3D Point Cloud Classification

Subjects: Computer Science, Artificial Intelligence

Contributor: Haoran Zhou, wenju wang, Gang Chen, Xiaolin Wang

Three-dimensional (3D) point cloud classification methods based on deep learning have good classification performance. The 3D point cloud is mainly collected by light detection and ranging (LiDAR) scanner, red, green, blue, and depth (RGB-D) camera, and other sensor equipment or obtained by model conversion using computer software.

Keywords: 3D point cloud classification ; convolutional neural network ; dynamic convolution

1. Background

With the rapid development of artificial intelligence and deep learning technology, breakthroughs have been made in 3D perception and understanding; one such breakthrough, the point cloud, contains abundant geometric, shape, and structure information. As shown in **Figure 1**. It is widely used in several fields, such as urban environment monitoring [1], urban form analysis [2], autonomous driving [3][4], computer vision [5][6], robotics [7], and reverse engineering modeling [8]. Research fields related to the 3D point cloud include registration, denoising, classification, target detection, and segmentation. Among these fields, the classification-based 3D point cloud is one of the most basic and significant. Hence, most research focuses on improving the accuracy and efficiency of classification-based point clouds [9][10][11][12][13][14][15].

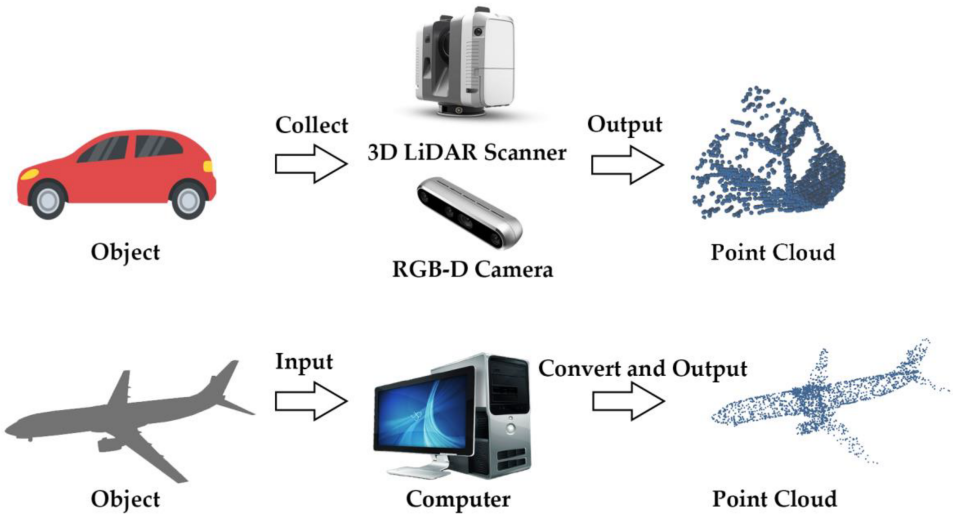

Figure 1. A series of examples of 3D point cloud acquisition.

Classification methods based on the 3D point cloud include voxel, direct point cloud, and multiview methods, which are the subject of much research.

2. Voxel-Based

Voxels, i.e., volume pixels, represent a 3D region with a constant scalar or vector [16]. Because voxels can represent complex objects with simplified and discrete units, such as particles, they have powerful capabilities in simulating the behavior of complex objects in the real world, while their structural representation is relatively simple. Plaza et al. [17] proposed a voxel-based 3D point cloud data classification method for natural environments that uses a multilayer perceptron (MLP) to conduct a statistical geometric analysis on the spatial distribution of internal points and voxel classification. The local spatial distribution characteristics around points are defined by the principal components of the point position covariance matrix. The combination of voxels and neural networks realizes faster computing speeds than other strategies; however, it does not remedy the essential shortcomings of voxels themselves, including intensive computation and time consumption. Liu et al. [18] presented a convolutional neural network (CNN) model, point-voxel, that captures the overall structure and details of an object with two modules, integrating the advantages of a point cloud and a

grid. It has less data overhead, better data regularity, and a lower memory consumption than other voxel models, as well as better classification accuracy than many voxel-based models; however, it still has the low efficiency and intrinsic stereotype of voxel convolution and the combination of voxels and a point cloud. Plaza et al. [19] proposed a general framework based on a voxel neighborhood, which was compared to different supervised learning classifiers. The framework uses simple features in the support region that are defined based on the voxel itself, and it assigns each non-overlapping voxel point in the regular grid to the same class, which improves the effectiveness of 3D space shape feature classification. Therefore, it is easier to reduce processing time and parallelize. Preliminary experiments have been conducted, and further analysis is required using diverse performance indicators, environments, and sensors. Liu et al. [20] proposed a generalized learning network based on voxels—VB-Net—and used it to classify 3D objects. By transforming the original point cloud to voxels, VB-Net can be used to extract features for the target classification of a generalized learning system. The technique significantly reduces the time required for system training, although the classification accuracy must be improved. In particular, with an increase in 3D object resolution, the classification accuracy decreases sharply. Hamada et al. [21] considered the change in voxel density along the depth direction and presented a 3D scene classification measure based on three-projection voxel expansion, which normalizes the scene according to the position and size and projects it onto three vertical planes. The algorithm applies deep learning to predict the category of each scene, combining three images, and greatly improves classification accuracy. However, three-projection voxel expansion must be standardized to make each 3D scene suitable for voxels. Apparently, if the 3D scene is large enough, then tri-projection voxel splatting (TVS) may not recognize small objects. Wang et al. [22] proposed a voxel-based CNN, NormalNet, that employs a reflection convolution concatenation (RCC) series module as a convolution layer to extract distinguishable features via a relatively good reflection number for 3D visual tasks, which significantly reduces the number of parameters and enhances network performance. Nevertheless, the model could still be further optimized since the best reflection number was not found in the key module RCC. Hui et al. [23] explored the unsuitability of binary voxels for 3D convolution representation. By assigning distance values to voxels, they improved the accuracy by about 30% and designed a fast, full connection and convolution hybrid cascade network for the classification of 3D objects. Its average reasoning time is faster than that of methods based on a point cloud and voxels, and its accuracy is also higher; however, the recognition rate of some hard samples (some sample data that take a high price to train and learn, yet usually obtain a large loss value and worse performance) in the deep network is lower than that of its shallow counterpart. Voxel-based classification has the disadvantage of a high storage overhead, and processing a large, complicated 3D image represented by direct voxels in 3D convolution requires significant graphics processing unit (GPU) resources and considerable computing time. This problem can be solved by reducing the resolution of the 3D image; however, this will decrease the accuracy of the final trained model.

3. Point Cloud-Based

A point cloud is a collection of data points defined by a coordinate system. Each point contains abundant information, including 3D coordinates (x, y, z), color, classification value, intensity, and time. This representation form has the characteristics of disorder, sparsity, rotation, and translation invariance. Compared with voxel methods, point cloud-based processing discards additional model transformations and directly considers the original point cloud as the processing object, which can reduce storage. Qi et al. [10] designed PointNet, a neural network that directly trains point clouds using a learned T-Net transformation matrix to ensure invariance in the specific spatial transformation and extract the features of point cloud data through an MLP. The final global features are obtained by maximum pooling for features in each dimension and sent to the MLP for the classification of 3D objects. Since the point cloud operates directly without complex preprocessing, the network architecture is efficient, and it resists disturbance. Yet, the network does not learn the connection among local points; hence, the model cannot capture information on local features. Zhao et al. [24] proposed a deep neural network that combines multiscale features with PointNet, adopting multiscale approaches to extract the neighborhood features of points and combining them with global features extracted by PointNet to classify LiDAR point clouds. The network has a good classification effect; however, it extracts local features with poor efficiency. Li et al. [25] proposed an optimization method based on PointNet to improve the accuracy of 3D classification models. The model can obtain more abstract features by increasing the number of hidden layers and can generate discriminant features by combining the Softmax and central loss functions. The improved model has better performance than the original PointNet. Nevertheless, it has not been studied in detail whether a more lightweight deep convolution network can be used to further optimize the model. Aiming to solve the occlusion problem of environment classification using a 1D signal and two-dimensional (2D) image, Zhang et al. [26] proposed directional PointNet to directly classify a 3D point cloud. The model utilizes the direction information of a point cloud to classify terrain in order to help wearable robot objects walk in complex environments. It provides a robust and efficient classification of the environment; however, its high classification accuracy is limited to specific application fields, and the generalization effect warrants further exploration. Because a PointNet network is unable to capture the local structure generated by metric space points, its capability to identify fine-grained

patterns and generalize complex scenes is limited. The PointNet++ hierarchical network model [11] was proposed to improve PointNet. The PointNet++ architecture adaptively combines multiscale features through a learning layer and combines the local point set density to effectively learn the point set features at a deep level for more accurate classification. The time consumption is greatly increased compared to PointNet, especially in preprocessing. Rivlin et al. [27] used 3D coordinates as class identifiers by comparing the attributes of shape moments and added a polynomial function of coordinates to accommodate higher-order shape moments. Their experiments demonstrated its improved memory usage and computational complexity; it was also able to classify rigid objects. However, this method has not been applied for use in other fields involving geometric analysis. Yang et al. [28] replaced the expensive multi-head attention mechanism with parameter-efficient group shuffle attention (GSA) to develop a converter of point attention that can process input data of different sizes and exhibits equivalence of transformation. Gumble subset sampling (GSS) was proposed for end-to-end learning and task-agnostic sampling. By selecting subsets in the representative hierarchy, the network can obtain a stronger representation of the input set at a lower computational cost. Experiments revealed the effectiveness and efficiency of the method in 3D image classification; however, GSS was not applied to general datasets, and its effectiveness and interpretability in hierarchical multi-instance learning were not explored. Zhao et al. [29] presented PointWeb to extract context features from a local neighborhood in the point cloud; PointWeb closely connects each point and uses an adaptive feature adjustment module to find the interaction among points. This framework can better learn the point representation for point cloud processing; however, its application to the understanding of 3D scenes needs further verification. Xie et al. [30] proposed a generation model of a disordered point cloud, where an energy function learns the coordinate coding of each point and then aggregates all individual point features into the energy of the whole point cloud. The model is trained by maximum likelihood learning based on Markov chain Monte Carlo (MCMC) and its variants, without an auxiliary network, and does not rely on the manual drawing of distance measurements to generate the point cloud. Thus, it is an efficient method for point cloud classification. However, the disturbance rejection of the model and the processing of point clouds with many outliers are unsatisfactory in practical applications. Yan et al. [31] proposed an end-to-end network called point adaptive sampling and local and nonlocal module (PointASNL) for robust point cloud processing in the presence of outliers and noise in the original point cloud. The network includes an adaptive sampling module and a local and nonlocal module. The first module adopts farthest point sampling (FPS) to sample the initial point cloud and reweights the neighborhood value, while the second captures neighborhood points and long-range dependencies. PointASNL has achieved a high level of robustness in point cloud classification; however, the fine-tuning strategy in the adaptive sampling module requires constant exploration. In the task of point cloud classification and segmentation, it is difficult to use geometric information to extract local features and correctly select important features. Accordingly, Jin et al. [32] proposed a graph-based neural network with an attention pooling strategy, named AGNet, which can effectively extract the spatial information of different distances and select the most crucial features, achieving a good classification and segmentation effect. Because of the influence of sensors, scenes, and other factors, the sparsity of collected point clouds in real environments is much different, which affects the accuracy of the final classification. Hence, complex preprocessing on point clouds is inevitable. Moreover, disturbances and outliers will inevitably appear in a real point cloud, i.e., a point may appear in a radius near its sampled area or anywhere in space and rotating the point cloud will represent the same 3D object with a different shape. Therefore, point cloud outliers and disturbances, as well as rigid transformation processing, make point cloud-based 3D object classification more intricate.

4. Multiview-Based

Multiview-based 3D point cloud classification involves the collection of 2D images of a 3D point cloud object from different viewpoints, which are sent to a CNN model for classification to produce the final representation. Chen et al. [4] studied 3D high-precision target detection in autonomous driving scenes and proposed a multiview 3D network consisting of sub-networks for 3D object representation and multiview feature fusion. Its deep fusion scheme combines regional features from multiple views, enabling the middle layers of different paths to interact. This improves the accuracy of the classification and reduces the amount of calculation. However, because the model utilizes a region-based fusion network, it is deficient in the extraction of feature information of objects from a global point of view. Using a large-scale ground truth dataset and a baseline view-based recognition method, Papadakis et al. [33] benchmarked multiple multiview hypothesis fusion schemes under various environmental assumptions and observation capabilities. Their experimental results highlighted significant aspects that should be considered in the design of a multiview-based recognition pipeline, while the analysis was only reflected in the 3D shape without considering texture characteristics. Cheng et al. [34] proposed a feature selection method that embeds low-rank constraints, sparse representation, and global and local structure learning in a unified framework; constructs a Laplace matrix based on the regularization term of a hypergraph; and applies a novel optimization algorithm to solve the objective function. The method achieves good classification performance on multiview datasets but does not extend to unsupervised and clustering learning. Inspired by the singular multiview convolution network structure, Pramerdorfer et al. [35] proposed a 3D bounding box approach that combines jointly classified objects

and regression models in a depth map. This method has excellent robustness to occlusion. It can process the views of encoded object geometry and occlusion information to output class scores and bounding box coordinates in world coordinates without post-processing. The model has excellent classification accuracy and a low regression error rate. It is worth researching whether its performance can be improved with different front-end architectures, such as ResNet, or by integrating the model into a deployed detection system based on Kinect to evaluate its performance. Feng et al. [12] presented a group view convolution neural network (GVCNN) for discriminative 3D shape description and hierarchical correlation modeling to better utilize the inherent hierarchical correlation and resolution between multiple views. This method introduces a hierarchical shape description framework, including views, groups, and shape level descriptors; considers the correlation among the views of each shape; and uses the group information for shape representation. It has realized performance improvements in 3D shape classification and retrieval tasks; however, it is not perfect with more complete views. Liu et al. [36] proposed a multiview hierarchical fusion network (MVHFN) that includes visual feature learning and multiview hierarchical fusion modules. In the first module, a 2D CNN extracts the visual features of multiple views drawn around a 3D object, while multiple view features are integrated as a compact descriptor in the second module. The model can find discriminant content by learning the cluster-level feature information, making full use of multiview and improving the classification accuracy. However, the number of views captured is low and the order of views is fixed, which cannot simulate ground object recognition in real 3D scenes well. Li et al. [37] proposed a probabilistic hierarchical model for multiview classification that learns a potential variable to fuse multiple features obtained from the same view, sensor, and morphology; applies the mapping matrix of a view to project the potential variable from the shared space to multiple observations; and employs expectation maximization (EM) to estimate parameters and potential variables. The network model can integrate multiview and feature data in multilevel, and the calculation of relevant parameters is repeatable and in line with common sense; however, the complexity of the model is high and its calculations are inefficient. He et al. [38] presented an online Bayesian multiview learning algorithm that learns the prediction subspace based on the principle of maximum margin, defines the potential marginal loss, and minimizes the learning problems associated with various Bayesian frameworks by using the theory of pseudo-likelihood and data enhancement. It obtains an approximate a posteriori change according to past samples. The model can attain higher classification performance than some advanced methods and can automatically infer weights and penalty parameters; however, the calculation is complex when the dataset is large. Li et al. [39] designed a Gaussian process latent variable model (GPVLM), which represents multiple views in a common subspace. It learns another projection from observed data to shared variables through view sharing and view-specific kernel parameters under a Gaussian process structure. Potential variables are changed to label information by Gaussian transformation, which reveals the correlation between views, and the model performs relatively well. Nevertheless, the radial basis function (RBF) cannot adapt to distributed and complex data, and multi-core learning is not considered. Yu et al. [40] proposed latent multiview CNN (LMVCNN), which utilizes predefined or random multiview images to identify 3D shapes and consists of three sub-convolution neural networks. The three CNN modules generate a multi-category probability distribution, build a potential vector to help the first CNN select the appropriate distribution, and output the category probability distribution of one view from another. LMVCNN has good discriminative ability for predefined and random views and can show excellent performance when the number of views is small; however, it does not solve the problem of 3D shape recognition without background interference. Hence, it is difficult to recognize objects in real 3D environments. Considering similarity measurements between image blocks, Yu et al. [13] proposed a multiview harmonized bilinear network (MHBN) for 3D object recognition. The model applies bilinear pooling to local convolution to obtain a compact global representation and produces a more discriminant representation by coordinating the singular values of set features. Its effectiveness in 3D object recognition was verified by experiments, where a high classification accuracy was achieved. Traditional methods invariably have certain disadvantages, such as multiview selection from a fixed viewpoint and static convolution for feature extraction. Therefore, if these disadvantages are overcome, the classification accuracy can be further improved. Multiview-based methods occupy less storage space than voxel and direct point cloud processing because they only require several 2D views. Converted 2D views can be adequately utilized by the current 2D CNN model; thus, training time is significantly reduced and the accuracy of model classification is the highest among the three different methods described above, namely voxel, point cloud, and multiview-based techniques.

However, most multiview-based methods [9][12][13] use traditional fixed-view projection when converting a 3D point cloud to 2D views, which can cause high similarity among view data. Therefore, the discriminative ability of the model decreases with multiple views on a test set, thus reducing the generalization ability of the model. Moreover, these methods often use a pretrained CNN backbone model to improve the efficiency of feature extraction; however, most backbone models adopt traditional static convolution, which is not adaptive to different data. Additionally, the use of only maximum and average pooling in feature fusion causes a considerable loss of detail and makes fusion inefficient. In order to solve these problems, researchers propose FSDCNet, a fusion of static and dynamic convolution networks for 3D point cloud

classification with high accuracy, high efficiency, and a strong generalization ability. The model has five advantages over the abovementioned deep learning-based methods:

(1)FSDCNet, a fusion static and dynamic convolution neural network, is proposed and applied to the classification of a 3D point cloud. The network adopts fixed and random viewpoint selection method for multiview generation. Meanwhile, it carries out local feature extraction by combining lightweight dynamic convolution with traditional static convolution in parallel. In addition, adaptive global attention pooling is used for global feature fusion. The network model is applicable not only to dense point clouds (ModelNet40), but also to sparse point clouds (Sydney Urban Objects). The experiments demonstrate that it can achieve state-of-the-art classification accuracy on two datasets. The algorithm framework, FSDCNet, is evidently different from other advanced algorithm frameworks.

(2)FSDCNet devises a view selection method of fixed and random viewpoints in the process of converting a point cloud into a multiview representation. The combination of random viewpoints avoids the problem of high similarity among views obtained by the traditional fixed viewpoint and improves the generalization of the model.

(3)FSDCNet constructs a lightweight dynamic convolution operator. The operator solves the problems of large parameters and expensive computational complexity in dynamic convolution, while maintaining the advantages of dynamic convolution; the complexity is only equivalent to that of traditional static convolution. Meanwhile, it can obtain abundant feature information in different receptive fields.

(4)FSDCNet proposes an adaptive fusion method of static and dynamic convolution. It can solve the problem of weak adaptability associated with traditional static convolution and extract more fine-grained feature information, which significantly improves the performance of the network model.

(5)FSDCNet designs adaptive global attention pooling to avert the low efficiency of global feature fusion with maximum and average pooling. It can integrate the most crucial details on multiple local views and improve the fusion efficiency of multiview features.

References

1. Zhang, J.X.; Lin, X.G. Advances in fusion of optical imagery and LiDAR point cloud applied to photogrammetry and remote sensing. Int. J. Image Data Fusion 2017, 8, 1–31.
2. Wentz, E.A.; York, A.M.; Alberti, M.; Conrow, L.; Fischer, H.; Inostroza, L.; Jantz, C.; Pickett, S.T.A.; Seto, K.C.; Taubenbock, H. Six fundamental aspects for conceptualizing multidimensional urban form: A spatial mapping perspective. Landsc. Urban Plan. 2018, 179, 55–62.
3. Yue, X.Y.; Wu, B.C.; Seshia, S.A.; Keutzer, K.; Sangiovanni-Vincentelli, A.L.; Assoc Comp, M. A LiDAR Point Cloud Generator: From a Virtual World to Autonomous Driving. In Proceedings of the 8th ACM International Conference on Multimedia Retrieval (ACM ICMR), Yokohama, Japan, 11–14 June 2018; pp. 458–464.
4. Chen, X.Z.; Ma, H.M.; Wan, J.; Li, B.; Xia, T. Multi-View 3D Object Detection Network for Autonomous Driving. In Proceedings of the 30th IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 6526–6534.
5. Braun, A.; Tuttas, S.; Borrmann, A.; Stilla, U. Improving progress monitoring by fusing point clouds, semantic data and computer vision. Autom. Constr. 2020, 116, 103210.
6. Jaritz, M.; Gu, J.Y.; Su, H. Multi-view PointNet for 3D Scene Understanding. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Korea, 27 October–2 November 2019; pp. 3995–4003.
7. Duan, H.; Wang, P.; Huang, Y.; Xu, G.; Wei, W.; Shen, X. Robotics Dexterous Grasping: The Methods Based on Point Cloud and Deep Learning. Front. Neurorobot. 2021, 15, 1–27.
8. Yang, Y.; Fang, H.R.; Fang, Y.F.; Shi, S.J. Three-dimensional point cloud data subtle feature extraction algorithm for laser scanning measurement of large-scale irregular surface in reverse engineering. Measurement 2020, 151, 107220.
9. Su, H.; Maji, S.; Kalogerakis, E.; Learned-Miller, E. Multi-view Convolutional Neural Networks for 3D Shape Recognition. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 11–18 December 2015; pp. 945–953.
10. Qi, C.R.; Su, H.; Mo, K.C.; Guibas, L.J. PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation. In Proceedings of the 30th IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 77–85.

11. Qi, C.R.; Yi, L.; Su, H.; Guibas, L.J. PointNet plus plus: Deep Hierarchical Feature Learning on Point Sets in a Metric Space. In Proceedings of the 31st Annual Conference on Neural Information Processing Systems (NIPS), Long Beach, CA, USA, 4–9 December 2017; pp. 4–9.
12. Feng, Y.F.; Zhang, Z.Z.; Zhao, X.B.; Ji, R.R.; Gao, Y. GVCNN: Group-View Convolutional Neural Networks for 3D Shape Recognition. In Proceedings of the 31st IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–23 June 2018; pp. 264–272.
13. Yu, T.; Meng, J.J.; Yuan, J.S. Multi-view Harmonized Bilinear Network for 3D Object Recognition. In Proceedings of the 31st IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–23 June 2018; pp. 186–194.
14. Wei, X.; Yu, R.X.; Sun, J. View-GCN: View-based Graph Convolutional Network for 3D Shape Analysis. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 14–19 June 2020; pp. 1847–1856.
15. Li, L.; Zhu, S.Y.; Fu, H.B.; Tan, P.; Tai, C.L. End-to-End Learning Local Multi-View Descriptors for 3D Point Clouds. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 14–19 June 2020; pp. 1916–1925.
16. Xiong, B.A.; Jiang, W.Z.; Li, D.K.; Qi, M. Voxel Grid-Based Fast Registration of Terrestrial Point Cloud. Remote Sens. 2021, 13, 1905.
17. Plaza, V.; Gomez-Ruiz, J.A.; Mandow, A.; Garcia-Cerezo, A.J. Multi-layer Perceptrons for Voxel-Based Classification of Point Clouds from Natural Environments. In Proceedings of the 13th International Work-Conference on Artificial Neural Networks (IWANN), Palma de Mallorca, Spain, 10–12 June 2015; pp. 250–261.
18. Liu, Z.J.; Tang, H.T.; Lin, Y.J.; Han, S. Point-Voxel CNN for Efficient 3D Deep Learning. In Proceedings of the 33rd Conference on Neural Information Processing Systems (NeurIPS), Vancouver, BC, Canada, 8–14 December 2019; pp. 965–975.
19. Plaza-Leiva, V.; Gomez-Ruiz, J.A.; Mandow, A.; Garcia-Cerezo, A. Voxel-Based Neighborhood for Spatial Shape Pattern Classification of Lidar Point Clouds with Supervised Learning. Sensors 2017, 17, 594.
20. Liu, Z.S.; Song, W.; Tian, Y.F.; Ji, S.M.; Sung, Y.S.; Wen, L.; Zhang, T.; Song, L.L.; Gozho, A. VB-Net: Voxel-Based Broad Learning Network for 3D Object Classification. Appl. Sci. 2020, 10, 6735.
21. Hamada, K.; Aono, M. 3D Indoor Scene Classification using Tri-projection Voxel Splatting. In Proceedings of the 10th Asia-Pacific-Signal-and-Information-Processing-Association Annual Summit and Conference (APSIPA ASC), Honolulu, HI, USA, 12–15 November 2018; pp. 317–323.
22. Wang, C.; Cheng, M.; Sohel, F.; Bennamoun, M.; Li, J. NormalNet: A voxel-based CNN for 3D object classification and retrieval. Neurocomputing 2019, 323, 139–147.
23. Hui, C.; Jie, W.; Yuqi, L.; Siyu, Z.; Shen, C. Fast Hybrid Cascade for Voxel-based 3D Object Classification. arXiv 2020, arXiv:2011.04522.
24. Zhao, Z.; Cheng, Y.; Shi, X.; Qin, X.; Sun, L. Classification of LiDAR Point Cloud based on Multiscale Features and PointNet. In Proceedings of the Eighth International Conference on Image Processing Theory, Tools and Applications (IPTA), Xi'an, China, 7–10 November 2018; p. 7.
25. Li, Z.Z.; Li, W.M.; Liu, H.Y.; Wang, Y.; Gui, G. Optimized PointNet for 3D Object Classification. In Proceedings of the 3rd European-Alliance-for-Innovation (EAI) International Conference on Advanced Hybrid Information Processing (ADHIP), Nanjing, China, 21–22 September 2019; pp. 271–278.
26. Kuangen, Z.; Jing, W.; Chenglong, F. Directional PointNet: 3D Environmental Classification for Wearable Robotics. arXiv 2019, arXiv:1903.06846.
27. Joseph-Rivlin, M.; Zvirin, A.; Kimmel, R. Momenet: Flavor the Moments in Learning to Classify Shapes. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Korea, 27 October–2 November 2019; pp. 4085–4094.
28. Yang, J.C.; Zhang, Q.; Ni, B.B.; Li, L.G.; Liu, J.X.; Zhou, M.D.; Tian, Q.; Soc, I.C. Modeling Point Clouds with Self-Attention and Gumbel Subset Sampling. In Proceedings of the 32nd IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 16–20 June 2019; pp. 3318–3327.
29. Hengshuang, Z.; Li, J.; Chi-Wing, F.; Jiaya, J. PointWeb: Enhancing Local Neighborhood Features for Point Cloud Processing. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 15–20 June 2019; pp. 5560–5568.
30. Xie, J.; Xu, Y.; Zheng, Z.; Zhu, S.-C.; Wu, Y.N. Generative PointNet: Deep Energy-Based Learning on Unordered Point Sets for 3D Generation, Reconstruction and Classification. In Proceedings of the IEEE/CVF Conference on Computer

Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 14976–14985.

31. Yan, X.; Zheng, C.D.; Li, Z.; Wang, S.; Cui, S.G. PointASNL: Robust Point Clouds Processing using Nonlocal Neural Networks with Adaptive Sampling. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 14–19 June 2020; pp. 5588–5597.
32. Jing, W.; Zhang, W.; Li, L.; Di, D.; Chen, G.; Wang, J. AGNet: An Attention-Based Graph Network for Point Cloud Classification and Segmentation. Remote Sens. 2022, 14, 1036.
33. Papadakis, P. A Use-Case Study on Multi-View Hypothesis Fusion for 3D Object Classification. In Proceedings of the 16th IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017; pp. 2446–2452.
34. Cheng, X.H.; Zhu, Y.H.; Song, J.K.; Wen, G.Q.; He, W. A novel low-rank hypergraph feature selection for multi-view classification. Neurocomputing 2017, 253, 115–121.
35. Pramerdorfer, C.; Kampel, M.; Van Loock, M. Multi-View Classification and 3D Bounding Box Regression Networks. In Proceedings of the 24th International Conference on Pattern Recognition (ICPR), Beijing, China, 20–24 August 2018; pp. 734–739.
36. Liu, A.A.; Hu, N.; Song, D.; Guo, F.B.; Zhou, H.Y.; Hao, T. Multi-View Hierarchical Fusion Network for 3D Object Retrieval and Classification. IEEE Access 2019, 7, 153021–153030.
37. Li, J.X.; Yong, H.W.; Zhang, B.; Li, M.; Zhang, L.; Zhang, D. A Probabilistic Hierarchical Model for Multi-View and Multi-Feature Classification. In Proceedings of the 32nd AAAI Conference on Artificial Intelligence/30th Innovative Applications of Artificial Intelligence Conference/8th AAAI Symposium on Educational Advances in Artificial Intelligence, New Orleans, LA, USA, 2–7 February 2018; pp. 3498–3505.
38. He, J.; Du, C.Y.; Zhuang, F.Z.; Yin, X.; He, Q.; Long, G.P. Online Bayesian max-margin subspace learning for multi-view classification and regression. Mach. Learn. 2020, 109, 219–249.
39. Li, J.X.; Li, Z.Q.; Lu, G.M.; Xu, Y.; Zhang, B.; Zhang, D. Asymmetric Gaussian Process multi-view learning for visual classification. Inf. Fusion 2021, 65, 108–118.
40. Yu, Q.; Yang, C.Z.; Fan, H.H.; Wei, H. Latent-MVCNN: 3D Shape Recognition Using Multiple Views from Pre-defined or Random Viewpoints. Neural Processing Lett. 2020, 52, 581–602.
