SCIENCE LETTERS

Available online at https://scilett-fsg.uitm.edu.my/

ISSN: 1675-7785
eISSN: 2682-8626

Science Letters 18(4) 2024, 1-6.

A 3D Point Cloud Human Body Parts Segmentation Using U-Net Based on Spherical Projection

Hanafie Rizal Fathoni, Triana Harmini and Oddy Virgantara Putra*

Departement of Informatics, Faculty of Science and Technology, Universitas Darussalam Gontor, Jl. Raya Siman Km. 5 Siman, Ponorogo, East Java, Indonesia

ARTICLE INFO

Article history:
Received 30 April 2024
Revised 31 August 2024
Accepted 30 September 2024
Online first
Published 28 October 2024

Keywords:
Lidar
Point Cloud
Segmentation
Spherical Projection
U-Net

DOI:
10.24191/sl.v18i4.9661

ABSTRACT

Light Detection and Ranging (LiDAR) has become crucial in human activity recognition. LiDAR reflects and receives laser beams from humans, including surrounding objects, visualizing the reflected results as a 3D Point Cloud. This often faces challenges like object shape variations and noise that can cause overlap, hindering segmentation and object recognition accuracy. The spherical projection approach is used to address these issues. This paper proposes a spherical projection for the segmentation approach, which comes as an answer. The methodology employed in the study focuses on effectively partitioning data derived from LiDAR technology, leveraging a unique spherical projection technique. This innovative approach involves the transformation of a three-dimensional Point Cloud into a two-dimensional projection, facilitating more efficient segmentation processes. Integrating U-Net, a deep learning model recognized for its prowess in image segmentation tasks, aims to enhance the accuracy and precision of delineating human body parts from LiDAR-generated data. The proposed method contributes to advancing the segmentation technique field and highlights the potential applications of U-Net in LiDAR data preprocessing. The model achieved an accuracy of 99.74% and Intersection Over Union (IoU) results of 87.92%.

INTRODUCTION

Technological advances have recently increased with technologies such as Light Detection and Ranging (LiDAR) sensors. LiDAR technology has been widely utilized to identify urban features, including roads and buildings [1], and is not limited to these uses. The application of this technology has penetrated various fields, especially the recognition of human activities [2]. The LiDAR sensor shoots a laser beam quickly to a surface and calculates the distance from the laser transmitter to the surface of an object, then calculates how long it takes for the laser light to return to the receptor. The visualization results of the reflection are called a 3D Point Cloud, a collection of 3D points gathered to form or describe an object.

* Corresponding author. *E-mail address*: oddy@unida.gontor.ac.id
https://doi.org/10.24191/sl.v18i4.9661

However, segmenting the human body using LiDAR laser capture data has its challenges because the 3D Point Cloud model data has unique data characteristics contained in the 3D Point Cloud model. One of the challenges is that the objects generated from the LiDAR sensor have different distances from an object to the scanning sensor. This will make it challenging to perform human body segmentation techniques resulting from LiDAR sensors due to the unorganized point cloud or 3D Point Cloud format [3]. Because the 3D *point cloud* has a variety of object shapes, noise can cause overlap and potentially hamper accuracy in performing segmentation techniques. Also, it always faces challenges such as noise and clutter interference [4]. To address these issues, *spherical projection* is a solution for *3D Point Cloud-based* human body segmentation. *Spherical Projection* is a technique of transforming data from 3D to 2D for segmentation needs. After 3D data is transformed into 2D data, the results of 2D can be applied to the segmentation techniques using the U-Net model to get good accuracy results in segmenting the human body.

The U-Net model in the CNN method has become the best method for image segmentation and classification of 3D point cloud objects from LiDAR [5]. Segmentation of instances in the image is part of the image processing process [6]. Image segmentation divides an image into various homogeneous regions based on certain similarity criteria between pixels and their neighbors. The output of this segmentation process is then used for advanced steps at a higher level in image analysis, such as image classification, image identification, and other processes [7]. This paper is divided into four sections. The first section will contain an introduction to this research, the second section will explain the materials and research methods, and the third section will explain the results of the discussion and conclusions of the researcher.

EXPERIMENTAL

The main objective of this research is to get accurate results in segmenting the human body using 3D *Point Cloud* data from *LiDAR* sensors using the U-Net method. The U-Net method effectively segments images by using encoders and decoders in the U-Net method. To achieve good performance in segmentation, the U-Net network model is applied to 3D data transformed into 2D data through Spherical Projection, as detailed in Fig. 1.

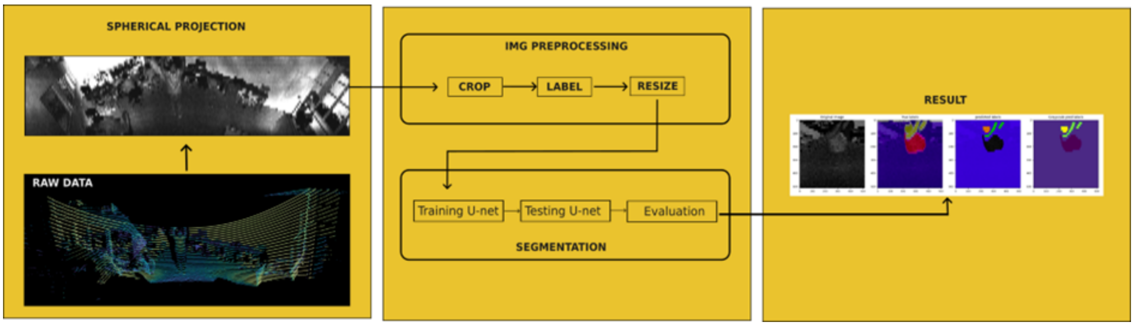

Fig. 1. The flow of 3D Point Cloud segmentation after passing the spherical projection stage is continued to the U-Net model training stage

Dataset

This study uses mid-range LiDAR OUSTER technology with type OS1-32 to collect the main data on human poses. The sensor has 32 rows of scanning, 2048 points in one column, a 45° field of view (FOV), 0.5 cm precision, and a 90-meter range. The subject used was a person who raised both hands, rotated them at 360^{0}, and recorded them for 9 seconds. The recorded data is outputted in Packet Capture (PCAP) format, and the resulting point cloud was visualized using Ouster Studio; the recorded duration output comes in

https://doi.org/10.24191/sl.v18i4.9661

Packet Capture (PCAP) format, and the resulting point cloud was visualized using Ouster Studio, as illustrated in Fig. 2.

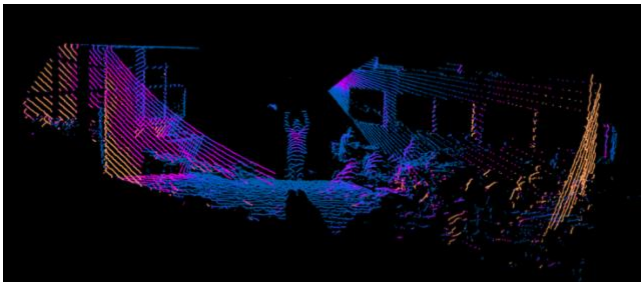

Fig. 2. PCAP file visualization results using Ouster Studio

In a recording of 9 seconds, the 3D data was converted into 2D image data for each transfer of the video frame, with a width of 2048 pixels and a height of 32 pixels. To produce 71 image data, manually label 71 image data and then proceed to the next stage. The mapping of the labeled dataset results is shown in Table 1. Data labeled and put into each folder is ready to be processed. Data processing is done on Kaggle. Therefore, for data to be processed, the author uploads the data to Kaggle as a dataset.

Table 1. Result dataset

Data	Real Image	Mask Image	Total Image
Data Train	71	71	142
Data Test	71	71	142
Data Val	71	71	142

Spherical Projection

This section describes transforming a 360^0-point cloud into a 2D spherical image data incorporated into the subsequent spherical view model block, as proposed by Ali Alnaggar [8]. At first, the 3D *point cloud* is mapped from *Euclidean* space (*x, y, z*) to spherical space (θ, φ, r) by applying equation 1. Next, the points are embedded into the 2D spherical image with dimensions (h, w) by discretizing the angles θ and φ using equation 2.

$$\begin{pmatrix}\theta\\ \emptyset\\ r\end{pmatrix} = \begin{pmatrix} arcsin\left(\frac{z}{\sqrt{x^2+y^2+z^2}}\right)\\ arctan(y,x)\\ \sqrt{x^2+y^2+z^2}\end{pmatrix} \quad (1) \qquad \begin{pmatrix}u\\ v\end{pmatrix} = \begin{pmatrix}\frac{1}{2}[1-\emptyset\pi^{-1}]w\\ [1-(\theta+f_{up})f^{-1}]h\end{pmatrix} \quad (2)$$

Where u and v represent the indices of points in the spherical image, and $f = f_{up} + f_{down}$ is the vertical field of view of the sensor. The mapping and discretization steps may result in some 3D points having the same u and v values. To solve this situation, 3D points closer to the LiDAR sensor are prioritized to be presented as 2D images by sorting them in descending order based on distance values.

Architecture U-Net

This research uses the Convolutional Neural Network method with U-Net architecture to segment 3D LiDAR image data converted into 2D format data. Researchers use Unet because Unet has been proven efficient in performing or handling various image segmentation needs [9]. This research uses 4 encoder-decoder blocks. Encoder with 2 convolution layers, 2 batch normalization, and 2 activation layers with relu activation. The decoder has 1 convolution transpose layer, 1 skip connection layer, 2 convolution layers, 2 batch normalization, and 2 activation screens. A block contains 2 convolution layers, 2 batch normalization layers, and 2 activation screens in the middle between the encoder-decoder. After the decoder, there is 1 convolution layer with 1x1 kernel and 1 activation with softmax. This research uses hyperparameters, as shown in Table 2.

Table 2. Result dataset

Sequential Parameter	Value
Convolutional Layer	18
Pooling Layer	4
Transpose Convolutional Layer	4
Drop out Rate	4
Activation Function	Relu, Softmax
Optimizer	Adam
Loss Function	Categorical Crossentropy
Epoch	50
Batch Size/Filter	64
Learning Rate	0.0001
Total Parameter	22.806.296

RESULTS AND DISCUSSION

Preprocessing Data

In this section, the 2D image data resulting from spherical projection with a width of 2048 pixels and a height of 32 pixels is cropped to a width of 126 pixels and a height of 32 pixels, then resize the image with a resolution of 512x512 pixels. This can be used so the trained model can produce maximum performance in testing and validation. The resized data is shown in Fig. 3.

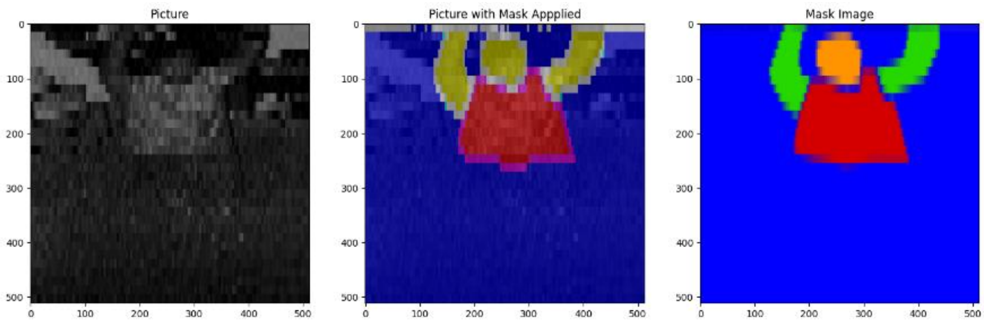

Fig. 3. Dataset of 3D to 2D transform result. (Left: Original image, Right: Image segmentation label)

The training process carried out in this research is run on Kaggle using the GPU that Kaggle has provided for free by implementing the U-Net model and hyper-parameters. Testing using the U-Net model with a relu activation function, epoch 50, can produce an accuracy of around 99.77%, as shown in Fig. 4, and the model evaluation results table is shown in Table 3.

https://doi.org/10.24191/sl.v18i4.9661

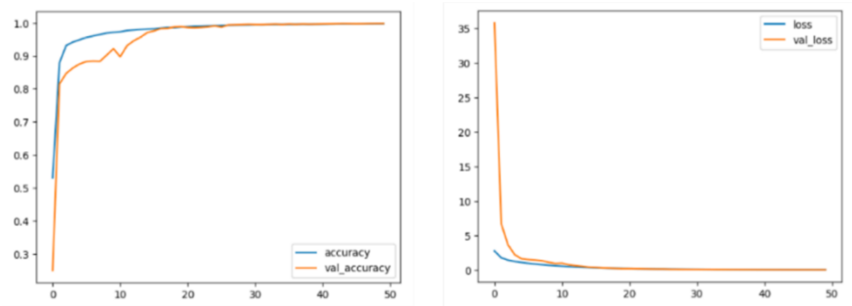

Fig. 4. Training graph using the U-Net model

Table 3. U-Net model evaluation results

Model Evaluation	Value
Vall Accuracy	0.9983
Accuracy	0.9977
Vall Loss	0.0282
Loss	0.0314
IoU	0.8792

The trained U-Net model is predicated on random images. The predicted image is compared with the label. The comparison image of the prediction and label is shown in Fig. 5.

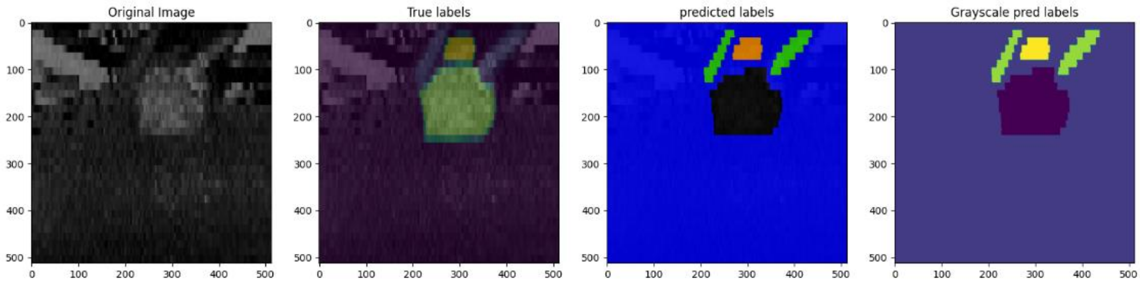

Fig. 5. Image segmentation prediction results using the U-Net model

CONCLUSION

This paper presents a 3D point cloud segmentation technique generated from LiDAR technology by converting 3D data into 2D data through the spherical projection technique. Then, the 3D data converted into 2D is continued to the data training and prediction stage using the U-Net model to get maximum accuracy results.

ACKNOWLEDGMENTS

The authors thank the Universitas Darussalam Gontor for providing this research's facilities and financial support.

AUTHOR'S CONTRIBUTION

Hanafie Rizal Fathoni conducted the research, and Oddy Virgantara Putra and Triana Harmini were the supervisors.

CONFLICT OF INTEREST STATEMENT

The authors affirm that there are no conflicts of interest related to the publication of this manuscript.

REFERENCES

[1] Liu, D., Chen, C., Xu, C., Qiu, R. C., & Chu, L. (2023). Self-Supervised Point Cloud Registration with Deep Versatile Descriptors for Intelligent Driving. *IEEE Transactions on Intelligent Transportation Systems*, *24*(9), 9767–9779. https://doi.org/10.1109/TITS.2023.3268273

[2] Rinchi, O., Ghazzai, H., Alsharoa, A., & Massoud, Y. (2023). LiDAR Technology for Human Activity Recognition: Outlooks and Challenges. *IEEE Internet of Things Magazine*, *6*(2), 143–150. https://doi.org/10.1109/iotm.001.2200199

[3] Lin, Y., Wang, C., Cheng, J., Chen, B., Jia, F., Chen, Z., & Li, J. (2015). Line segment extraction for large scale unorganized point clouds. *ISPRS Journal of Photogrammetry and Remote Sensing*, *102*, 172–183. https://doi.org/10.1016/j.isprsjprs.2014.12.027

[4] Putra, O.V., Riansyah, M.I., Rahmanti, F.Z., Priyadi, A., Wulandari, D.P., Ogata, K., Yuniarno, E.M., & Purnomo, M.H. (2024). Enhancing LiDAR-Based Object Recognition Through a Novel Denoising and Modified GDANet Framework. *IEEE Access, 12*, 7285-7297.

[5] Putra, O. V., Riansyah, M. I., Riandini, Priyadi, A., Yuniarno, E. M., & Purnomo, M. H. (2023). Fuzzy Lightweight CNN for Point Cloud Object Classification based on Voxel. *TENCON 2023 - 2023 IEEE Region 10 Conference (TENCON)*, 685–690. https://doi.org/10.1109/TENCON58879.2023.10322519

[6] Nezla, N. A., Mithun Haridas, T. P., & Supriya, M. H. (2021). Semantic Segmentation of Underwater Images using UNet architecture based Deep Convolutional Encoder Decoder Model. *2021 7th International Conference on Advanced Computing and Communication Systems, ICACCS 2021*, *March*, 28–33. https://doi.org/10.1109/ICACCS51430.2021.9441804

[7] Delibasoglu, I., & Cetin, M. (2020). Building segmentation with Inception-Unet and classical methods. *2020 28th Signal Processing and Communications Applications Conference, SIU 2020 - Proceedings*, *October 2020*. https://doi.org/10.1109/SIU49456.2020.9302155

[8] Ali Alnaggar, Y., Afifi, M., Amer, K., & Elhelw, M. (2021). Multi projection fusion for real-time semantic segmentation of 3D LiDAR point clouds. *Proceedings - 2021 IEEE Winter Conference on Applications of Computer Vision, WACV 2021*, 1799–1808. https://doi.org/10.1109/WACV48630.2021.00184

[9] Mcglinchy, J., Johnson, B., Muller, B., Joseph, M., & Diaz, J. (2019). Application Of Unet Fully Convolutional Neural Network to Impervious Surface Segmentation in Urban Environment From High Resolution Satellit Imagery University of Colorado at Boulder , CIRES Earth Lab Analytics Hub University of Colorado at Boulder , Pro. *IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium*, 3915–3918.