Road type classification of MLS point clouds using deep learning

Bai, Q.; Lindenbergh, R. C.; Vijverberg, J.; Guelen, J. A.P.

DOI
10.5194/isprs-archives-XLIII-B2-2021-115-2021

Publication date
2021

Document Version
Final published version

Published in
International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives

ROAD TYPE CLASSIFICATION OF MLS POINT CLOUDS USING DEEP LEARNING

Q. Bai[1, 2*], R. C. Lindenbergh[1], J. Vijverberg[2], J. A. P. Guelen[2]

[1] Dept. of Geoscience and Remote Sensing, Delft University of Technology, The Netherlands
- (q.bai@student.tudelft.nl, r.c.lindenbergh@tudelft.nl)
[2] Cyclomedia Technology - (QBai, JVijverberg, JGuelen)@cyclomedia.com

Commission II, WG II/3

KEY WORDS: Mobile mapping, point clouds, semantic segmentation, road type, local feature aggregation, deep learning.

ABSTRACT:

Functional classification of the road is important to the construction of sustainable transport systems and proper design of facilities. Mobile laser scanning (MLS) point clouds provide accurate and dense 3D measurements of road scenes, while their massive data volume and lack of structure also bring difficulties in processing. 3D point cloud understanding through deep neural networks achieves breakthroughs since PointNet and arouses wide attention in recent years. In this paper, we study the automatic road type classification of MLS point clouds by employing a point-wise neural network, RandLA-Net, which is designed for consuming large-scale point clouds. An effective local feature aggregation (LFA) module in RandLA-Net preserves the local geometry in point clouds by formulating an enhanced geometric feature vector and learning different point weights in a local neighborhood. Based on this method, we also investigate possible feature combinations to calculate neighboring weights. We train on a colorized point cloud from the city of Hannover, Germany, and classify road points into 7 classes that reveal detailed functions, i.e., *sidewalk*, *cycling path*, *rail track*, *parking area*, *motorway*, *green area*, and *island without traffic*. Also, three feature combinations inside the LFA module are examined, including the geometric feature vector only, the geometric feature vector combined with additional features (e.g., color), and the geometric feature vector combined with local differences of additional features. We achieve the best overall accuracy (86.23%) and mean IoU (69.41%) by adopting the second and third combinations respectively, with additional features including Red, Green, Blue, and intensity. The evaluation results demonstrate the effectiveness of our method, but we also observe that different road types benefit the most from different feature settings.

1. INTRODUCTION

Automation of road information extraction is of great significance to economic and social development. Road type, which indicates the function of a road segment, is key to various applications, including autonomous driving, inspections of infrastructures, and decision making of companies and governments (Zhu et al., 2012). LiDAR provides accurate 3D point measurements and is illumination invariant, showing a strong ability for mapping. Similar to image recognition, point cloud processing also benefits from the rapid development of deep learning techniques (Liu et al., 2019). However, it is still challenging to interpret 3D point clouds using neural networks due to their irregular data structure.

Determining the type of each road point is consistent with the aim of point cloud semantic segmentation. Recent studies on semantic segmentation of point clouds using deep learning mainly consist of two kinds of methods, i.e., projection-based and point-based methods. In projection-based methods, point clouds are first projected onto 2D planes (i.e., images) (Wu et al., 2018) or converted into voxels (i.e., 3D grids) (Riegler et al., 2017). Through achieving a regularly aligned data format, 2D or 3D convolutional neural networks (CNN) can be applied. Although these methods address the problem of unorganized point clouds indirectly, some spatial information is lost and additional computational resources are needed during pre-processing.

As an active research topic in this area, point-based neural networks can directly consume and model 3D point data. PointNet (Qi et al., 2017a), the pioneering work among these methods, employs a series of shared multi-layer perceptrons (MLP) to learn higher-dimensional features for each point. Then these per-point features are aggregated by applying a symmetric function (e.g., max-pooling), ensuring that point cloud processing is irrelevant to the point order. However, PointNet does not consider local structures inside the point cloud, limiting its performance in complex scenes (Qi et al., 2017b). Starting from PointNet, many networks are proposed combining MLPs with local feature aggregation. The local feature aggregation module aims to extract prominent features from a point neighborhood, thereby exploiting wider contextual information around each point. The choice of input features for local aggregation has a great impact on its effectiveness. PointNet++ uses relative coordinates in a local region together with additional point features (e.g., R, G, B), while DGCNN (Wang et al., 2019) constructs the concatenation of all original and relative features as input. RandLA-Net (Hu et al., 2020), which is adopted in this study, employs more complex encoding for relative coordinates to capture geometric details. The encoded geometric feature, together with additional point features, is then used to achieve local feature aggregation. Different from PointNet++ and DGCNN, which use a symmetric function to aggregate input features indistinguishably, RandLA-Net represents local information as a weighted sum of all neighboring point features, making the choice of input features even more crucial for capturing the local geometry.

Some of the aforementioned methods have already verified their model performance on benchmark MLS point cloud datasets

* Corresponding author

like SemanticKITTI (Behley et al., 2019). The labeling of these datasets covers the whole road scene including road surface, cars, buildings, etc. However, further research on detailed classification focusing on different road types is still needed. Some studies also evaluate different input features of the local aggregation module (Widyaningrum et al., 2021). It turns out that one fixed feature combination is not optimal for all datasets (Liu et al., 2020). To find a proper setting for road type classification, it is important to conduct more experiments.

The main contributions of this paper are:

- We achieve detailed road type classification in dense urban areas by applying RandLA-Net.
- We assess how features should be combined to achieve weights of neighboring points when aggregating local information in point clouds.

This paper is organized as follows: Section 2 illustrates the dataset employed in this study. Section 3 describe the methodology, including the data pre-processing procedure, details of the neural network, and adopted evaluation metrics. Experiment results are discussed in Section 4. Finally, Section 5 presents the drawn conclusions.

2. DATASET AND STUDY AREA

The MLS point cloud used in this paper is acquired by Cyclomedia's proprietary recording system (Cyclomedia, 2021), in the city of Hannover, Germany. This system is mainly composed of 5 high-resolution cameras and a Velodyne HDL-32E LiDAR sensor (Velodyne, 2010). Figure 1 shows the trajectory of the recording vehicle, which is about 16 km in length. The original LiDAR point cloud has an average point spacing of 1 cm and is colorized by panoramic images obtained at the same time.

Ground truth annotations of the MLS point cloud contain 9 road classes: *sidewalk*, *cycling path*, *rail track*, *parking area*, *motorway*, *green area*, *island without traffic*, *pedestrian area* (car-free zones) and *others*. The order of these classes also reveals the priority of labeling. For example, if a motorway is crossing a rail track in a point cloud, corresponding points will be labeled as *rail track*.

3. METHODOLOGY

This study investigates the capability of a deep neural network, i.e., RandLA-Net, for classifying 3D point clouds into different road types and evaluates the performance of different feature combinations in the local feature aggregation module. Our methodology mainly includes data pre-processing, training with RandLA-Net, and evaluation.

3.1 Data pre-processing

To handle the sheer volume of the acquired MLS point cloud, we first downsample it using grid sampling, with a grid size of 0.1 m. Figure 2a presents an example of the colorized point cloud after downsampling. Afterwards, non-road points are removed by a ground filtering approach (Isenburg, 2014), as shown in Figure 2b. The label of each point is then achieved through overlaying the ground truth annotations, which are polygons stored in the shapefile format, on the road point cloud

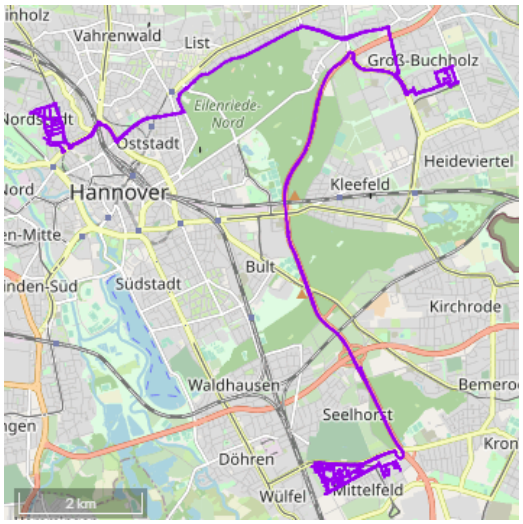


Figure 1. Trajectory of the recording vehicle shown in purple, with a length about 16 km. The base map is provided by © OpenStreetMap.

(see Figure 2d). Also, points belonging to *pedestrian area* and *others* have similar appearance as *sidewalk*. Considering that they are detected with the help of other information like a road sign in practice, *pedestrian area* and *others* are merged into the *sidewalk* to ease the training.[(1)]

After pre-processing, the point cloud dataset used for our study has a total number of 74,629,166 points, with 9 attributes, i.e., x, y, z, R, G, B, intensity, return number, and number of returns. Besides, the distribution of points in 7 road types is illustrated in Table 1, in which a class imbalance issue can be observed. *Motorway* contains a dominant number of points. *sidewalk* and *green area* are also frequently seen in the data. By contrast, *rail track* has the least amount of points. The MLS point cloud after pre-processing is vertically split into 39 tiles, with 29 tiles for training and 10 for testing.

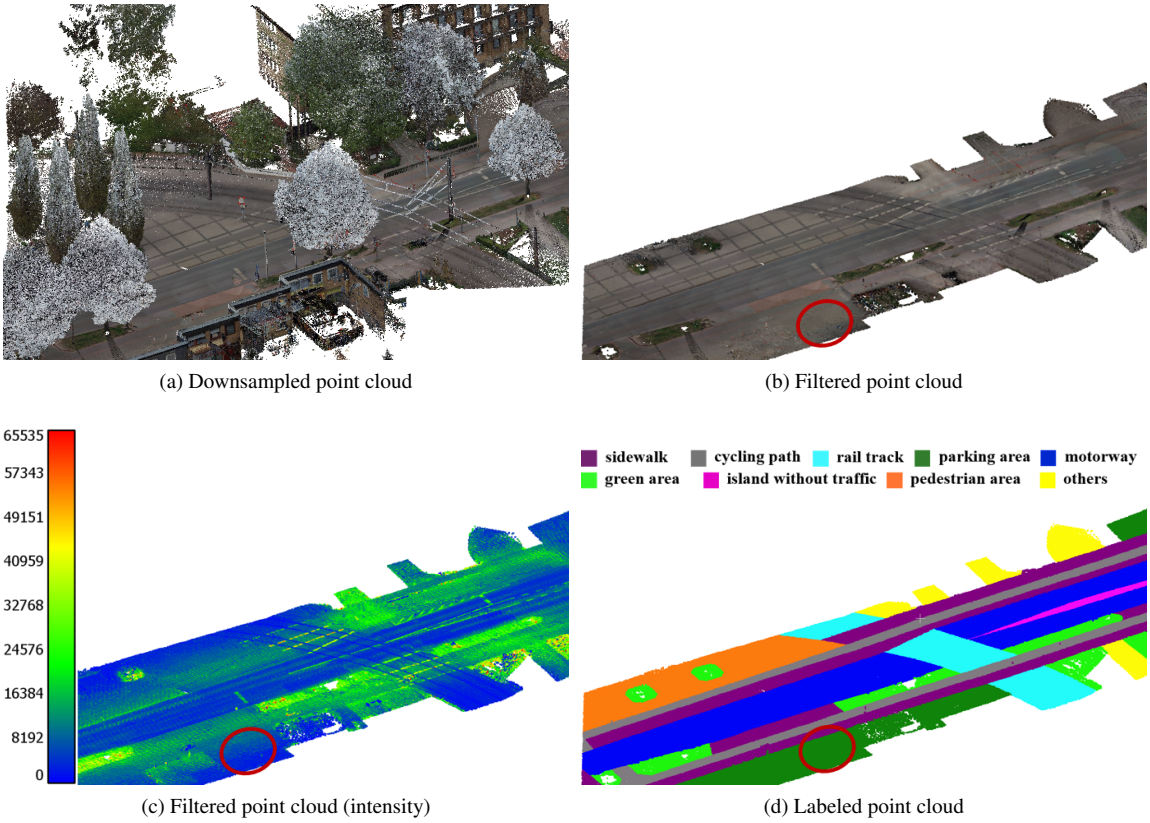
sidewalk	cycling path	rail track	parking area
17.27	3.22	0.54	3.92
motorway	green area	island without traffic	
32.68	14.12	2.88	

Table 1. Number of points ($\cdot 10^6$) in each road type.

3.2 RandLA-Net

We implement RandLA-Net (Hu et al., 2020) for road type classification in this study. RandLA-Net is a point-wise neural network and follows an encoder-decoder hierarchical design (see Figure 3). Given a point cloud with a large number of points, the points are progressively downsampled in each encoding layer and upsampled again in decoding layers to preserve the original resolution in final predictions. To achieve processing efficiency, random sampling is chosen as the downsampling strategy. Since random sampling drops points non-selectively, each neural layer also contains an effective local feature aggregation (LFA) module to summarize neighborhood information without losing important point features. The LFA module

(1) When using (x, y, z, R, G, B) and the original feature combination in the LFA module of RandLA-Net, the mean IoU is improved by 10.97% after merging the labels.



(a) Downsampled point cloud (b) Filtered point cloud

(c) Filtered point cloud (intensity) (d) Labeled point cloud

Figure 2. Downsampled, filtered and labeled point cloud. Circle areas highlight the similarity between some parking areas and sidewalks.

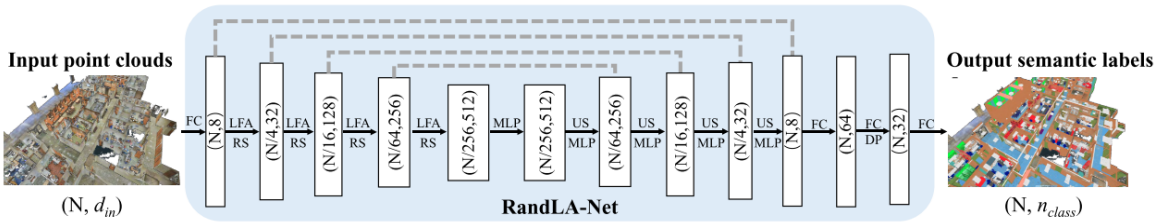


Figure 3. Overview of the RandLA-Net architecture (Hu et al., 2020). N indicates the number of input points. FC: Fully Connected layer, LFA: Local Feature Aggregation, RS: Random Sampling, MLP: shared Multi-Layer Perceptron, US: Up-sampling, DP: Dropout.

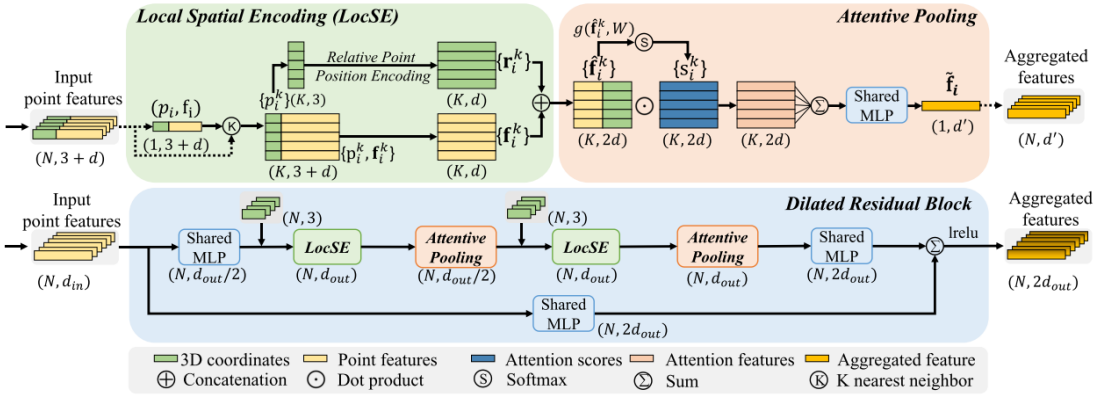


Figure 4. Components of an encoding layer in RandLA-Net (Hu et al., 2020). Top: Local Spatial Encoding (LocSE) block which transforms the input features and Attentive Pooling block which aggregates the local information based on weighing the neighboring points. Bottom: Two pairs of LoSE and Attentive Pooling blocks are stacked together to increase the receptive field, which forms the Dilated Residual Block of each encoding layer.

is the key to modeling and perceiving the local geometry of point clouds. Moreover, the neighborhood around each point is selected using K-Nearest Neighbor (KNN) in RandLA-Net.

As shown in Figure 4 (top), the LFA module consists of two components, i.e., Local Spatial Encoding (LocSE) and Attentive Pooling. Within LocSE, coordinates of the input points are first transformed to a higher dimensional geometric feature vector r_i^k according to:

$$r_i^k = MLP(p_i \oplus p_i^k \oplus (p_i - p_i^k) \oplus \|p_i - p_i^k\|), \quad (1)$$

where MLP = multi-layer perceptrons
$i \in \{1, 2, ..., N\}$
N = the total number of points
$k \in \{1, 2, ..., K\}$
K = the number of nearest neighbors
p_i = coordinates of the centered point
p_i^k = coordinates of one neighboring point
$\oplus$ = concatenation operation
$\|\cdot\|$ = Euclidean distance

The geometric feature vector r_i^k and additional features f_i^k (e.g, R, G, B) are then concatenated as $\hat{f}_i^k$, which is the input of Attentive Pooling.

The aim of Attentive Pooling is to aggregate the enhanced point feature $\hat{f}_i^k$ in the neighborhood to achieve local contextual information for each point. Neural networks like PointNet++ and DGCNN apply a symmetric function (e.g., max-pooling and $\sum$) as the aggregation function, which is simple but inevitably processes the neighboring points indistinguishably, causing a certain loss of geometric information. The Attentive Pooling in RandLA-Net, instead, learns different weights s_i^k of the neighboring points through a MLP, as indicated by $g(\hat{f}_i^k, W)$ in Figure 4 (top). The neighborhood features are subsequently aggregated by taking a weighted sum. Moreover, RandLA-Net applies the LFA module twice in each layer to effectively increase the receptive field of the network, as shown in Figure 4 (bottom).

In this study, we use a colorized MLS point cloud. Apart from the geometric vector r_i^k, how to combine additional features (e.g., R, G, B) to aggregate local information in road scenes remains to be discussed. In urban areas, road objects like *motorway* and *green area* have totally different appearance. Their variation in color can also differ a lot. Additionally, as indicated in Figure 2c, intensity values of the vegetation (*green area*) present distinct characteristics. Thus, it might be beneficial to include these features or even their local differences as additional information sources to help distinguish road types.

However, it may happen that the surface material of two adjacent road segments (e.g., *sidewalk* and *parking area* shown in circled areas of Figure 2) are the same, making the appearance and reflection values of different road objects very similar. In this case, it is possible that using only geometric features can reduce class confusion and acquire more accurate results.

Based on these assumptions, we compare three feature combinations to calculate neighboring weights in the local feature aggregation module of RandLA-Net, which refer to the choice of $\hat{f}_i^k$ in $g(\hat{f}_i^k, W)$:

1. $\mathbf{r_i^k}$: Geometric feature vector only.
2. $\mathbf{r_i^k} \oplus \mathbf{f_i^k}$: Geometric feature vector r_i^k concatenated with additional features f_i^k, which is the original implementation of RandLA-Net.
3. $\mathbf{r_i^k} \oplus (\mathbf{f_i} - \mathbf{f_i^k})$: Geometric feature vector r_i^k concatenated with relative additional features $(f_i - f_i^k)$.

We also consider two settings of the additional features f_i^k, i.e., (R, G, B) and (R, G, B, I), with I indicating the intensity. The intensity feature is a more stable attribute compared to RGB values since it is not affected by illumination conditions during recording.

3.3 Evaluation metrics

To evaluate and compare the performance of different feature combinations illustrated in Section 3.2, we determine the following evaluation metrics in this study, which are commonly used in the semantic segmentation task:

- Overall accuracy (OA), which measures the proportion of correctly classified points among all input points.
- Mean Intersection over Union (mIoU), which is the mean value of Intersection over Union (IoU) in each class, with IoU defined as:

$$\text{IoU} = \frac{\text{Overlap of the predicted and ground truth}}{\text{Union of the predicted and ground truth}}. \quad (2)$$

4. RESULTS AND DISCUSSION

In this section, we first compare the evaluation results for three feature combinations in the local feature aggregation module of RandLA-Net in Section 4.1. Section 4.2 shows the impact of adding intensity features on the overall performance, as well as the results of several specific road types. Finally, we discuss the importance of defining appropriate road classes that represent distinct functions in Section 4.3.

Table 2 and Table 3 summarize the quantitative results of road type classification with different feature combinations in the LFA module.

	$\mathbf{f_i^k}$	OA	mIoU
$\mathbf{r_i^k}$	RGB	84.01%	67.05%
$\mathbf{r_i^k} \oplus \mathbf{f_i^k}$		83.58%	64.07%
$\mathbf{r_i^k} \oplus (\mathbf{f_i} - \mathbf{f_i^k})$		**85.66%**	**69.17%**
$\mathbf{r_i^k}$	RGBI	85.57%	68.75%
$\mathbf{r_i^k} \oplus \mathbf{f_i^k}$		**86.23%**	68.26%
$\mathbf{r_i^k} \oplus (\mathbf{f_i} - \mathbf{f_i^k})$		86.09%	**69.41%**

Table 2. Comparison of the overall accuracy (OA) and the mean Intersection over Union (mIoU) among different setups in the LFA module.

Figure 5 and Figure 6 also illustrate part of the results in our test area. Compared to the ground truth labeling, each feature setup achieves reasonable predictions in general. However, since RandLA-Net learns and aggregates point features in a local neighborhood, inaccurate geometric shapes along object edges can be observed. Moreover, a feature combination that improves the overall accuracy does not ensure performance gains on each road type.

	$\mathbf{f_i^k}$	sidewalk	cycling path	rail track	parking area	motorway	green area	island without traffic
$\mathbf{r_i^k}$	RGB	64.7%	50.6%	**87.1%**	31.2%	82.2%	85.4%	68.2%
$\mathbf{r_i^k} \oplus \mathbf{f_i^k}$	RGB	63.4%	48.6%	75.7%	32.8%	82.3%	83.5%	62.3%
$\mathbf{r_i^k} \oplus (\mathbf{f_i} - \mathbf{f_i^k})$	RGB	**67.5%**	**57.9%**	85.6%	**35.2%**	**83.6%**	**85.5%**	**68.9%**
$\mathbf{r_i^k}$	RGBI	67.3%	50.4%	82.0%	34.2%	85.1%	**87.0%**	**75.3%**
$\mathbf{r_i^k} \oplus \mathbf{f_i^k}$	RGBI	**68.3%**	**53.5%**	82.7%	**36.3%**	**86.0%**	85.5%	65.5%
$\mathbf{r_i^k} \oplus (\mathbf{f_i} - \mathbf{f_i^k})$	RGBI	67.6%	53.2%	**87.0%**	34.6%	85.4%	86.9%	71.1%

Table 3. IoU of each class among different setups in the local feature aggregation module.

4.1 Comparison between different feature combinations

In the case of using RGB features, neighboring weights obtained with the combination of geometric feature vector r_i^k and local feature differences $(f_i - f_i^k)$ result in a dominant advantage in both evaluation metrics. Adding intensity, $\mathbf{r_i^k} \oplus (\mathbf{f_i} - \mathbf{f_i^k})$ helps to achieve the best mIoU of 69.41%, but the gap between it and other feature combinations is much smaller than that shown when only adopting RGB features.

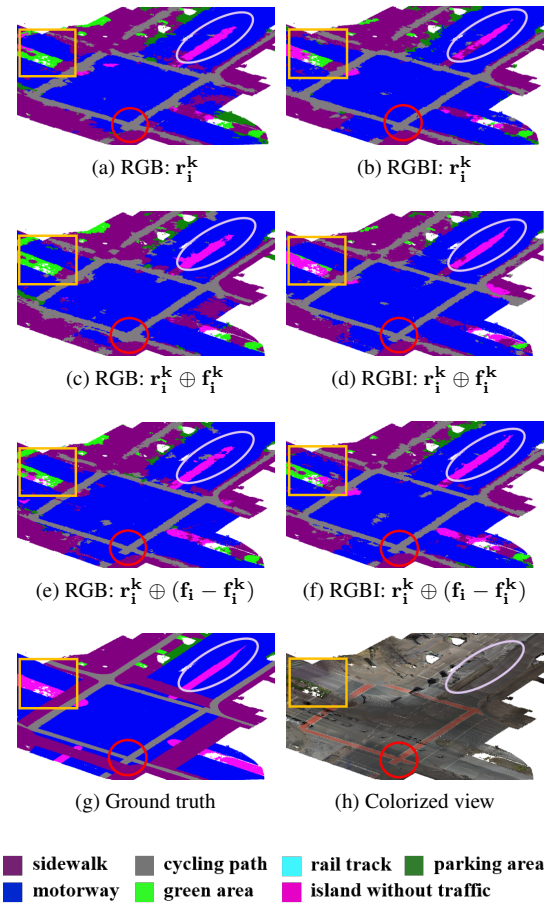


Figure 5. Comparison of road type classification results using different feature combinations to weigh the neighboring points. Holes in the dataset are caused by the removal of cars in pre-processing. Rectangles, ellipses, and circles highlight the differences between results with different feature combinations and the ground truth.

As illustrated in Table 3, the best performance on *cycling path* is achieved when combining geometric feature vector and color difference. The red circled area in Figure 5h shows part of a cycling path painted in two colors. RGB difference in the local region helps to highlight the color variation within one object and produces a clear outline of the cycling path in Figure 5e. However, both the cycling path and motorway in this figure are made of asphalt, so involving the local difference of intensity (i.e., $\mathbf{r_i^k} \oplus (\mathbf{f_i} - \mathbf{f_i^k})$) does not bring an advantage compared to using original intensity values (i.e., $\mathbf{r_i^k} \oplus \mathbf{f_i^k}$), which is also supported by the IoU results in Table 3.

Moreover, Table 3 demonstrates the effectiveness of using the geometric feature r_i^k only in the classification of *green area* and *island without traffic*. As shown in the boxed area of Figure 6h, there is a vegetation stripe next to the southern border of the rail track. Figure 6b indicates difficulties in distinguishing both classes. Due to the illumination condition, the hue of this figure is slightly dark, reducing the contrast in the appearance of *green area* and *rail track*. Eliminating the effect of RGB features when weighing neighboring points helps to highlight the difference in geometrical shapes of objects (see Figure 6a).

For the class *island without traffic*, using only the geometric vector r_i^k shows a dominant advantage. *Island without traffic* refers to areas that channel traffic, which is always slightly higher than the surrounding road surface. As shown in the white circled area in Figure 5, the traffic island has a very similar color as the motorway, which brings confusion in Figure 5c.

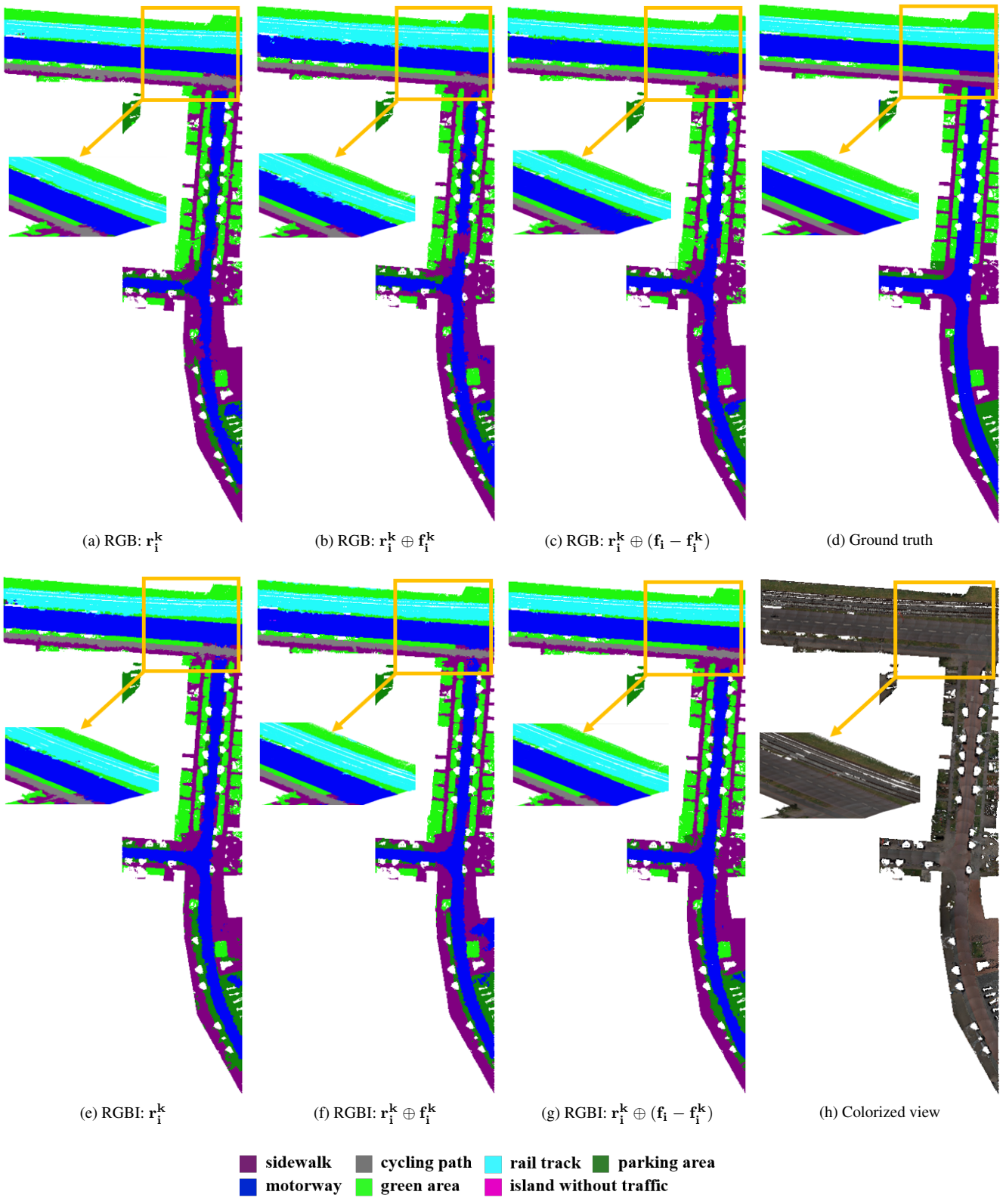
Also, one can see that only the geometric feature vector does not provide enough information for the network when adjacent objects are made of the same material but have a difference in color, especially for classes (e.g., *parking area*) that are sometimes identified by paintings in specific colors.

However, the segmentation performance on the *motorway* class is only slightly affected by the feature combination of weighing the neighboring points, which can also be explained by the object properties. *Motorway* has the most simple geometric characteristics among all these classes and is more invariant than additional features like RGB.

4.2 Impact of intensity

Comparisons of mIoU in Table 2 suggest that adding intensity features is beneficial when classifying different road types of 3D point clouds. Intensity brings effective information in training the model. Only relying on RGB features is not enough to distinguish some classes. First, there exist some traffic islands that are covered with vegetation (see boxed areas in Figure 5), resulting in *island without traffic* misclassified as *green area* if only RGB features are used to weigh neighboring points.

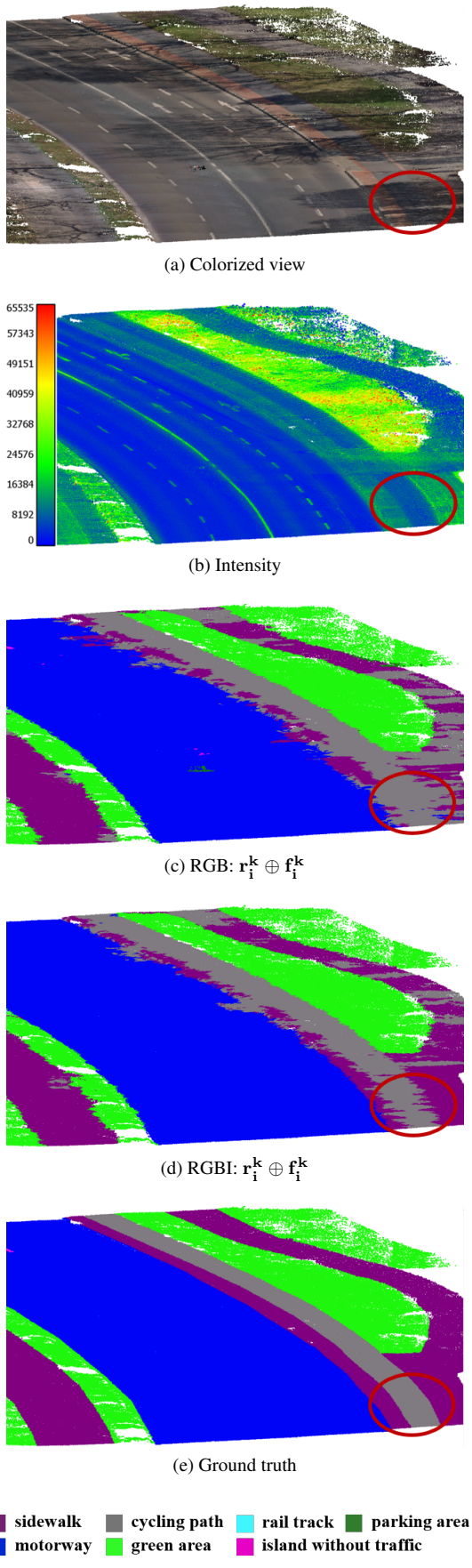
Also, point colors are easily affected by the change of illumination (see Figure 7a), while intensity values are more stable in case of shadows (see Figure 7b). Classification results in Figure 7c indicates that shadows cause confusions between the *sidewalk* and *cycling path* with additional features (R, G, B). Such confusions are largely reduced in Figure 7d, when the intensity feature is also considered.



(a) RGB: $\mathbf{r_i^k}$ (b) RGB: $\mathbf{r_i^k} \oplus \mathbf{f_i^k}$ (c) RGB: $\mathbf{r_i^k} \oplus (\mathbf{f_i} - \mathbf{f_i^k})$ (d) Ground truth

(e) RGBI: $\mathbf{r_i^k}$ (f) RGBI: $\mathbf{r_i^k} \oplus \mathbf{f_i^k}$ (g) RGBI: $\mathbf{r_i^k} \oplus (\mathbf{f_i} - \mathbf{f_i^k})$ (h) Colorized view

Figure 6. Comparison of road type classification results using different feature combinations to weigh the neighboring points.



(a) Colorized view

(b) Intensity

(c) RGB: $\mathbf{r_i^k} \oplus \mathbf{f_i^k}$

(d) RGBI: $\mathbf{r_i^k} \oplus \mathbf{f_i^k}$

(e) Ground truth

Figure 7. Comparison of road type classification results with additional features (R, G, B) and (R, G, B, I). Circle areas indicate the impact of the intensity feature.

In the case of *sidewalk*, *parking area*, and *motorway*, the original feature combination (i.e., $\mathbf{r_i^k} \oplus \mathbf{f_i^k}$) in the LFA module of RandLA-Net gives the best result when intensity is also used as input for the network, as indicated in Table 3. This tells us that intensity has a larger impact on the performance of these classes than the choice of feature combination in local information aggregation.

4.3 Definition of road types

As discussed in previous sections, some road classes in our dataset have the same material type or even appearance, which confuses the classification task to some extent. For instance, the vertical motorway in Figure 6 looks very similar to the sidewalk next to it. The horizontal motorway in this figure, on the other hand, has a different color. Moreover, in our dataset there exists a priority list in labeling, e.g., a road object should be classified as *sidewalk* even though it is also used as *motorway* (see 5), which is due to the importance of promoting green transportation in large cities nowadays.

Indeed, when defining the road type, the usage of a road segment is the most meaningful for the human being and practical applications like urban planning. However, a road class definition with high complexity might harm the performance of deep neural networks.

5. CONCLUSIONS

In this study, a deep neural network designed for the semantic segmentation of large-scale point clouds, RandLA-Net, is employed to classify road types of a colorized MLS point cloud. Considering the key component in RandLA-Net, which is the local feature aggregation (LFA) module, three feature combinations used to calculate point weights in a local neighborhood are assessed and compared. The difference in using RGB and RGBI features in road type classification is also discussed.

Through our experiments, RandLA-Net is demonstrated to be applicable to the road type classification task. The best mIoU (69.41%) is achieved when combining the enhanced geometric feature vector and local differences of RGBI features. The geometric feature vector adopted by RandLA-Net is powerful in modeling the 3D geometry and learning the local shapes of road objects, especially *island without traffic*. Using feature difference instead of the feature itself (which is the original implementation of RandLA-Net) makes it easier to detect complex objects in our dataset, like *cycling path* painted in various colors. Moreover, intensity, an important LiDAR feature, adds effective information to the neural network and helps to overcome the negative effect of illumination changes in the environment, which improves the overall performance of RandLA-Net.

In the pre-processing step, we apply grid sampling with a grid size of 0.1 m, which helps to avoid the problem of varying densities in point clouds and does not harm the local structure of road segments. As future work, more investigation on the effect of downsampling strategies can be conducted. Also, although RandLA-Net aims to process neighboring points indistinguishably through learning different weights, there is still space in improving the delineation between objects in the classification results. The feasibility of RandLA-Net on larger datasets and comparisons to other methods (e.g., image-based methods) should also be further studied. Additionally, urban scenes designed for modern life always show complex characteristics,

bringing difficulties to the automatic detection of objects like road segments. Definition of the road types determines information input to deep neural networks and affects how the scene is modeled. Dividing the road classes in a balanced way, to account for both the test accuracy and practical usage, needs a more detailed discussion in future research.

REFERENCES

Behley, J., Garbade, M., Milioto, A., Quenzel, J., Behnke, S., Stachniss, C., Gall, J., 2019. SemanticKITTI: A Dataset for Semantic Scene Understanding of LiDAR Sequences. *Proc. of the IEEE/CVF International Conf. on Computer Vision (ICCV)*.

Cyclomedia, 2021. Capture data from public spaces. https://www.cyclomedia.com/us/product/data-capture/data-capture. (Accessed on April 2021).

Hu, Q., Yang, B., Xie, L., Rosa, S., Guo, Y., Wang, Z., Trigoni, N., Markham, A., 2020. RandLA-Net: Efficient Semantic Segmentation of Large-Scale Point Clouds. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*.

Isenburg, M., 2014. LAStools - Efficient tools for LiDAR processing. http://rapidlasso.com/LAStools. (Accessed on February 2021).

Liu, W., Sun, J., Li, W., Hu, T., Wang, P., 2019. Deep learning on point clouds and its application: A survey. *Sensors*, 19(19), 4188.

Liu, Z., Hu, H., Cao, Y., Zhang, Z., Tong, X., 2020. A closer look at local aggregation operators in point cloud analysis. *European Conference on Computer Vision*, Springer, 326–342.

Qi, C. R., Su, H., Mo, K., Guibas, L. J., 2017a. PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 77–85.

Qi, C. R., Yi, L., Su, H., Guibas, L. J., 2017b. PointNet++: Deep Hierarchical Feature Learning on Point Sets in a Metric Space. NIPS'17, Curran Associates Inc.

Riegler, G., Osman Ulusoy, A., Geiger, A., 2017. Octnet: Learning Deep 3D Representations at High Resolutions. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 3577–3586.

Velodyne, 2010. HDL-32E: High Resolution Real-Time 3D Lidar Sensor. https://velodynelidar.com/products/hdl-32e/. (Accessed on April 2021).

Wang, Y., Sun, Y., Liu, Z., Sarma, S. E., Bronstein, M. M., Solomon, J. M., 2019. Dynamic Graph CNN for Learning on Point Clouds. *ACM Transactions On Graphics (tog)*, 38(5), 1–12.

Widyaningrum, E., Bai, Q., Fajari, M. K., Lindenbergh, R. C., 2021. Airborne Laser Scanning Point Cloud Classification Using the DGCNN Deep Learning Method. *Remote Sensing*, 13(5). https://www.mdpi.com/2072-4292/13/5/859.

Wu, B., Wan, A., Yue, X., Keutzer, K., 2018. Squeezeseg: Convolutional Neural Nets with Recurrent CRF for Real-time Road-object Segmentation from 3D LiDAR Point Cloud. *2018 IEEE International Conference on Robotics and Automation (ICRA)*, IEEE, 1887–1893.

Zhu, Q., Chen, L., Li, Q., Li, M., Nüchter, A., Wang, J., 2012. 3D LIDAR point cloud based intersection recognition for autonomous driving. *2012 IEEE Intelligent Vehicles Symposium*, 456–461.