Rotation Invariant Convolutions for 3D Point Clouds Deep Learning

Zhiyuan Zhang[1] Binh-Son Hua[2] David W. Rosen[1] Sai-Kit Yeung[3]

[1]Singapore University of Technology and Design [2]The University of Tokyo

[3]Hong Kong University of Science and Technology

3DV 2019 International Conference on 3D Vision Québec City, Canada September 16 - 19, 2019

Motivations and Contributions

Motivations:

- Most existing neural networks for 3D point cloud can NOT guarantee rotation invariance.
- Networks trained with simple ration augmentations are NOT able to generalize well to unseen rotations.
- A few existing convolution operators that allows rotation invariance still can NOT achieve consistent predictions with arbitrarily rotated data.

Main Contributions:

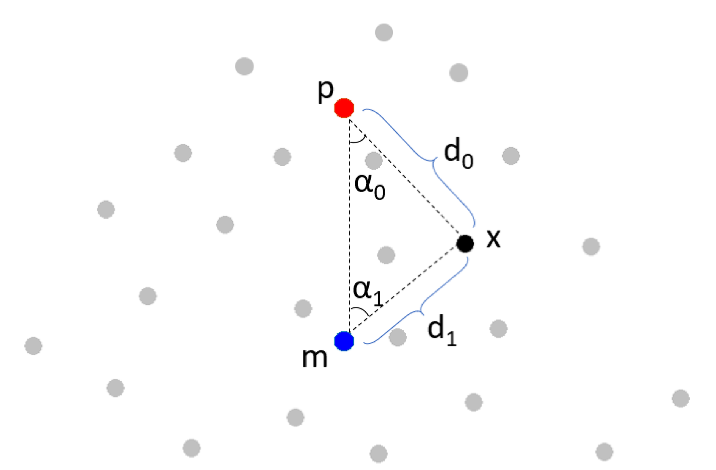
- A robust feature extraction scheme suitable for convolution that supports both rotation and translation invariant features based on low-level geometric cues;
- A novel convolution operator that is agnostic to both point cloud rotations and point orders.
- A simple binning approach that can be seamlessly combined with the feature extraction step;
- A compact convolutional neural network for object classification and object part segmentation achieving highly consistent and accurate performance under different rotations.

Method

Summary:

Rotation Invariant Features (RIF) → Rotation Invariant Convolution Operator (RIConv) → Rotation Invariant Neural Network (RINet)

Rotation Invariant Features:

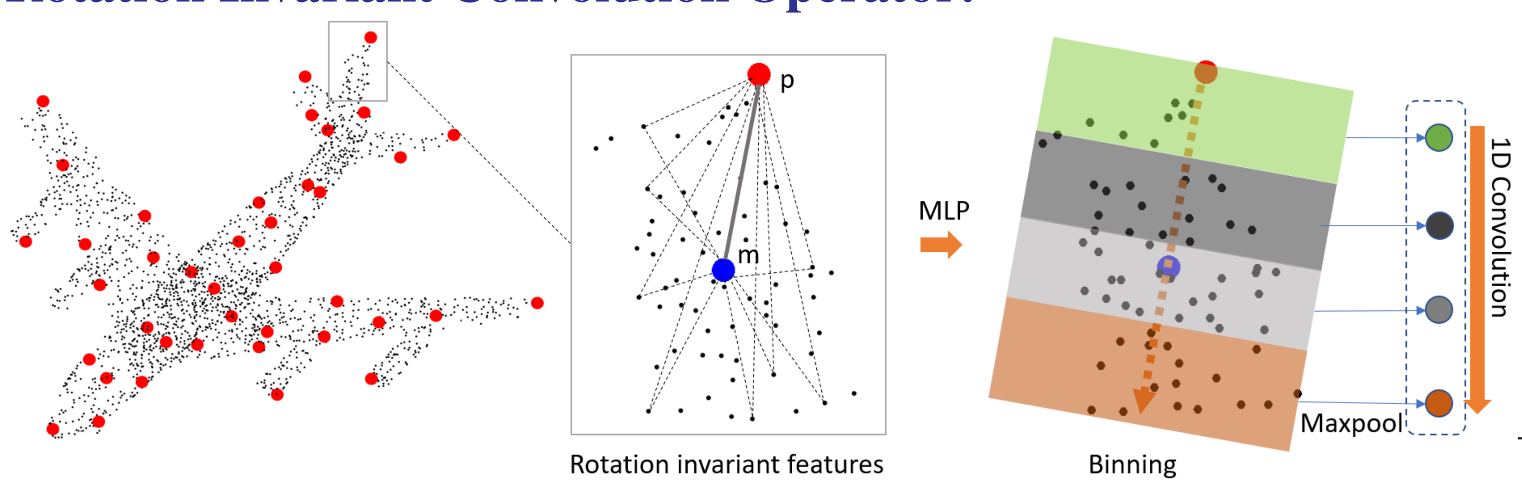


- d_0: distance between x and p
- d_1: distance between x and m
- α_0: angle between $\overrightarrow{xp}$ and $\overrightarrow{pm}$
- α_1: angle between $\overrightarrow{xm}$ and $\overrightarrow{pm}$
- $RIF(x; \overrightarrow{pm}) = [d_0, d_1, \alpha_0, \alpha_1]$

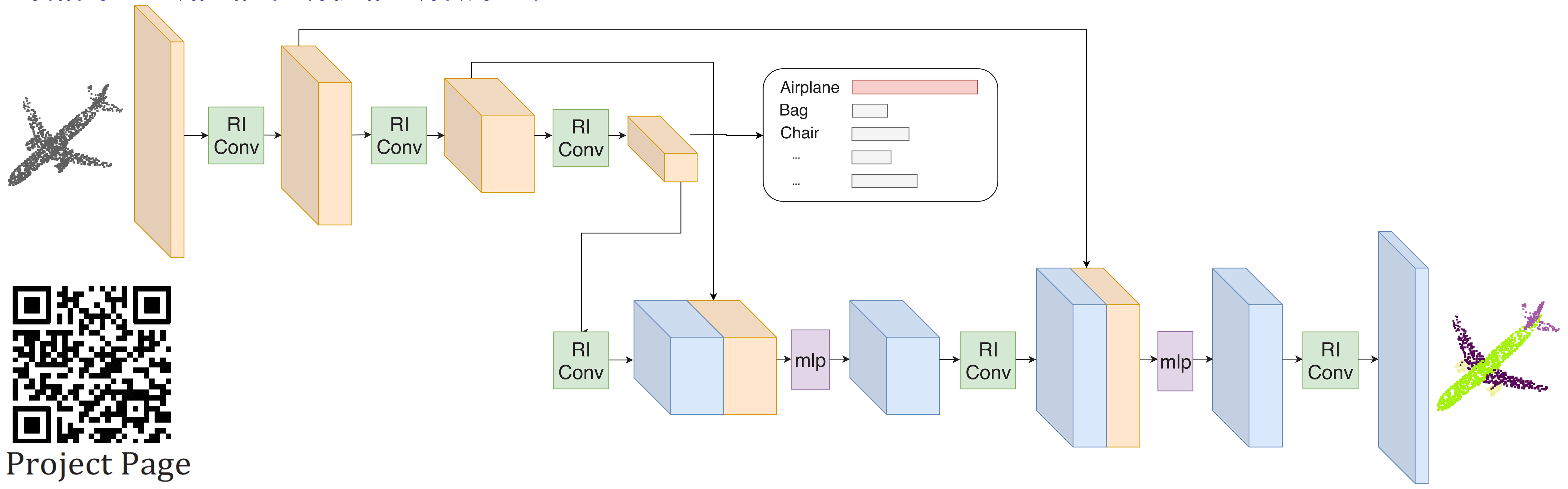
Algorithm 1 RIConv operator.

Input: Reference point p, point set P, features F_{prev};
Output: Convoluted features F;
$m \leftarrow avg(P)$; * Compute the centroid of P
$\overrightarrow{pm} \leftarrow m - p$; * Determine the reference orientation
$f_r \leftarrow \{RIF(x; \overrightarrow{pm}) : \forall x \in P\}$; * Find rotation invariant features
$F_r \leftarrow mlp(f_r)$; * Transform each feature f_r to high-dimensional feature F_r
$F_{in} \leftarrow [F_{prev}, F_r]$; * Concatenate the local and previous layer features
$\{S\} \leftarrow P$; * Divide local space into s bins along $\overrightarrow{pm}$
$\{F_{pool}\} \leftarrow \{\text{maxpool}(\{F_{in}(x) : \forall x \in s\}) : \forall s \in S\}$ * Max pool features for each bin of $\{S\}$
$F \leftarrow \text{conv}(\{F_{pool}\})$; * 1D convolution of the bin features
return F;

Rotation Invariant Convolution Operator:



Rotation Invariant Neural Network:



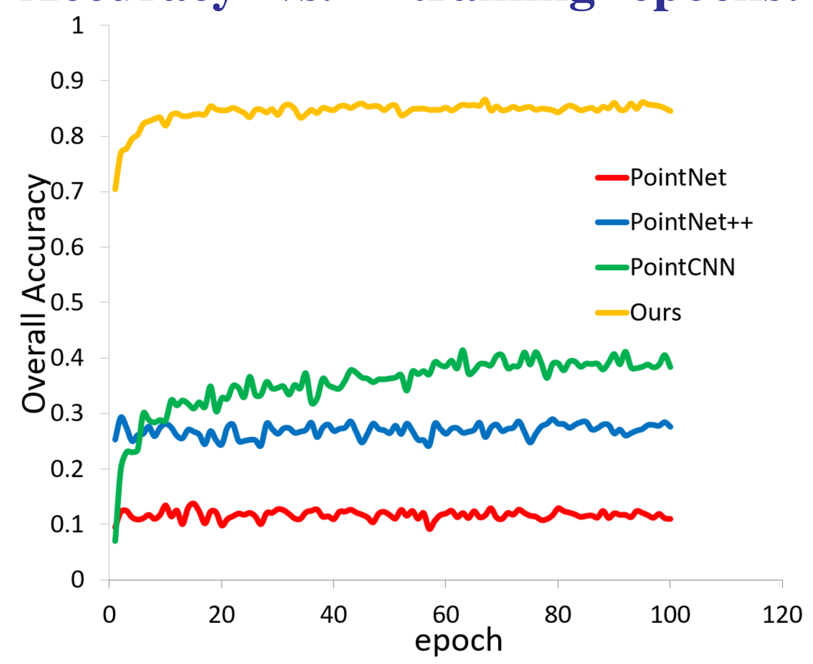
Project Page

Experiments & Results

Classification results on ModelNet40 dataset:

Method	Input	Input size	Parameters	z/z	SO3/SO3	z/SO3	Acc. std.
VoxNet	voxel	30^3	0.9M	83.0	**87.3**	-	3.0
SubVolSup	voxel	30^3	17M	88.5	82.7	36.6	28.4
SubVolSup MO	voxel	30^3	17M	89.5	85.0	45.5	24.2
Spherical CNN	voxel	2×64^2	0.5M	88.9	86.9	78.6	5.5
MVCNN 12x	view	12×224^2	99M	89.5	77.6	70.1	9.8
MVCNN 80x	view	80×224^2	99M	90.2	86.0	81.5	4.3
PointNet	xyz	1024×3	3.5M	87.0	80.3	12.8	41.0
PointNet++	xyz	1024×3	1.4M	89.3	85.0	28.6	33.8
PointCNN	xyz	1024×3	0.60M	**91.3**	84.5	41.2	27.2
Ours	xyz	1024×3	0.70M	86.5	86.4	**86.4**	**0.1**

Accuracy vs. training epochs:

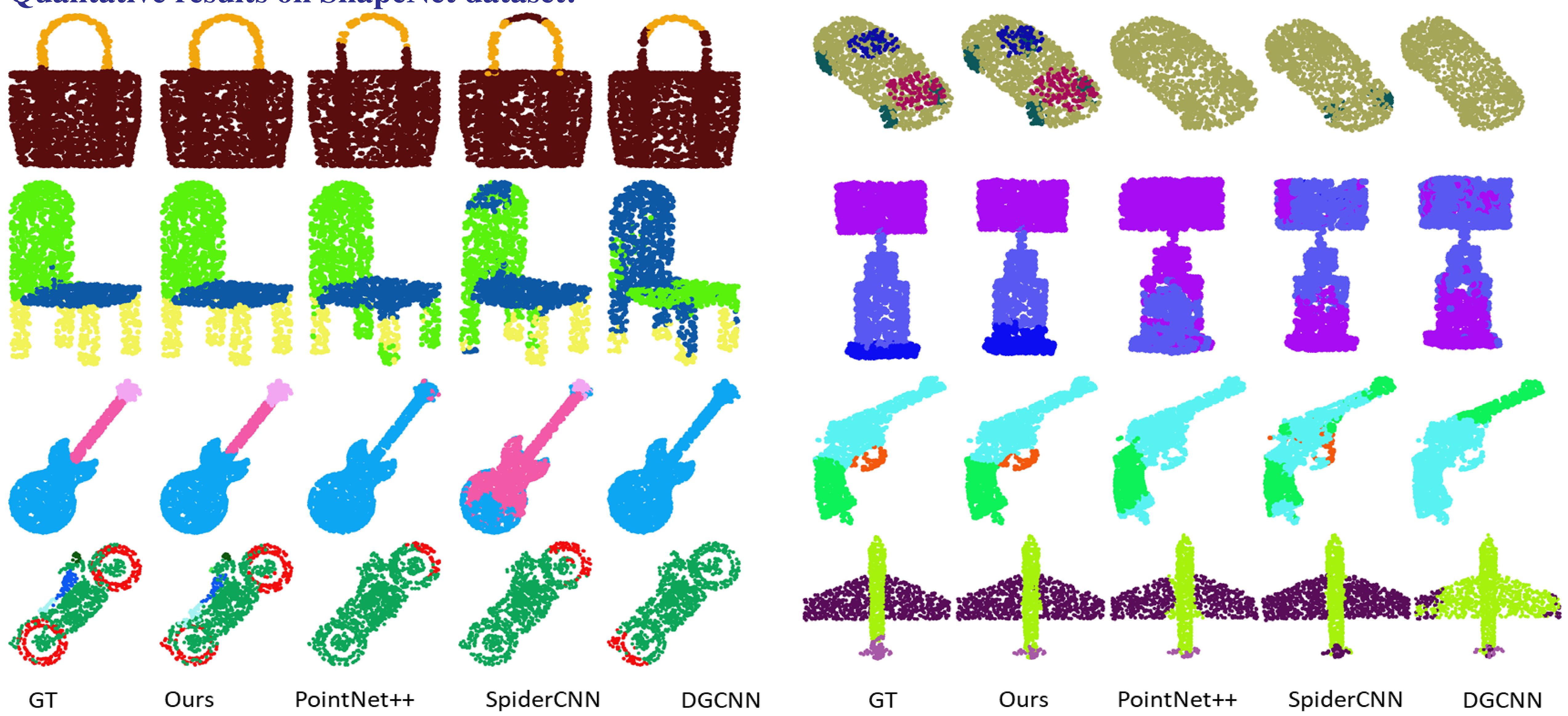


Part Segmentation results on ShapeNet dataset:

SO3/SO3	mIoU	aero	bag	cap	car	chair	earph.	guitar	knife	lamp	laptop	motor	mug	pistol	rocket	skate	table
PointNet	74.4	**81.6**	68.7	74.0	70.3	87.6	68.5	**88.9**	80.0	74.9	83.6	56.5	77.6	75.2	**53.9**	**69.4**	79.9
PointNet++	**76.7**	79.5	71.6	**87.7**	**70.7**	**88.8**	64.9	88.8	78.1	79.2	**94.9**	54.3	**92.0**	76.4	50.3	68.4	81.0
PointCNN	71.4	78.0	80.1	78.2	68.2	81.2	70.2	82.0	70.6	68.9	80.8	48.6	77.3	63.2	50.6	63.2	**82.0**
DGCNN	73.3	77.7	71.8	77.7	55.2	87.3	68.7	88.7	**85.5**	**81.8**	81.3	36.2	86.0	77.3	51.6	65.3	80.2
SpiderCNN	72.3	74.3	72.4	72.6	58.4	82.0	68.5	87.8	81.3	71.3	94.5	45.7	88.1	**83.4**	50.5	60.8	78.3
Ours	75.5	80.6	**80.2**	70.7	68.8	86.8	**70.4**	87.2	84.3	78.0	80.1	**57.3**	91.2	71.3	52.1	66.6	78.5

z/SO3	mIoU	aero	bag	cap	car	chair	earph.	guitar	knife	lamp	laptop	motor	mug	pistol	rocket	skate	table
PointNet	37.8	40.4	48.1	46.3	24.5	45.1	39.4	29.2	42.6	52.7	36.7	21.2	55.0	29.7	26.6	32.1	35.8
PointNet++	48.2	51.3	66.0	50.8	25.2	66.7	27.7	29.7	65.6	59.7	70.1	17.2	67.3	49.9	23.4	43.8	57.6
PointCNN	34.7	21.8	52.0	52.1	23.6	29.4	18.2	40.7	36.9	51.1	33.1	18.9	48.0	23.0	27.7	38.6	39.9
DGCNN	37.4	37.0	50.2	38.5	24.1	43.9	32.3	23.7	48.6	54.8	28.7	17.8	74.4	25.2	24.1	43.1	32.3
SpiderCNN	42.9	48.8	47.9	41.0	25.1	59.8	23.0	28.5	49.5	45.0	**83.6**	20.9	55.1	41.7	36.5	39.2	41.2
Ours	**75.3**	**80.6**	**80.0**	**70.8**	**68.8**	**86.8**	**70.3**	**87.3**	**84.7**	**77.8**	80.6	**57.4**	**91.2**	**71.5**	**52.3**	**66.5**	**78.4**

Qualitative results on ShapeNet dataset:



Acknowledgement: The authors acknowledge support from the SUTD Digital Manufacturing and Design Centre (DManD) funded by the Singapore National Research Foundation. This project is also partially supported by Singapore MOE Academic Research Fund MOE2016-T2-2-154 and Singapore NRF under its Virtual Singapore Award No. NRF2015VSGAA3DCM001-014.