pillar. The resulting output is of size (C, P). Moreover, we introduce a convolutional layer of kernel $1 \times N$ with the batch norm and ReLu to learn the potential spatial features of neighbor points in a pillar. Two kinds of high-level features are fused by a 1×1 convolutional layer. Finally, the pillar features are attached to their corresponding locations to gain a pseudo image of size (H, W, C), where H and W indicate the height and width of the pseudo image. 3D-MiniNet [26] implemented a partially similar module for the spherical projection space but demanded more parameters.

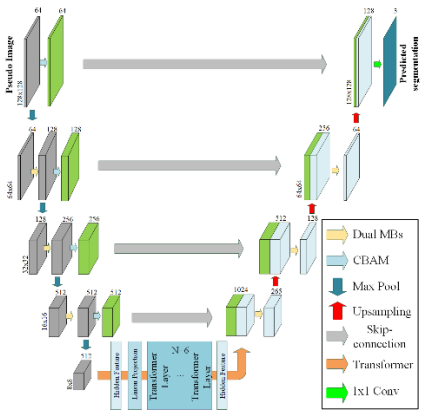

Figure 3. The architecture of the proposed segmentation network. Input is a pseudo image, and output is the predicted segmentation. Each cube represents a multi-channel feature map. The Numbers on the left side are the size of feature maps. The above numbers are the channels of feature maps.

4.2 Segmentation Network

Since upper layers, including high-resolution feature maps, can extract features without the high learning capacity, we exploit MBs in the upper layers to reduce parameters. In contrast, the bottom layer requires more complex and expensive operations to extract high-level features, where we employ Transformers for the low-resolution feature map. Moreover, we place CBAM before each skip-connection to boost performance. As shown in Figure 3, the encoder consists of four components: Dual MBs, CBAMs, maxpooling, and skip-connections. Dual MBs extract features in upper layers. Then, different-scale feature maps go through CBAMs to be reassigned weights from the inherent relationship of pillars. The CBAMs enhanced feature maps are to concatenate corresponding feature maps on the decoder side via skip-connections. The Transformer is located at the bottom path. At the decoder side, bilinear upsampling operations increase the feature map twofold size. Doubled feature maps are concatenated with the corresponding self-attention features maps via skip-connections from the encoder side. Eventually, a 1×1 convolution as the output layer generates a single classification map of predicted labels.

5. EXPERIMENTS

We conducted experiments on semanticKITTI benchmark dataset and our synthesized rough road dataset, respectively. Besides, selected state-of-the-art approaches of geometry-based and learning-based were involved compared to our proposed approach. Further, we provided an ablation study on the Transformer sub-module in segmentation network and normal vector features.

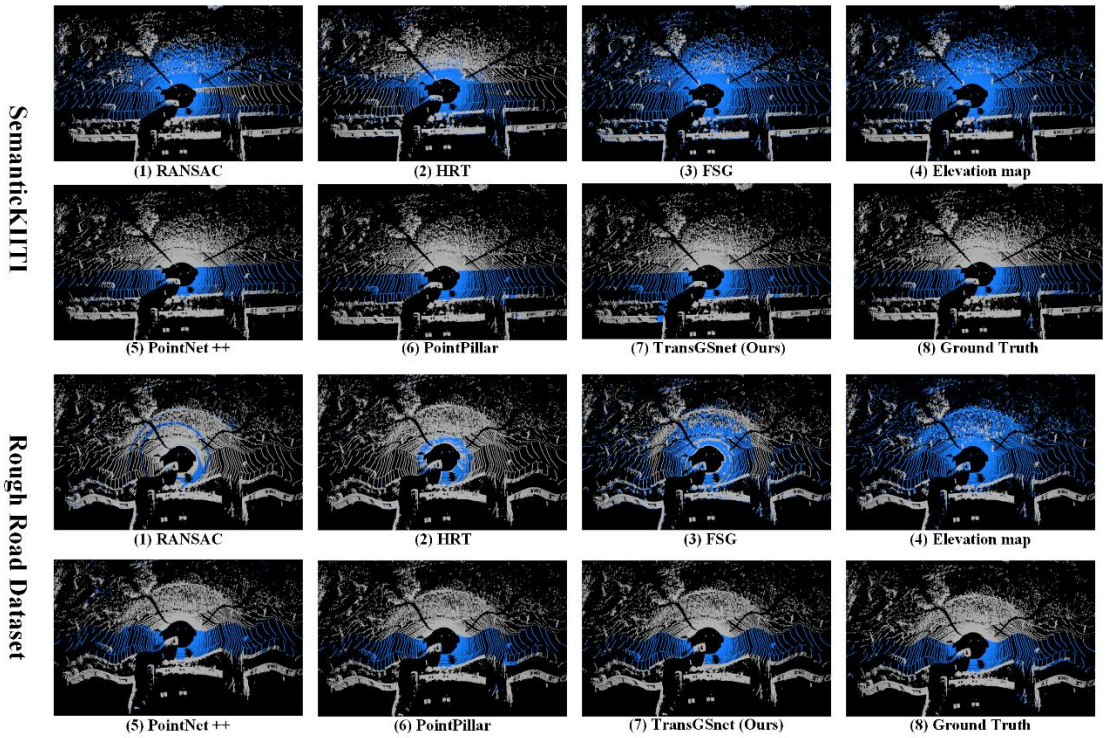

Figure 4. Visualized results on the test set are shown as point clouds, where blue points are the ground, gray points are the non-ground. (Visualized results may vary depending on different frames.)

5.1 Dataset

SemanticKITTI dataset, widely used in 3D object detection tasks, provides 28 annotated classes for 23,190 frames of sequences 00-10 in KITTI dataset. The synthetic dataset keeps the same data format as semanticKITTI and is available to train or evaluate other ground segmentation approaches. So, we could use the same points cataloged by road, sidewalk, parking, and other ground in both datasets to generate the ground truth of ground pillars. In both semanticKITTI dataset and the rough road dataset, sequences 00-07 and 09-10 were utilized for training, and sequence 08 for testing.

5.2 Experiments on semanticKITTI

We compared our TransGSnet against both geometry-based approaches [2], [19], [20], [27] and learning-based approaches [5], [23], [28]. The visualized ground segmentation results on the No.194 frame in sequence No.8 are shown in Fig. 3. The overall performance of learning-based approaches surpassed that of geometry-based approaches on this frame. We then evaluated the ground segmentation performance with the evaluation metrics, as shown in Table 1. In learning-based approaches, although PointNet++, the point-based approach, achieves the best performance in terms of accuracy, mIoU, F-1 score, and FPR, its runtime is too low to be employed in practice. Among the rest discretization-based approaches, our proposed TransGSnet shows the competitive performance.

5.3Experiments on rough road dataset

We observed the significant performance gap between two groups of approaches on the rough road dataset, as shown in Figure 3. For geometry-based approaches, the ground modeling approaches [2], [19], [27] showed poor performance largely owing to undulated roads, and the elevation approaches [20] also perform poorly due to the road slope. Nevertheless, learning-based approaches maintain good performance. PointNet++ achieves the best performance but with a high

computational cost and runtime. GndNet is absent in comparison to the rough road because it relies on their customized ground truth, which we can not convert to rough road one. Our TransGSnet obtains a better evaluation on segmentation than PointPillars and shows a performance-cost balance guaranteeing the applicability in practice.

TABLE 1 . COMPARISON OF TRANSGSNET WITH STATE-OF-THE-ART APPROACHES ON THE TEST SET OF SEMANTICKITTI/ ROUGH ROAD DATASET

Dataset	Approach	Accuracy	mIoU	F-1 Score	Number of Parameters	Runtime
SemanticKITTI	Elevation Map [20]	0.841	0.711	0.781	-	-
	RANSAC [19]	0.809	0.663	0.681	-	-
	HRT [2]	0.839	0.699	0.77	-	-
	FSG [27]	0.869	0.759	0.825	-	-
	PointNet++ [23]	**0.954**	**0.898**	**0.965**	1.47M	1.7fps
	PointPillars [5]	0.875	0.748	0.803	**0.19M**	143.8fps
	GndNet [8]	0.853	0.738	0.806	-	**174.2fps**
	TransGSnet (Ours)	0.876	0.748	0.805	15.03M	78.7fps
Rough Road Dataset	Elevation Map	0.786	0.614	0.667	-	-
	RANSAC	0.653	0.377	0.207	-	-
	HRT	0.757	0.519	0.481	-	-
	FSG	0.743	0.563	0.627	-	-
	PointNet++	**0.877**	**0.740**	**0.909**	1.47M	1.8fps
	PointPillars	0.857	0.719	0.778	**0.19M**	**151.5fps**
	TransGSnet (Ours)	0.869	0.732	0.79	15.03M	67.9fps

5.4 Ablation study

Transformer sub-module is the key component in our segmentation network. To study its effect on performance improvement, we compared it with standard U-Net and attention U-Net. Comparison results are reported in Table 2. Because the Transformer builds underlying relations in global feature representations, we could see an improvement in segmentation performance.

TABLE 2 . COMPARISON OF TRANSGSNET WITH STATE-OF-THE-ART APPROACHES ON THE TEST SET OF SEMANTICKITTI/ ROUGH ROAD DATASET

Dataset	Segmentation backbone	Accuracy	mIoU	F-1 Score	Number of Parameters	Runtime
semanticKITTI	UNet [7]	0.872	0.741	0.796	34.56M	**157.3fps**
	Attention UNet [18]	0.853	0.738	0.806	34.91M	131.1fps
	TransGSnet (Ours)	**0.876**	**0.748**	**0.805**	**15.03M**	78.7fps
Rough Road Dataset	UNet	0.862	0.729	0.789	34.56M	**137.8fps**
	Attention UNet	0.857	0.719	0.778	34.91M	113.2fps
	TransGSnet (Ours)	**0.869**	**0.732**	**0.790**	**15.03M**	67.9fps

Normal vector features were emphasized in the previous section. We evaluated our framework with and without normal vector features on the rough road dataset. Table 3 clearly illustrates the normal vector features as additional inputs improve performance for rough roads.

TABLE 3. RESULTS OF WITHOUT OR WITH NORMAL FEATURES ON THE TEST SET OF THE ROUGH ROAD DATASET

Augmentation	Accuracy	mIoU	F-1 Score	TPR	FPR
w/o Normals	0.855	0.713	0.768	0.787	0.871
w/ Normals	**0.869**	**0.736**	**0.790**	**0.812**	**0.879**

6. CONCLUSION

This paper focuses on ground segmentation on rough roads that are more challenging than flat roads. Our proposed TransGSnet framework overcomes the road changes in height and slope and achieves an outstanding balance of performance-cost on rough roads. Besides, experiments show that the Transformer sub-module enhanced the encoder representation ability, resulting in competitive performance. However, the Transformer we adopt needs a proper pre-trained model, and its parameter size is much larger than that of CNN. Recent works involve matrix multiplications, hierarchical feature maps, or shift windows to implement the Transformer efficiently. The segmentation performance can be enhanced by replacing the standard transformer model with these improved transformers for ground segmentation. In addition, another 2D partition strategy can be adapted to avoid numerous null pillars in future work.

Acknowledgment: This work was supported by Institute for Information & communications Technology Promotion (IITP) grantfunded by the Korea government (MSIT) (No.2020-0-00440, Development of artificial intelligence technologythat continuously improves itself as the situation changes in the real world). We thank JieCheng for sharing partial comparison data.

REFERENCES

[1] Guo, Y., Wang, H., Hu, Q., Liu, H., Liu, L., and Bennamoun, M., "Deep learning for 3d point clouds: Asurvey," IEEE transactions on pattern analysis and machine intelligence(2020).

[2] Liu, K., Wang, W., Tharmarasa, R., Wang, J., and Zuo, Y., "Ground surface filtering of 3d point clouds based on hybrid regression technique," IEEE Access **7**, 23270–23284 (2019).

[3] Behley, J., Garbade, M., Milioto, A., Quenzel, J., Behnke, S., Stachniss, C., and Gall, J., "Semantickitti: A dataset for semantic scene understanding of lidar sequences," in [Proceedings of the IEEE/CVF InternationalConference on Computer Vision], 9297–9307 (2019).

[4] Du, Y., Liu, C., Wu, D., and Jiang, S., "Measurement of international roughness index by using-axis accelerometers and gps," Mathematical Problems in Engineering (2014).

[5] Lang, A. H., Vora, S., Caesar, H., Zhou, L., Yang, J., and Beijbom, O., "Pointpillars: Fast encoders for object detection from point clouds," in [Proceedings of the IEEE/CVF Conference on Computer Vision andPattern Recognition], 12697–12705 (2019).

[6] Qi, C. R., Su, H., Mo, K., and Guibas, L. J., "Pointnet: Deep learning on point sets for 3d classification and segmentation," in [Proceedings of the IEEE conference on computer vision and pattern recognition], 652–660 (2017).

[7] Ronneberger, O., Fischer, P., and Brox, T., "U-net: Convolutional networks for biomedical image segmentation," MICCAI, 234–241, Springer (2015).

[8] Chen, J., Lu, Y., Yu, Q., Luo, X., Adeli, E., Wang, Y., Lu, L., Yuille, A. L., and Zhou, Y., "Transunet:Transformers make strong encoders for medical image segmentation," arXiv preprint :2102.04306(2021).

[9] Otsuki, K., Iwamoto, Y., Chen, Y.-W., Furukawa, A., and Kanasaki, S., "Cine-mr image segmentation for assessment of small bowel motility function using 3d u-net," Journal of Image and Graphics **7** (4), 134–139(2019).

[10] Wu, J., Li, G., Lu, H., and Kamiya, T., "A supervoxel classification based method for multi-organ segmentation from abdominal ct images," Journal of Image and Graphics **9** (1) (2021).

[11] Enokiya, Y., Iwamoto, Y., Chen, Y.-W., and Han, X.-H., "Automatic liver segmentation using u-net with wasserstein gans," Journal of Image and Graphics **7**, 94–101 (2018).

[12] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., and Polosukhin,I., "Attention is all you need," in [Advances in neural information processing systems], 5998–6008 (2017).

[13] Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M.,Minderer, M., Heigold, G., Gelly, S., et al., "An image is worth 16x16 words: Transformers for image recognition at scale," arXiv preprint :2010.11929(2020).

[14] Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., and Wojna, Z., "Rethinking the inception architecturefor computer vision," ," in [Proceedings of the IEEE conference on computer vision and pattern recognition], 2818–2826 (2016).

[15] Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., and Adam,H., "Mobilenets: Efficient convolutionalneural networks for mobile vision applications," arXiv preprint :1704.04861(2017).

[16] Gadosey, P. K., Li, Y., Agyekum, E. A., Zhang, T., Liu, Z., Yamak, P. T., and Essaf, F., "Sd-unet:Stripping down u-net for segmentation of biomedical images on platforms with low computational budgets," Diagnostics **10** (2), 110 (2020).

[17] Bello, I., Zoph, B., Vaswani, A., Shlens, J., and Le, Q. V., "Attention augmented convolutional networks," in [Proceedings of the IEEE/CVF international conference on computer vision], 3286–3295 (2019).

[18] Trebing, K., Stanczyk, T., and Mehrkanoon, S., "Smaat-unet: Precipitation nowcasting using a small attention-unet architecture," Pattern Recognition Letters (2021).

[19] Li, L., Yang, F., Zhu, H., Li, D., Li, Y., and Tang, L., "An improved ransac for 3d point cloud plane segmentation based on normal distribution transformation cells," Remote Sensing **9** (5), 433 (2017).

[20] Thrun, S., Montemerlo, M., Dahlkamp, H., Stavens, D., Aron, A., Diebel, J., Fong, P., Gale, J., Halpenny,M., Hoffmann, G., et al., "Stanley: The robot that won the darpa grand challenge,"Journal of fieldRobotics **23** (9), 661–692 (2006).

[21] Byun, J., Na, K.-i., Seo, B.-s., and Roh, M., "Drivable road detection with 3d point clouds based on the mrf for intelligent vehicle," in [Field and Service Robotics], 49–60, Springer (2015).

[22] Rummelhard, L., Paigwar, A., N`egre, A., and Laugier, C., "Ground estimation and point cloud segmentationusing spatiotemporal conditional random field," in [2017 IEEE Intelligent Vehicles Symposium (IV)], 1105–1110, IEEE (2017).

[23] Qi, C. R., Yi, L., Su, H., and Guibas, L. J., "Pointnet++: Deep hierarchical feature learning on point setsin a metric space,"arXiv preprint arXiv:1706.02413(2017).

[24] Zhou, Y. and Tuzel, O., "Voxelnet: End-to-end learning for point cloud based 3d object detection," in [Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition], 4490–4499 (2018).

[25] Zhou, Q.-Y., Park, J., and Koltun, V., "Open3d: A modern library for 3d data processing,"arXiv preprint:1801.09847 (2018).

[26] Alonso, I., Riazuelo, L., Montesano, L., and Murillo, A. C., "3d-mininet: Learning a 2d representationfrom point clouds for fast and efficient 3d lidar semantic segmentation,"IEEE Robotics and AutomationLetters5(4), 5432–5439 (2020).

[27] Himmelsbach, M., Hundelshausen, F. V., and Wuensche, H.-J., "Fast segmentation of 3d point clouds forground vehicles," in [2010 IEEE Intelligent Vehicles Symposium], 560–565, IEEE (2010).

[28] Paigwar, A., Erkent, Ö., Gonz´alez, D. S., and Laugier, C., "Gndnet: Fast ground plane estimation andpoint cloud segmentation for autonomous vehicles," in [IEEE/RSJ International Conference on IntelligentRobots and Systems (IROS)], (2020).