점운증강을 위한 프로젝션 손실

오신모*, 이효종*
*전북대학교 컴퓨터공학과

e-mail: chenmou0410@gmail.com
hlee@jbnu.ac.kr

Projection Loss for Point Cloud Augmentation

Chenmou Wu*, Hyo-Jone Lee*
*Div. of Computer Science & Engineering, Chonbuk National University

Abstract

Learning and analyzing 3D point clouds with deep networks is challenging due to the limited and irregularity of the data. In this paper, we present a data-driven point cloud augmentation technique. The key idea is to learn multilevel features per point and to reconstruct to a similar point set. Our network is applied to a projection loss function that encourages the predicted points to remain on the geometric shapes with a particular target. We conduct various experiments using ShapeNet part data to evaluate our method and demonstrate its possibility. Results show that our generated points have a similar shape and are located closer to the object.

1. Introduction

Point cloud, as a fundamental 3D representation, is widely used in various real-world surrounding perception researches such as auto-driven, SLAM and UAV. Recently, researches [2,3,4] on 3D perception attracts much attention with the availability of ShapeNet datasets [5] and the advent of deep leaning. These works achieve impressive results for 3D object classification, reconstruction, and semantic scene segmentation.

In these works, without using the traditional method, the features are commonly extracted from the raw 3D point set by using deep neural networks. However, existing point cloud datasets are not enough to satisfy this data-driven method. Compared with image or video datasets [6][7] in computer vision, point cloud-related datasets are minimal.

In this paper we are interested in point cloud-based data augmentation problem: give a random set of points with a label, generate a corresponding 3D model based on point cloud. This augmentation problem is similar in spirit to image augmentation. However, dealing with 3D points rather than a 2D grid of pixels poses new challenges. The problem is, as an irregular data format, point clouds do not have any order and regular grid. Unlike the image plane, the spatial features of the point cloud are extracted from its coordinates. Therefore, the geometric transformation (e.g., rotation and translation), which widely used in image augmentation, is not suitable for points clouds in 3D space. Furthermore, downsampling or upsampling only can reduce or improve the precision of the model. The geometric structure of the point cloud ar e not changed.

To solve the above challenges, we present a data-driven point cloud autoencoder network. Our network is with a projection loss that encourages real object image. Inspired by Pointnet [1], convolution layers and deconvolution layer use T-net architecture in our network. The key idea is using point cloud autoencoder to get the basic shape and projection loss to improve the reality of the generated model.

2. Related work

As irregular data, points in the point cloud are unordered and independent. It is hard to extract the features by using the traditional convolution network directly. To get the spatial information in the point cloud, some earlier works propose to convert point cloud into other 3D representations such as volumetric grids [8, 9, 10, 11] to process.

PointNet [1] is the pioneer that adopt a deep learning model to directly process point clouds in the convolution network of the point cloud. Detailly, they use the channel-wise max-pooling to aggregate per-point features into a global descriptor vector. The problem with PointNet is that the local features are not well extracted, because the point clouds are inconsistent in each local uniformity. PointNet++ [2] uses a multi-scale grouping and multi-resolution grouping to solve this problem. A similar permutation equivariant layer [12] is also proposed at almost the same time as [1], with the significant difference that the permutation equivariant layer is max-normalized. Although the max-pooling idea is proven to be effective, it suffers from the lack of ConvNet-like hierarchical feature aggregation. SO-Net [13] builds SOM to simulate the spatial distribution of point clouds. They extract hierarchical features from both per-points and SOM nodes to improve network performance. The PointCNN [3] proposes to learn an X transform based on the input points and then use it to simultaneously weight the input features associated with the points and rearrange them into potentially implied canonical sequences. PU-Net [14] is applied at a patch-level, with a joint loss function that encourages the upsampled points to remain on the underlying surface with a uniform distribution.

3. Method Overview

System overview

Given a set of randomly distribution 3D points with label, our network aims to output a point cloud with the similar geometric shape of the target object. The overview of our framework is illustrated in Fig.1. We adopt the autoencoder architecture that includes five convolution layers, five deconvolution layers and four fully connected layers. The convolution method is inspired by PointNet [1].

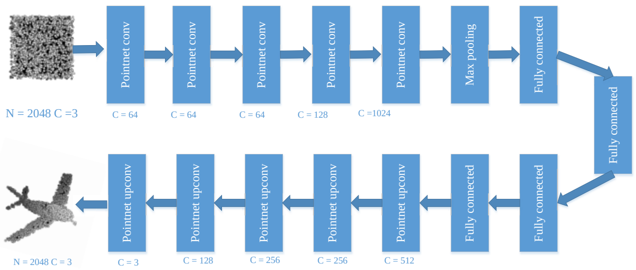

Fig.1. Overview of our framework. The whole network

Initial point cloud

Our model does not require any prior knowledge of the 3D shape and always deform from an initial point cloud model with average size placed at the standard location in the camera coordinate. The model is centered at 0.6m in front of the camera with 0.3m, 0.3m,0.4m as the radius of three axes. The model is generated by Gaussian distribution and contains 2048 points. We use this model to initialize our input, where the original feature contains only the 3D coordinate of each point.

Loss function

We adopt two kinds of losses to guarantee the property of the output model. The Chamfer loss [9] is used to constrain the location of points in the point cloud. The projection loss is used to reshape the geometric structure of the object.

Chamfer Loss: The Chamfer distance can measure the distance of each point to the other set:

$$l_c = \sum_p \min_q \|p - q\|_2^2 + \sum_q \min_p \|p - q\|_2^2$$

As a common loss function used in point cloud-based deep network, the Chamfer loss can regress the points close to its correct position. However, it is not sufficient to produce excellent shape.

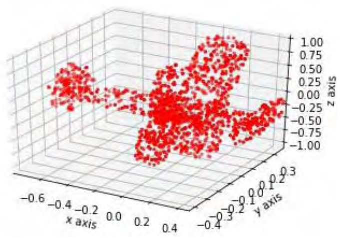

Fig.2. The normalized point cloud in 3D world coordinates.

Projection loss: The prior knowledge and imagination of the structure can help human quickly figure out the shape of an object. The projection loss injects such prior information in neural networks.

Projection loss measures the inconsistency of geometric shapes between the predicted points P_{pred} and ground-truth P_{GT} in different projections. As shown in Fig. 2, we first normalize the point clouds to be centered at the origin of the world coordinate. The numbers of points in P_{GT} and P_{pred} are both pre-assigned to 2048. Second, We orthogonally project P_{pred} and P_{GT} toward different image planes at the same time. The projection of P_{pred} and P_{GT} are denoted as I_{pred} and I_{GT} , respectively. To measure geometric inconsistency between P_{pred} and P_{GT}, we compare the pair of projection from the X-Y axis, Y-Z axis, and Z-X axis. An example is shown in Fig.3. The projection loss function defined as

$$L_{proj} = \sum_{axis} w_{axis} \frac{I_{pred}^{axis} \cap I_{GT}^{axis}}{I_{GT}^{axis}}$$

where, $axis \in$ (X-Y, Y-Z, Z-X) is the projected image plane and $I_{pred}^{axis} \cap I_{GT}^{axis}$ is the overlapping area of I_{pred} and I_{GT} in this axis plane. w_{axis} means weights of loss in three image planes.

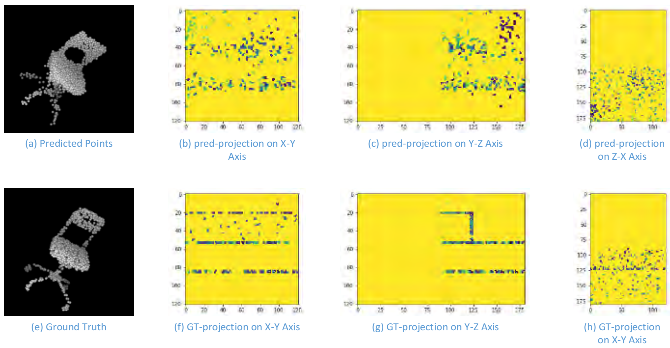

Fig. 3. (a) is the predicted points. (b)&(f) show the 2D projection of predicted point cloud and ground truth on x-axis and y-axis. (c)&(g) show the 2D projection of predicted point cloud and ground truth on y-axis and z-axis. (d)&(h) show the 2D projection of predicted point cloud and ground truth on x-axis and z-axis. (f) is the ground-truth.

4. Experiments

Our network is implemented with tensorflow on a NVIDIA GTX TITAN X. We perform our experiments on the ShapeNet dataset [2], which has a large collection of textured CAD models.

The whole network is trained in an end-to-end fashion using Adam [15] optimizer with batch size 32 and an initial learning rate of 0.001. Batch-normalization and ReLU activation are applied to every layer. Weights of the X-Y axis, Y-Z axis and Z-X axis are respectively set as 0.4, 0.4, 0.2.

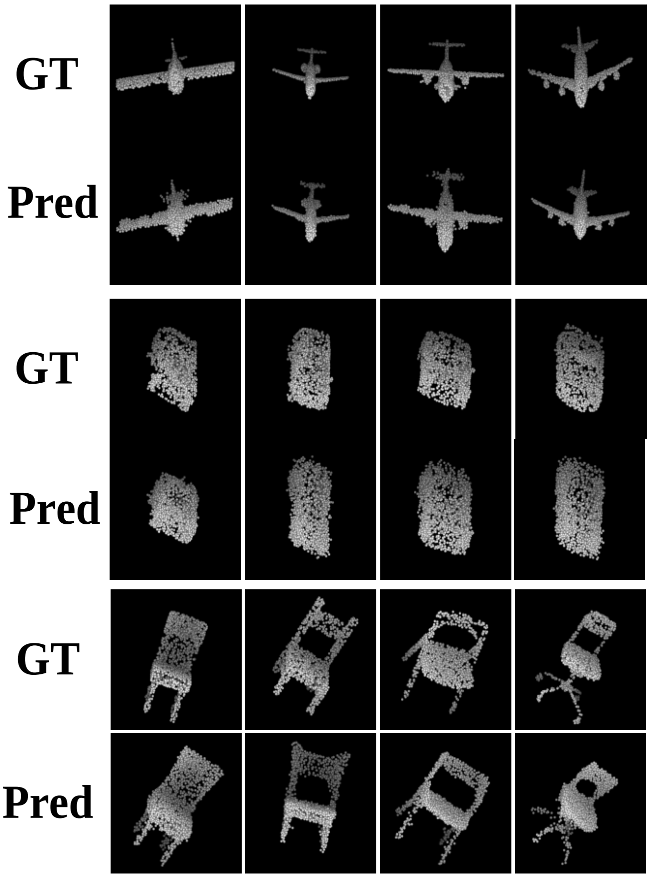

Fig. 4. Examples of point cloud autoencoder results. First row: input point clouds of size 1024. Second row: reconstructed point clouds of size 1280. From top to bottom: airplane, car, chair.

It is difficult to provide quantitative comparison for the point cloud autoencoder task because little research has been done on this topic. The most related work is the point set generation network [16] and the point cloud generative models [17]. Examples of our reconstructed ShapeNet point clouds are visualized in Fig. 4, where 2048 points recovered from the deconvolution branch. The overall testing Chamfer distance is 0.0039. The convolution branch recovers the main body of the object, while the more flexible fully connected branch focuses on details such as the legs of a table. Nevertheless, many finer details are lost.

5. Conclusion

In this paper, we present a deep network for point cloud augmentation, with the goal of generating a similar and uniform set of points from a randomly set of points. We have presented the projection loss to reconstruct point cloud from three global perspectives. Projection loss stimulates the network to reconstruct 3D object regarding semantic information from ground truth. Results and analysis in the experiment section show that the model trained by our projection loss achieves good performance on ShapeNet dataset.

6. Acknowledgeemtn

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (GR 2016R1D1A3B03931911).

Reference

[1] Qi C R, Su H, Mo K, et al. Pointnet: Deep learning on point sets for 3d classification and segmentation[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017: 652-660.
[2] C. R. Qi, L. Yi, H. Su, and L. J. Guibas. PointNet++: Deep hierarchical feature learning on point sets in a metric space. In Int. Conf. on Neural Information Processing Systems (NIPS), 2017.
[3] Li Y, Bu R, Sun M, et al. PointCNN: Convolution On X-transformed Points[C]. Advances in Neural Information Processing Systems. 2018: 828-838.
[4] Fan H, Su H, Guibas L J. A point set generation network for 3d object reconstruction from a single image[C]. Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 605-613.
[5] Chang, A.X., Funkhouser, T., Guibas, L., Hanrahan, P., Huang, Q., Li, Z., Savarese, S., Savva, M., Song, S., Su, H., et al.: Shapenet: An information-rich 3d model repository. arXiv (2015)
[6] Deng J, Dong W, Socher R, et al. Imagenet: A large-scale hierarchical image database[C]. 2009 IEEE conference on computer vision and pattern recognition. Ieee, 2009: 248-255.
[7] Soomro K, Zamir A R, Shah M. UCF101: A dataset of 101 human actions classes from videos in the wild[J]. arXiv preprint arXiv: 1212.0402, 2012.
[8] D. Maturana and S. Scherer. Voxnet: A 3D convolutional neural network for real-time object recognition. In Int. Conf. on Intell. Robots and Systems (IROS), 2015.
[9] G. Riegler, A. O. Ulusoys, and A. Geiger. OctNet: Learning deep 3D representations at high resolutions. In IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), 2017.
[10] A. Dai, C. R. Qi, and M. Niessner. Shape completion using 3D-Encoder-Predictor CNNs and shape synthesis. In IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), 2017.
[11] Z. Wu, S. Song, A. Khosla, F. Yu, L. Zhang, X. Tang, and J. Xiao. 3D ShapeNets: A deep representation for volumetric shapes. In IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), 2015.
[12] S. Ravanbakhsh, J. Schneider, and B. Poczos. Deep learning with sets and point clouds. arXiv preprint arXiv:1611.04500, 2016.
[13] Li J, Chen B M, Hee Lee G. So-net: Self-organizing network for point cloud analysis[C]. Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 9397-9406.
[14] Yu L, Li X, Fu C W, et al. Pu-net: Point cloud upsampling network[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018: 2790-2799.
[15] D. Kingma and J. Ba. Adam: A method for stochastic optimization. *arXiv* preprint arXiv:1412.6980, 2014.
[16] P. Achlioptas, O. Diamanti, I. Mitliagkas, and L. Guibas. Learning representations and generative models for 3d point clouds. arXiv preprint arXiv:1707.02392, 2017.
[17] H. Fan, H. Su, and L. Guibas. A point set generation network for 3d object reconstruction from a single image. 2017.