PointCutMix: Regularization Strategy for Point Cloud Classification

Jinlai Zhang 1 * Lvjie Chen 2 * Bo Ouyang 2 Binbin Liu 2 Jihong Zhu 1,2,3 Yujing Chen 1 Yanmei Meng 1 Danfeng Wu 3

Abstract

3D point cloud analysis has received increasing attention in recent years, however, the diversity and availability of point cloud datasets are still limited. We therefore present PointCutMix, a simple but effective method for augmentation in point cloud. In our method, after finding the optimal assignment between two point clouds, we replace some points in one point cloud by its counterpart point in another point cloud. Our strategy consistently and significantly improves the performance across various models and datasets. Surprisingly, when it is used as a defense method, it shows far superior performance to the SOTA defense algorithm. The code is available at: https://github.com/cuge1995/PointCutMix.

1. Introduction

With the rapid development of autonomous driving and robotics industries, making machines understand the real three-dimensional world has become a guarantee for safe and efficient task execution (Guo et al., 2020). Point cloud, as a commonly used format for 3D data representation, which can be obtained directly by LiDAR sensors and was widely used in industry and 3D computer vision community (Lang et al., 2019; Chen et al., 2019; Rao et al., 2020). With the pioneering work of Pointnet (Qi et al., 2017a), deep learning has boost the performance of many 3D computer vision tasks, such as 3D object detection (Shi et al., 2020; Bhattacharyya & Czarnecki, 2020), point cloud segmentation (Liu et al., 2020b) and point cloud classification (Qi et al., 2017b; Liu et al., 2019c; Wang et al., 2019). However, for those networks which over millions of parameters, one common drawback is that they can easily over-fit when training data are scarce (Jing & Tian, 2020). Considering the point cloud datasets are smaller compared to the image datasets (Chen et al., 2020), researchers has explored several data augmentation techniques on point cloud. The rotation, scaling, and/or jittering are commonly used for point cloud analysis (Yan et al., 2020; Liu et al., 2020b). However, those augmentations ignore the shape complexity of the samples (Li et al., 2020), and thus lead to insufficient training.

Mixed sample data augmentation (MSDA) has received increasing interest in the image domain over past few years (Harris et al., 2020). However, point cloud is unordered, which is different from the image, and thus MSDA strategies in image domain cannot directly applied to point cloud.

In this paper, inspired by the success of MSDA in image domain, we investigated the MSDA in point cloud. Specifically, we proposed PointCutMix, a new augmentation strategy of point cloud samples. By replacing some points in one point cloud with its optimal assignment points in another point cloud, we formulate two PointCutMix methods. The first is we randomly selected some points in one point cloud and replaced with its optimal assignment points in another point cloud. The second is we randomly selected one point in one point cloud, then finding the n nearest neighbors points of this point, we formulate those point sets S and finally replaced with the optimal assignment points of S in another point cloud. Experimental results showed that clear improvements in shape classification on ModelNet40 (Wu et al., 2015).

2. Related Work

Deep learning on point cloud. Pointnet (Qi et al., 2017a) firstly processes point cloud using deep neural networks directly, the shared pointwise multi-layer perceptions (MLPs) followed by the max-pooling operation was used for point cloud learning. After that, several recent works (Qi et al., 2017b; Yan et al., 2020; Zhao et al., 2019; Yang et al., 2019b) mainly focus on how to efficiently capture local features. Some others investigated convolutional kernels for 3D point clouds. Liu et al. (Liu et al., 2019c) proposed RS-CNN, which implemented the convolution using an MLP in the

---

1Equal contribution 1College of Mechanical Engineering, Guangxi University, Nanning, China 2Department of Computer Science, Tsinghua University, Beijing, China 3Department of Precision Instrument, Tsinghua University, Beijing, China. Correspondence to: Jihong Zhu <jzhu@tsinghua.edu.cn>.

Proceedings of the 38th International Conference on Machine Learning, Copyright 2021 by the author(s).
local subset of points. DensePoint (Liu et al., 2019b) defines a single-layer perceptron with a nonlinear activator as convolution. In KPConv (Thomas et al., 2019), by using a set of learnable kernel points, the rigid and deformable Kernel point convolution operators were proposed. Another researchers has explored the graph-based networks, where each point in a point cloud was considered as a vertex of a graph. DGCNN (Wang et al., 2019) constructed a graph in the feature space, and MLP is used for each edge. To simplify the process of points agglomeration, the Dynamic Points Agglomeration Module based on graph convolution was proposed by Liu et al. (Liu et al., 2019a) In RGCNN (Te et al., 2018), a graph was constructed by connecting each point with all other points in the point cloud. To utilize the local structural information, LocalSpecGCN (Wang et al., 2018) used the spectral convolution network to a local graph.

Our work aim to improve the existing network’s regulazation and generalization.

**Mixed sample data augmentation.** Mixed sample data augmentation (MSDA) mixed training samples according to some rules (Harris et al., 2020), and then using the mixed data to train the model. An ideal model would learn multiple features in a balanced way (Taghanaki et al., 2020), obviously, MSDA is a balanced way. MSDA is mainstream data augmentation pipeline and has dominated modern image classification for years (Harris et al., 2020; Guo et al., 2019; Zhang et al., 2017; Yun et al., 2019; Verma et al., 2019). Among them, CutMix shows impressive performance improvement across different datasets and networks. Our work can be viewed as an extension of CutMix (Yun et al., 2019) in point cloud.

**Data augmentation on point cloud.** The random rotation, random jittering, random scaling are commonly used in point cloud learning (Qi et al., 2017a;b). Obviously, the data augmentation in point cloud have not been studied systematically compared to the image domain. Recently, PointAugment (Li et al., 2020) and PointMixup (Chen et al., 2020) were proposed for point cloud data augmentation. PointAugment is the first auto-augmentation framework for point cloud, which optimizes the augmentor and classifier networks jointly. However, the complicated adversarial training process make it less practical. PointMixup extents Mixup (Zhang et al., 2017) to point cloud by interpolation between point cloud samples. However, for point cloud networks like Pointnet++ (Qi et al., 2017b) and RS-CNN (Liu et al., 2019c) that local features are important for point cloud learning, it is easy fall into locally ambiguous and unnatural. In this paper, we proposed PointCutMix, which naturally combine two point clouds. We hope our methods could help future research in 3D computer vision community.

![Figure 1. Some mixed samples](image)

The key idea of PointCutMix is to create a new training point cloud $(\tilde{x}, \tilde{y})$ given two distinct training point clouds $(x_1, y_1)$ and $(x_2, y_2)$. Here, $\tilde{y}$ is the training label and the training point cloud $x \in \mathbb{R}^{N \times d}$ with $N$ points, and each point having $d$-dimensional features. However, different from CutMix (Yun et al., 2019) in image domain, the point clouds are permutation-invariant and orderless, we therefore needing to define the one-to-one correspondence between points in two point clouds (Similar to the one-to-one correspondence between pixels in two images). Inspired by PointMixup (Chen et al., 2020) and MSN (Liu et al., 2020a), we define the one-to-one correspondence between points in two point clouds $F$ as the assignment of Earth Mover’s Distance (EMD) function. Finally, the combining operation is defined as follows,

$$\tilde{x} = B \odot x_1 + (1 - B) \odot F(x_2)$$

$$\tilde{y} = \lambda y_1 + (1 - \lambda)y_2$$

where $F(x_2)$ is the assigned point cloud of point cloud $x_2$, and $B \in \{0,1\}^N$ denotes a binary mask indicating which points belong to either of the two point cloud, 1 is a binary mask filled with ones and $\odot$ is the general element-wise multiplication. $\lambda \in [0,1]$ denotes the cutmix ratio, where $\lambda$ is sampled from the beta distribution $Beta(\beta, \beta)$.

In this paper, we propose two methods to sample the binary mask $B$. In the first method, a randomly sampled subset $x_1^*$ with $n = \lfloor \lambda \times N \rfloor$ points was sampled from $x_1$. Those points are marked 0 in $B$. Thus, the combining operation becomes

$$\tilde{x} = B \odot x_1 + (1 - B) \odot F(x_2)$$

$$\tilde{y} = \lambda y_1 + (1 - \lambda)y_2$$

where $\lambda_{rd} = n/N$, we abbreviate this method as PointCutMix-R.

In the second method, we randomly sample one point $p$ from $x_1$, and then finding its $n = \lfloor \lambda \times N - 1 \rfloor$ nearest neighbors, we combine $p$ and its $n$ nearest neighbors to form $x_1^*$ and marked those points as 0 in $B$. Therefore, the combining
PointCutMixoperation becomes
\[
\tilde{x} = B \odot x_1 + (1 - B) \odot F(x_2)
\]
\[
\tilde{y} = \lambda_{knn} y_1 + (1 - \lambda_{knn}) y_2
\]
where \(\lambda_{knn} = (n + 1)/N\), we abbreviate this method as PointCutMix-K.

4. Experiments
We now turn to the evaluation of PointCutMix.

4.1. Datasets and Classifiers

Datasets. We evaluated our PointCutMix on the ModelNet40 (Wu et al., 2015) and ModelNet10 (Wu et al., 2015) dataset. ModelNet40 contains 12311 samples in 40 categories, among them, 9843 for training and 2468 for testing. ModelNet10 contains a total of 4899 samples in 10 categories, of which 3991 samples are used for training and the rest are used for testing. We also performed extensive experiments on ScanObectNN (Uy et al., 2019) and SHREC16 (Savva et al., 2016) dataset further. ScanObectNN contains 2902 samples in 15 categories, which was sampled from real world scanning. SHREC16 is the largest dataset which contains 41313 samples in 55 categories, split into 36148 for training and 5165 for testing.

Classifiers. Our work can be implemented into any point-based point cloud classification networks, we therefore select the most popular networks in in 3D computer vision community (Li et al., 2020; Zhao et al., 2020), which are Pointnet (Qi et al., 2017a), Pointnet++ (Qi et al., 2017b), DGCNN (Wang et al., 2019) and RS-CNN (Liu et al., 2019c).

4.2. Implementation details
Our work was implemented by PyTorch (Paszke et al., 2017) on a workstation with four NVIDIA GeForce GTX 2080Ti GPUs. All networks take 1024 points as input. We set 300 training epochs with a batch size 16. We use the Adam optimizer with a learning rate of 0.001 and a decay rate of 0.5 every 20 epochs for RS-CNN (Liu et al., 2019c), Pointnet (Qi et al., 2017a) and Pointnet++ (Qi et al., 2017b), which is the same as the original configuration of the released paper and code. We trained DGCNN with SGD optimizer, the initial learning rate is 0.1 and the momentum of SGD is 0.9, the cosine annealing strategy is used to decay learning rate.

4.3. Evaluation results
Table 1 illustrate the classification accuracy of ModelNet40 dataset under various data augmentation methods, including conventional data augmentation (baseline) (Qi et al., 2017b), PointMixup (Chen et al., 2020), PointCutMix-K and PointCutMix-R. We observe that the PointCutMix-K consistently outperform other data augmentation methods.

| Method          | Pointnet | Pointnet++ | RS-CNN  | DGCNN  |
|-----------------|----------|------------|---------|--------|
| baseline        | 89.50    | 90.68      | 91.65   | 92.30  |
| PointMixup      | 88.21    | 91.61      | 91.85   | 92.13  |
| PointCutMix-R   | 88.65    | 91.61      | 91.85   | 92.54  |
| PointCutMix-K   | **89.71**| **92.34**  | **92.02**| **92.63**|

Table 2. Classification accuracy of ModelNet40 under point dropping attack (Zheng et al., 2019) on DGCNN.

| Method          | Baseline | PointMixup | PointCutMix-R | PointCutMix-K |
|-----------------|----------|------------|---------------|---------------|
| Acc(%)          | 43.80    | 73.38      | 79.38         | **79.94**     |

Deep neural networks are vulnerable to adversarial examples, which have studied extensively in 2D images (Dong et al., 2018; Akhtar & Mian, 2018). Recently, point perturbation attack (Xiang et al., 2019), kNN attack (Tsai et al., 2020) and point dropping attack (Zheng et al., 2019) were proposed for 3D point cloud. However, for the point perturbation attack and the kNN attack, they didn’t perform normalization of point clouds during the attack and the generated point clouds may not center within a unit sphere. But in this paper, our methods are trained after normalization of point clouds. Therefore, we only consider the point dropping attack in our robustness test.

We report the recognition accuracy after the point dropping attack on ModelNet40 test set in Table 2. It should be noticed that our baseline is more weak than models trained without conventional data augmentation, which is lower than IF-Defense (Wu et al., 2020) baseline. However, our PointCutMix-K and PointCutMix-R improve the robustness significantly, they still have nearly 80% accuracy to adversarial attacks.

Interestingly, we find that if our pretrained models using PointCutMix-K and PointCutMix-R augmentation are used as defense methods, it outperform state-of-the-art algorithm IF-Defense (Wu et al., 2020) significantly. Specifically, we first get the adversarial samples of the DGCNN model in IF-Defense baseline, then we perform classification on this dataset using the pretrained models trained with PointCutMix-K and PointCutMix-R augmentation. The results are shown in Table 3. Where SRS (Yang et al., 2019a), SOR (Zhou et al., 2019) and DUP-Net (Zhou et al., 2019) are defense method developed recently. From the results, we
Table 3. Classification accuracy of ModelNet40 under various defense methods on DGCNN. * denotes results report in IF-Defense (Wu et al., 2020) paper.

| Defense methods   | Acc(%) |
|-------------------|--------|
| No defense        | 55.06  |
| SOR*              | 59.36  |
| DUP-Net*          | 36.02  |
| SRS*              | 23.82  |
| IF-Defense*       | 73.30  |
| PointMixup        | 83.02  |
| PointCutMix-K     | 85.74  |
| PointCutMix-R     | **86.18** |

can see that, PointCutMix can be used as a defense method. It just uses a small amount of computing power to classify adversarial samples, which is a more natural defense method.

4.5. Discussion and Future work

Overall, the PointCutMix-R and PointCutMix-K show gratifying performance in various datasets and models.

In the future, we plan to extend our work to point cloud segmentation (Qi et al., 2017a) and 3D object detection (Shi et al., 2020). However, due to the point cloud is different from image, there are still some challenges.

5. Conclusion

In this paper, we proposed PointCutMix, a regularization strategy for point cloud classification. We conducted extensive experiments on our method, the results shows that PointCutMix improve the performance of four point cloud classification networks on four 3D benchmark significantly. Moreover, the robustness test shows that our work improve the robustness significantly, and outperform SOTA defense algorithm if used as defense method.

References

Akhtar, N. and Mian, A. Threat of adversarial attacks on deep learning in computer vision: A survey. IEEE Access, 6:14410–14430, 2018.

Bhattacharyya, P. and Czarnecki, K. Deformable pv-r cnn: Improving 3d object detection with learned deformations. arXiv preprint arXiv:2008.08766, 2020.

Chen, Y., Liu, S., Shen, X., and Jia, J. Fast point r-cnn. In Proceedings of the IEEE International Conference on Computer Vision, pp. 9775–9784, 2019.

Chen, Y., Hu, V. T., Gavves, E., Mensink, T., Mettes, P., Yang, P., and Snoek, C. G. Pointmixup: Augmentation for point clouds. arXiv preprint arXiv:2008.06374, 2020.

Dong, Y., Liao, F., Pang, T., Su, H., Zhu, J., Hu, X., and Li, J. Boosting adversarial attacks with momentum. In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 9185–9193, 2018.

Guo, H., Mao, Y., and Zhang, R. Mixup as locally linear out-of-manifold regularization. In Proceedings of the AAAI Conference on Artificial Intelligence, volume 33, pp. 3714–3722, 2019.

Guo, Y., Wang, H., Hu, Q., Liu, H., Liu, L., and Bennamoun, M. Deep learning for 3d point clouds: A survey. IEEE transactions on pattern analysis and machine intelligence, 2020.

Harris, E., Marcu, A., Painter, M., Niranjan, M., and Hare, A. P.-B. J. Fmix: Enhancing mixed sample data augmentation. arXiv preprint arXiv:2002.12047, 3, 2020.

Jing, L. and Tian, Y. Self-supervised visual feature learning with deep neural networks: A survey. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020.

Lang, A. H., Vora, S., Caesar, H., Zhou, L., Yang, J., and Beijbom, O. Pointpillars: Fast encoders for object detection from point clouds. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 12697–12705, 2019.

Li, R., Li, X., Heng, P.-A., and Fu, C.-W. Pointaugment: an auto-augmentation framework for point cloud classification. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 6378–6387, 2020.

Liu, J., Ni, B., Li, C., Yang, J., and Tian, Q. Dynamic points agglomeration for hierarchical point sets learning. In Proceedings of the IEEE International Conference on Computer Vision, pp. 7546–7555, 2019a.

Liu, M., Sheng, L., Yang, S., Shao, J., and Hu, S.-M. Morphing and sampling network for dense point cloud completion. In Proceedings of the AAAI Conference on Artificial Intelligence, volume 34, pp. 11596–11603, 2020a.

Liu, Y., Fan, B., Meng, G., Lu, J., Xiang, S., and Pan, C. Densepoint: Learning densely contextual representation for efficient point cloud processing. In Proceedings of the IEEE International Conference on Computer Vision, pp. 5239–5248, 2019b.

Liu, Y., Fan, B., Xiang, S., and Pan, C. Relation-shape convolutional neural network for point cloud analysis. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 8895–8904, 2019c.
Liu, Z., Hu, H., Cao, Y., Zhang, Z., and Tong, X. A closer look at local aggregation operators in point cloud analysis. In European Conference on Computer Vision, pp. 326–342. Springer, 2020b.

Paszke, A., Gross, S., Chintala, S., Chanan, G., Yang, E., DeVito, Z., Lin, Z., Desmaison, A., Antiga, L., and Lerer, A. Automatic differentiation in pytorch. 2017.

Qi, C. R., Su, H., Mo, K., and 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, pp. 652–660, 2017a.

Qi, C. R., Yi, L., Su, H., and Guibas, L. J. Pointnet++: Deep hierarchical feature learning on point sets in a metric space. Advances in neural information processing systems, 30:5099–5108, 2017b.

Rao, Y., Lu, J., and Zhou, J. Global-local bidirectional reasoning for unsupervised representation learning of 3d point clouds. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 5376–5385, 2020.

Savva, M., Yu, F., Su, H., Aono, M., Chen, B., Cohen-Or, D., Deng, W., Su, H., Bai, S., Bai, X., et al. Shrec16 track: largescale 3d shape retrieval from shapenet core55. In Proceedings of the eurorraphics workshop on 3D object retrieval, volume 10, 2016.

Shi, S., Guo, C., Jiang, L., Wang, Z., Shi, J., Wang, X., and Li, H. Pv-rcnn: Point-voxel feature set abstraction for 3d object detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 10529–10538, 2020.

Taghanaki, S. A., Hassani, K., Jayaraman, P. K., Khasahmadi, A. H., and Custis, T. Pointmask: Towards interpretable and bias-resilient point cloud processing. arXiv preprint arXiv:2007.04525, 2020.

Te, G., Hu, W., Zheng, A., and Guo, Z. Rgcn: Regularized graph cnn for point cloud segmentation. In Proceedings of the 26th ACM international conference on Multimedia, pp. 746–754, 2018.

Thomas, H., Qi, C. R., Deschaud, J.-E., Marcotegui, B., Goulette, F., and Guibas, L. J. Kpconv: Flexible and deformable convolution for point clouds. In Proceedings of the IEEE International Conference on Computer Vision, pp. 6411–6420, 2019.

Tsai, T., Yang, K., Ho, T.-Y., and Jin, Y. Robust adversarial objects against deep learning models. In Proceedings of the AAAI Conference on Artificial Intelligence, volume 34, pp. 954–962, 2020.

Uy, M. A., Pham, Q.-H., Hua, B.-S., Nguyen, T., and Yeung, S.-K. Revisiting point cloud classification: A new benchmark dataset and classification model on real-world data. In Proceedings of the IEEE International Conference on Computer Vision, pp. 1588–1597, 2019.

Verma, V., Lamb, A., Beckham, C., Najafi, A., Mitliagkas, I., Lopez-Paz, D., and Bengio, Y. Manifold mixup: Better representations by interpolating hidden states. In International Conference on Machine Learning, pp. 6438–6447. PMLR, 2019.

Wang, C., Samari, B., and Siddiqi, K. Local spectral graph convolution for point set feature learning. In Proceedings of the European conference on computer vision (ECCV), pp. 52–66, 2018.

Wang, Y., Sun, Y., Liu, Z., Sarma, S. E., Bronstein, M. M., and Solomon, J. M. Dynamic graph cnn for learning on point clouds. Acm Transactions On Graphics (tog), 38 (5):1–12, 2019.

Wu, Z., Song, S., Khosla, A., Yu, F., Zhang, L., Tang, X., and Xiao, J. 3d shapenet: A deep representation for volumetric shapes. In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 1912–1920, 2015.

Wu, Z., Duan, Y., Wang, H., Fan, Q., and Guibas, L. J. If-defense: 3d adversarial point cloud defense via implicit function based restoration. arXiv preprint arXiv:2010.05272, 2020.

Xiang, C., Qi, C. R., and Li, B. Generating 3d adversarial point clouds. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 9136–9144, 2019.

Yan, X., Zheng, C., Li, Z., Wang, S., and Cui, S. 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, pp. 5589–5598, 2020.

Yang, J., Zhang, Q., Fang, R., Ni, B., Liu, J., and Tian, Q. Adversarial attack and defense on point sets. arXiv preprint arXiv:1902.10899, 2019a.

Yang, J., Zhang, Q., Ni, B., Li, L., Liu, J., Zhou, M., and Tian, Q. Modeling point clouds with self-attention and gumbel subset sampling. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3323–3332, 2019b.

Yun, S., Han, D., Oh, S. J., Chun, S., Choe, J., and Yoo, Y. Cutmix: Regularization strategy to train strong classifiers with localizable features. In Proceedings of the IEEE International Conference on Computer Vision, pp. 6023–6032, 2019.
Zhang, H., Cisse, M., Dauphin, Y. N., and Lopez-Paz, D. mixup: Beyond empirical risk minimization. arXiv preprint arXiv:1710.09412, 2017.

Zhao, H., Jiang, L., Fu, C.-W., and Jia, J. Pointweb: Enhancing local neighborhood features for point cloud processing. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 5565–5573, 2019.

Zhao, Y., Wu, Y., Chen, C., and Lim, A. On isometry robustness of deep 3d point cloud models under adversarial attacks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 1201–1210, 2020.

Zheng, T., Chen, C., Yuan, J., Li, B., and Ren, K. Pointcloud saliency maps. In Proceedings of the IEEE International Conference on Computer Vision, pp. 1598–1606, 2019.

Zhou, H., Chen, K., Zhang, W., Fang, H., Zhou, W., and Yu, N. Dup-net: Denoiser and upsampler network for 3d adversarial point clouds defense. In Proceedings of the IEEE International Conference on Computer Vision, pp. 1961–1970, 2019.