International Journal on Advanced Science Engineering Information Technology

Vol.13 (2023) No. 6
ISSN: 2088-5334

Analysis and Evaluation of PointNet for Indoor Office Point Cloud Semantic Segmentation

Calvin Wijaya [a], Harintaka [a,*]

[a] *Department of Geodetic Engineering, Faculty of Engineering, Universitas Gadjah Mada, Yogyakarta, 55281, Indonesia*
*Corresponding author: *harintaka@ugm.ac.id*

***Abstract*—Indoor modeling is one of the primary sources of information in building management due to the increased use of BIM in the AEC industry. The indoor model can be acquired with several survey instruments, but TLS is the most popular resulting point cloud that can be processed into a 3D model. However, the process commonly still uses inefficient manual methods. Point cloud data have irregular, unordered, unstructured characteristics, making them more challenging to process. The deep learning algorithm can be a solution to solve the problem. PointNet is the first deep learning algorithm that directly accepts point cloud data as input. This study aims to analyze and evaluate the office indoor point cloud segmentation using PointNet. The office indoor point cloud data was acquired using TLS and then pre-processed for deep learning input. Transfer learning strategy is used as a weight initialization technique. The pre-trained model was trained with the S3DIS dataset and then fine-tuned to segment nine indoor classes in this study. The result shows PointNet achieves 85% overall accuracy and 66% average class IoU score to predict indoor classes using this study's point cloud data. Geometry control shows that the predicted point cloud has an RMSE score of 1.8 cm, meaning the geometries of the segmented point cloud are accurate. Using the transfer learning method has increased the performance of the deep learning model. Further research is needed to evaluate the model thoroughly using more training and evaluation data and different transfer learning strategies.**

***Keywords*— Point cloud; deep learning; PointNet; semantic segmentation; TLS.**

Manuscript received 15 Apr. 2023; revised 21 Aug. 2023; accepted 9 Sep. 2023. Date of publication 31 Dec. 2023.

I. INTRODUCTION

Indoor modeling has rapidly gained the interest of researchers and developers and has gained tremendous attention in the last decade [1]. The massive need for and use of BIM for the architecture, engineering, and construction (AEC) industries also increases the critical value of an indoor model. It supports a digital representation that gives detailed information about a room. The indoor model has a specific function [2] and is one key building management aspect. The 3D indoor model can be acquired using several 3D data, one being the point cloud. The point cloud is the most common three-dimensional data type used to create BIM and map indoor models [3].

The point cloud is a set of points defined in 3D space [4]–[6]. It usually consists of a million points, including 3D position information and additional information such as radiometric or color information and density [7]. Point clouds have become essential data lately because of the increased availability of acquisition devices. It became popular because it is used in various applications, such as robotics, autonomous driving, and augmented-virtual reality [4]. The point cloud is also a basis data for representing 3D, aside from depth images, meshes, and volumetric grids. The point cloud presentation preserves original geometric information without discretization, making it a preferred representation and 3D data format in several applications [8].

The point cloud data can be acquired mainly with four techniques: image-derived methods, Light Detection and Ranging (LiDAR) systems, Red Green Blue – Depth (RGB-D) cameras, and Synthetic Aperture Radar (SAR) systems [9]. However, LiDAR or laser scanners have become the favorite method for point cloud acquisition, especially in urban applications. 3D laser scanning also become a standard procedure for architecture, reconstruction, and BIM [10]. Recent advances in laser scanner systems allow the automatic, fast, efficient, and accurate collection of point clouds with a high level of detail [6], [11].

The process of semantically generating 3D indoor models from a point cloud is formally known as Scan-to-BIM. This process usually needs segmented point cloud data into distinct subsets and then populating the segmented scene using a predetermined set of 3D objects [12]. The point cloud data needs to be segmented into several categories. In other words, the point cloud is segmented to the semantic level, referred to as the point cloud semantic segmentation process [3]. However, the Scan-to-BIM and point cloud segmentation processes mostly remain manual. This surely is a time-consuming and inefficient method. Besides consisting of millions of data, the point cloud also has several characteristics which complicate the process. The point cloud data is irregular, unstructured, and unordered [4], making it more challenging to process.

The emergence and rapid use of artificial intelligence, machine learning, and deep learning have changed how a problem can be solved. As classified by Nguyen and Le [13], machine learning algorithms for point cloud segmentation can be classified into five methods: edge-based, region-based, attribute-based, model-based, and graph-based. Each of them has a different approach to the segment point cloud. Liu et al. [5] also summarized machine learning algorithms for point cloud segmentation into two main groups. The first involves mathematical models and geometric shapes for segmentation, such as region-growing algorithms, model fitting algorithms or least squares fitting algorithms, Hough transforms [14], and Random Sample Consensus (RANSAC) [15]–[21] algorithms. However, each algorithm has its limitations for point cloud processing, and while deep learning is growing rapidly, advances in deep learning have started taking a role in point cloud segmentation.

The deep learning method takes much attention due to its ability to process huge and complex data. However, it still faces challenges in applying deep learning for point cloud processing due to the unstructured characteristics of point cloud [4], [22]. Several paradigms for point cloud segmentation use deep learning, projection-based, discretization-based, point-based, and hybrid [8]. Projection-based and discretization-based works by projecting or discrete the point cloud data into regular representation, while point-based works directly on the point cloud. The hybrid method combines projection or discretization with the point-based method.

Wu et al. [23] were the first to use a neural network to transform point clouds into a 3D voxel grid, creating 3D ShapeNet. Since then, several researchers have created and used various algorithms for point cloud segmentation, such as VoxNet [24], VV-Net [25], SEGCloud [26], Fully-Convolutional Point Networks (FCPN) [27], ScanComplete [28], Convolutional Neural Network [29], Dense Conditional Random Forest (DCRF) [30], etc. However, all voxelization processes to transform a point cloud into a 3D voxel grid introduce discretization artifacts and information loss. It also leads to high memory and computational cost if using a high-resolution voxel or loss of details if using a low-resolution.

Multi-view projection transforms 3D point cloud data into multi-view 2D images, and then each image is processed based on 2D CNN [9]. The most influential multi-view deep learning is Multi-View Convolutional Neural Network (MVCNN) [31], although it was not used for point cloud segmentation. It influences the use of multi-view representation for point cloud segmentation, such as SnapNet [32], SqueezeSeg [33], and RangeNet++ [34]. The projection-based model performs better than the discretization-based model but still introduces information loss due to the transformation from 3D to 2D. Finding suitable and proper images to feed into 2D CNN is also tricky.

Several research studies also try to combine different methods to resolve the limitations of both methods. Qi et al. [35] combine CNN-based volumetric representations with multi-view representations for extensive analysis. Meanwhile, Dai and Nießner [36] created a 3DMV algorithm that combines multi-view and voxel discretization in a continuous network. Although it combines both methods, there are possibilities of information loss. To solve the problem, we are to use raw point cloud data as input directly.

The first neural network to use the point cloud directly as input for deep learning processing is PointNet, which was created by Qi et al. [22]. PointNet processes point clouds directly without projection or discretization into voxel. It respects the permutation invariance of point input to solve unordered characteristics of the point cloud. PointNet marks a new era of deep learning for point cloud processing that uses point cloud directly. It became a bedrock and pioneering framework for most methods now due to its simplicity but an effective network [3], [4], [8], [9], [37]. PointNet can be used for several tasks such as classification, part segmentation, semantic segmentation, including Scan-to-BIM purposes [3], and registration [38].

Most of the research and study done for PointNet commonly uses Stanford 3D Indoor Scene (S3DIS) dataset for point cloud segmentation. The S3DIS is commonly used as a dataset and evaluates the PointNet, so the deep learning model is newly made for each research. This research does not use point S3DIS for the process and instead uses data acquired from TLS. This research also does not create a new model of PointNet; instead, it uses the pre-trained model. Several experiments have been done in this research that aim to analyze and evaluate the abilities of PointNet.

II. Materials and Method

A. Study Area

The research is located at the Engineering and Research Innovation Centre (ERIC), Faculty of Engineering, UGM. Despite having several floors, this research only uses certain rooms with some reason and consideration. Certain rooms with different sizes, geometry, and complexity were chosen as study areas to simulate deep learning capability with various conditions. With different situations in the rooms, it can be used to analyze and evaluate deep learning performance in different circumstances.

The rooms scanned have a variety of complexities, referring to interior fittings in the room. Certain rooms may not have any interior inside, while others are full of it. The scanned rooms consist of 14 office rooms, one hall area, one stair area, and four corridors. Most scanned rooms are typical rooms such as the meeting rooms, secretary rooms, and lab, while other rooms, such as the pantry and toilet, are not scanned. Each room scanned has a different level of detail and complexity, used to test the accuracy of the deep learning model to segment indoor point cloud effectively in various rooms.

B. Research Data

The data includes point cloud as main data acquired from TLS scanning, S3DIS pre-trained model, tie point, and ground truth data. The indoor point cloud data in the research location was acquired using Leica RTC360 TLS as the primary data for semantic segmentation. A pre-trained model from S3DIS was also used in the research to accelerate the training process and improve accuracy.

Another data used is Ground Control Points (GCP) acquired from GNSS and terrestrial measurement. A GCP is used to georeferenced the point cloud data from the local coordinate system into the ground coordinate. Ground truth data was also collected from geometry measurements of the indoor situation. It is then used in the quality control process to compare real-world geometry and attributes with the segmented point cloud.

C. Research Methodology

The methodology used in the research is generally divided into three main steps: preparation, data acquisition, processing, and analysis. The preparation step is the first stage, which consists of preparing the instrument, forming the acquisition team, scheduling, administering, and conducting the preliminary survey. Preparation is important to ensure data acquisition times are effective and efficient. A preliminary survey is also needed to map rooms inside the study area. All instruments used are prepared in this step. Data acquisition is the next stage to collect all research data. After collecting research data, the processing stage can start with the pre-processing point cloud, segmentation process, and quality control process. The flowchart of the research methodology is shown in Fig. 1.

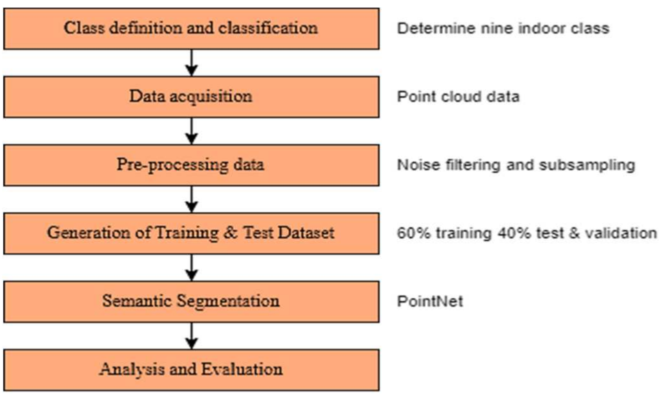

Fig. 1 Research flowchart

A detailed explanation from the research flowchart is described below:

1) Class Classification: Class classification and definition aim to create a classification schema of what object or indoor element will be segmented from the point cloud. The final segmented point cloud class is customized for indoor office situations referencing S3DIS classes. This research will classify point clouds into nine classes: ceiling, floor, wall, column, window, door, table, chair, and interior. The interior class includes any object inside the rooms except tables and chairs, which means various objects such as books, bookcases, lamps, vases, boxes, and cupboards. This research does not create a new deep-learning model but only fine-tunes a pre-trained model. Other research that uses the PointNet model for segmentation and the S3DIS dataset aims to create a new and improve the deep-learning model, for example, Point Transformer by Zhao et al. [39] and A-CNN by Komarichev et al. [40]. The indoor class used in this research is defined in Table I below.

TABLE I
CLASS CLASSIFICATION

Label	Class	Definition
(0)	Ceiling	The top part of the room surface covers the upper limits of a room
(1)	Floor	The bottom part of the room, level base
(2)	Wall	The surfaces that enclose the room
(3)	Column	A vertical compression member in structures, the effective length of which exceeds three times its lateral dimension
(4)	Window	An opening formed in a wall, admits daylight through some transparent material
(5)	Door	Accessible or movable barriers through the wall
(6)	Table	Furniture with a raised flat top, supported most by 4 legs
(7)	Chair	A seat, typically designed for one people
(8)	Interior	All other furniture besides the table and chair

The S3DIS dataset is commonly used as a benchmark dataset to see the effectiveness of a created deep-learning model. The pre-trained model of the S3DIS dataset classifies point clouds into 13 classes, while this research only uses 9 classes, so there must be a fine-tuning process to change the k number of classes in the model.

2) Data Acquisition: This step aims to collect the point cloud data in the research location. To collect point cloud data, Leica RTC360 was used. It is a portable and easy-to-use TLS with high accuracy, making it suitable for indoor situations. Each room has several scan stations depending on room size and complexity. A room with several scan stations is registered directly in the application field using the Cloud-to-Cloud method.

Fig. 2 Point cloud acquisition process (Source: Author Documentation)

There are 54 scan stations in the research location, consisting of 19 stations on the first floor and 35 others on the second floor. The point cloud from the acquisition instrument has a specific data format and must be processed with Leica Cyclone Register360. The point cloud from the acquisition process has 3D local coordinates, color (RGB), and intensity information.

3) Point Cloud Pre-Processing: This process aims to ensure the point cloud data is ready for segmentation. The point cloud pre-processing step consists of noise filtering, subsampling, and georeferencing point cloud. Each of them is essential and has a specific purpose. Noise filtering is used to remove noise and outliers from the data, subsampling is used to reduce the total number of data to accelerate time processing, and georeferencing to take the local system into the ground coordinate system.

Noise filtering was done using a SOR filter, noise filter, or manual segmenting out the noise. The SOR filter uses a statistical approach to delete outliers using the average distance of a point to its neighbors and reject points that are farther than average plus standard deviation [41]. The method eliminates points far from their neighbors based on statistical calculations [5]. Meanwhile, the radius noise filter considers the geometric position between points to detect and delete noise [41]. The noise or outliers in point cloud data are caused by various things such as sensors or environmental factors [42]. The presence of noise can cause calculations and subsequent processing of point clouds to become inaccurate [5].

The next pre-processing step is a georeferencing point cloud. The process included in the alignment method and operator needs to pick a minimum of 4 equivalent pairs of points between the tie point in the point cloud and input manually coordinates of the tie point sequentially. After the alignment process, the final Root Mean Square Error (RMSE) is 19.47 mm, which means it is very small.

4) Training and Test Dataset: The deep learning architecture needs a training dataset to learn and create a model and needs a test dataset to evaluate the model. Training data in point cloud segmentation is the point cloud data containing its information (3D coordinates and additional information such as color) and a ground truth label. Deep learning will use the information to learn from geometric features derived from 3D coordinates and radiometric features derived from color corresponding with specific label classes. The training process then creates "knowledge" stored in a model and uses it for prediction.

The process of creating a training and test dataset label was done precisely using segment tools in CloudCompare. The operator selects and defines which class belongs to or represents each indoor class, then segments out the rest of the point cloud data. The process is then repeated until all indoor elements in the room are labeled out for every room, and several rooms are chosen for training and others for the test process. The sequence of indoor class or the label is the same as described in Table I. A sample of input point cloud data in RGB and its GT after the labeling process is shown in Fig. 3 below.

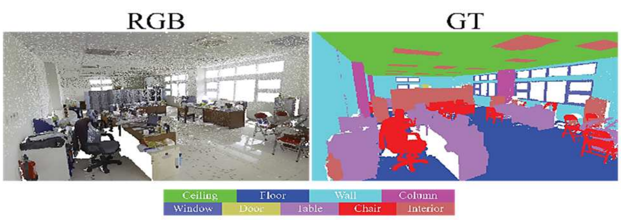

Fig. 3 Point cloud labeling for Ground Truth (GT) (Source: Author Documentation)

This research uses 63% of point cloud data for the training dataset and 37% for others for the test dataset. The point cloud distribution for the training and test dataset is shown in Fig. 4. The ceiling, floor, and wall classes have the highest number for both training and test data, while other classes have fewer numbers. The reason is that those three classes are the main structure of the rooms that must exist in every room. Several rooms can be identified with no column, door, table, chair, or interior, but all rooms have ceiling, floor, and wall.

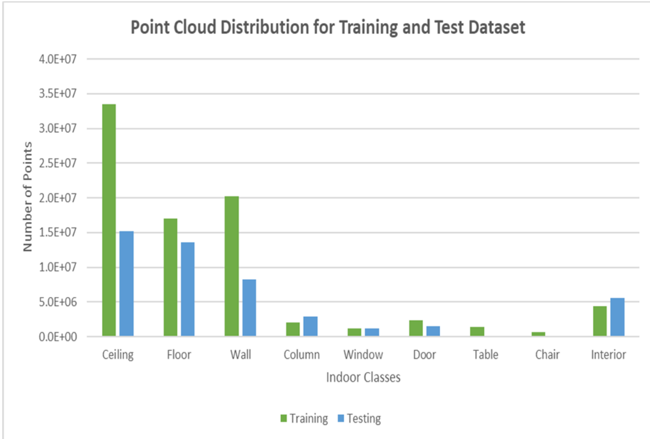

Fig. 4 Point cloud distribution for training-testing data

5) Point Cloud Semantic Segmentation: This step is the focus of research, where pre-processed point clouds are then segmented automatically. The predicted segmentation is then analyzed for evaluation. The PointNet architecture, as shown in Fig. 5, is used to segment the point cloud due to its ability to process huge amounts of data. PointNet is the first neural network that directly processes point clouds, which respects the permutation invariance of points in the input [22]. The architecture is chosen because it directly processes the point cloud without regularization or first transforms it into a regular structure. PointNet is not a new algorithm in point cloud processing but can still be used and become a backbone for many architectures today. The research uses and modifies script code created by Yan [43]. The code uses Pytorch as a framework, and this research uses Python 3.8 as a programming language.

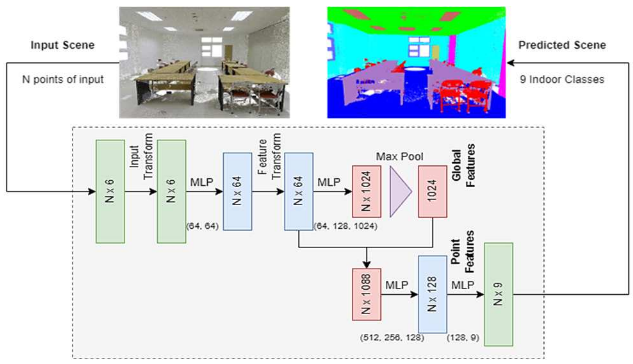

Fig. 5 Workflow in the segmentation process. (Source: Authors Documentation)

6) Performance Evaluation: Performance evaluation aims to ensure that the quality of the segmentation process is correct and accurate. This research has done two types of performance evaluation to evaluate both model and geometric of the segmented point cloud. Each of them has its objective to evaluate the quality of segmentation. The model evaluation aims to ensure the trained model has the highest accuracy for prediction. The evaluation is used to determine which model has the highest Intersection over Union (IoU) score. The best model is then used for prediction. This research uses the IoU score as a performance evaluation, as seen in Equation 1.

$$IoU = \frac{Area\ of\ Overlap}{Area\ of\ Union} = \frac{TP}{(TP+FP+FN)} \tag{1}$$

In segmentation problems, the IoU score is the standard performance measure that is commonly used to evaluate model accuracy [44]. True Positive (TP) is defined as the area of intersection between ground truth and segmentation mask, or in other words, is the overlap area. It is then divided by the union area total of True Positive (TP), False Positive (FP), and False Negative (FN). IoU score ranges from 0 to 1. An IoU score greater than 0.5 is considered a good prediction.

The geometric evaluation aims to keep segmented point clouds with accurate geometry like width, length, and scale. It is done by comparing ground truth geometry with measured geometry from the point cloud. The ground truth geometry is acquired using Leica Distometer™ D2. Using a Distometer to measure geometry helps reduce blunders due to misreading when using a manual method like measuring tape. This research measures 70 samples of geometric dimension used to compare with model measurement. The difference between them was then calculated to find Root Mean Square Error (RMSE).

III. Results and Discussion

A. Semantic Segmentation

The deep learning process started with training to create a model and use the model for prediction. Fig. 6 shows training accuracy and mean loss score information for each training epoch. The line in the chart has an opposite direction, where the training tends to go upward while the mean loss tends to go down. This means that for each training epoch, its accuracy in creating an accurate model is getting higher than the previous epoch, and the loss score is decreased. The training accuracy and mean loss score are defined as the difference in learned knowledge of the model with ground truth labels in training data. The difference is adjusted and then propagated back into the neural network.

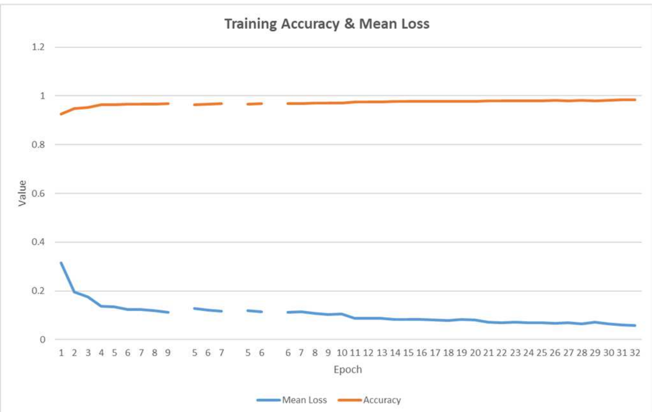

Fig. 6 Training accuracy and mean loss score

Training accuracy gets a 0.92 score at the first epoch and 0.31 for the loss score. It indicates the training accuracy is still getting a high score despite being the first epoch. The reason is possibly due to using a pre-trained model or transfer learning strategies that were applied in this research. The weight parameters have an initial value from training before using S3DIS data, then re-train to get a fine-tuned model. The training accuracy gets higher for each epoch and reaches a 0.98 score in the final epoch, and the mean loss gets a 0.05 score. This means the training process creates an accurate model where it is getting better for each epoch.

The training accuracy and loss score obtained from the training process indicates that the neural network learns to extract information from the input data, specifically the training data. The knowledge inside the model was then tested with a test dataset and compared between them. The process then gives a mIoU score, as shown in Fig. 7. The evaluation process is repeated for each epoch until the mIoU score is the highest possible. Fig. 7 shows that the mIoU score gradually increased from the first epoch to the final epoch. The best mIoU score is 0.47, lower than expected. It can happen due to several factors, but in this case, it is possibly caused by the minimum training and test data dataset. The difference in the situation between the office and hall-corridor area in the research area also possibly became the reason it did not get the mIoU score as expected.

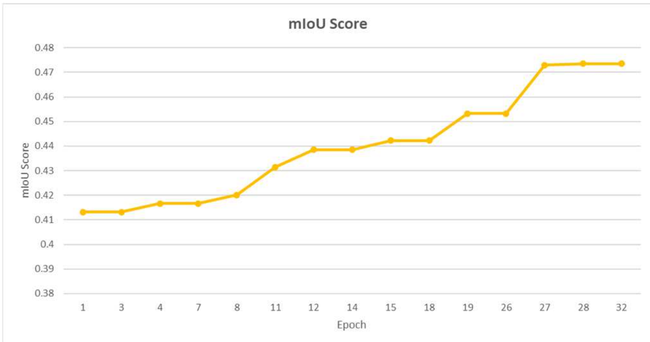

Fig. 7 mIoU Score

The training process creates two models, a normal model and the best model which has the highest mIoU score. The best model is used to predict other point cloud data besides training and test datasets. The sample prediction of point cloud segmentation into nine indoor classes is shown in Fig. 8. Fig. 8 shows that the trained or best model successfully predicts the point cloud data. It means the training process is accurate enough to create a deep learning model for point cloud segmentation even though the test is not as expected.

The prediction of point cloud has also been segmented into nine classes as defined. This indicates the successful use of the transfer learning strategy to use a pre-trained model to re-train the whole network. It changes the defined class in the pre-trained model from 13 classes as S3DIS class into nine classes as defined in this research. Although the prediction is not as clean as the ground truth, it gives an insight into how powerful and capable a deep learning model can be used for point cloud segmentation.

B. *Intersection over Union Analysis*

The Intersection over Union (IoU) score is a common method used for the analysis and evaluates the accuracy of the segmentation process. After segmentation, the score is calculated by comparing the ground truth label and the prediction class. Mathematically, it was calculated using True Positives (TP), False Positives (FP), and False Negatives (FN) of the data. The matrix version of the IoU score is mIoU (see equation 1), which has the same concept to compare ground truth and prediction. The score measures model accuracy and evaluates the whole scene prediction.

The evaluation measured by the IoU score defines how many predicted point clouds are correct according to the test dataset. Fig. 9 gives the overall IoU score per class after evaluation. From Fig. 9, it can be seen that most classes have high IoU scores, which indicates that the predicted point cloud using the trained model is the same as the ground truth label of the test dataset. The ceiling, floor, wall, door, table, and chair classes get IoU scores above 0.5 indicates a good prediction. Meanwhile, three other classes, column, window, and interior, get IoU scores varying from 0.3 to 0.4. However, it still achieves quite near the threshold for them. The structure class, ceiling, floor, and wall get high IoU scores, comparable with their training accumulation.

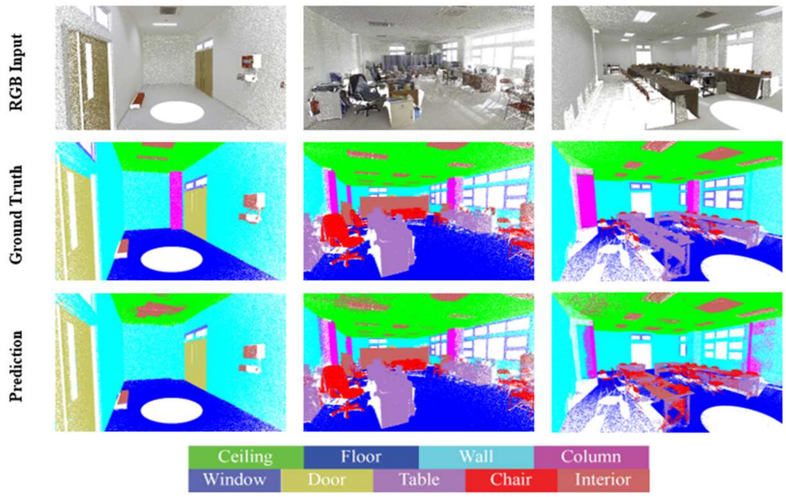

Fig. 8 Visual comparison of RGB input, GT, and the prediction (Source: Author Documentation)

The column gets the lowest IoU score compared to another class, indicating the model still has difficulty finding the class. It can be caused by the column data itself having the same geometric and radiometric attributes as the wall class. The column has a planar surface and the same color as the wall in most rooms, making it difficult for the algorithm to differentiate them. It can occur for windows because of the window's geometry, consisting of many frames. The interior class cannot get the threshold, possibly caused by too many variations of interior classes in training data. This research classifies small indoor elements inside the room except for table and chair as an interior class, making the training data for the interior vary, which makes learning inconsistent in the trained model.

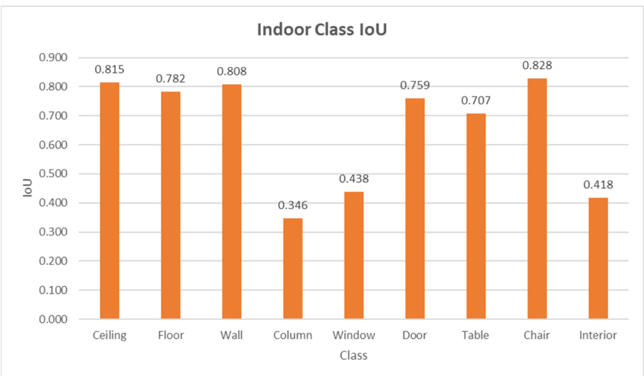

Fig. 9 Indoor class IoU per class

Besides being calculated per class, the IoU score can be evaluated per room, as shown in Fig. 10. It gives clear information on which rooms get high and low IoU scores or which rooms get well-predicted using a trained model. Most of the rooms get high IoU scores, indicating a good prediction. A room with office situations gets a higher score than a non-situated one. An area such as a hall, corridor, or stair gets a lower score than indoor rooms. The reason is that the training dataset contains mostly indoor office situations. It makes the trained model know how to predict the office indoor rooms, giving it a higher IoU score. In other words, if the model has already seen a similar situation with training data, it can be more accurate to predict or segment it.

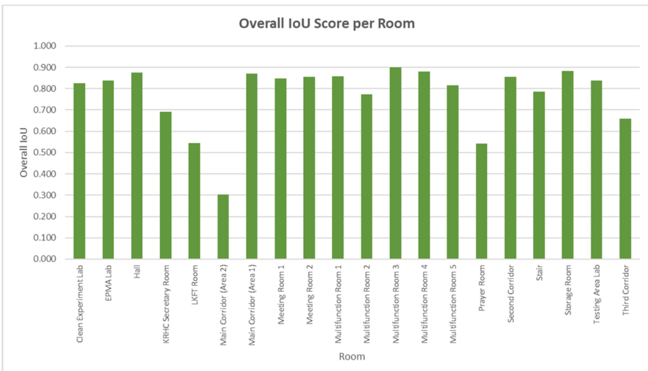

Fig. 10 Indoor class IoU per room

The overall accuracy and class average IoU after the prediction are quite high; both are shown in Table II. The prediction gets a higher score than previous research that also uses the PointNet model. The transfer learning strategies used possibly became the reason this research gets higher accuracy and IoU score. The advantage of using the transfer learning method is an accuracy improvement due to the loss score having been minimized from previous training. Although it is not a fair comparison due to the different data used, it compares how transfer learning can boost the deep learning model's accuracy.

TABLE II
OVERALL ACCURACY COMPARISON

Research	OA	Class Avg IoU	Data
Qi et al. [22]	78.62	47.71	S3DIS
Yan [43]	78.90	43.70	S3DIS
This research	84.57	65.58	Research data

C. *Performance Evaluation*

This research used two performance evaluations to evaluate the model and geometry of the point cloud after segmentation. Based on previous explanations, choosing the right training and test dataset affects both the model and final prediction. The training data influences the trained model or how the model is trained. The trained model is then used to predict the test dataset, and the prediction is compared with GT labels for evaluation. Well-chosen and generated training and test datasets have an impact on the results.

After training with the training dataset, the trained model stores information, knowledge, and patterns. It means the information inside the training dataset affects the trained model. The pattern inside is what is extracted for the model. The model itself is then tested or evaluated using a test dataset by comparing the label between the prediction created by the model and the ground truth. It means there is a correlation between the training and test datasets. If the training data has different information and patterns from the test dataset, it possibly gets a low evaluation score due to the model not being able to predict correctly. This is because the different pattern between both datasets makes the model difficult to predict. This emphasizes the importance of training and testing datasets for the deep learning process.

Geometry evaluation aims to ensure the predicted model still holds the original geometry or has the correct geometries of point cloud data [45]. This is important because the indoor point cloud data is commonly used as geometry information of BIM and as digital twin models, so the geometries of the room must be proportionally correct. The evaluation is done by comparing ground truth geometries of a room (width, length, height) with the corresponding pair in the model. The statistics of geometry evaluation are shown in Table III.

TABLE III
GEOMETRY QUALITY CONTROL

Geometry Eval	Values (m)
Maximum difference	0.0394
Minimum difference	0.0007
Average difference	0.0154
Deviation	0.0101
RMSE	0.0184

The average difference between the ground truth and the model is only 1.5 ± 1 cm, meaning the model is geometrically accurate. From the statistics, it can be known that the predicted model has accurate geometry and is approximately like a real-life room. The difference between real-life and model in geometric aspect can be caused by several factors, such as the instrument's accuracy for point cloud acquisition, tie point measurement, geometry evaluation measurement, and then the accuracy when georeferencing data and last accuracy when measuring model geometry in software. While the instruments used in this research have high accuracy, there is a possibility of an error caused by the operator when georeferencing and measuring the model.

D. *Discussion*

This study uses the deep learning PointNet model for point cloud segmentation and transfer learning strategies for weight initialization. The point cloud segmentation process aims to separate point cloud data into several subsets according to the semantic meaning of the points [8]. The term segmentation is widely used in computer vision and deep learning applications [22], [46] and has similar meaning with point cloud classification [47], [48] or point cloud labeling [49], [50] in remote sensing and photogrammetry.

This research indeed successfully uses PointNet to predict the point cloud data into nine indoor classes. The evaluation score also shows that the overall accuracy and IoU score proves the prediction is accurate enough. Compared with Qi et al. [22], this research achieves higher accuracy and IoU score. Xiong and Wang [3] use PointNet also for semantic segmentation S3DIS dataset and create a BIM model using the Dynamo plugin in Revit. The research shows a similar pattern where the overall accuracy is good, especially for segmenting walls, floors, and ceilings. The finding is similar to what this research shows in Fig. 9, where the structural room is well segmented. While another class is still worthy of recognition, it is not clear as structural elements.

The finding of Haznedar et al. [51] is parallel with the explanation of the training and test dataset in the previous section. Haznedar et al. [51] use PointNet to segment the heritage point cloud into five classes. However, the results did not achieve sufficient accuracy due to the deformation and deterioration of the existing buildings in training data. If the PointNet algorithm is trained with restitution data, it gives higher accuracy. This research also shows that the mIoU score is lower than expected as shown in Fig. 7, the minimum training and test datasets and the different conditions or situation between them cause the reason. It emphasizes the important role of training and testing datasets for deep learning. The statement about training data was also provided by Nurunnabi et al. [52], who use PointNet for large-scale outdoor point cloud semantic segmentation.

IV. CONCLUSION

The PointNet deep learning architecture can be used to predict or segment indoor point cloud data into several indoor classes. It receives point cloud data input directly, making it efficient because the data does not need to be transformed into regular representations first. The IoU score per class shows that the PointNet accurately predicts the room's structural elements, such as walls, ceilings, and structures. The IoU score per room shows that it can predict better for office rooms than for areas such as halls and corridors. This is related to the condition of the training and test dataset. Both datasets have an essential role in the deep learning process, either in training or testing and evaluation. The mIoU score still does not provide a sufficient score, probably related to the training and test dataset used in this research. However, the geometry of the segmented point cloud has been checked and represents an accurate geometry for indoor modeling.

Acknowledgment

This research can be done with help from Leica Geosystems Indonesia, which provides the TLS instrument and equipment. The authors also want to thank the Department of Geodetic Engineering, Universitas Gadjah Mada, for financial support for this research.

References

[1] F. Mortari, S. Zlatanova, L. Liu, G. Sithole, and J. Zhao, *Space subdivision for indoor applications*, no. December. 2014. doi: 10.13140/2.1.2914.2081.

[2] S. Zlatanova, U. Isikdag, and M. S. Fine, "3D Indoor Models and Their Applications," *Encycl. GIS*, pp. 1–12, 2015, doi: 10.1007/978-3-319-23519-6.

[3] Z. Xiong and T. Wang, "Research on BIM Reconstruction Method Using Semantic Segmentation Point Cloud Data Based on PointNet," *IOP Conf. Ser. Earth Environ. Sci.*, vol. 719, no. 2, 2021, doi:10.1088/1755-1315/719/2/022042.

[4] S. A. Bello, S. Yu, C. Wang, J. M. Adam, and J. Li, "Review: Deep learning on 3D point clouds," *Remote Sens.*, vol. 12, no. 11, 2020, doi:10.3390/rs12111729.

[5] S. Liu, M. Zhang, P. Kadam, and C.-C. J. Kuo, *3D point cloud analysis : traditional, deep learning, and explainable machine learning methods*. Switzerland: Springer International Publishing, 2021.

[6] Z. Ning, L. Tang, S. Qi, and Y. Liu, "Deep Learning on 3D Point Cloud for Semantic Segmentation," *Smart Innov. Syst. Technol.*, vol. 250, no. September, pp. 275–282, 2022, doi: 10.1007/978-981-16-4039-1_27.

[7] M. E. Atik, Z. Duran, and D. Z. Seker, "Machine learning-based supervised classification of point clouds using multiscale geometric features," *ISPRS Int. J. Geo-Information*, vol. 10, no. 3, 2021, doi:10.3390/ijgi10030187.

[8] Y. Guo, H. Wang, Q. Hu, H. Liu, L. Liu, and M. Bennamoun, "Deep Learning for 3D Point Clouds: A Survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, pp. 1–1, 2020, doi: 10.1109/tpami.2020.3005434.

[9] Y. Xie, J. Tian, and X. X. Zhu, "Linking Points With Labels in 3D: A Review of Point Cloud Semantic Segmentation," *arXiv*, no. August, 2019.

[10] G. Rocha, L. Mateus, J. Fernández, and V. Ferreira, "A scan-to-bim methodology applied to heritage buildings," *Heritage*, vol. 3, no. 1, pp. 47–65, 2020, doi: 10.3390/heritage3010004.

[11] H. Macher, T. Landes, and P. Grussenmeyer, "From Point Clouds to Building Information Models: 3D Semi-Automatic Reconstruction of Indoors of Existing Buildings," *Appl. Sci.*, vol. 7, no. 10, Oct. 2017, doi: 10.3390/app7101030.

[12] Y. Perez-Perez, M. Golparvar-Fard, and K. El-Rayes, "Segmentation of point clouds via joint semantic and geometric features for 3D modeling of the built environment," *Autom. Constr.*, vol. 125, no. December 2020, p. 103584, 2021, doi: 10.1016/j.autcon.2021.103584.

[13] A. Nguyen and B. Le, "3D point cloud segmentation: A survey," 2013 6th IEEE Conference on Robotics, Automation and Mechatronics (RAM), Nov. 2013, doi: 10.1109/ram.2013.6758588.

[14] R. O. Duda and P. E. Hart, "Use of the Hough Transformation to Detect Lines and Curves in Pictures," *Commun. ACM*, vol. 15, no. 1, pp. 11–15, 1972, doi: 10.1145/361237.361242.

[15] F. Poux, C. Mattes, and L. Kobbelt, "Unsupervised Segmentation of Indoor 3D Point Cloud: Application to Object-Based Classification," *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci. - ISPRS Arch.*, vol. 44, no. 4/W1, pp. 111–118, 2020, doi: 10.5194/isprs-archives-XLIV-4-W1-2020-111-2020.

[16] L. S. Runceanu, S. Becker, N. Haala, and D. Fritsch, "Indoor Point Cloud Segmentation for Automatic Object Interpretation," Publikationen der Deutschen Gesellschaft für Photogrammetrie, Fernerkundung und Geoinformation e.V., pp. 147–159, 2017.

[17] T. Watanabe, "A Fuzzy RANSAC Algorithm Based on Reinforcement Learning Concept," in *IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, 2013, pp. 1–6.

[18] M. A. Fischler and R. C. Bolles, "RANSAC: Random Sample Paradigm for Model Consensus: A Apphcatlons to Image Fitting with Analysis and Automated Cartography," *Graph. Image Process.*, vol. 24, no. 6, pp. 381–395, 1981.

[19] Ruwen Schnabel, Roland Wahl, and Reinhard Klein, "Efficient RANSAC for Point-Cloud Shape Detection," *Comput. Graph. Forum*, vol. 26, no. 2, pp. 214–226, 2007.

[20] L. Li, F. Yang, H. Zhu, D. Li, Y. Li, and L. Tang, "An improved RANSAC for 3D point cloud plane segmentation based on normal distribution transformation cells," *Remote Sens.*, vol. 9, no. 5, 2017, doi: 10.3390/rs9050433.

[21] A. Adam, E. Chatzilari, S. Nikolopoulos, and I. Kompatsiaris, "H-RANSAC: A Hybrid Point Cloud Segmentation Combining 2D and 3D Data," *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.*, vol. 4, no. 2, pp. 1–8, 2018, doi: 10.5194/isprs-annals-IV-2-1-2018.

[22] C. R. Qi, H. Su, K. Mo, and L. J. Guibas, "PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation," *Comput. Vis. Pattern Recognit.*, 2017.

[23] Z. Wu *et al.*, "3D ShapeNets: A deep representation for volumetric shapes," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 07-12-June, pp. 1912–1920, 2015, doi:10.1109/CVPR.2015.7298801.

[24] D. Maturana and S. Scherer, "VoxNet: A 3D Convolutional Neural Network for Real-Time Object Recognition," *Int. Conf. Intell. Robot. Syst.*, pp. 922–928, 2015, [Online]. Available: http://www.thepositiveencourager.global/the-mentoring-approach/

[25] H. Y. Meng, L. Gao, Y. K. Lai, and Di. Manocha, "VV-net: Voxel VAE net with group convolutions for point cloud segmentation," *Proc. IEEE Int. Conf. Comput. Vis.*, vol. 2019-Octob, pp. 8499–8507, 2019, doi: 10.1109/ICCV.2019.00859.

[26] L. Tchapmi, C. Choy, I. Armeni, J. Gwak, and S. Savarese, "SEGCloud: Semantic segmentation of 3D point clouds," *Proc. - 2017 Int. Conf. 3D Vision, 3DV 2017*, pp. 537–547, 2018, doi:10.1109/3DV.2017.00067.

[27] D. Rethage, J. Wald, J. Sturm, N. Navab, and F. Tombari, "Fully-Convolutional Point Networks for Large-Scale Point Clouds," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 11208 LNCS, pp. 625–640, 2018, doi:10.1007/978-3-030-01225-0_37.

[28] A. Dai, D. Ritchie, M. Bokeloh, S. Reed, J. Sturm, and M. Niebner, "ScanComplete: Large-Scale Scene Completion and Semantic Segmentation for 3D Scans," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pp. 4578–4587, 2018, doi:10.1109/CVPR.2018.00481.

[29] K. Babacan, L. Chen, and G. Sohn, "Semantic Segmentation Of Indoor Point Clouds Using Convolutional Neural Network," *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.*, vol. 4, no. 4W4, pp. 101–108, 2017, doi: 10.5194/isprs-annals-IV-4-W4-101-2017.

[30] Y. Rao, M. Zhang, Z. Cheng, J. Xue, J. Pu, and Z. Wang, "Semantic Point Cloud Segmentation Using Fast Deep Neural Network and DCRF," *Sensors*, vol. 21, no. 8, pp. 1–16, 2021, doi:10.3390/s21082731.

[31] H. Su, S. Maji, E. Kalogerakis, and E. Learned-Miller, "Multi-view convolutional neural networks for 3D shape recognition," *Proc. IEEE Int. Conf. Comput. Vis.*, vol. 2015 Inter, pp. 945–953, 2015, doi:10.1109/ICCV.2015.114.

[32] A. Boulch, J. Guerry, B. Le Saux, and N. Audebert, "SnapNet: 3D point cloud semantic labeling with 2D deep segmentation networks," *Comput. Graph.*, vol. 71, pp. 189–198, 2018, doi:10.1016/j.cag.2017.11.010.

[33] B. Wu, A. Wan, X. Yue, and K. Keutzer, "SqueezeSeg: Convolutional Neural Nets with Recurrent CRF for Real-Time Road-Object Segmentation from 3D LiDAR Point Cloud," in *Proceedings - IEEE International Conference on Robotics and Automation*, 2018, pp. 1887–1893. doi: 10.1109/ICRA.2018.8462926.

[34] A. Milioto, I. Vizzo, J. Behley, and C. Stachniss, "RangeNet ++: Fast and Accurate LiDAR Semantic Segmentation," *IEEE Int. Conf. Intell. Robot. Syst.*, no. i, pp. 4213–4220, 2019, doi:10.1109/IROS40897.2019.8967762.

[35] C. R. Qi, H. Su, M. Niebner, A. Dai, M. Yan, and L. J. Guibas, "Volumetric and multi-view CNNs for object classification on 3D data," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2016-Decem, pp. 5648–5656, 2016, doi:10.1109/CVPR.2016.609.

[36] A. Dai and M. Nießner, "3DMV: Joint 3D-multi-view prediction for 3D semantic scene segmentation," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 11214 LNCS, pp. 458–474, 2018, doi: 10.1007/978-3-030-01249-6_28.

[37] B. Zhang, S. Huang, W. Shen, and Z. Wei, "Explaining the pointnet: What has been learned inside the pointnet?," *IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Work.*, vol. 2019-June, pp. 71–74, 2019.

[38] A. Zaganidis, L. Sun, T. Duckett, and G. Cielniak, "Integrating Deep Semantic Segmentation into 3-D Point Cloud Registration," *IEEE Robot. Autom. Lett.*, vol. 3, no. 4, pp. 2942–2949, 2018, doi:10.1109/LRA.2018.2848308.

[39] H. Zhao, L. Jiang, J. Jia, P. Torr, and V. Koltun, "Point Transformer," *Proc. IEEE Int. Conf. Comput. Vis.*, pp. 16239–16248, 2021, doi: 10.1109/ICCV48922.2021.01595.

[40] A. Komarichev, Z. Zhong, and J. Hua, "A-CNN: Annularly convolutional neural networks on point clouds," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2019-June, pp. 7413–7422, 2019, doi: 10.1109/CVPR.2019.00760.

[41] Daniel, "CloudCompare Wiki," 2015. https://www.cloudcompare.org/doc/wiki/index.php/Main_Page (accessed Jan. 01, 2023).

[42] M. Weinmann, B. Jutzi, C. Mallet, and M. Weinmann, "Geometric Features And Their Relevance For 3D Point Cloud Classification," *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.*, vol. 4, no. 1W1, pp. 157–164, 2017, doi: 10.5194/isprs-annals-IV-1-W1-157-2017.

[43] X. Yan, "Pointnet/Pointnet++ Pytorch," *https://github.com/yanx27/Pointnet_Pointnet2_pytorch*, 2019. (accessed Jan. 05, 2023).

[44] M. A. Rahman and Y. Wang, "Optimizing intersection-over-union in deep neural networks for image segmentation," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 10072 LNCS, pp. 234–244, 2016, doi:10.1007/978-3-319-50835-1_22.

[45] C. Wijaya, "Point Cloud Semantic Segmentation for Indoor Modeling Using PointNet," Unpublished Master Thesis of Geomatics Engineering, Department of Geodetic Engineering, Universitas Gadjah Mada, 2023.

[46] L. Loic and M. Simonovsky, "Large Scale PCSS with Superpoint Graphs," *Comput. Vis. Pap. Res.*, pp. 4558–4567, 2018, doi: 10.1109/CVPR.2018.00479.

[47] M. Weinmann, A. Schmidt, C. Mallet, S. Hinz, F. Rottensteiner, and B. Jutzi, "Contextual classification of point cloud data by exploiting individual 3D neigbourhoods," *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.*, vol. 2, no. 3W4, pp. 271–278, 2015, doi:10.5194/isprsannals-II-3-W4-271-2015.

[48] J. Zhang, X. Lin, and X. Ning, "SVM-Based classification of segmented airborne LiDAR point clouds in urban areas," *Remote Sens.*, vol. 5, no. 8, pp. 3749–3775, 2013, doi: 10.3390/rs5083749.

[49] A. Boulch, B. Le Saux, and N. Audebert, "Unstructured point cloud semantic labeling using deep segmentation networks," *Eurographics Work. 3D Object Retrieval, EG 3DOR*, vol. 2017-April, pp. 17–24, 2017, doi: 10.2312/3dor.20171047.

[50] H. S. Koppula, A. Anand, T. Joachims, and A. Saxena, "Semantic labeling of 3D point clouds for indoor scenes," *Adv. Neural Inf. Process. Syst. 24 25th Annu. Conf. Neural Inf. Process. Syst. 2011, NIPS 2011*, pp. 1–9, 2011.

[51] B. Haznedar, R. Bayraktar, A. E. Ozturk, and Y. Arayici, "Implementing PointNet for point cloud segmentation in the heritage context," *Herit. Sci.*, vol. 11, no. 1, pp. 1–18, 2023, doi:10.1186/s40494-022-00844-w.

[52] A. Nurunnabi, F. N. Teferle, J. Li, R. C. Lindenbergh, and S. Parvaz, "Investigation Of Pointnet For Semantic Segmentation Of Large-Scale Outdoor Point Clouds," in *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives*, 2021, vol. 46, no. 4/W5-2021, pp. 397–404. doi:10.5194/isprs-Archives-XLVI-4-W5-2021-397-2021.