A Review of Point Cloud 3D Object Detection Methods Based on Deep Learning

Xiyuan Wang [1] [0009-0009-8927-7359], Jie Lin*[1] [0009-0006-2516-3692],

Longrui Yang[1][0009-0003-9466-9513], Sicong Wang[1] [0009-0009-8367-9146]

[1] College of Electrical Engineering and Information Engineering, Lanzhou University of Technology Lanzhou,730050, China

*449066528@qq.com

Abstract. Based on introducing the coupling relationship between deep learning and three-dimensional point clouds, this paper reviews the three characteristics and research problems of point clouds, randomness, sparsity, and unstructuredness, and discusses three-dimensional point cloud target detection based on deep neural networks, including point cloud detection techniques following graph convolution, detection techniques following the original point cloud, and detection algorithms based on fusion processing of graph convolution and the original point cloud. Focusing on future research direction and development, the field of point cloud analysis is currently undergoing further development through the application of deep learning techniques.

Keywords: 3D Point Cloud,Deep Leaning, Semantic Segmentation,Target Detection.

1 Introduction

Three-dimensional target detection[1-3] is a complex process of interpret-ing environmental information and the spatial location of the target based on the geometric model and mathematical information obtained by three-dimensional intelligent sensors. In this way, we could obtain profitable data such as the type, location, and direction of motion of the desired target. In re-cent years, with the increasing requirement for the diagnostic accuracy and stabilization of the detection of objects in three-dimensional space, traditional two-dimensional object detection methods have difficulty meeting the needs of researchers. At present, three-dimensional object detection technology based on deep neural frameworks features the advantages of economical acquisition, fast recognition speed, and exact measurements, which provide great help for feature extraction and accurate classification of point clouds[4-8]. Due to the swift progress of three-dimensional innovation in recent years, three-dimensional object detection has gradually become an important research field and hot industry for researchers[9-10].

The perceptual images and 3D point clouds obtained based on lidar have difficulties such as sparsity, disorder, and unstructured, which makes it difficult for researchers to further decompose and process the original point clouds. Over the past few years, the combination of deep learning and 3D laser point clouds has been continuously valued. and supported by researchers. [11-13] Through its rich, diverse, and nonstructural characteristics, point clou. d data have been efficiently used in feature processing and feature classification and have achieved rapid development in 3D data processing based on point clouds. Concurrently, due to its excellent classification accuracy index and real-time performance, it has attracted the attention of scholars in many fields, such as autonomous driving[14-15], model reconstruction[16], and ground inspection[17].

2 Relationship Between Deep Learning and Point Cloud

The process of recognizing and categorizing different semantic regions within a point cloud is known as three-dimensional point cloud object recognition. In the research of traditional point cloud classification methods, most of the features used by researchers are based on the three-dimensional structure of the local areas of the point cloud, such as lattice constant, curvature, normal vector, and spatial distribution. For this reason, researchers have developed many point descriptors and selected the appropriate point cloud classifier for predicting the semantic labels of the point cloud, such as support vector machine[18], random forest[19], AddBoost[20], Gaussian mixture model[21], and JointBoost[22]. However, this kind of artificial extraction method mainly depends on the judg-ment of the experimenter and does not fully consider the objectivity and stabil-ity of the adjacent point cloud. Alterations to this aspect can significantly im-pact the classification outcomes of the point cloud. Later, researchers devel-oped methods to correlate data information before and. after, such as the Mar-kov random field[23-25] and discriminative random field[26-27]. These meth-ods effectively improve the classification effect and reduce the cost of task completion. These methods also show different levels of constraint effects un-der different scenarios and constraints, and the universality and accuracy in complex environments are not. strong, which is also a major difficulty in re-search classification and model d. esign[28]. Over the past few years, ad-vancements in iterative updates for computer arithmetic processing, as well as the creation and expansion of a large three-dimensional scene database, have accelerated the development and implementation of. deep learning tools within three-dimensional point cloud areas. As a result, traditional feature classifiers and classification techniques are becoming outdated and are being replaced by newer, more advanced skills. Initially,. researchers usually use the method of projecting the original point cloud data onto a two-dimensional image when preprocessing the point cloud. This method[29-33] of three-dimensional projec-tion to two-dimensional projection often loses critical information in the process of processing, which eventually leads to incomplete model training and deci-sion-making errors, which greatly limits the

performance of the model. For this reason, researchers decided to start from the source of the three-dimensional point cloud and directly input the original point cloud to avoid losing key infor-mation. In 2017, Reference[34] first proposed a new network architecture called PointNet, which was introduced in the field of point cloud analysis. This network is unique in that it can directly process raw point clouds and has since gained widespread attention and usage in various applications, such as point cloud categorization, semantic segmentation, and object detection.

Deep learning has emerged as a powerful technique for point cloud 3D object detection, enabling impressive advancements in various domains such as scene classification and object detection. Aiming at the method of point cloud 3D object detection based on deep learning, this paper uses a knowledge graph to collect data, process information and summarize conclusions and expounds the structural characteristics and research contents of this hot field to study. and reveal the development trend and prospect of this direction in the field over the last couple of years. Based on data from the Web of Science(WoS), this paper summarizes and plans the papers from 2016 to 2023. The keywords are'Three-dimensional Point Cloud ','Point Cloud Based on Deep Learning ','Object Recognition from Point Cloud ', and' Semantic Labelling of Point Cloud '. A total of 1067 Chinese literature and 1193 English literature records (SCI source journ. als, EI source journals, CCSCI WoS core set) were obtained, including many articles and reviews. Through literature collation, refinement, and filtering, 366 closely related Chinese studies and 535 closely related English studies were finally obtained.

By analysing the search results and refining the keyword content, the research hotspots are visually displayed. Figure 1 shows the research hotspots of papers obtained from WoS. In Figure 1, the size of the circle and the density between the associated lines directly represent the degree of the close correlation between research ho. tspots. In this way, the links between hot research fields are effectively clustered together.

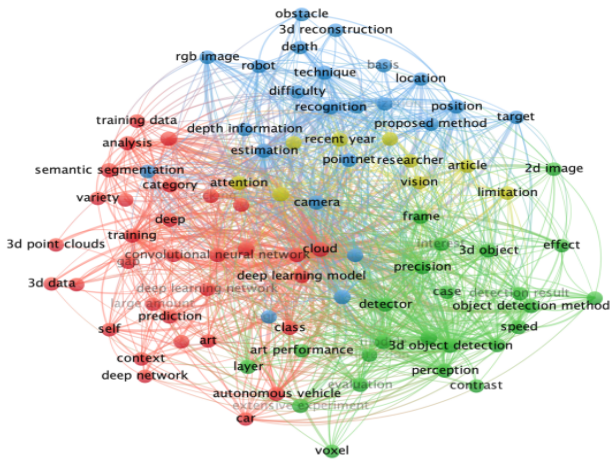

Fig. 1. WoS keyword co-occurrence analysis

Through visual analysis, it can be seen that the exploration in this branch mainly includes 3D laser point clouds, deep neural networks, 3D target detec-tion, and point cloud processing. At present, some review articles have discussed and summarized point cloud target detection based on deep learning[35]. Based on previous work and existing research, this paper enriches and improves the content summary of the deep learning method for 3D target detection in the point cloud field.

3 Point Cloud Processing Method

In recent years, the training and information processing of 3D laser point clouds[36-39] has become a research hotspot at home and abroad. The algo-rithms for detecting three-dimensional objects from point clouds based on deep learning can be divided into three categories: classification processing methods based on graph convolution, object detection methods based on the origina. l point cloud, and detection. methods based on fusion graph features and original point cloud information.

3.1 Based on Graph Convolution

The early VoxNet algorithm has many problems, such as high computational and memory costs and large model capacity constraints. The processing effect of the depth map has not been improved. In 2018, Reference[40] improved and proposed a

VoxelNet with end-to-end results on this basis, which not only improves the connection between the front and back features of the point cloud in the environment but also expands the learning range of the broader visual features, as shown in Figure 2. In the same year, Zhang proposed graph convo-lutional neural networks, Graph-CNN (Point-VGG), by using graph convolution and point cloud downsampling. The network combines the comparison of the features of convolutional kernels, max-pooling layers, and densely connected layers. Reference[41] proposed Point-GNN, which connects and aggregates the feature relationship between adjacent regions and point cloud center points, to perform point cloud target detection quickly and effectively.

In addition, DGCNN[42], MVCNN[43], and other algorithms are similar to the methods of traditional convolutional neural network models, convolution operations are performed directly on point cloud images, and good experi-mental results are also obtained. Figure 3 displays the flow chart for the MVCNN model.

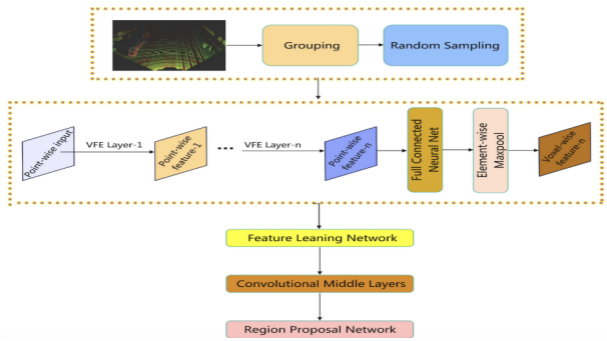

Fig. 2. Voxel Network Framework

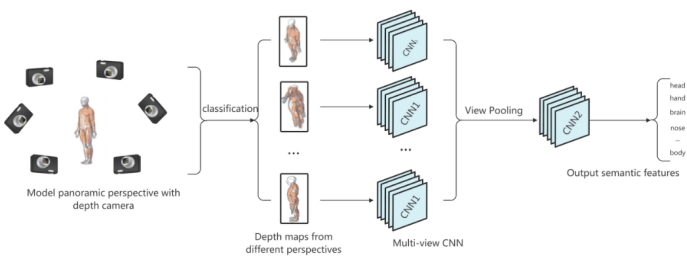

Fig. 3. MVCNN Network Framework

3.2 Based on Original Point Cloud

Based on the point cloud data of image processing, there are always prob-lems such as information loss. Therefore, researchers decided to process the original point cloud data directly. In the early stage, Reference[36] pioneered the point cloud feature learning network PointNet, which used the spatial trans-formation network to solve the problem of point cloud rotation invariance. This contributed significantly to the establishment of fundamental principles for the study of original point cloud manipulation. Therefore, researchers have pro-posed more feature classification networks based on PointNet++, such as PointConv, PointWeb, and PointRCNN[44-45]. The algorithm flow is shown in Figure 4 and Figure 5. Reference[35] proposes an adaptive feature learning mechanism to automatically extract and learn point cloud features by extract-ing local context information. Based on the pooling function aggregation of the original point cloud, Reference[46] predicted the extracted semantic features combined with spatial features and refined the specific location of the detection target in the three-dimensional coordinate system.

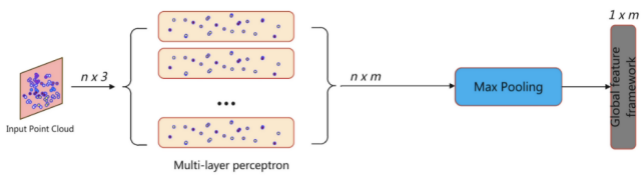

Fig. 4. PointNet Network Framework

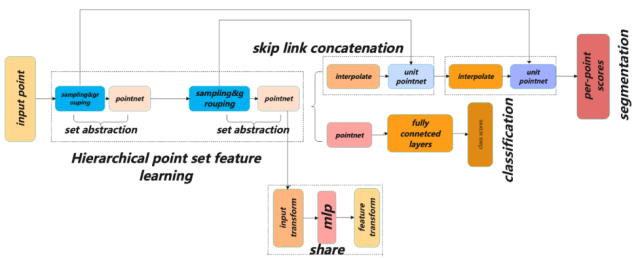

Fig. 5. PointNet++ Network Framework

3.3 Fusion Processing Algorithm

It can be seen from the above methods that both graph convolution-based and original point cloud processing-based methods can obtain rich semantic features and obtain extraction effects at different speeds according to the semantic resolution. Therefore, researchers have proposed a fusion algorithm that combines the ad-

vantages of graph convolution and original point clouds and uses image processing as an auxiliary to fuse the results of image data and point cloud detection on the existing 3D object detection technology.

The MV3D model proposed by Reference[47] in the early stage is one of the mainstream methods of early fusion processing. The convolution layer is used to analyse both the top-down and forward-facing views of the bird's. perspective, fuse with the depth RGB image through the pooling function, predict the target category and return the detection box. Due to the low accuracy of the early MV3D model and the loss of key information, reference [23] proposed the AVOD neural network structure model. The algorithm flow is shown in Figure 6. More underlying information and semantic information are restored, and the model accuracy is further improved. Inspired by graph convolution and semantic segmentation, Reference[49] proposed a multitask multisensor detection model in 2019. This method fuses the ROI pooling region with the features detected by the sensor system and uses the deep convolution network to complete the image information to obtain better point cloud feature fusion information. In 2020, Reference[48] unified the image of the deep learning network with the three-dimensional target data of the point cloud. Based on a new decision function Hough, the upsampling points and downsampling points of the point cloud are distinguished, and the geometric characteristics of the red–green-blue (RGB) visual data and the point cloud are combined or merged.

4 Generalize

Utilizing deep learning techniques for the identification of three-dimensional point clouds not only improves the operation ability of the system to process point cloud information and strengthens the flexibility and expansibility of a detection system but also improves the participation ability of the original point cloud data infor-mation. As an emerging area of study, this research direction will promote the rapid development of the future point cloud field, which has great development space and application advantages. Based on the quantitative analysis of the literature from 2018 to 2023 obtained by WoS, this paper first demonstrates the coupling relation-ship between deep learning and point cloud research from different perspectives, including semantic segmentation, target detection, and point cloud classification. This paper expounds on the advantages and research progress of the exploration of deep learning-based methodologies for researching three-dimensional point cloud data, discusses the development process of point cloud feature extraction and deep learning utilized for detecting targets and focuses on the main position and role of this field in the whole point cloud research field. In this paper, based on different processing methods, the three-dimensional point cloud target detection method based on deep learning atmosphere based on graph convolution processing method, based on the original point cloud processing approach and. based on graph convolution and original point cloud data fusion processing method in different categories show their respective advantages and disadvantages.

In the future, including larger and more complex environmental point cloud data technology, how to support the rapid development of point cloud research with more efficient, more accurate, and more intelligent methods and technologies and realize the full mining of point cloud value through the coupling of deep learning technology is a research topic with broad scientific research prospects and major national needs in the future.

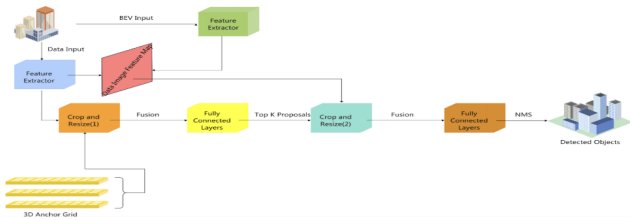

Fig. 6. AVOD Network Framework

5 Prospect

In general, the current methods based on graph convolution, original point cloud, or fusion processing have achieved different advantages. The method based on graph convolution has high detection efficiency and fast recognition. Converting a three-dimensional point cloud into a two-dimensional image can significantly reduce the time required for processing but also loses much key information. The method based on the original point cloud retains all key information well, increases the opera-tion cost and is difficult to calculate. The method based on fusion processing is cur-rently a relatively safe and efficient method. It uses image information as an aid to retain the original information of the point cloud while ensuring high efficiency. Based on the current research progress and problems, this paper proposes the follow-ing suggestions for future research:

- The input of the original point cloud is preprocessed to retain key information while removing redundant scene information. To enhance both the accuracy and efficiency of detection...
- The filter and recognition framework is redesigned for the point cloud effect of image processing, and the accuracy of the two-dimensional converter is opti-mized to achieve higher precision levels for the model.
- Because the hardware computing cost is too high and the image auxiliary func-tion of fusion processing makes up for this shortcoming well, the fusion pro-cessing method will be a main development trend and the current research trend in target detection using three-dimensional point clouds.

Reference

1. Li, B., Ouyang, W., Sheng, L., et al. (2019). Gs3D: An efficient 3D object detection framework for autonomous driving. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 1019-1028).
2. Zhou, Y., & Tuzel, O. (2018). Voxelnet: End-to-end learning for point cloud based 3D object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 4490-4499).
3. Ku, J., Mozifian, M., Lee, J., et al. (2018). Joint 3D proposal generation and object detection from view aggregation. In 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) (pp. 1-8). IEEE.
4. Girshick, R. (2015). Fast R-CNN. In IEEE International Conference on Computer Vision (ICCV) (pp. 1440-1448).
5. Ren, S., He, K., Girshick, R., et al. (2015). Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. IEEE Transactions on Pattern Analysis and Machine Intelligence, 39(6), 1137-1149.
6. Redmon, J., Divvala, S., Girshick, R., et al. (2015). You Only Look Once: Unified, Real-Time Object Detection. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 779-788).
7. Liu, W., Anguelov, D., Erhan, D., et al. (2016). SSD: Single Shot MultiBox Detector. In 14th European Conference on Computer Vision (ECCV) (pp. 21-37).
8. Redmon, J., & Farhadi, A. (2017). YOLO9000: Better, Faster, Stronger. In 30th IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 6517-6525).
9. Ma, X., & Hovy, E. (2016). End-to-End Sequence Labelling via Bi-Directional LSTM-CNNs-CRF. In 54th Annual Meeting of the Association for Computational Linguistics (ACL) (pp. 1064-1074).
10. Yoon, S., & Kim, E. (2017). Temporal Classification Error Compensation of Convolutional Neural Network for Traffic Sign Recognition. In International Conference on Control Engineering and Artificial Intelligence (CCEAI).
11. Zhou, Y., & Tuzel, O. (2018). VoxelNet: End-to-End Learning for Point Cloud Based 3D Object Detection. In IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 4490-4499).
12. Chen, X., Ma, H., Wan, J., Li, B., & Xia, T. (2017). Multi-View 3D Object Detection Network for Autonomous Driving. In 30th IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 6526-6534).
13. Qi, C. R., Su, H., Mo, K., & Guibas, L. J. (2017). PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation. In 30th IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 77-85).
14. Kim, K., Kim, C., Jang, C., Kim, J., & Kim, H. (2020). Deep learning-based dynamic object classification using LiDAR point cloud augmented by layer-based accumulation for intelligent vehicles. Expert Systems with Applications, 113861.
15. Zermas, D., Izzat, I., & Papanikolopoulos, N. (2017). Fast segmentation of 3D point clouds: A paradigm on lidar data for autonomous vehicle applications. In 2017 IEEE International Conference on Robotics and Automation (ICRA) (pp. 5067-5073).
16. Bisheng, Y., Ronggang, H., Jianping, L., Jian, Y., & Jiayuan, L. (2016). Automated reconstruction of building LoDs from airborne LiDAR point clouds using an improved morphological scale space. Remote Sensing, 9(1), 14.
17. Ene, L. T., Næsset, E., Gobakken, T., & Gregoire, T. G. (2017). Large-scale estimation of change in aboveground biomass in miombo woodlands using airborne laser scanning and national forest inventory data. Remote Sensing of Environment, 188, 106-117.
18. Chen, C., Li, X., Belkacem, A. N., Zhang, H., & Xiang, S. (2019). The mixed kernel function SVM-based point cloud classification. International Journal of Precision Engineering and Manufacturing, 20(5), 737-747.
19. Ni, H., Lin, X., & Zhang, J. (2017). Classification of ALS point cloud with improved point cloud segmentation and random forests. Remote Sensing, 9(3), 288.
20. Weinmann, M., Jutzi, B., Hinz, S., & Mallet, C. (2015). Semantic point cloud interpretation based on

optimal neighborhoods, relevant features and efficient classifiers. ISPRS Journal of Photogrammetry and Remote Sensing, 105(7), 286-304.

21. Chan, C. W., & Paelinckx, D. (2008). Evaluation of random forest and adaboost tree-based ensemble classification and spectral band selection for ecotope mapping using airborne hyperspectral imagery. Remote Sensing of Environment, 112(6), 2999-3011.
22. Lalonde, J. F., Unnikrishnan, R., Vandapel, N., & Hebert, M. (2005). Scale selection for classification of point-sampled 3D surfaces. In The Fifth International Conference on 3D Digital Imaging and Modelling (3DIM'05) (pp. 285-292). IEEE.
23. Han, Y., Sun, H., Lu, Y., Zhong, R., Ji, C., & Xie, S. (2022). 3D Point Cloud Generation Based on Multi-Sensor Fusion. Applied Sciences, 12(19).
24. Niemeyer, J., Rottensteiner, F., & Soergel, U. (2014). Contextual classification of LiDAR data and building object detection in urban areas. ISPRS Journal of Photogrammetry and Remote Sensing, 87, 152-165.
25. Munoz, D., Bagnell, J. A., Vandapel, N., & Hebert, M. (2009). Contextual classification with functional maxmargin Markov networks. In 2009 IEEE Conference on Computer Vision and Pattern Recognition (pp. 975-982).
26. Shapovalov, R., Velizhev, E., & Barinova, O. (2010). Nonassociative Markov networks for 3D point cloud classification. In International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences.
27. Munoz, D., Bagnell, J. A., Vandapel, N., & Hebert, M. (2009). Contextual classification with functional max-margin markov networks. In 2009 IEEE Conference on Computer Vision and Pattern Recognition (pp. 975-982).
28. Niemeyer, J., Rottensteiner, F., & Soergel, U. (2014). Contextual classification of lidar data and building object detection in urban areas. Isprs Journal of Photogrammetry & Remote Sensing, 87(1), 152-165.
29. Maturana, D., & Scherer, S. (2015). Voxnet: A 3D convolutional neural network for real-time object recognition. In 2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) (pp. 922-928).
30. Wu, Z., Song, S., Khosla, A., et al. (2015). 3D shapenets: A deep representation for volumetric shapes. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 1912-1920).
31. Cohen, T. S., Geiger, M., Köhler, J., et al. (2018). Spherical CNNs. arXiv preprint arXiv:1801.10130.
32. You, Y., Lou, Y., Liu, Q., et al. (2020). Pointwise rotation-invariant network with adaptive sampling and 3D spherical voxel convolution. In Proceedings of the AAAI Conference on Artificial Intelligence (pp. 12717-12724).
33. Riegler, G., Osman Ulusoy, A., & Geiger, A. (2017). Octnet: Learning deep 3D representations at high resolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 3577-3586).
34. Wang, Y., Tian, Y., Li, G., et al. (2011). A review of 3D object detection based on convolutional neural network. Pattern Recognition and Artificial Intelligence, 34(12), 1103-1119.
35. Guo, Y. L., Wang, H., Hu, Q., et al. (2019). Deep learning for 3D point clouds: A survey. arXiv preprint arXiv:1912.12033.
36. Qi, C. R., Su, H., Mo, K., et al. (2017). Pointnet: deep learning on point sets for 3D classification and segmentation. IEEE, 2017, 652-660.
37. Blanco, L., García Sellés, D., Guinau, M., et al. (2022). Machine Learning-Based Rockfalls Detection with 3D Point Clouds, Example in the Montserrat Massif (Spain). Remote Sensing, 14(17).
38. Dabetwar, S., Kulkarni, N. N., Angelosanti, M., Niezrecki, C., & Sabato, A. (2022). Sensitivity analysis of unmanned aerial vehicle-borne 3D point cloud reconstruction from infrared images. Journal of Building Engineering, 58.
39. Li, T., Zhao, Z., Luo, Y., Ruan, B., Peng, D., Cheng, L., & Shi, C. (2022). Gait Recognition Using Spatio-Temporal Information of 3D Point Cloud via Millimeter Wave Radar. Wireless Communications and Mobile Computing, 2022.
40. Maturana, D., & Scherer, S. (2015). Voxnet: A 3D convolutional neural network for real-time object recognition. In 2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) (pp. 922-928). IEEE.

41. Zhou, Y., & Tuzel, O. (2018). Voxelnet: End-to-end learning for point cloud based 3D object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 4490-4499).
42. Kalogerakis, E., Averkiou, M., Maji, S., et al. (2017). 3D shape segmentation with projective convolutional networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 3779-3788).
43. I.C.R., Su, H., Niessner, M., et al. (2016). Volumetric and Multi-View CNNs for Object Classification on 3D Data. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 5648-5656).
44. Nguyen Duc-Phong, Berg Paul, Debbabi Bilel, Nguyen Tan-Nhu, Tran Vi-Do, Nguyen Ho-Quang, Dakpé Stéphanie, & Dao Tien-Tuan. (2023). Automatic part segmentation of facial anatomies using geometric deep learning toward a computer-aided facial rehabilitation. Engineering Applications of Artificial Intelligence, 119.
45. Hao, H., Yu, J., Yin, L., Cai, G., Zhang, S., & Zhang, H. (2023). An improved PointNet++ point cloud segmentation model applied to automatic measurement method of pig body size. Computers and Electronics in Agriculture, 205.
46. Shi, S., Wang, X., & Li, H. (2019). PointRCNN: 3D object proposal generation and detection from point cloud. In Proc of the 32nd IEEE/CVF Conf on Computer Vision and Pattern Recognition (pp. 770-779). Piscataway, NJ: IEEE.
47. Chen, Y., Liu, S., Shen, X., et al. (2019). Fast point r-cnn. In Proceedings of the IEEE/CVF International Conference on Computer Vision (pp. 9775-9784).
48. Yan, Y., Mao, Y., & Li, B. (2018). Second: Sparsely embedded convolutional detection. Sensors, 18(10), 3337.
49. Mac, G., Guoy, Y., Yang, J., et al. (2018). Learning multiview representation with LSTM for 3D shape recognition and retrieval. IEEE Transactions on Multimedia, 21(5), 1169-1182.