

On Automatic Data Augmentation for 3D Point Cloud Classification

Wanyue Zhang*
wzhang@mpi-inf.mpg.de

Xun Xu†
xux@i2r.a-star.edu.sg

Fayao Liu
liu_fayao@i2r.a-star.edu.sg

Le Zhang‡
zhangleuestc@gmail.com

Chuan-Sheng Foo
foo_chuan_sheng@i2r.a-star.edu.sg

Institute for Infocomm Research
A*STAR, Singapore

Abstract

Data augmentation is an important technique to reduce overfitting and improve learning performance, but existing works on data augmentation for 3D point cloud data are based on heuristics. In this work, we instead propose to automatically learn a data augmentation strategy using bilevel optimization. An augmentor is designed in a similar fashion to a conditional generator and is optimized by minimizing a base model’s loss on a validation set when the augmented input is used for training the model. This formulation provides a more principled way to learn data augmentation on 3D point clouds. We evaluate our approach on standard point cloud classification tasks and a more challenging setting with pose misalignment between training and validation/test sets. The proposed strategy achieves competitive performance on both tasks and we provide further insight into the augmentor’s ability to learn the validation set distribution.

1 Introduction

Understanding 3D point clouds is crucial to a wide range of applications including autonomous driving, robotics and human-computer interaction. Recent progress in this area is attributed to training deep neural networks on labeled data [22, 23]. In order to increase the diversity of training data to aid generalization, data augmentation has been used in the training of neural networks for 3D point clouds to much success [8, 12, 33]. Current data augmentation methods for point clouds can be broadly classified into heuristic methods and learning-based methods. In the former category, mixup [8, 11, 32] between multiple shapes is often used to synthesize novel training examples. Despite being able to provide consistent

performance improvements, potentially due to stronger regularization with label smoothing, the data sample after mixup often does not resemble a realistic object. For instance, when applied to a plane and a car, mixup does not yield a semantically meaningful object; the reason behind the success of mixup remains unclear. Learning-based augmentation methods, on the other hand, automatically learn an augmentation strategy instead of hand-engineering one. PointAugment [12] trains a neural network as the augmentor to predict the rotation and additive jittering parameters for each sample. A heuristic learning objective encouraging augmented samples to be different from the original is used to update the parameters of the augmentor. While PointAugment has showed improved performance over baseline methods, its learning objective does not necessarily guarantee that the augmented sample will benefit the main task.

By contrast, in the image domain, recent automatic data augmentation methods adopt a bilevel optimization framework [6, 13, 14] where the learnt augmentation strategy minimizes a base model’s loss on validation data while the model is trained using this strategy. Solving the bilevel optimization problem then involves iterating between training the base model and updating augmentation policies. Inspired by the recent success of bilevel optimization for automatic data augmentation on images [13, 14], we propose a novel 3D point cloud augmentation method, **Automatic data augmentation for 3D Point Cloud data (AdaPC)**. We define the augmentation procedure as sampling from a sequence of transformation operations each parameterized by a distribution, then apply the transformations to each training sample. To allow for a high degree of flexibility for augmentation, we use a neural network termed as the *augmentor* to model such distributions. The augmentor parameters are then learned using a bilevel optimization framework by minimizing a base classifier’s loss on validation data when the model is trained using the augmentor. For computational efficiency, the hypergradients, i.e the gradients of the augmentor parameters w.r.t. the validation loss, are approximated by a one-step unrolling algorithm [15] and augmentor parameters are updated with gradient descent.

AdaPC has several advantages over existing methods for data augmentation on 3D point clouds. First, compared to heuristic approaches [6, 13], AdaPC is able to learn the optimal augmentation for a particular dataset as it directly minimizes the validation loss, while heuristic approaches may only work well in certain cases and require extensive manual hyperparameter tuning. Second, compared to PointAugment [12], AdaPC optimizes the augmentor in the direction that minimizes validation error which makes it more likely to generalize to test data. More importantly, if the training set distribution does not exactly match that of the validation/test set (e.g. due to pose mismatch), AdaPC can learn to bridge the gap by capturing any misalignment with the augmentor; we demonstrate that AdaPC is indeed superior in an experimental setup with non-fixed pose. Insight into the effectiveness of the augmentor is also provided by analyzing the distributions learnt by the augmentor.

We summarize the contributions of this work below:

- We propose an automatic data augmentation method for 3D point cloud classification by training the augmentor and classifier jointly through bilevel optimization. Compared to existing augmentation methods on 3D point clouds, we provide a more principled way for learning data augmentation strategies.
- We achieve superior results on two standard 3D point cloud classification benchmarks compared to other data augmentation methods.
- We demonstrate that our augmentor is able to learn meaningful augmentation policies

when there is a pose mismatch between training and validation/test sets and provide insight into the distribution of augmentation learned under this scenario.

2 Related work

2.1 Data Augmentation for 3D Point Clouds

Deep learning methods for 3D point clouds [1, 21, 22, 23, 25] commonly incorporate small perturbations such as scaling, jittering, point dropout, flipping and rotation to enhance the diversity of the training set. Recently, more sophisticated data augmentation strategies have been studied. Inspired by the success of mixing up images as augmentation [27, 32], PointMixUp [3] proposed to interpolate two point clouds by finding a shortest path between the two shapes on the manifold. PointCutMix [33] further extended CutMix [31] by first finding correspondences between 2 point clouds and then swapping selected regions. Despite demonstrating improvements on standard point cloud datasets, both works are based on heuristics and do not explain the success of the proposed methods. Orthogonal to heuristic augmentation strategies, PointAugment [12] formulates augmentation as learning an augmentor which is reminiscent of PointNet. However, it employs a hand-engineered adversarial loss function to train the augmentor which does not necessarily guarantee that the learned augmentation is beneficial and often leads to unstable results. For 3D object detection, PA-AUG [5] divides object bounding boxes into fixed partitions and stochastically applies 5 basic transformations such as noise, points dropout, sampling, cutmix [31], cutmixup [30]. Progressive population based augmentation (PPBA) [4] uses evolutionary search to narrow down the search space of optimal augmentation policies. In this work, we propose to formulate learning data augmentation in a more principled way. The base model’s loss on a validation set is exploited to guide the learning of augmentation strategies that results in improved performance and better generalization to more challenging scenarios.

2.2 Automatic Data Augmentation

Hand-crafted augmentation rules can be ineffective in selecting the optimal combination of augmentation policies as well as their magnitudes. SmartAugment [11] tackles the problem by training an augmentor to combine samples from the same class alongside the target network. There is also a line of GANs-based approaches [11, 19, 20, 24] which generate augmented samples according to the feedback from the main task. Another class of methods learns an augmentation strategy based on the validation set. AutoAugment [6] adopts reinforcement learning (RL) to select augmentation policies which improve validation accuracy. RandAugment [7] eliminates the search phase used by previous automatic augmentation methods and conducts a grid search to find the augmentation policies. In light of slow convergence of RL-based approaches, FastAutoAugment [12] uses Bayesian optimization to efficiently learn augmentation strategies. DADA [13] further boosted efficiency by introducing a gradient-based policy searching paradigm and employed bilevel optimization to update augmentation policies. DABO [18] proposed a differentiable augmentor which is capable of generating affine and color transformations. MetaMixUp [16] focused on finding a dynamic and adaptative interpolation ratio in an online fashion. In this work, we take a similar approach to optimize the augmentor based on the validation performance using bilevel optimization. For computational efficiency, the hypergradients are approximated using the algorithm described in [15].

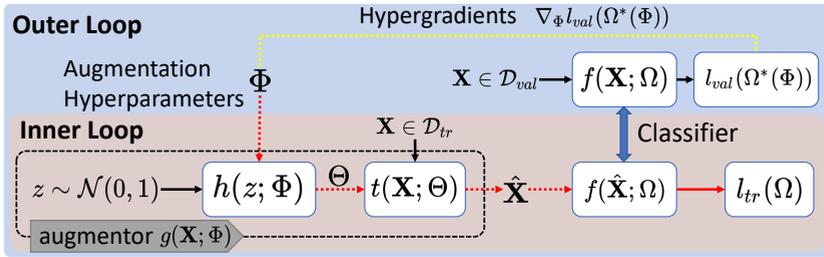


Figure 1: Bilevel optimization framework for automatic data augmentation. The solid red arrow indicates a forward pass in the inner loop. The red dashed arrow indicates the update path for outer loop optimization. We use the yellow dashed line to highlight that the hypergradients are approximated instead of being computed by automatic differentiation.

3 Methodology

We first provide an overview of AdaPC in Fig. 1. During training, AdaPC iterates between an inner loop and outer loop optimization. The inner loop optimization involves updating the base classifier network $f(\cdot; \Omega)$ by minimizing the training loss l_{tr} . The forward pass for inner loop optimization is indicated by the solid red arrow. The outer loop optimization updates the data augmentation parameters (or hyperparameters) Φ . Since the validation loss $l_{val}(\Omega^*(\Phi))$ implicitly depends on Φ , we approximate the hypergradients following [13]. The outer loop update path is indicated by the red dashed arrows. In the following section, we first introduce our bilevel optimization framework in Sect. 3.1 followed by the specific designs of the augmentor in Sect. 3.2

3.1 Learning Data Augmentation with Bilevel Optimization

We introduce the bilevel optimization framework for learning data augmentation. We define the augmentation as a mapping $g: \mathbf{X} \rightarrow \hat{\mathbf{X}}$ parameterized by Φ (the augmentation hyperparameters), which is to be learned. This augmentation often involves randomness with instantiations explained in Sect. 3.2. Let $\mathcal{D}_{tr} = \{\mathbf{X}_{tr}, y_{tr}\}$ and $\mathcal{D}_{val} = \{\mathbf{X}_{val}, y_{val}\}$ denote training and validation sets, $f: \mathbf{X} \rightarrow y$ parameterized by Ω denote the classification network, $l_{val}(\Omega; \mathcal{D}_{val}) = \frac{1}{|\mathcal{D}_{val}|} \sum_{\mathbf{X}_i, y_i \in \mathcal{D}_{val}} CE(f(\mathbf{X}_i; \Omega), y_i)$ denote the validation loss and $l_{tr}(\Omega, \Phi; \mathcal{D}_{tr})$ denote the training loss. Since augmentation is turned off during validation, l_{val} does not explicitly depend on Φ . We now formally write the bilevel optimization objective in Eq. 1 where the left problem is often called the outer loop and the right problem is the inner loop.

$$\min_{\Phi} l_{val}(\Omega^*(\Phi); \mathcal{D}_{val}), \quad s.t. \quad \Omega^*(\Phi) = \arg \min_{\Omega} l_{tr}(\Omega, \Phi; \mathcal{D}_{tr}) \quad (1)$$

The above objective aims to find the optimal augmentation hyperparameters Φ such that when the model is trained with Φ (to obtain model parameters Ω^*) its loss on the validation data is minimized. If we assume that the validation set has the same distribution as the test set, this formulation is more principled than PointAugment [14] in guaranteeing better generalization. Similar optimization problems have been formulated for neural architecture search [15] and automatic data augmentation [16]. Solving the inner loop until convergence (training the model to convergence) followed by updating the outer loop leads to a very

inefficient optimization paradigm as the inner loop usually takes many steps to converge. We therefore instead used a more efficient approximation algorithm proposed in [15].

Let ξ denote the learning rate for the inner loop optimizer. One step of gradient descent on the classification network parameters can then be written as $\Omega' = \Omega - \xi \nabla_{\Omega} l_{tr}(\Phi)$. By approximating the best response (optimal) model parameters Ω^* to hyperparameters Φ with parameters obtained from one step of gradient descent Ω' , we obtain approximate hypergradients using the chain rule as follows:

$$\nabla_{\Phi} l_{val}(\Omega^*(\Phi)) \approx \nabla_{\Phi} l_{val}(\Omega'(\Phi)) = -\xi \nabla_{\Phi, \Omega}^2 l_{tr}(\Omega, \Phi) \nabla_{\Omega'} l_{val}(\Omega') \quad (2)$$

We further use a central finite difference to approximate the Hessian-vector product. Let ε be a small perturbation, $\Omega^+ = \Omega + \varepsilon \nabla_{\Omega'} l_{val}(\Omega')$ and $\Omega^- = \Omega - \varepsilon \nabla_{\Omega'} l_{val}(\Omega')$. Then,

$$-\xi \nabla_{\Phi, \Omega}^2 l_{tr}(\Omega, \Phi) \nabla_{\Omega'} l_{val}(\Omega') \approx -\xi \frac{\nabla_{\Phi} l_{tr}(\Omega^+, \Phi) - \nabla_{\Phi} l_{tr}(\Omega^-, \Phi)}{2\varepsilon}. \quad (3)$$

We follow DARTS [15] to set ξ to the learning rate of the classifier and ε to $\frac{0.01}{\|\nabla_{\Omega'} l_{val}(\Omega')\|}$. Since the hypergradients tend to be noisy if the hyperparameters are initialized far from a local optimum, we include an L2 regularization term, $l_{reg} = \|\Theta - \hat{\Theta}\|_2^2$, to the output of the augmentation network to smooth the hypergradients. Prior knowledge $\hat{\Theta}$ about the hyperparameters can be incorporated using this regularizer, e.g. for scaling, $\hat{\theta} = 1$ and for rotation, $\hat{\theta} = 0$. Overall, the final hypergradients with regularization are given by Eq. 4 where λ is the weight of the regularization term, and the training procedure is presented in Algo. 1.

$$\nabla_{\Phi} l_{val}(\Omega^*(\Phi)) \approx -\xi \nabla_{\Phi, \Omega}^2 l_{tr}(\Omega, \Phi) \nabla_{\Omega'} l_{val}(\Omega') + 2\lambda(\Theta - \hat{\Theta}) \quad (4)$$

Algorithm 1: Algorithm for learning model weights Ω and augmentor weights Φ

Initialize classifier weights Ω and augmentor hyperparameters Φ ;

while not converged **do**

 One-step inner loop update $\Omega' = \Omega - \xi \nabla_{\Omega} l_{tr}(\Phi)$;

 One-step outer loop update $\Phi' = \Phi - \alpha \nabla_{\Phi} l_{val}(\Omega'(\Phi))$;

 Update classifier weights $\Omega = \Omega - \xi \nabla_{\Omega} l_{tr}(\Phi')$;

end

return Ω, Φ

3.2 Augmentation Operations

In this section, we introduce the design of an augmentor which can be easily incorporated into the bilevel optimization framework. Inspired by recent studies into deep learning on 3D point clouds [12], we apply two types of augmentation operations. First, we apply a point-wise jittering $\mathbf{J} \in \mathbb{R}^{N \times 4}$ (in homogeneous coordinates) which can be seen as a deformation of the original 3D shape. To limit the learnable parameters for jittering, we sample jittering from a uniform distribution $\mathcal{U}(\theta_j, \theta_j + v)$ where θ_j is a learnable parameter and v determines a fixed range. Then, we further apply a rigid transformation, represented in homogeneous coordinates as $\mathbf{H} \in \mathbb{R}^{4 \times 4}$ to the perturbed 3D shape. Instead of directly regressing a transformation matrix like PointAugment, we parameterize $\mathbf{H} = \mathbf{T} \cdot \mathbf{R} \cdot \mathbf{S}$ where \mathbf{R} is a rotation matrix parameterized by $\theta_r \in \mathbb{R}^1$, \mathbf{S} is a scaling matrix parameterized by $\theta_s \in \mathbb{R}^3$ and \mathbf{T} is a translation matrix parameterized by $\theta_t \in \mathbb{R}^3$. Such a decomposition allows the augmentor

to directly regress individual parameters; as a result we can easily apply regularization to individual operations and each operation can be randomly dropped out to create more diverse augmentations. In contrast, it is difficult to incorporate explicit regularization when directly regressing onto \mathbf{H} . For brevity, we denote $\Theta = [\theta_j, \theta_r, \theta_s, \theta_t] \in \mathbb{R}^8$ and the transformation operation is described by Eq. 5 as follows:

$$t(\mathbf{X}; \Theta) = (\mathbf{X} + \mathbf{J}(\Theta)) \cdot \mathbf{H}(\Theta)^\top. \quad (5)$$

One approach to obtain Θ is to learn a fixed vector that applies to all training examples. However, this does not allow randomness in the augmentation procedure. Another approach is to model Θ using a parametric distribution $p(\Theta)$, e.g. Gaussian or uniform distribution, which is easy to parameterize and sample from. For example, if a Gaussian distribution is used, we parameterize it as $\theta_i \sim \mathcal{N}(\mu, \sigma)$ where μ and σ are hyperparameters to optimize and $\Omega = [\mu, \sigma]$. We discuss such augmentor designs in more detail in Section S1 of the supplementary material. Despite being more flexible, it is not trivial to specify the most appropriate distribution without any prior knowledge. To further increase the flexibility of the distribution used to model Θ , we propose to use a neural network denoted as $h(z; \Phi)$, as illustrated in Fig. 1. The neural network takes a random noise vector with components sampled from a zero-mean, unit variance Gaussian distribution $z \sim \mathcal{N}(0, 1)$ and transforms it into augmentation parameters $\Theta = h(z; \Phi)$. This network is similar to a generator in a GAN [9] which maps a simple distribution to a more complicated one. The generator is trained by optimizing the performance on the validation set. Overall, the augmentor $g(\mathbf{X}; \Phi)$ can be seen as a conditional generator [10] and augmentation is regarded as a conditional generative process.

4 Experiments

4.1 Datasets

We evaluate on **ModelNet40** [24] for shape classification. It consists of 12311 CAD models from 40 categories of which 9843 models are used for training and 2468 for testing. We use a pre-processed dataset which contains uniformly sampled points from the mesh surfaces as input to our network. Following the standard practice, we use 1024 points per shape.

To test the robustness of our method on objects cropped from real 3D scans, we use the OBJ_ONLY variant provided by **ScanObjectNN** [26] without the background. OBJ_ONLY contains objects with noisy surfaces of non-uniform density from 15 categories.

Our method requires a validation set which is not provided by the above 2 datasets. Hence, we use 90% of the original training data as the new training data, and the remaining 10% as the validation data. To ensure a fair comparison with other methods, we randomly re-split the training and validation set every epoch so that that our network has access to the entire training data across epochs.

4.2 Implementation details

The initial learning rate for both classifier ξ and augmentor α are 0.001 with Adam optimizer ($\beta_1 = 0.9, \beta_2 = 0.999$). The classifier learning rate is halved every 20 epochs. The classifier and augmentor are updated every minibatch and we train for 300 epochs in total. We use a batch size of 24 for PointNet and 12 for PointNet++ and DGCNN. The regularization weight λ is a manually tunable hyperparameter. A proper selection of λ is vital to the stability of training. We use $\lambda = 0.5$ for ModelNet40 and $\lambda = 10.0$ for ScanObjectNN. We use cross-entropy loss for classification. We adopt other training practices from PointAugment for

fair comparison, e.g. adding a term to penalize the feature difference between the original and augmented samples, apply dropout to the augmentation types and mix original data and augmented data during training.

4.3 Automatic Augmentation for Classification

We first report results on standard benchmark datasets for classification on point clouds, ModelNet40 and ScanObjectNN using PointNet as the backbone [22] in Tab. 1. We compare the following approaches in this experiment. First, we evaluate three baseline methods with a fixed predefined augmentor, referred to as Fixed Aug. in Tab. 1. Vanilla PointNet without any augmentation, PointNet with predefined scaling and translation as augmentation, referred to as Sampling-S,T in Tab. 1 and PointNet with predefined scaling, translation and rotation, referred to as Sampling-S,T,R are respectively evaluated. For predefined scaling (S), we adopt the practice in PointAugment [22] by randomly sampling from a uniform distribution $\mathcal{U}(0.67, 1.5)$. For predefined translation (T), we uniformly sample from $\mathcal{U}(-0.2, 0.2)$. For predefined rotation about the y-axis (R_y), we sample from $\mathcal{N}(0, 0.06^2)$ with output clipped at 0.18. For predefined jittering (J), we sample from $\mathcal{N}(0, 0.01^2)$ with output clipped at 0.05. We further evaluate state-of-the-art point cloud data augmentation approaches, including PointMixup [9] (PointMixup-Mixup) and PointAugment [22] (Sampling-S,T-PointAug.-R,J). We notice from the released code that sampling from predefined scaling and translation is adopted by PointAugment. Finally, we evaluate two variants of our proposed approach: one that includes scaling, translation and rotation operations (AdaPC-S,T,R) and another that further includes jittering (AdaPC-S,T,R,J).

We make the following observations from Tab.1. First, with predefined augmentation strategies, classification performance is consistently improved compared to the baseline. The improvement is more significant on the dataset scanned from the real world (ScanObjectNN). This is partially explained by the increased diversity of data samples in the real-world dataset. Second, when jittering is naively incorporated, performance drops on ScanObjectNN as seen by comparing Sampling-S,T (80.20%) v.s. Sampling-S,T,J (76.93%). This is consistent with many observations on data augmentation for 3D point clouds that jittering may not be helpful. PointAugment managed to learn jittering policies with heuristic objectives but still failed to improve over the baseline (Sampling-S,T) on ScanObjectNN. In contrast, with a more principled augmentation objective, our AdaPC achieves the best performance on both datasets with S,T,R. We would like to highlight that AdaPC’s 0.7% improvement over PointAugment is non-trivial given the limited diversity and highly controlled setting of the ModelNet40 dataset. When jittering is incorporated, AdaPC further improves on ScanObjectNN. Finally, we evaluate AdaPC on two additional backbones, PointNet++ and DGCNN and present results in Tab. 2. AdaPC achieves results superior to, if not comparable with, the predefined augmentation baseline, PointMixup and PointAugment.

4.4 Classification under Non-Fixed Poses

In the previous setting, the poses of CAD models are all perfectly aligned, thus reducing the effect of data augmentation. Thus, here we consider a more challenging scenario where both training, validation and test data are not perfectly aligned and then investigate the effectiveness of learning the augmentation. Since variations in scaling and translation can often be eliminated by normalization, we consider rotation about the y-axis (gravity) only in this experiment.

		ModelNet40				ScanObjectNN	
		Predefined Aug.		Learnable Aug.		Test Acc.	Test Acc.
		Method	Ops	Method	Ops		
Existing	-	-	-	-	-	89.58	77.96
	Sampling	S, T	-	-	-	90.55	80.20
	Sampling	S, T, R	-	-	-	90.51	77.96
	Sampling	S, T, J	-	-	-	90.88	76.93
	PointMixup	Mixup	-	-	-	89.90	-
	Sampling	S, T	PointAug.	R, J	-	90.90	79.34
Ours	-	-	AdaPC	S, T, R	-	91.61	79.86
	-	-	AdaPC	S, T, R, J	-	90.80	81.75

Table 1: Results on 3D point cloud classification. Method and Ops indicate respectively the method adopted for augmentation and the specific operations used for augmentation.

		PointNet++				DGCNN	
		Predefined Aug.		Learnable Aug.		Test Acc.	Test Acc.
		Method	Ops	Method	Ops		
-	-	-	-	-	-	91.41	91.85
Sampling	S, T, R	-	-	-	-	92.26	92.38
PointMixup	Mixup	-	-	-	-	92.70	-
Sampling	S, T	PointAugment*	R, J	-	-	92.82	92.70
-	-	AdaPC	S, T, R	-	-	92.94	92.82

Table 2: Additional backbones on ModelNet40 (*for fair comparison, we report the reproduced results for PointAugment under the setting where batch size is 12 and Adam optimizer is used.)

Concretely, we randomly rotate each example by sampling an angle $\theta_r \sim \mathcal{N}(1.0, 0.2^2)$. We compare AdaPC to vanilla PointNet without augmentation, PointNet using rotation augmentation with angle $\theta \sim \mathcal{U}(-\delta, \delta)$ for several values of δ , PointAugment [10] (PointAug.-R,J), a state-of-the-art SO(3) rotation invariant network VectorNeurons [8] and our final model (AdaPC-S,T,R). The results are shown in Tab. 3. We observe that including augmentation with a small rotation ($\delta = 0.1$) increases baseline model test accuracy from 87.07% (without augmentation) to 89.54%, while large rotations ($\delta = 3.14$) harm performance; our results suggest that including small pose perturbations is helpful for generalization. PointAugment produces results close to the best manually defined augmentation but is still outperformed by AdaPC. Finally, we observe a much weaker performance by VectorNeurons despite its robustness to SO(3) rotation.

4.5 Pose Mismatch between Training and Testing

In this section, we simulate the challenging setting where there is a pose mismatch between training and validation/testing sets, which may occur in practice as detecting and aligning poses for 3D point clouds is a non-trivial task [28]. We rotate the test and validation sets by 0.35/0.70/1.05/1.4 radians about the y-axis while the training set remains unchanged. We note that the validation and test sets should be drawn from the same distribution for bilevel optimization to work. Here we assume no prior knowledge on rotation angles and hence remove the regularization term. We evaluate 4 methods in this experiment: 1) the “None” baseline that does not apply any augmentation, 2) “Same rotation”, that applies the ground-truth rotation to the training data and thus serves as an upper bound, 3) “Random rotation”,

Augmentation		Test Acc.
Method	Ops	
-	-	87.07
Sampling	$R(\theta_r \sim \mathcal{U}(-0.05, 0.05))$	89.42
Sampling	$R(\theta_r \sim \mathcal{U}(-0.1, 0.1))$	89.54
Sampling	$R(\theta_r \sim \mathcal{U}(-0.17, 0.17))$	88.81
Sampling	$R(\theta_r \sim \mathcal{U}(-3.14, 3.14))$	77.87
PointAug.	R, J	89.54
VectorNeurons	-	81.68
AdaPC	S, T, R	90.92

Table 3: Classification results on ModelNet40 under non-fixed poses.

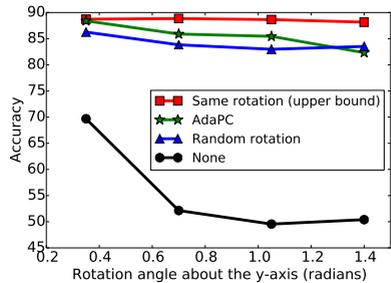


Figure 2: Test accuracy on ModelNet40 at various test rotations

that applies a random rotation $\theta_r \sim \mathcal{U}(-3.14, 3.14)$ to the training data, and 4) AdaPC with only rotation augmentation. Fig. 2 shows the test accuracy on ModelNet40 at various rotation angles. We see that AdaPC (green line) largely outperforms “Random rotation” and is always much better than the no-augmentation “None” baseline.

We further provide insight into the distribution learned by the proposed augmentor by visualizing the augmentor’s output distribution over the course of training in Fig. 3. At the beginning, the augmentor’s output closely resembles a Gaussian distribution. However, as training progresses, the mean of the augmentor distribution converges to the ground-truth pose. The spread of the distribution at the best epoch suggests that randomness is essential to the success of augmentation.

We also explored if the augmentor can be trained using a subset of the training data and transferred to the full dataset under this setting, as this could accelerate training of the augmentor. Results from this experiment are discussed in Section S2 of the supplementary material and suggest the potential viability of such an approach.

Aug. Ops	S	-	✓	-	-	-	✓	✓	✓
	T	-	-	✓	-	-	✓	✓	✓
	R	-	-	-	✓	-	-	✓	✓
	J	-	-	-	-	✓	-	-	✓
ModelNet40	89.58	90.19	90.15	90.15	89.82	90.84	91.61	90.80	
ScanObjectNN	77.96	77.62	80.72	75.59	76.62	78.14	79.86	81.75	

Table 4: Ablation study on individual augmentation operations.

4.6 Ablation Study

4.6.1 Different combinations of augmentation operations

Here we explore the effectiveness of various combinations of augmentation operations on both ModelNet40 and ScanObjectNN datasets. From the results in Tab. 4.5, we see that using each of scaling, translation, rotation, and jittering alone outperforms the no-augmentation baseline on ModelNet40, and that the best results are achieved by combining the first 3 types

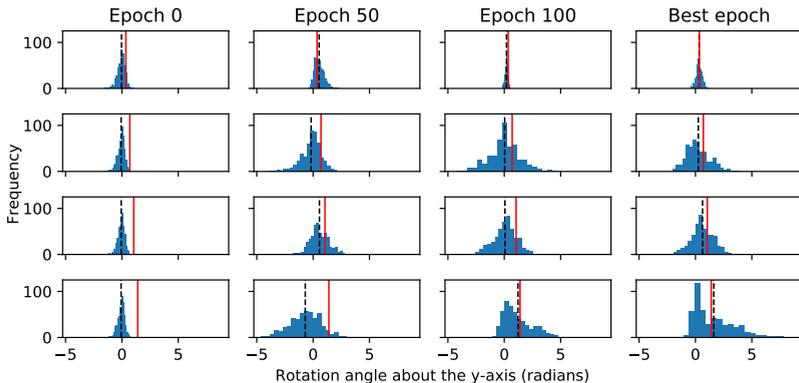


Figure 3: Rotation distributions learned by the augmentor over training epochs. The dashed black vertical line indicates the mean of the learned distribution and the red vertical line indicates the ground-truth rotation applied to the test set. From top to bottom, a 0.35/0.70/1.05/1.4 radian rotation is respectively applied to the validation/test sets.

of augmentations (91.61% on ModelNet40). We observe a slight drop in performance on ModelNet40 when further including jittering. One possible reason for this drop in performance is that ModelNet40 consists of points uniformly sampled from CAD models that are noise-free. Hence, training samples augmented with jittering deviate from the distribution of noise-free test samples, causing a drop in performance. In contrast, ScanObjectNN is cropped from real-world scenes, thus shapes are naturally contaminated by noise and jittering is more effective. Our observations suggest that research into domain-specific augmentation strategies will be valuable.

5 Conclusion

Data augmentation plays an important role for deep learning on 3D point clouds. Different from existing heuristic augmentation methods, we propose an automatic data augmentation method, AdaPC, based on a more principled bilevel optimization framework. In AdaPC, an augmentor network is used to predict transformation parameters for augmentation. This augmentor is then learned by minimizing the validation loss. AdaPC achieved very competitive results on two point cloud classification datasets. We also demonstrate that AdaPC is effective in a pose mismatch scenario, where the mean of the learned augmentation distribution converged to the true pose. Our work suggests that future works on point cloud augmentation should include evaluations on challenging real-world datasets and settings to demonstrate the true value of data augmentation.

Acknowledgement: This research is supported by the Agency for Science, Technology and Research, Singapore (A*STAR) under its Career Development Award (Grant No. 202D8243). Discussions with Xiatian Zhu are gratefully acknowledged.

References

- [1] Antreas Antoniou, Amos Storkey, and Harrison Edwards. Data augmentation generative adversarial networks. *arXiv preprint arXiv:1711.04340*, 2017.
- [2] Matan Atzmon, Haggai Maron, and Yaron Lipman. Point convolutional neural networks by extension operators. *ACM Transactions on Graphics*, 2018.
- [3] Yunlu Chen, Vincent Tao Hu, Efstratios Gavves, Thomas Mensink, Pascal Mettes, Pengwan Yang, and Cees GM Snoek. Pointmixup: Augmentation for point clouds. In *European Conference on Computer Vision*, 2020.
- [4] Shuyang Cheng, Zhaoqi Leng, Ekin Dogus Cubuk, Barret Zoph, Chunyan Bai, Jiquan Ngiam, Yang Song, Benjamin Caine, Vijay Vasudevan, Congcong Li, et al. Improving 3d object detection through progressive population based augmentation. In *European Conference on Computer Vision*, 2020.
- [5] Jaeseok Choi, Yeji Song, and Nojun Kwak. Part-aware data augmentation for 3d object detection in point cloud. *arXiv preprint arXiv:2007.13373*, 2020.
- [6] Ekin D Cubuk, Barret Zoph, Dandelion Mane, Vijay Vasudevan, and Quoc V Le. Autoaugment: Learning augmentation policies from data. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018.
- [7] Ekin D Cubuk, Barret Zoph, Jonathon Shlens, and Quoc V Le. Randaugment: Practical automated data augmentation with a reduced search space. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2020.
- [8] Congyue Deng, Or Litany, Yueqi Duan, Adrien Poulénard, Andrea Tagliasacchi, and Leonidas Guibas. Vector neurons: A general framework for so (3)-equivariant networks. *arXiv preprint arXiv:2104.12229*, 2021.
- [9] Ian J Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron C Courville, and Yoshua Bengio. Generative adversarial nets. In *NIPS*, 2014.
- [10] Dogyoon Lee, Jaeha Lee, Junhyeop Lee, Hyeongmin Lee, Minhyeok Lee, Sungmin Woo, and Sangyoung Lee. Regularization strategy for point cloud via rigidly mixed sample. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021.
- [11] Joseph Lemley, Shabab Bazrafkan, and Peter Corcoran. Smart augmentation learning an optimal data augmentation strategy. *IEEE Access*, 2017.
- [12] Ruihui Li, Xianzhi Li, Pheng-Ann Heng, and Chi-Wing Fu. Pointaugment: an auto-augmentation framework for point cloud classification. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020.
- [13] Yonggang Li, Guosheng Hu, Yongtao Wang, Timothy Hospedales, Neil M Robertson, and Yongxin Yang. Differentiable automatic data augmentation. In *European Conference on Computer Vision*, 2020.

- [14] Sungbin Lim, Ildoo Kim, Taesup Kim, Chiheon Kim, and Sungwoong Kim. Fast auto-augment. In *Advances in Neural Information Processing Systems*, 2019.
- [15] Hanxiao Liu, Karen Simonyan, and Yiming Yang. Darts: Differentiable architecture search. In *International Conference on Learning Representations*, 2018.
- [16] Zhijun Mai, Guosheng Hu, Dexiong Chen, Fumin Shen, and Heng Tao Shen. Metamixup: Learning adaptive interpolation policy of mixup with metalearning. *IEEE Transactions on Neural Networks and Learning Systems*, 2021.
- [17] Mehdi Mirza and Simon Osindero. Conditional generative adversarial nets. *arXiv preprint arXiv:1411.1784*, 2014.
- [18] Saypraseuth Mounsaveng, Issam Laradji, Ismail Ben Ayed, David Vazquez, and Marco Pedersoli. Learning data augmentation with online bilevel optimization for image classification. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2021.
- [19] Seongkyu Mun, Sangwook Park, David K Han, and Hanseok Ko. Generative adversarial network based acoustic scene training set augmentation and selection using svm hyper-plane. *Proc. DCASE*, 2017.
- [20] Luis Perez and Jason Wang. The effectiveness of data augmentation in image classification using deep learning. *Convolutional Neural Networks Vis. Recognit*, 2017.
- [21] Anh Viet Phan, Minh Le Nguyen, Yen Lam Hoang Nguyen, and Lam Thu Bui. Dgcnn: A convolutional neural network over large-scale labeled graphs. *Neural Networks*, 2018.
- [22] Charles R Qi, Hao Su, Kaichun Mo, and Leonidas J Guibas. Pointnet: Deep learning on point sets for 3d classification and segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017.
- [23] Charles R Qi, Li Yi, Hao Su, and Leonidas J Guibas. Pointnet++: Deep hierarchical feature learning on point sets in a metric space. In *Advances in Neural Information Processing Systems*, 2017.
- [24] Fabio Henrique Kiyoyiti dos Santos Tanaka and Claus Aranha. Data augmentation using gans. *arXiv preprint arXiv:1904.09135*, 2019.
- [25] Hugues Thomas, Charles R Qi, Jean-Emmanuel Deschaud, Beatriz Marcotegui, François Goulette, and Leonidas J Guibas. Kpconv: Flexible and deformable convolution for point clouds. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019.
- [26] Mikaela Angelina Uy, Quang-Hieu Pham, Binh-Son Hua, Thanh Nguyen, and Sai-Kit Yeung. 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*, 2019.
- [27] Vikas Verma, Alex Lamb, Christopher Beckham, Amir Najafi, Ioannis Mitliagkas, David Lopez-Paz, and Yoshua Bengio. Manifold mixup: Better representations by interpolating hidden states. In *International Conference on Machine Learning*, 2019.

- [28] He Wang, Srinath Sridhar, Jingwei Huang, Julien Valentin, Shuran Song, and Leonidas J. Guibas. Normalized object coordinate space for category-level 6d object pose and size estimation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019.
- [29] Zhirong Wu, Shuran Song, Aditya Khosla, Fisher Yu, Linguang Zhang, Xiaoou Tang, and Jianxiong Xiao. 3d shapenets: A deep representation for volumetric shapes. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015.
- [30] Jaejun Yoo, Namhyuk Ahn, and Kyung-Ah Sohn. Rethinking data augmentation for image super-resolution: A comprehensive analysis and a new strategy. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020.
- [31] Sangdoon Yun, Dongyoon Han, Seong Joon Oh, Sanghyuk Chun, Junsuk Choe, and Youngjoon Yoo. Cutmix: Regularization strategy to train strong classifiers with localizable features. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019.
- [32] Hongyi Zhang, Moustapha Cisse, Yann N Dauphin, and David Lopez-Paz. mixup: Beyond empirical risk minimization. In *International Conference on Learning Representations*, 2017.
- [33] Jinlai Zhang, Lvjie Chen, Bo Ouyang, Binbin Liu, Jihong Zhu, Yujing Chen, Yanmei Meng, and Danfeng Wu. Pointcutmix: Regularization strategy for point cloud classification. *arXiv preprint arXiv:2101.01461*, 2021.