

Point Cloud Object Classifications

Subjects: Computer Science, Artificial Intelligence

Contributor: Hawking Lai, K. L. Eddie Law

A point cloud is a set of individual data points in a three-dimensional (3D) space. Proper collection of these data points may create an identifiable 3D structure, map, or model.

Keywords: point cloud classification ; deep learning ; lightweight model

1. Introduction

A point cloud is a set of individual data points in a three-dimensional (3D) space. Proper collection of these data points may create an identifiable 3D structure, map, or model. The effective analysis of point cloud data ^[1], such as semantic segmentation and object detection, is crucial for applications in 3D computer vision, including robotic sensing and autonomous driving. In general, the data points in point cloud contain geometric information, such as geometric coordinates, and other optional information (e.g., colors, normal vectors, etc.). However, the disordered and irregular nature of point cloud data makes analysis challenging. The recent advancement in deep learning has been phenomenal. Among them, the PointNet ^{[2][3]} was introduced as a solution to overcome the three non-trivial issues of permutation invariance, interactions between points, and transformation invariance. While newer designs, such as PointConv ^[4], KPConv ^[5], DeepGCNs ^[6], and PointTransformer ^[7], have shown better performance, they are expensive to run, thus limiting their practical uses. Hence, there is a need for lightweight and faster applications while still offering high precision and accuracy.

In 3D computer vision, methods of learning geometric features evolve from landmark MLP models (PointNet, PointNet++) to using convolutions, graphs, and transformer models. Though these newer designs obtain better performance, they mostly execute on expensive commodities, e.g., the high-end GPU card with a large amount of memory. It is atypical to deploy a heavy, slow, but high-performance application in 3D shape classification, object detection, etc. Instead, a lighter, faster, and relatively high-performance application is what users desire in general. In addition, limited improvement for some newer models can be observed through the experiments of benchmark datasets, such as ModelNet40 ^[8], ScanObjectNN ^[9], S3DIS ^[10], etc. Recently, many latest models such as PointNeXt ^[11], and PointMLP ^[12], begin to rethink network design and training strategies. They show that advanced extractors (MLPs, convolutions, graphs, attentions) may not be the main factors to improve overall performance. For instance, PointNeXt performed lots of experiments to assess different training strategies, and their results showed that training strategies are factors in raising the ranks in benchmark tests.

2. Deep Learning on Point Cloud

There are two main types of 3D data representations: point-cloud-based and non-point-cloud-based. For the point cloud data, PointNet is the baseline model for processing the point cloud data. Another possible choice is to convert the point cloud to voxel data ^[13] or multi-view images ^[14] or other methods; for example, implicit representations ^[15] could be used to enrich point clouds for better 3D analysis. Traditionally, voxel-based structures are used to study 3D representations, but they require extensive computations to carry out 3D convolutions or run neural networks on 2D multi-view images. On the other hand, multi-view images are a desirable option because they allow for the use of well-designed 2D neural networks to interpret 3D visual data. It is generally observed that state-of-the-art works can be broadly categorized into two types: those that use pre-training with point cloud data and those that do not. This trend is also reflected in the benchmarks for classification models.

Therefore, classification models can be broadly categorized into two types: those that use pre-training methods with point cloud data and those that do not. Generally speaking, pre-training methods have an advantage in terms of performance, if we do not consider the voting strategy. The use of large models trained with multimodality, which connect computer vision and natural language processing, has allowed models to reach new heights. However, it is also possible to achieve very good performance by training models from scratch without using pre-training. In addition, models trained from scratch

tend to have significantly fewer parameters than those that use pre-training methods. Here are some nice SOTA works using pre-training: RECON ^[16] unifies the data scaling capacity of contrastive models and the representation generalization of generative models to achieve a new state-of-the-art in 3D representation learning on ScanObjectNN. I2P-MAE ^[17] is an alternative method to obtain high-quality 3D representations from 2D pre-trained models via image-to-point masked autoencoders, which leverage well-learned 2D knowledge to guide 3D masked autoencoding. ULIP ^[18] is a framework that pre-trains 3D models on object triplets from images, texts, and 3D point clouds by leveraging a pre-trained vision-language model, achieving state-of-the-art performance on 3D classification tasks. These pre-training methods almost aim to improve the representation learning of point clouds for better performance on downstream tasks.

The basic point cloud data structure contains points, and each point is like a pixel in a picture. PointNet ^[2] is one deep learning neural network model to process the point cloud data. Vanilla PointNet uses an MLP and max pooling to learn a feature. Apart from MLP, there are other approaches using attention (also known as transformers), convolutions, and graphs to produce good benchmark results. Unlike MLPs, these design models are usually complicated, and their approaches are tedious to work on. Hence, most accepted approaches use MLPs, such as the PointNeXt ^[11], PointMLP, and RepSurf ^[19], etc. PointNeXt reuses the architecture of PointNet++ ^[3] with some modifications and uses training strategies to achieve good performance. Then the PointMLP is totally a different architectural MLP design that is not based on the PointNet model, and RepSurf focuses mostly on the point cloud representation. Potentially, a good point cloud representation (triangular RepSurf and umbrella RepSurf) may lead a selected model (e.g., PointNet++) to achieve excellent performance which is comparable to that of the state-of-the-art design. Recent advancements in hyperbolic neural networks (HNN) ^[20] have led to several studies attempting to project data into the hyperbolic space to improve performance, such as HyCoRe ^[21], which proposes a new method of embedding point cloud classifier features into hyperbolic space with explicit regularization to capture the inherent compositional nature of 3D objects, achieving a significant improvement in the performance of supervised models for point cloud classification.

In summary, there are various ways to process point cloud data using deep learning models, and the choice of model depends on the specific application and performance requirements. While PointNet is a popular choice for point cloud processing, other models such as PointNeXt, PointMLP, and RepSurf can also achieve excellent results with some modifications.

3. PointMLP

PointMLP ^[12] is an extension of the traditional MLP architecture that incorporates residual blocks for better performance on point cloud analysis tasks. The key idea behind PointMLP is that local geometric information may not be as important as previously thought and that simple yet effective methods such as the geometric affine sampling technique can achieve high accuracy without the need for complex geometric feature extraction. This approach not only improves robustness but also allows for faster computation and reduced model complexity. The use of residual blocks in PointMLP further enhances the network's ability to learn complex representations by enabling it to capture both short-term and long-term dependencies. Additionally, the FPS sampling with the geometric affine method allows for better handling of variations in scale, rotation, and translation, making PointMLP suitable for a wide range of point cloud analysis tasks.

Overall, The main contribution of PointMLP is its use of residual blocks in an MLP framework for point cloud analysis, challenging the notion that local geometric information is necessary for accurate analysis. Additionally, the use of the geometric affine sampling method enhances the robustness and accuracy of the model. PointMLP represents a promising advancement in the field of deep learning for point cloud analysis.

4. GhostMLP

GhostMLP is a novel neural network architecture that introduces a new residual MLP module that significantly reduces the number of parameters without compromising the classification performances. The GhostMLP is evaluated and validated with the two popular point cloud classification datasets: ScanObjectNN and ModelNet40. The results show that GhostMLP and its variant GhostMLP-S achieve competitive performance and fast inference compared with other state-of-the-art methods and their baseline. GhostMLP also could be to other tasks such as part segmentation on ShapeNet-Part and classification on mobile lidar scan (MLS) data, demonstrating its versatility and efficiency for different point cloud recognition applications.

References

1. Camuffo, E.; Mari, D.; Milani, S. Recent Advancements in Learning Algorithms for Point Clouds: An Updated Overview. *Sensors* 2022, 22, 1357.
2. Qi, C.R.; Su, H.; Mo, K.; Guibas, L.J. Pointnet: Deep learning on point sets for 3D classification and segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 652–660.
3. Qi, C.R.; Yi, L.; Su, H.; Guibas, L.J. Pointnet++: Deep hierarchical feature learning on point sets in a metric space. *Adv. Neural Inf. Process. Syst.* 2017, 30, 5105–5114.
4. Wu, W.; Qi, Z.; Fuxin, L. Pointconv: Deep convolutional networks on 3D point clouds. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 9621–9630.
5. Thomas, H.; Qi, C.R.; Deschaud, J.E.; Marcotegui, B.; Goulette, F.; Guibas, L.J. Kpconv: Flexible and deformable convolution for point clouds. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Republic of Korea, 27 October–2 November 2019; pp. 6411–6420.
6. Li, G.; Muller, M.; Thabet, A.; Ghanem, B. Deepgcns: Can gcns go as deep as cnns? In Proceedings of the Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Republic of Korea, 27 October–2 November 2019; pp. 9267–9276.
7. Zhao, H.; Jiang, L.; Jia, J.; Torr, P.H.; Koltun, V. Point transformer. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Virtual Conference, 11–17 October 2021; pp. 16259–16268.
8. Wu, Z.; Song, S.; Khosla, A.; Yu, F.; Zhang, L.; Tang, X.; Xiao, J. 3D shapenets: A deep representation for volumetric shapes. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 1912–1920.
9. Uy, M.A.; Pham, Q.H.; Hua, B.S.; Nguyen, T.; Yeung, S.K. Revisiting point cloud classification: A new benchmark dataset and classification model on real-world data. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Republic of Korea, 27 October–2 November 2019; pp. 1588–1597.
10. Armeni, I.; Sener, O.; Zamir, A.R.; Jiang, H.; Brilakis, I.; Fischer, M.; Savarese, S. 3D semantic parsing of large-scale indoor spaces. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 1534–1543.
11. Qian, G.; Li, Y.; Peng, H.; Mai, J.; Hammoud, H.A.A.K.; Elhoseiny, M.; Ghanem, B. PointNeXt: Revisiting PointNet++ with Improved Training and Scaling Strategies. *arXiv* 2022, arXiv:2206.04670.
12. Ma, X.; Qin, C.; You, H.; Ran, H.; Fu, Y. Rethinking network design and local geometry in point cloud: A simple residual mlp framework. *arXiv* 2022, arXiv:2202.07123.
13. Zhou, Y.; Tuzel, O. Voxelnet: End-to-end learning for point cloud based 3d object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 4490–4499.
14. Hamdi, A.; Giancola, S.; Ghanem, B. Mvtn: Multi-view transformation network for 3D shape recognition. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, BC, Canada, 11–17 October 2021; pp. 1–11.
15. Yang, Z.; Ye, Q.; Stoter, J.; Nan, L. Enriching Point Clouds with Implicit Representations for 3D Classification and Segmentation. *Remote Sens.* 2022, 15, 61.
16. Qi, Z.; Dong, R.; Fan, G.; Ge, Z.; Zhang, X.; Ma, K.; Yi, L. Contrast with Reconstruct: Contrastive 3D Representation Learning Guided by Generative Pretraining. *arXiv* 2023, arXiv:2302.02318.
17. Zhang, R.; Wang, L.; Qiao, Y.; Gao, P.; Li, H. Learning 3D Representations from 2D Pre-trained Models via Image-to-Point Masked Autoencoders. *arXiv* 2022, arXiv:2212.06785.
18. Xue, L.; Gao, M.; Xing, C.; Martín-Martín, R.; Wu, J.; Xiong, C.; Xu, R.; Niebles, J.C.; Savarese, S. ULIP: Learning Unified Representation of Language, Image and Point Cloud for 3D Understanding. *arXiv* 2022, arXiv:2212.05171.
19. Ran, H.; Liu, J.; Wang, C. Surface Representation for Point Clouds. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 18942–18952.
20. Ganea, O.; Bécigneul, G.; Hofmann, T. Hyperbolic neural networks. *Adv. Neural Inf. Process. Syst.* 2018, 31, 1–11.
21. Montanaro, A.; Valsesia, D.; Magli, E. Rethinking the compositionality of point clouds through regularization in the hyperbolic space. *arXiv* 2022, arXiv:2209.10318.

