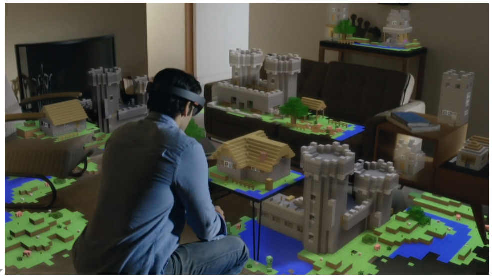
ELEC5307: Applications II

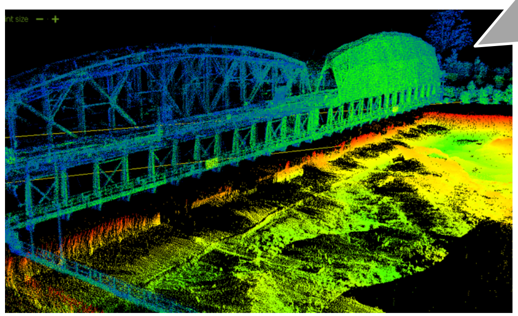
Many slices are taken from Charles Qi @Stanford

Background

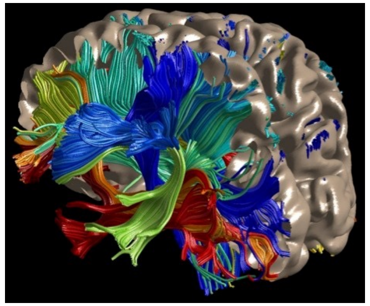
Robotics



Augmented Reality



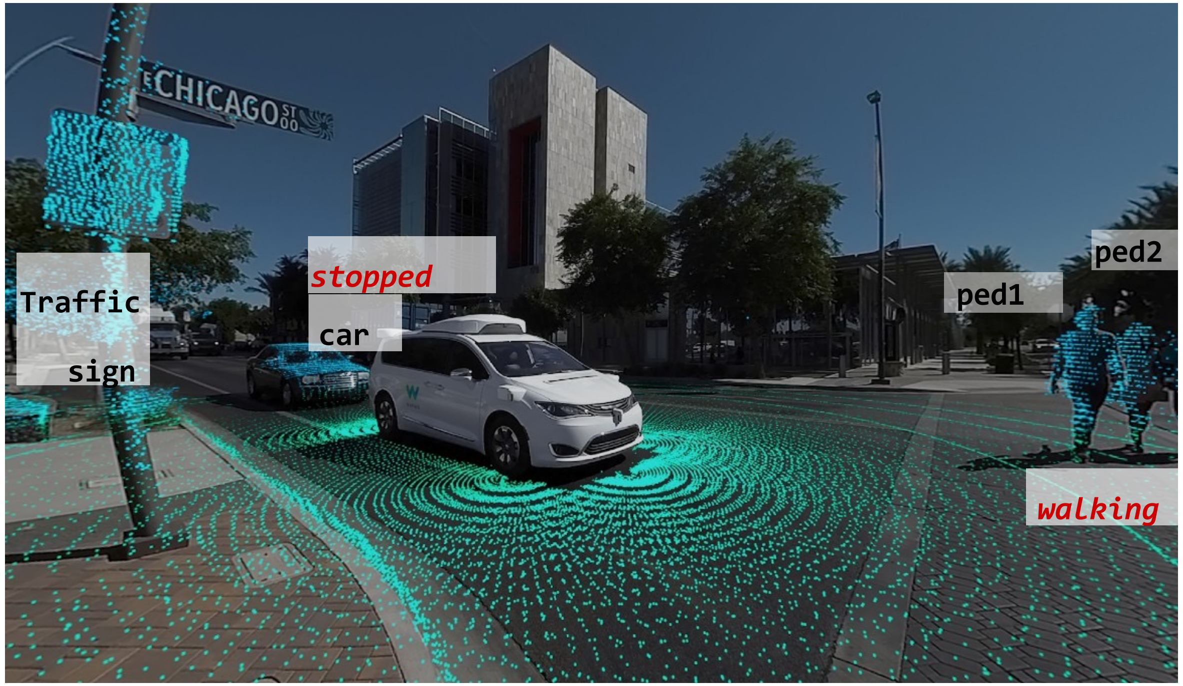
Autonomous driving



Medical Image Processing

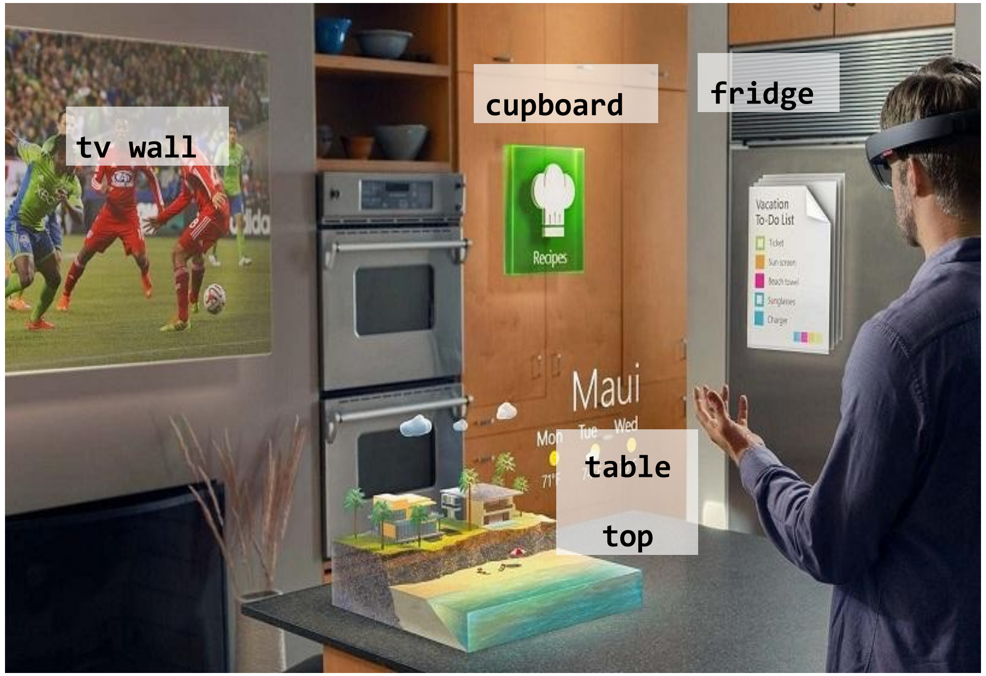
Background

Self-Driving Cars



source: Waymo

Augmented Reality

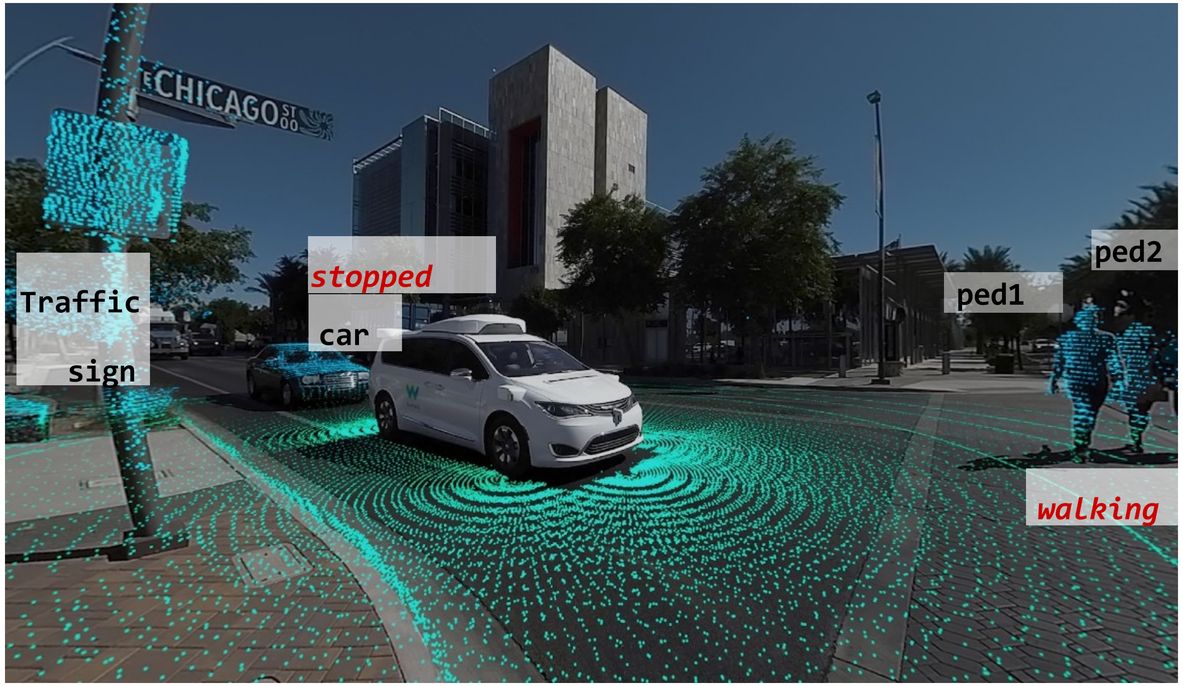


source: Microsoft HoloLens

Require **data-driven method** to **process and understand 3D data**

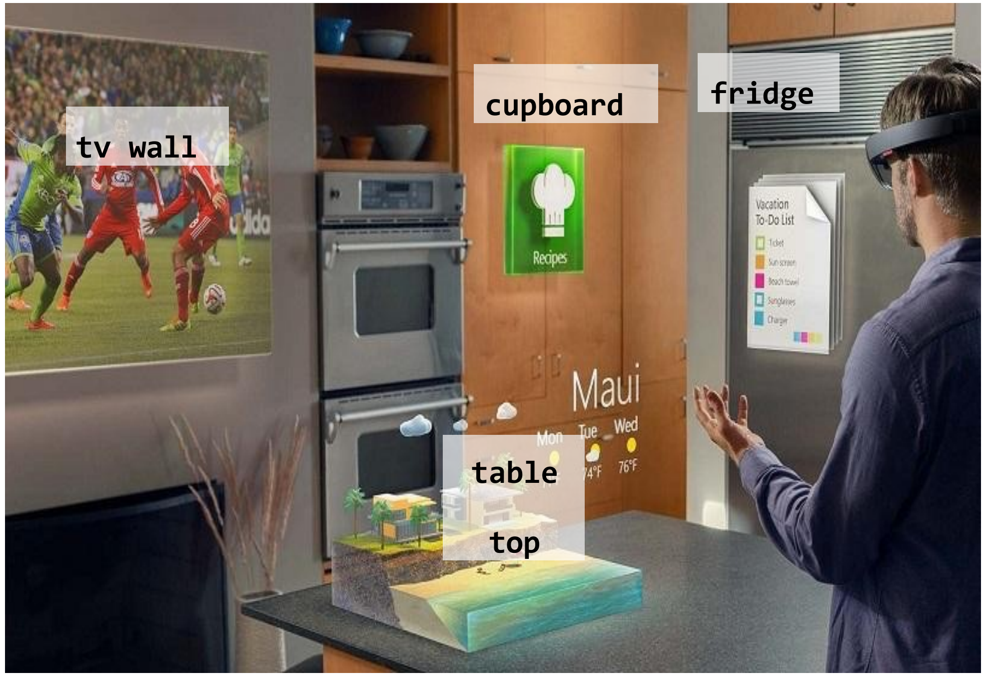
Background

Self-Driving Cars



source: Waymo

Augmented Reality

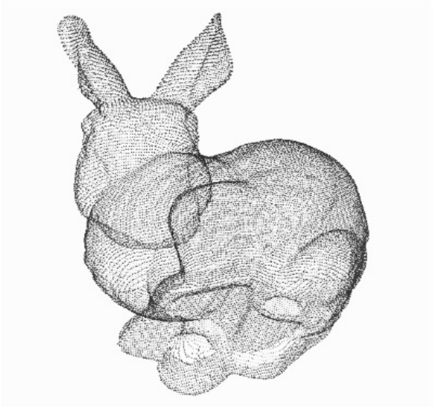


source: Microsoft HoloLens

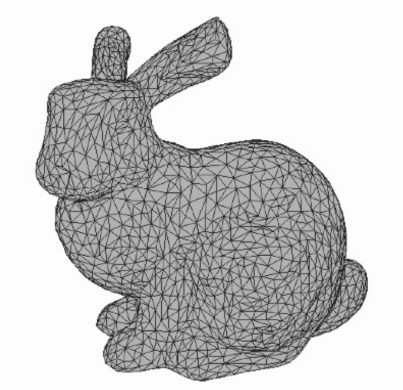
Need for 3D Deep Learning!

Background

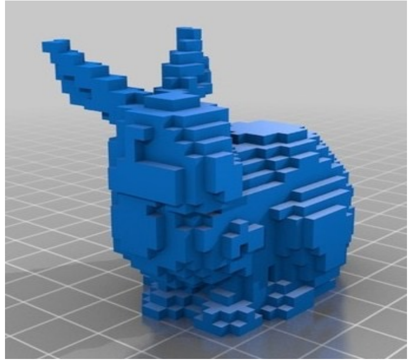
3D Representations



Point Cloud

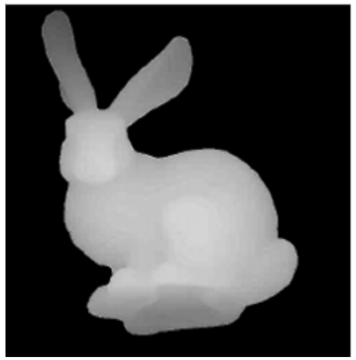


Mesh



Volumetric

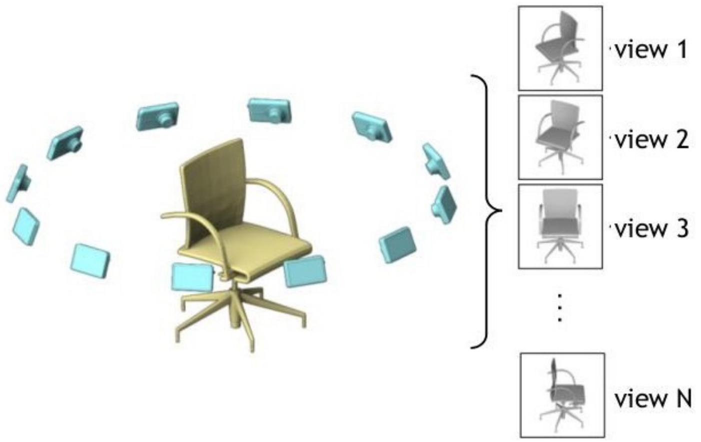
Multi-View Images



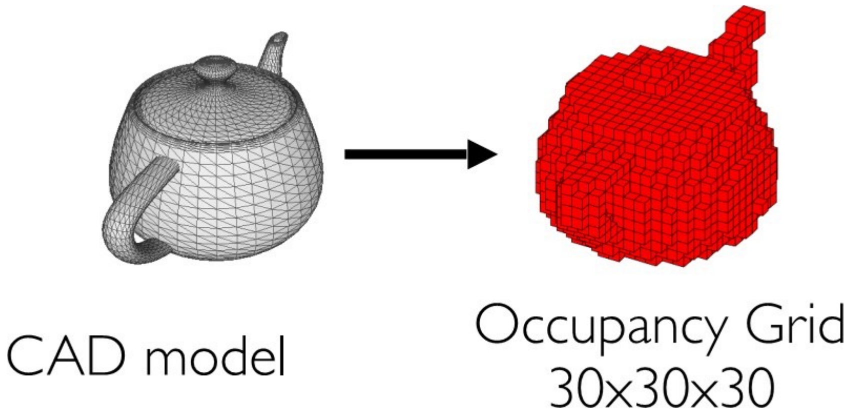
Depth Map

...

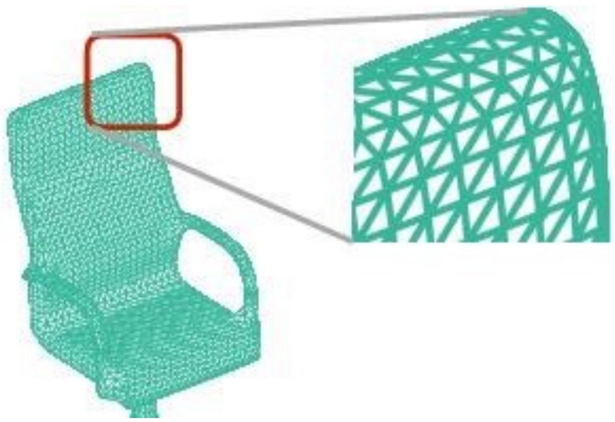
Background



multi-view images + 2D CNN



volumetric data + 3D CNN



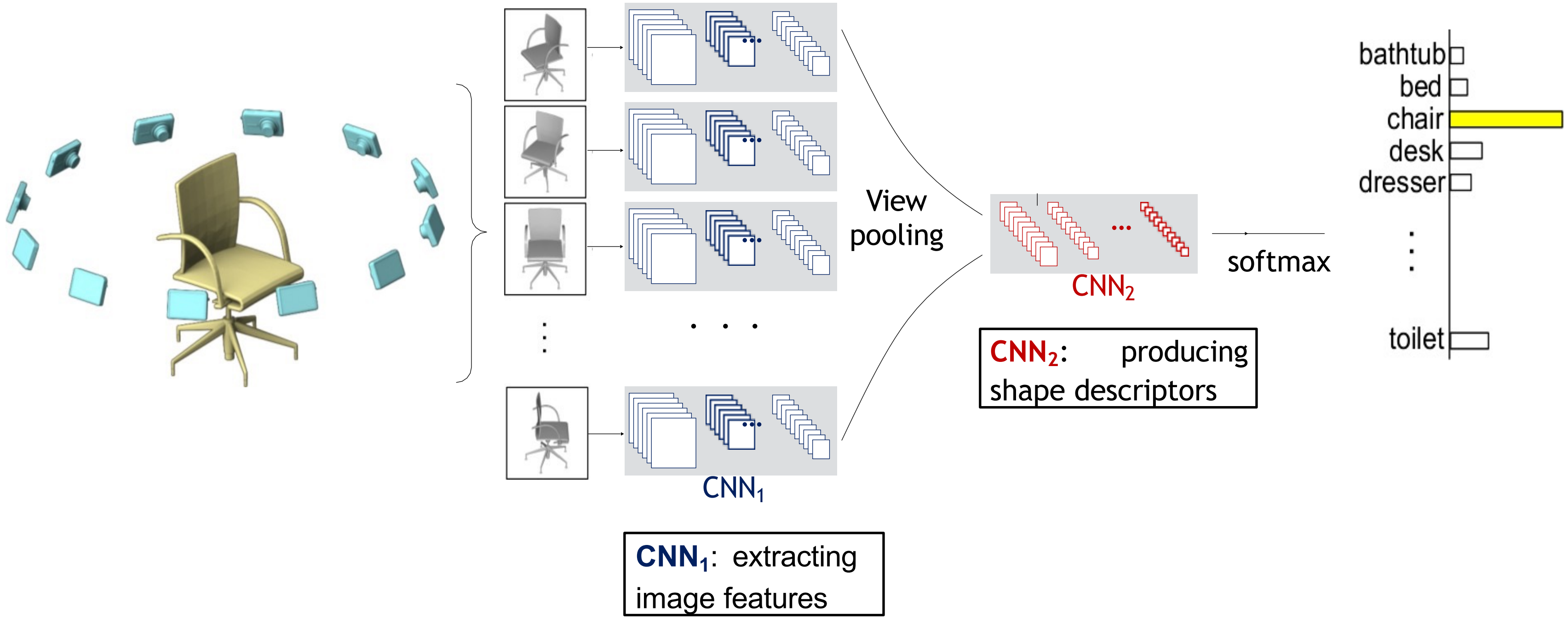
mesh data + DL (GNN) ?

image depth + CNN

point cloud + DL (CNN) ?

Background

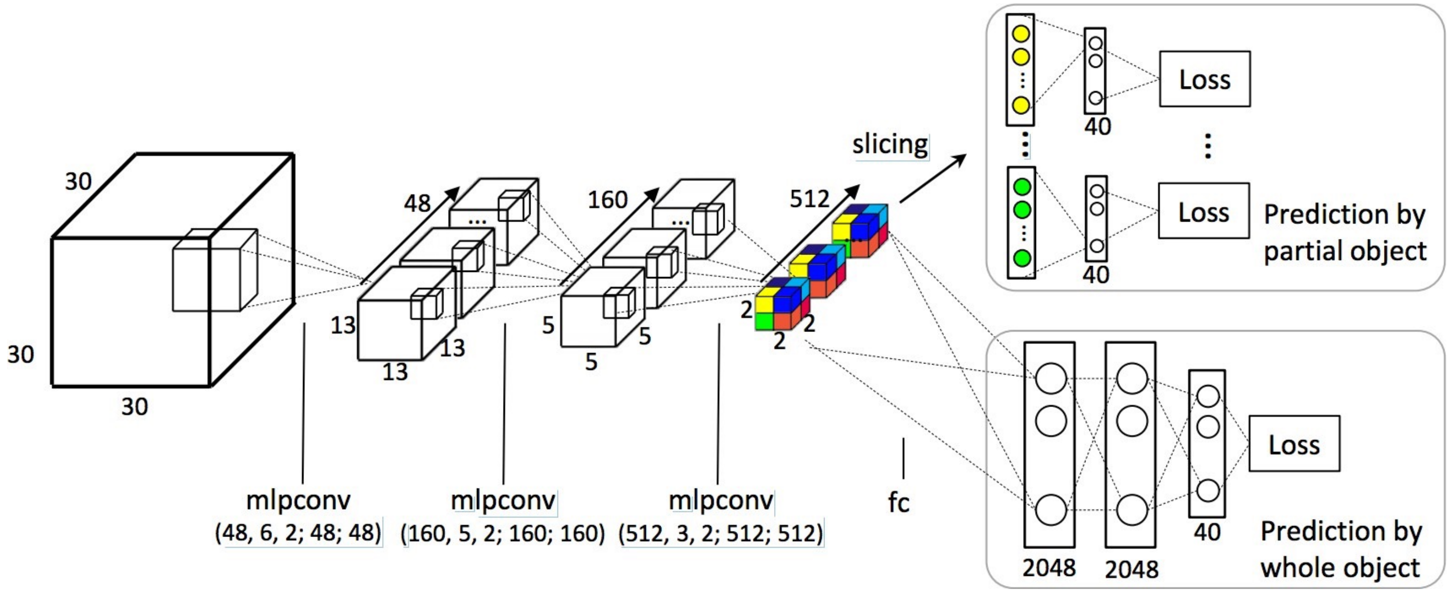
Multi-view CNN



Su et al., "**Multi-view Convolutional Neural Networks for 3D Shape Recognition**", *ICCV 2015*

Background

3D CNN for volumetric data

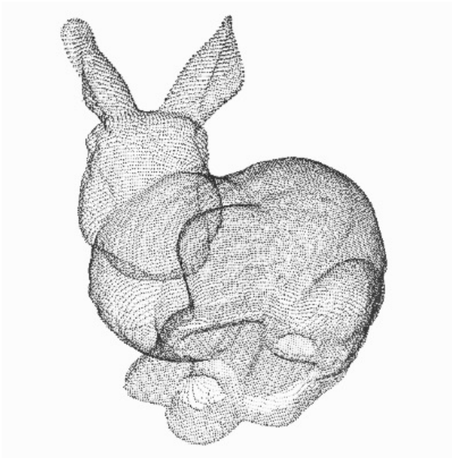


Background

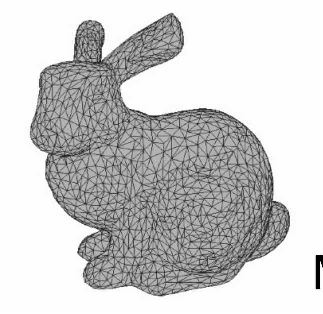
- Point cloud is close to raw sensor data
- Point cloud is representationally simple

LiDAR

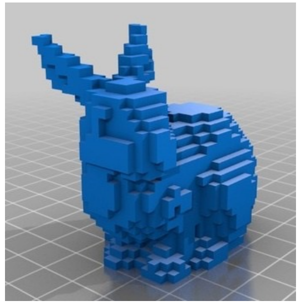
Depth Sensor



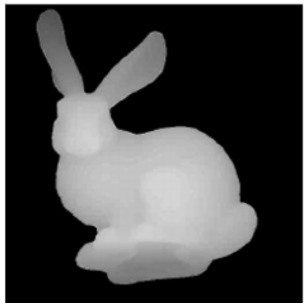
Point Cloud



Mesh



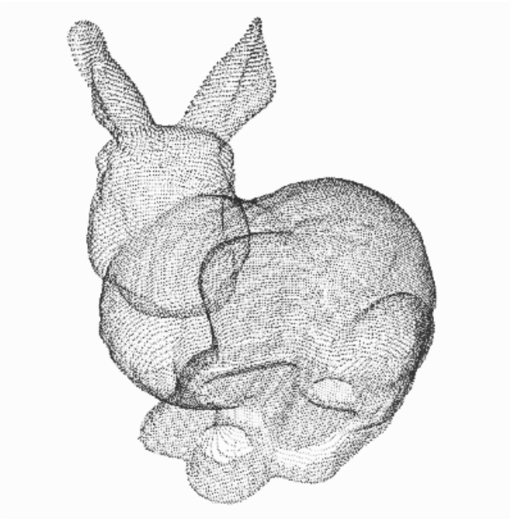
Volumetric



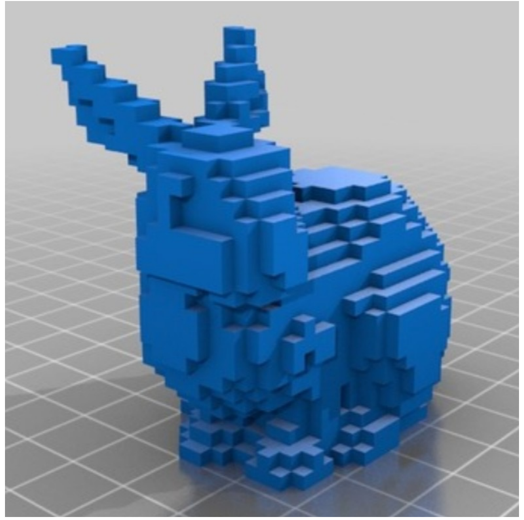
Depth Map

Background

Point cloud is **converted to other representations**
before it is input to a deep neural network



voxelization →



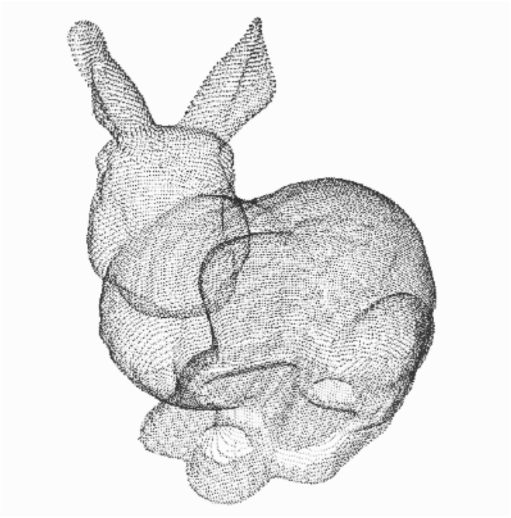
+ 3D CNN

irregular
point cloud

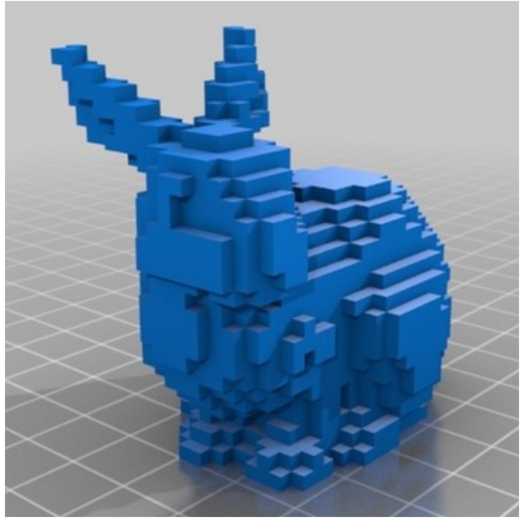
regular
volumetric array

Background

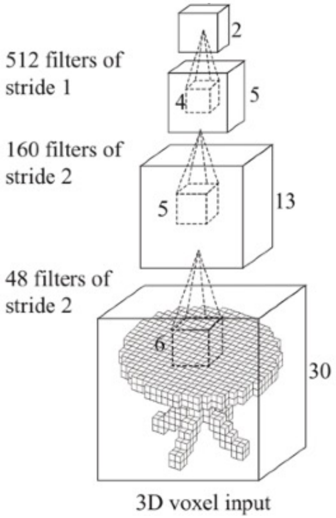
Point cloud is **converted to other representations** before it is input to a deep neural network



voxelization



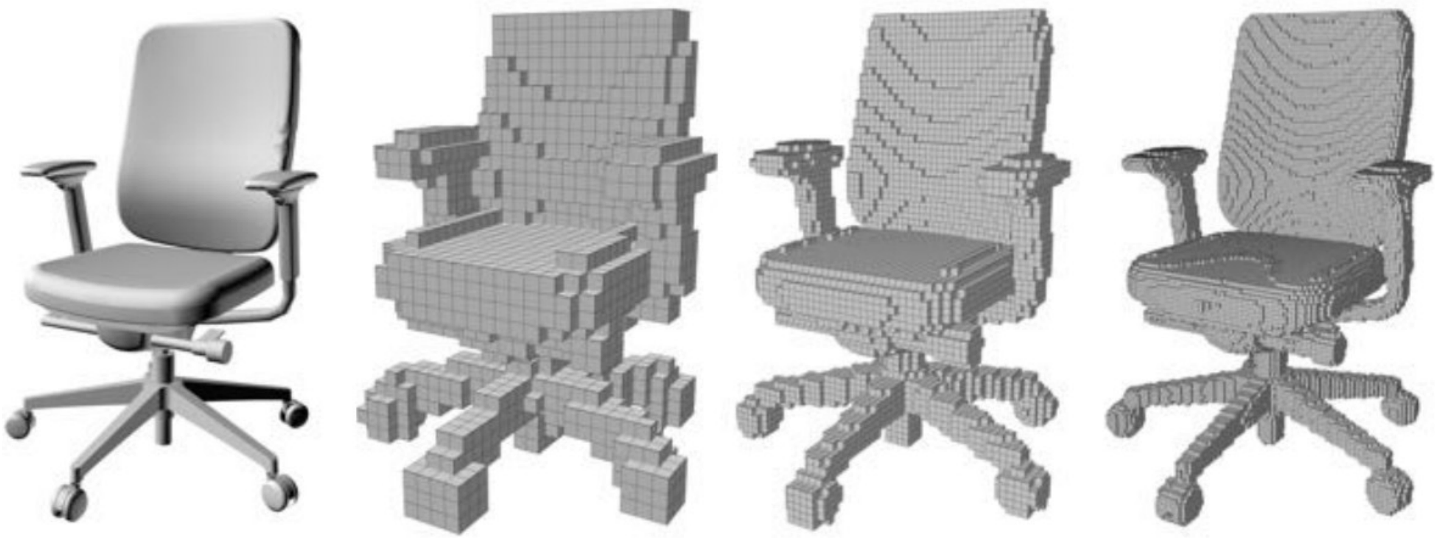
+



Con: High space & time complexity -- 3D convolution !("^{3})!

Quantization error in voxelization!

Background



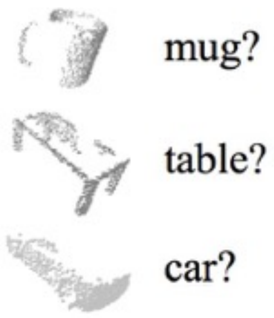
Occupancy:	10.41%	5.09%	2.41%
Resolution:	32	64	128

Background

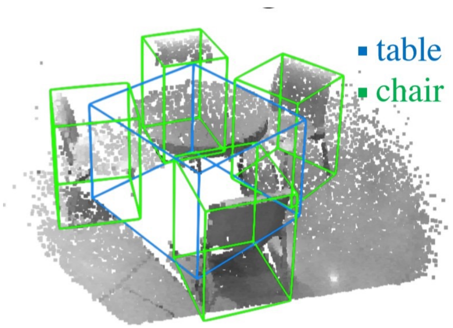
Can we achieve effective **feature learning directly** on point clouds?

Background

Recognition tasks on 3D point clouds

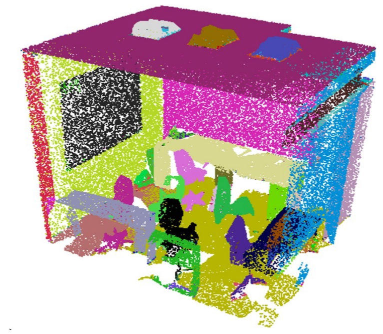


Classification



Object Detection

Part Segmentation

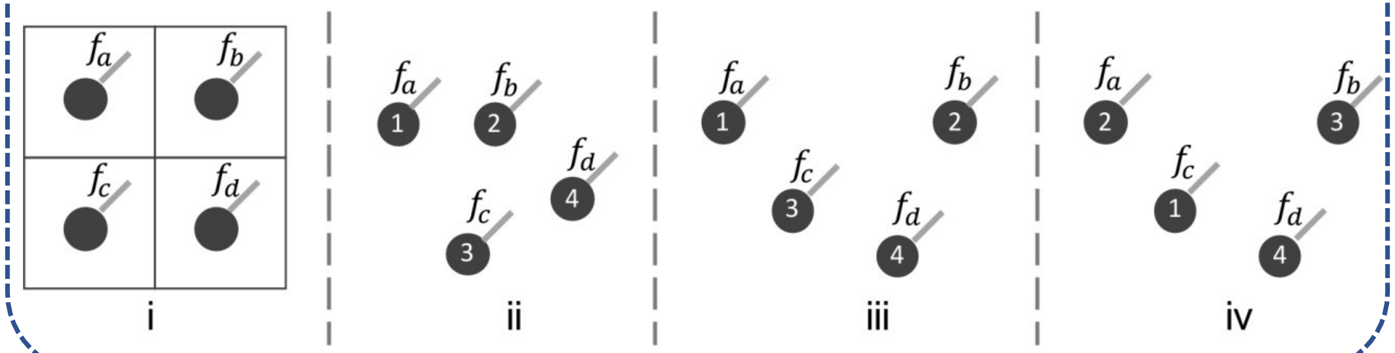


Semantic Segmentation

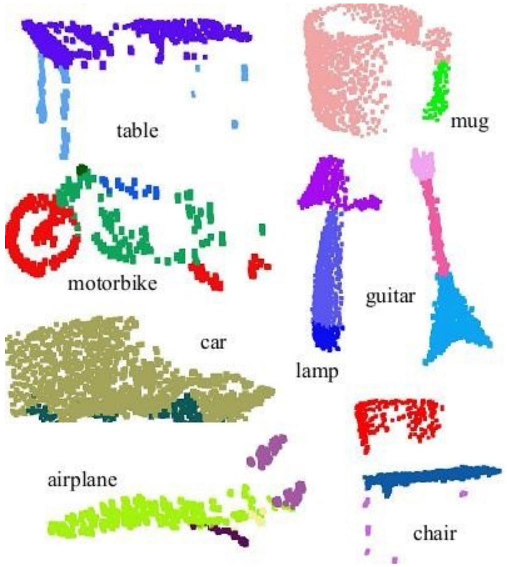
...

Background

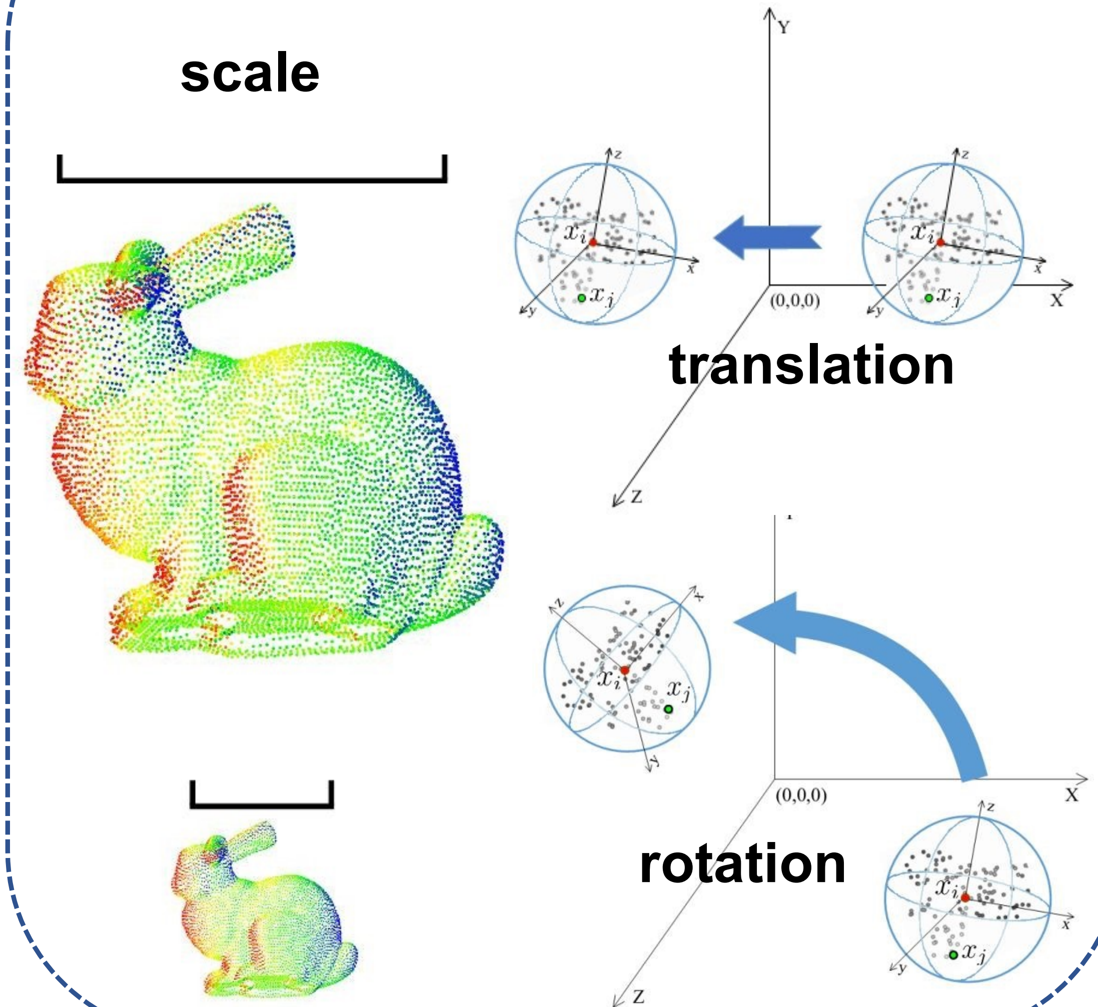
Irregular (unordered): permutation invariance



Robustness to corruption, outlier, noise; partial data

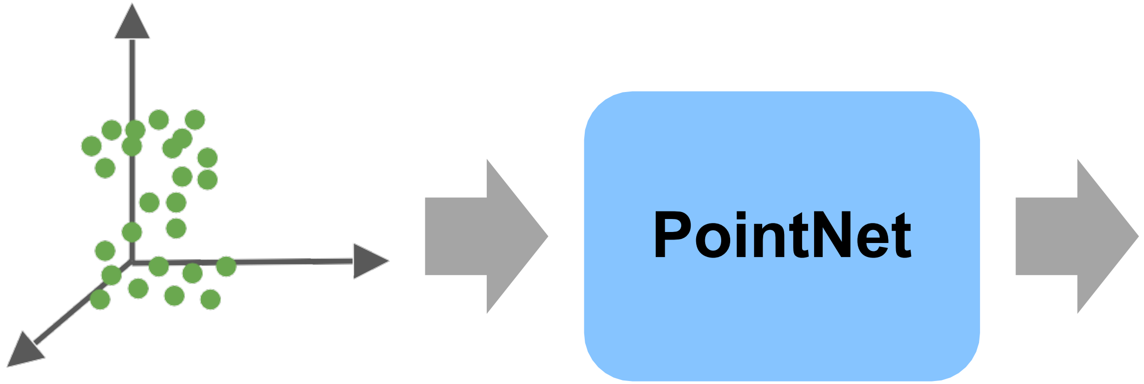


Robustness to rigid transformations



PointNet

- **End-to-end learning** for irregular point data
- **Unified** framework for various tasks



Object Classification

Object Part Segmentation

Semantic Scene Parsing

...

*[1] **Charles R. Qi**, Hao Su, Kaichun Mo, Leonidas J. Guibas. PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation. (CVPR'17)*

PointNet

Challenges

Permutation Invariance

Point cloud is a set of unordered points

Transformation Invariance

Point cloud rotations should not alter classification results

PointNet

Challenges

Permutation Invariance

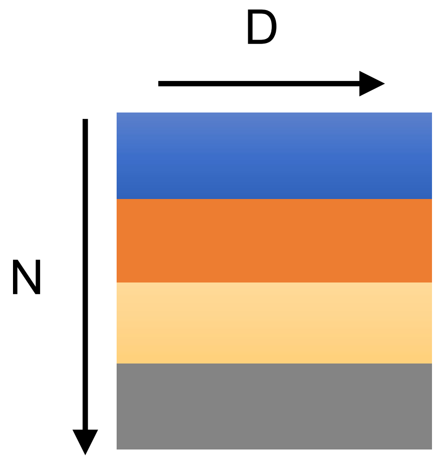
Point cloud is a set of unordered points

Transformation Invariance

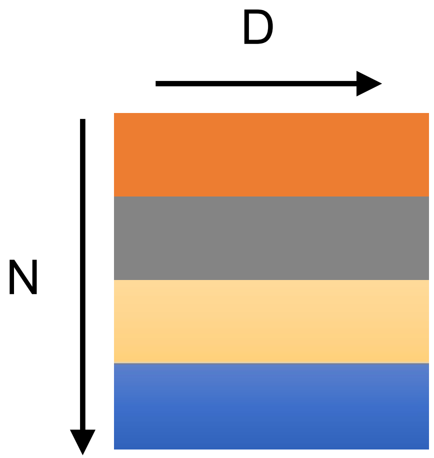
Point cloud rotations should not alter classification results

PointNet

Point cloud: N **unordered** points, each represented by a D dim vector



represents the same **set** as



Model needs to be invariant to N! permutations

PointNet

Three strategies to achieve permutation invariance:

- Sort input into a canonical order

- Treat the input as a sequence to train an RNN
  & augment the training data with all permutations

- Use a simple symmetric function to aggregate point information

PointNet

Permutation Invariance: Symmetric Function

$$f(x_1, x_2, \ldots, x_n) \equiv f(x_{\pi_1}, x_{\pi_2}, \ldots, x_{\pi_n})$$

Examples:

$$f(x_1, x_2, \ldots, x_n) = \max\{x_1, x_2, \ldots, x_n\}$$

$$f(x_1, x_2, \ldots, x_n) = x_1 + x_2 + \ldots + x_n$$

…

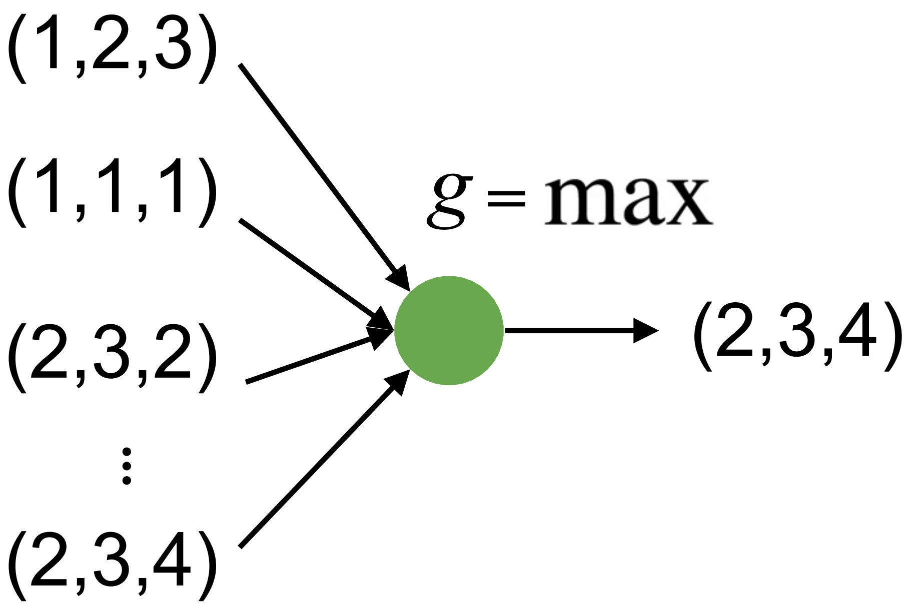
How can we construct a family of symmetric functions by neural networks?

PointNet

Construct Symmetric Functions by Neural Networks

Simplest form: directly aggregate all points with a symmetric operator g

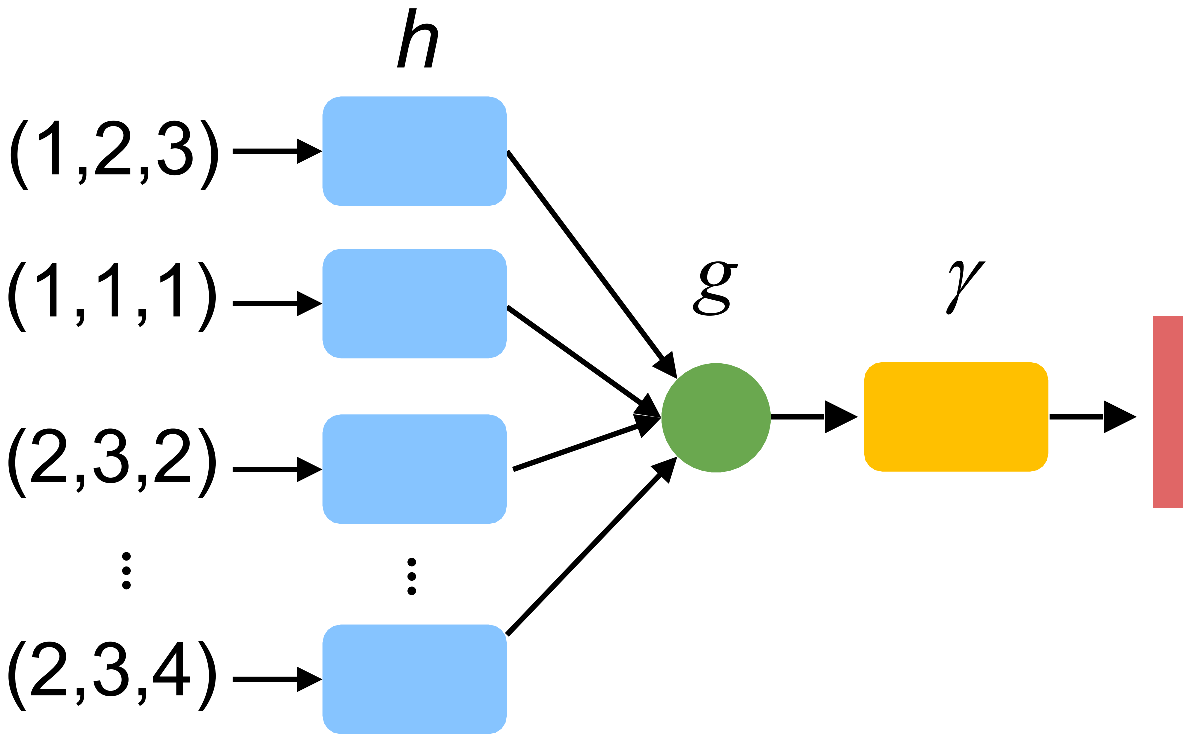
Just discovers very extreme/naive property of the geometry.



PointNet

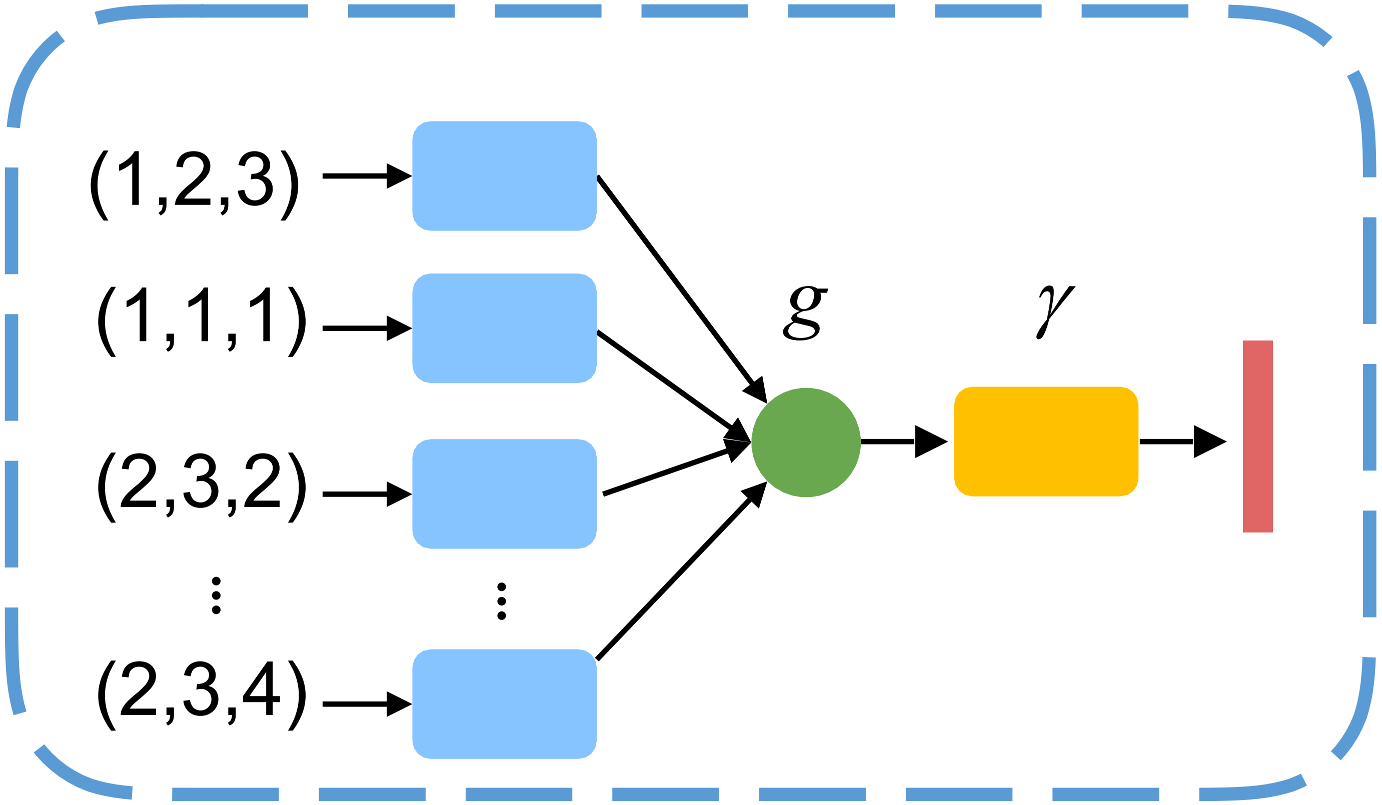
Embed points to a high-dim space before aggregation.

Aggregation in the (redundant) high-dim space preserves interesting properties of the geometry.



PointNet

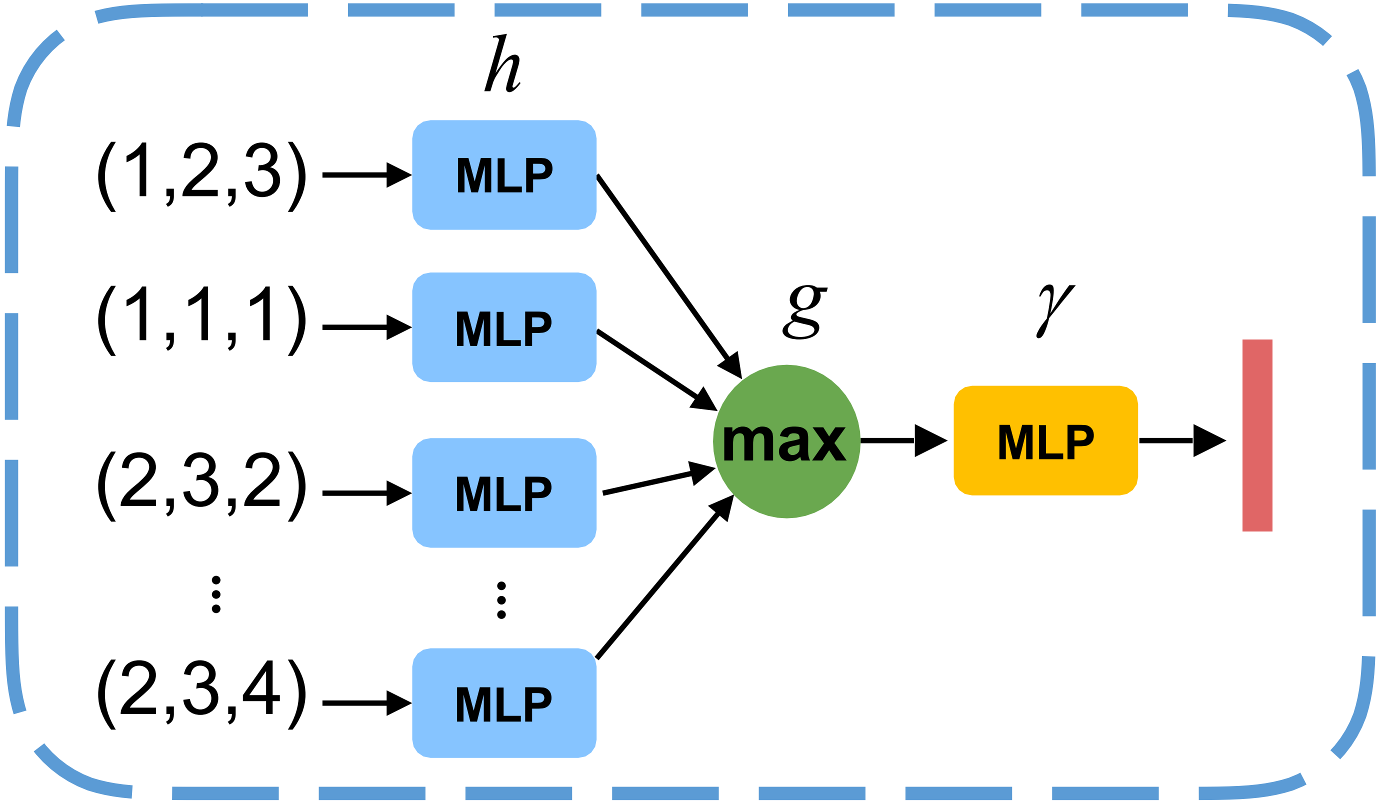
$$f(x_1, x_2, \ldots, x_n) = \gamma \mapsto g(h(x_1), \ldots, h(x_n))$$ is symmetric if g is symmetric



PointNet (vanilla)

PointNet

$$f(x_1, x_2, \ldots, x_n) = \gamma \mapsto g(h(x_1), \ldots, h(x_n))$$ is symmetric if g is symmetric



PointNet (vanilla)

PointNet

Challenges

Permutation Invariance

Point cloud is a set of unordered points

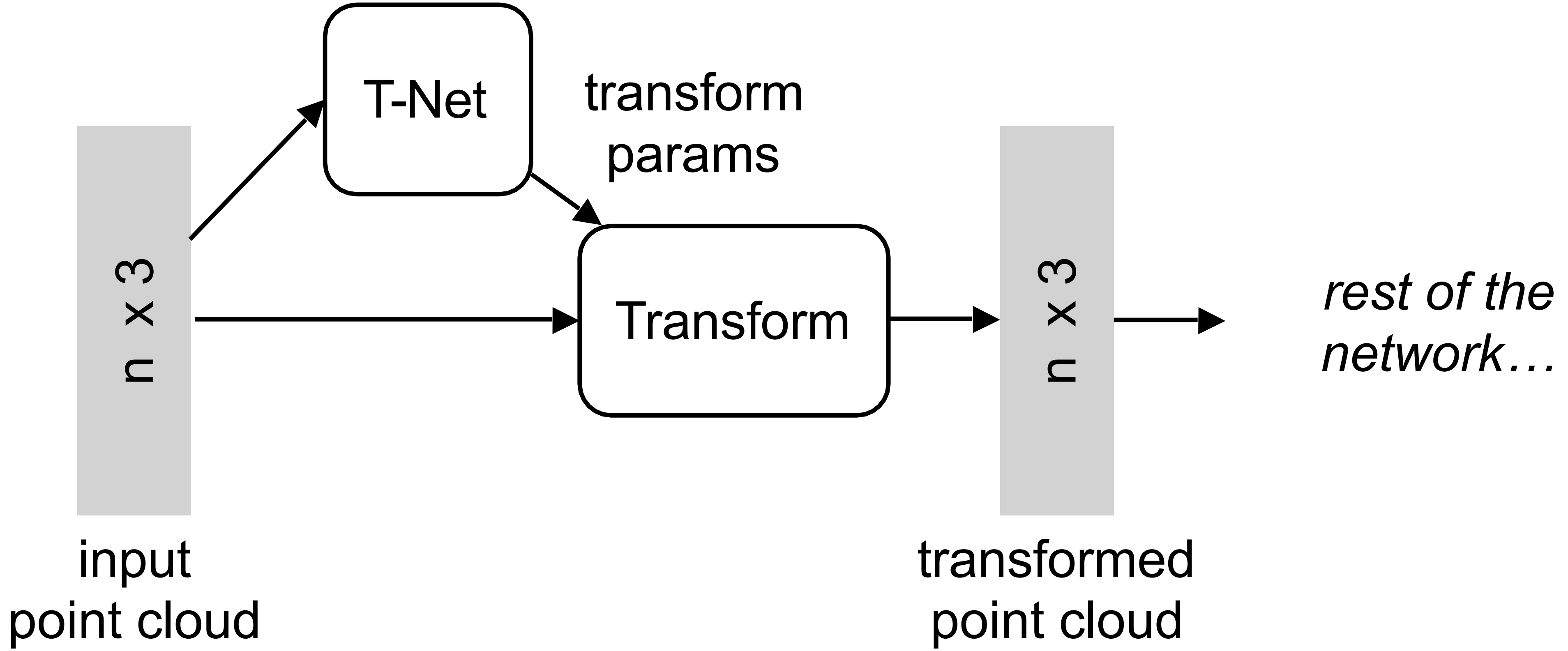
Transformation Invariance

Point cloud rotations should not alter classification results

PointNet

Input Alignment by Transformer Network

Idea: Data dependent transformation for automatic alignment

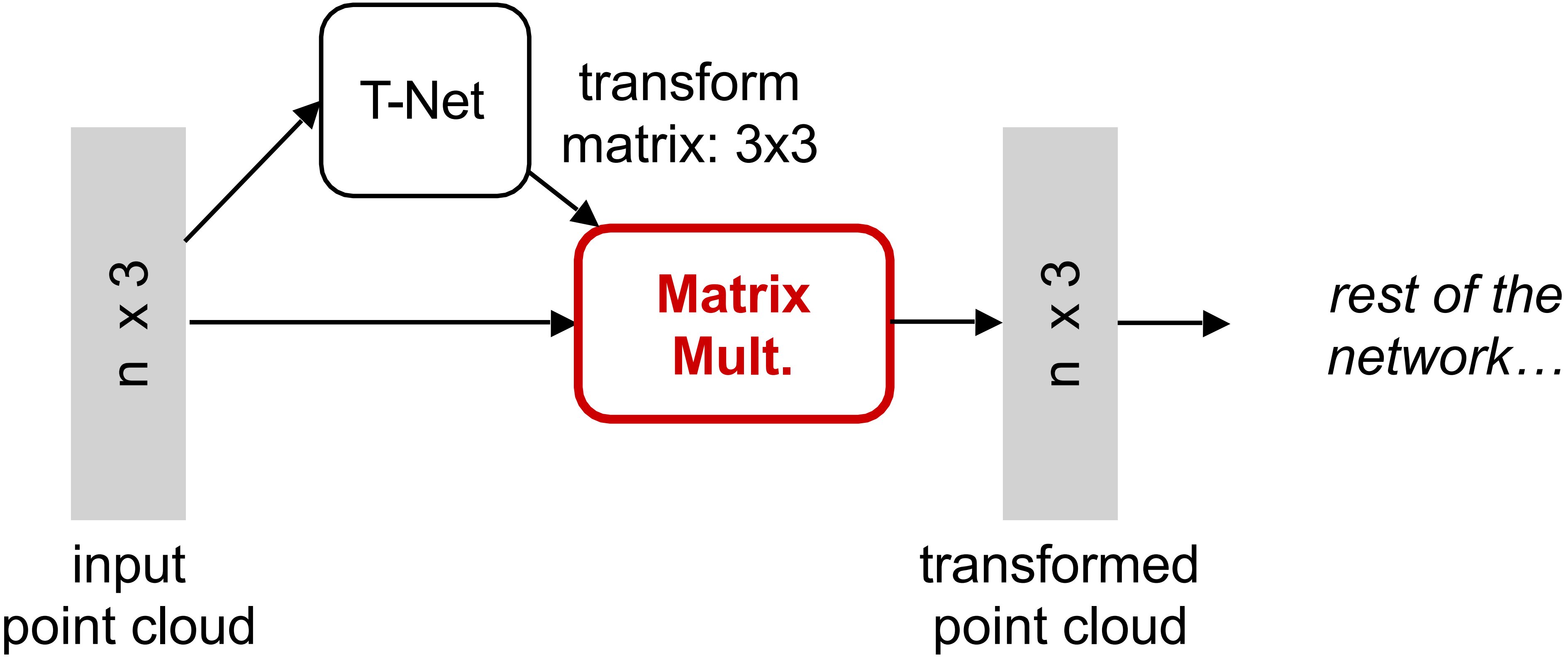


PointNet

Input Alignment by Transformer Network

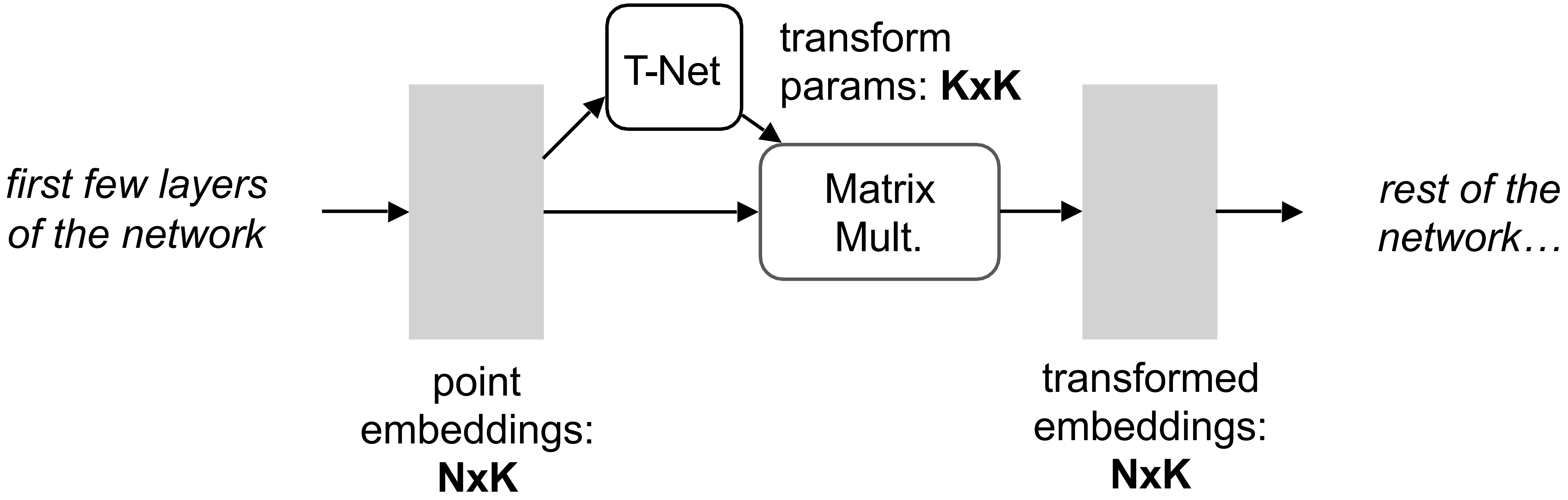
Idea: Data dependent transformation for automatic alignment

The transformation is just matrix multiplication!



PointNet

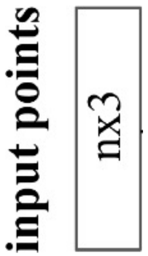
Embedding Space Alignment



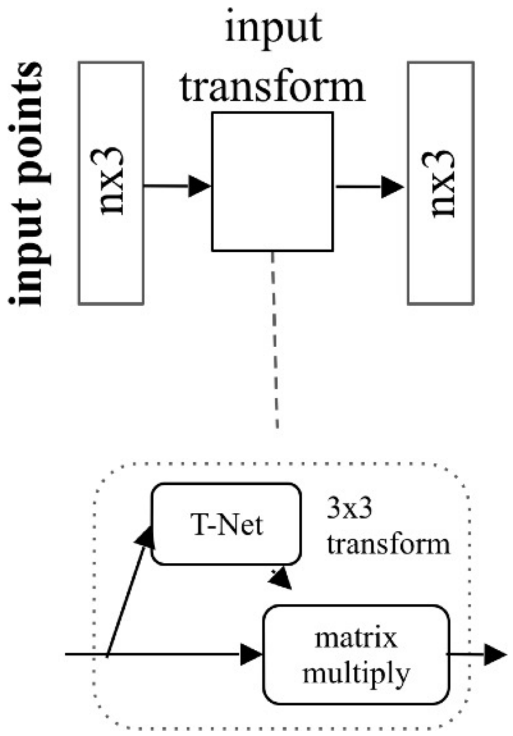
Regularization loss:

Transform matrix close to orthogonal: $L_{reg} = \|I - AA^T\|_F^2$

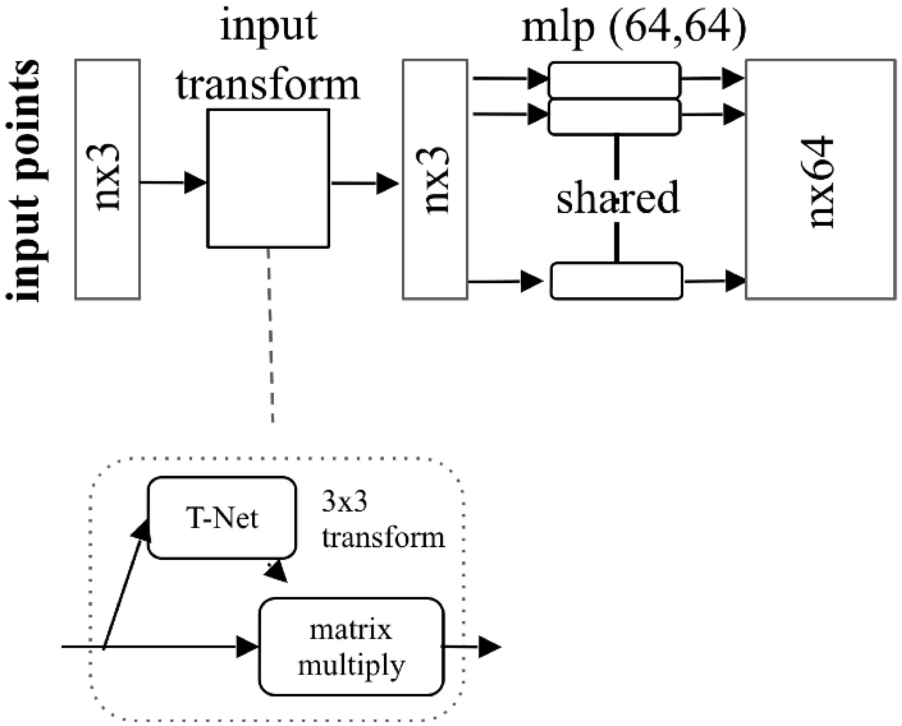
PointNet Classification Network



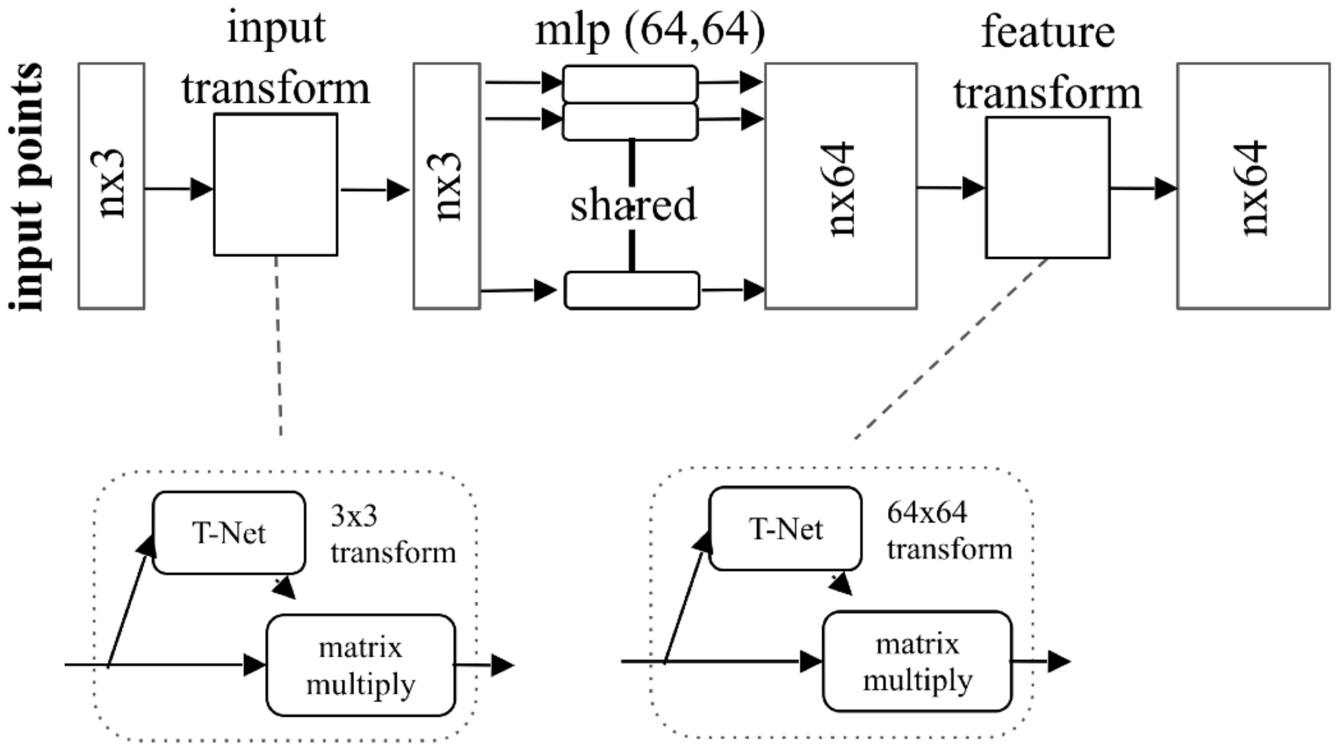
PointNet Classification Network



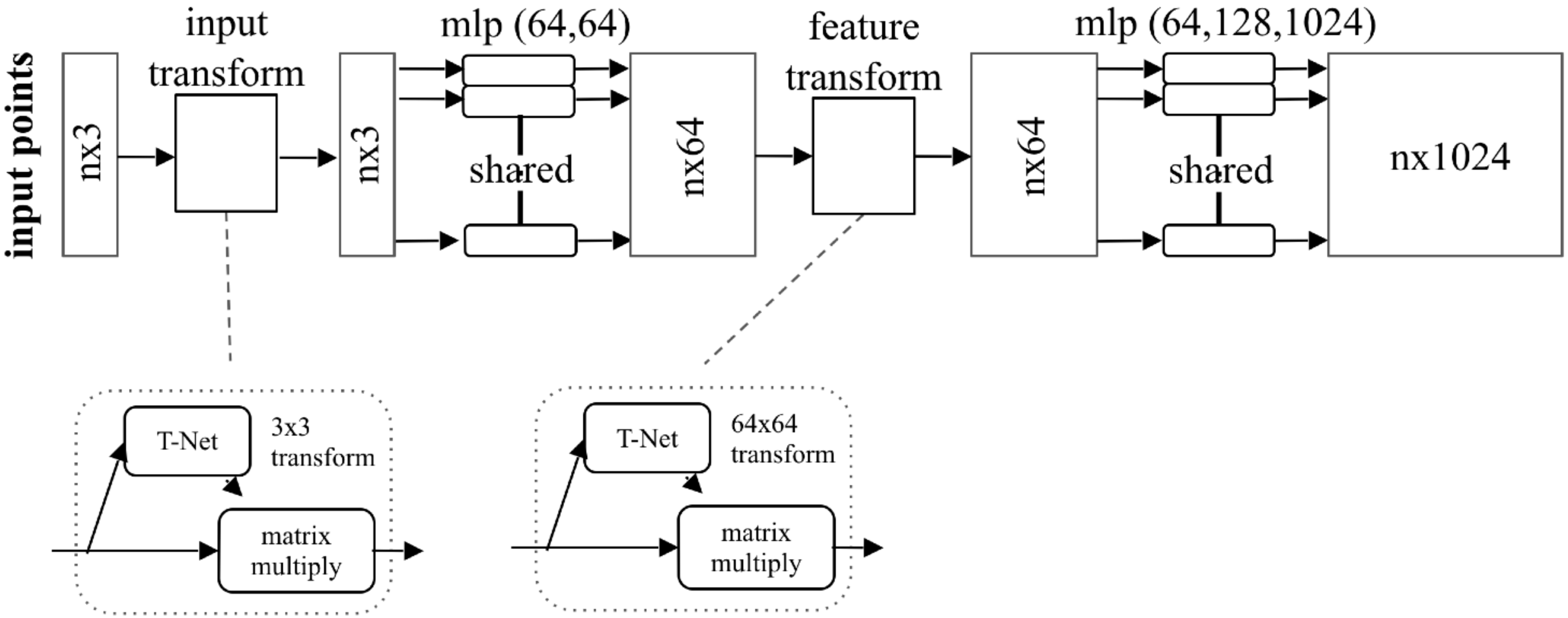
PointNet Classification Network



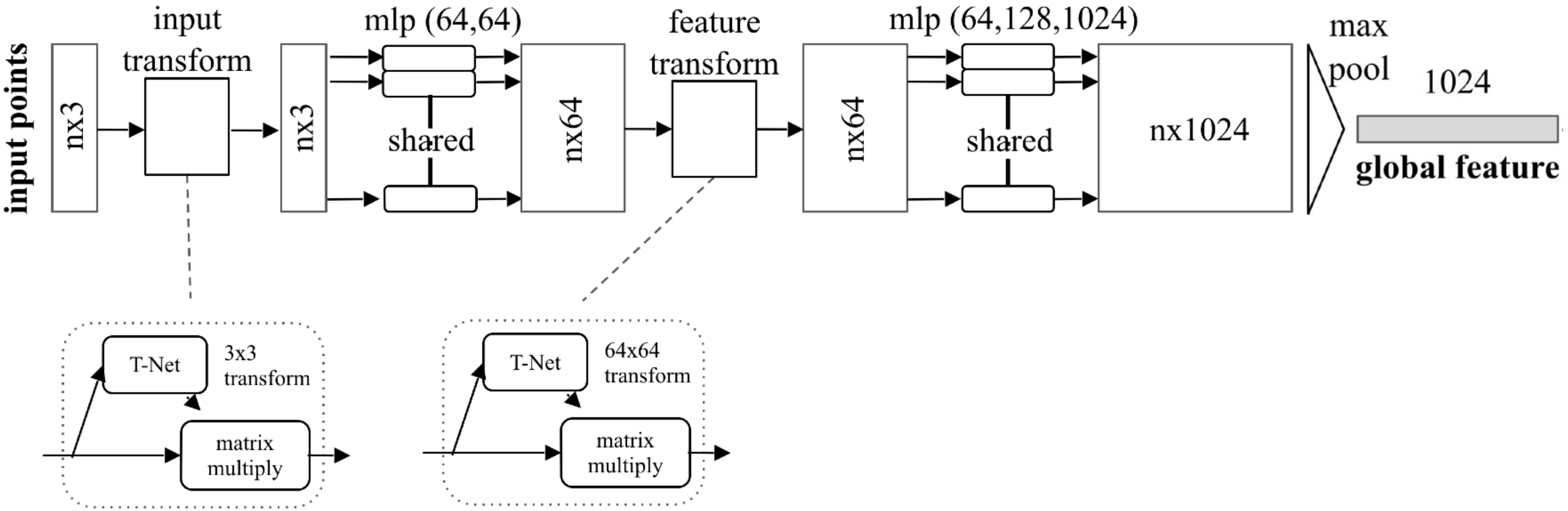
PointNet Classification Network



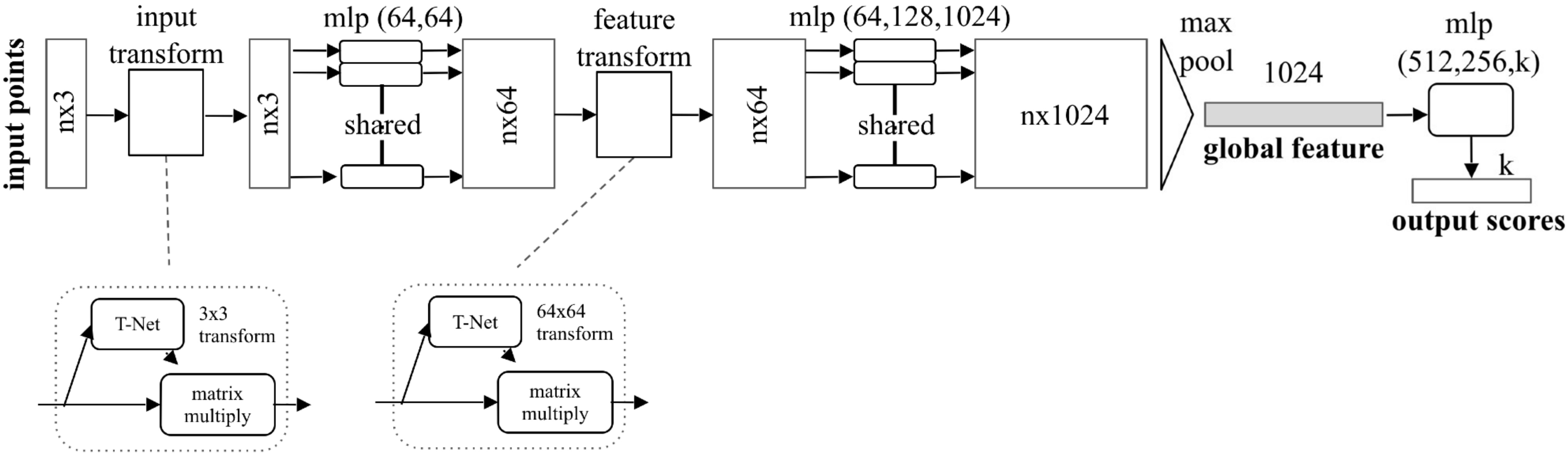
PointNet Classification Network



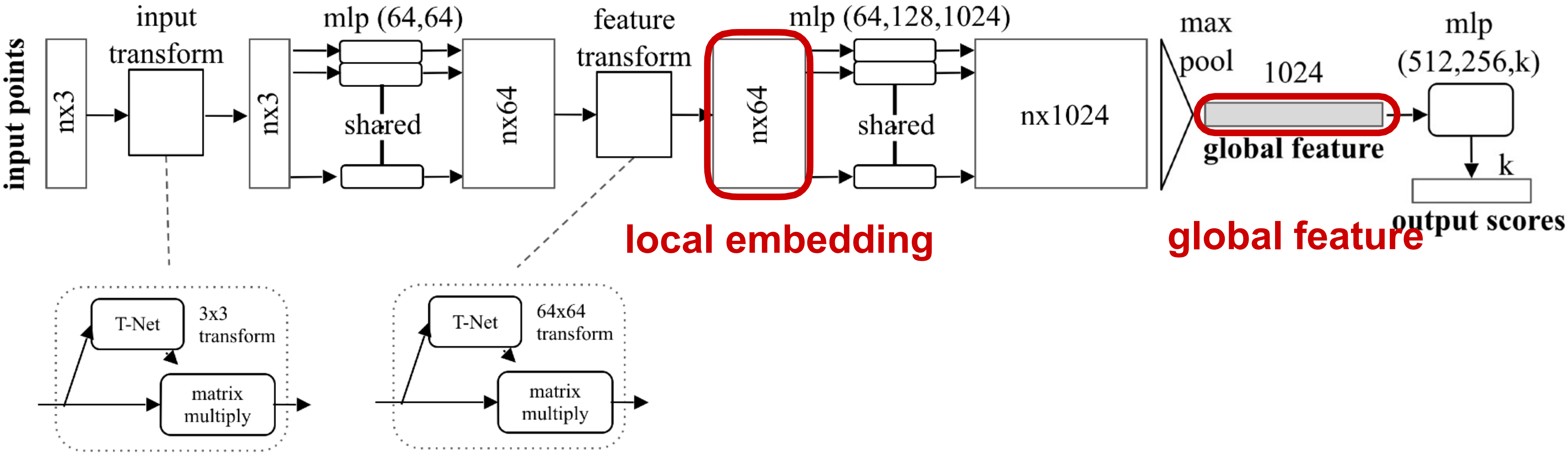
PointNet Classification Network



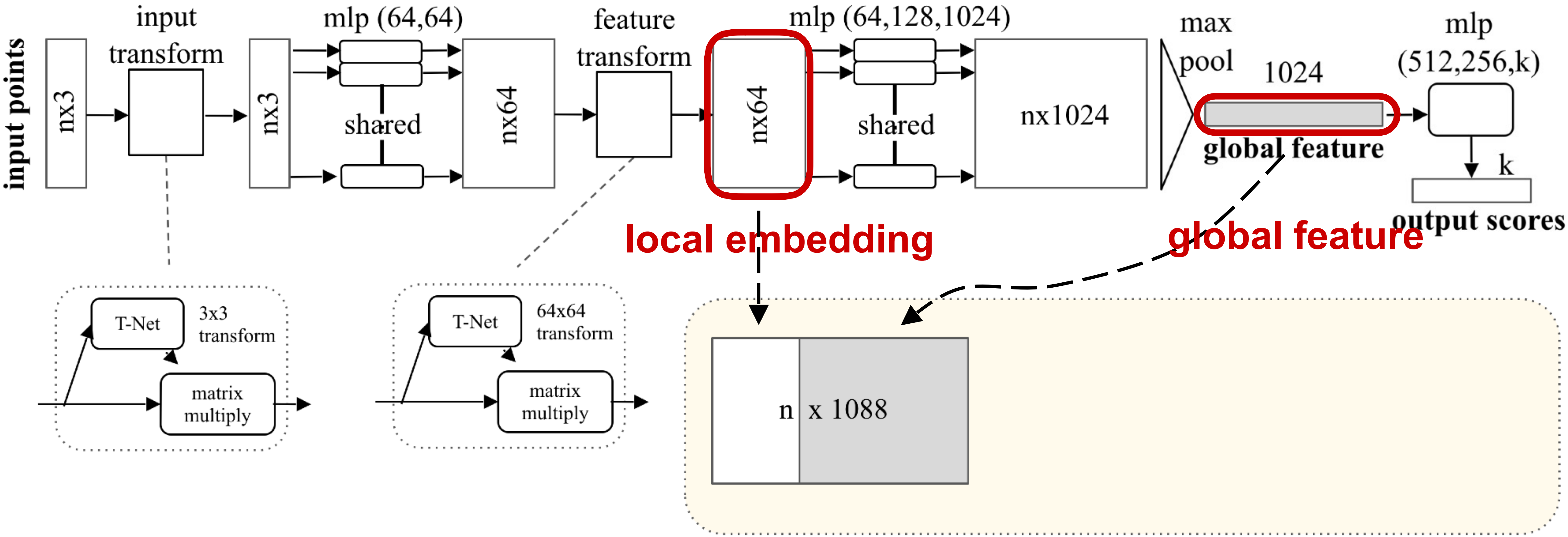
PointNet Classification Network



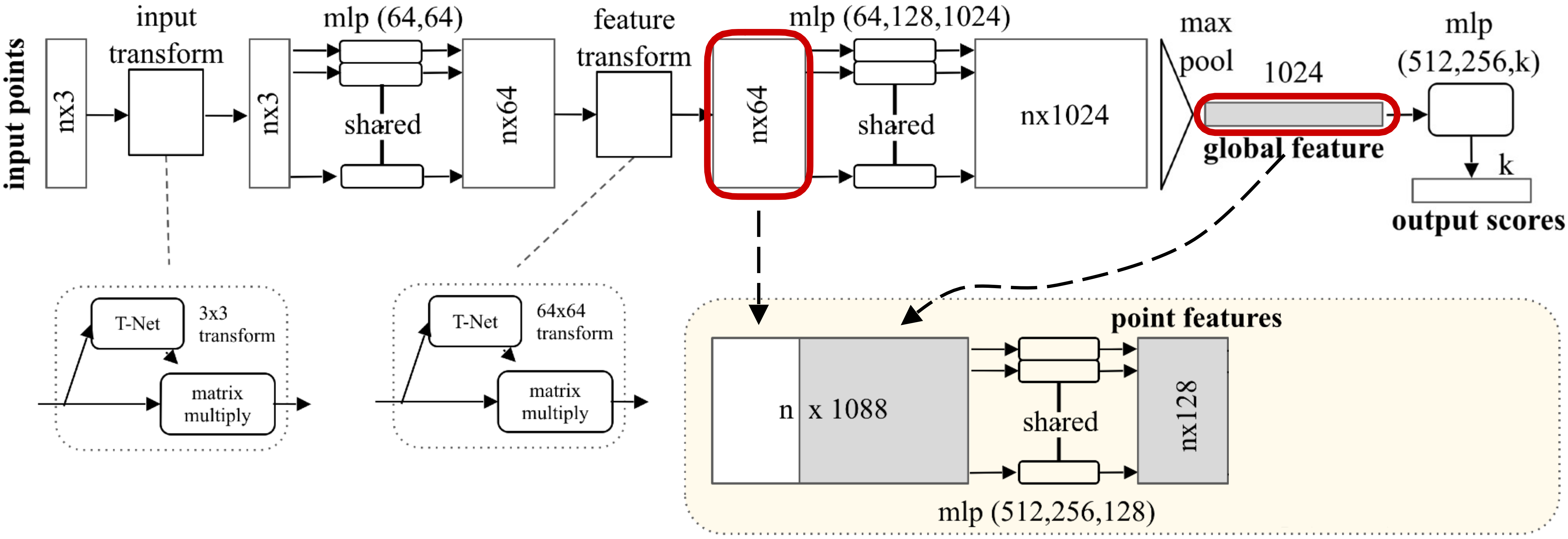
PointNet Classification Network



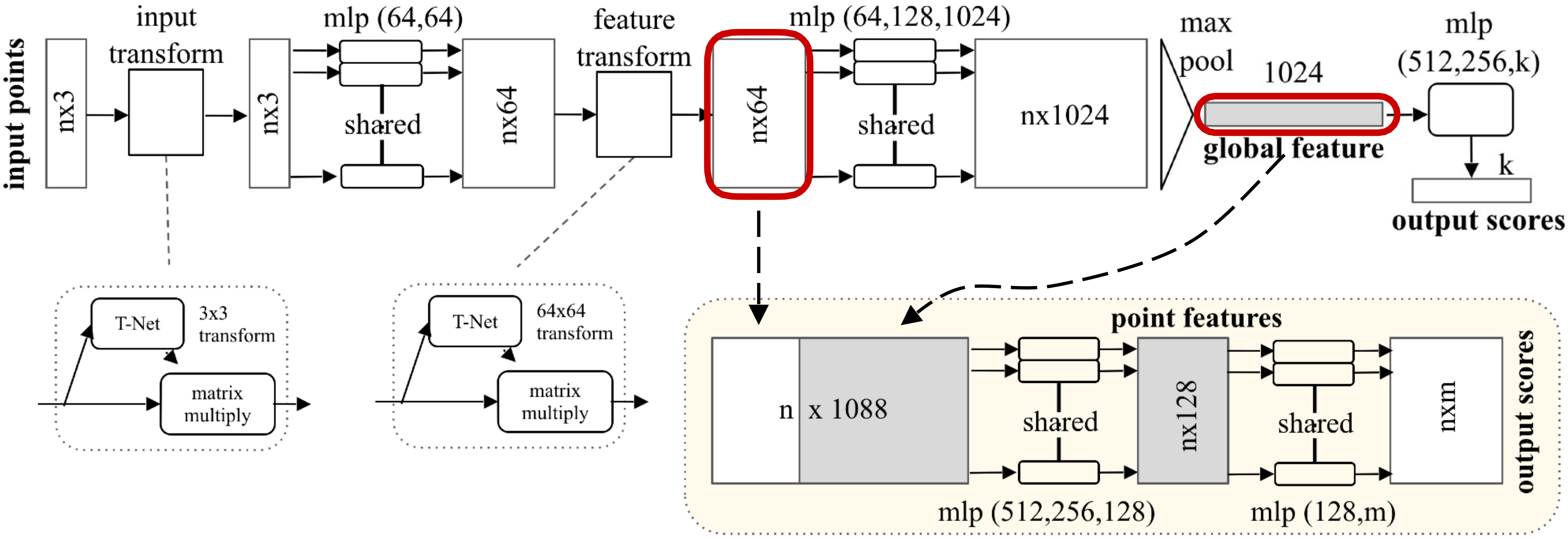
PointNet Classification Network



PointNet Classification Network



Extension to PointNet Segmentation Network



PointNet

Results on Object Classification

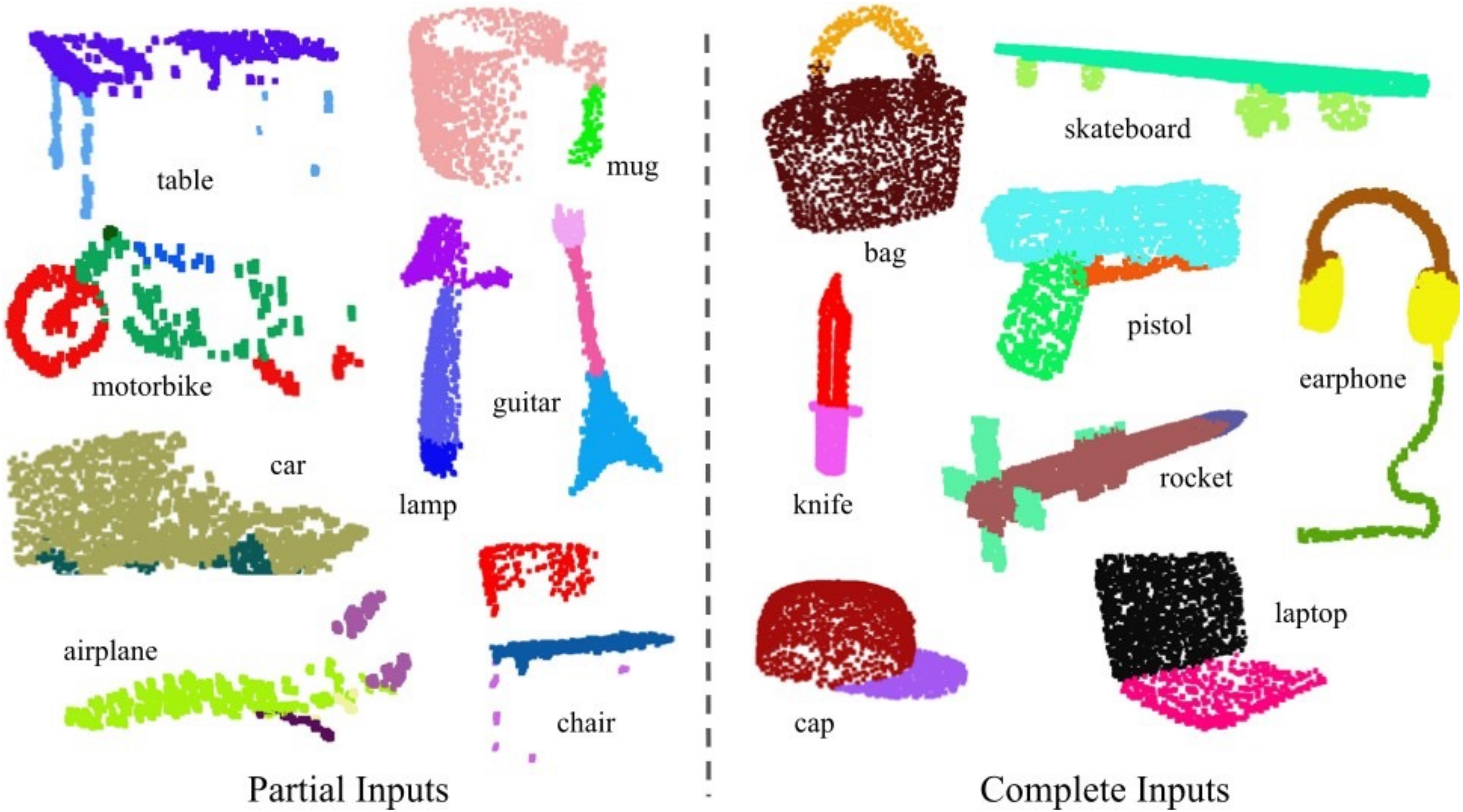
	input	#views	accuracy avg. class	accuracy overall
SPH [12]	mesh	-	68.2	-
3DShapeNets [29]	volume	1	77.3	84.7
VoxNet [18]	volume	12	83.0	85.9
Subvolume [19]	volume	20	86.0	**89.2**
LFD [29]	image	10	75.5	-
MVCNN [24]	image	80	**90.1**	-
Ours baseline	point	-	72.6	77.4
Ours PointNet	point	1	86.2	**89.2**

3D CNNs

dataset: ModelNet40; metric: 40-class classification accuracy (%)

PointNet

Results on Object Part Segmentation



PointNet

Results on Object Part Segmentation

	mean	aero	bag	cap	car	chair	ear phone	guitar	knife	lamp	laptop	motor	mug	pistol	rocket	skate board	table
# shapes		2690	76	55	898	3758	69	787	392	1547	451	202	184	283	66	152	5271
Wu [28]	-	63.2	-	-	-	73.5	-	-	-	74.4	-	-	-	-	-	-	74.8
Yi [30]	81.4	81.0	78.4	77.7	**75.7**	87.6	61.9	**92.0**	85.4	**82.5**	**95.7**	**70.6**	91.9	**85.9**	53.1	69.8	75.3
3DCNN	79.4	75.1	72.8	73.3	70.0	87.2	63.5	88.4	79.6	74.4	93.9	58.7	91.8	76.4	51.2	65.3	77.1
Ours	**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**

dataset: ShapeNetPart; metric: mean IoU (%)

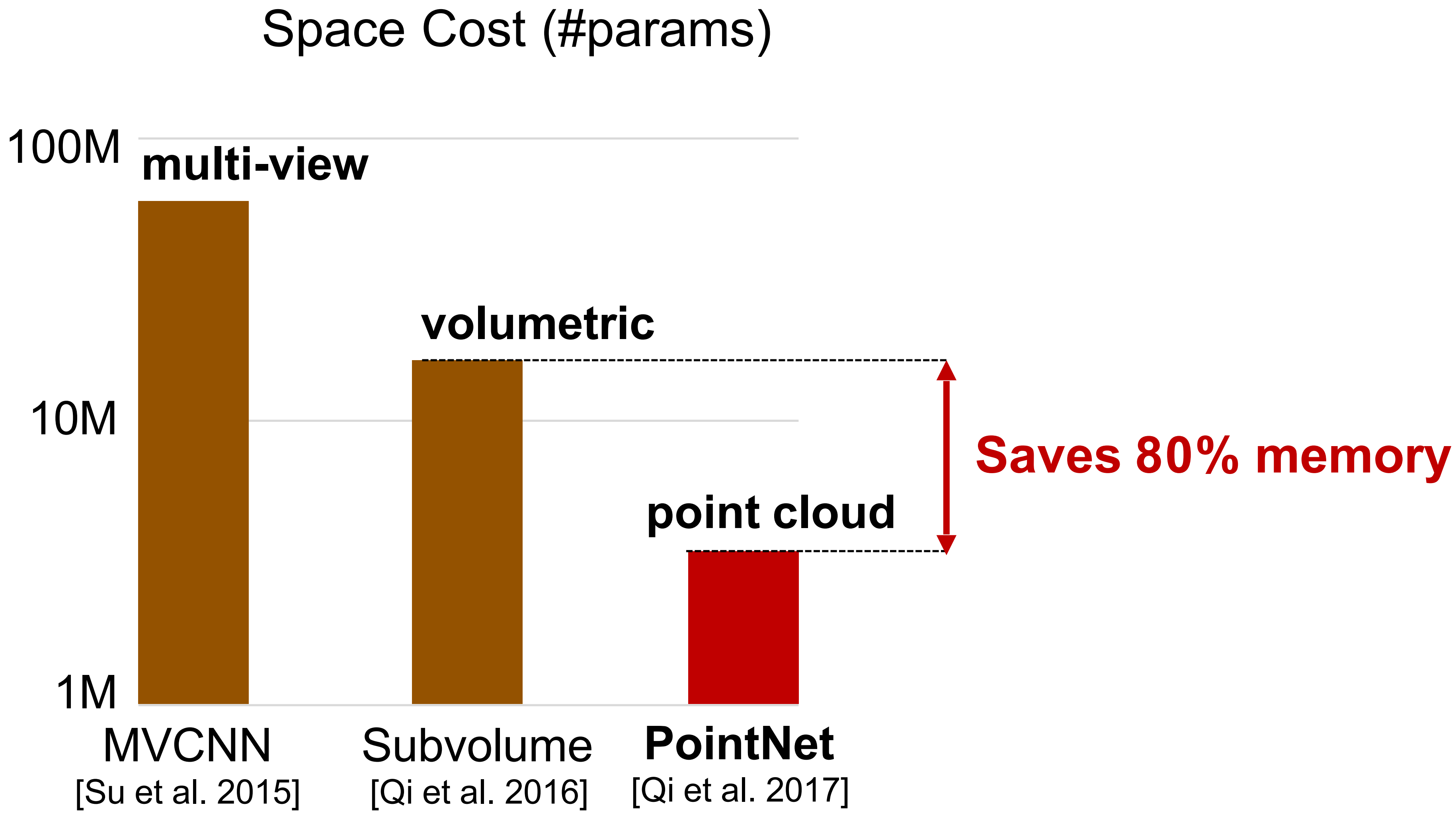
PointNet

Results on Semantic Scene Parsing

dataset: Stanford 2D-3D-S (Matterport scans)

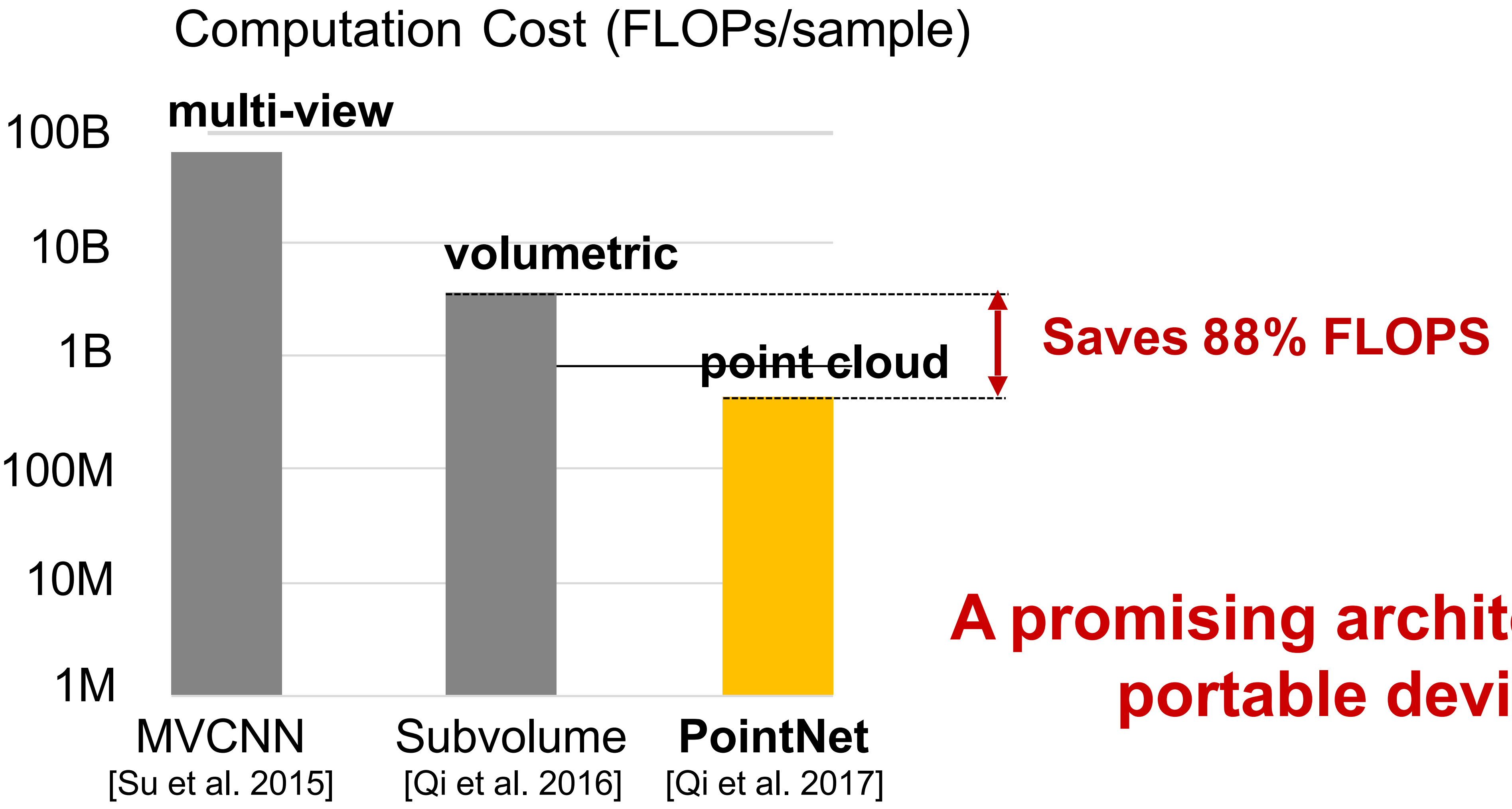
PointNet

PointNet is Light-Weight and Fast



PointNet

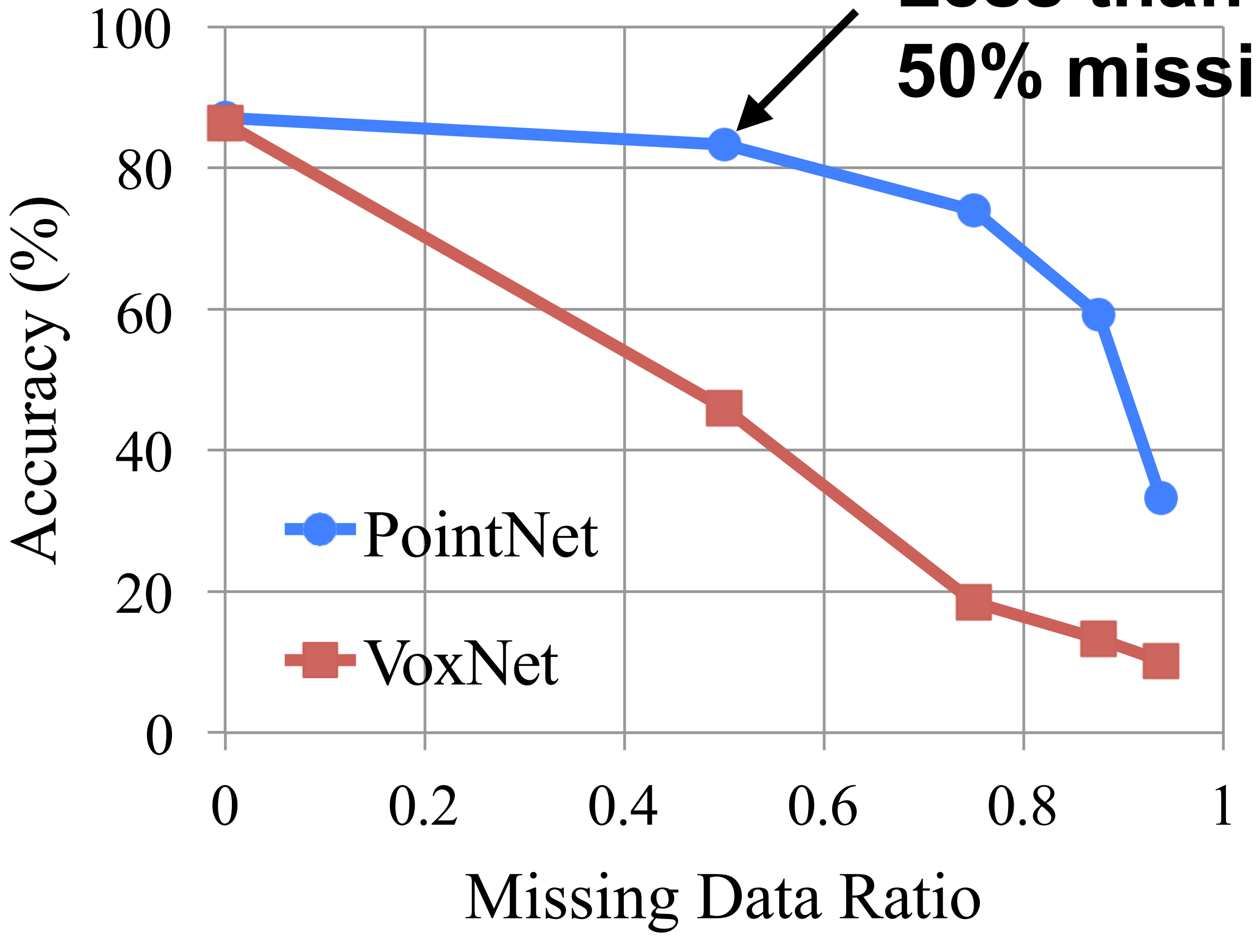
PointNet is Light-Weight and Fast



A promising architecture for portable devices!

PointNet

PointNet is Robust to Data Corruption



Less than 2% accuracy drop with 50% missing data

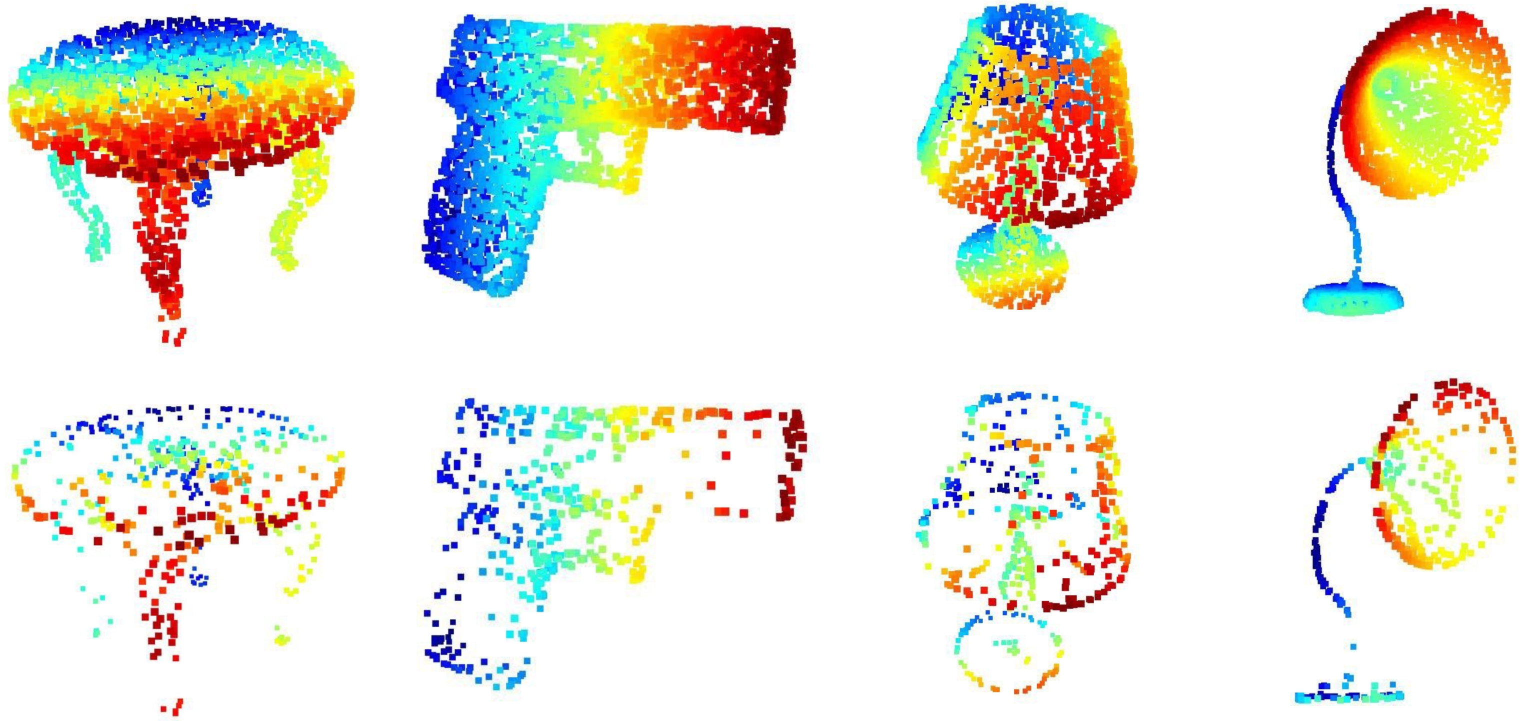
Why is PointNet so robust to missing data?

dataset: ModelNet40; metric: 40-class classification accuracy (%)

PointNet

Original Shape

Critical Points

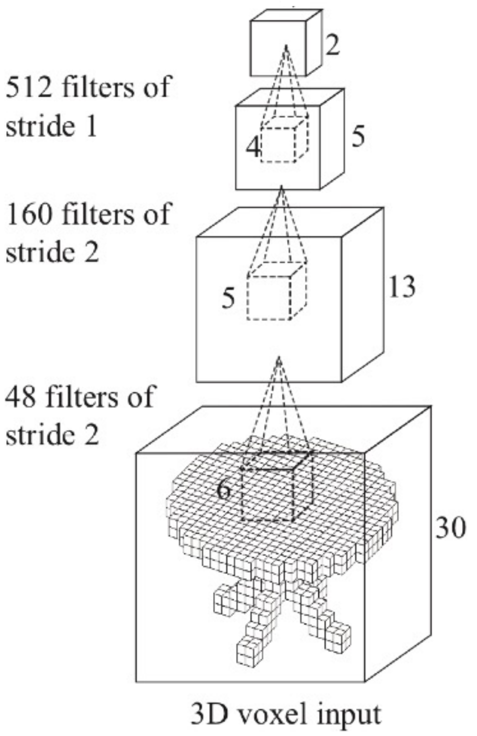


PointNet learns to pick perceptually interesting points!

From PointNet to PointNet++

Limitations of PointNet

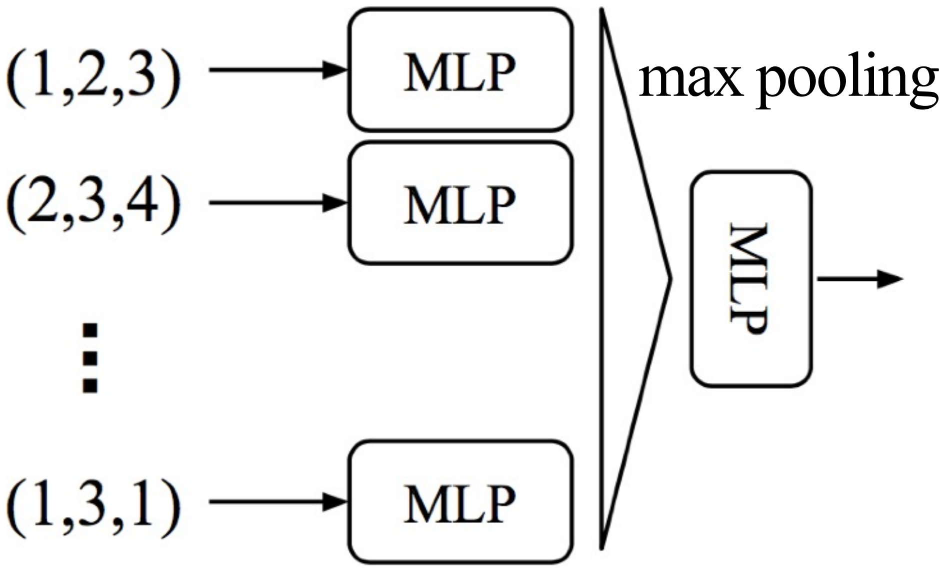
Hierarchical feature learning
multiple levels of abstraction



3D CNN [Wu et al.2015]

v.s.

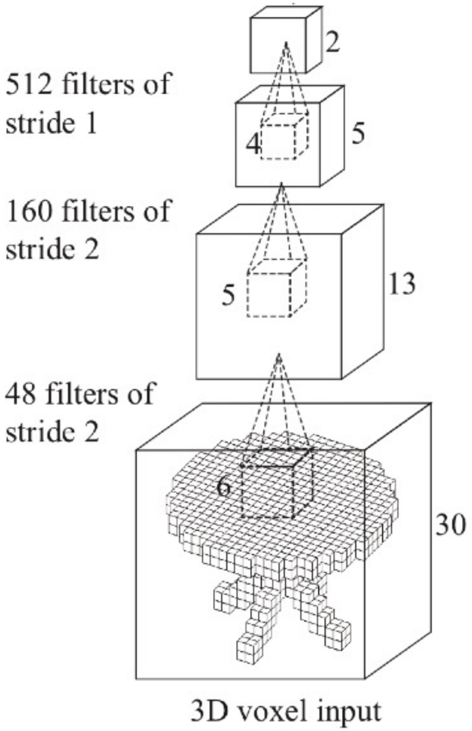
Global feature learning
either **one** point or **all** points



PointNet (vanilla) [Qi et al.2017]

Limitations of PointNet

Hierarchical feature learning
multiple levels of abstraction



3D CNN [Wu et al.2015]

v.s.

Global feature learning
either **one** point or **all** points

No local context

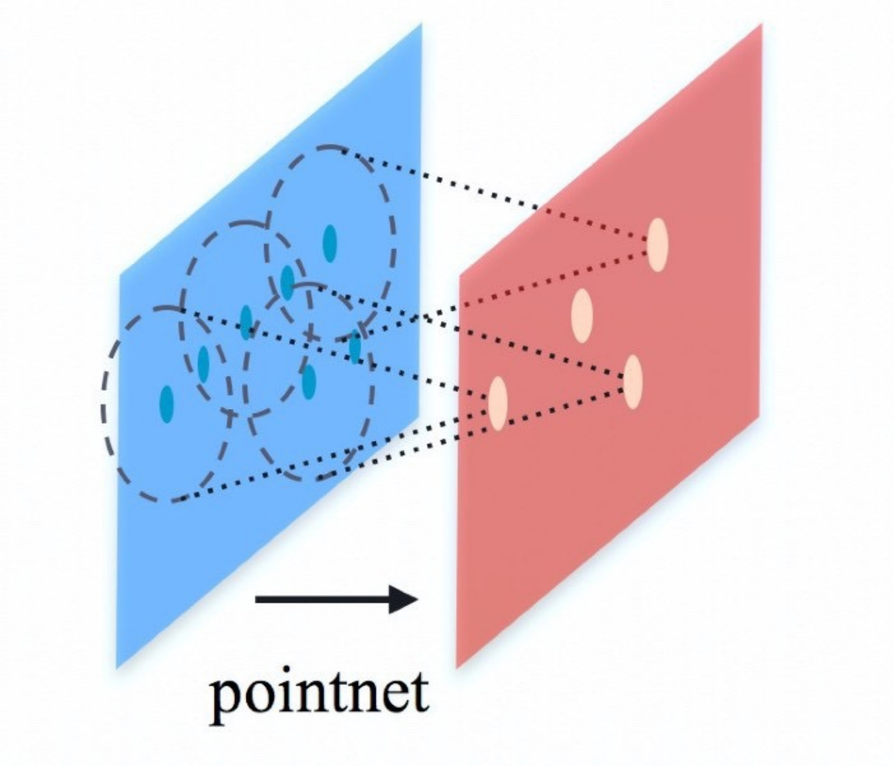
Limited translation invariance

PointNet (vanilla) [Qi et al.2017]

PointNet++

Basic idea: Recursively apply pointnet at local regions.

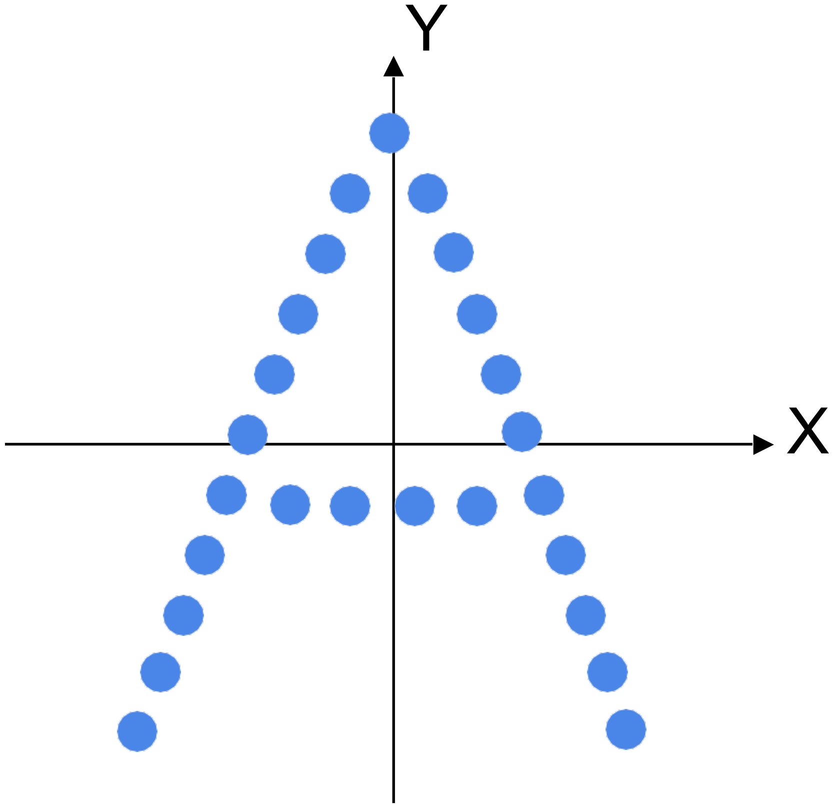
✓ Hierarchical feature learning
✓ Translation invariant
✓ Permutation invariant



*[2] **Charles R. Qi**, Li Yi, Hao Su, Leonidas Guibas. PointNet++: Deep Hierarchical Feature Learning on Point Sets in a Metric Space (NIPS'17)*

PointNet++

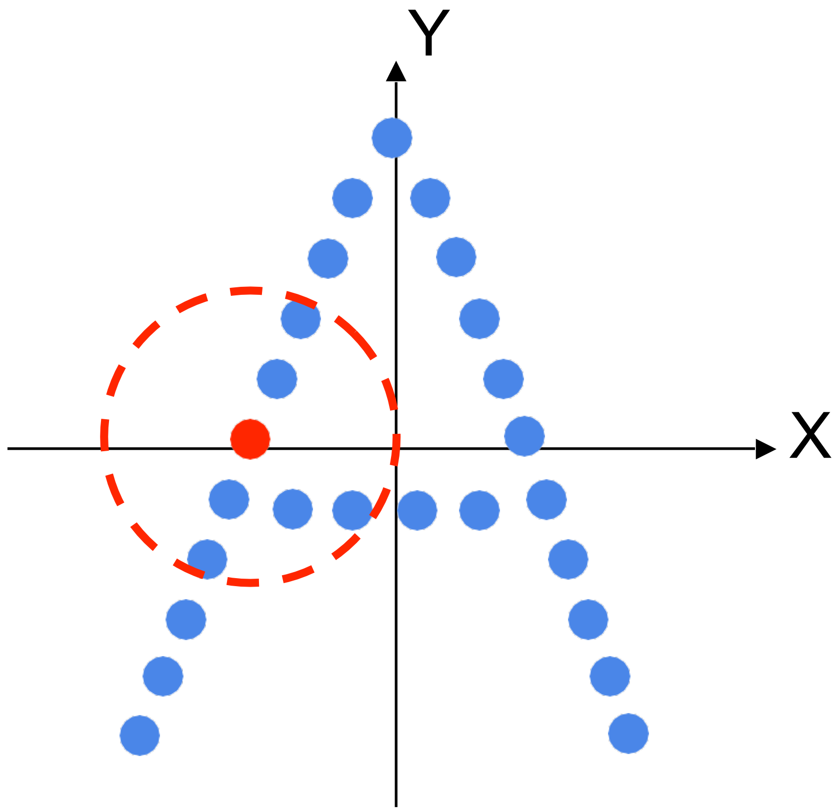
Hierarchical Point Feature Learning



N points in (X,Y)

PointNet++

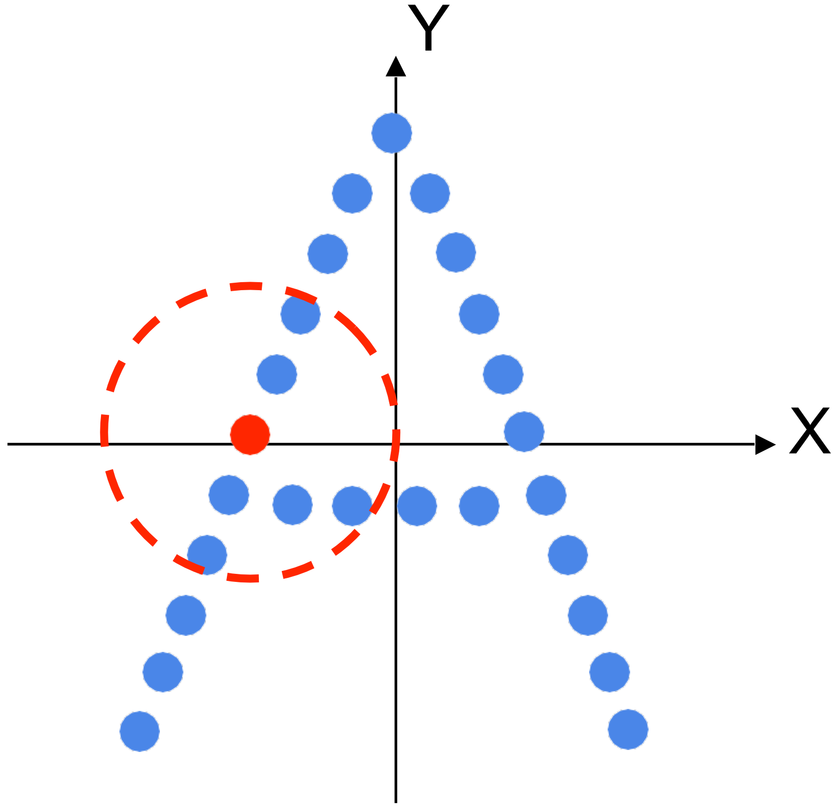
Hierarchical Point Feature Learning



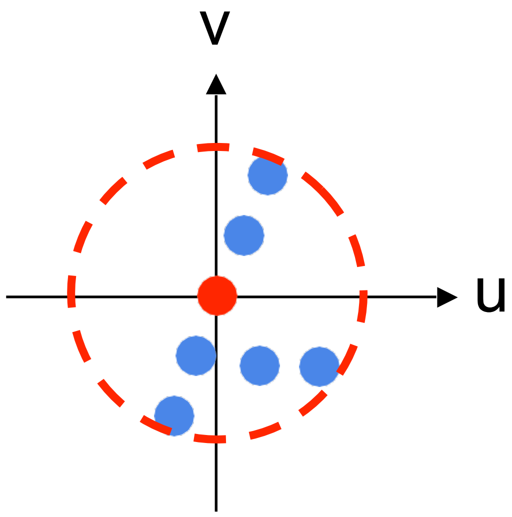
N points in (X,Y)

PointNet++

Hierarchical Point Feature Learning



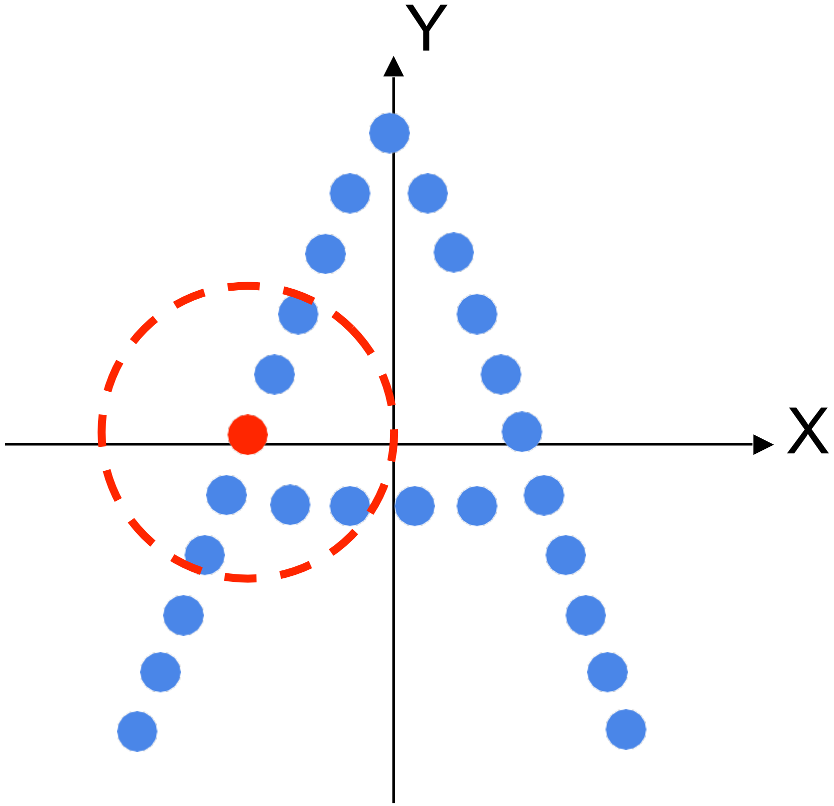
N points in (X,Y)



k points in local coordinates (u,v)

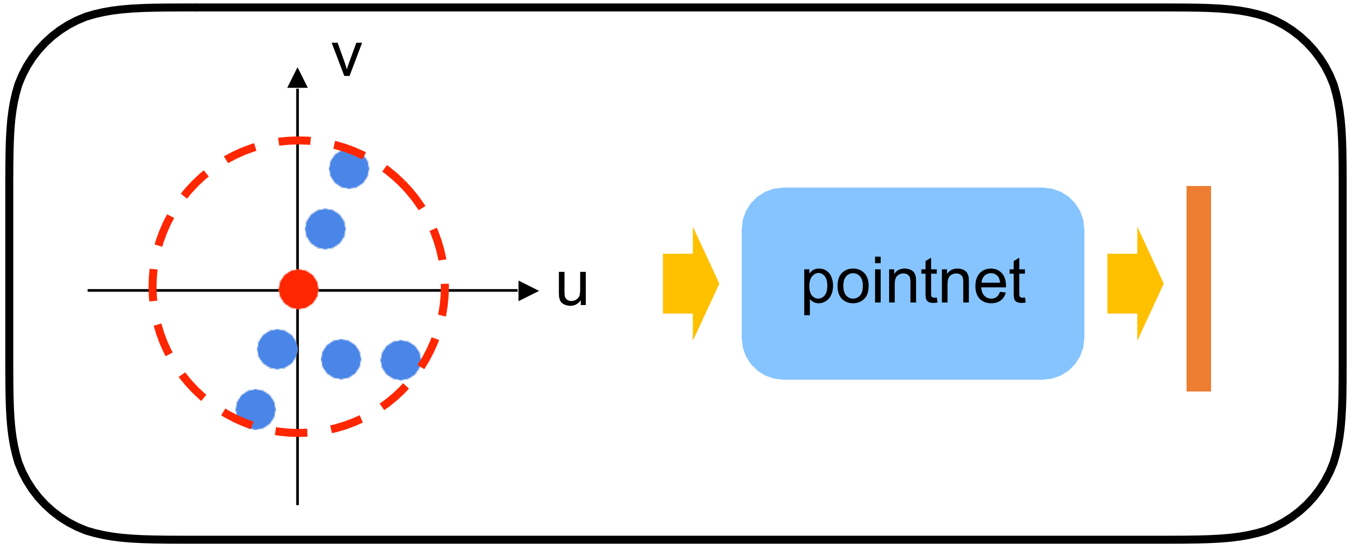
PointNet++

Hierarchical Point Feature Learning



N points in (X,Y)

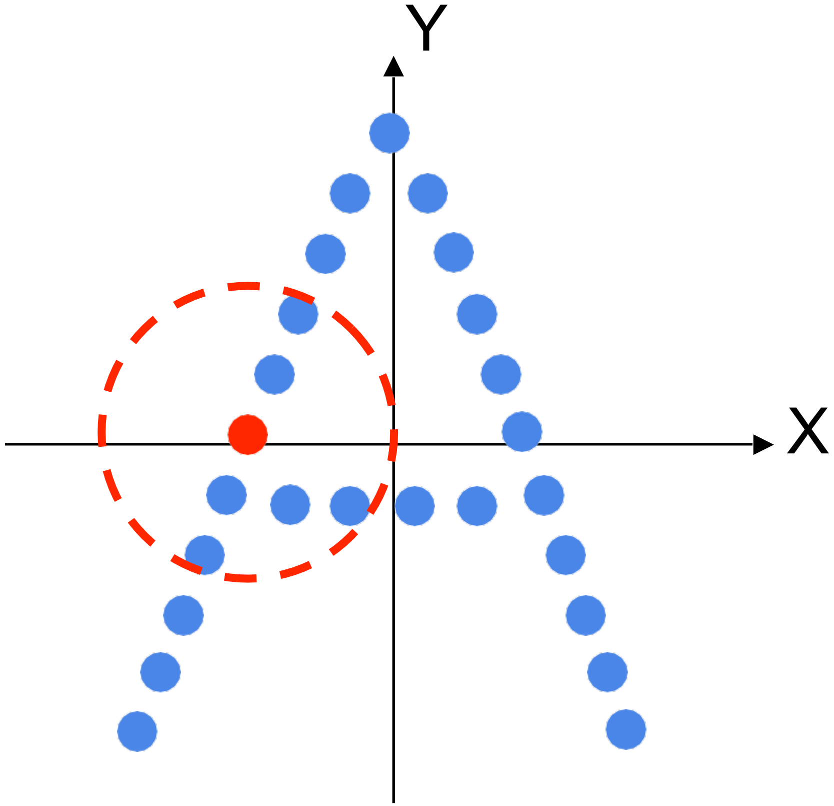
Apply pointnet at a local region



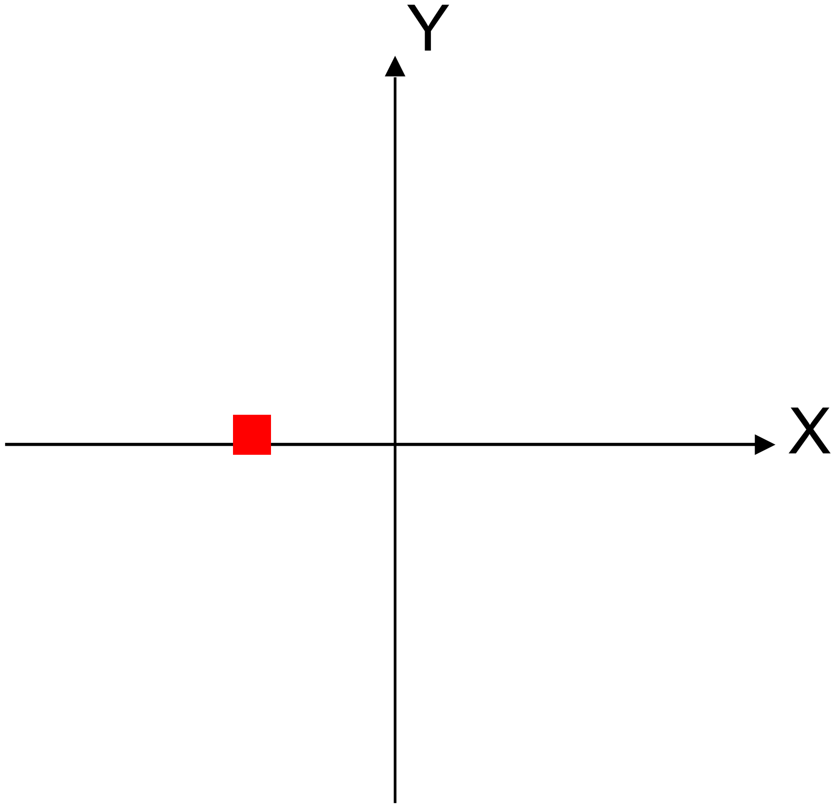
k points in local coordinates (u,v)

PointNet++

Hierarchical Point Feature Learning



N points in (X,Y)



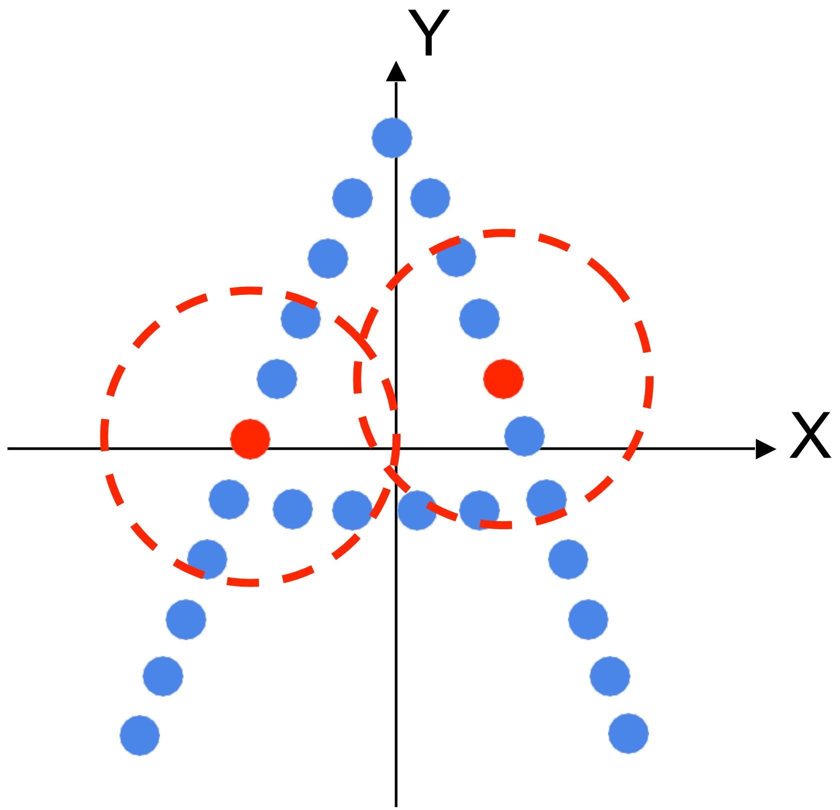
points in (X,Y, **F**)

Euclidean space

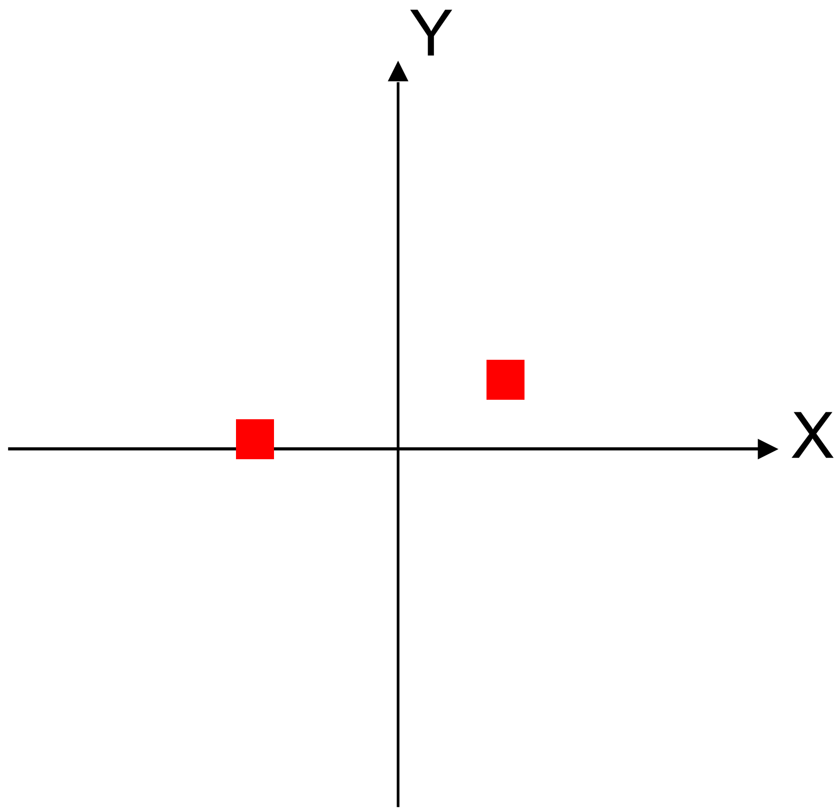
high-dim feature space

PointNet++

Hierarchical Point Feature Learning



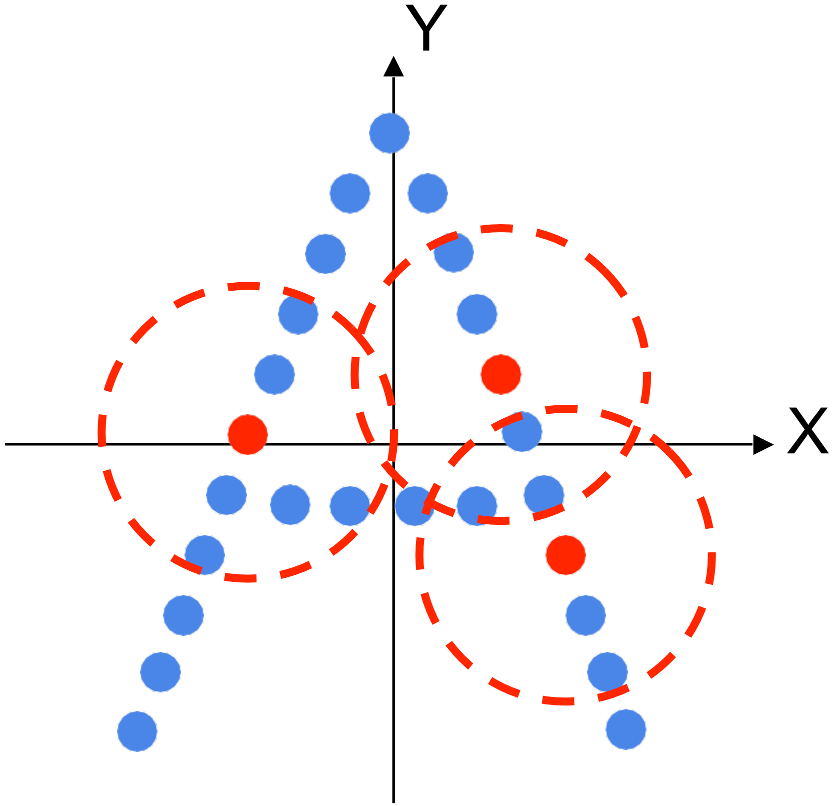
N points in (X,Y)



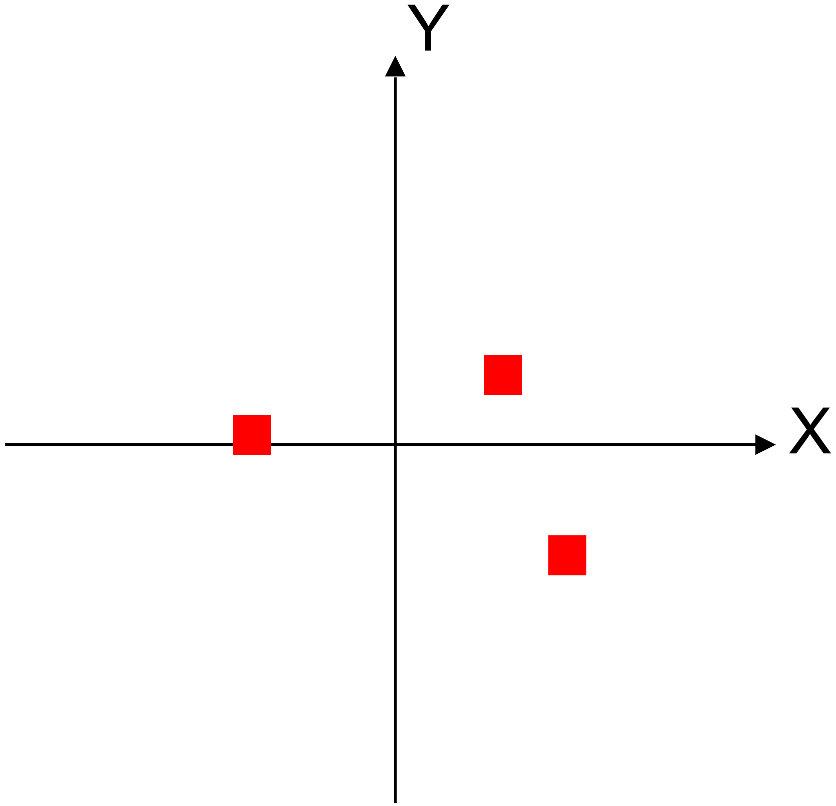
points in (X,Y, **F**)

PointNet++

Hierarchical Point Feature Learning

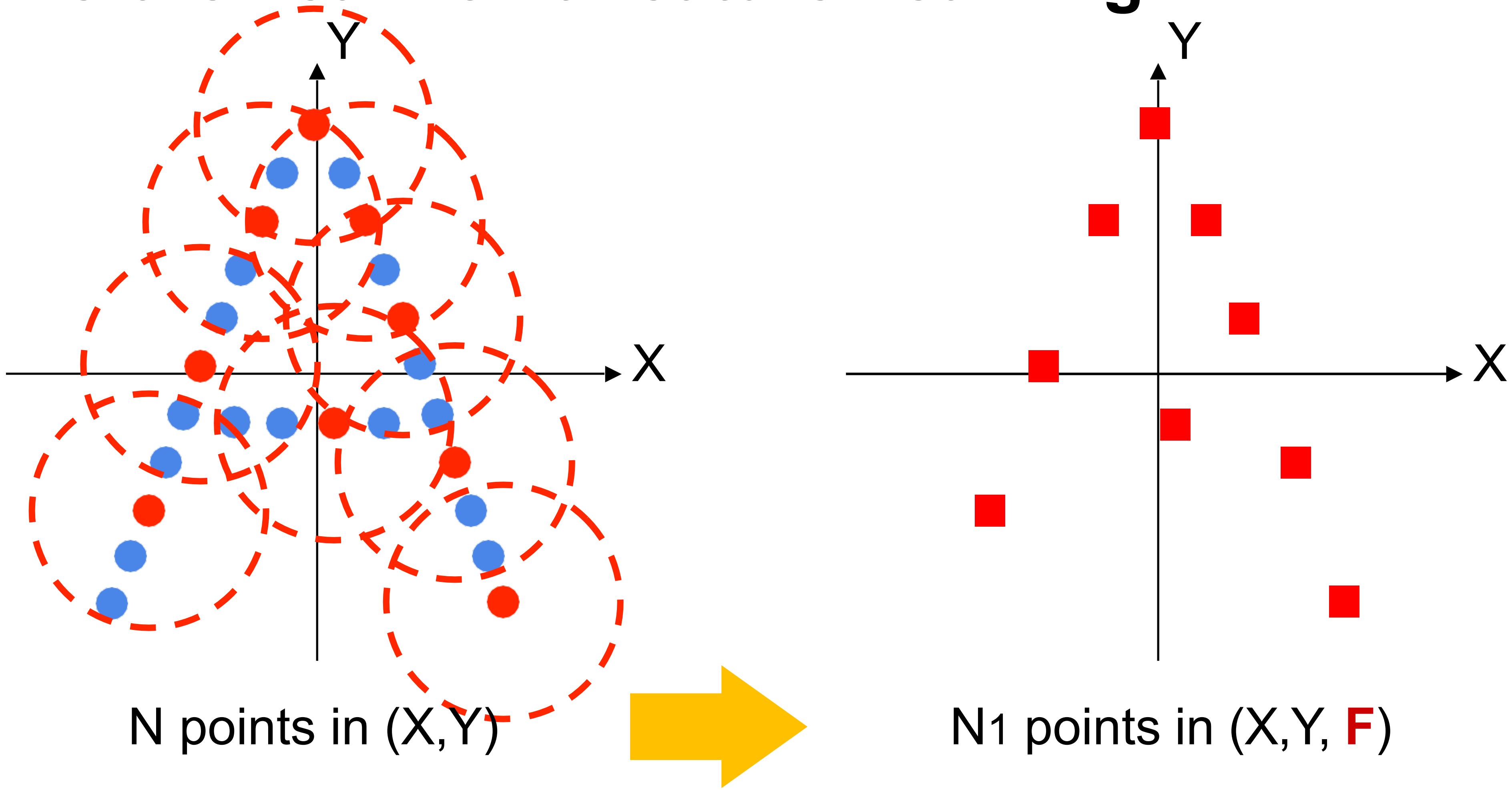


N points in (X,Y)



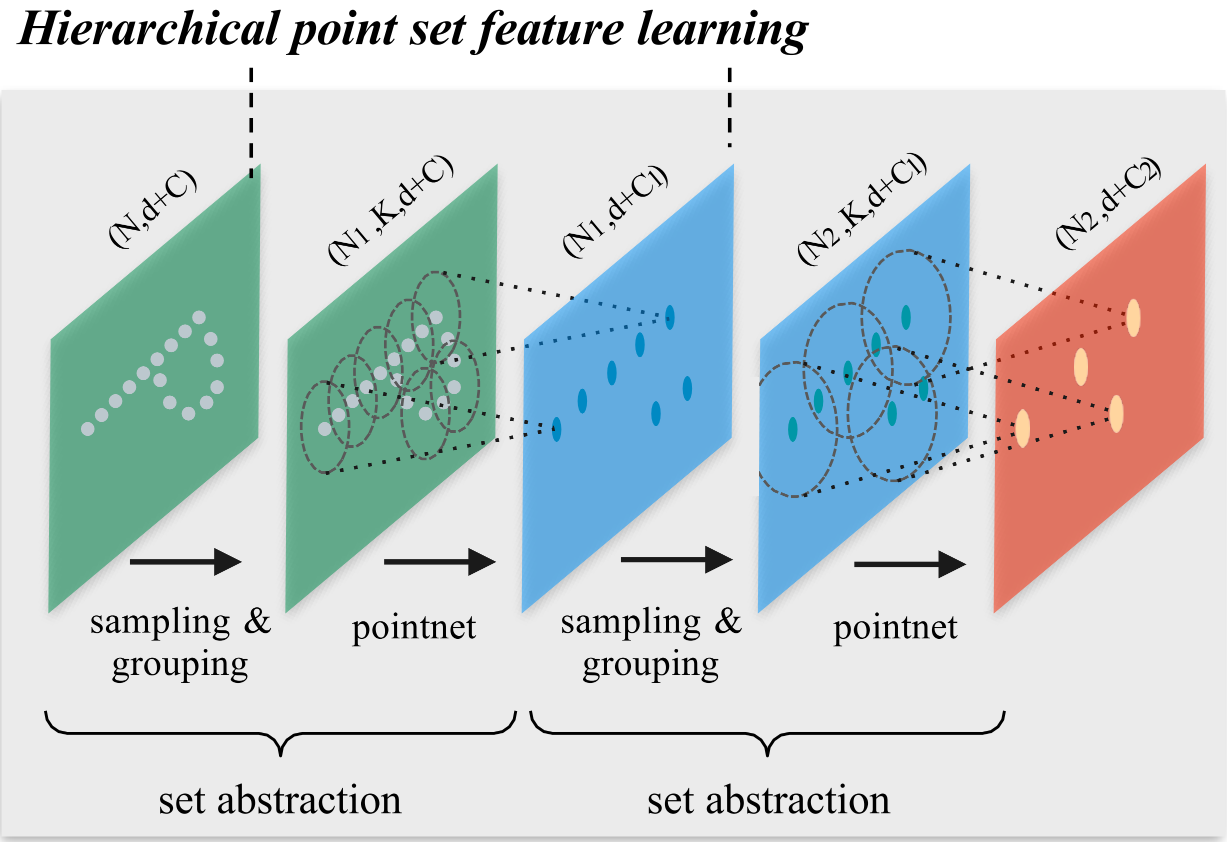
points in (X,Y, **F**)

PointNet++
Hierarchical Point Feature Learning

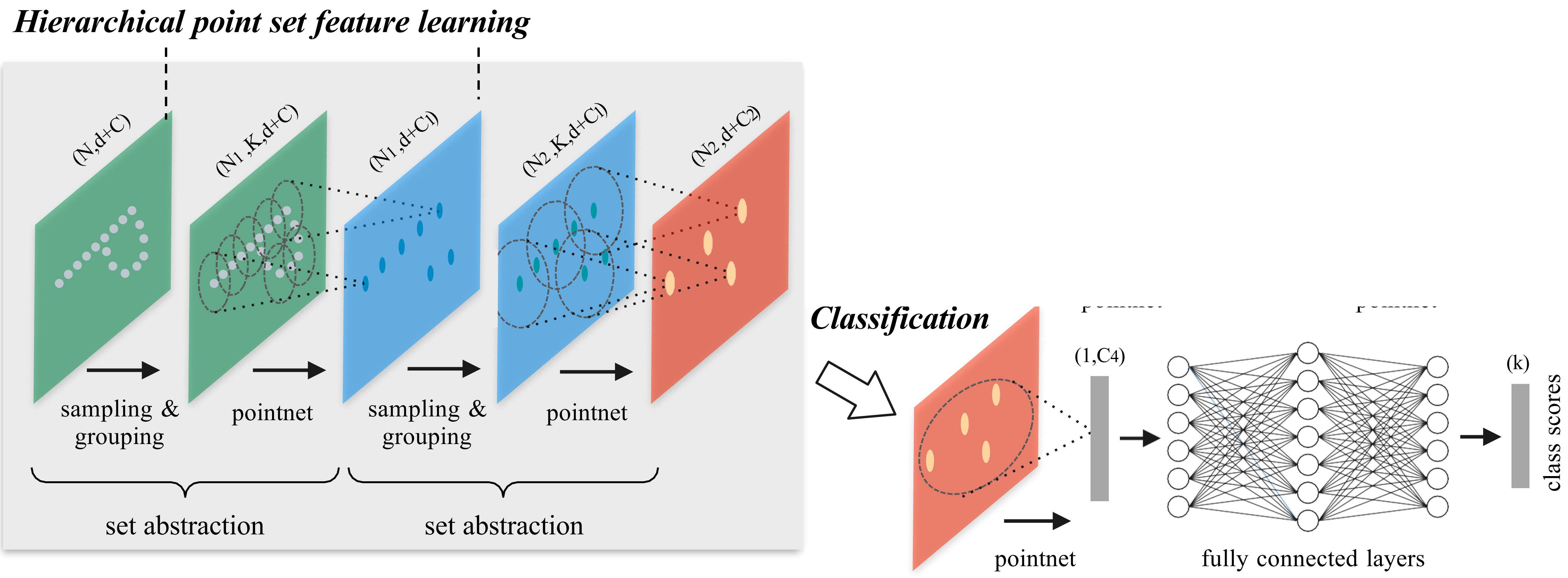


Set Abstraction: farthest point sampling + grouping + pointnet

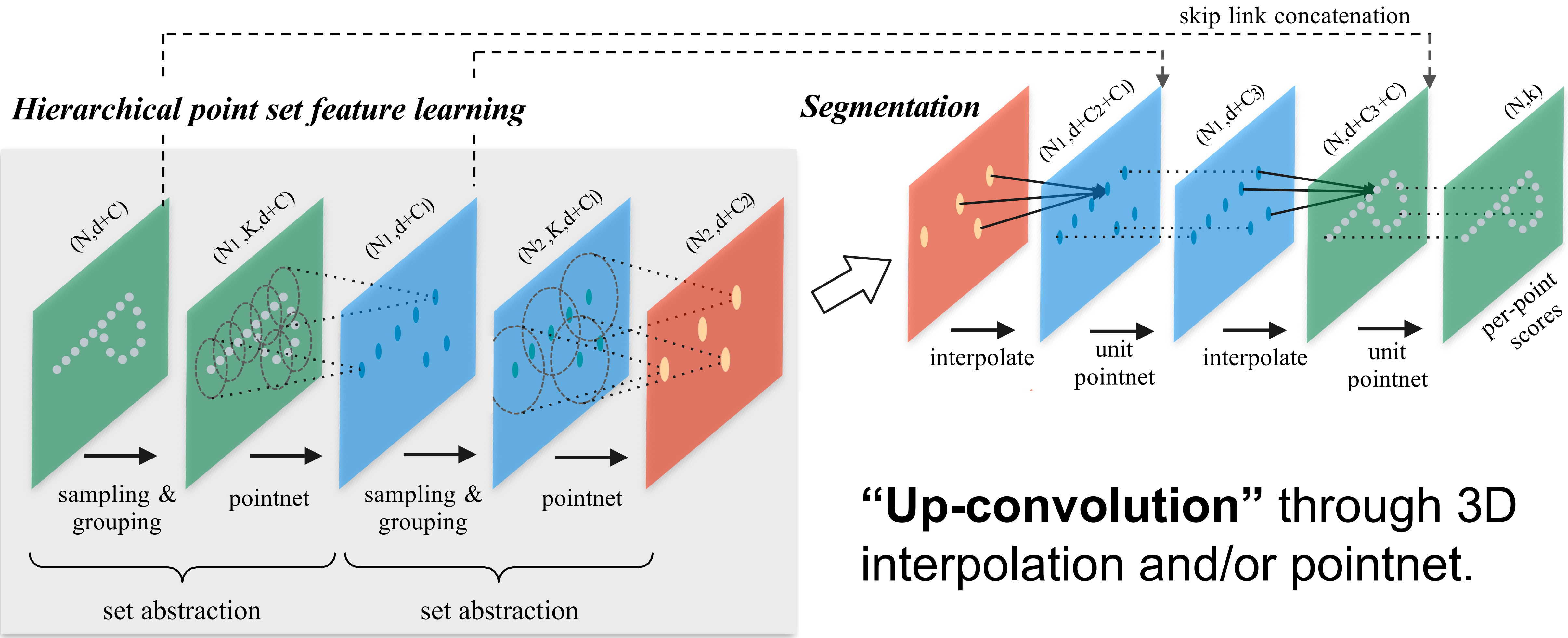
PointNet++ for Classification and Segmentation



PointNet++ for Classification and Segmentation



PointNet++ for Classification and Segmentation



"Up-convolution" through 3D interpolation and/or pointnet.

PointNet++

Kernel Size Design in a Hierarchical Network

- In CNN, small kernels are usually better

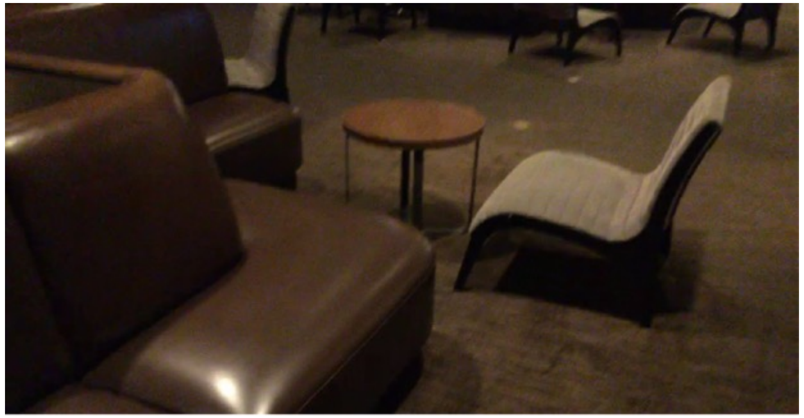
Karen Simonyan & Andrew Zisserman, *Very Deep Convolutional Networks for Large-scale Image Recognition*, ICLR 2015

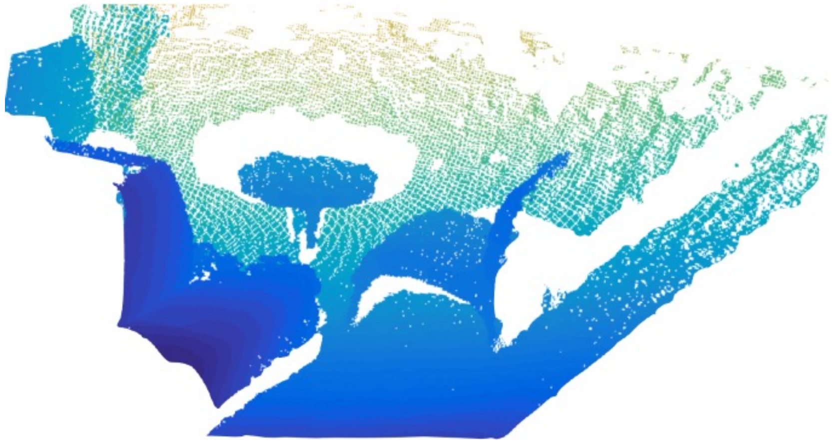
- Is it also true for point cloud learning?

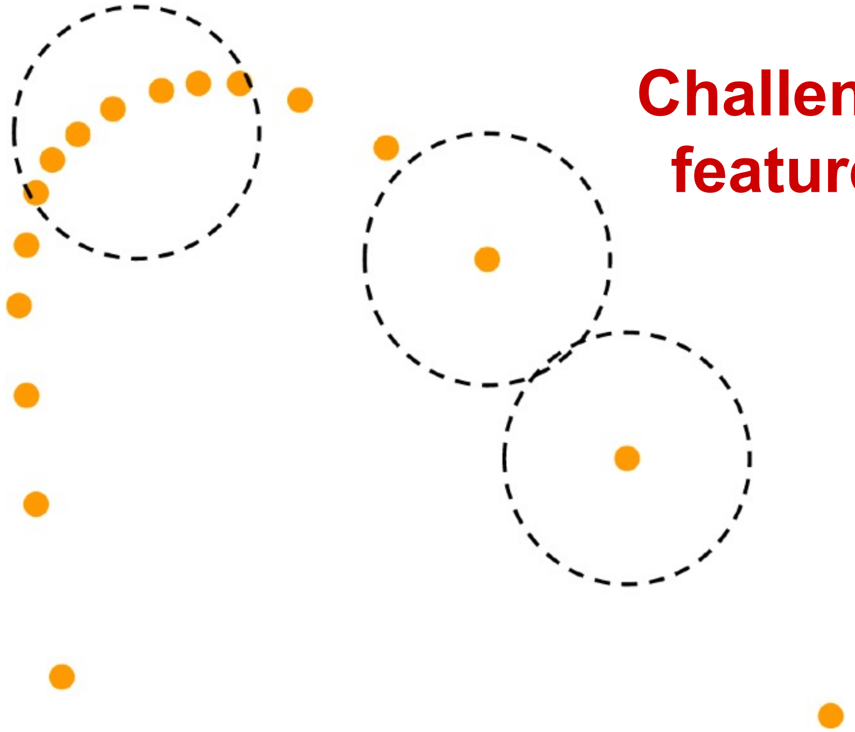
Non-uniform Sampling Density in Point Clouds

Density variation is a common issue in 3D point cloud processing

- perspective effect, radial density variation, motion etc.



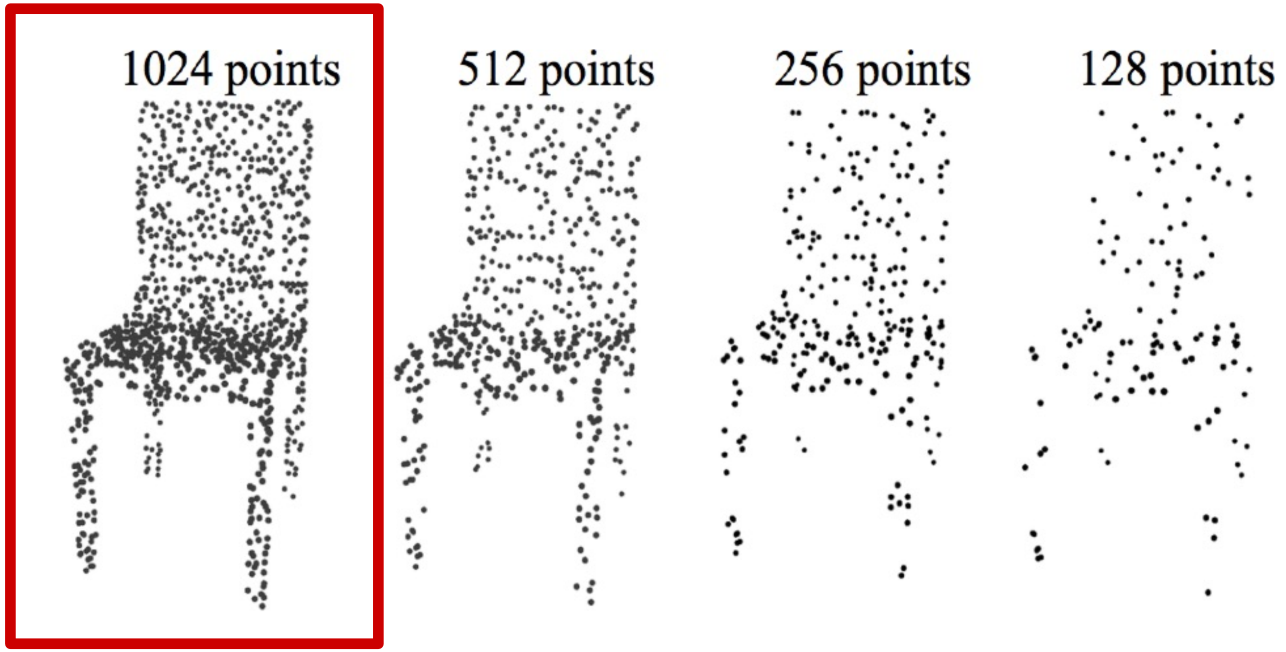


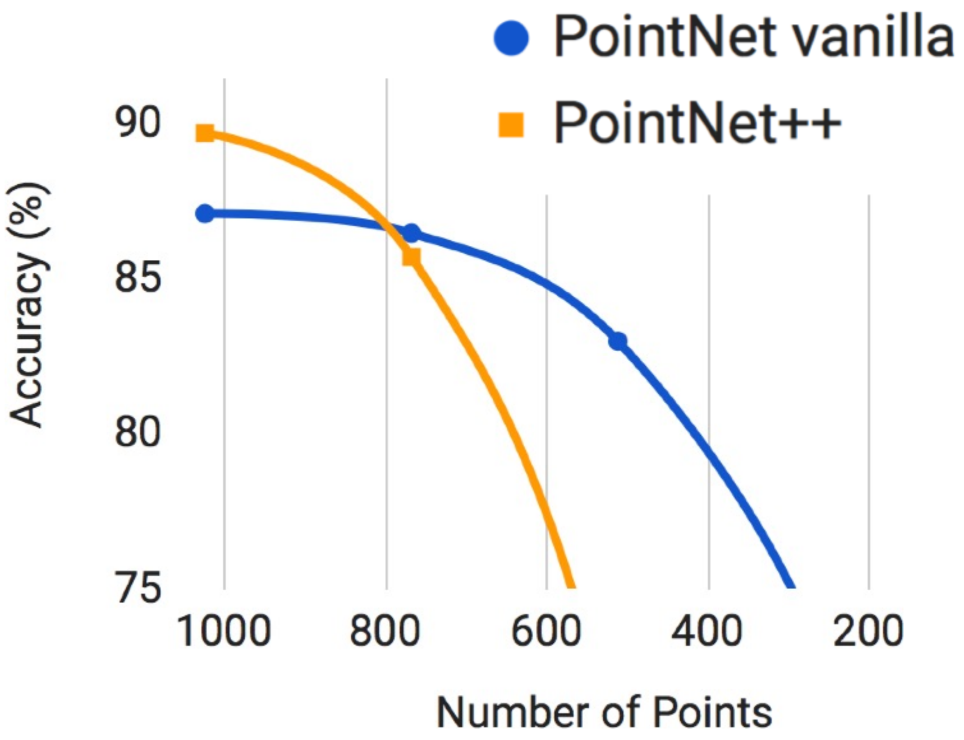


Challenge for local feature learning!

PointNet++

Density Variation Affects Hierarchy

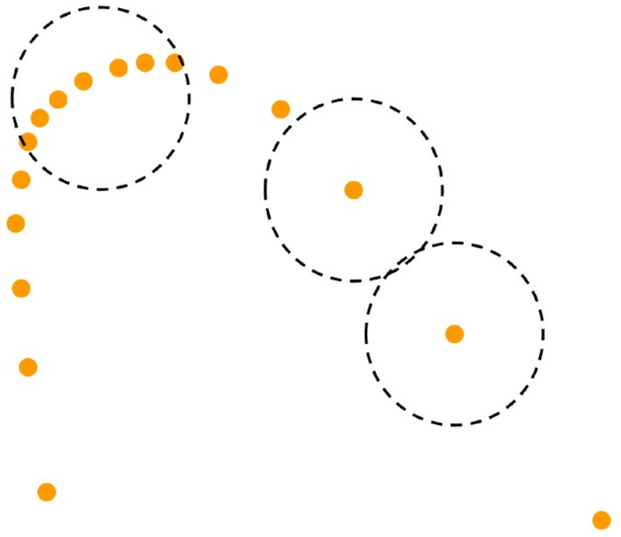


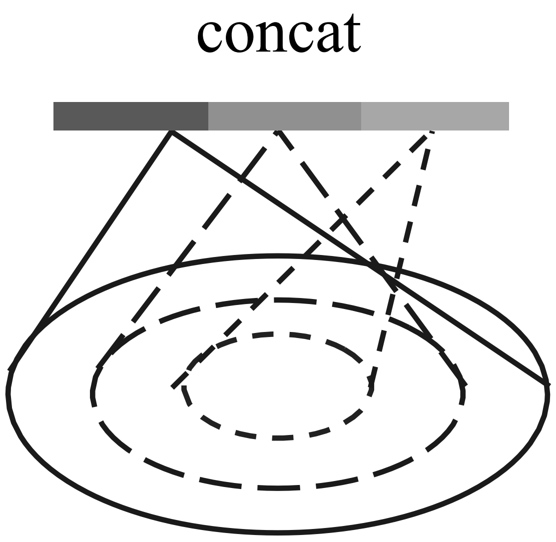


Small kernels suffer from varying densities!

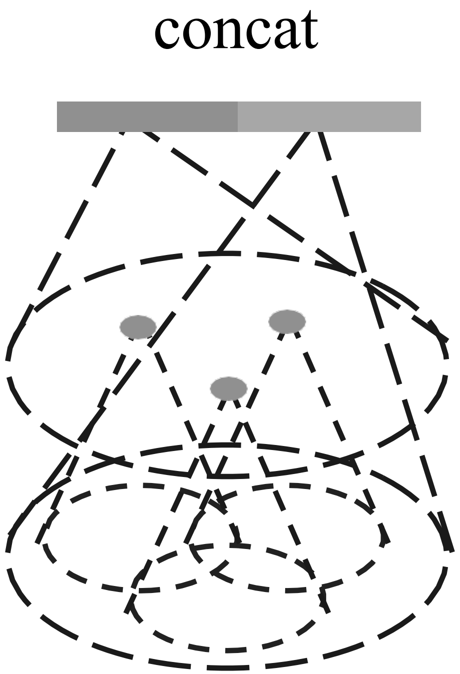
PointNet++

Robust learning under varying sampling density





(a)



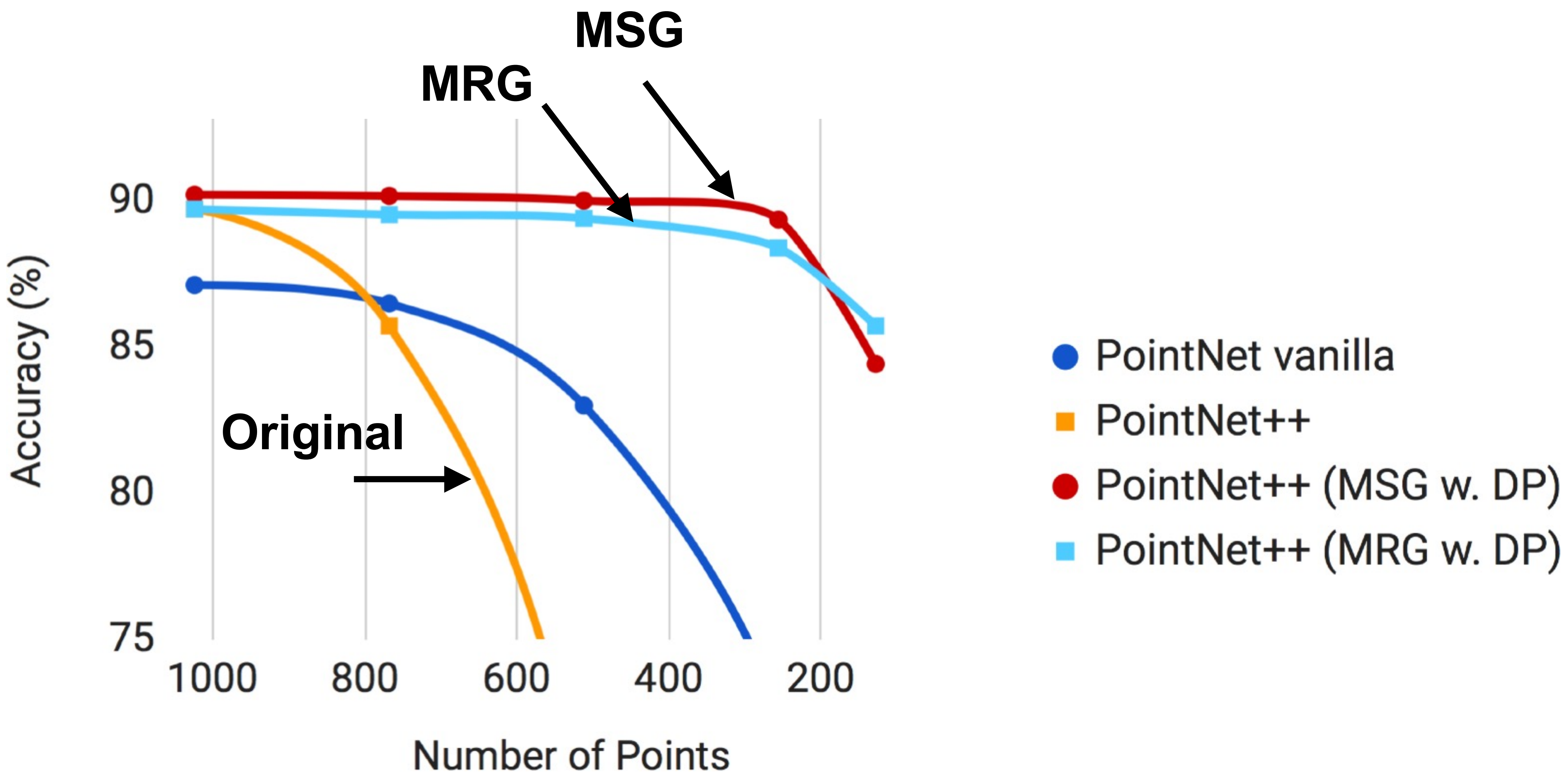
(b)

Multi-scale grouping (MSG) Multi-res grouping (MRG)

During Training: input point dropout with random dropout ratio

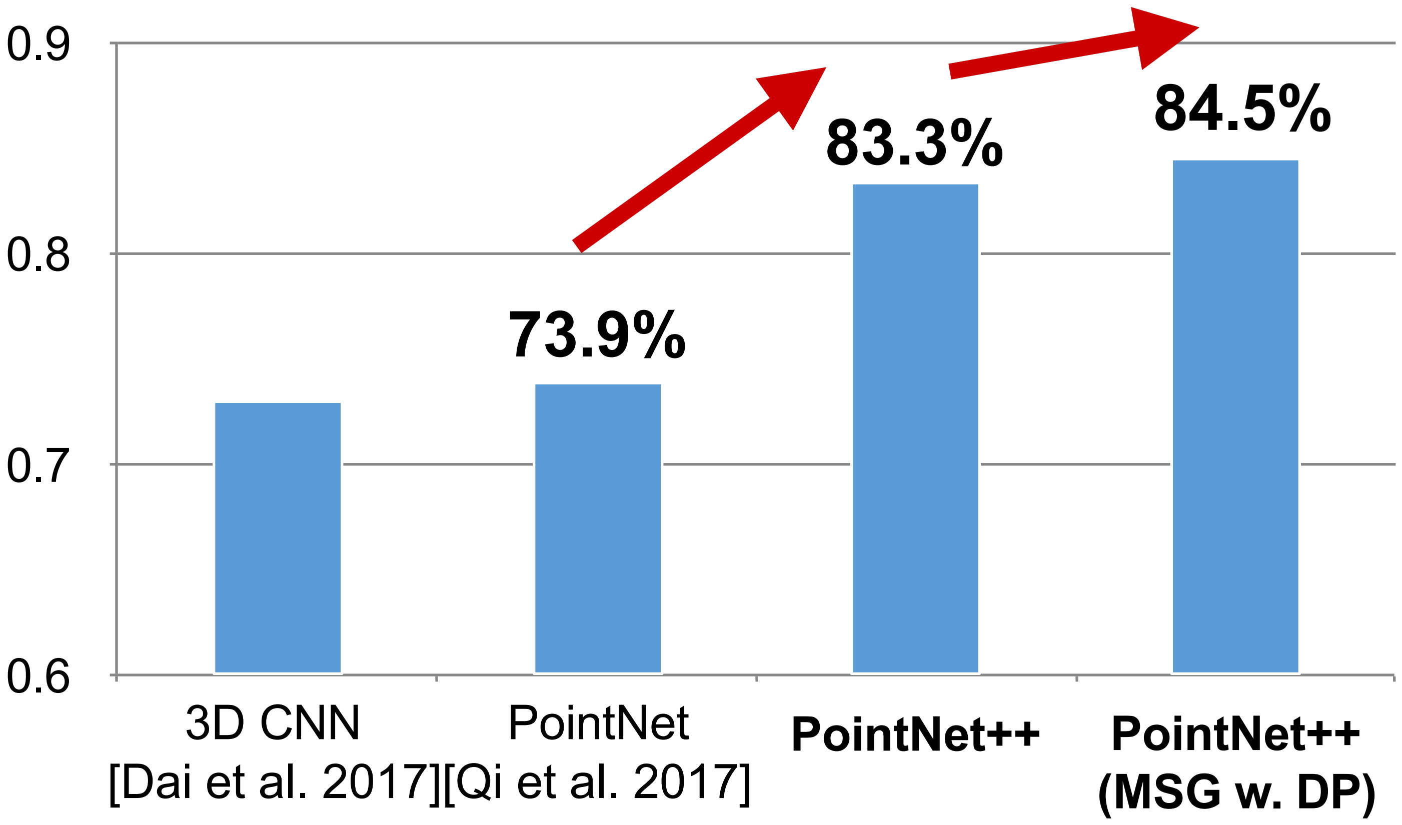
PointNet++

Robust learning under varying sampling density



PointNet++ Results: Scene Parsing

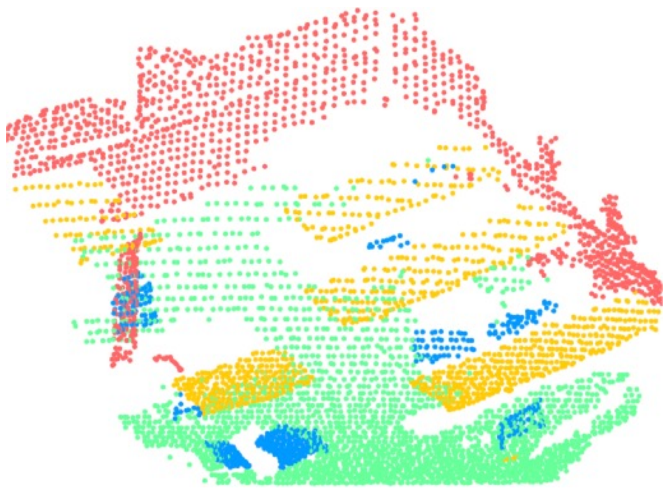
Better accuracy with hierarchical learning.

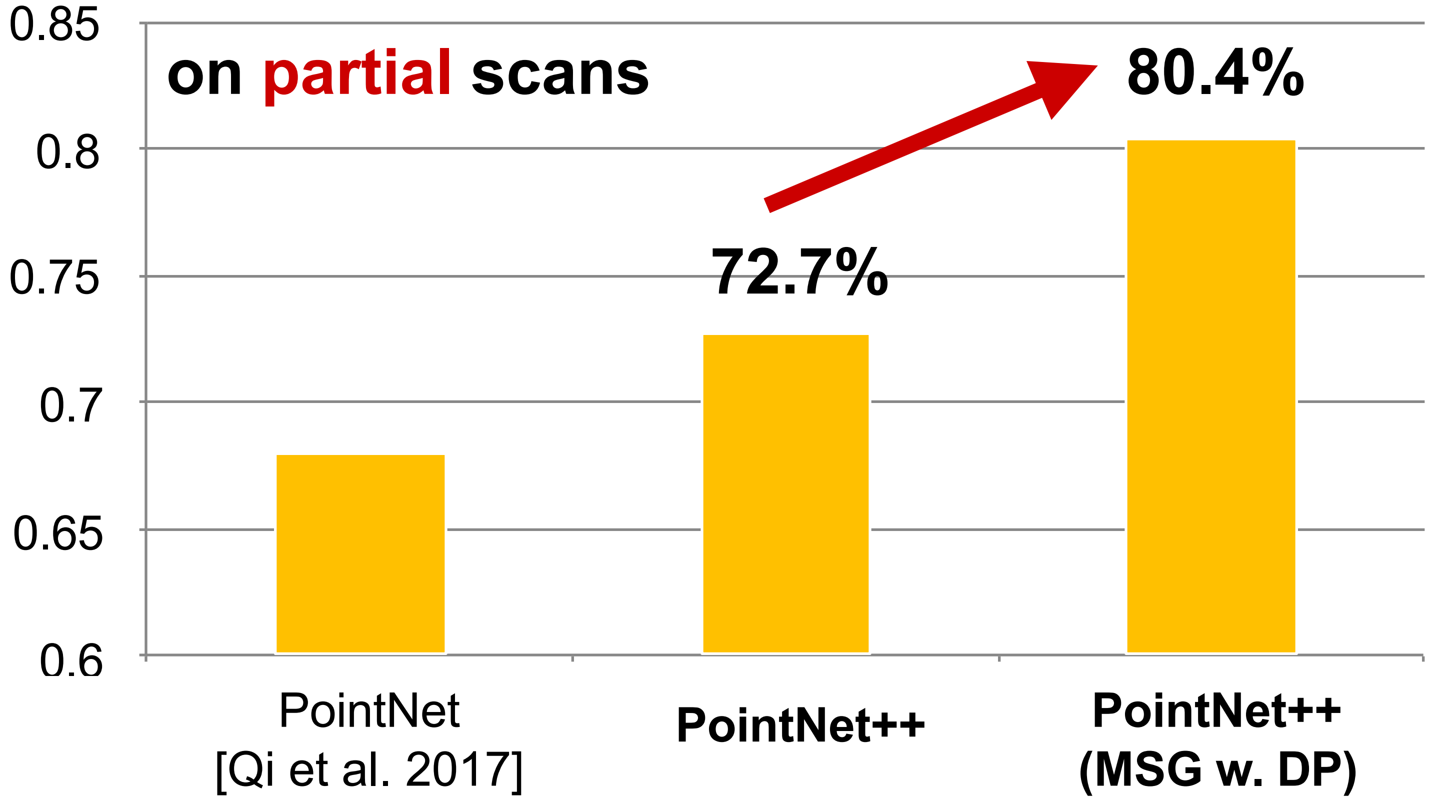


dataset: ScanNet; metric: per-point semantic classification accuracy (%)

PointNet++ Results: Scene Parsing

Robust layers for non-uniform densities (MSG) helps a lot.





dataset: ScanNet; metric: per-point semantic classification accuracy (%)

PointNet++ Results: Non-Euclidean Space

For organic shape recognition, PointNet++ can generalize to non-Euclidean space

- ❖ intrinsic point features (HKS, WKS, Gaussian curvature)
- ❖ intrinsic distance metric (geodesic)

(a) Horse (b) Cat (c) Horse

	Metric space	Input feature	Accuracy (%)
DeepGM [13]	-	Intrinsic features	93.03
Ours	Euclidean	XYZ	60.18
	Euclidean	Intrinsic features	94.49
	Non-Euclidean	Intrinsic features	**96.09**

dataset: SHREC15; metric: shape classification accuracy (%)