人工智能前沿学生论坛

第25期

SFFAI
三维视觉之点云识别

讲　者：饶永铭（清华大学）
刘永成（自动化所）
主持人：蒋正锴（自动化所）

论坛分享 - II

- 分享人：<u>**刘永成**</u>

刘永成，中科院自动化所，模式识别国家重点实验室15级在读博士，研究兴趣包括三维点云处理、图像分割、场景分类等。以第一作者在CVPR、ACM MM等国际会议上发表论文3篇，国际顶级期刊上发表论文1篇。以第一完成人获国际图像分割竞赛冠军1次，国内目标检测竞赛亚军、季军各一次。

- 分享题目：<u>Geometric Relation Learning in 3D Point Cloud Analysis</u>
- 报告简介：

三维点云来自距离度量空间，这意味着每一个点并非孤立存在，相邻的点形成一个有意义的几何形状。因此，对点间几何关系进行建模非常重要。本次分享将回顾近年来使用深度学习进行点间关系学习的经典论文，并介绍我们的CVPR 2019 Oral工作Relation-Shape Convolutional Neural Network for Point Cloud Analysis。我们提出了一种几何关系卷积方法，并用该卷积搭建了一个关系形状卷积神经网络RS-CNN 。RS-CNN在三个主流的点云分析任务上均取得了SOTA，同时也很鲁棒。并且，RS-CNN不仅可以从点云中学习隐含的3D形状，还能从点云的2D投影空间中推理3D形状。

- Spotlight：
  - 1. 回顾点间关系学习的经典论文
  - 2. CVPR 2019 Oral工作：RS-CNN

Geometric Relation Learning in 3D Point Cloud Analysis

Yongcheng Liu

2019.04

中国科学院自动化研究所
Institute of Automation, Chinese Academy of Sciences

Outline

1 Introduction

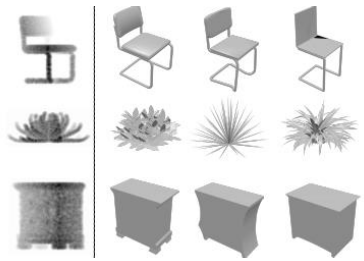
2 Related work

3 RS-CNN

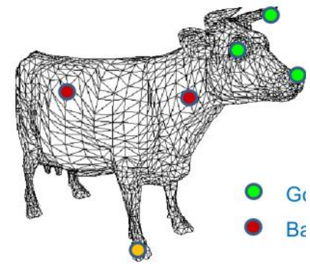
Introduction *Point cloud analysis*

 → lamp

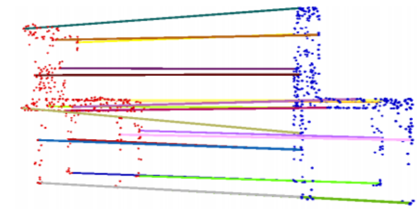
shape classification



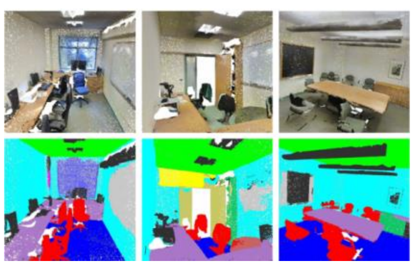
shape retrieval



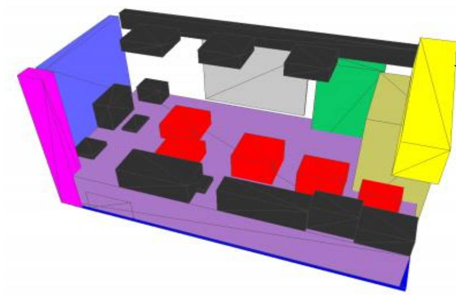
keypoint detection



shape correspondence



semantic segmentation



object detection

normal estimation

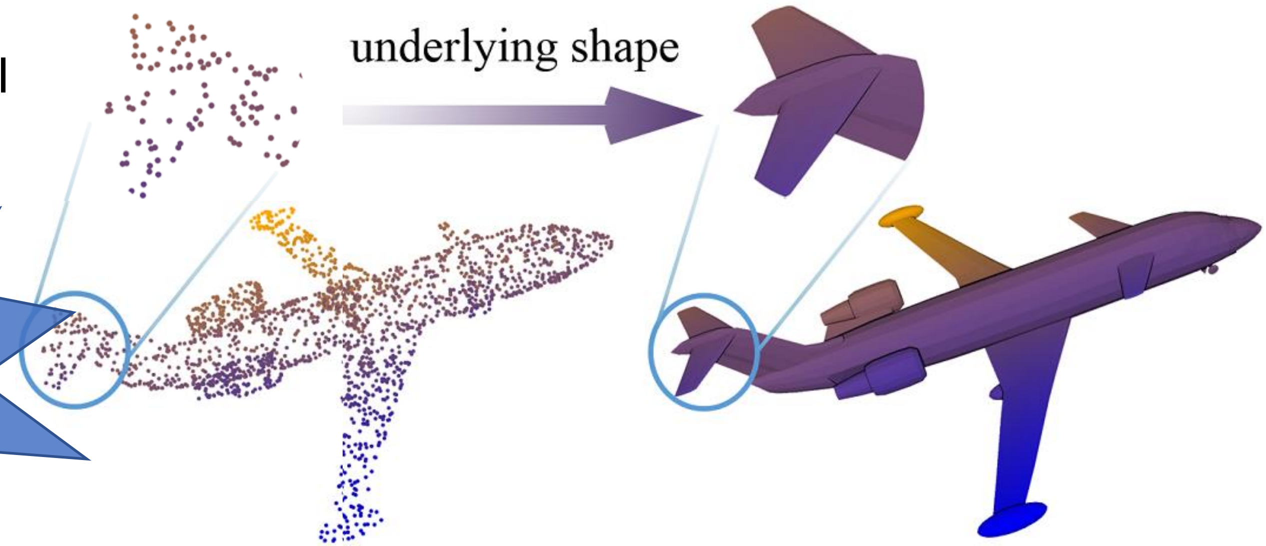
High-level shape understanding

……

Introduction *Shape understanding*

(x, y, z) + additional features

Deep Learning



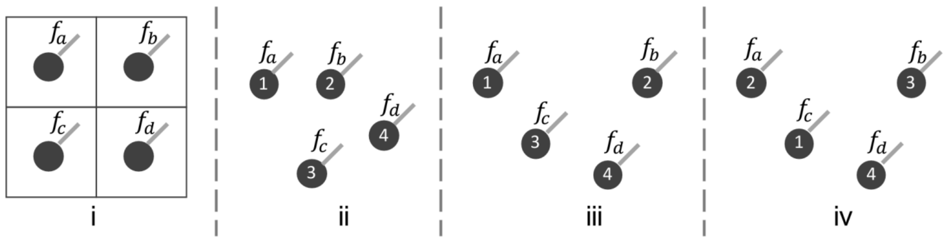
CNN, local-to-global manner for semantic learning

Points are not isolated → Analyzing each point independently maybe suboptimal

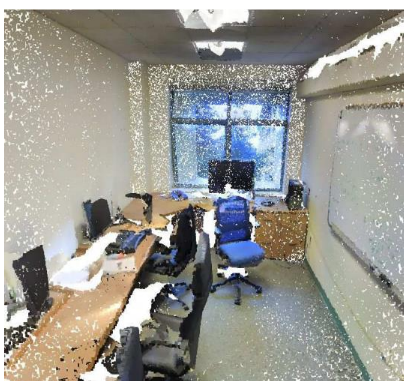
Nearby points form a meaningful shape → Abstract points’ relation (interaction)

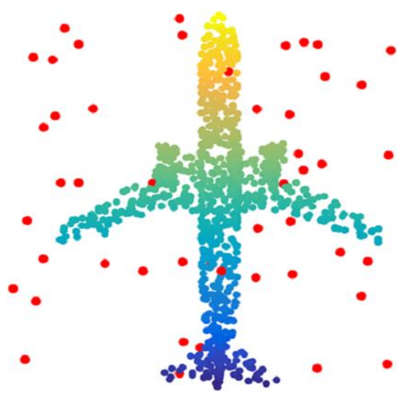
Introduction *Some challenges*

Irregular (unordered): permutation invariance

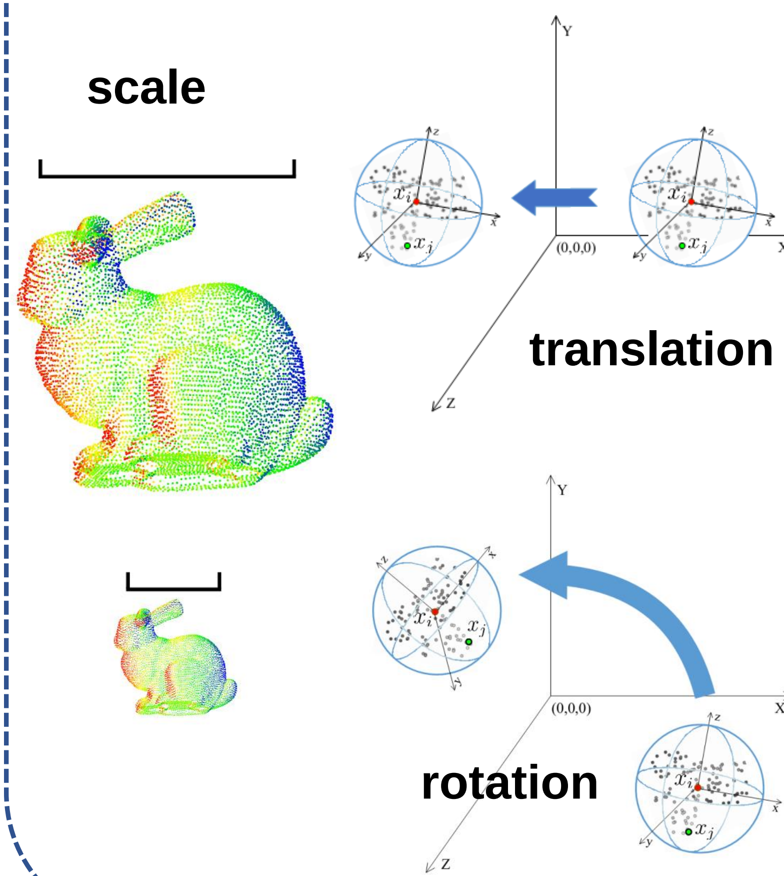


Robustness to corruption, outlier, noise

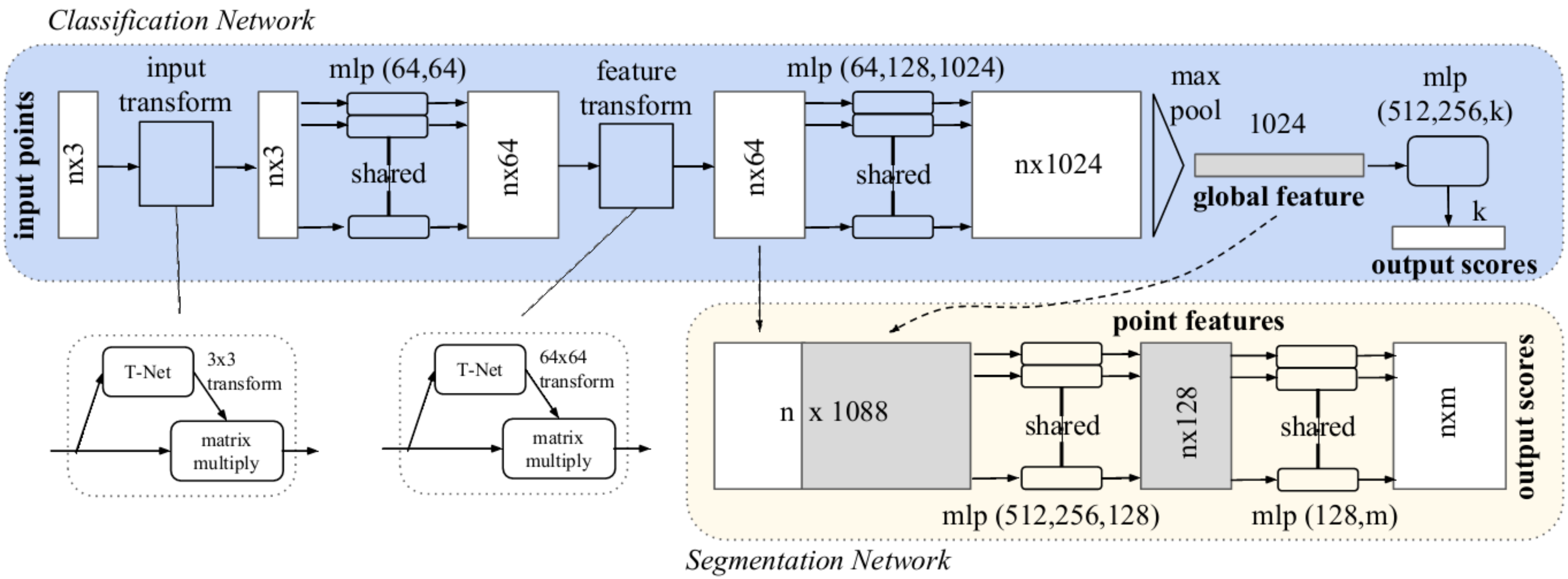




Robustness to rigid transformations



Related Work *PointNet: permutation invariance*

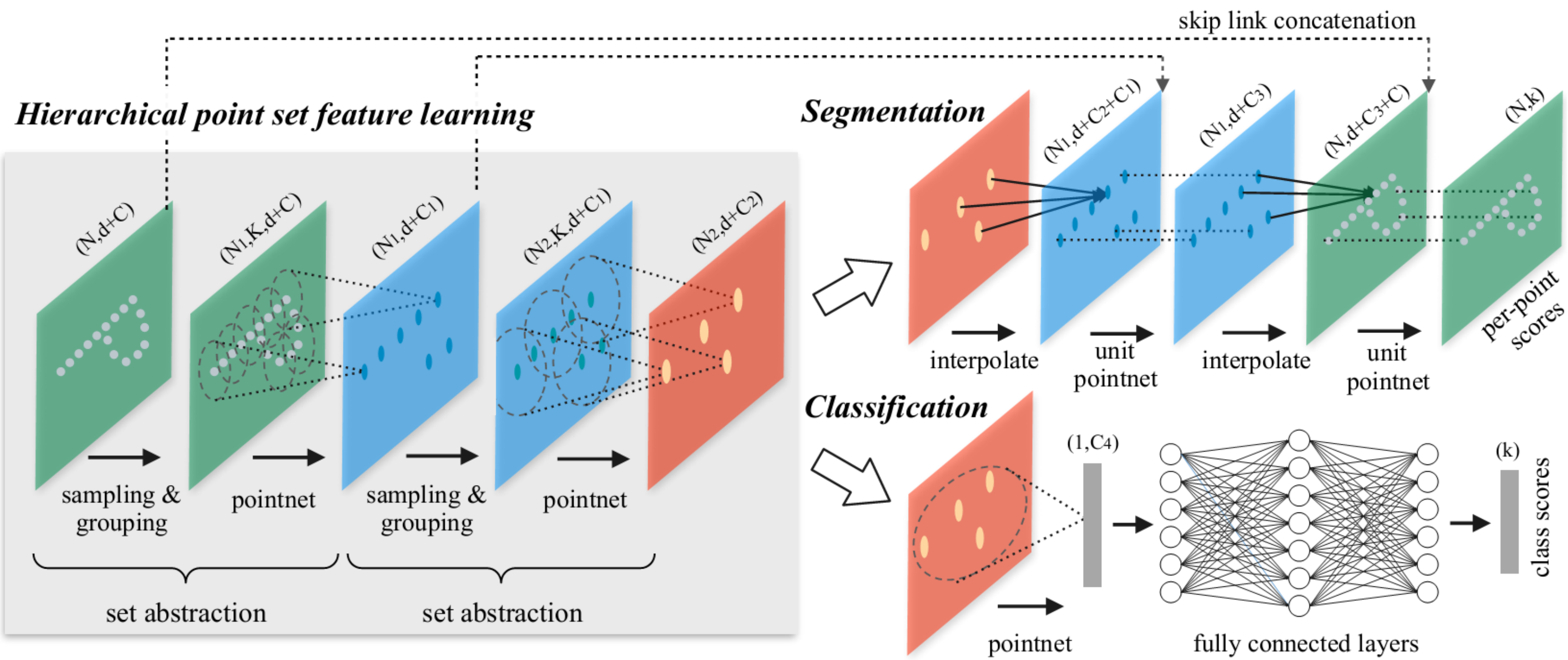


Shared MLP + max pool

No local patterns capturing

Qi et al. PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation. CVPR 2017.

Related Work PointNet++: local to global



Sampling + Grouping + PointNet

Implicit relation modeling – maybe suboptimal

Qi et al. PointNet++: Deep Hierarchical Feature Learning on Point Sets in a Metric Space. NIPS 2017.

Related Work *Relation modeling*

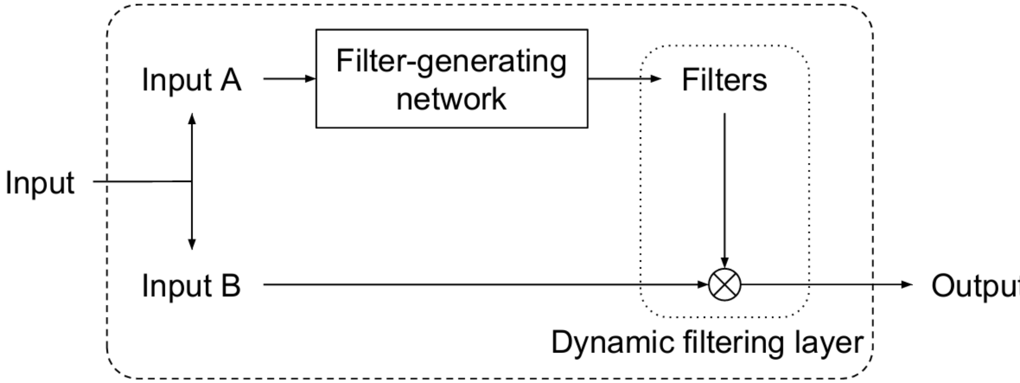
In a traditional convolutional layer, the learned filters stay fixed after training.

$$y = f(\sum \mathbf{W} * \mathbf{X} + b)$$

The filters in DFN are generated dynamically conditioned on an input.

$$y = f(\sum \boxed{g_{\boldsymbol{\theta}}(\mathbf{X}_{\mathrm{A}})} * \mathbf{X}_{\mathrm{B}} + b)$$

Dynamic Filter Networks (DFN)



Model parameter: $\boldsymbol{\theta}$

Filter: $g_{\boldsymbol{\theta}}(\mathbf{X}_{\mathrm{A}})$

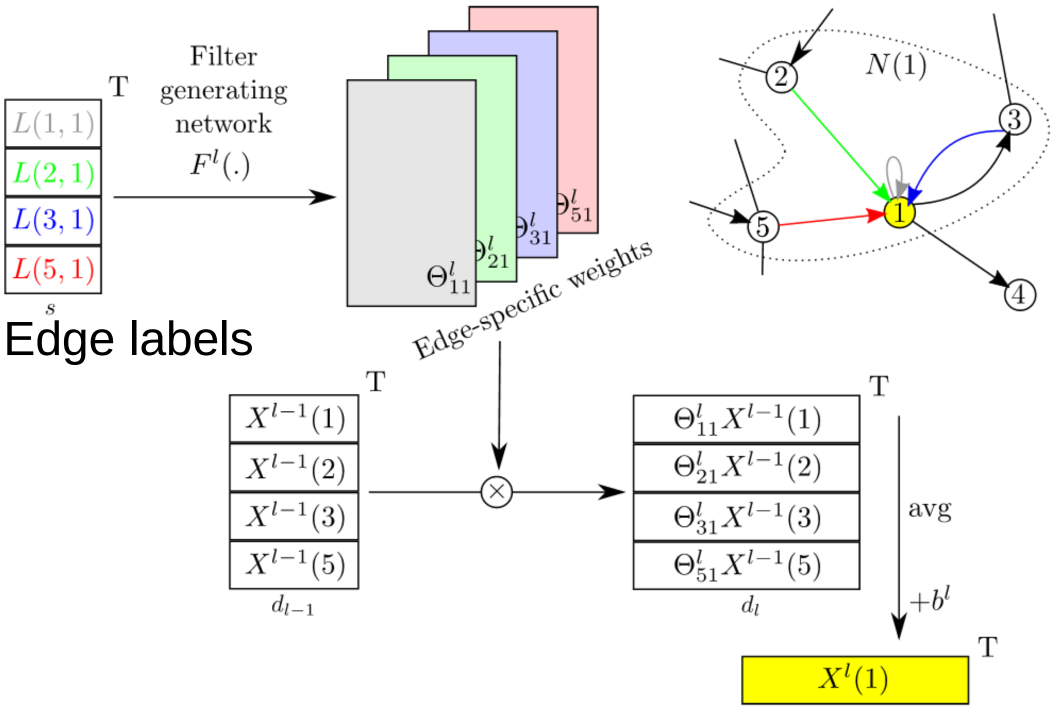
Brabandere et al. Dynamic Filter Networks. NIPS 2016.

Related Work *Relation modeling*

Dynamic Edge-Conditioned Filters in graph CNN

The current formulations of graph convolution do not exploit edge labels, which results in an overly homogeneous view of local graph neighborhoods.

The first to apply graph convolution to point cloud classification.



Edge labels

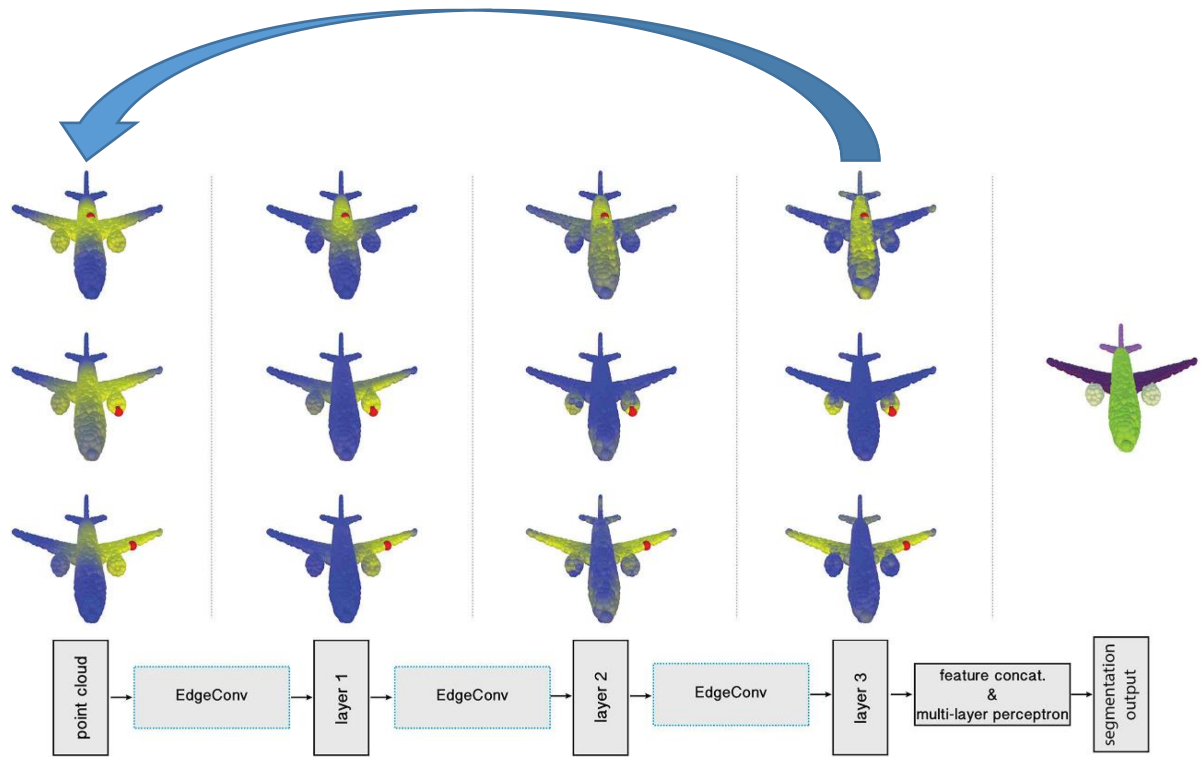
Simonovsky et al. Dynamic Edge-Conditioned Filters in Convolutional Neural Networks on Graphs. CVPR 2017.

Related Work *Relation modeling*

Dynamic Graph CNN (DGCNN)

Points in high-level feature space captures semantically similar structures.

Despite a large distance between them in the original 3D space.



Wang et al. Dynamic Graph CNN for Learning on Point Clouds. 2018.

Related Work *Relation modeling*

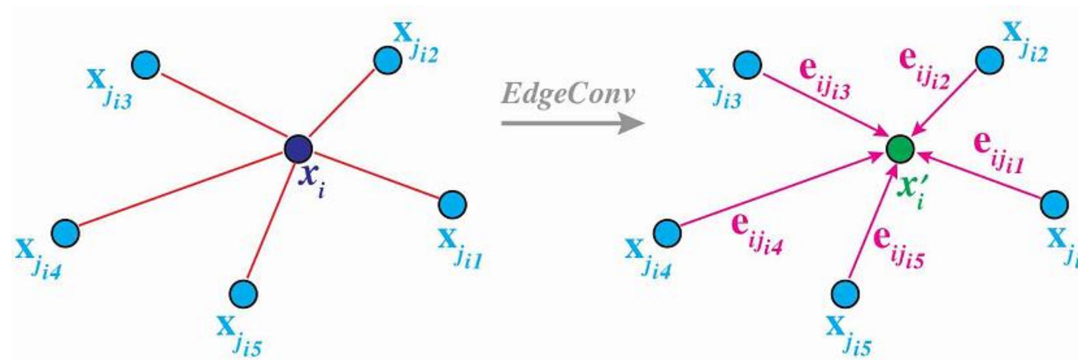
global info. local info.

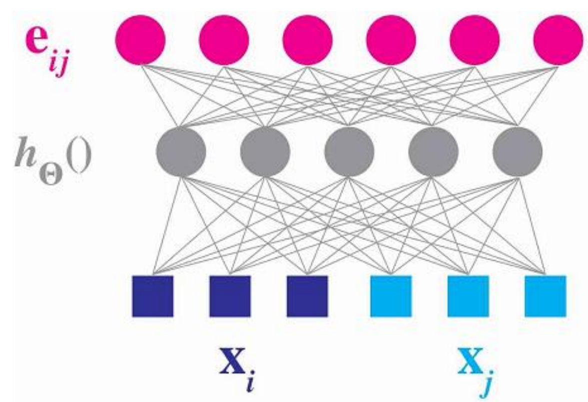
$$h_{\Theta}(\boxed{x_i}, \boxed{x_j - x_i})$$

$$x_i' = \overset{\text{max}}{\underset{j:(i,j)\in\mathcal{E}}{\square}} h_{\Theta}(x_i, x_j).$$

- Neighbors are found in feature space
- Learn from semantically similar structures

DGCNN —— EdgeConv





Wang et al. Dynamic Graph CNN for Learning on Point Clouds. 2018.

Relation-Shape Convolutional Neural Network for Point Cloud Analysis (RS-CNN)

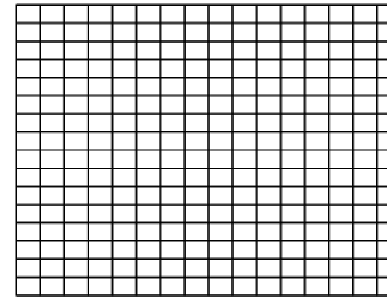
Yongcheng Liu, Bin Fan, Shiming Xiang, Chunhong Pan

CVPR 2019 Oral Presentation

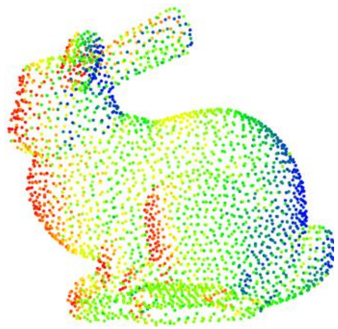
Project Page: https://yochengliu.github.io/Relation-Shape-CNN/

RS-CNN *Motivation*

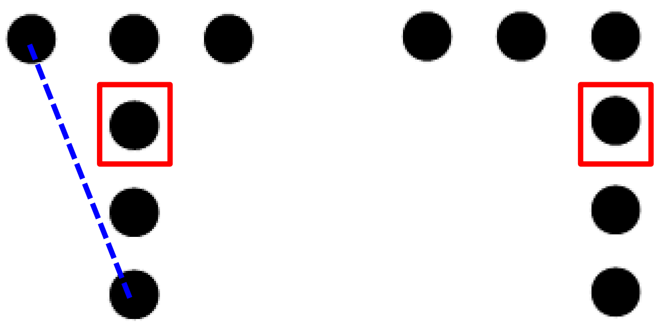
2D image



3D point cloud



3D **Shape** Learning



Relation Learning

Deep Learning (CNN)

RS-CNN *Method*

Relation-Shape Convolution (**RS-Conv**)

local point subset $P_{\text{sub}} \subset \mathbb{R}^3$ → spherical neighborhood: $x_i + x_j \in \mathcal{N}(x_i)$

$$\mathbf{f}_{P_{\text{sub}}} = \sigma(\mathcal{A}(\{\mathcal{T}(\mathbf{f}_{x_j}), \forall x_j\}))^1, \; d_{ij} < r \; \forall x_j \in \mathcal{N}(x_i) \qquad y = \sigma(\sum \mathbf{W} * \mathbf{X})$$

$\mathcal{T}$: feature transformation $\mathcal{A}$: feature aggregation

- Permutation invariance: only when *A* is symmetric and *T* is shared over each point
- Limitations of CNN: weight is not shared $\mathcal{T}(\mathbf{f}_{x_j}) = \mathbf{w}_j \cdot \mathbf{f}_{x_j}$
  gradient only w.r.t single point - implicit
- Conversion: learn from relation $\mathcal{T}(\mathbf{f}_{x_j}) = \mathbf{w}_{ij} \cdot \mathbf{f}_{x_j} = \mathcal{M}(\mathbf{h}_{ij}) \cdot \mathbf{f}_{x_j}$

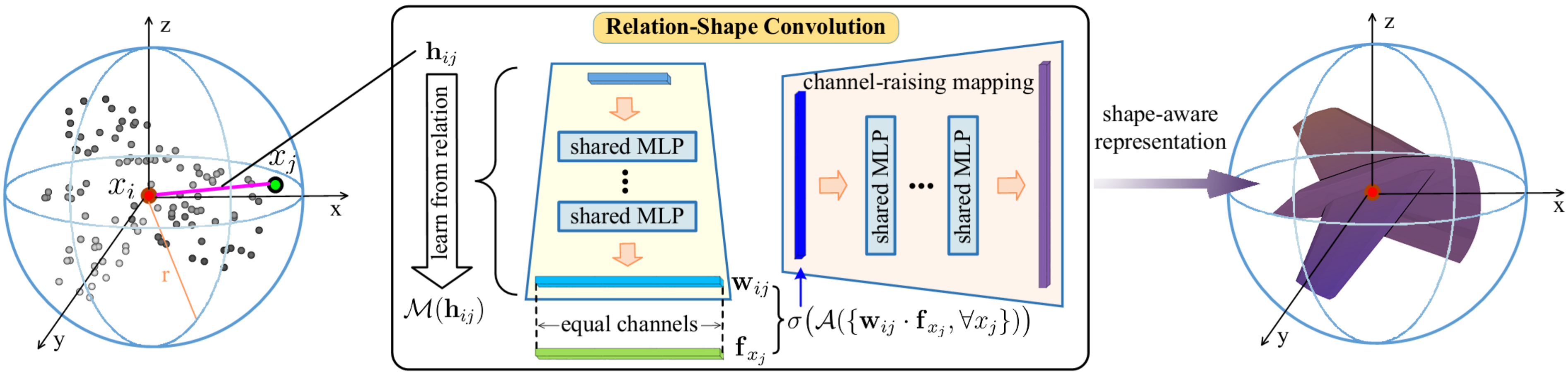
$\mathbf{h}_{ij}$: predefined geometric priors
→ low-level relation

$$\mathbf{f}_{P_{\text{sub}}} = \sigma(\mathcal{A}(\{\mathcal{M}(\mathbf{h}_{ij}) \cdot \mathbf{f}_{x_j}, \forall x_j\}))$$

$\mathcal{M}$: mapping function(shared MLP)
→ high-level relation

RS-CNN *Method*

$$\mathbf{f}_{P_{\text{sub}}} = \sigma\big(\mathcal{A}(\{\mathcal{M}(\mathbf{h}_{ij}) \cdot \mathbf{f}_{x_j}, \ \forall x_j\})\big)$$



high-level relation encoding + channel raising mapping

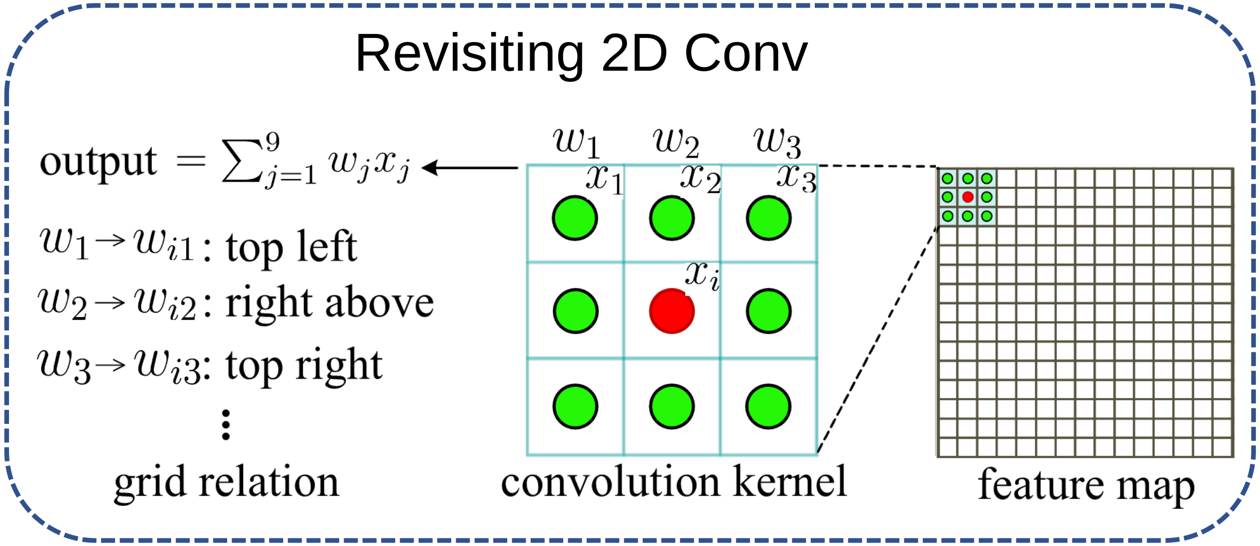
low-level relation $\mathbf{h}_{ij}$: (3D Euclidean distance, $x_i - x_j$, x_i, x_j)

10 channels

RS-CNN *RS-Conv: Properties*

$$\mathbf{f}_{P_{\text{sub}}} = \sigma\big(\mathcal{A}(\{\mathcal{M}(\mathbf{h}_{ij}) \cdot \mathbf{f}_{x_j}, \ \forall x_j\})\big)$$

- ✓ Permutation invariance
- ✓ Robustness to rigid transformation in Relation Learning, e.g., 3D Euclidean distance
- ✓ Points' interaction
- ✓ Weight sharing

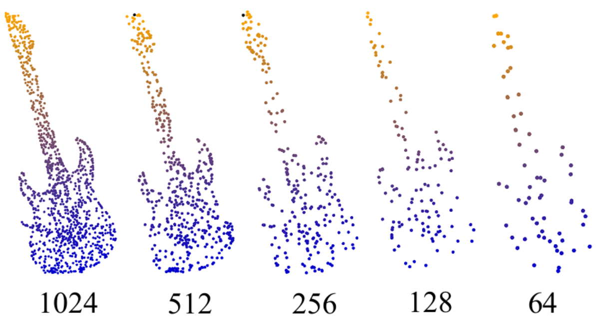


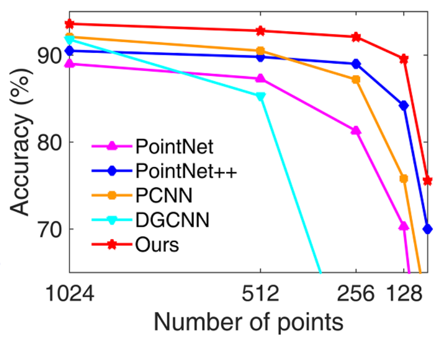
RS-Conv with relation learning is more general and can be applied to model 2D grid spatial relationship.

RS-CNN *Shape classification*

ModelNet40 benchmark

Robustness to sampling density





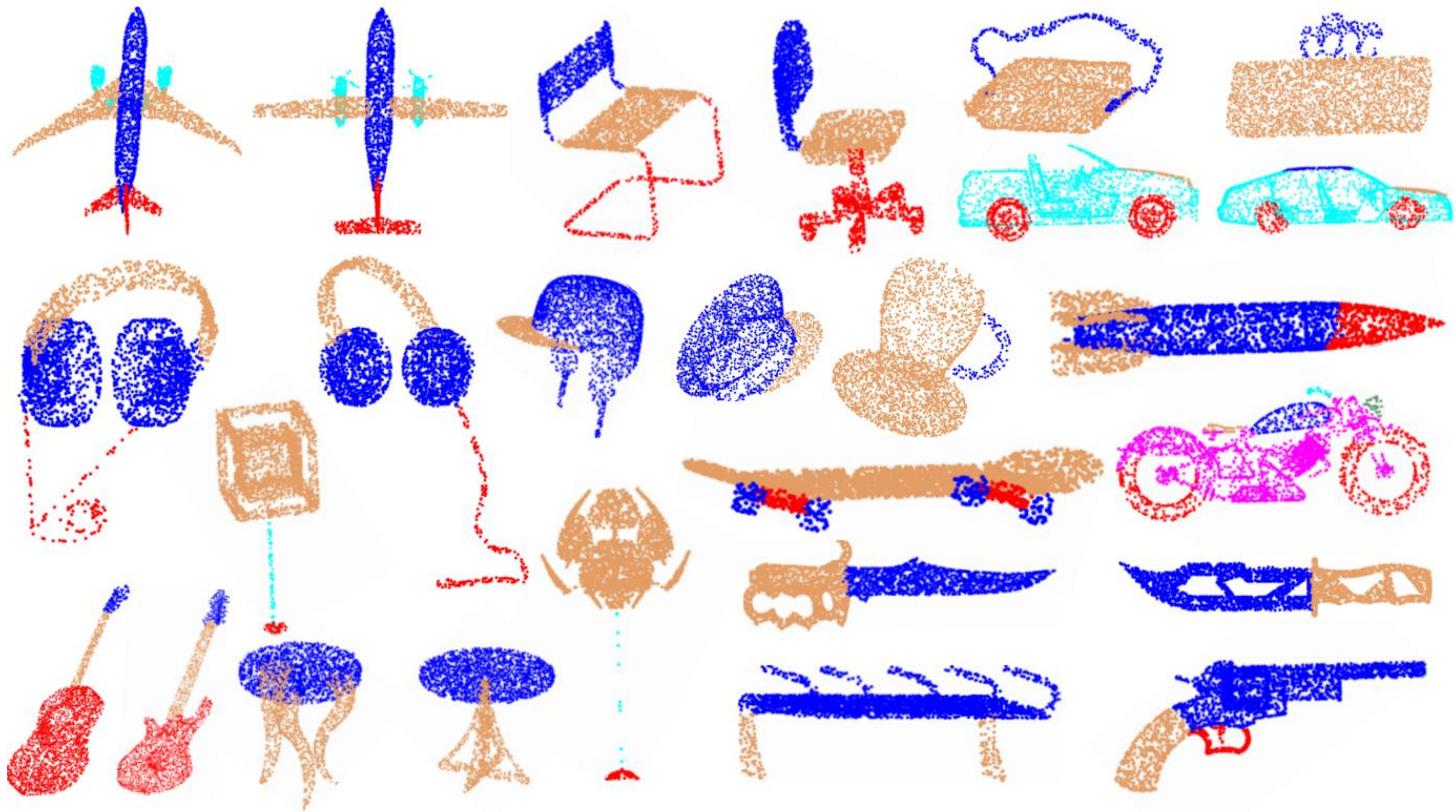
method	input	#points	acc.
Pointwise-CNN [10]	xyz	1k	86.1
Deep Sets [48]	xyz	1k	87.1
ECC [31]	xyz	1k	87.4
PointNet [24]	xyz	1k	89.2
SCN [44]	xyz	1k	90.0
Kd-Net(depth=10) [16]	xyz	1k	90.6
PointNet++ [26]	xyz	1k	90.7
KCNet [30]	xyz	1k	91.0
MRTNet [3]	xyz	1k	91.2
Spec-GCN [38]	xyz	1k	91.5
PointCNN [21]	xyz	1k	91.7
DGCNN [41]	xyz	1k	92.2
PCNN [1]	xyz	1k	92.3
Ours	**xyz**	**1k**	**93.6**
SO-Net [19]	xyz	2k	90.9
Kd-Net(depth=15) [16]	xyz	32k	91.8
O-CNN [39]	xyz, nor	-	90.6
Spec-GCN [38]	xyz, nor	1k	91.8
PointNet++ [26]	xyz, nor	5k	91.9
SpiderCNN [45]	xyz, nor	5k	92.4
SO-Net [19]	xyz, nor	5k	93.4

RS-CNN *ShapePart Segmentation*

method	input	class mIoU	instance mIoU	air plane	bag	cap	car	chair	ear phone	guitar	knife	lamp	laptop	motor bike	mug	pistol	rocket	skate board	table
Kd-Net [16]	4k	77.4	82.3	80.1	74.6	74.3	70.3	88.6	73.5	90.2	87.2	81.0	94.9	57.4	86.7	78.1	51.8	69.9	80.3
PointNet [24]	2k	80.4	83.7	83.4	78.7	82.5	74.9	89.6	73.0	91.5	85.9	80.8	95.3	65.2	93.0	81.2	57.9	72.8	80.6
RS-Net [11]	-	81.4	84.9	82.7	**86.4**	84.1	78.2	90.4	69.3	91.4	87.0	83.5	95.4	66.0	92.6	81.8	56.1	75.8	82.2
SCN [44]	1k	81.8	84.6	83.8	80.8	83.5	79.3	90.5	69.8	**91.7**	86.5	82.9	96.0	69.2	93.8	82.5	**62.9**	74.4	80.8
PCNN [1]	2k	81.8	85.1	82.4	80.1	85.5	79.5	90.8	73.2	91.3	86.0	85.0	95.7	73.2	94.8	83.3	51.0	75.0	81.8
SPLATNet [34]	-	82.0	84.6	81.9	83.9	88.6	79.5	90.1	73.5	91.3	84.7	84.5	**96.3**	69.7	**95.0**	81.7	59.2	70.4	81.3
KCNet [30]	2k	82.2	84.7	82.8	81.5	86.4	77.6	90.3	76.8	91.0	87.2	84.5	95.5	69.2	94.4	81.6	60.1	75.2	81.3
DGCNN [41]	2k	82.3	85.1	**84.2**	83.7	84.4	77.1	90.9	78.5	91.5	87.3	82.9	96.0	67.8	93.3	82.6	59.7	75.5	82.0
Ours	**2k**	**84.0**	**86.2**	83.5	84.8	**88.8**	**79.6**	**91.2**	**81.1**	91.6	**88.4**	**86.0**	96.0	**73.7**	94.1	**83.4**	60.5	**77.7**	**83.6**
PointNet++ [26]	2k,nor	81.9	85.1	82.4	79.0	87.7	77.3	90.8	71.8	91.0	85.9	83.7	95.3	71.6	94.1	81.3	58.7	76.4	82.6
SyncCNN [47]	mesh	82.0	84.7	81.6	81.7	81.9	75.2	90.2	74.9	93.0	86.1	84.7	95.6	66.7	92.7	81.6	60.6	82.9	82.1
SO-Net [19]	1k,nor	80.8	84.6	81.9	83.5	84.8	78.1	90.8	72.2	90.1	83.6	82.3	95.2	69.3	94.2	80.0	51.6	72.1	82.6
SpiderCNN [45]	2k,nor	82.4	85.3	83.5	81.0	87.2	77.5	90.7	76.8	91.1	87.3	83.3	95.8	70.2	93.5	82.7	59.7	75.8	82.8

class mIoU 1.7↑ instance mIoU 1.1↑

Best results over 10 categories



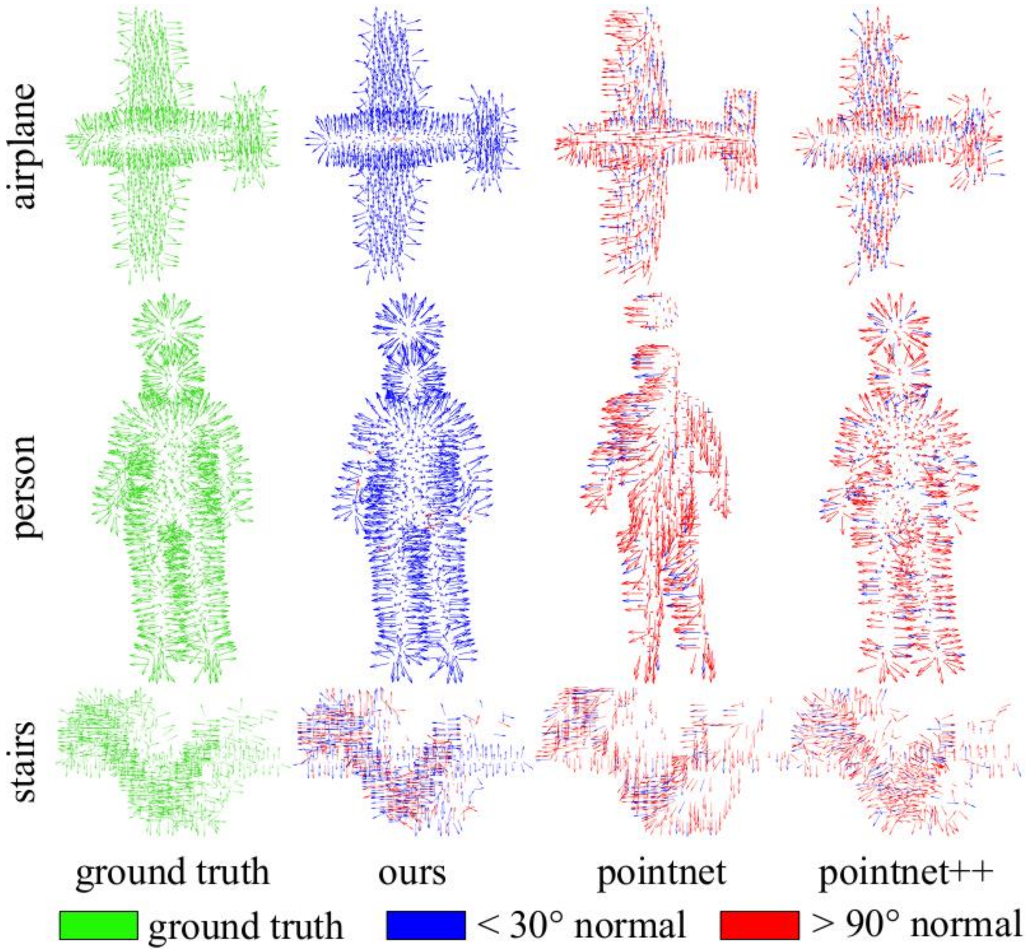
Diverse, confusing shapes

RS-CNN *Normal estimation*

Table 3. Normal estimation error on ModelNet40 dataset.

dataset	method	#points	error
ModelNet40	PointNet [1]	1k	0.47
	PointNet++ [1]	1k	0.29
	PCNN [1]	1k	0.19
	Ours	**1k**	**0.15**

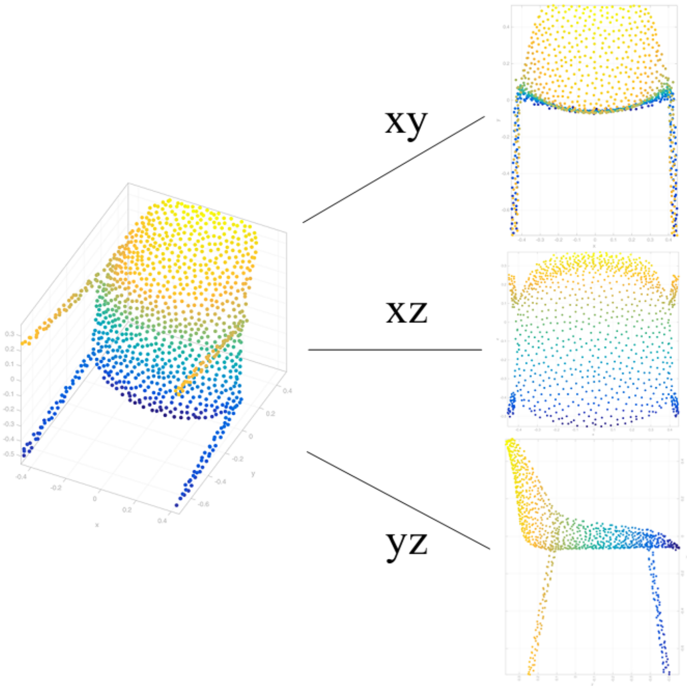
less effective for some intractable shapes, such as spiral stairs and intricate plants



RS-CNN *Geometric priors*

$$\mathbf{f}_{P_{\text{sub}}} = \sigma\big(\mathcal{A}(\{\mathcal{M}(\mathbf{h}_{ij}) \cdot \mathbf{f}_{x_j},\ \forall x_j\})\big)$$

model	low-level relation $\mathbf{h}$	channels	acc.
A	(3D-Ed)	1	92.5
B	(3D-Ed, $x_i - x_j$)	4	93.0
C	(3D-Ed, $x_i - x_j$, x_i, x_j)	10	**93.6**
D	(3D-cosd, x_i^{nor}, x_j^{nor})	7	92.8
E	(2D-Ed, $x_i' - x_j'$, x_i', x_j')	10	$\approx$ 92.2



low-level relation $\mathbf{h}$	channels	acc.
(XY-Ed, $x_i^{\text{xy}} - x_j^{\text{xy}}$, x_i^{xy}, x_j^{xy})	10	92.1
(XY-Ed, $x_i^{\text{xz}} - x_j^{\text{xz}}$, x_i^{xz}, x_j^{xz})	10	92.1
(XY-Ed, $x_i^{\text{yz}} - x_j^{\text{yz}}$, x_i^{yz}, x_j^{yz})	10	92.2
fusion of above three views		92.5

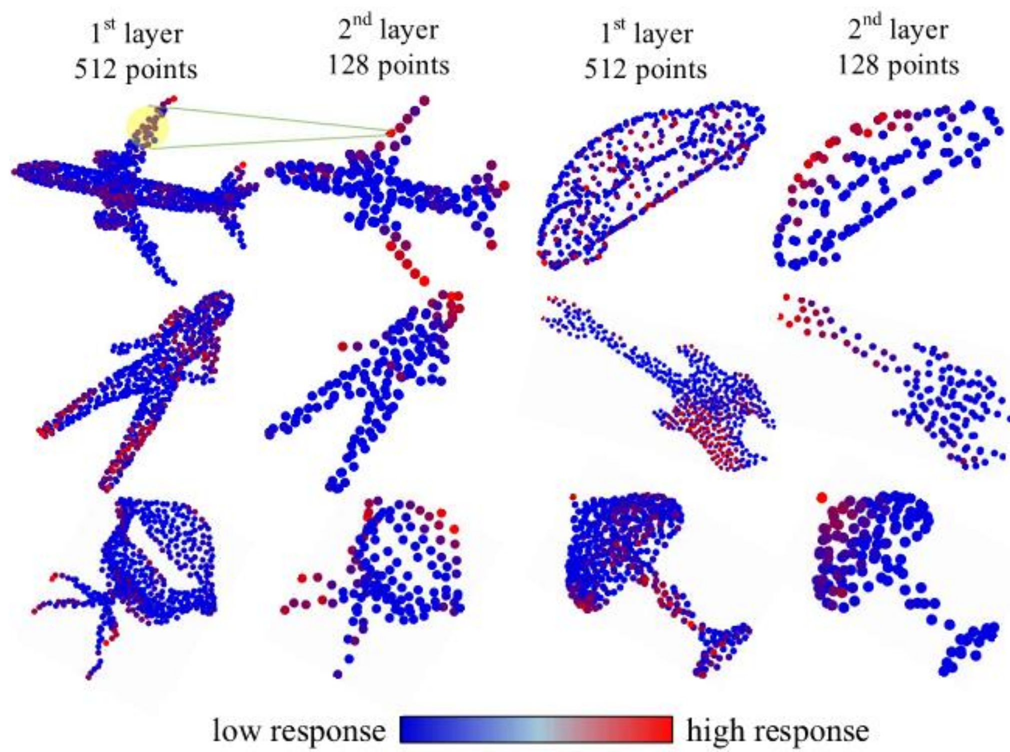
RS-CNN *Model analysis*

Robustness to point permutation and rigid transformation

relation: 3D Euclidean distance

method	acc.	perm.	+0.2	-0.2	90°	180°
PointNet [24]	88.7	88.7	70.8	70.6	42.5	38.6
PointNet++ [26]	88.2†	88.2	88.2	88.2	47.9	39.7
Ours	**90.3†**	**90.3**	**90.3**	**90.3**	**90.3**	**90.3**

$$\mathbf{f}_{P_{\text{sub}}} = \sigma\big(\mathcal{A}(\{\mathcal{M}(\mathbf{h}_{ij}) \cdot \mathbf{f}_{x_j},\ \forall x_j\})\big)$$



Model complexity

method	#params	#FLOPs/sample
PointNet [24]	3.50M	440M
PointNet++ [21]	1.48M	1684M
PCNN [21]	8.20M	**294M**
Ours	**1.41M**	295M

Relation-Shape Convolutional Neural Network for Point Cloud Analysis

We did a little exploration in **Geometric Relation Learning** for point cloud analysis.

Thanks

Speaker : Yongcheng Liu