Dynamic Graph CNN for Learning on Point Clouds

Credit: Yue Wang[1], Yongbin Sun[1], Ziwei Liu[2], Sanjay E. Sarma[1], Michael M. Bronstein[3], Justin M. Solomon[1]

[1]MIT

[2]UC Berkeley

[3]USI/TAU/Intel

Presenter: Fuwen Tan

https://qdata.github.io/deep2Read

Point Cloud Representation of 3D Shape

$$\mathbf{X} = \{\mathbf{x}_1, \cdots, \mathbf{x}_n\} \subseteq \mathbb{R}^F \tag{1}$$

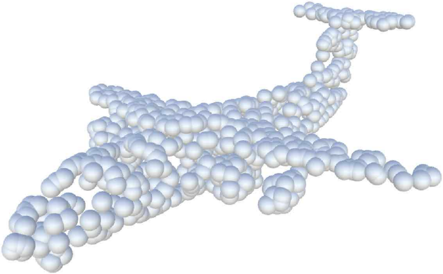


Figure: Point Cloud representation of a *plane*. Each point vector may encode multiple attributes, e.g. 3D coordinate, surface normal, color, etc.

Tasks

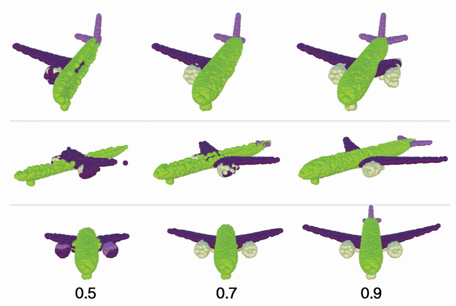
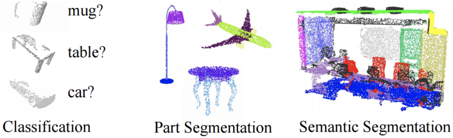


Figure: Class-specific part segmentation



EdgeConv

$$
\begin{aligned}
\mathbf{e_{ij}} &= h_\theta(\mathbf{x}_i, \mathbf{x}_j - \mathbf{x}_i) \\
&= \mathbf{W}_c([\mathbf{x}_i; \mathbf{x}_j - \mathbf{x}_i]) \\
\mathbf{x}_i^{out} &= \max_{j:(i,j)\in E}\{\mathbf{e_{ij}}\}
\end{aligned}
$$

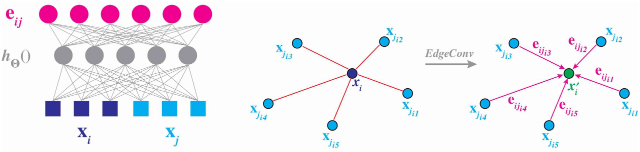


Figure: Edge Convolution: a symmetry function for the two vertices.

How to define E (the edge set)?

- k-nn in the **feature** space ($\mathbf{x}_i \in \mathbb{R}^F$)
- the main distinction from previous works
- each layer has a different graph, which will change after each training iteration

Dynamic Graph CNNs

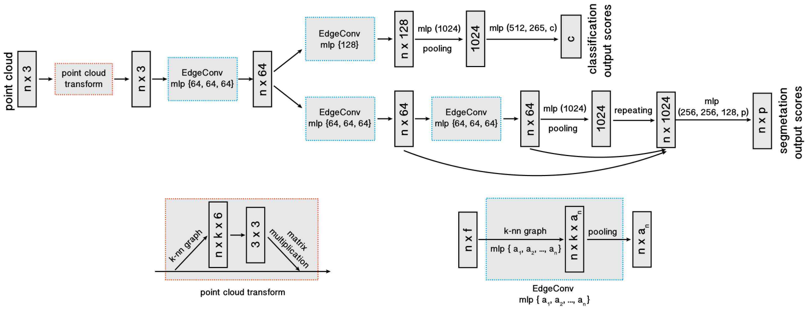
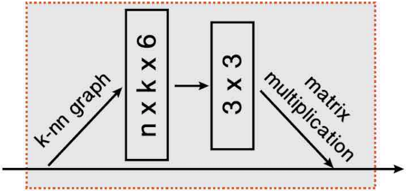


Figure: Overview

Point Cloud Transformation

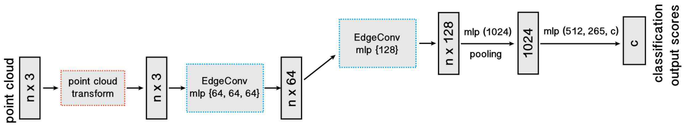
- Proposed in PointNet [6]
- Align the local neighborhood of a point to a canonical space by applying an estimated 3x3 matrix
- Similar with the spatial transformer network in 2D



point cloud transform

Shape recognition: implementation

- K=20
- Each EdgeConv block has a shared edge function h_θ
- Short-cut connections for multi-scale feature aggregations (not clear)
- ReLU+BatchNorm after each layer
- 0.5 Dropout rate for the last two fc layers
- A variant version (Baseline): no point cloud transformer and using fixed graph



Shape recognition: experiment

- Dataset: ModelNet40 [12]
  - 9843/2468 CAD shapes
  - 40 categories
  - 1024 points sampled for each shape and normalized to the unit sphere

Shape recognition: results

	Mean Class Accuracy	Overall Accuracy
3DShapeNets [12]	77.3	84.7
VoxNet [5]	83.0	85.9
Subvolume [7]	86.0	89.2
ECC [10]	83.2	87.4
PointNet [6]	86.0	89.2
PointNet++ [8]	-	90.7
Kd-Net (depth 10) [4]	-	90.6
Kd-Net (depth 15) [4]	-	91.8
Ours (baseline)	88.8	91.2
Ours	**90.2**	**92.2**

Table: Classification results on ModelNet40.

Shape recognition: model complexity

	Model size(MB)	Forward time(ms)	Accuracy(%)
PointNet (Baseline)	9.4	11.6	87.1
PointNet	40	25.3	89.2
PointNet++	12	163.2	90.7
Ours (Baseline)	11	29.7	91.2
Ours	21	94.6	92.2

Table: Complexity, forward time and accuracy of different models

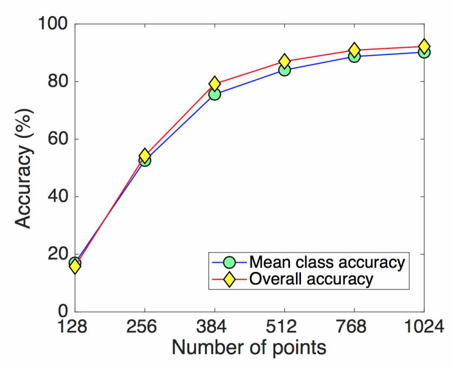
Shape recognition: ablation study

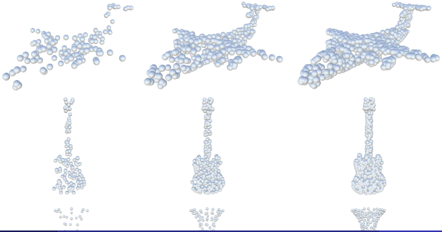
- Centralization: $h_\theta(\mathbf{x}_i, \mathbf{x}_j - \mathbf{x}_i)$ vs $h_\theta(\mathbf{x}_i, \mathbf{x}_j)$

CENT	DYN	XFORM	MEAN CLASS ACCURACY(%)	OVERALL ACCURACY(%)
x			88.8	91.2
x	x		88.8	91.5
x		x	89.6	91.9
	x	x	89.8	91.9
x	x	x	90.2	92.2

Table: Effectiveness of different components. CENT denotes centralization, DYN denotes dynamical graph recomputation, and XFORM denotes the use of a spatial transformer.

Shape recognition: ablation study





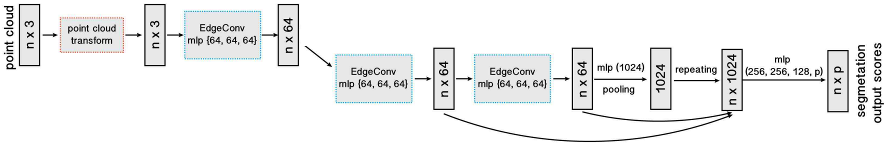
Shape recognition: ablation study

Number of nearest neighbors (k)	Mean Class Accuracy(%)	Overall Accuracy(%)
5	88.0	90.5
10	88.8	91.4
20	90.2	92.2
40	89.2	91.7

Table: Results of our model with different numbers of nearest neighbors.

Part segmentation: implementation

- K=30
- similar with the shape recognition model



Part segmentation: experiment

- Dataset: ShapeNet part dataset [11]
  - 16881 3D shapes
  - splits defined in [2]
  - 16 categories
  - 50 parts in total
  - 2048 points sampled for each shape
  - evalution metric: IoU on points

Part segmentation: results

	MEAN	AREO	BAG .	CAP	CAR	CHAIR	EAR PHONE	GUITAR	KNIFE	LAMP	LAPTOP	MOTOR	MUG	PISTOL	ROCKET	SKATE BOARD	TABLE	**WINNING CATEGORIES**
# SHAPES		2690	76	55	898	3758	69	787	392	1547	451	202	184	283	66	152	5271	
POINTNET [6]	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	1
POINTNET++ [8]	**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	5
KD-NET [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	0
LOCALFEATURENET [9]	84.3	**86.1**	73.0	54.9	**77.4**	88.8	55.0	90.6	86.5	75.2	**96.1**	57.3	91.7	**83.1**	53.9	72.5	**83.8**	5
OURS	**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	**6**

Table: Part segmentation results on ShapeNet part dataset. Metric is mIoU(%) on points.

Part segmentation: ablation study

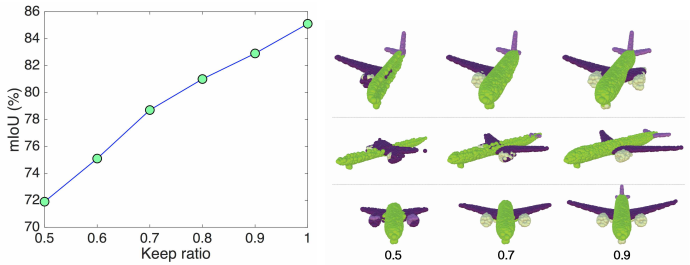


Figure: **Left:** The mean IoU (%) improves when the ratio of kept points increases. Points are dropped from one of six sides (top, bottom, left, right, front and back) randomly during evaluation process. **Right:** Part segmentation results on partial data. Points on each row are dropped from the same side. The keep ratio is shown below the bottom row. Note that the segmentation results of turbines are improved when more points are included.

Indoor scene segmentation: experiment

- Dataset: S3DIS [1]
  - 6 indoor areas
  - 272 rooms in total
  - 16 semantica categories
  - 9D feature vector: XYZ, normalized XYZ, color
  - 4096 points sampled for each shape during training, all points are used during testing
  - evalution metric: IoU on points

Indoor scene segmentation: results

	MEAN IOU	OVERALL ACCURACY
POINTNET (BASELINE) [6]	20.1	53.2
POINTNET [6]	47.6	78.5
MS + CU(2) [3]	47.8	79.2
G + RCU [3]	49.7	81.1
OURS	**56.1**	**84.1**

Table: 3D semantic segmentation results on S3DIS. MS+CU for multi-scale block features with consolidation units; G+RCU for the grid-blocks with recurrent consolidation Units.

Conclusion

- Simple, effective, maybe not very efficient
- The performance looked good at the submitted time (Jan. 2018)
- Not in good shape yet

Iro Armeni, Ozan Sener, Amir R. Zamir, Helen Jiang, Ioannis Brilakis, Martin Fischer, and Silvio Savarese.
3d semantic parsing of large-scale indoor spaces.
In *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition*, 2016.

Angel X. Chang, Thomas Funkhouser, Leonidas Guibas, Pat Hanrahan, Qixing Huang, Zimo Li, Silvio Savarese, Manolis Savva, Shuran Song, Hao Su, Jianxiong Xiao, Li Yi, and Fisher Yu.
ShapeNet: An Information-Rich 3D Model Repository.
Technical Report arXiv:1512.03012 [cs.GR], Stanford University — Princeton University — Toyota Technological Institute at Chicago, 2015.

Francis Engelmann, Theodora Kontogianni, Alexander Hermans, and Bastian Leibe.
Exploring spatial context for 3d semantic segmentation of point clouds.
In *IEEE International Conference on Computer Vision, 3DRMS Workshop, ICCV*, 2017.

Roman Klokov and Victor Lempitsky.

Escape from cells: Deep kd-networks for the recognition of 3d point cloud models.

In *The IEEE International Conference on Computer Vision (ICCV)*, Oct 2017.

D. Maturana and S. Scherer.

VoxNet: A 3D Convolutional Neural Network for Real-Time Object Recognition.

In *IROS*, 2015.

Charles R Qi, Hao Su, Kaichun Mo, and Leonidas J Guibas.

Pointnet: Deep learning on point sets for 3d classification and segmentation.

arXiv preprint arXiv:1612.00593, 2016.

Charles R Qi, Hao Su, Matthias Niessner, Angela Dai, Mengyuan Yan, and Leonidas J Guibas.

Volumetric and multi-view cnns for object classification on 3d data.

arXiv preprint arXiv:1604.03265, 2016.

Charles R Qi, Li Yi, Hao Su, and Leonidas J Guibas.

Pointnet++: Deep hierarchical feature learning on point sets in a metric space.
arXiv preprint arXiv:1706.02413, 2017.

Yiru Shen, Chen Feng, Yaoqing Yang, and Dong Tian.
Neighbors do help: Deeply exploiting local structures of point clouds.
CoRR, abs/1712.06760, 2017.

Martin Simonovsky and Nikos Komodakis.
Dynamic edge-conditioned filters in convolutional neural networks on graphs.
In *CVPR*, 2017.

Li Yi, Vladimir G. Kim, Duygu Ceylan, I-Chao Shen, Mengyan Yan, Hao Su, Cewu Lu, Qixing Huang, Alla Sheffer, and Leonidas Guibas.
A scalable active framework for region annotation in 3d shape collections.
ACM Transactions on Graphics (SIGGRAPH ASIA), 2016.

A. Khosla F. Yu L. Zhang X. Tang J. Xiao Z. Wu, S. Song.
3d shapenets: A deep representation for volumetric shapes.

In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015.