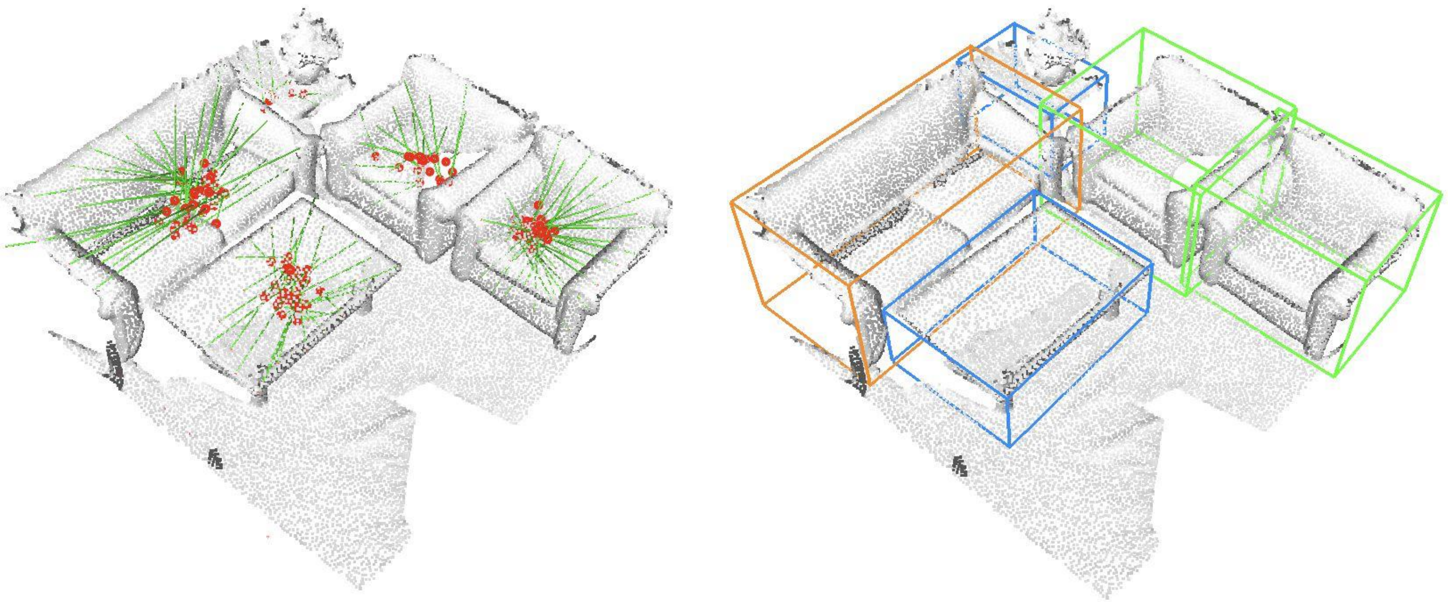
Course Timetable

		Niloy	Iasonas	Paul	Nils	Leo
Introduction	9:00	X				
Neural Network Basics	~9:15		X			
Supervised Learning in CG	~9:50	X				
Unsupervised Learning in CG	~10:20			X		
Learning on Unstructured Data	~10:55					X
Learning for Simulation/Animation	~11:35				X	
Discussion	12:05	X	X	X	X	X

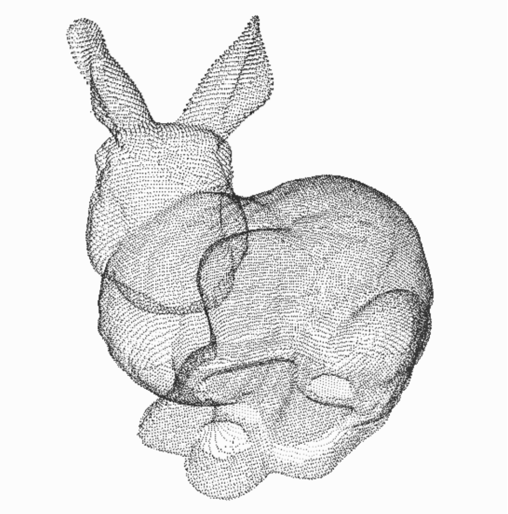
Deep Learning for Point Cloud Data



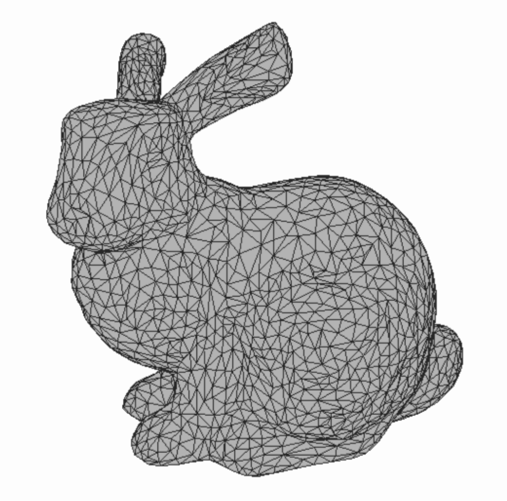
Leonidas Guibas
Stanford University
Facebook AI Research

Leonidas Guibas Laboratory
Geometric Computing

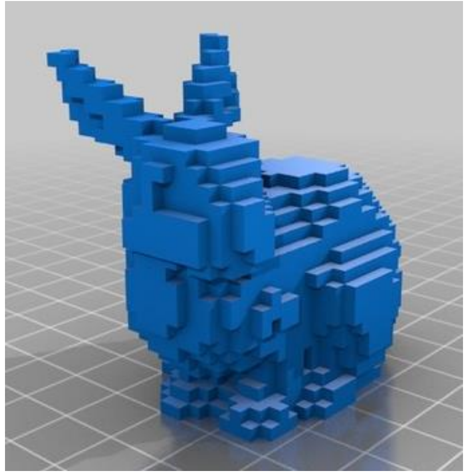
Multiple 3D Representations



Point Cloud



Surface Mesh



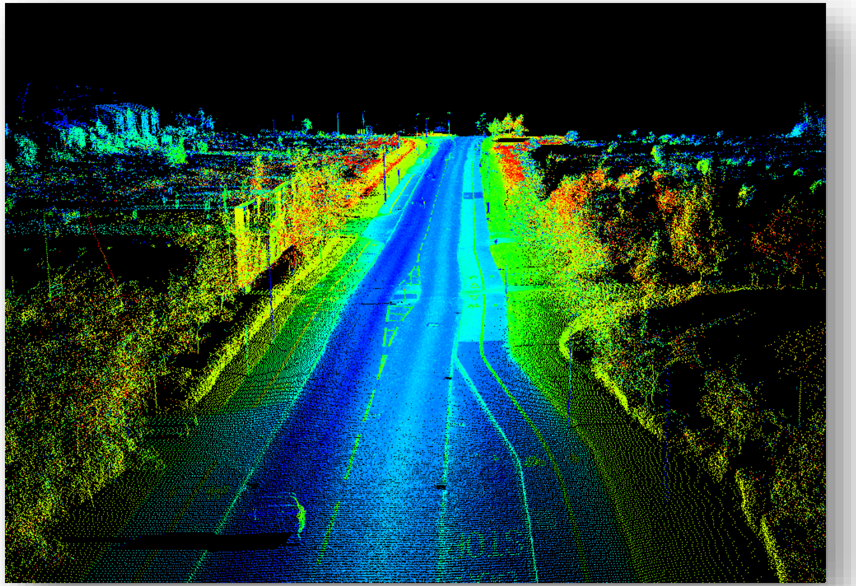
Volumetric

Multi-View
Images

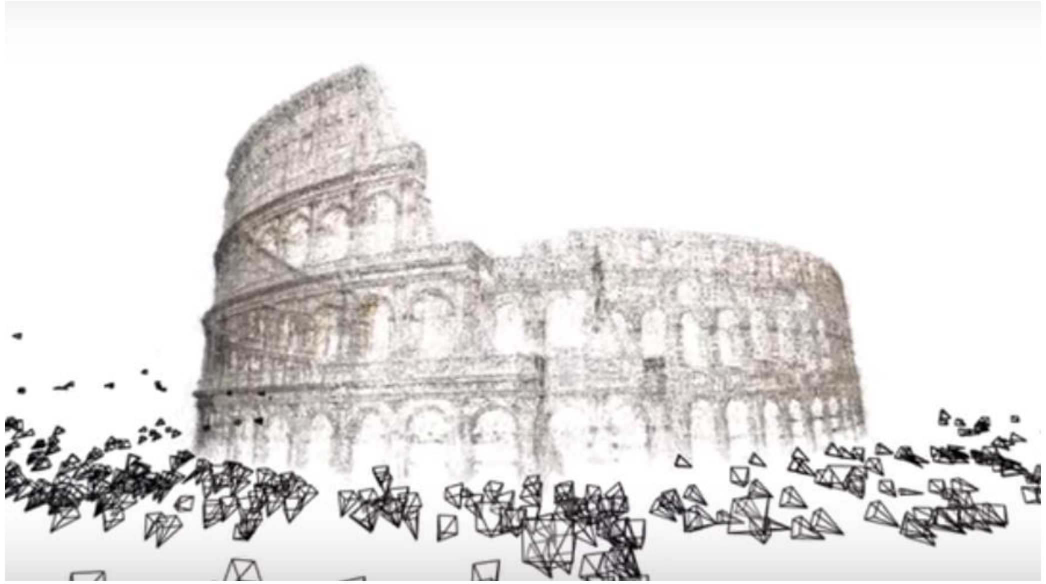
RGB(D)

…

Point Clouds



Lidar point clouds (LizardTech)



Structure from motion (Microsoft)

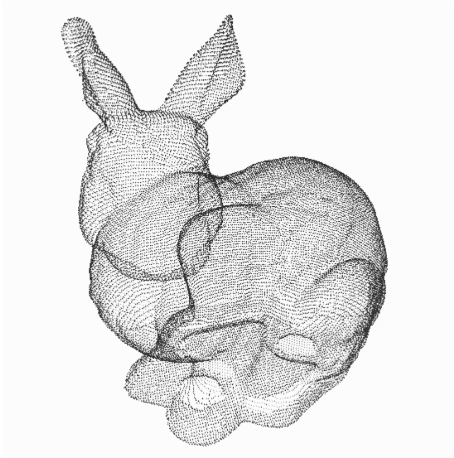
Depth camera (Intel)

A Common 3D Representation: Point Cloud

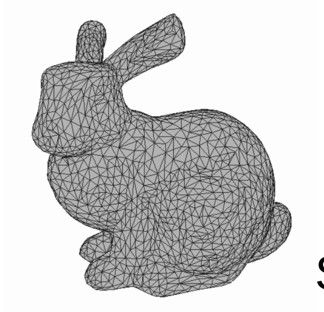
- **Point clouds are close to raw sensor data**
- **Point clouds are representationally simple**

LiDAR

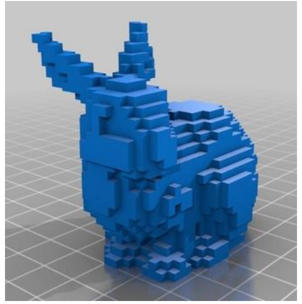
Depth Sensor



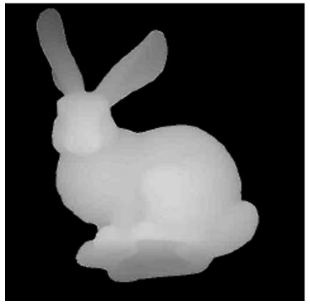
Point Cloud



Surface Mesh



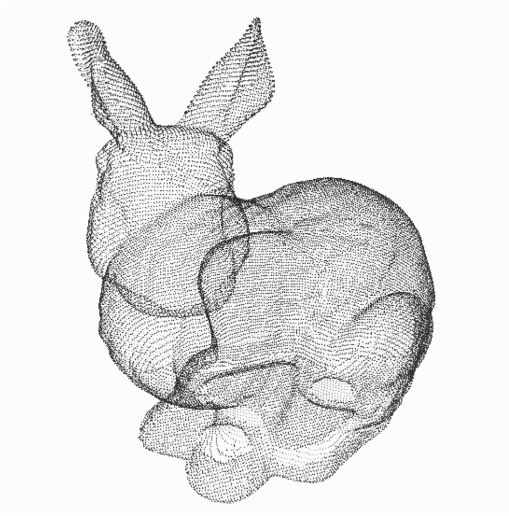
Volumetric



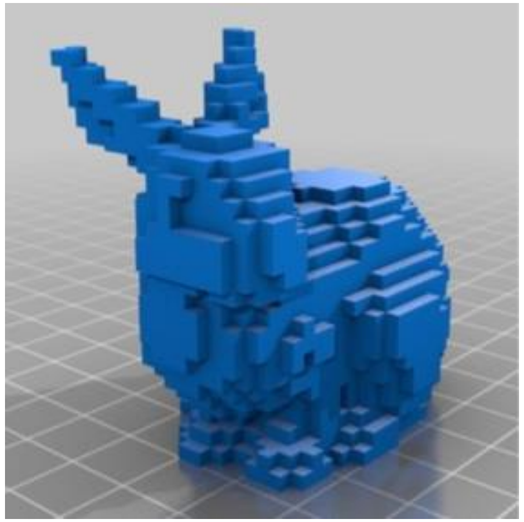
Depth Map

Early Work on 3D Learning

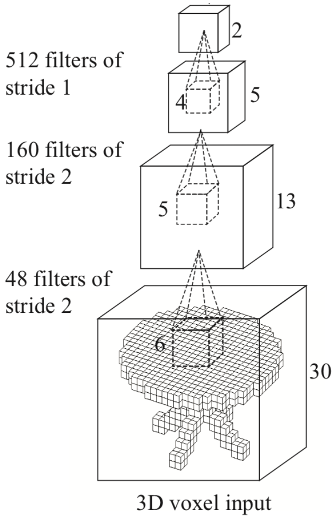
Point clouds were **converted to other regular representations** before input to a deep neural network



voxelization →



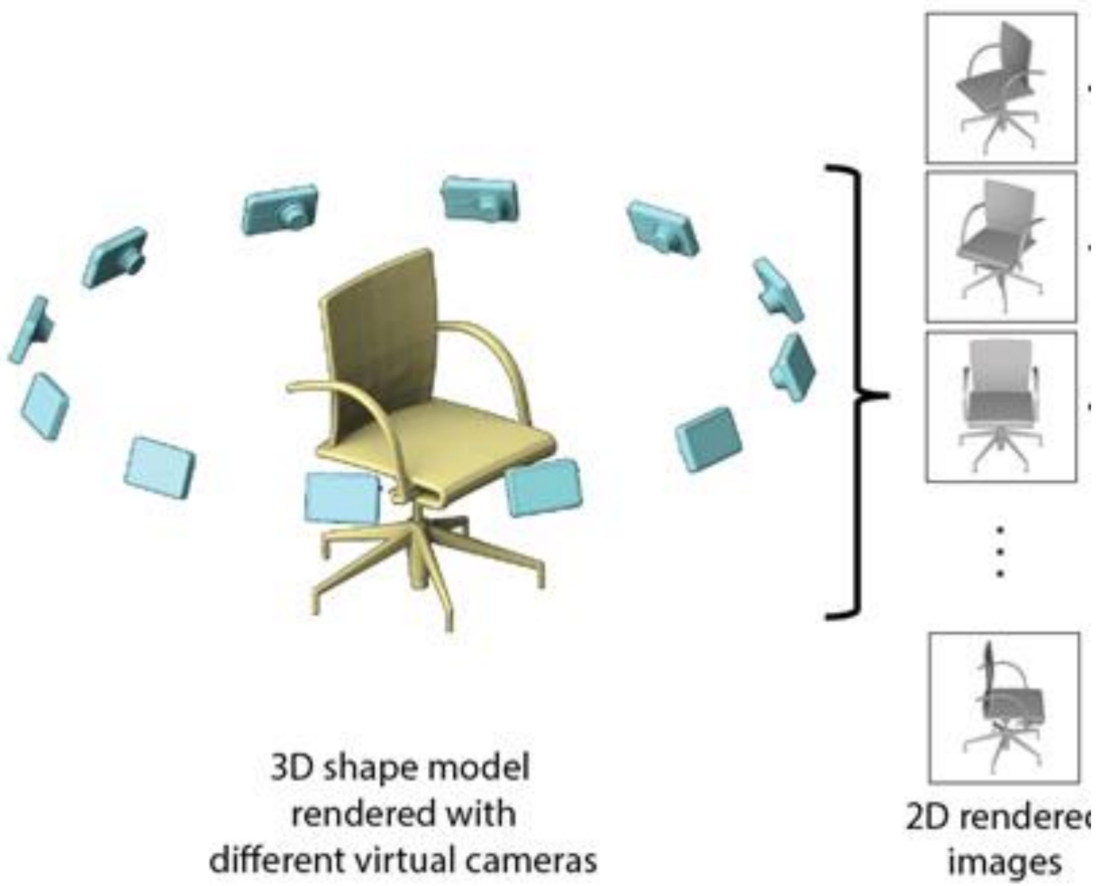
+



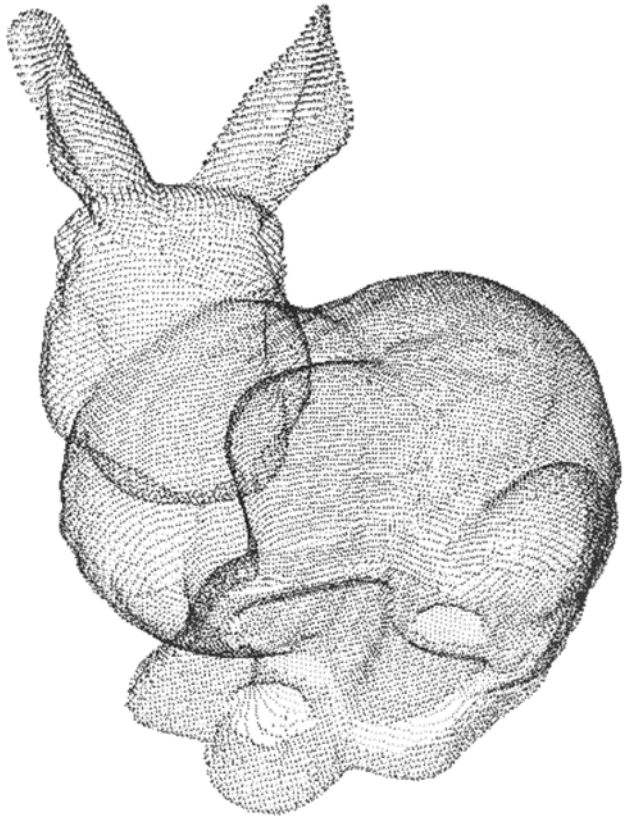
Con: High space & time complexity -- 3D convolution $O(N^3)$
Quantization errors in voxelization

Earlier Work

Point clouds were **converted to other regular representations** before input to a deep neural network



Multiview Images



Research Question:

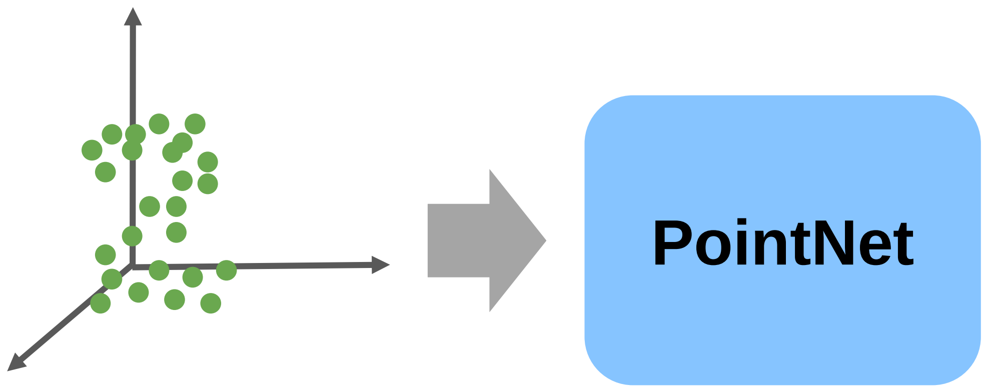
Can we achieve effective **feature learning directly on irregular point clouds**?

Talk Ouline

- Survey of PointNet, PointNet++ architectures (~2017)
- Since the original PointNet work, an explosion of activity in this area -- very brief survey
- Applications to outdoor and indoor object detection and navigation, point cloud synthesis

PointNet Architeture Review

End-to-end learning for irregular point data

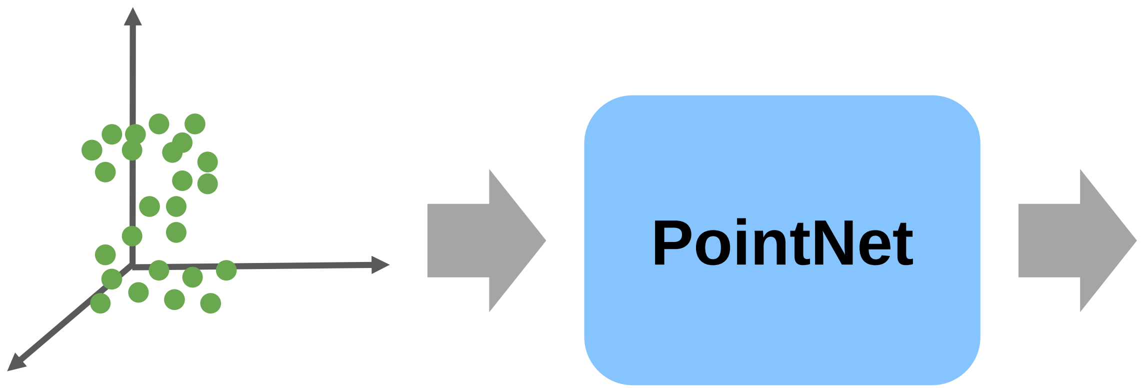


Charles R. Qi, Hao Su, Kaichun Mo, Leonidas J. Guibas. PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation. (CVPR'17)

PointNet Architecture Review

End-to-end learning for irregular point data

Unified framework for various tasks



Object Classification

Object Part Segmentation

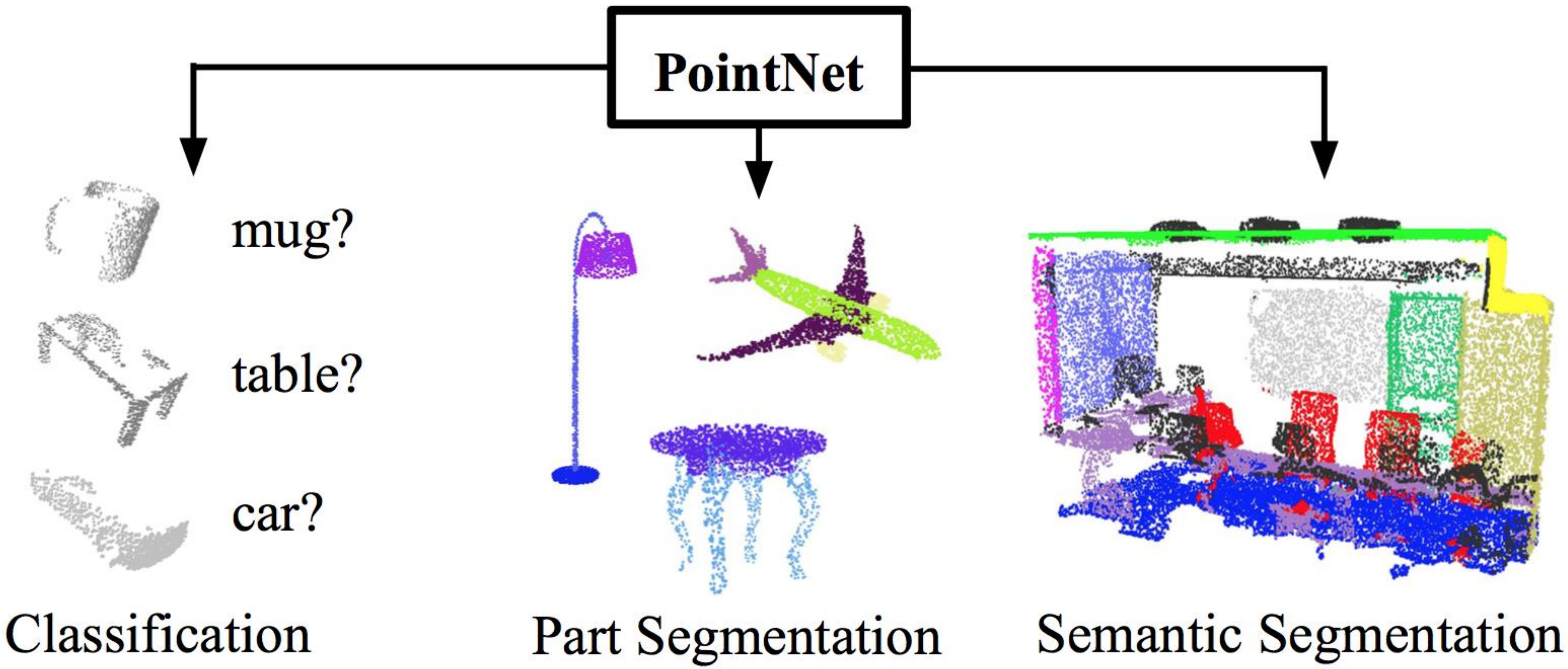
Semantic Scene Parsing

...

PointNet Architecture Review

End-to-end learning for irregular point data

Unified framework for various tasks



Challenges

The model has to respect key properties of point clouds:

Point Permutation Invariance

Point cloud is a set of unordered points

Spatial Transformation Invariance

Point cloud rigid motions should not alter classification results

Challenges

The model has to respect key properties of point clouds:

Point Permutation Invariance

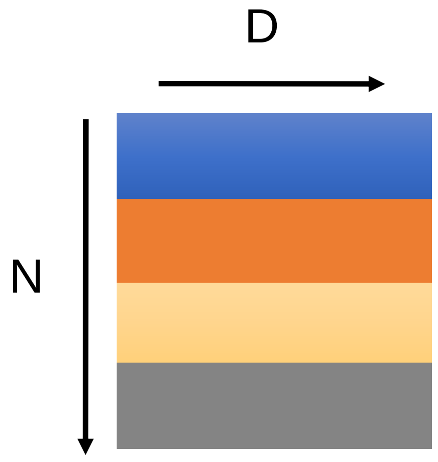
Point cloud is a set of unordered points

Spatial Transformation Invariance

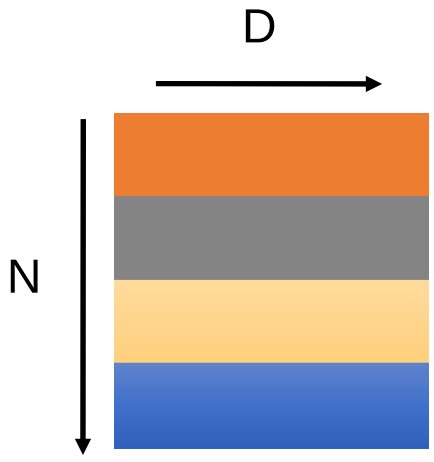
Point cloud rigid motions should not alter classification results

Unordered Input

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



represents the same **set** as



Model needs to be invariant to N! permutations

Permutation Invariance: Symmetric Function

$$f(x_1, x_2, \ldots, x_n) \equiv f(x_{\pi_1}, x_{\pi_2}, \ldots, x_{\pi_n}), \quad x_i \in \mathbb{R}^D$$

Examples:

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

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

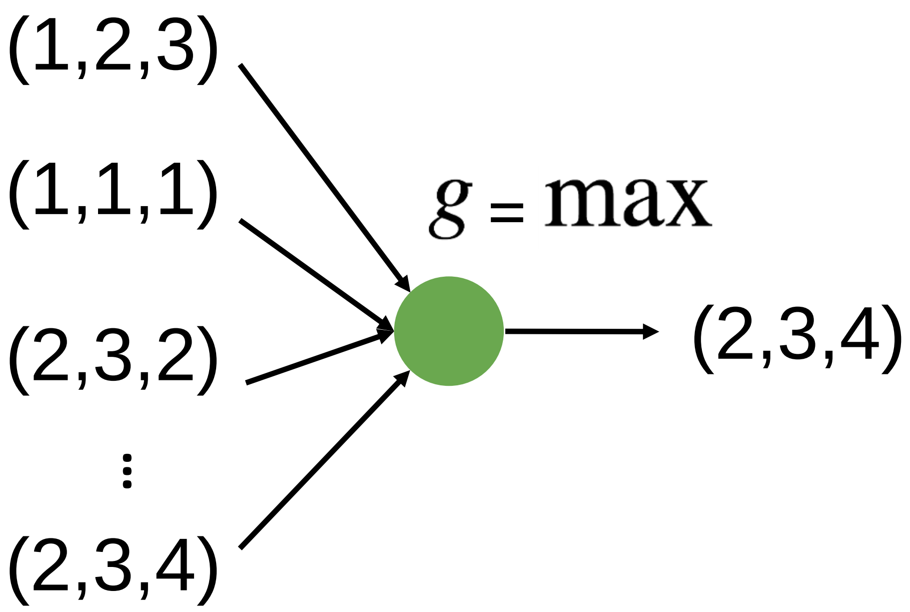
…

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

Construct Symmetric Functions by Neural Networks

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

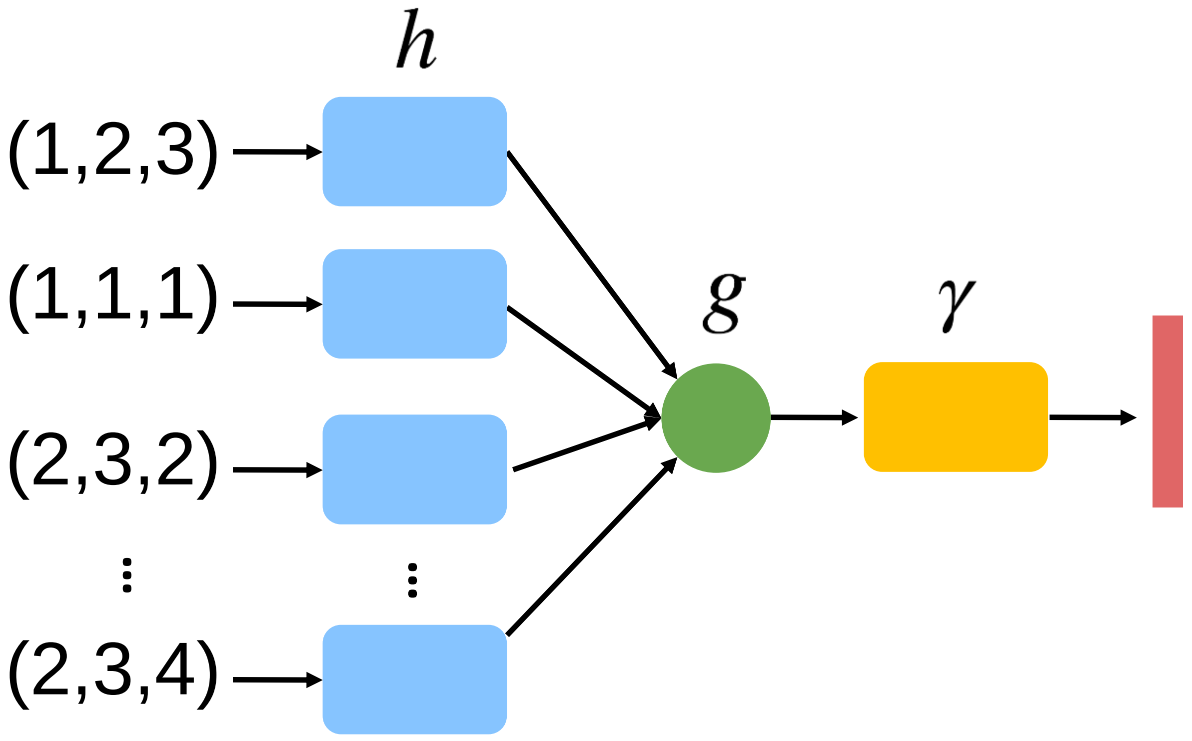
Just discovers simple extreme/aggregate properties of the geometry.



Construct Symmetric Functions by Neural Networks

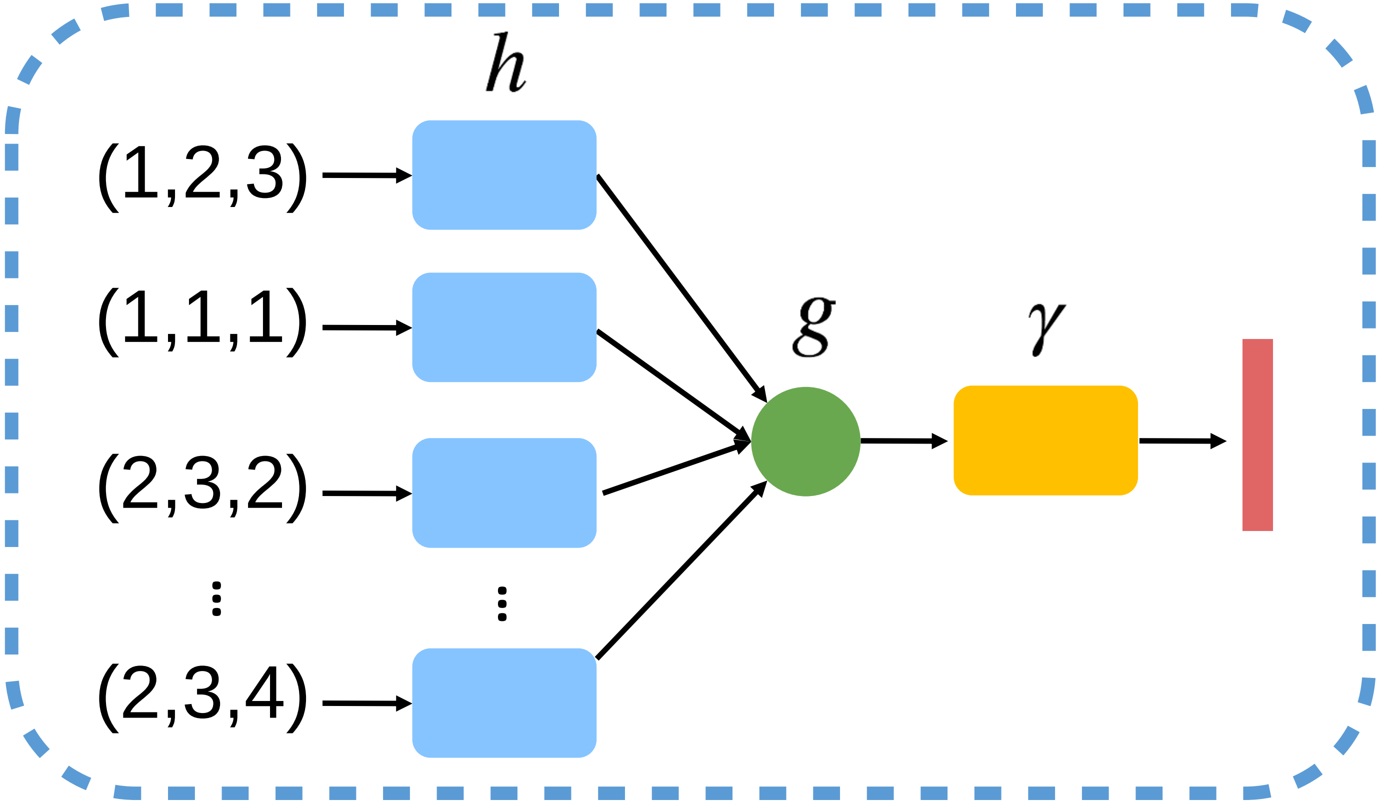
Embed points to a high-dim space before aggregation.

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



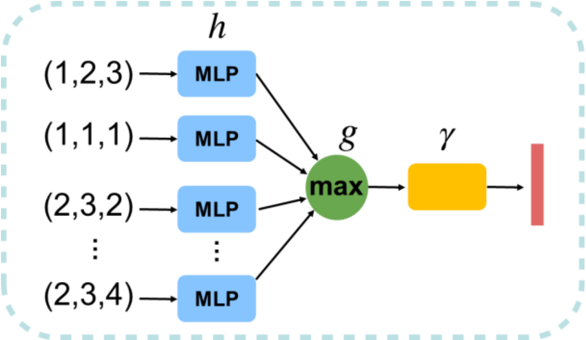
Construct Symmetric Functions by Neural Networks

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



PointNet (vanilla)

Symmetric Functions: Polynomials



$$2\sum_{i\neq j} x_i x_j = (\sum_i x_i)^2 - \sum_i x_i^{\,2} \qquad \sum_{i\neq j}(x_i - x_j)^2 = 3\sum_i x_i^{\,2} - (\sum_i x_i)^2$$

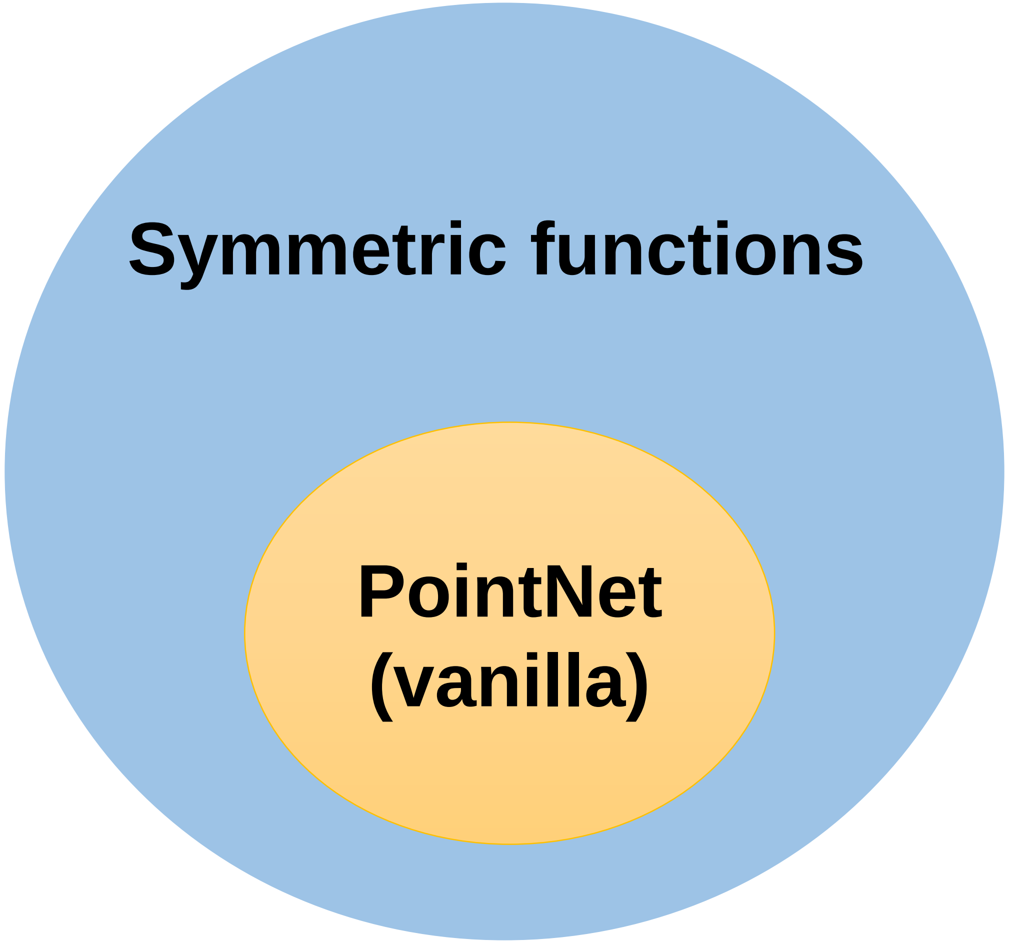
- In fact, any symmetric polynomial in the x_i can be expressed as a polynomial in sums of the form

$$\sum_i x_i^{\,k}$$

and can be computed by

$$f(x_1, x_2, \ldots, x_n) = \gamma \circ g(h(x_1), \ldots, h(x_n))$$

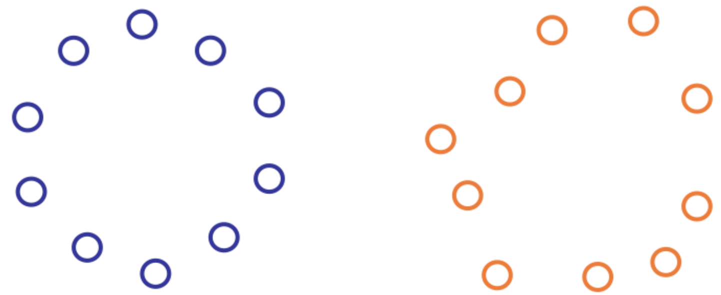
What Symmetric Functions Can Be Constructed By PointNet?



PointNet as a Universal Approximation to Set Functions

Hausdorff continuous:

$f:2^{\mathcal{X}} \to \mathbb{R}$ is a continuous set function w.r.t Hausdorff distance



S $\quad$ S' (perturbed)

if $d_{Hausdorff}(S,S') \approx 0$, then $f(S) \approx f(S')$

Theorem

A Hausdorff continuous set function $f:2^{\mathcal{X}} \to \mathbb{R}$ can be arbitrarily approximated by PointNet.

$$\left| f(S) - \boxed{\gamma\left(\underset{x_i \in S}{\mathrm{MAX}}\{h(x_i)\}\right)} \right| < \epsilon$$

$S \subseteq R^d$

PointNet (vanilla)

Voxel occupancy maps

Challenges

The model has to respect key properties of point clouds:

Point Permutation Invariance

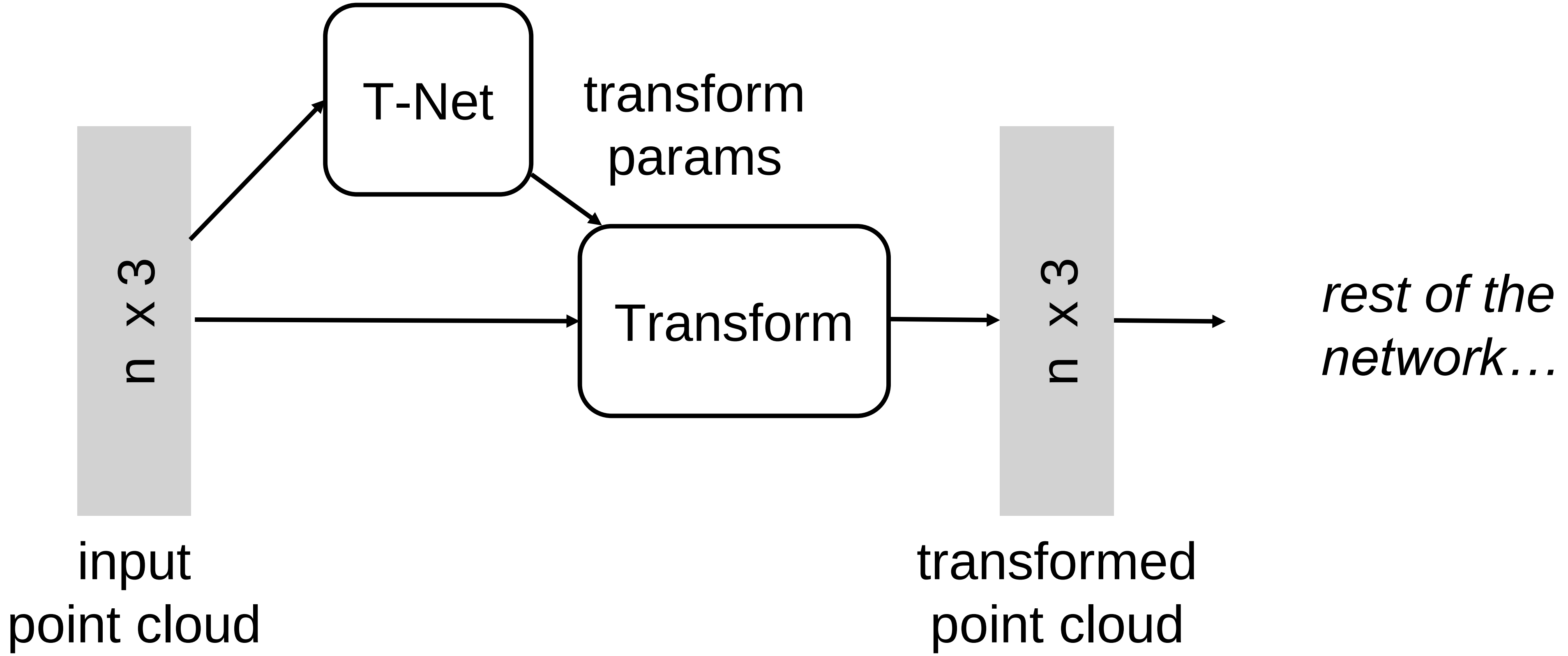
Point cloud is a set of unordered points

Transformation Invariance

Point cloud rigid motions should not alter classification results

Input Alignment by Transformer Network

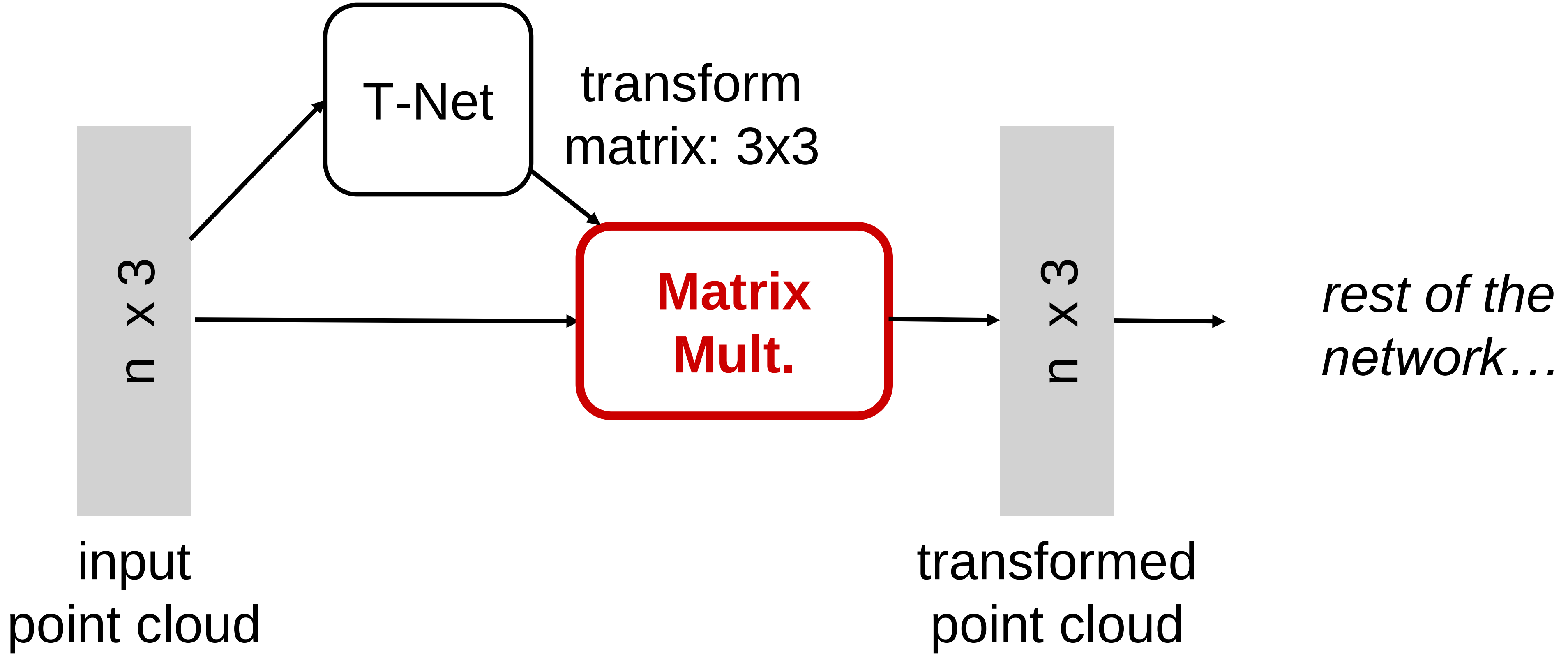
Idea: Data dependent transformation for automatic alignment



Input Alignment by Transformer Network

Idea: Data dependent transformation for automatic alignment

The transformation is just matrix multiplication!

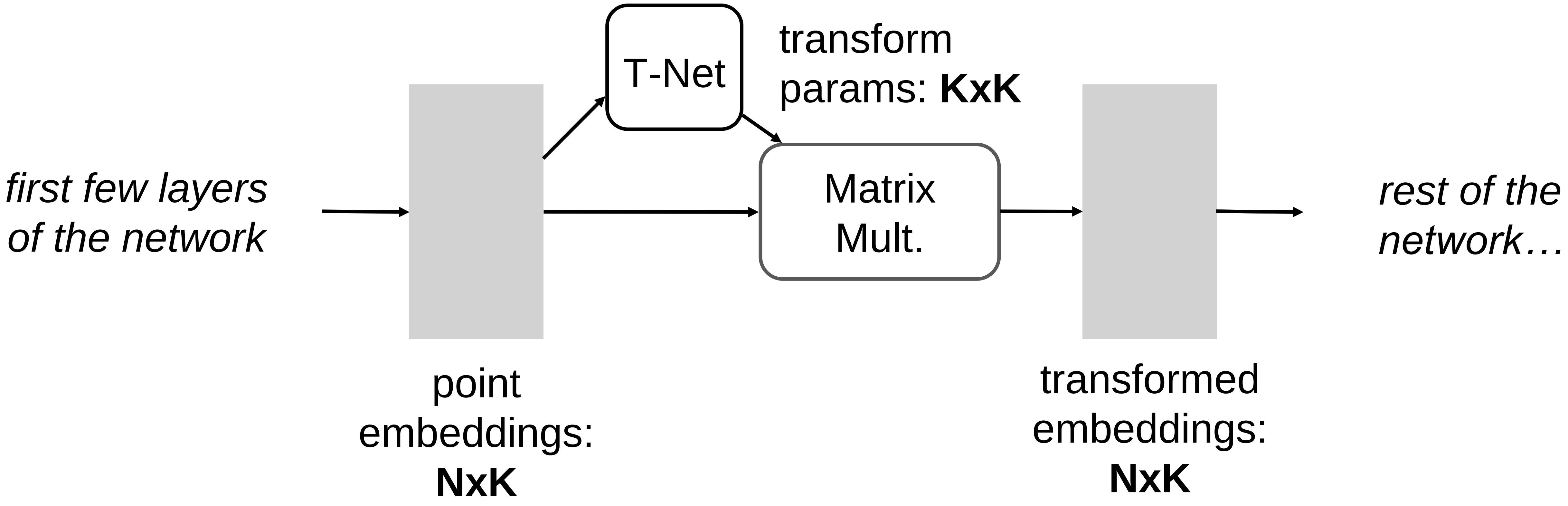


Embedding Space Alignment

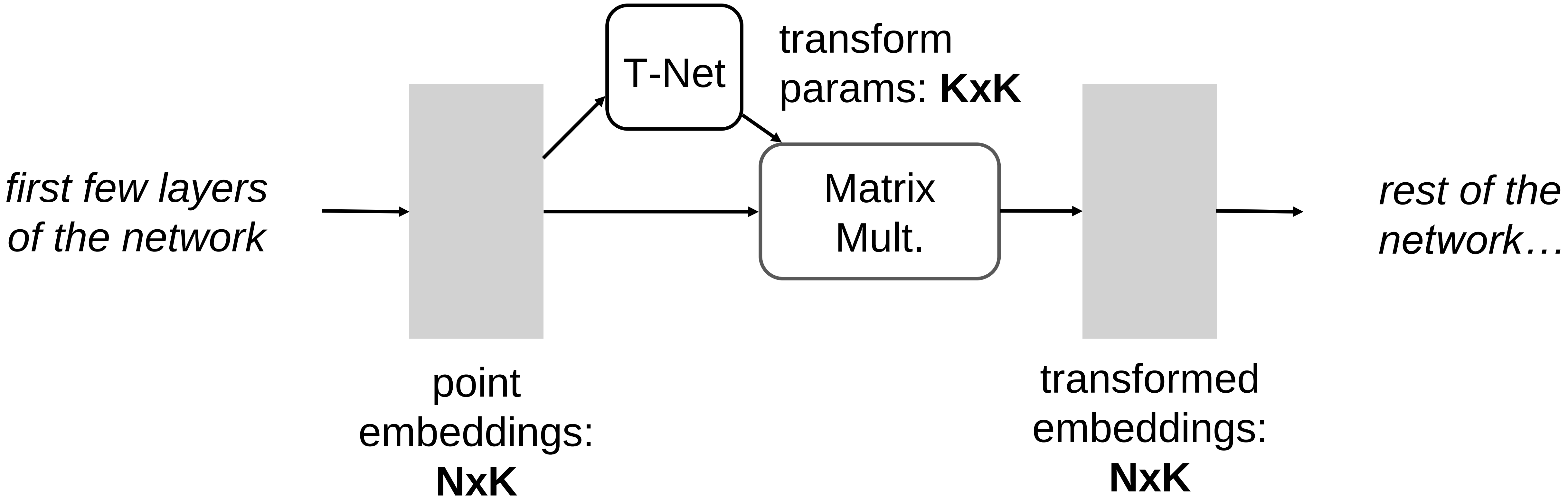
first few layers of the network

point embeddings: **NxK**

Embedding Space Alignment



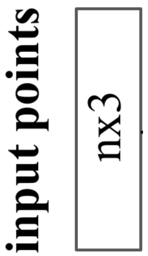
Embedding Space Alignment



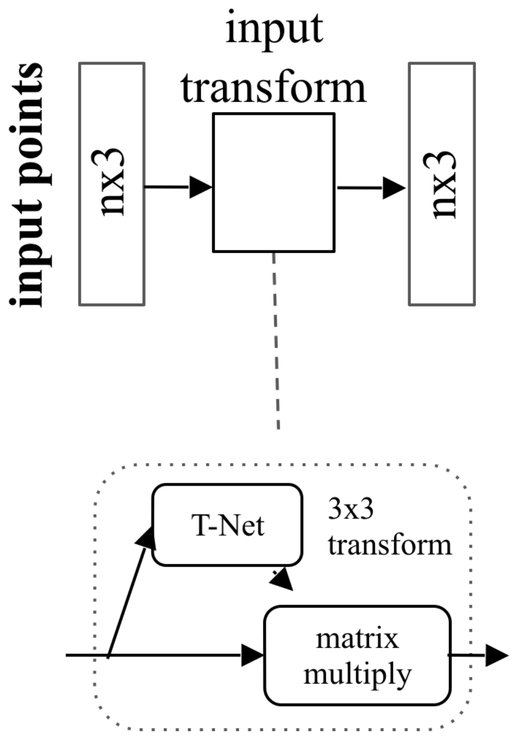
Regularization loss:

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

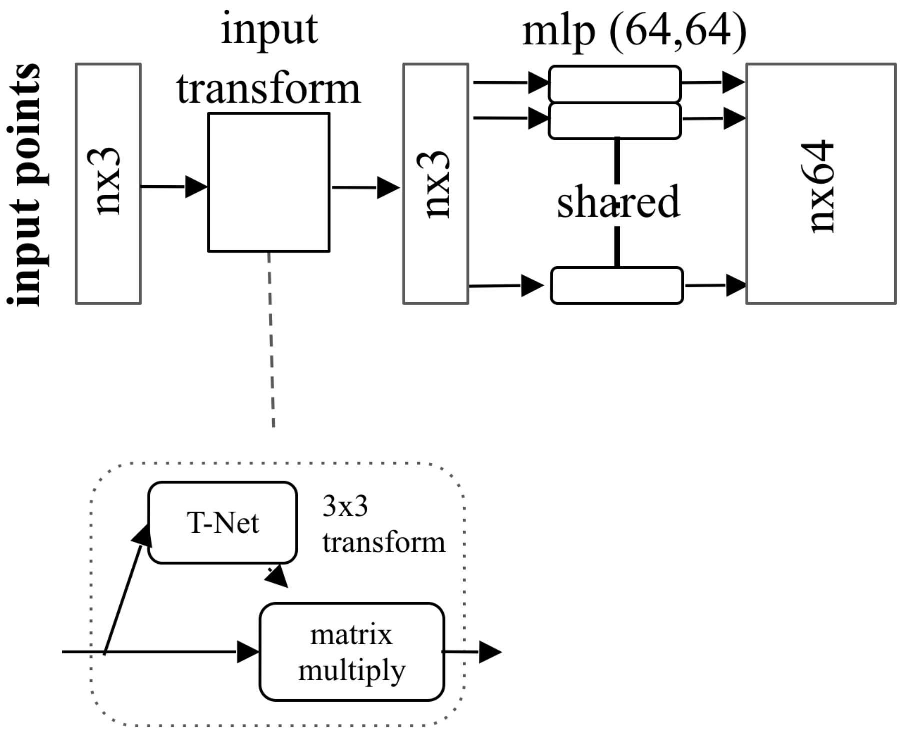
PointNet Classification Network



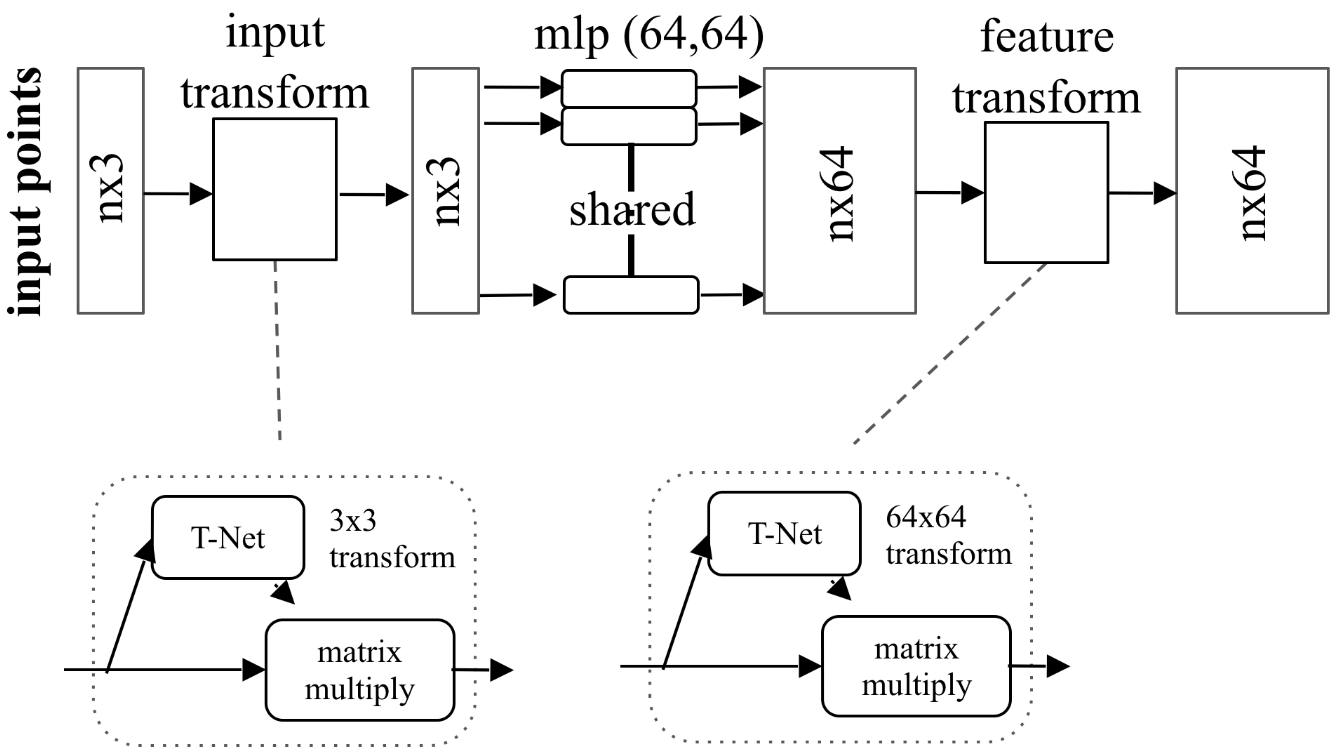
PointNet Classification Network



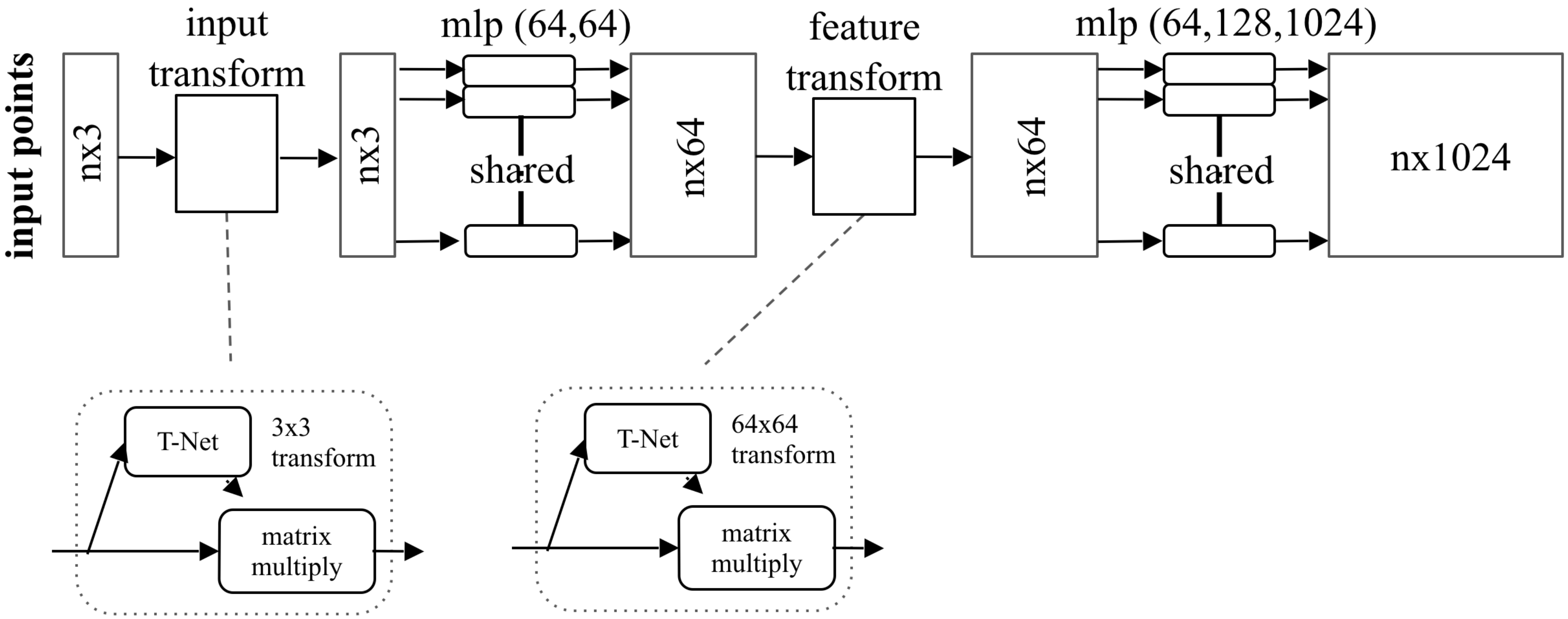
PointNet Classification Network



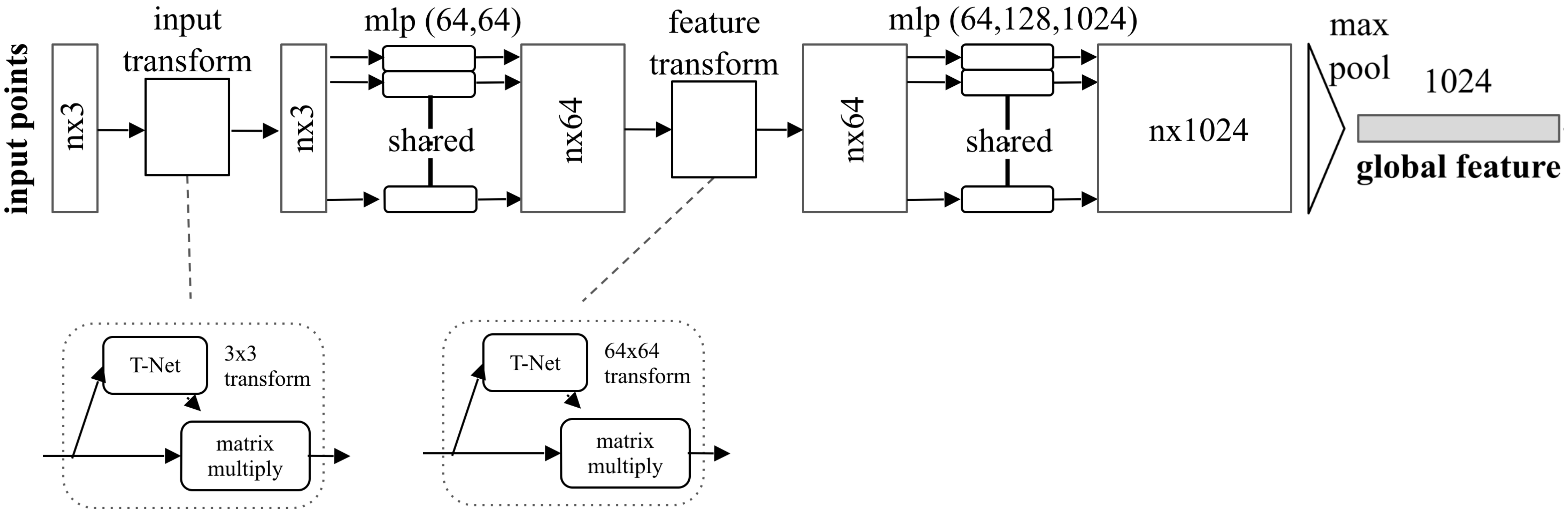
PointNet Classification Network



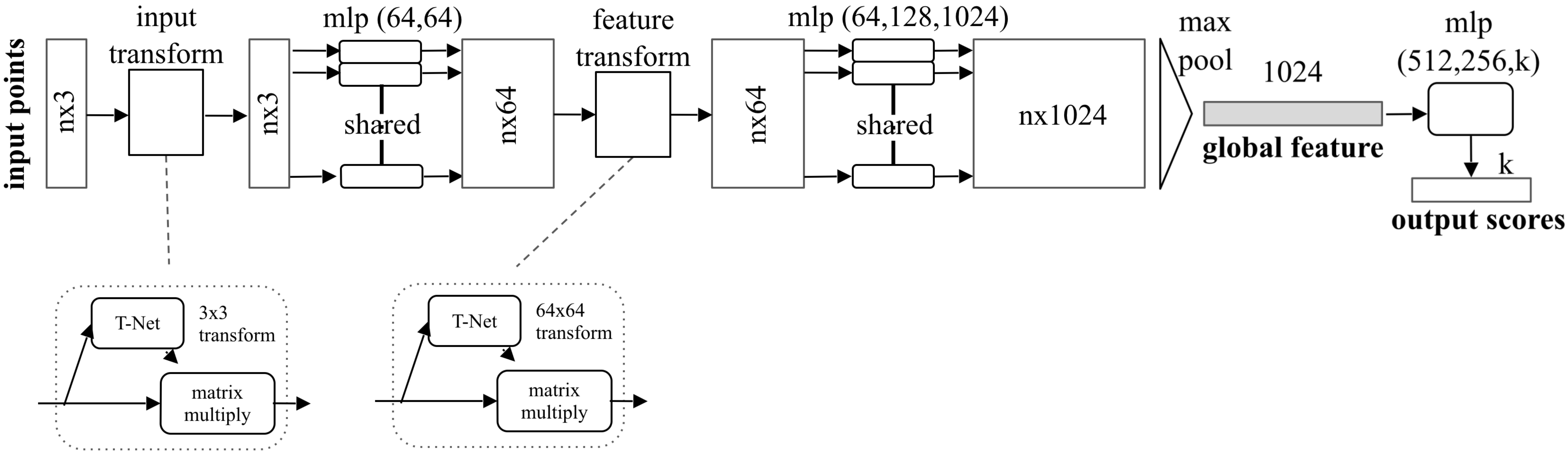
PointNet Classification Network



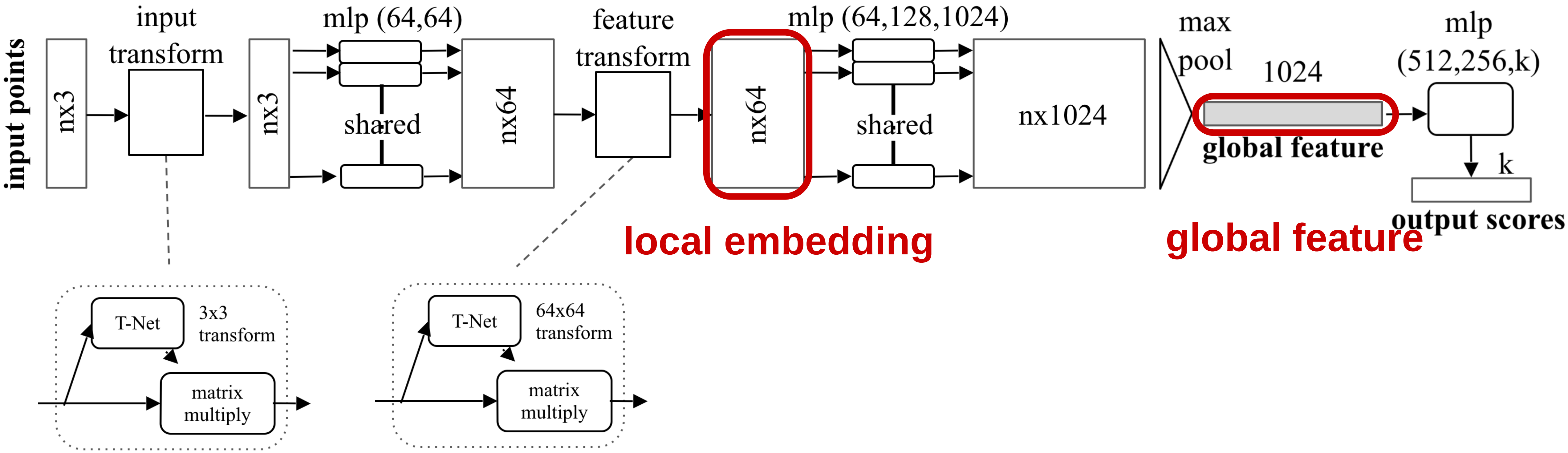
PointNet Classification Network



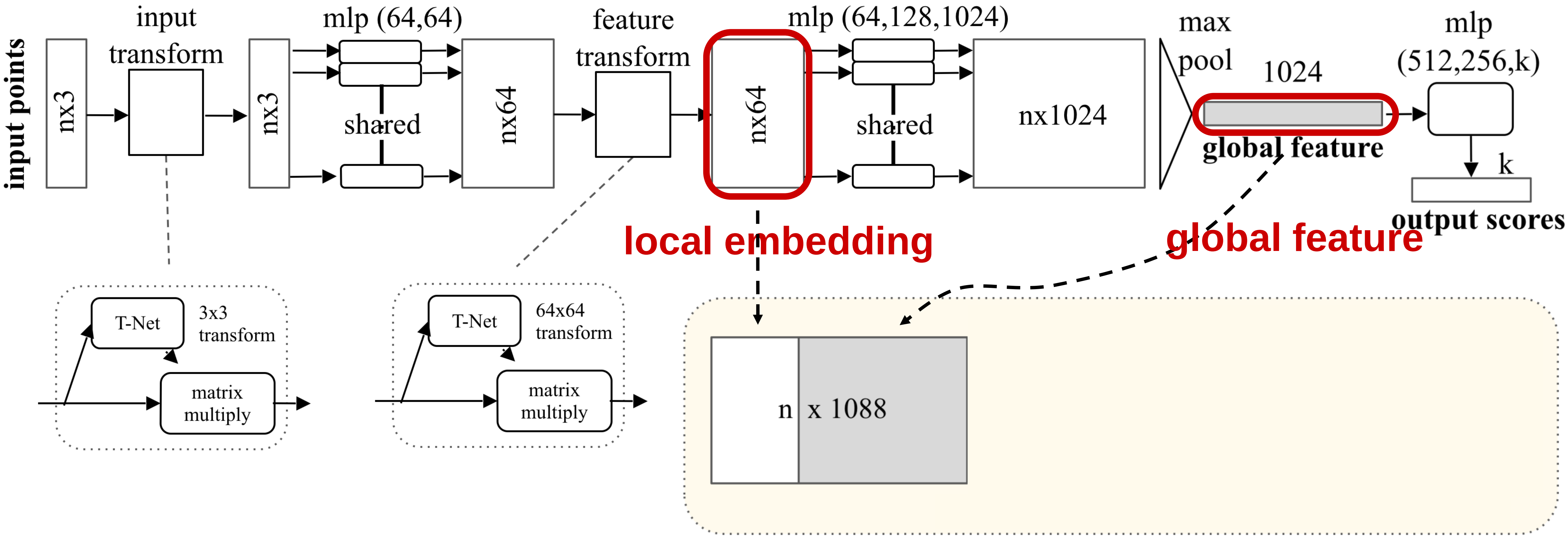
PointNet Classification Network



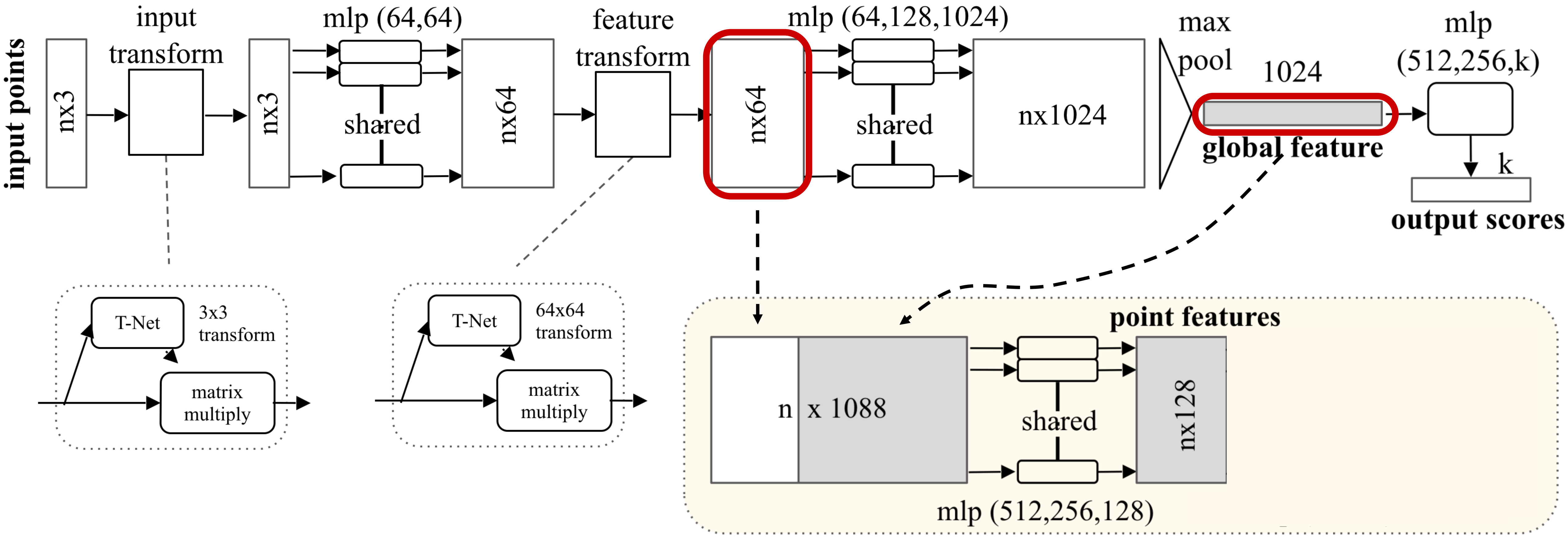
Extension to PointNet Segmentation Network



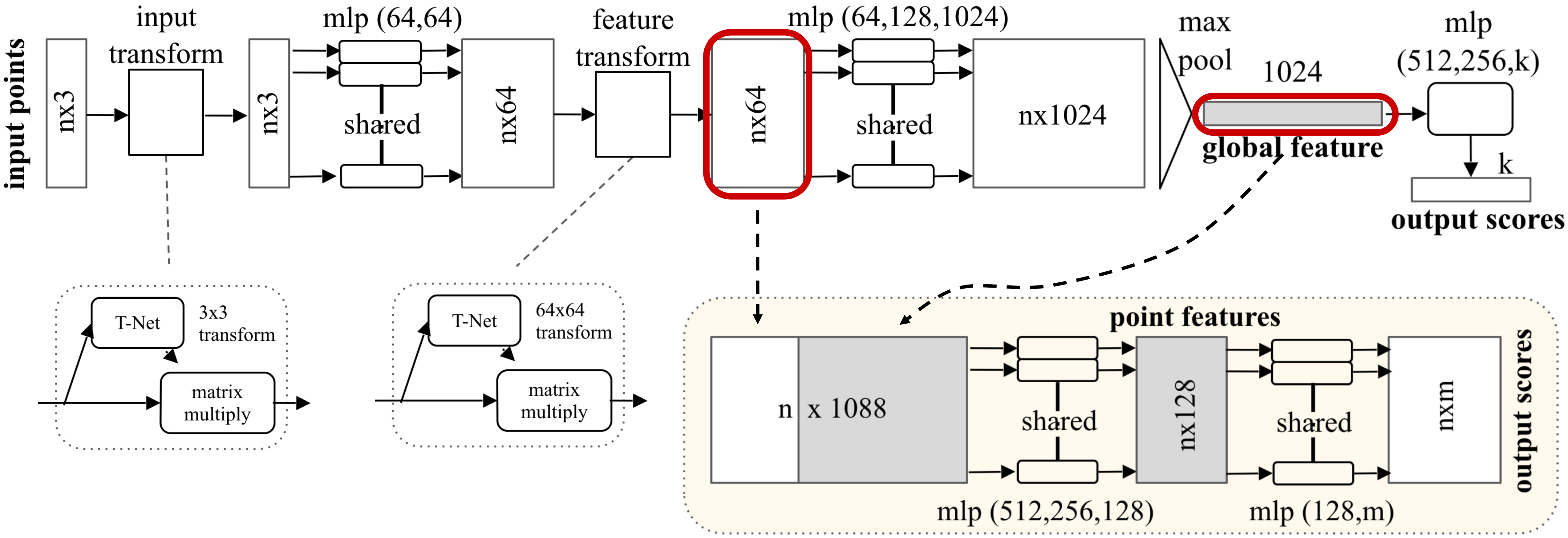
Extension to PointNet Segmentation Network



Extension to PointNet Segmentation Network



Extension to PointNet Segmentation Network



Results

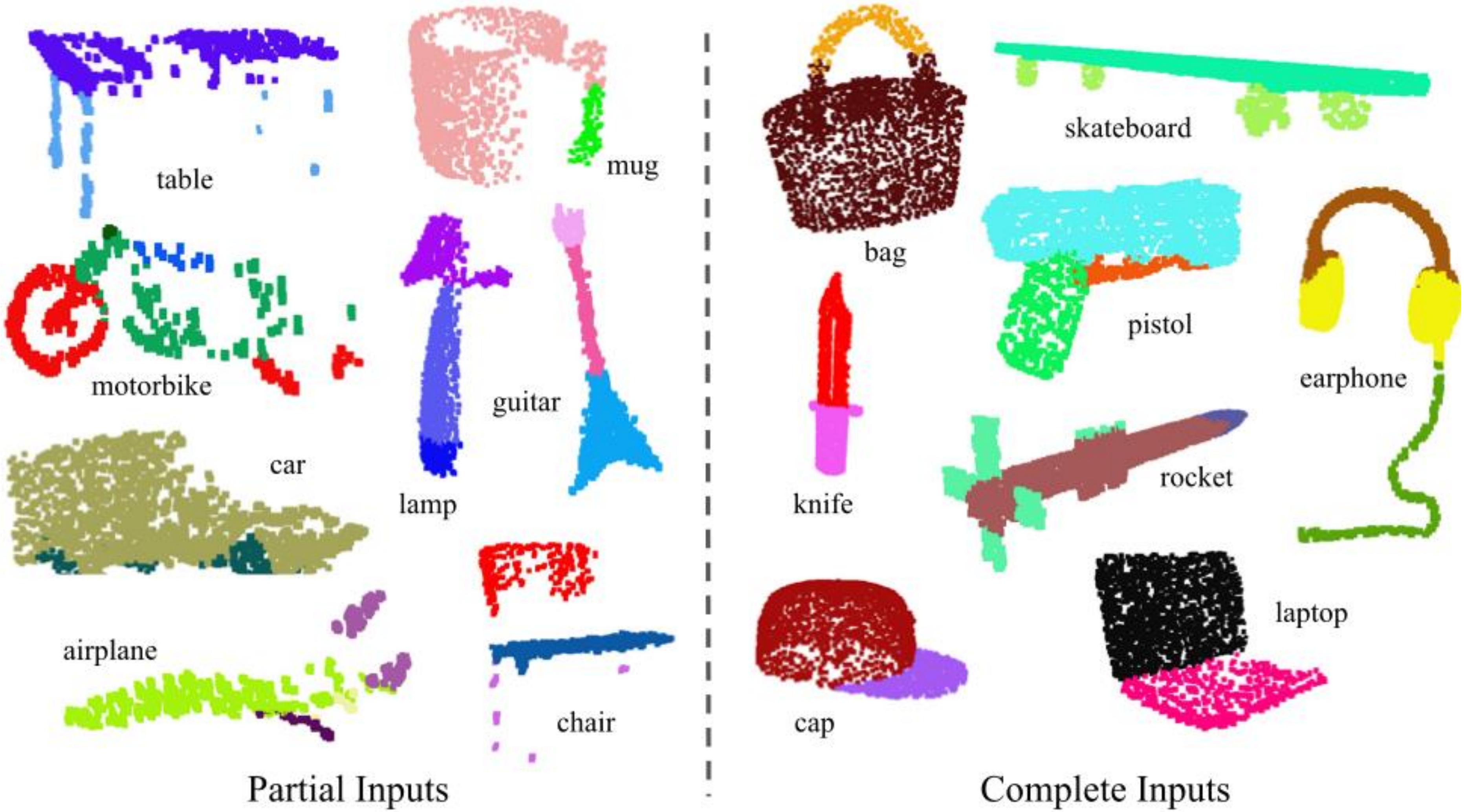
Results on Object Classification

3D CNNs

	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**

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

Results on Object Part Segmentation



Results on Object Part Segmentation

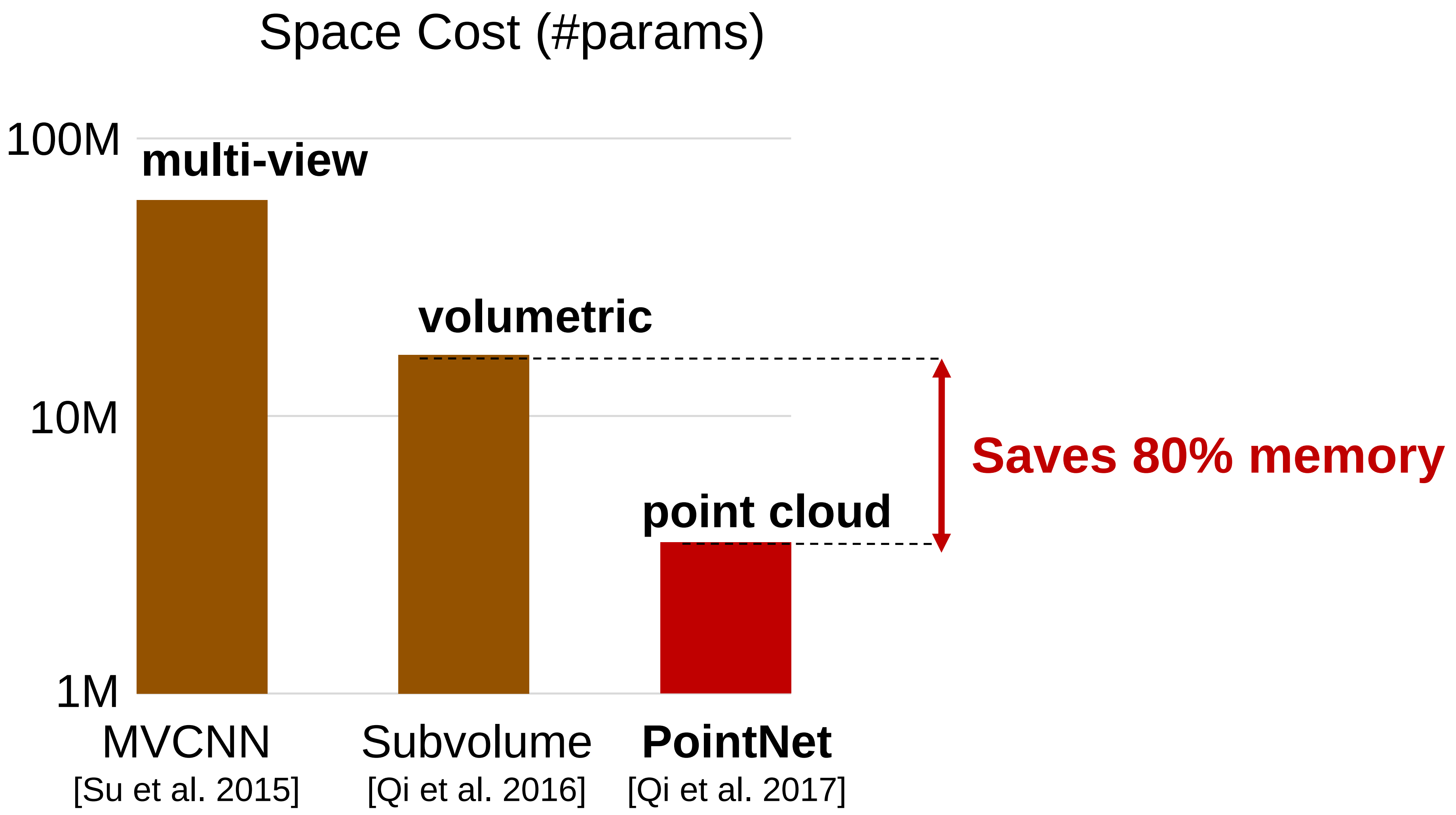
	mean	aero	bag	cap	car	chair	ear phone	guitar	knife	lamp	laptop	motor	mug	pistol	rocket	skate board	table
# shapes		2690	76	55	898	3758	69	787	392	1547	451	202	184	283	66	152	5271
Wu [28]	-	63.2	-	-	-	73.5	-	-	-	74.4	-	-	-	-	-	-	74.8
Yi [30]	81.4	81.0	78.4	77.7	**75.7**	87.6	61.9	**92.0**	85.4	**82.5**	**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 (%)

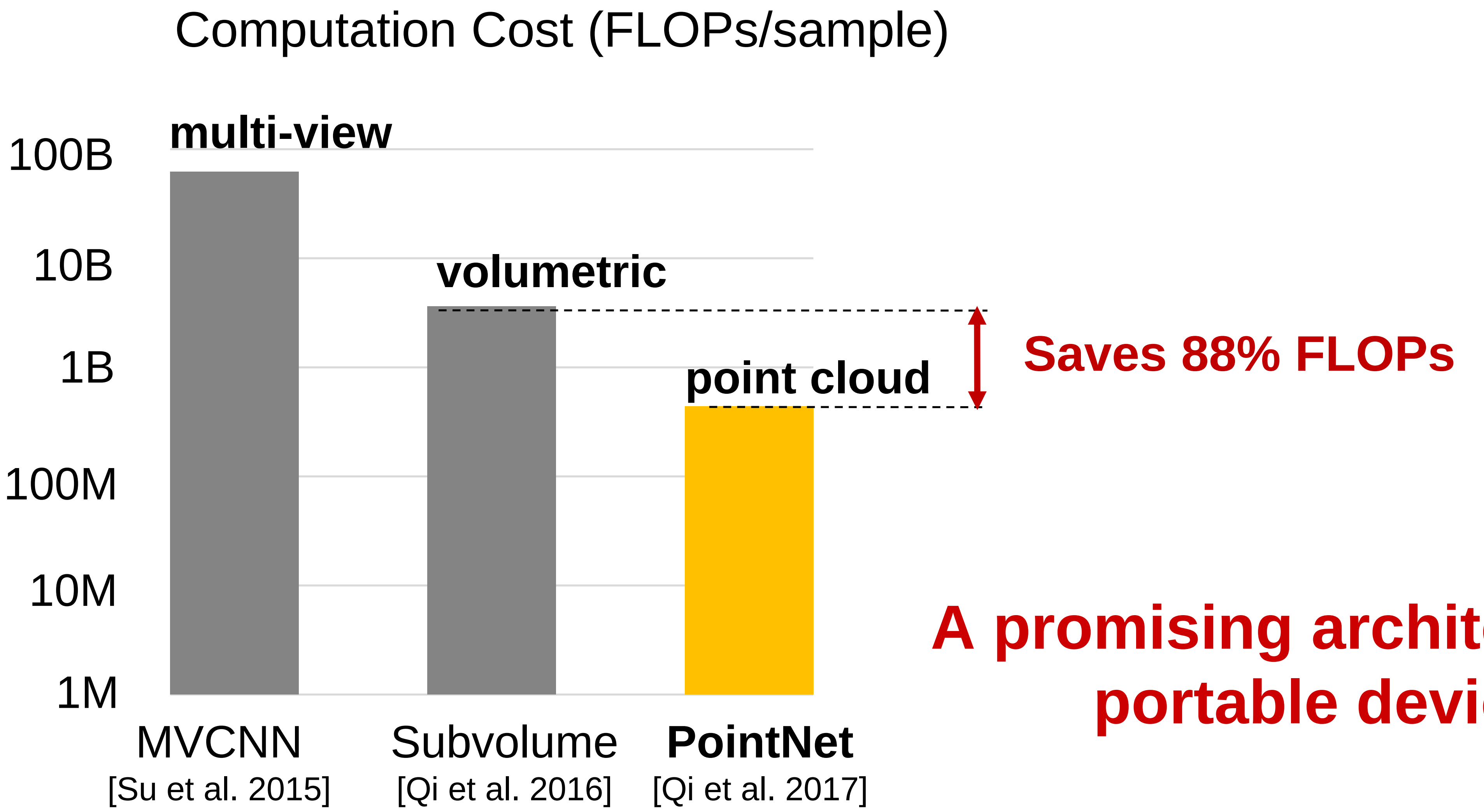
Results on Semantic Scene Parsing

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

PointNet is Light-Weight and Fast

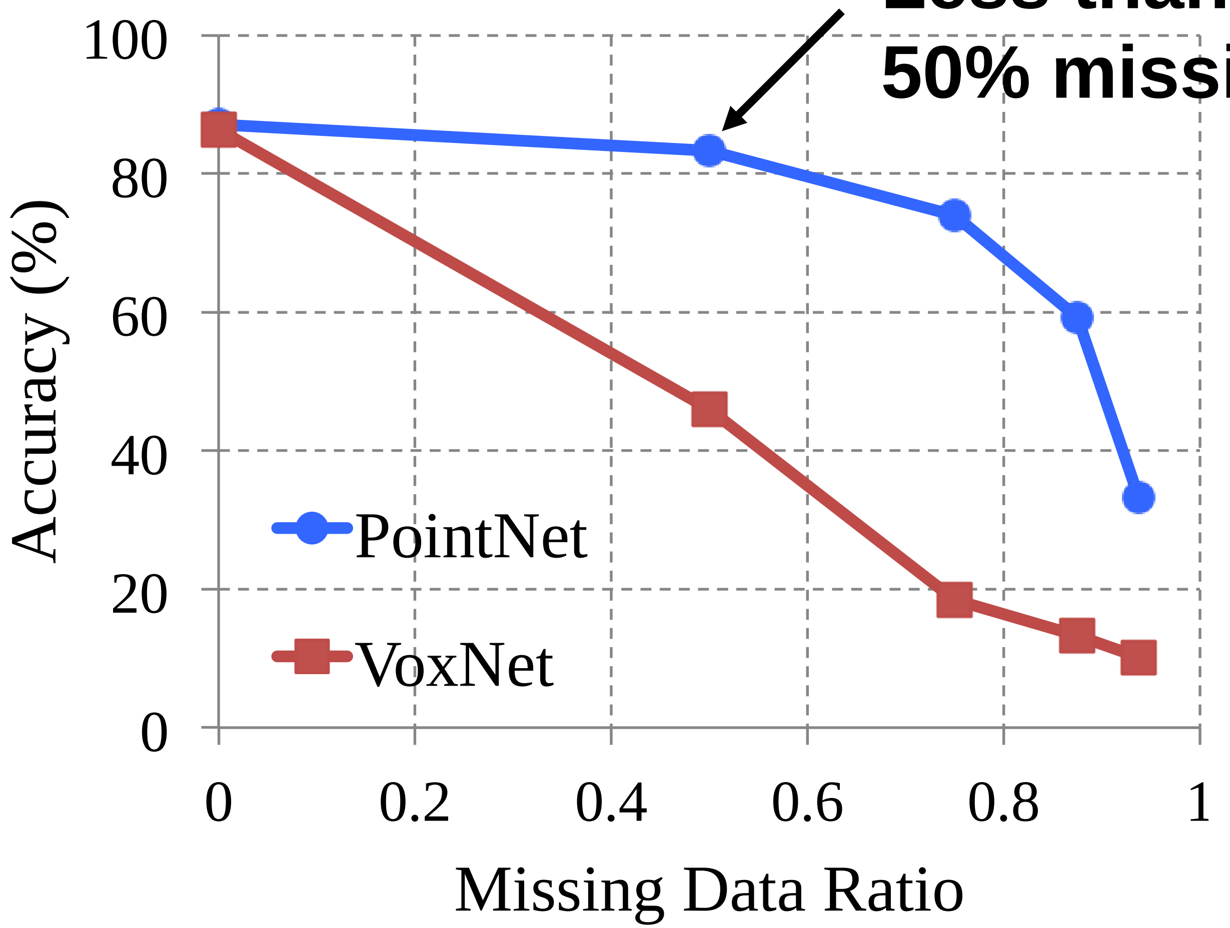


PointNet is Light-Weight and Fast



A promising architecture for portable devices!

PointNet is Robust to Data Corruption



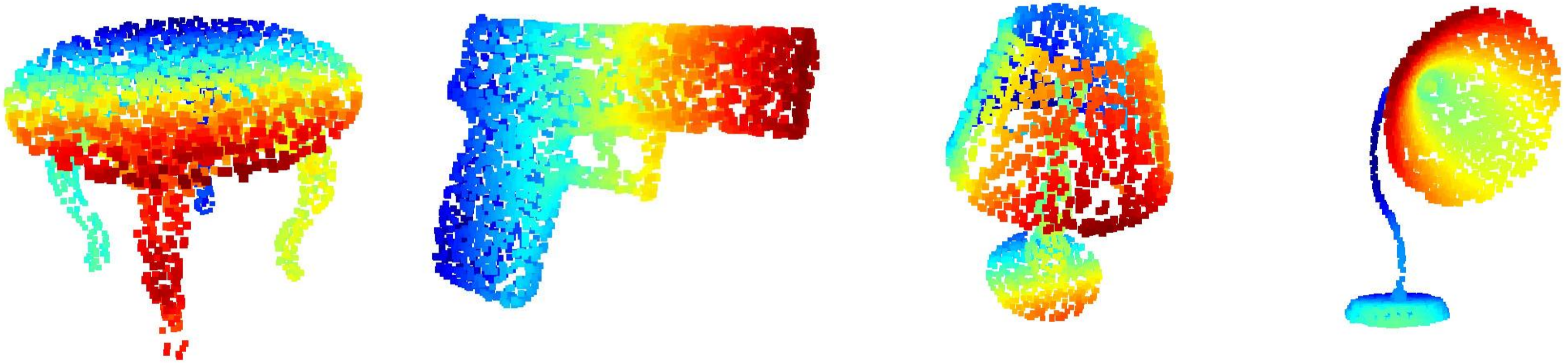
Less than 2% accuracy drop with 50% missing data

Why is PointNet so robust to missing data?

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

Visualizing Global Point Cloud Features

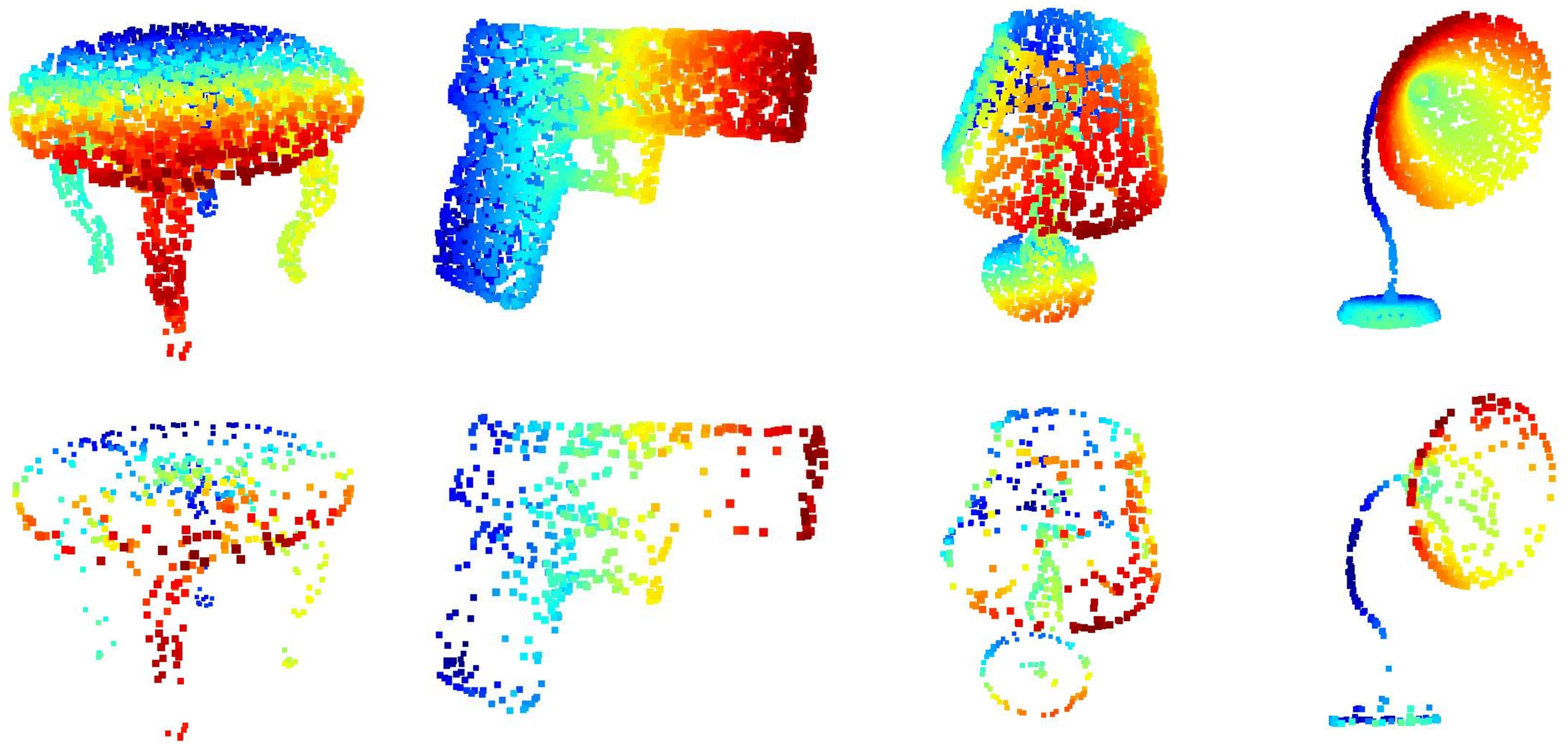
Original Shape



Visualizing Global Point Cloud Features

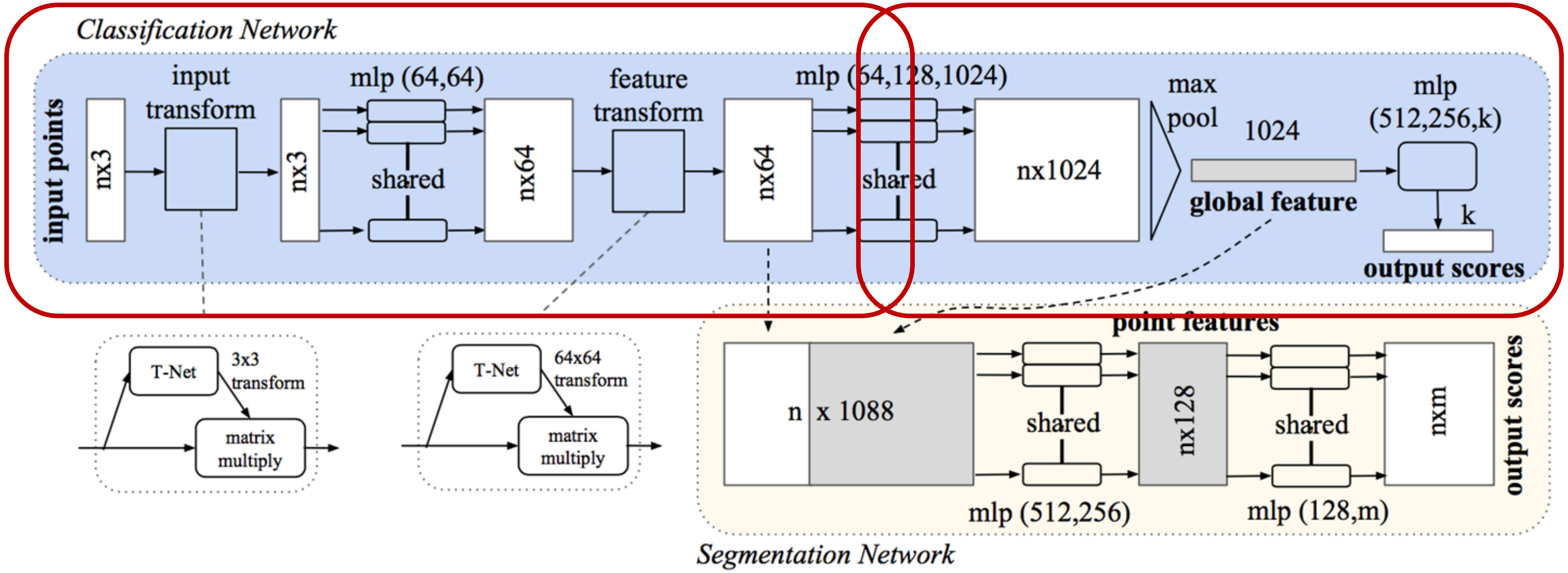
Original Shape

Critical Points



PointNet learns to pick perceptually interesting points!

Learning Interesting Points

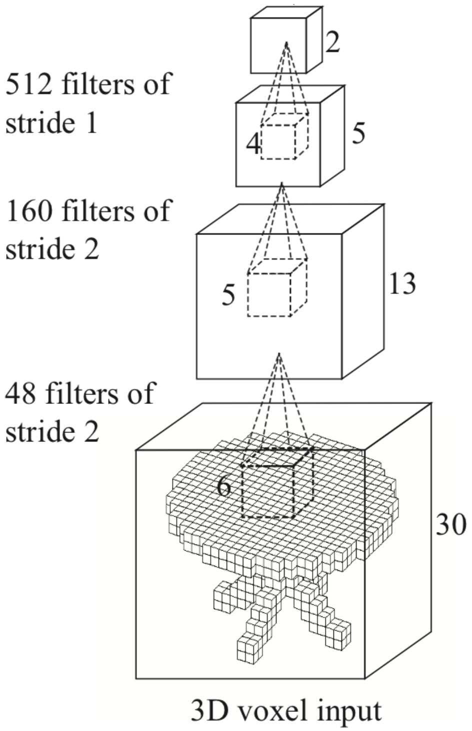


- Pointnet learns optimization criteria, which in turn pick interesting points

From PointNet to PointNet++

Limitations of PointNet

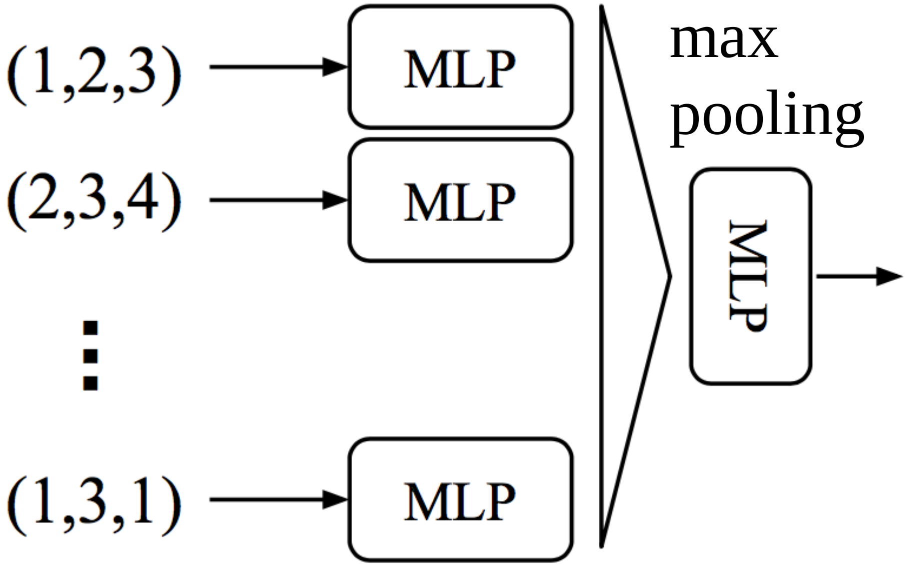
Hierarchical feature learning
multiple levels of abstraction



3D CNN [Wu et al.2015]

v.s.

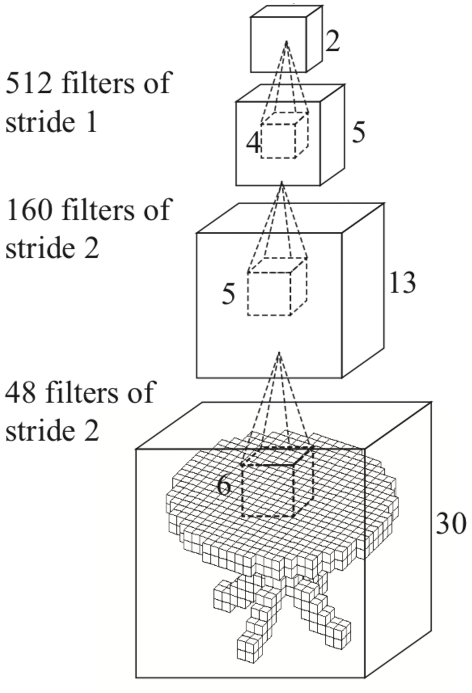
Global feature learning
either **one** point or **all** points



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 local invariance

PointNet (vanilla) [Qi et al.2017]

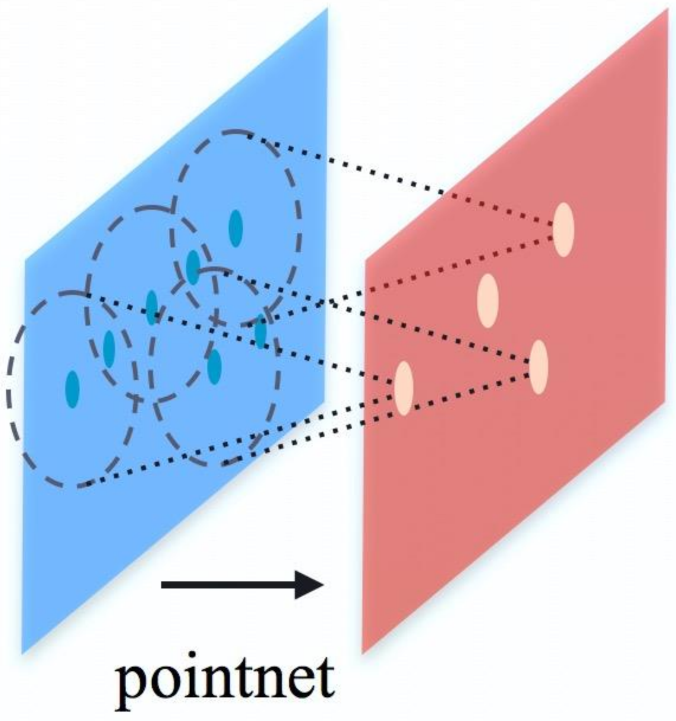
PointNet++

Basic idea: Recursively apply pointnet at local regions.

✓ Hierarchical feature learning

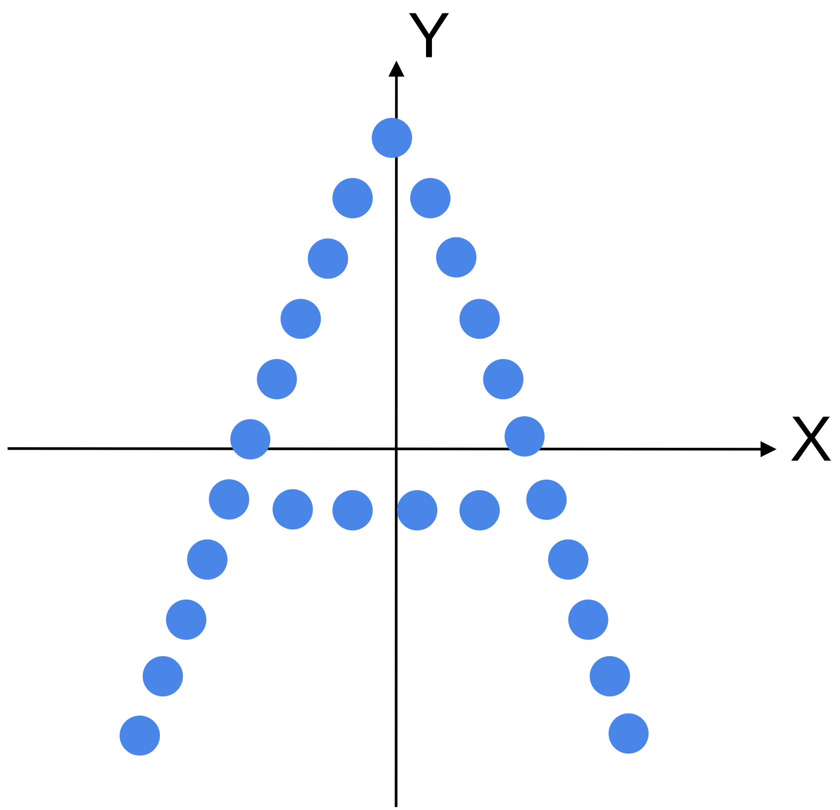
✓ Local translation invariance

✓ Permutation invariance



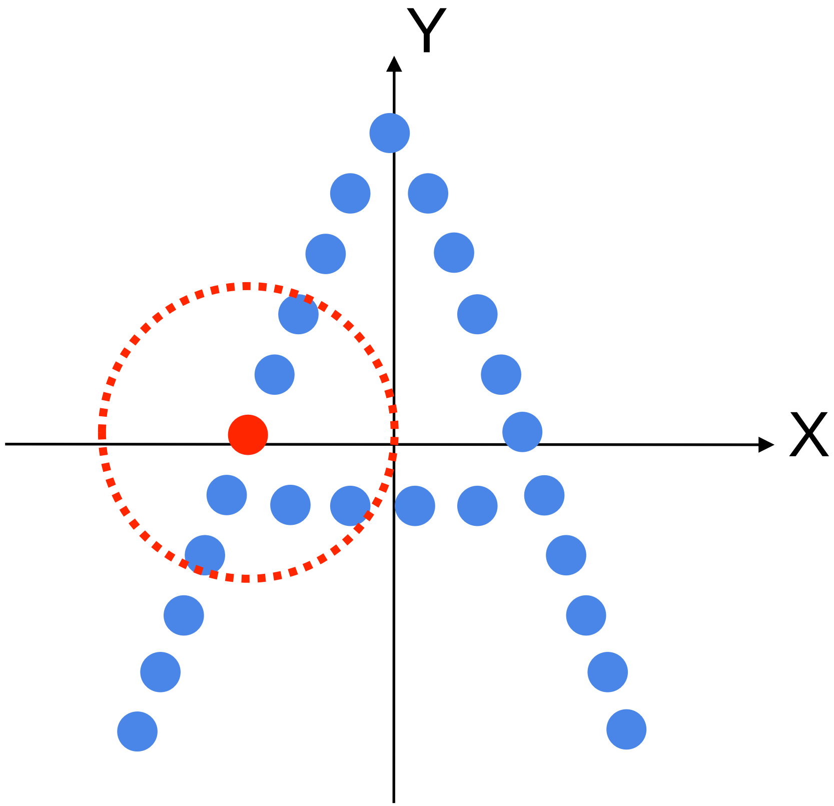
[2] *Charles R. Qi, Li Yi, Hao Su, Leonidas Guibas. PointNet++: Deep Hierarchical Feature Learning on Point Sets in a Metric Space (NIPS'17)*

Hierarchical Point Feature Learning



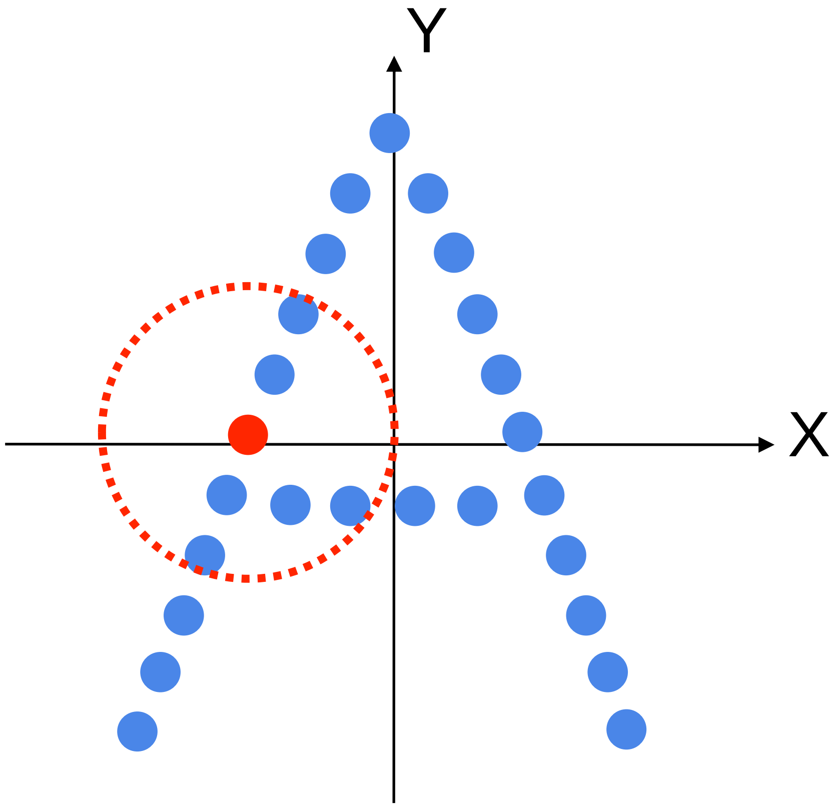
N points in (X,Y)

Hierarchical Point Feature Learning

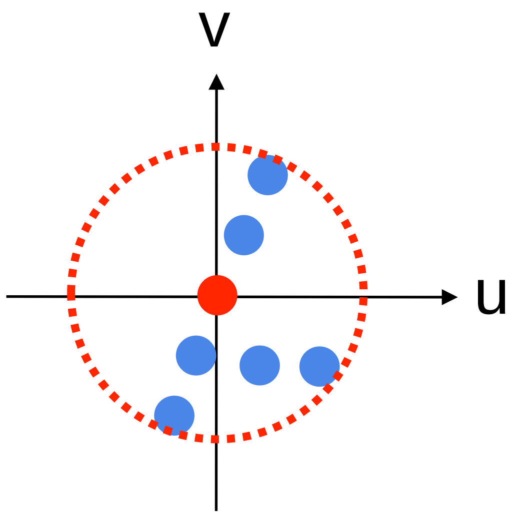


N points in (X,Y)

Hierarchical Point Feature Learning

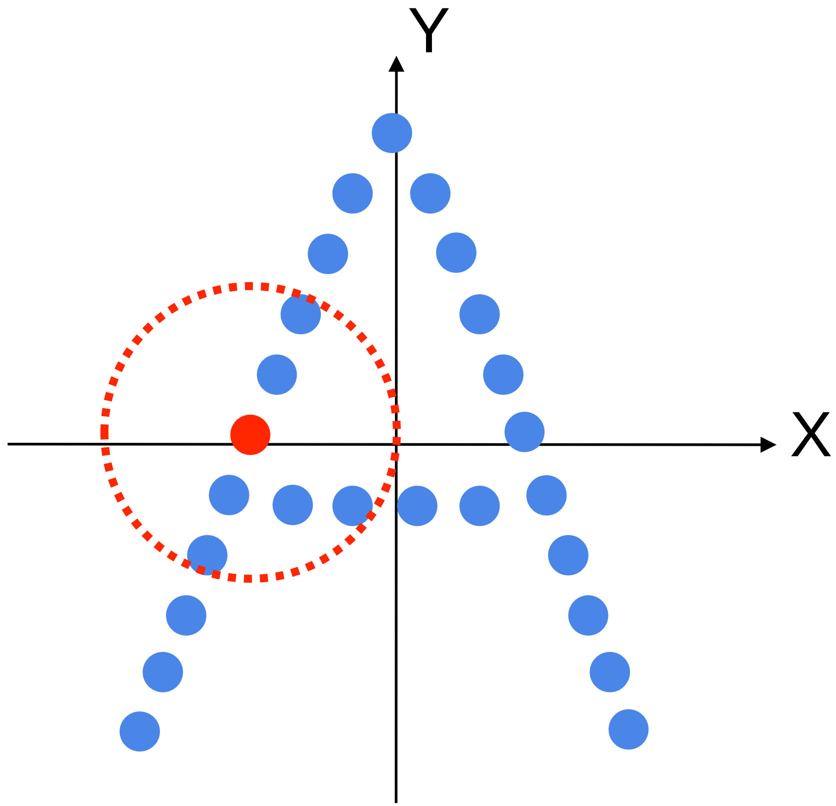


N points in (X,Y)



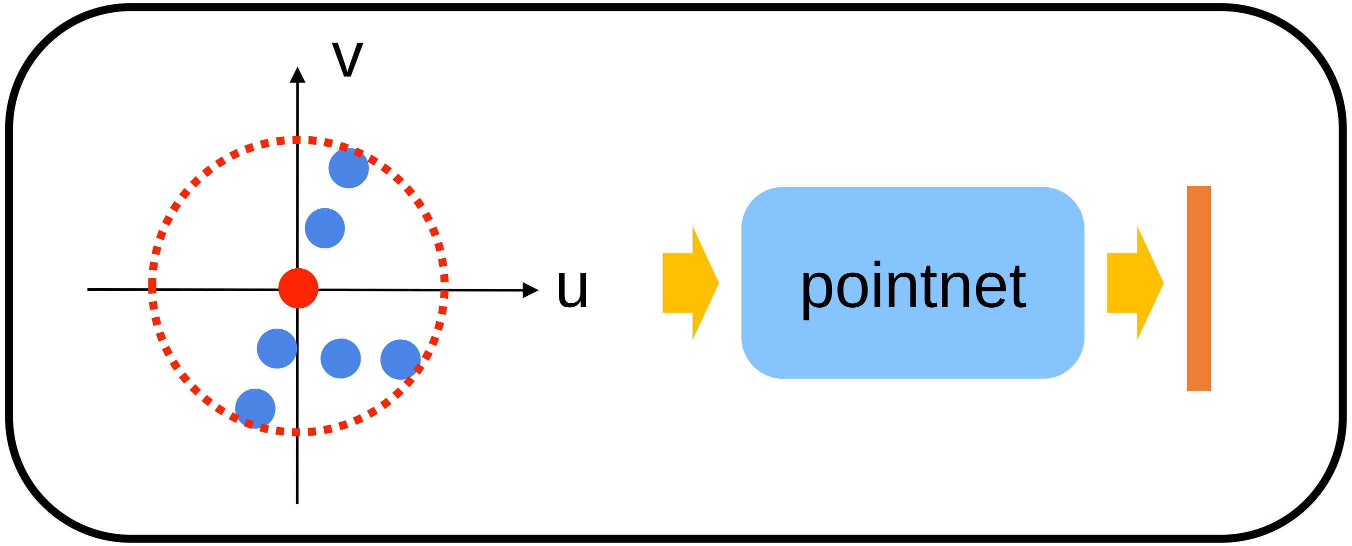
k points in local coordinates (u,v)

Hierarchical Point Feature Learning



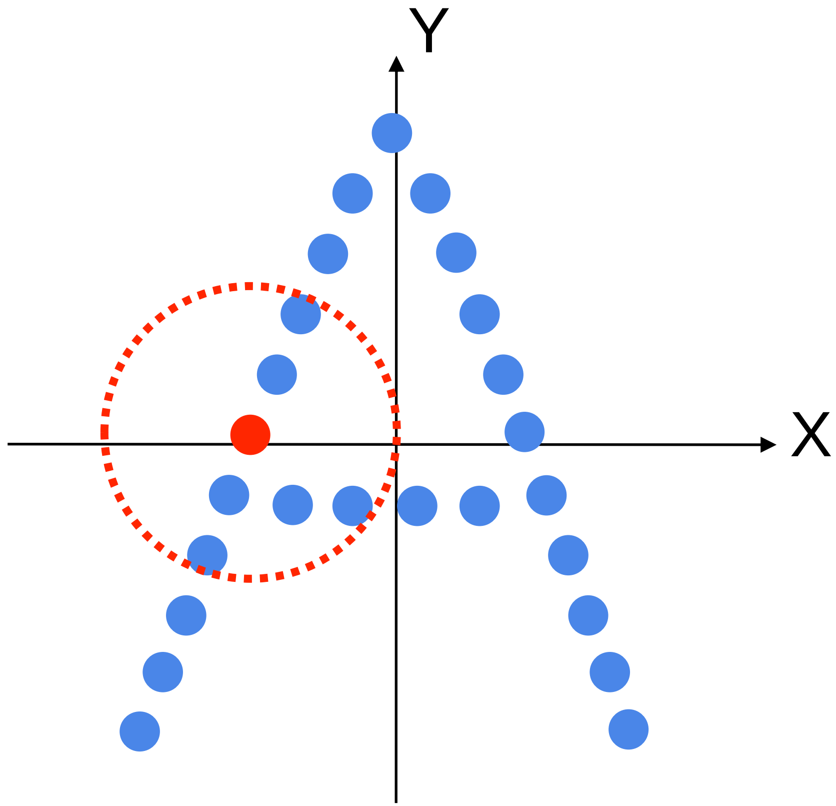
N points in (X,Y)

Apply pointnet at a local region

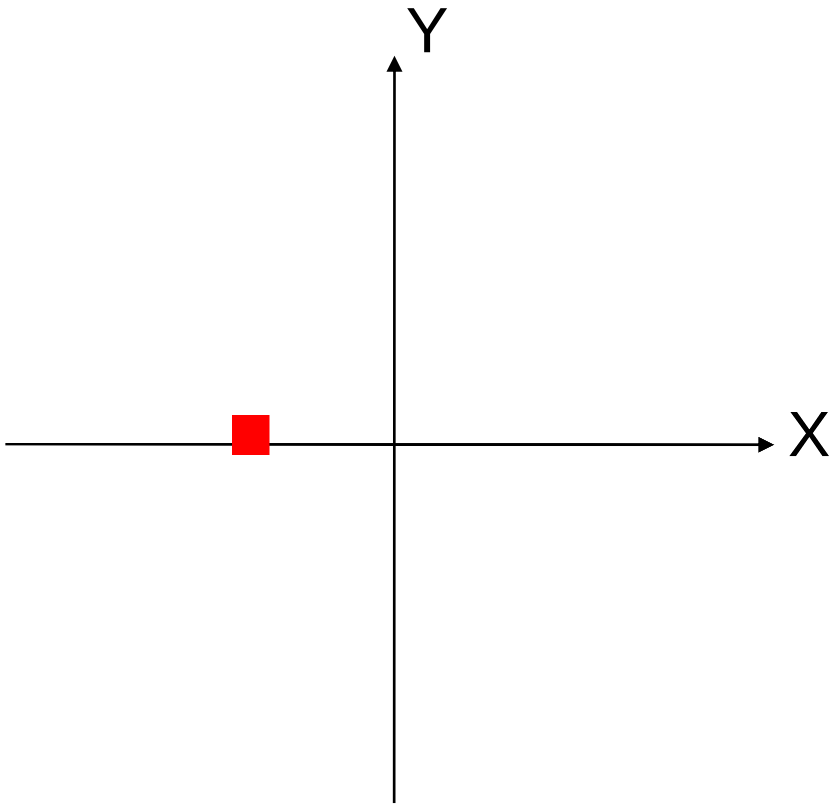


k points in local coordinates (u,v)

Hierarchical Point Feature Learning



N points in (X,Y)

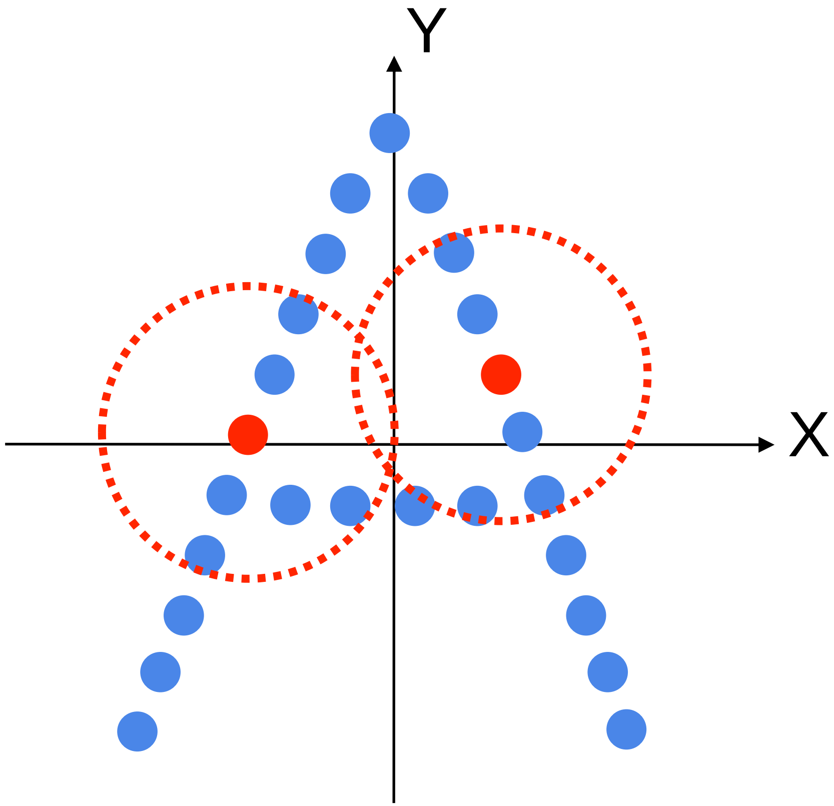


points in (X,Y, **F**)

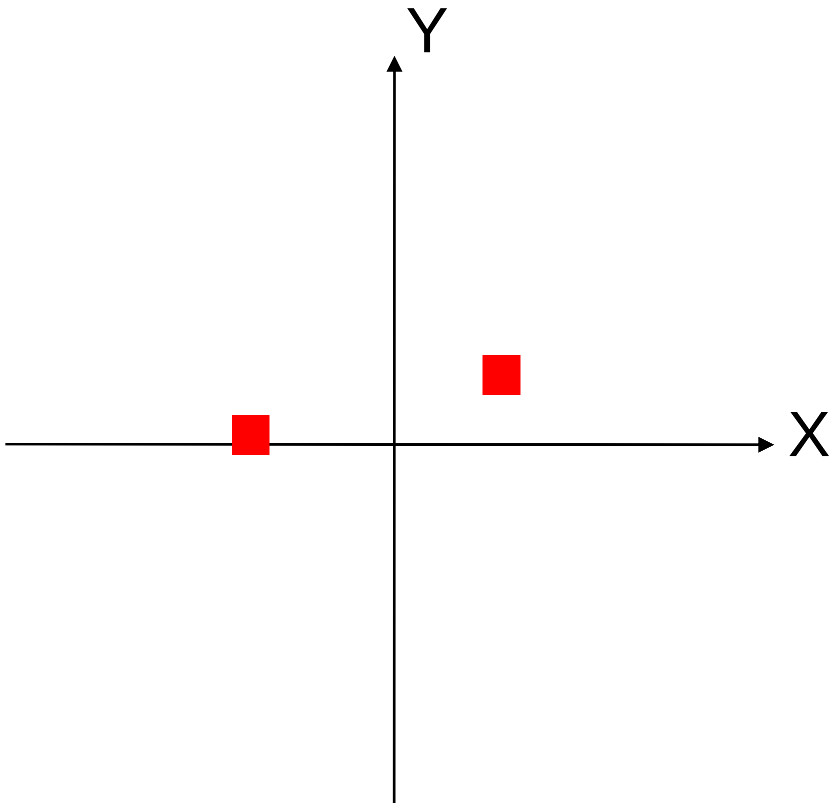
Euclidean space

high-dim feature space

Hierarchical Point Feature Learning

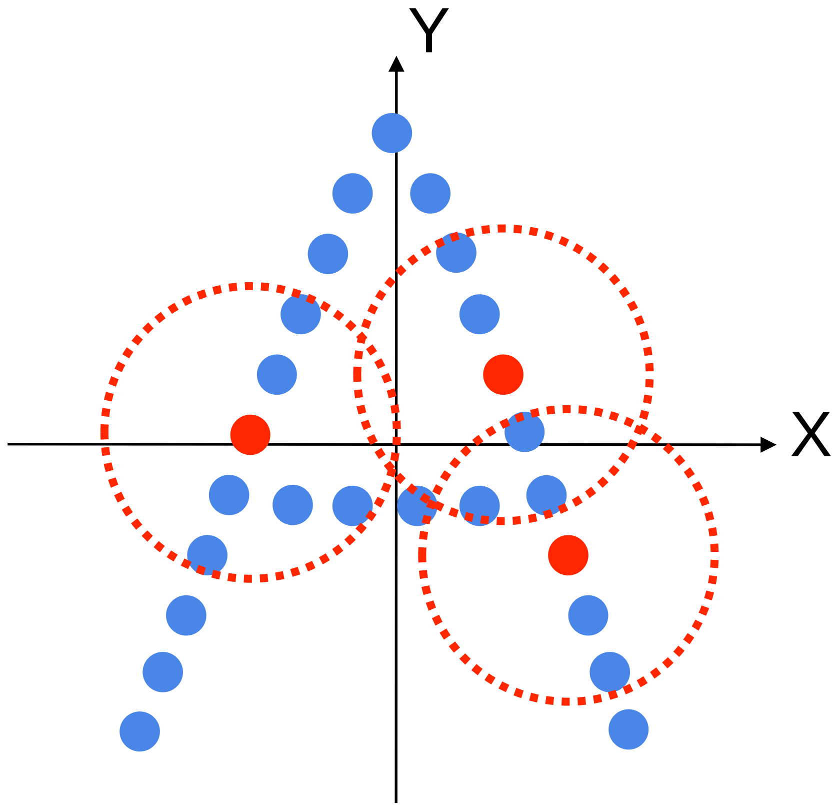


N points in (X,Y)

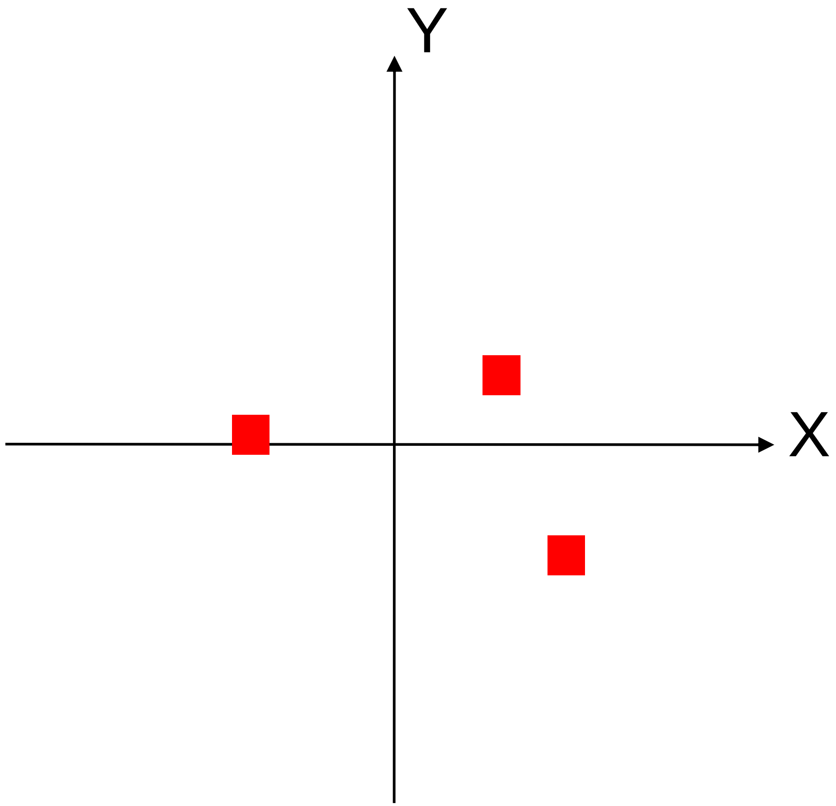


points in (X,Y, **F**)

Hierarchical Point Feature Learning

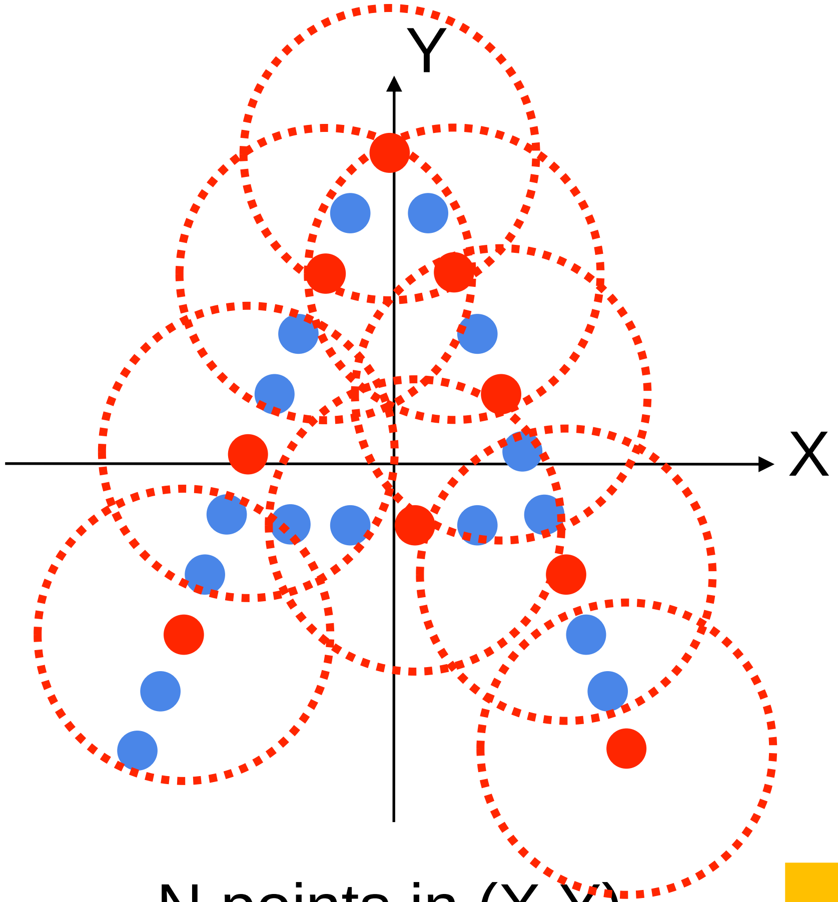


N points in (X,Y)

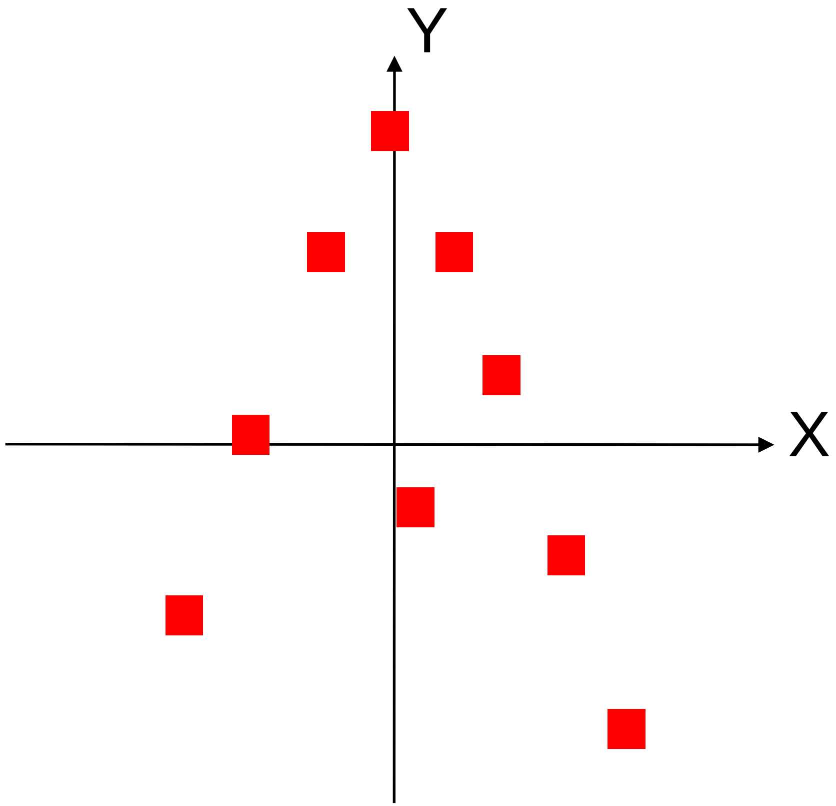


points in (X,Y, **F**)

Hierarchical Point Feature Learning



N points in (X,Y)

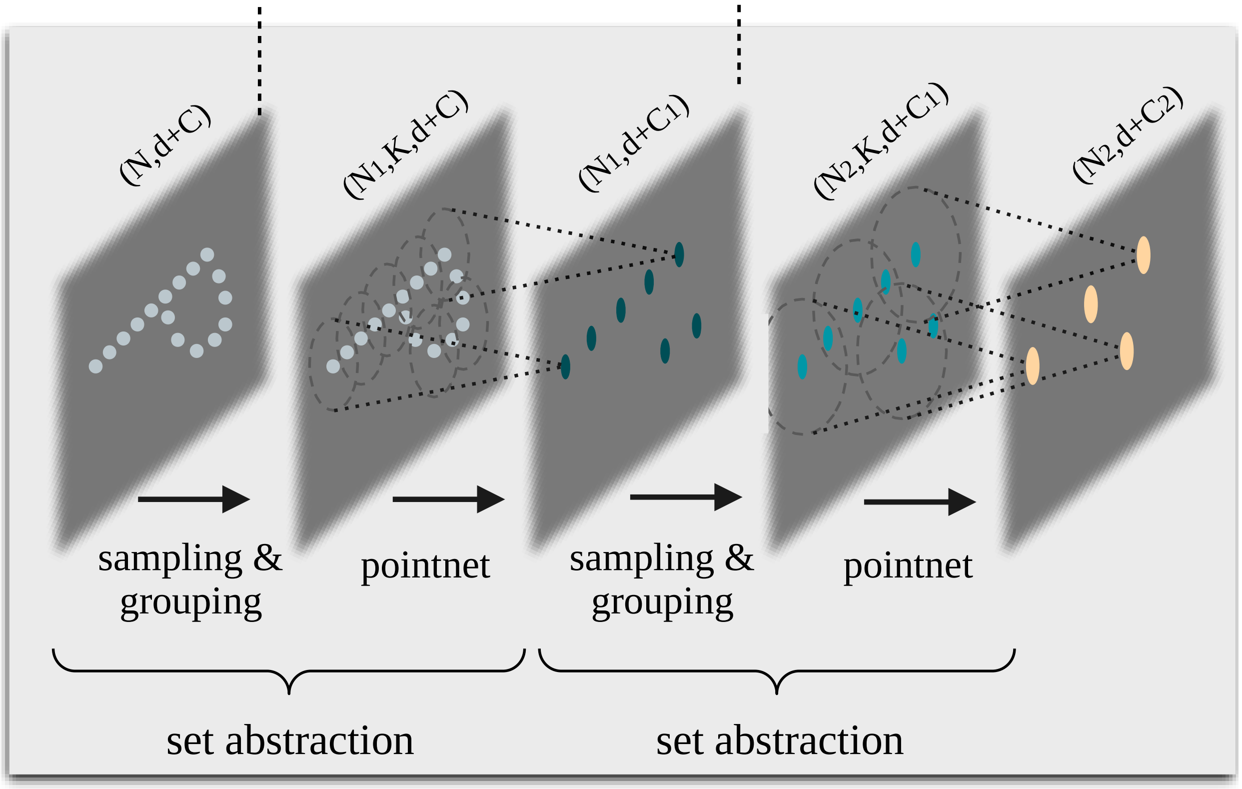


N1 points in (X,Y, **F**)

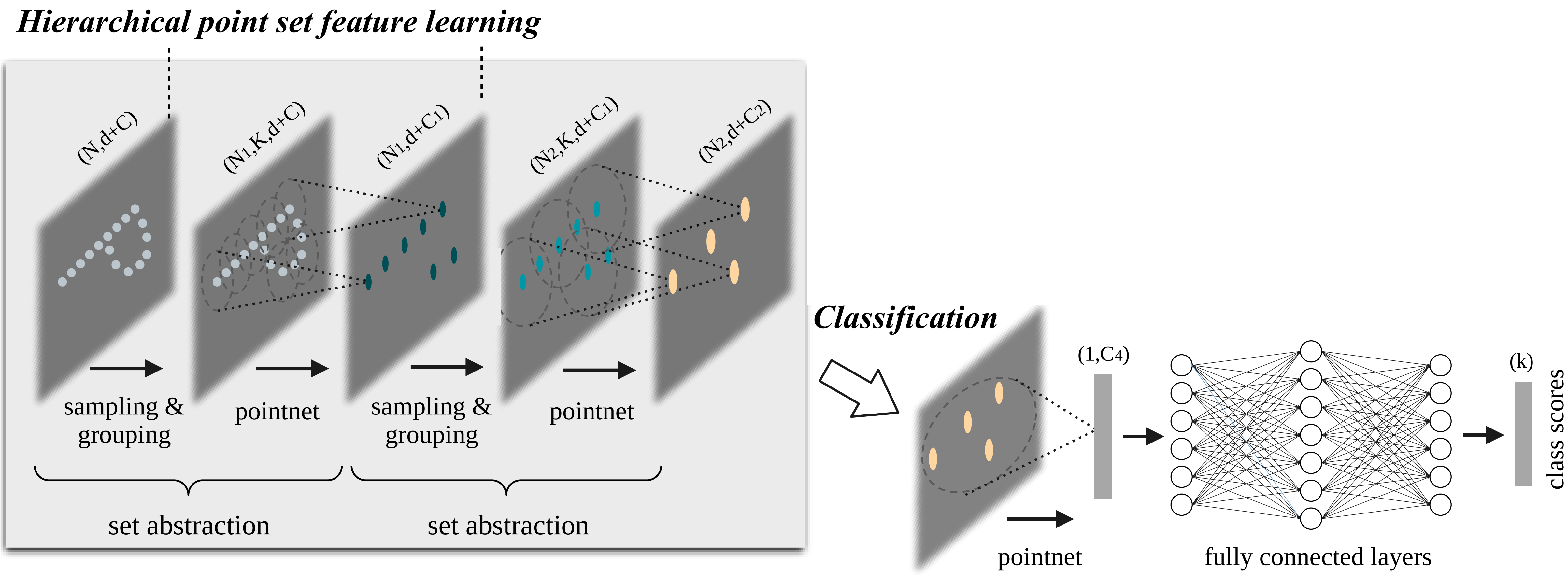
Set Abstraction: farthest point sampling + grouping + pointnet

PointNet++ for Classification and Segmentation

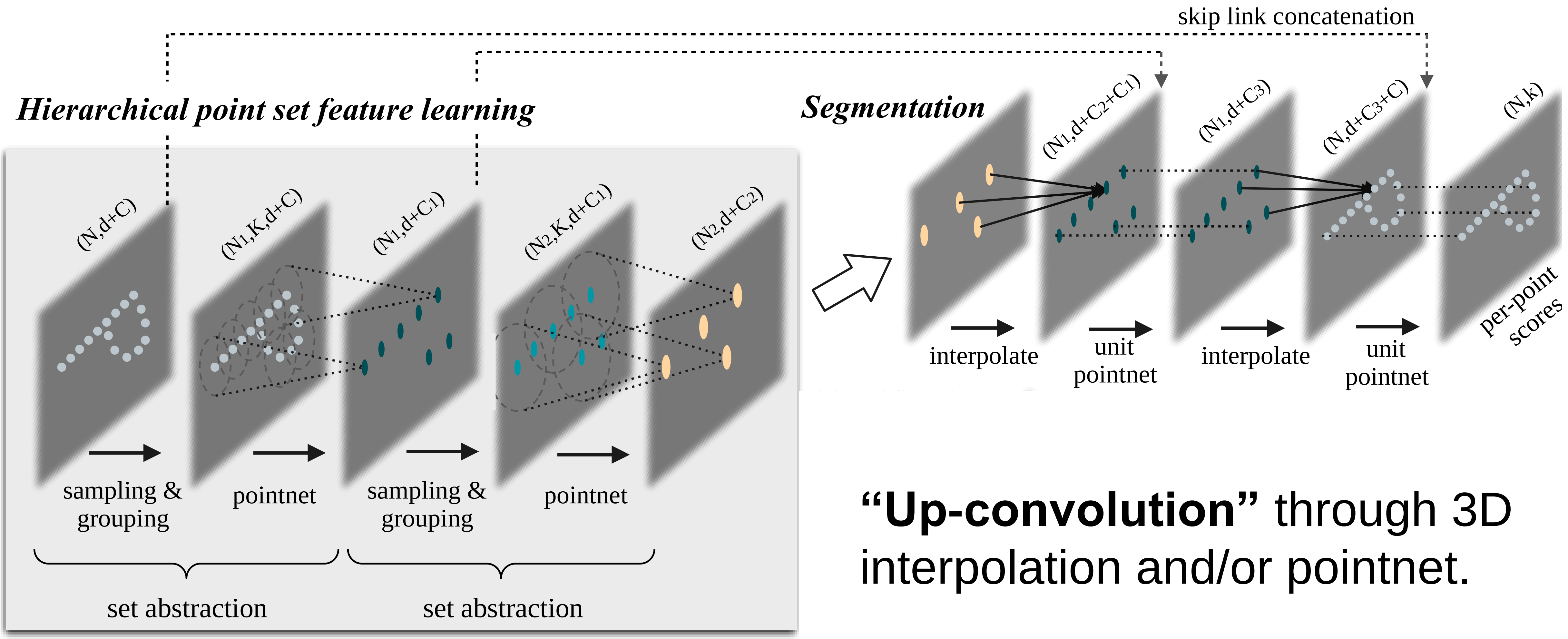
Hierarchical point set feature learning



PointNet++ for Classification and Segmentation



PointNet++ for Classification and Segmentation

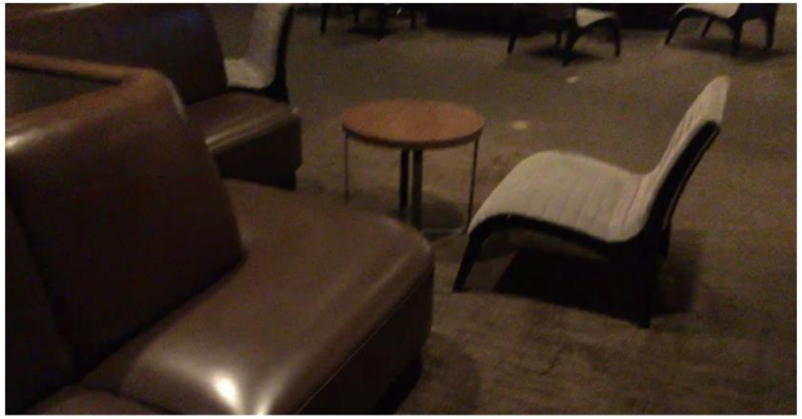


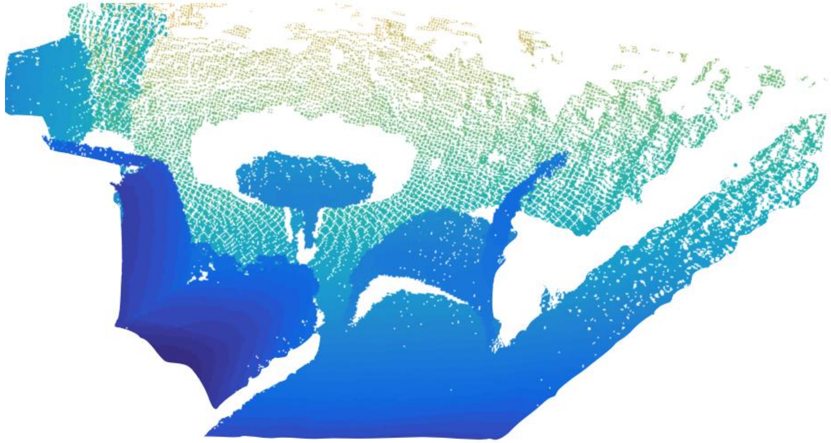
“Up-convolution” through 3D interpolation and/or pointnet.

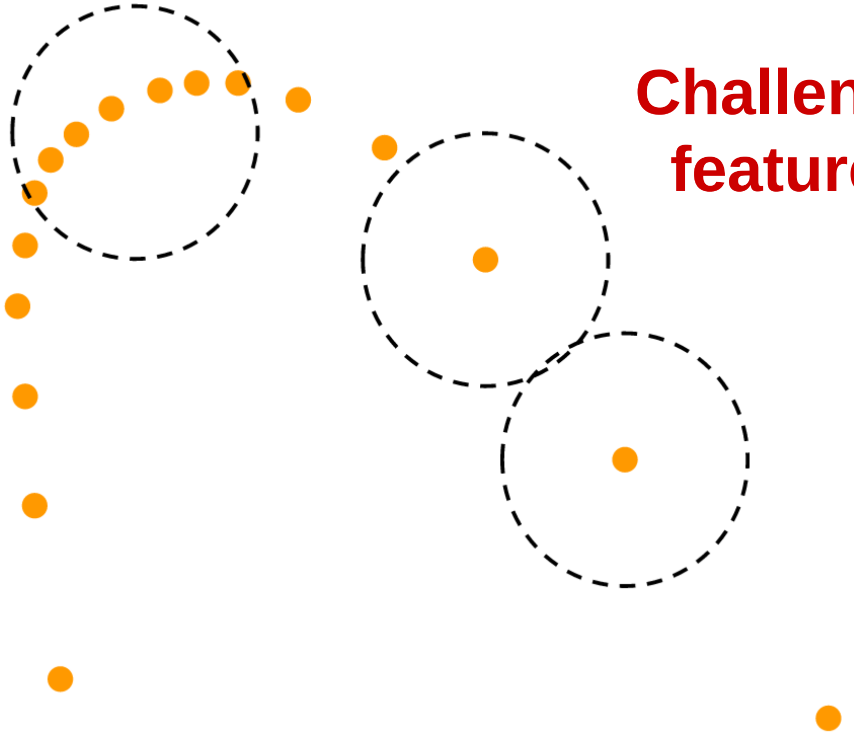
Non-uniform Sampling Density in Point Clouds

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

- perspective effect, radial density variation, motion etc.

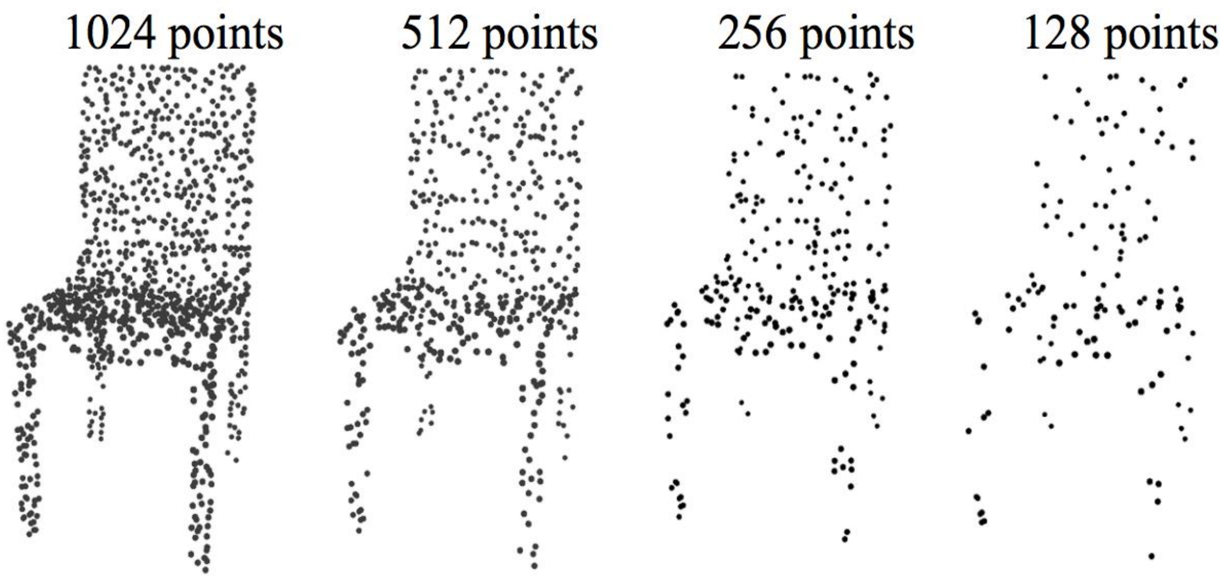


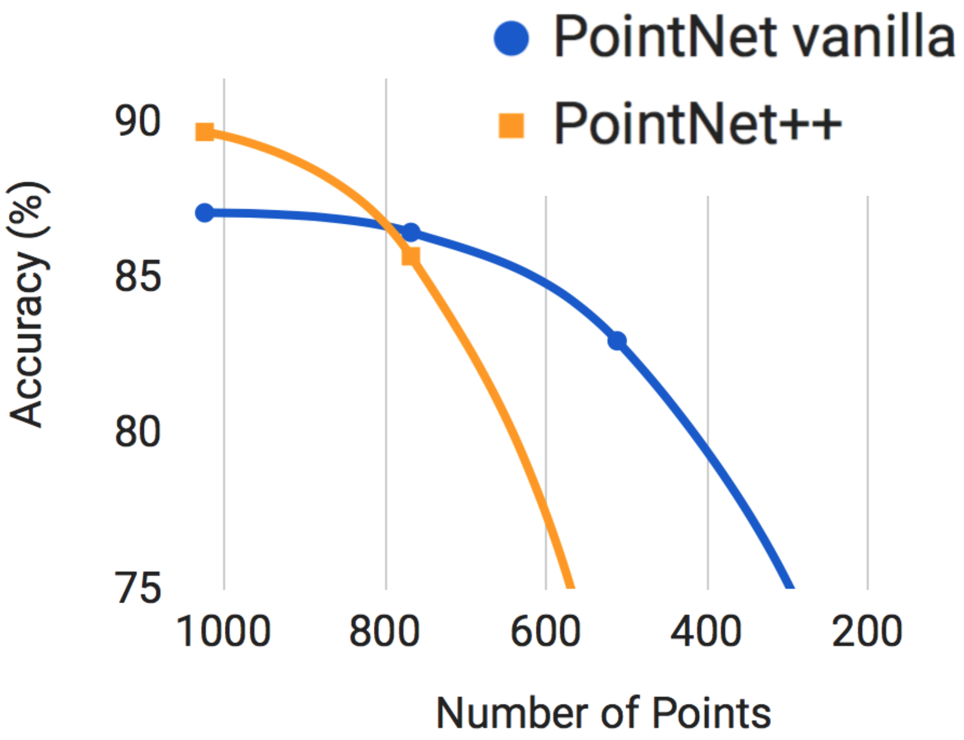




Challenge for local feature learning!

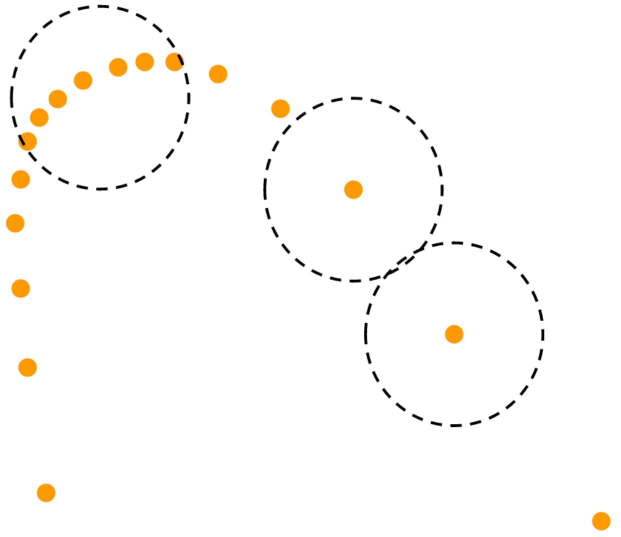
Density Variation Affects Hierarchy

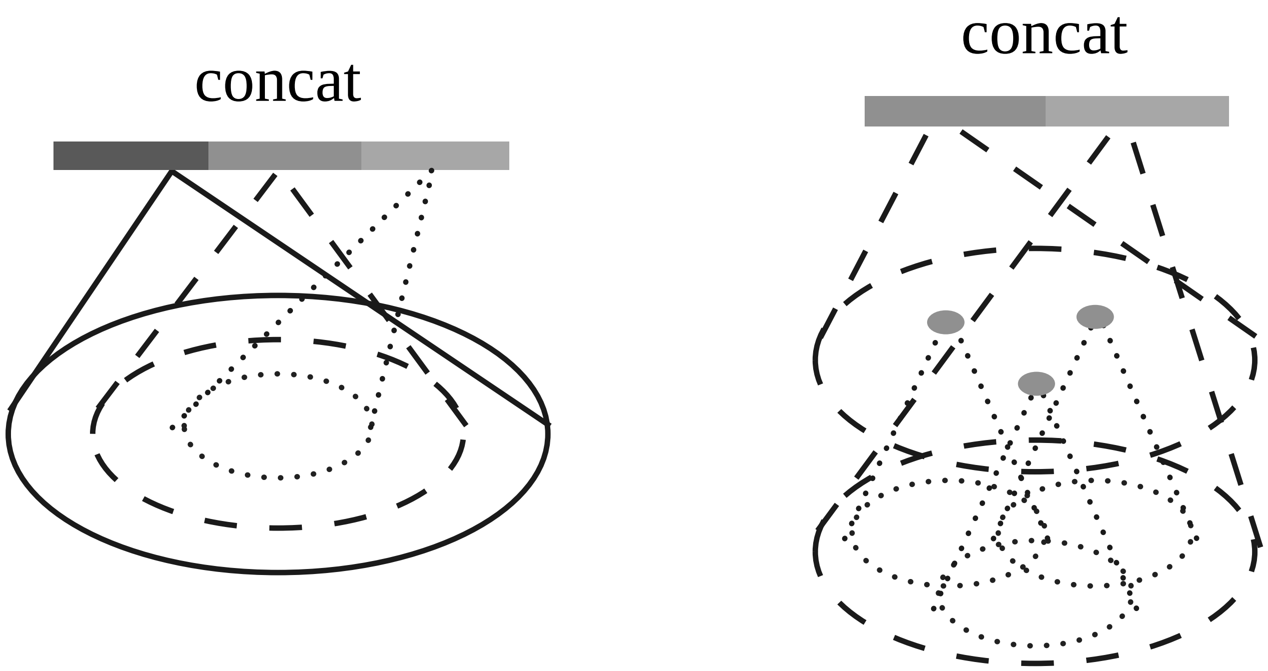




Small kernels suffer from varying densities!

Robust Learning Under Varying Sampling Density



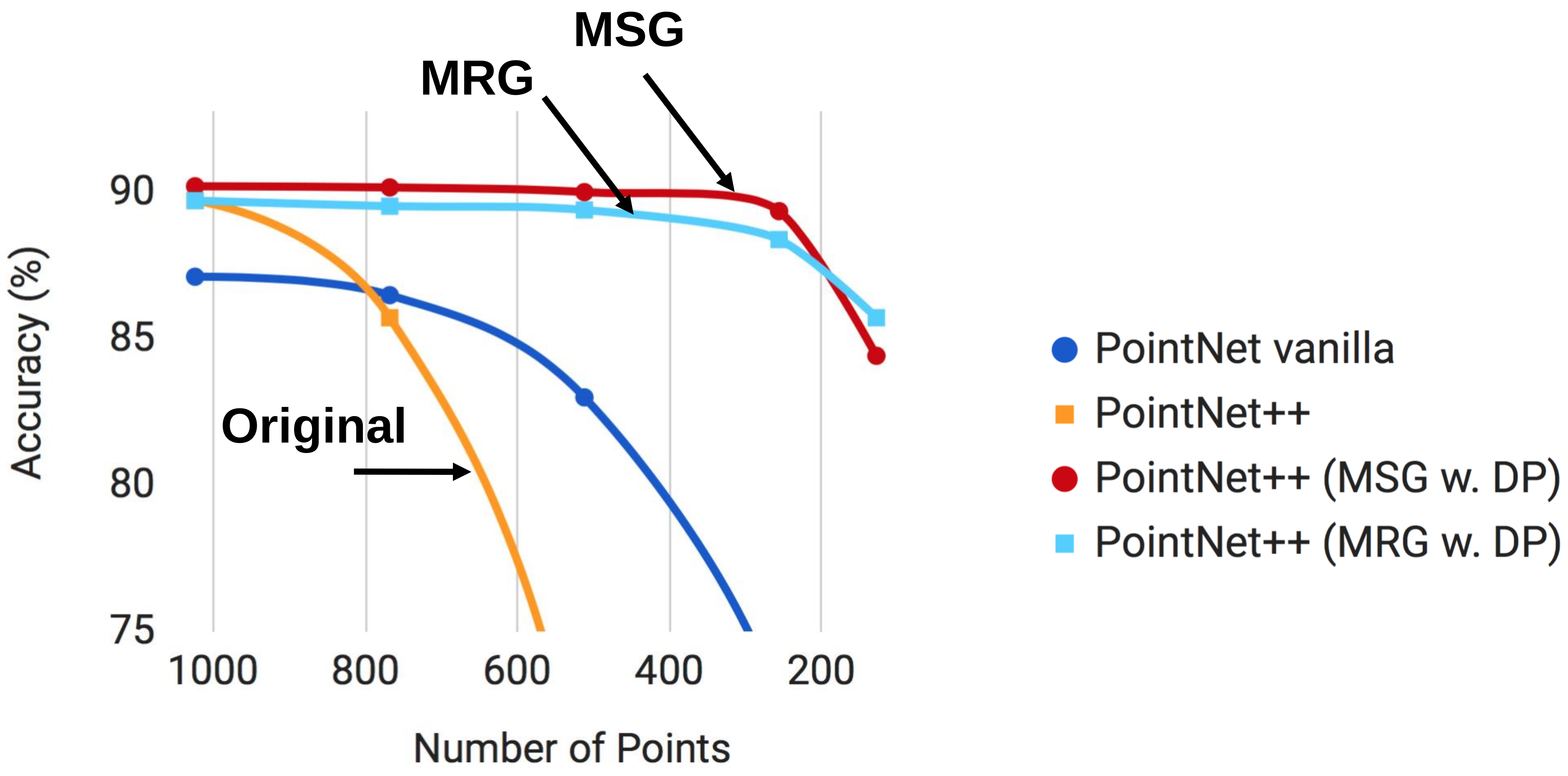


(a) (b)

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

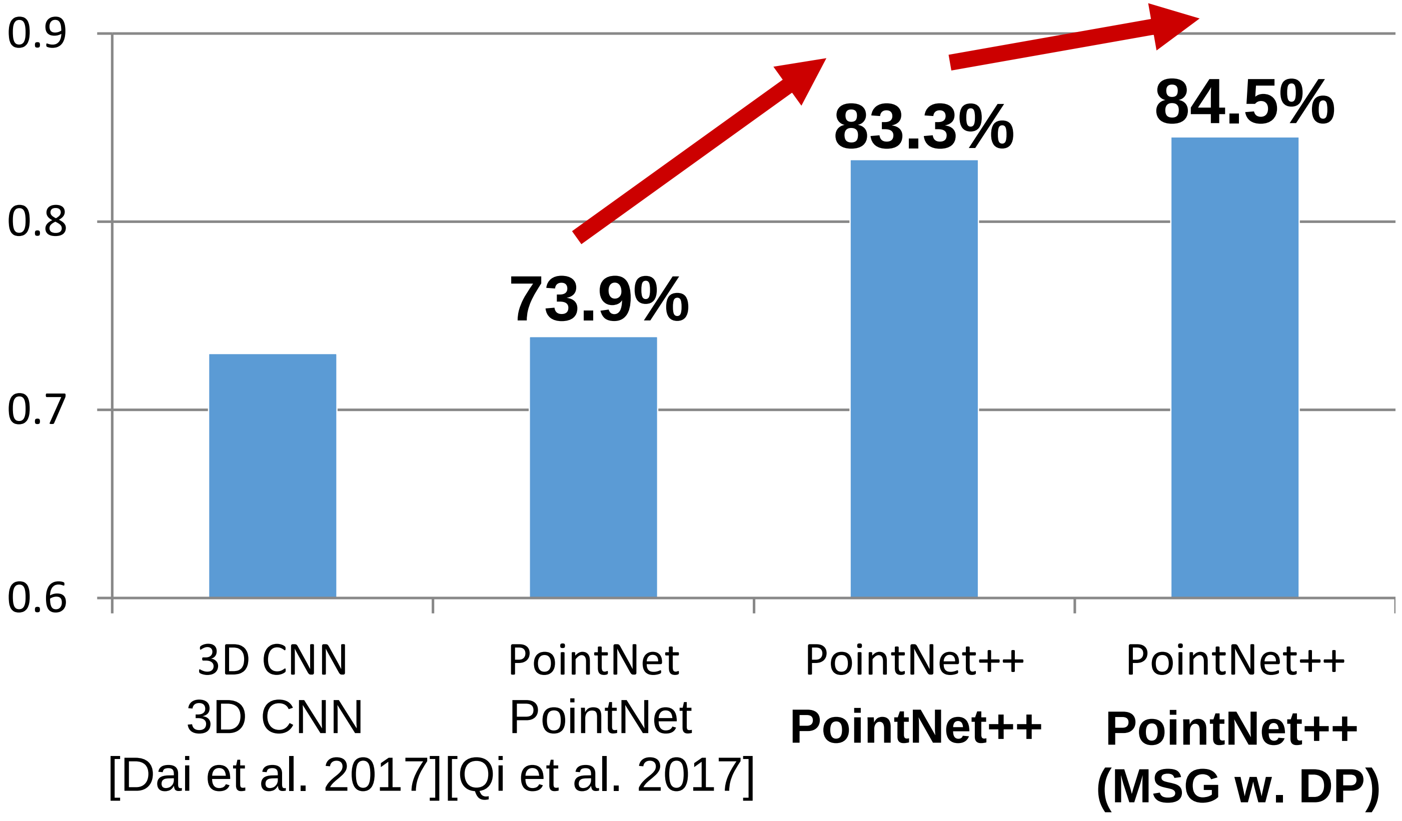
During Training: input point dropout with random dropout ratio

Robust Learning Under Varying Sampling Density



PointNet++ Results: Scene Parsing

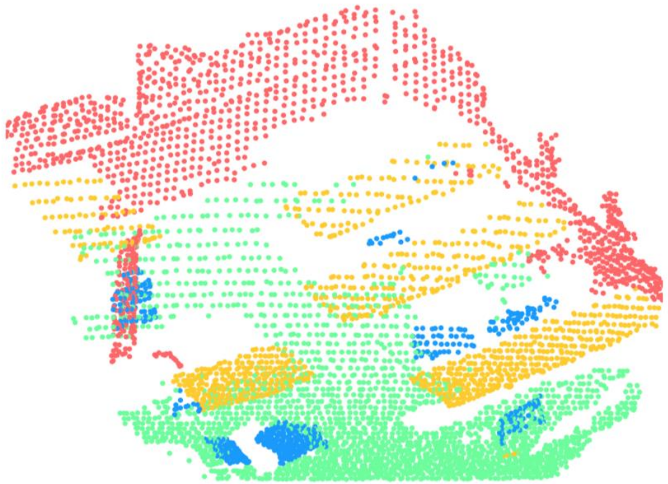
Better accuracy with hierarchical learning.

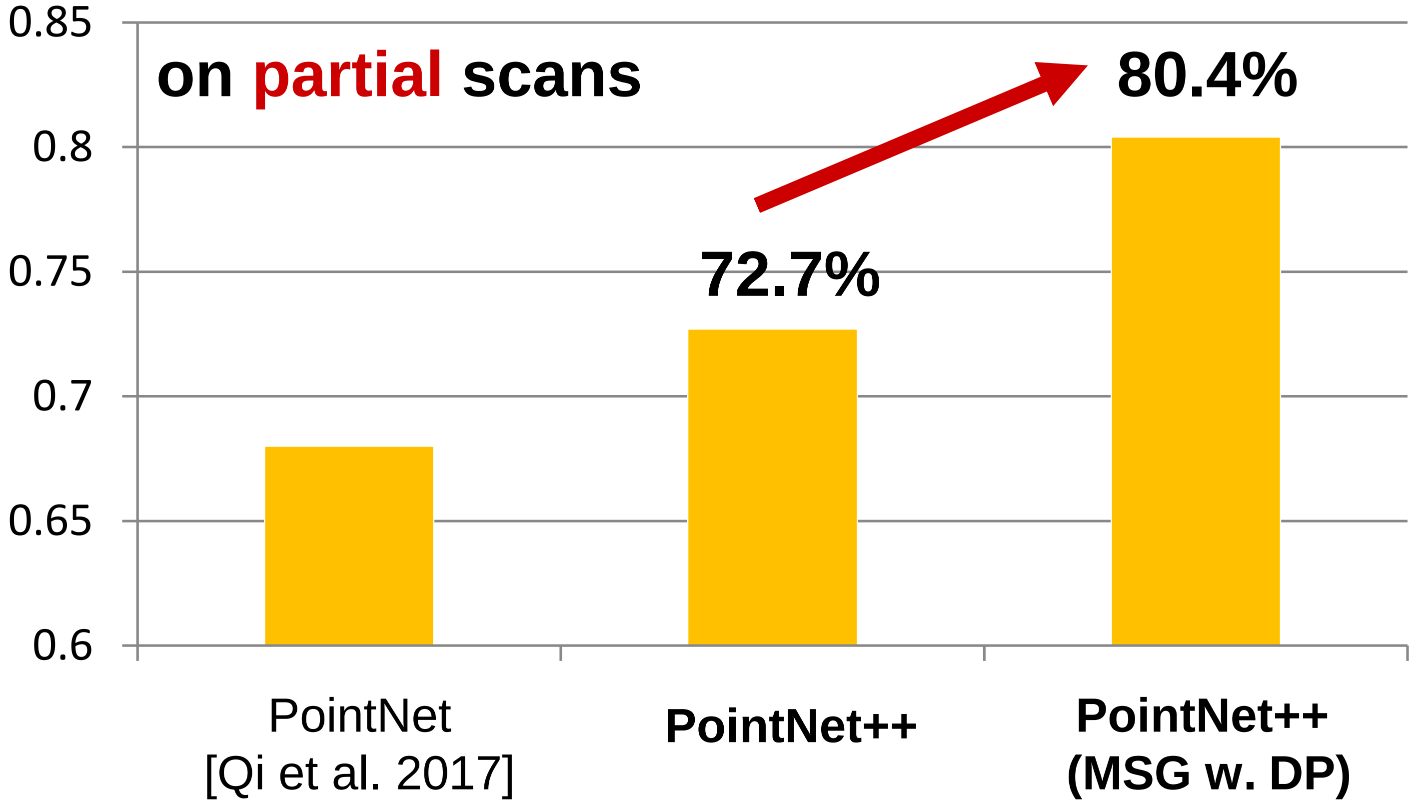


dataset: ScanNet; metric: per-point semantic classification accuracy (%)

PointNet++ Results: Scene Parsing

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



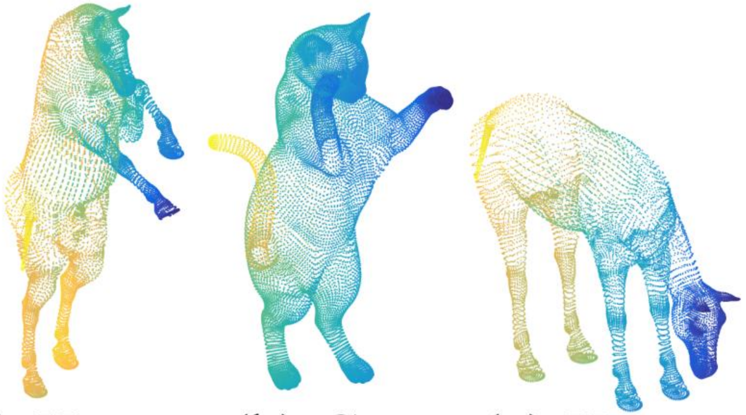


dataset: ScanNet; metric: per-point semantic classification accuracy (%)

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)



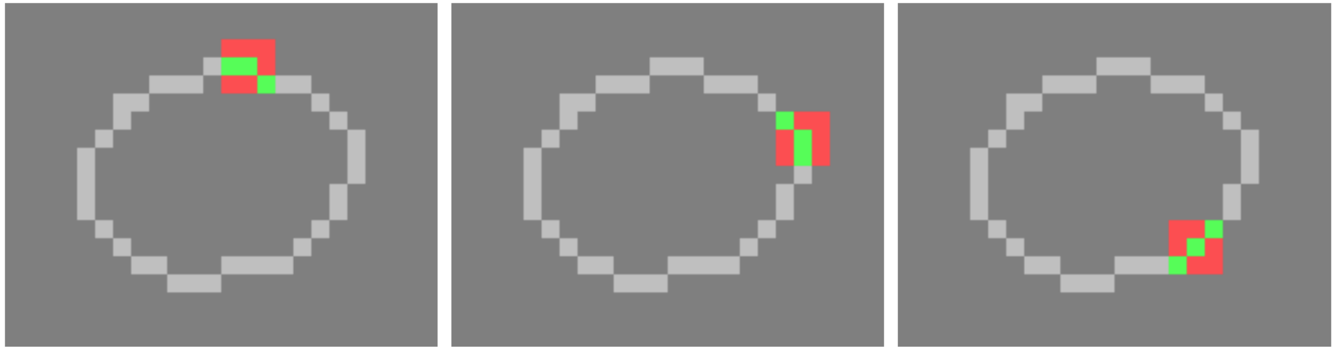
(a) Horse (b) Cat (c) Horse

	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 (%)

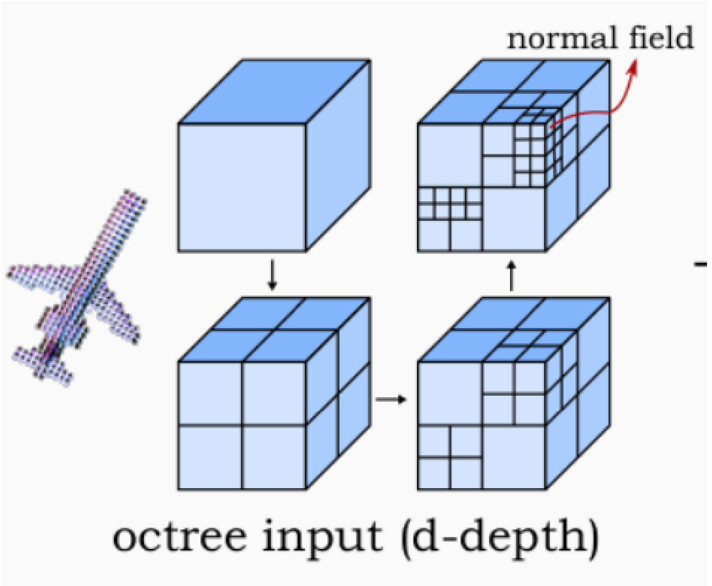
More Types of
Deep Networks Related to Point Clouds

Sparse 3D CNNs



Submanifold Sparse Convolutional Networks

[Graham et al. 2017]



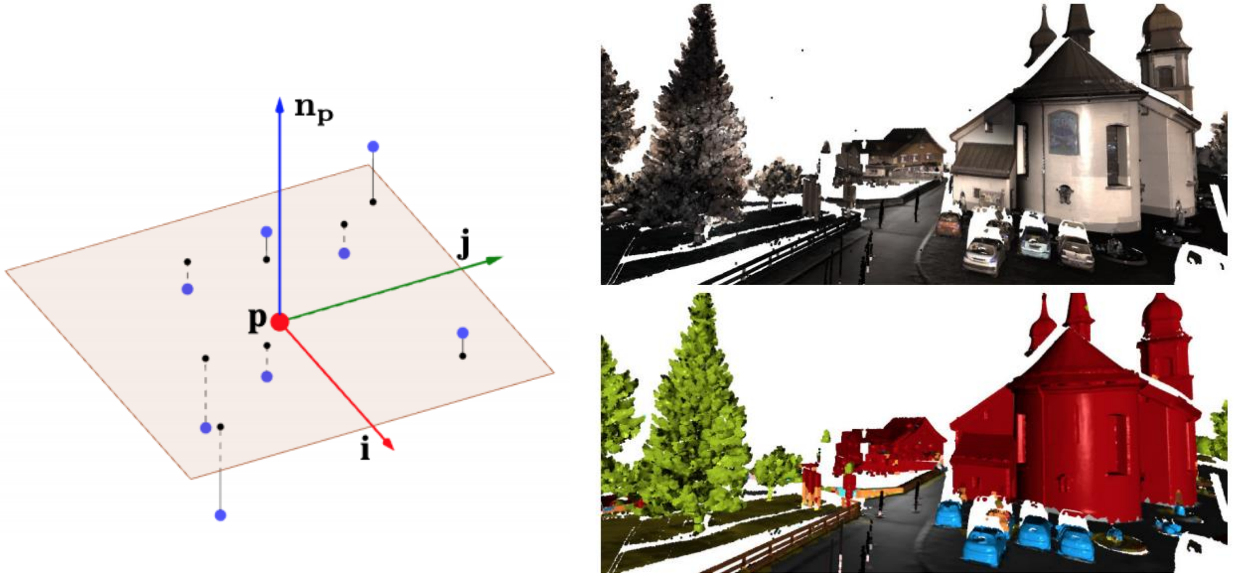
OctNet

[Riegler et al. 2017]

O-CNN: Octree based Convolutional Neural Networks

[Wang et al. 2017]

Surface-Based Convolutions

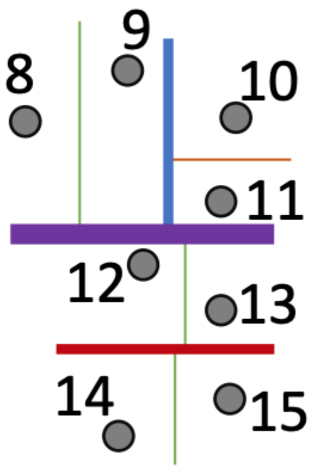


Surface Convolution

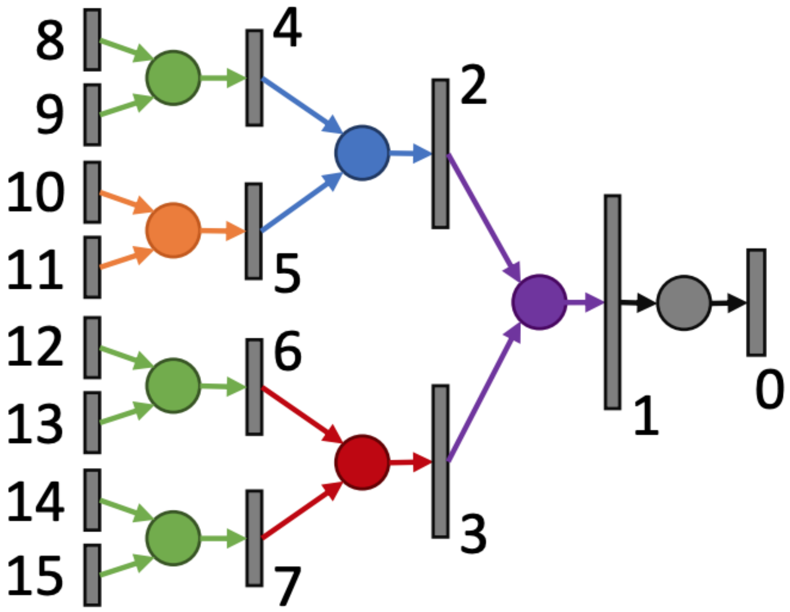
Tangent Convolutions
[Tatarchenko et al. 2018]

SurfConv
[Chu et al. 2018]

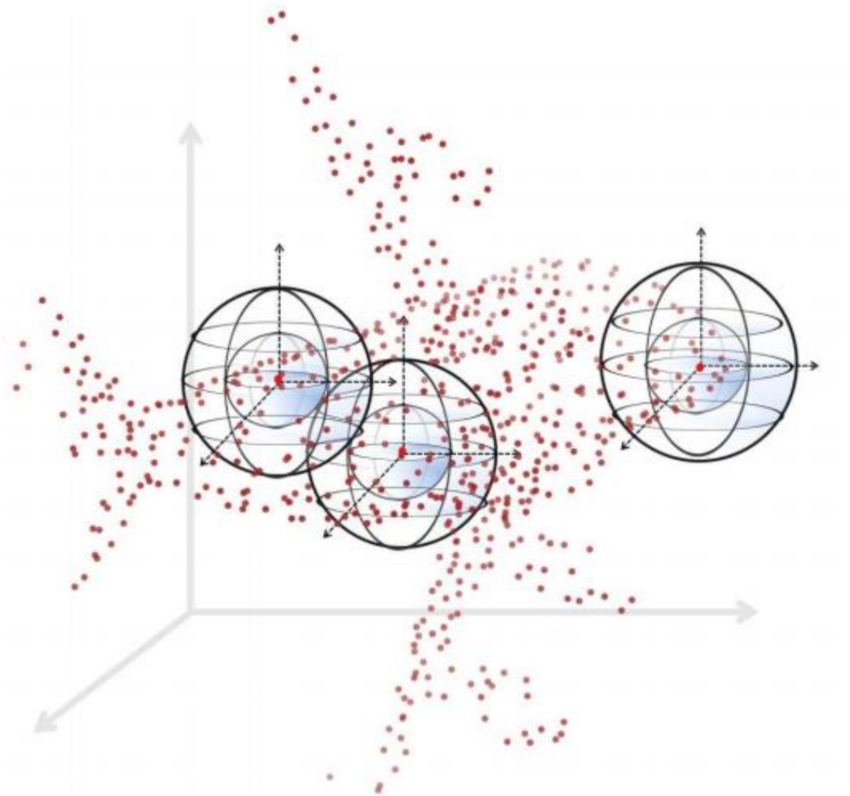
Classic Spatial Representations + NN



kd-tree



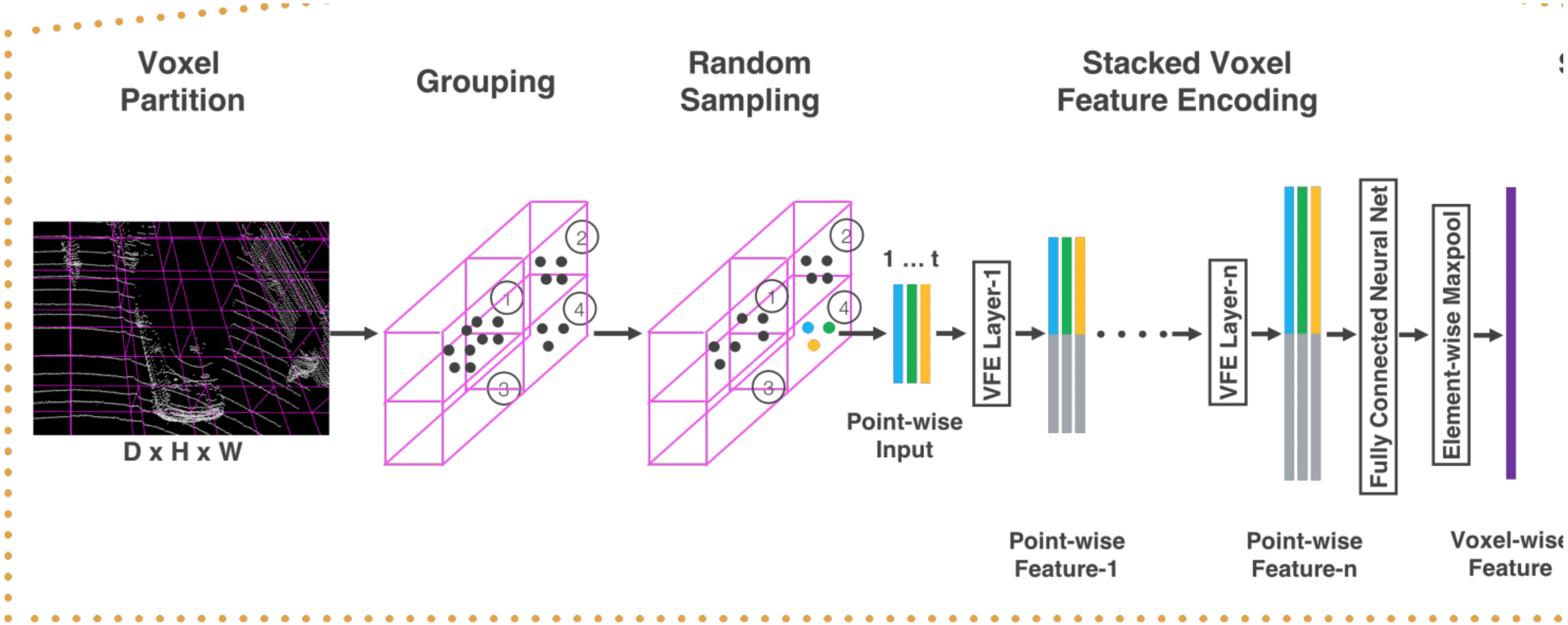
Kd-Network
[Klokov et al. 2017]

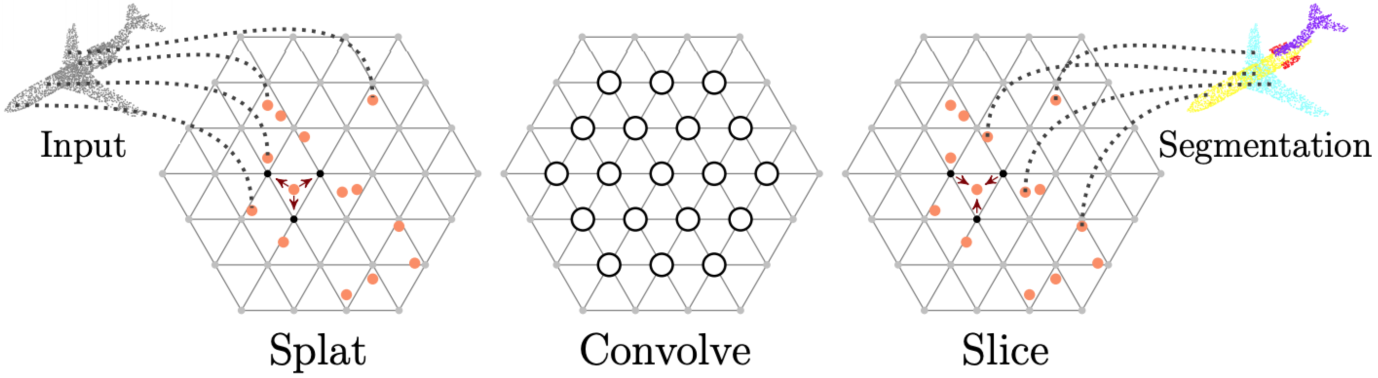


ShapeContextNet
[Xie et al. 2018]

Hybrid Networks: Grids + Points

VoxelNet
[Zhou et al. 2018]

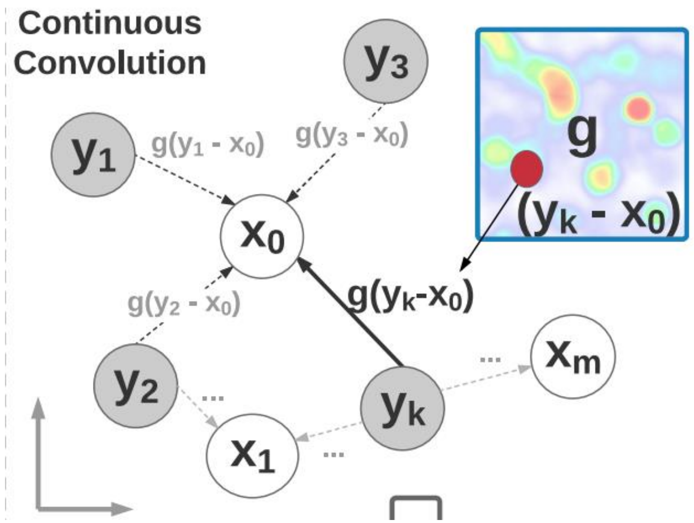




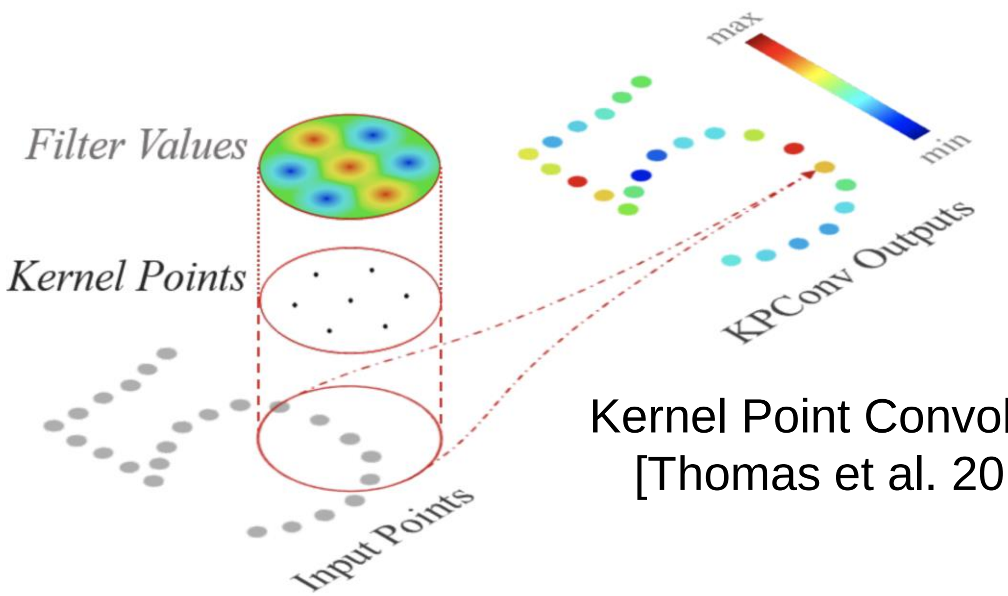
Bilateral convolution layers on a sparse lattice

SPLATNet
[Su et al. 2018]

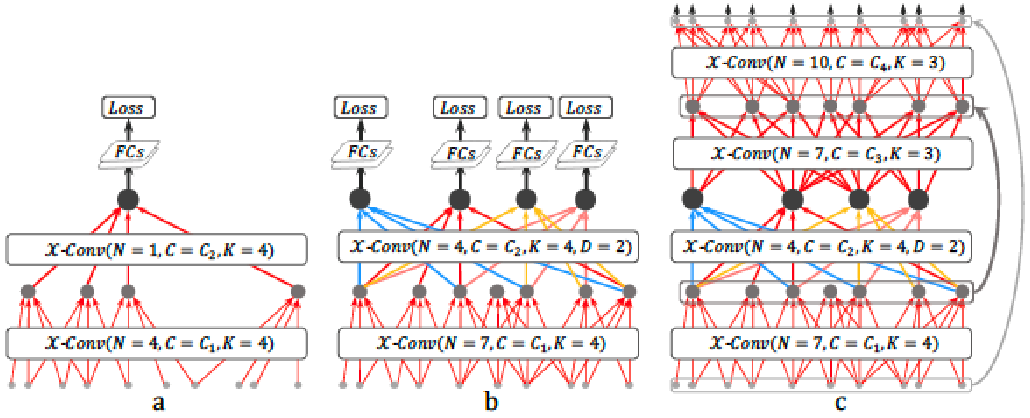
Point Cloud Convolution Variants



Deep Parametric Continuous Convolution
[Wang et al. 2018]

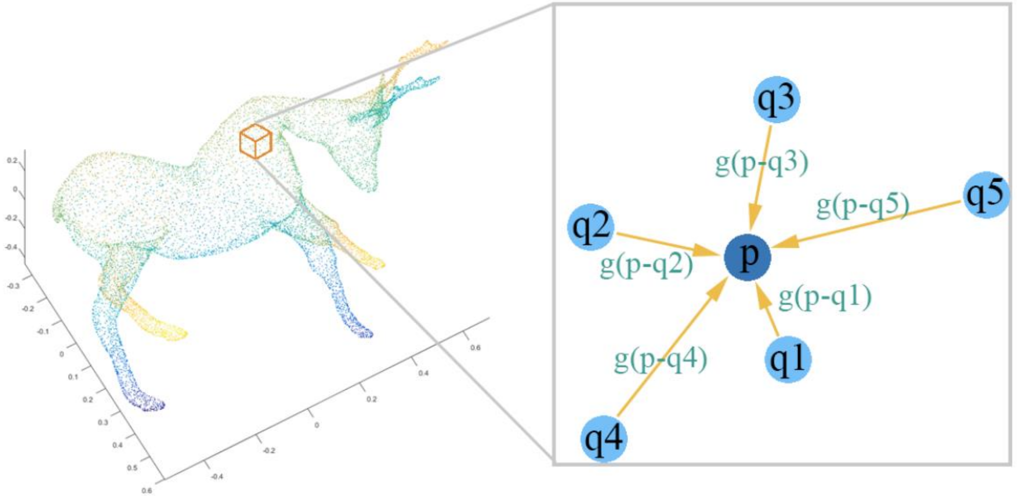


Kernel Point Convolution
[Thomas et al. 2019]

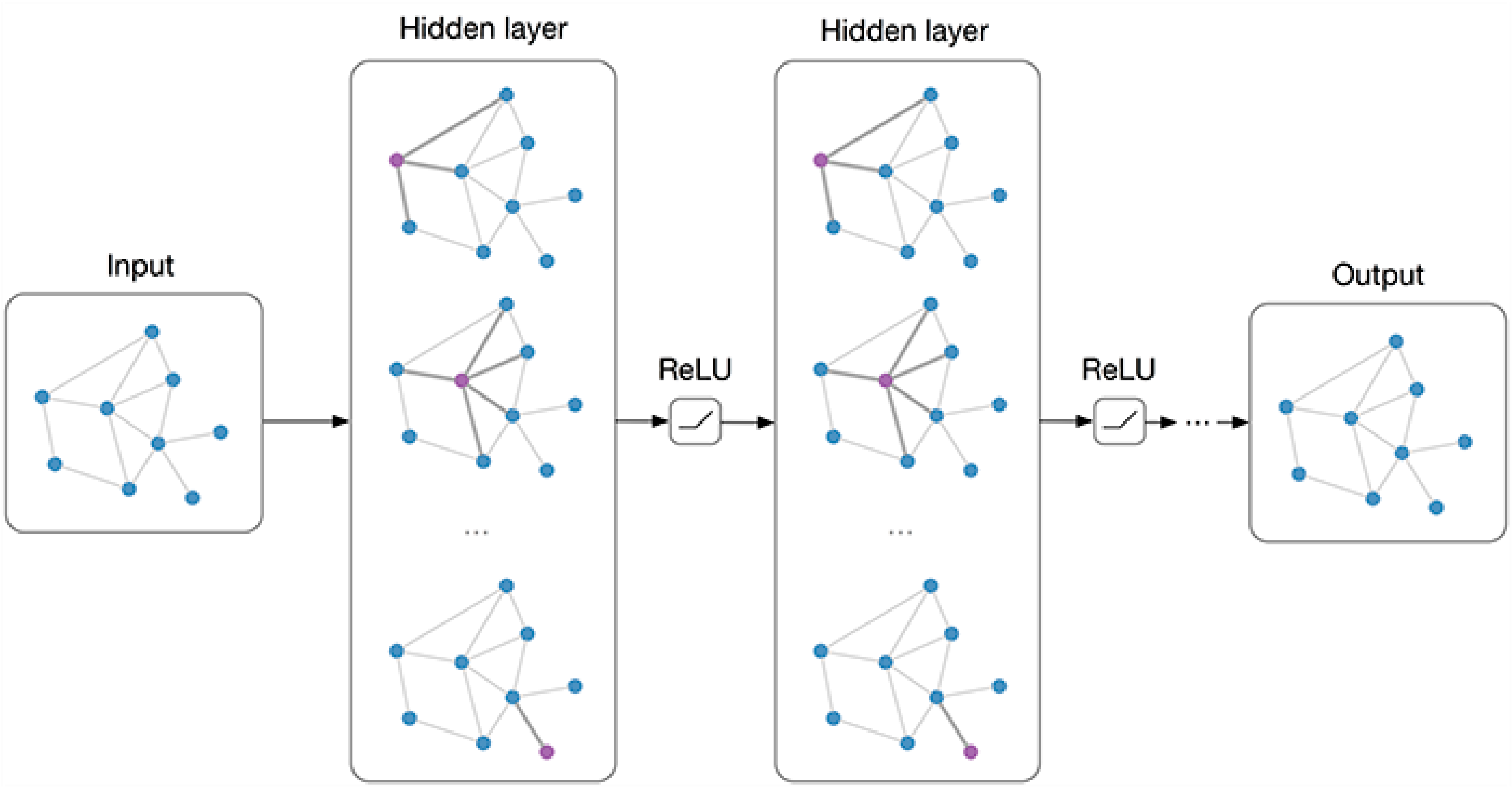


PointCNN
[Li et al., 2018]

SpiderCNN
[Xu et al. 2018]



Graph Neural Networks



https://tkipf.github.io/graph-convolutional-networks/

Which Network Architecture Is Best?

- Any distance metric among points?
- 3D points or higher-dim points?
- Single object or multi-object?
- Depth image or fused point clouds?
- Care about efficiency?

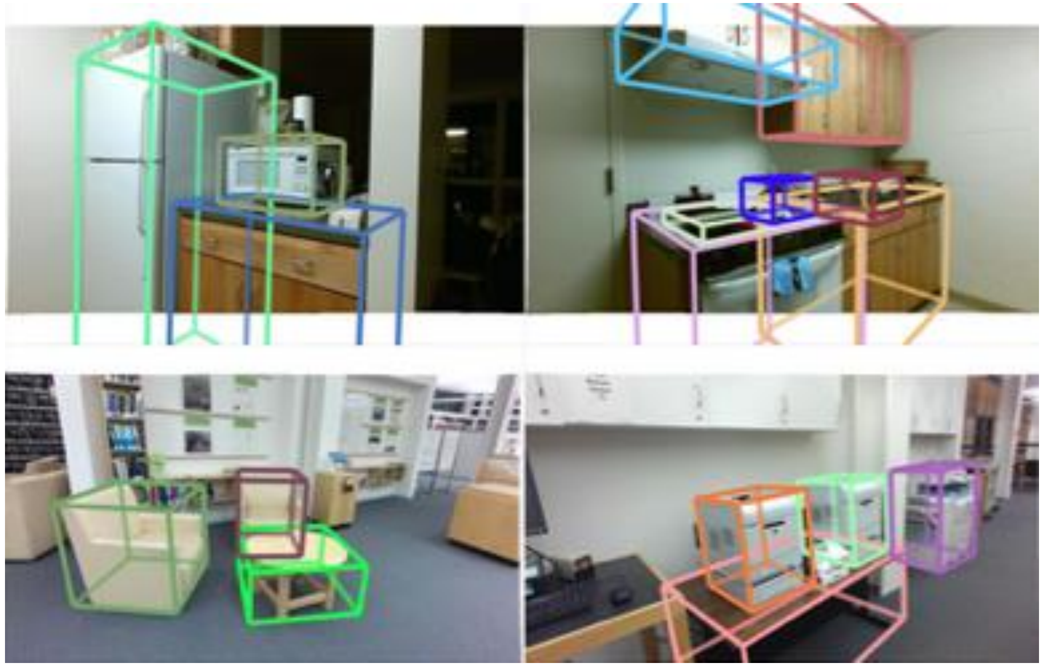
Is there a universally best architecture?

3D Scene Understanding with PointNet and PointNet++

3D Scene Understanding with PointNets

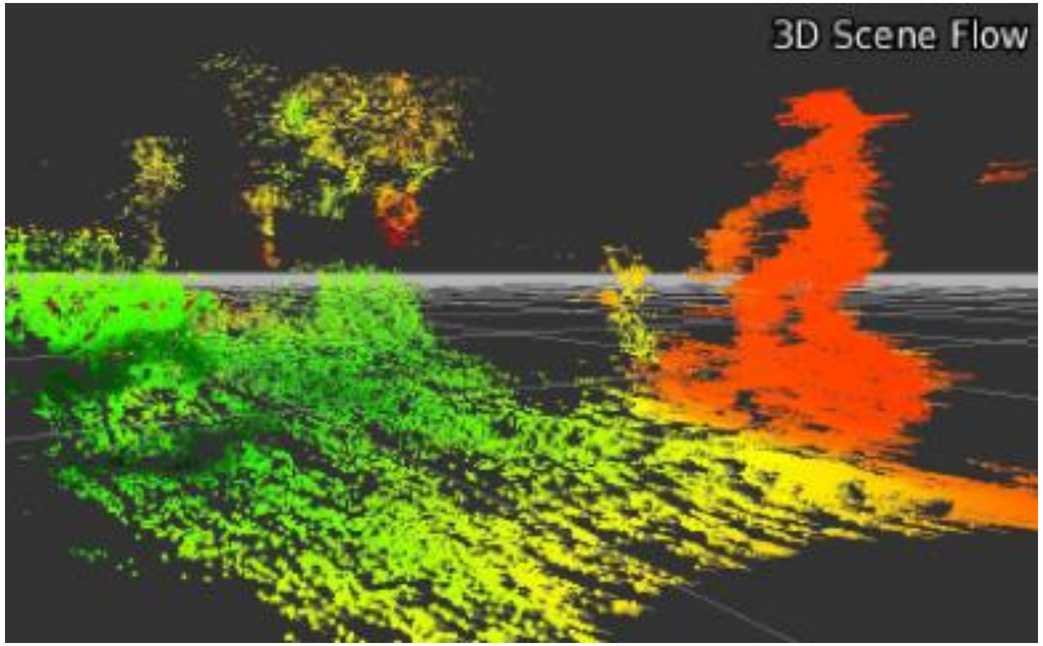
- PointNet and PointNet++ lead to new **3D centric approaches** to scene understanding

3D Object Detection



source: SUN RGB-D by Song et al.

3D Scene Flow

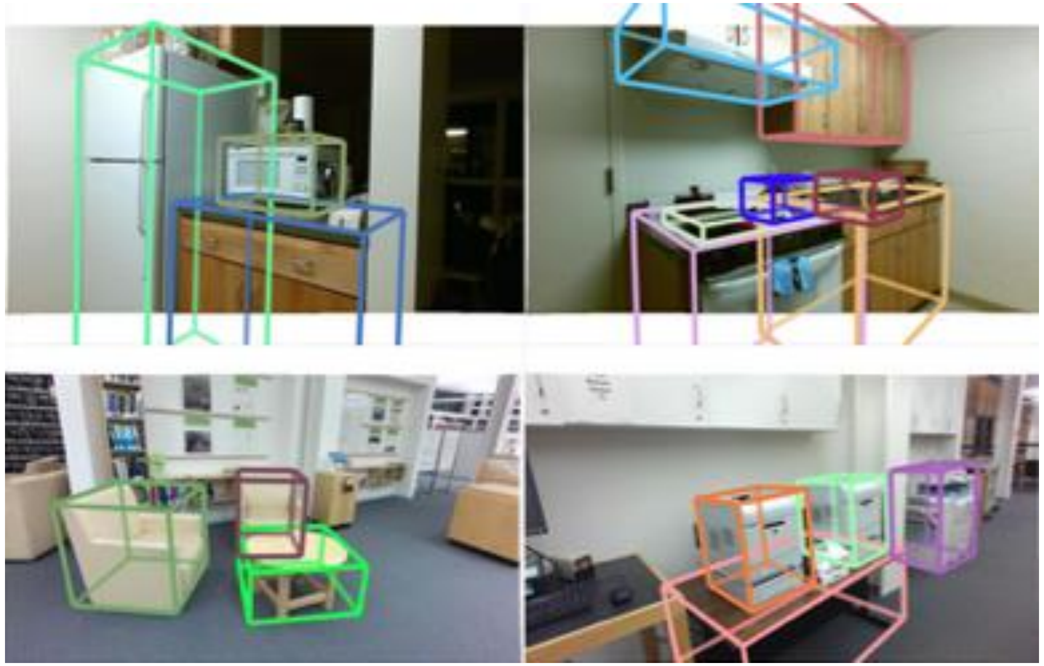


source: Wedel et al.

3D Scene Understanding with PointNets

- PointNet and PointNet++ lead to new **3D centric approaches** to scene understanding

3D Object Detection



source: SUN RGB-D by Song et al.

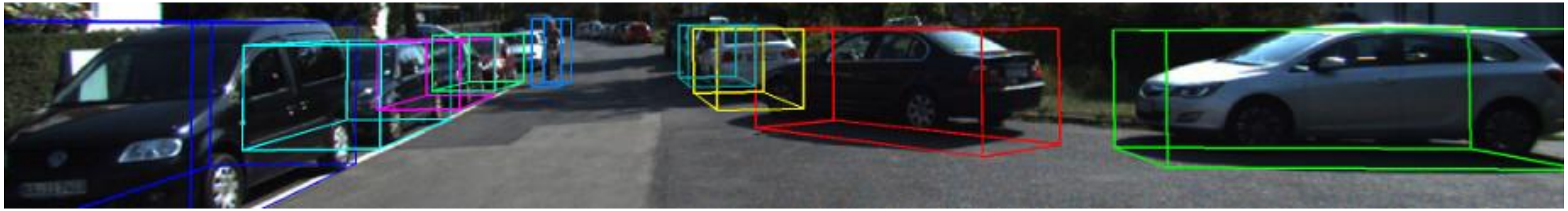
3D Scene Flow

source: Wedel et al.

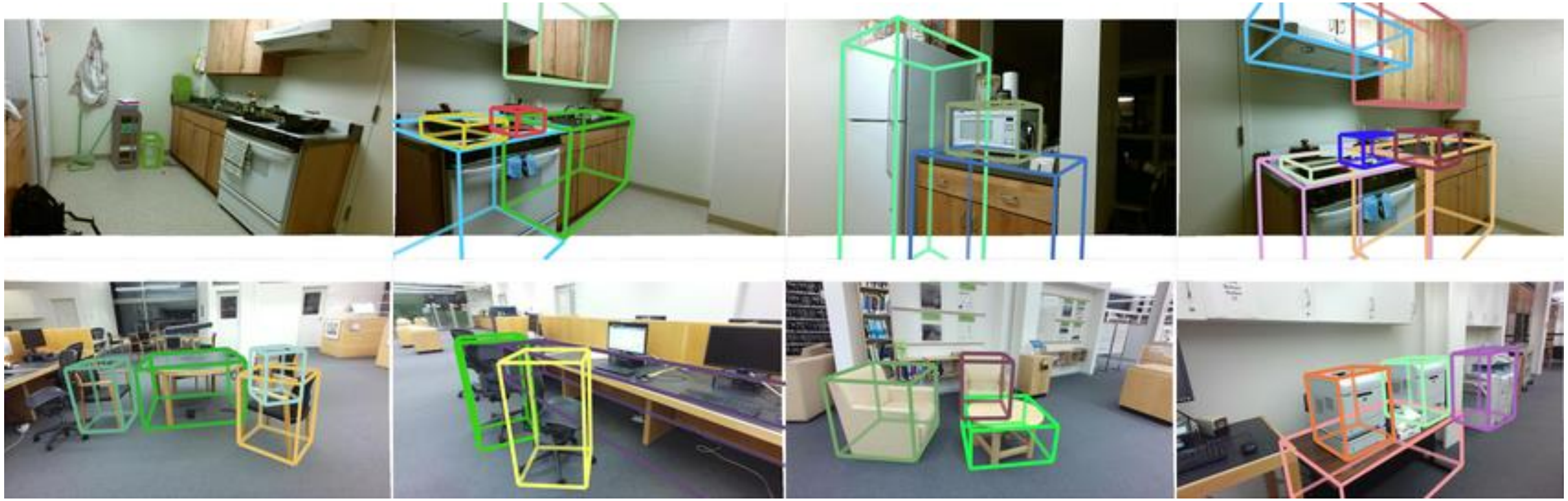
3D Object Detection

- Input: RGB-D data
- Output: 3D bounding boxes of objects

KITTI:



SUN RGB-D:



3D Object Detection

- Input: RGB-D data
- Output: 3D (amodal) bounding boxes of objects

KITTI:

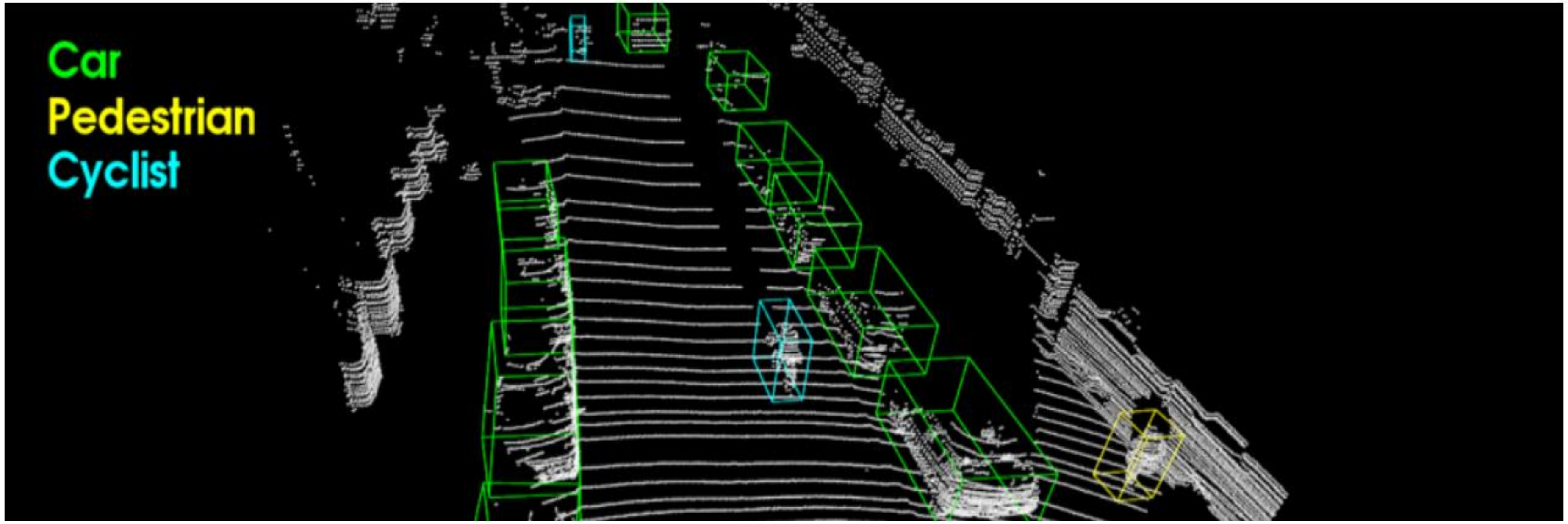
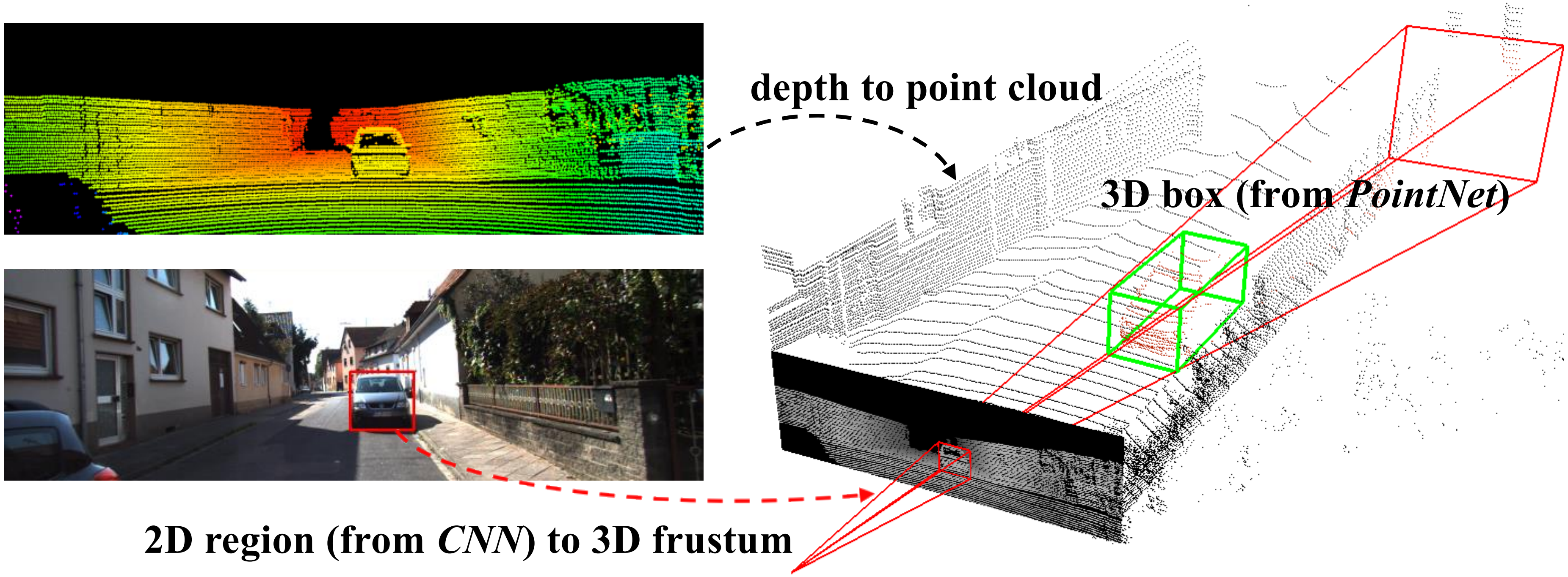


Figure from VoxelNet [Zhou et al. 2018]

Frustum PointNets for 3D Object Detection

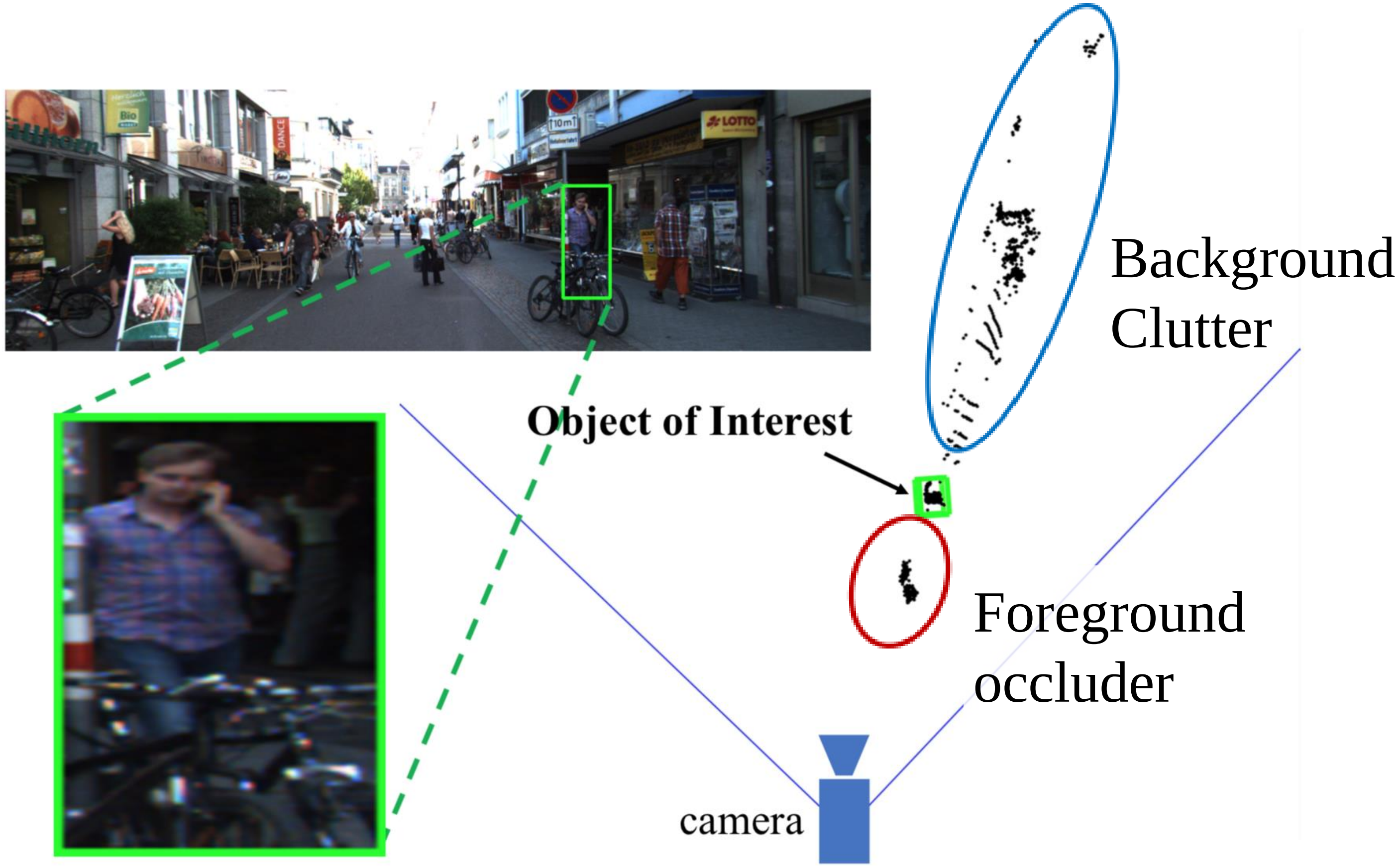


+ *Leveraging mature 2D detectors* *for region proposal. greatly reducing 3D search space.*

+ *Solving 3D detection problem with 3D data and 3D deep learning.*

Charles R. Qi, Wei Liu, Chenxia Wu, Hao Su, Leonidas Guibas. Frustum PointNets for 3D Object Detection from RGB-D Data (CVPR 2018)

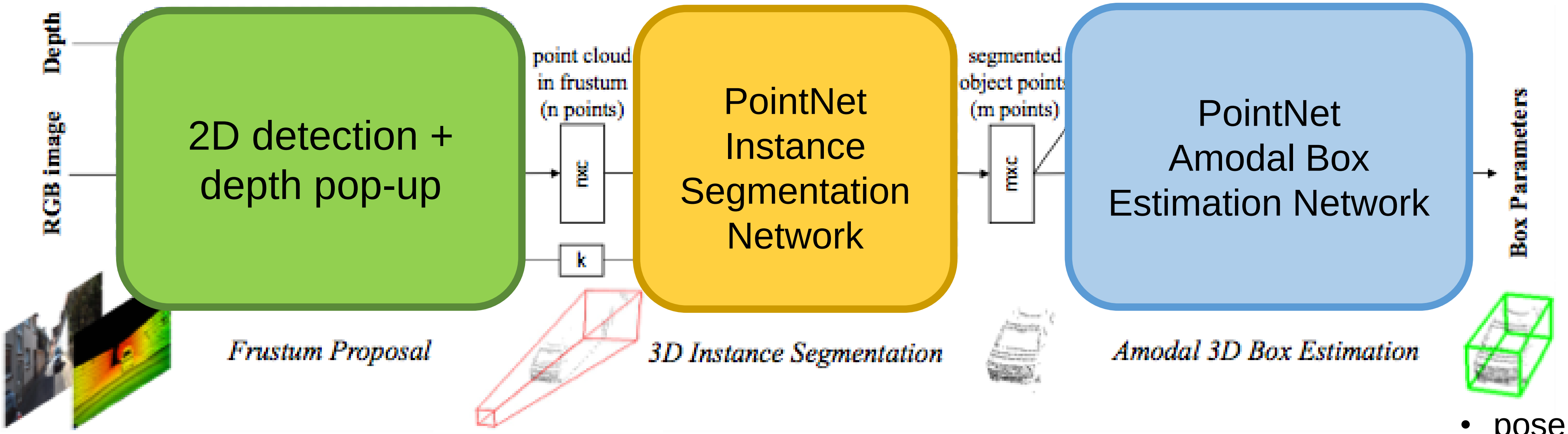
Frustum-based 3D Object Detection: Challenges



- Occlusion and clutter is common in frustum point clouds
- Large range of point depths

Frustum PointNets

Use **PointNets** for **data-driven** object detection in frustums.



- pose
- size
- center

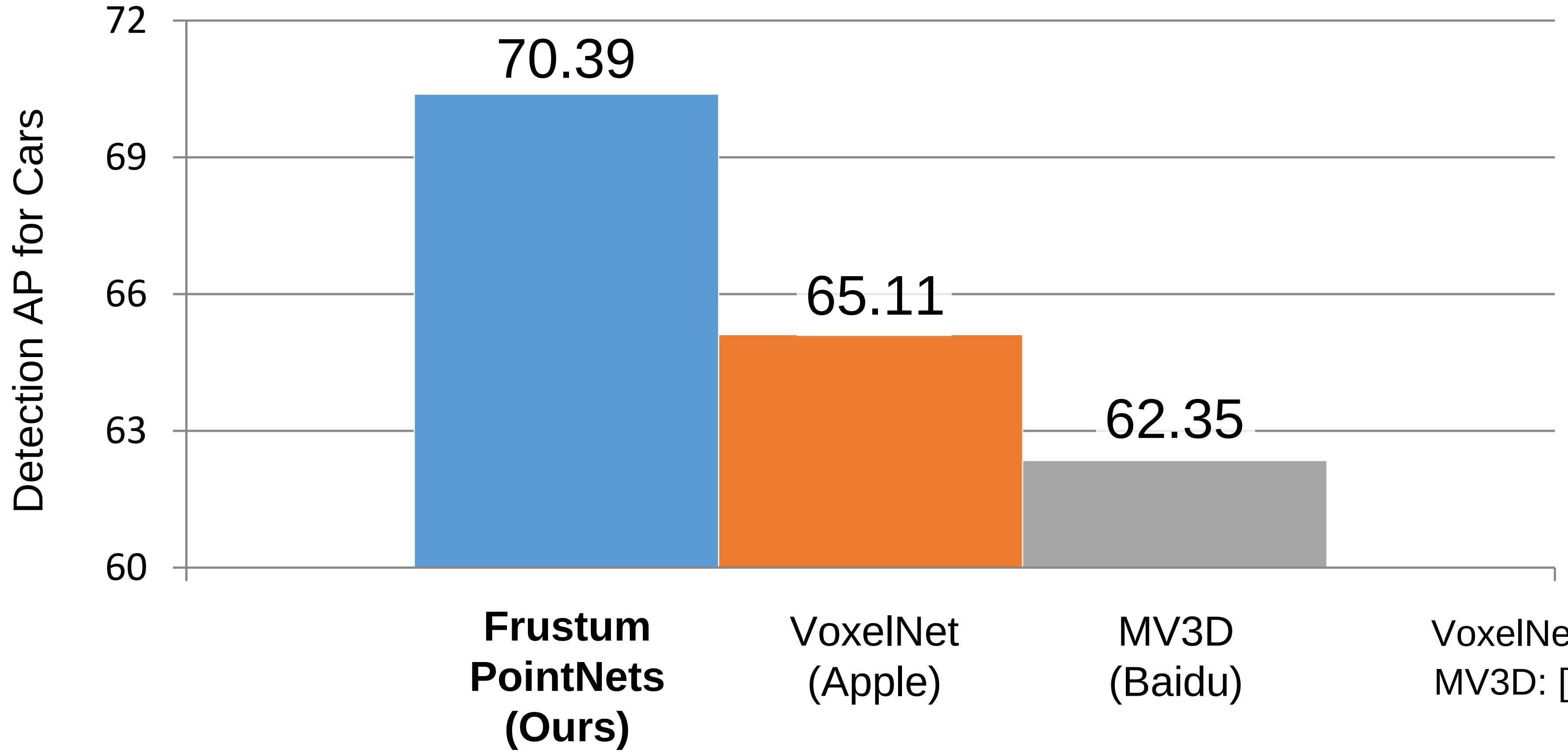
Frustum PointNets: Key to Success

Respect and exploit 3D

- **Use each modality (image, points) for what it's best at** — using 3D representation and 3D deep learning for the 3D problem.
- **Canonicalize the problem** — exploiting geometric transformations in point clouds.

KITTI Results: Quantitative

Leading performance on KITTI benchmark

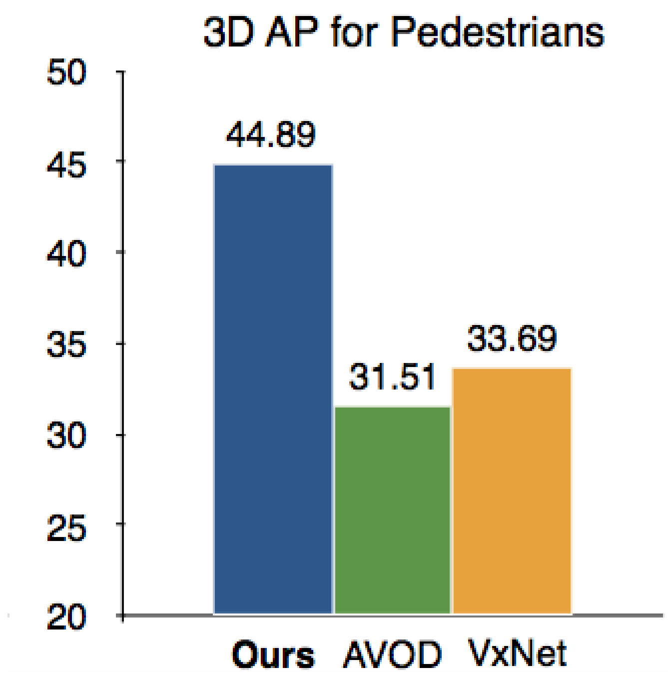


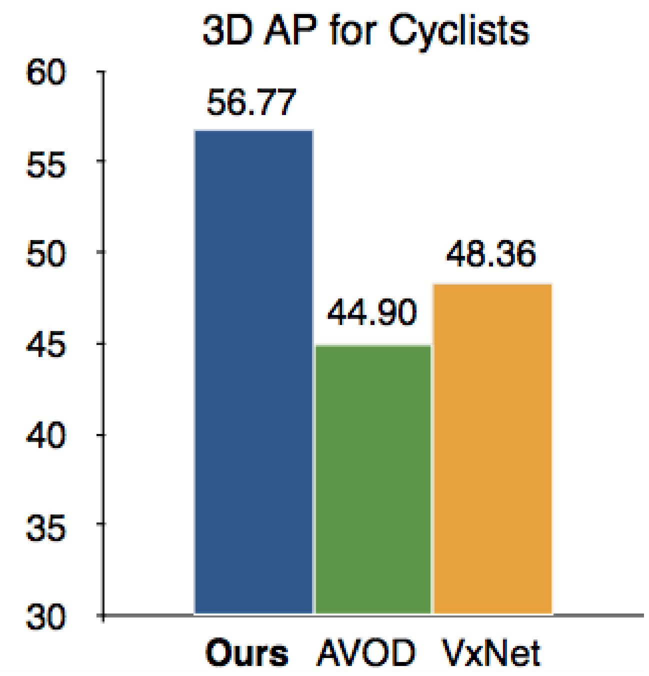
VoxelNet: [Zhou et al. 2018]
MV3D: [Chen et al. 2017]

KITTI Results: Quantitative

Leading performance on KITTI benchmark

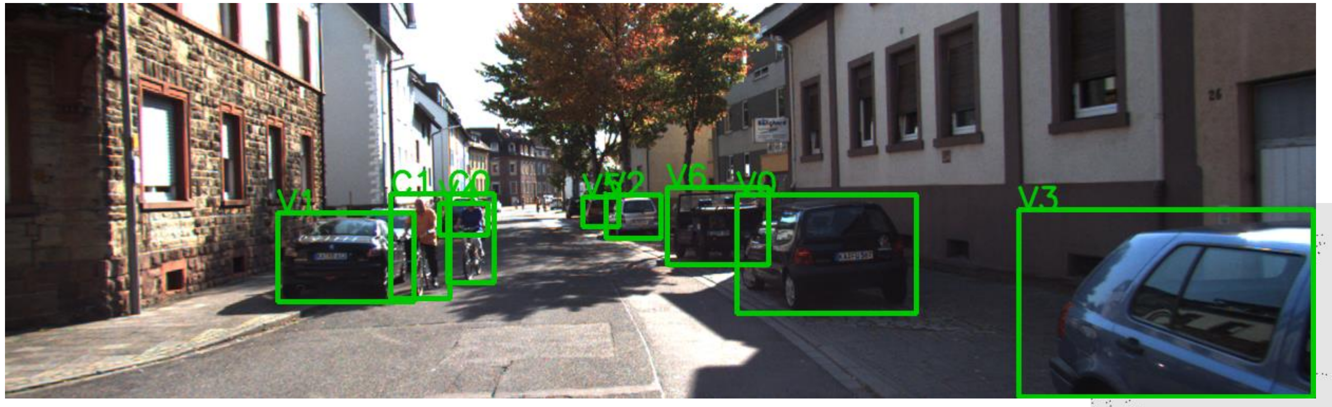
Especially leading at smaller objects (pedestrians and cyclists)
– hard to localize with 3D proposals only.

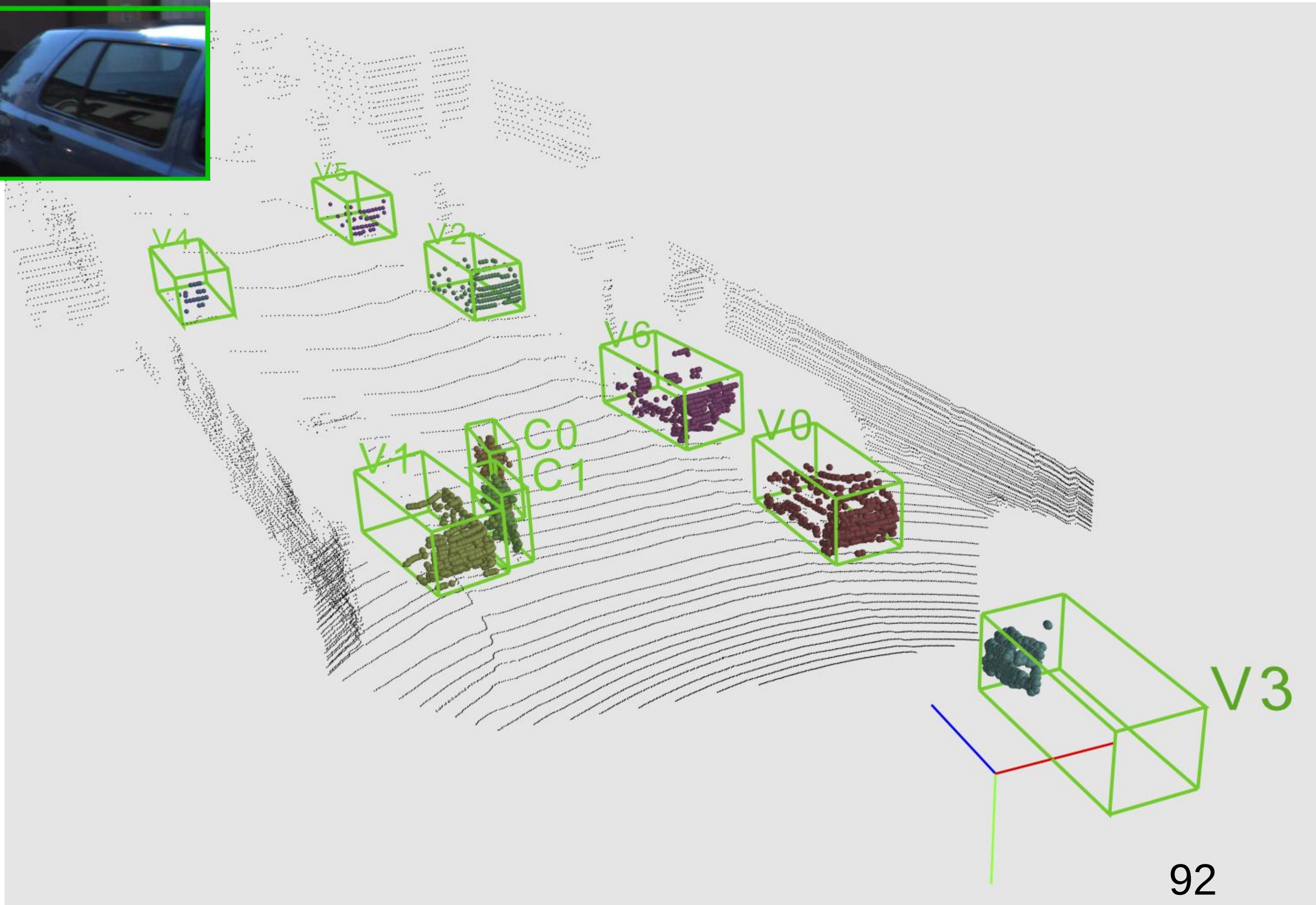




AVOD: [Ku et al. 2018]
VxNet: [Zhou et al. 2017]

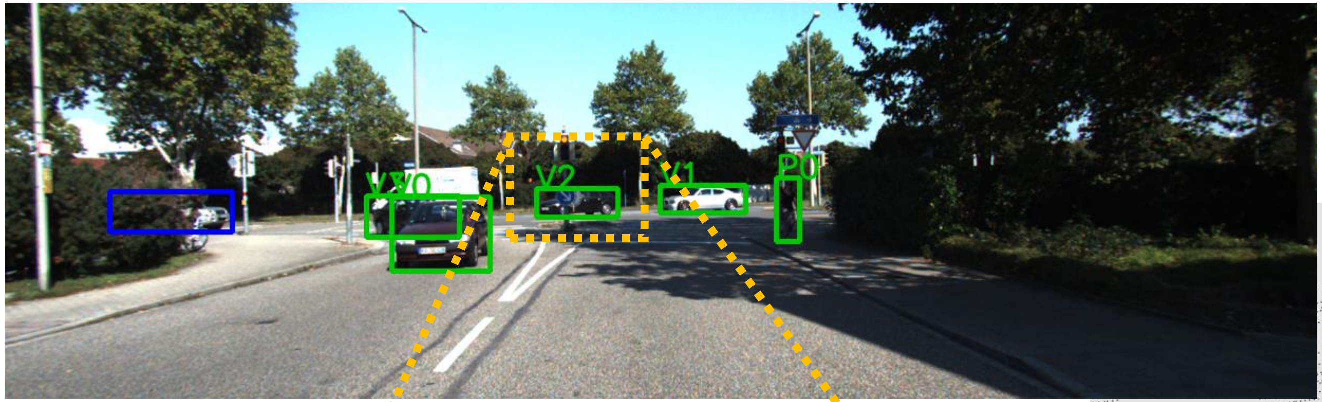
KITTI Results: Qualitative



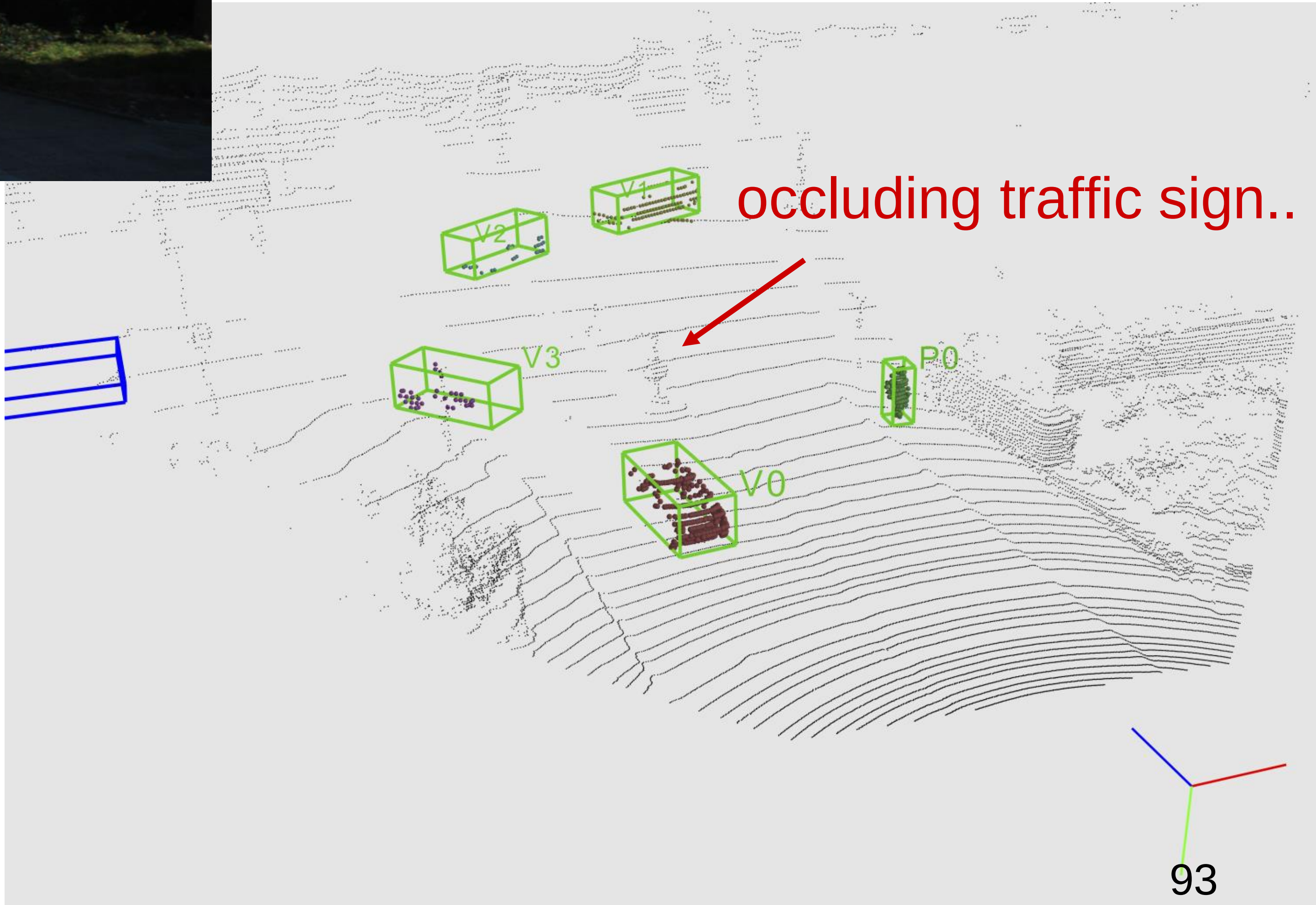


Remarkable box estimation accuracy even with a dozen of points or with very partial point clouds.

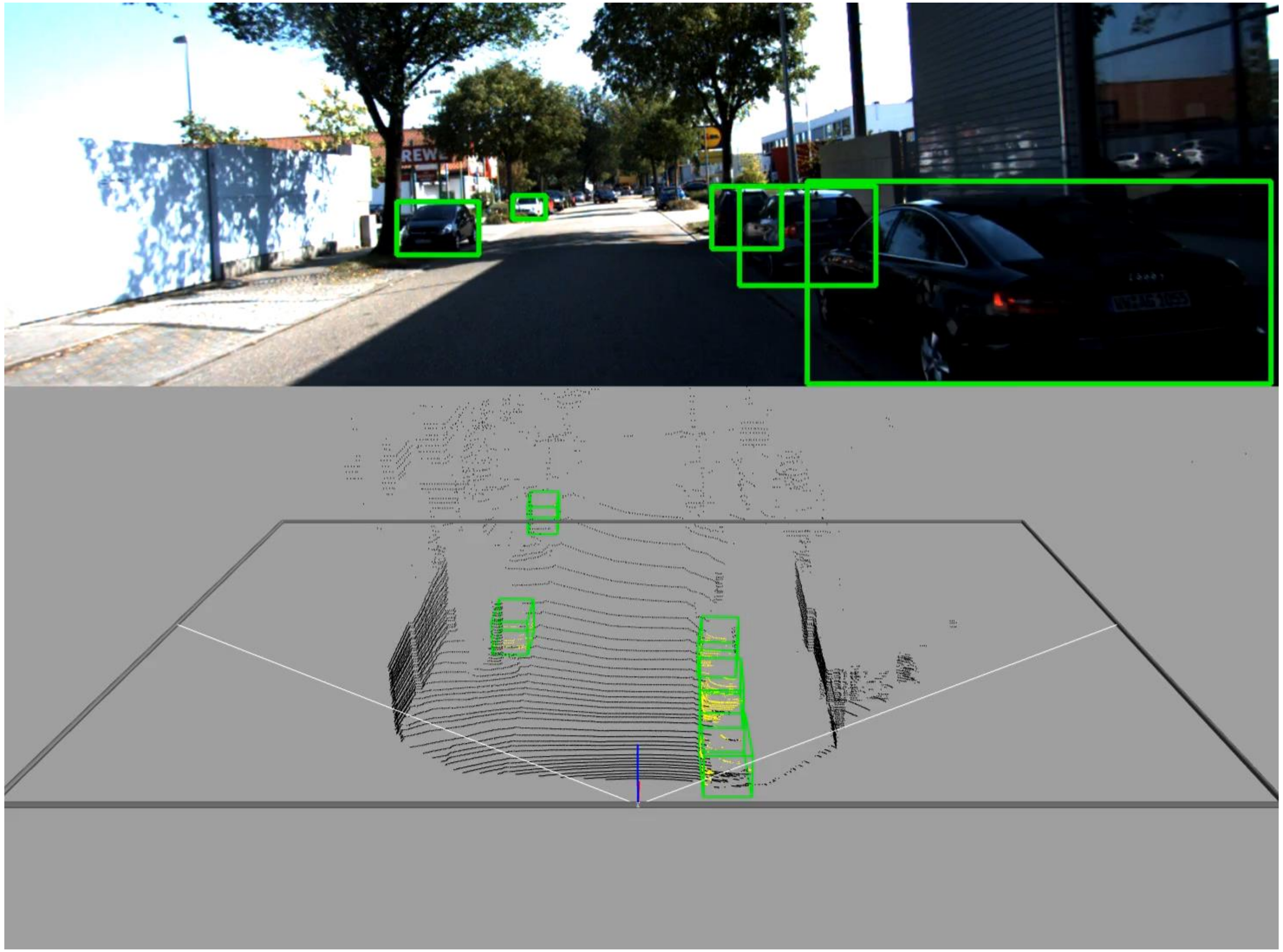
KITTI Results: Qualitative



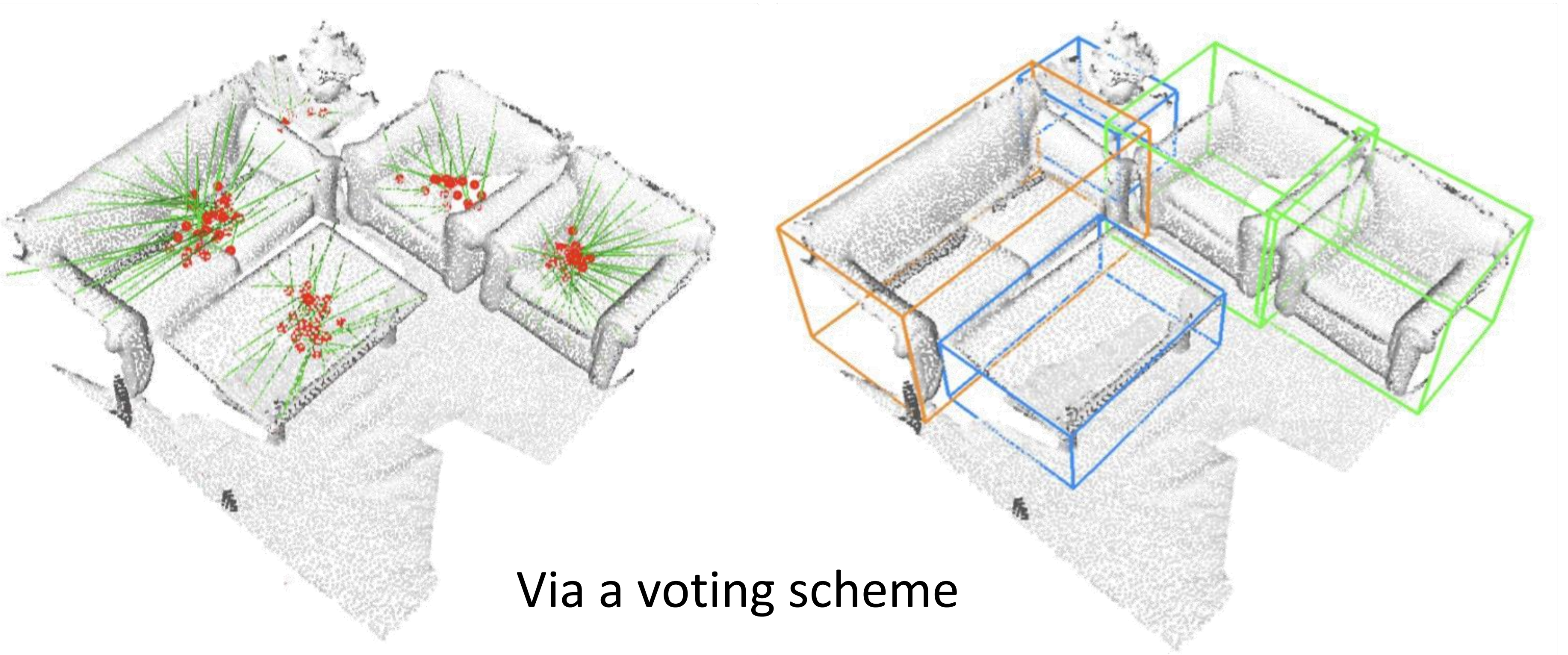
Correct segmentation in point clouds with heavy occlusion.



KITTI Results: Example

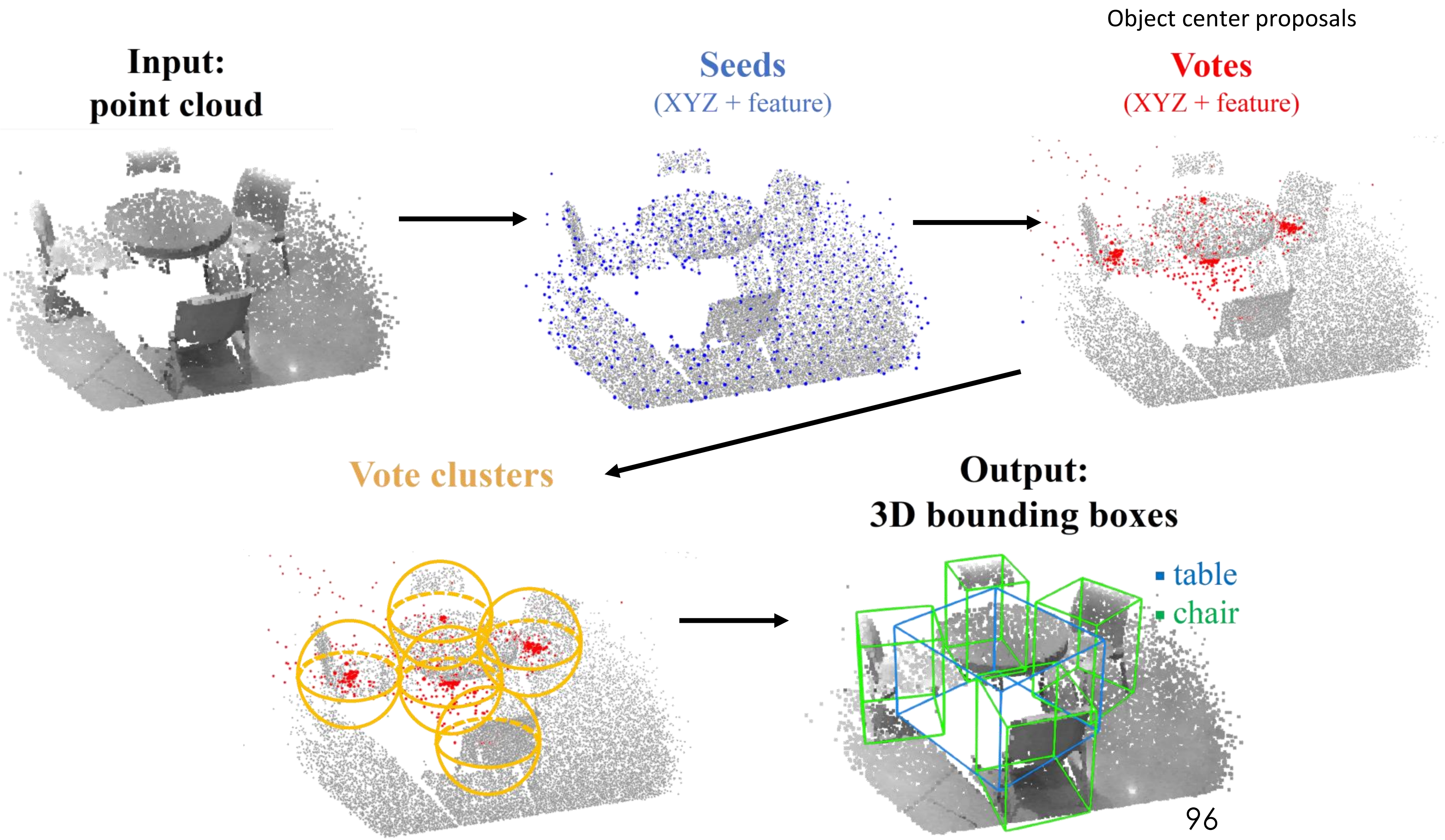


Point Cloud Amodal Bounding Box Detection (Indoor)

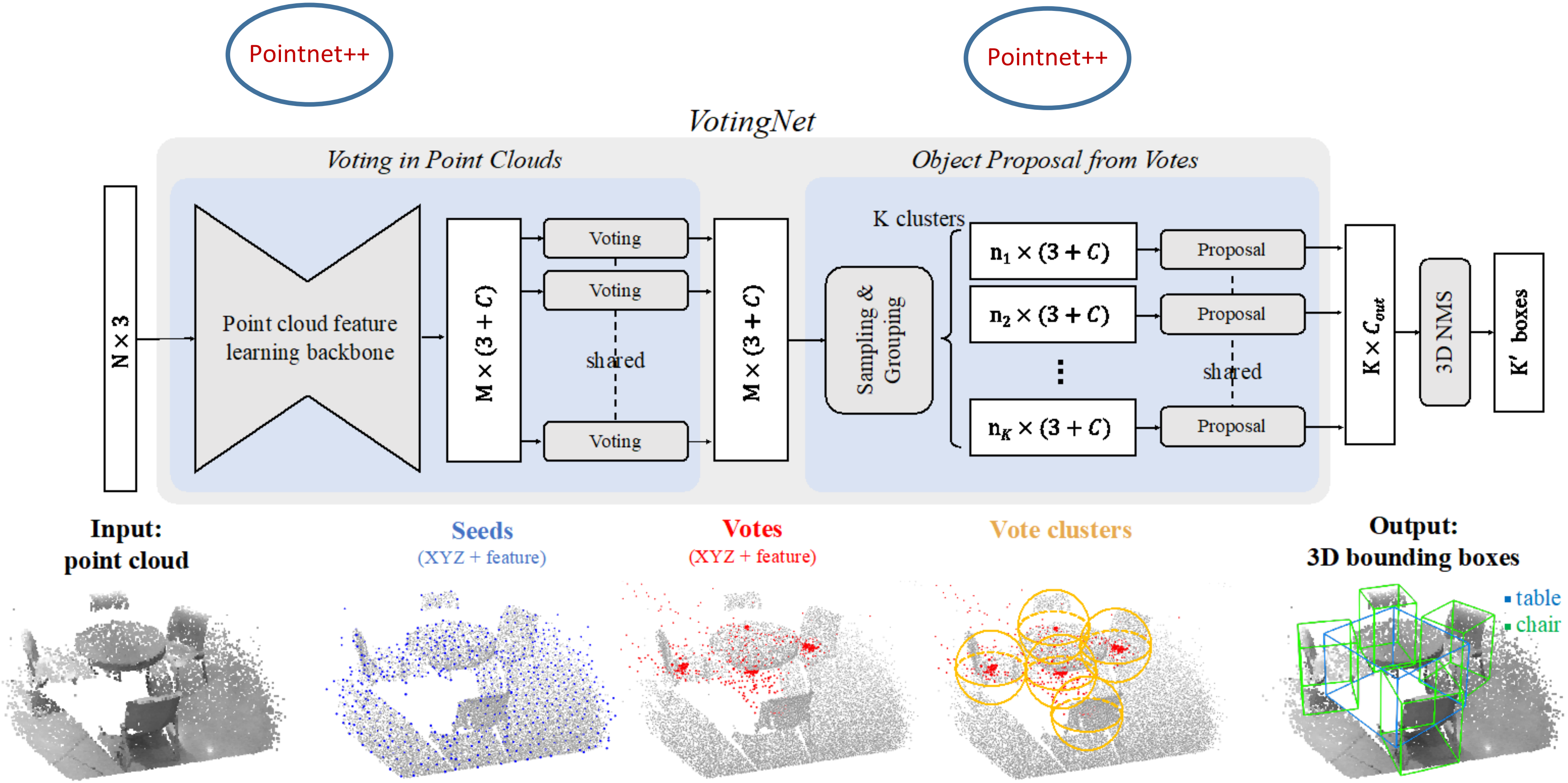


Via a voting scheme

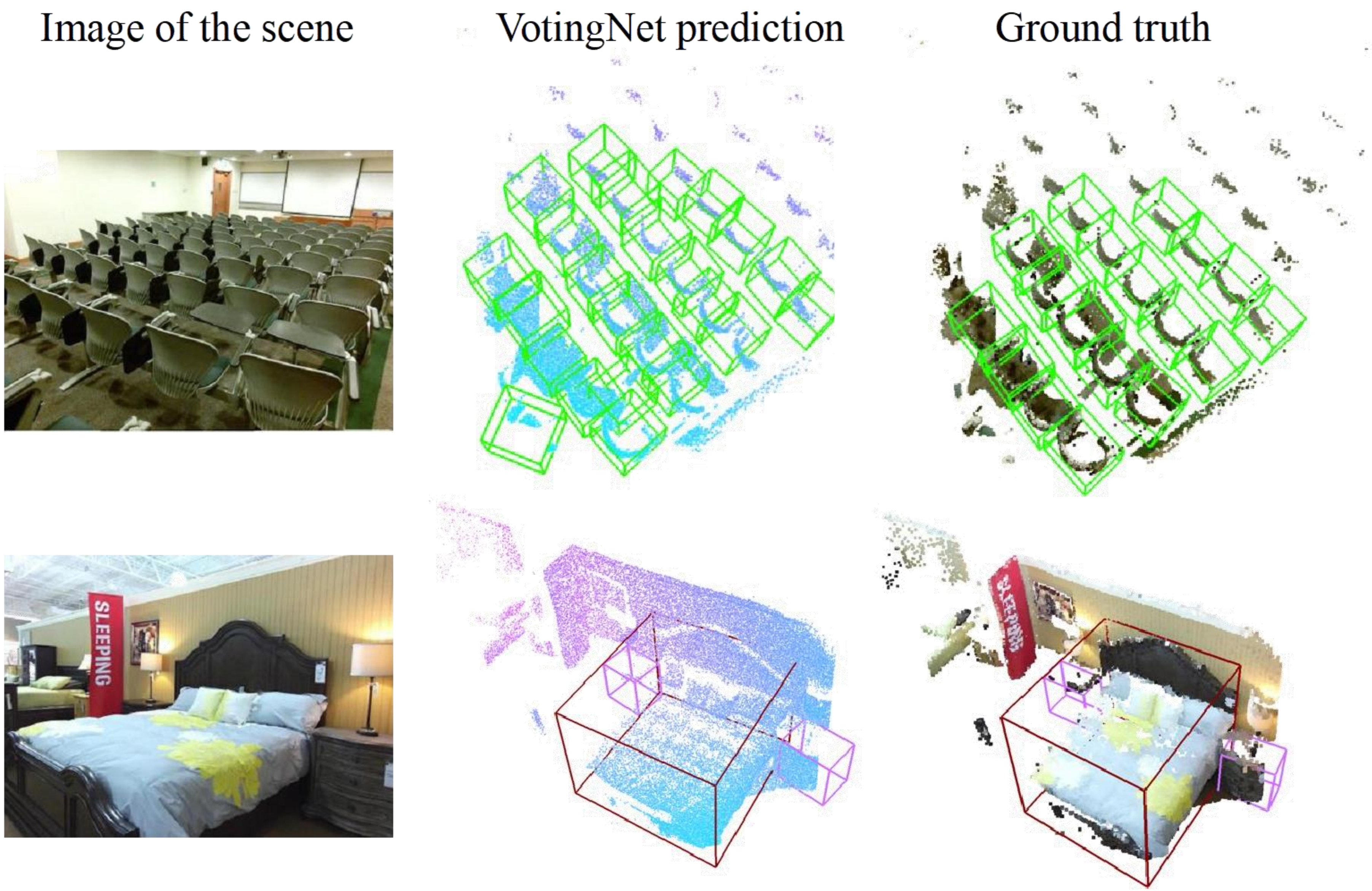
Deep Hough Voting



Deep Hough Voting



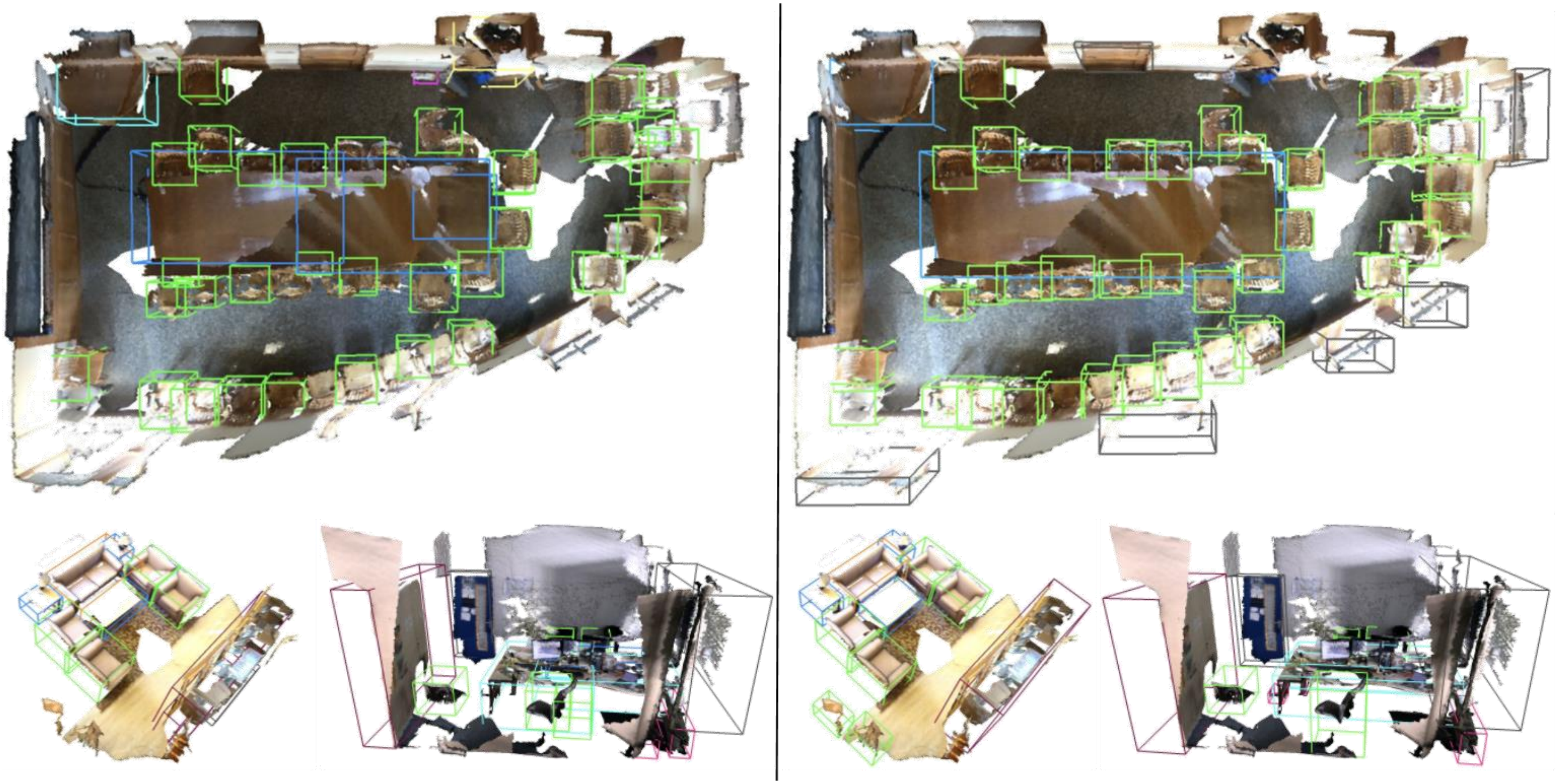
Results on SUN RGB-D



Results on ScanNet

VotingNet prediction

Ground truth



Quantitative Results

SUN RGB-D

	Input	bathtub	bed	bookshelf	chair	desk	dresser	nightstand	sofa	table	toilet	mAP
DSS [37]	Geo + RGB	44.2	78.8	11.9	61.2	20.5	6.4	15.4	53.5	50.3	78.9	42.1
COG [33]	Geo + RGB	58.3	63.7	31.8	62.2	**45.2**	15.5	27.4	51.0	**51.3**	70.1	47.6
2D-driven [17]	Geo + RGB	43.5	64.5	31.4	48.3	27.9	25.9	41.9	50.4	37.0	80.4	45.1
F-PointNet [30]	Geo + RGB	43.3	81.1	**33.3**	64.2	24.7	**32.0**	58.1	61.1	51.1	**90.9**	54.0
VotingNet (ours)	Geo only	**74.4**	**83.0**	28.8	**75.3**	22.0	29.8	**62.2**	**64.0**	47.3	90.1	**57.7**

ScanNetV2

	Input	mAP@0.25	mAP@0.5
DSS [37]	Geo + RGB	15.2	6.8
MRCNN 2D-3D [10]	Geo + RGB	17.3	10.5
F-PointNet [30]	Geo + RGB	19.8	10.8
GSPN [47]	Geo + RGB	30.6	17.7
3D-SIS [11]	Geo + 1 view	35.09	18.66
3D-SIS [11]	Geo + 3 views	36.64	19.04
3D-SIS [11]	Geo + 5 views	40.22	22.53
3D-SIS [11]	Geo only	25.36	14.60
VotingNet (ours)	Geo only	**46.75**	**24.65**

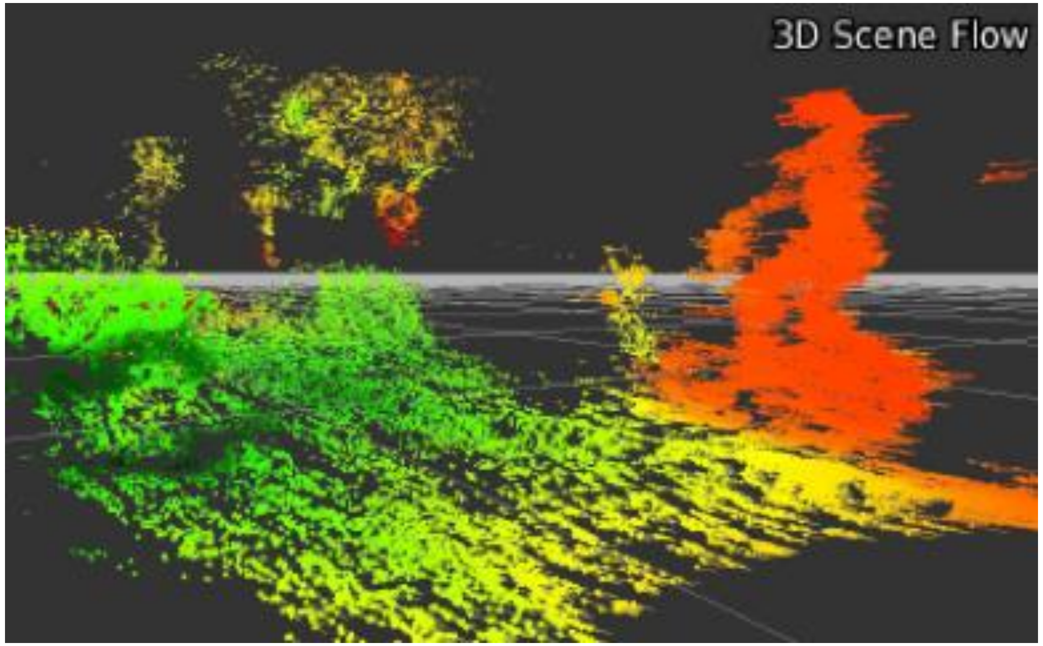
3D Scene Understanding with PointNets

- PointNet and PointNet++ lead to new **3D centric approaches** to scene understanding

3D Object Detection

source: SUN RGB-D by Song et al.

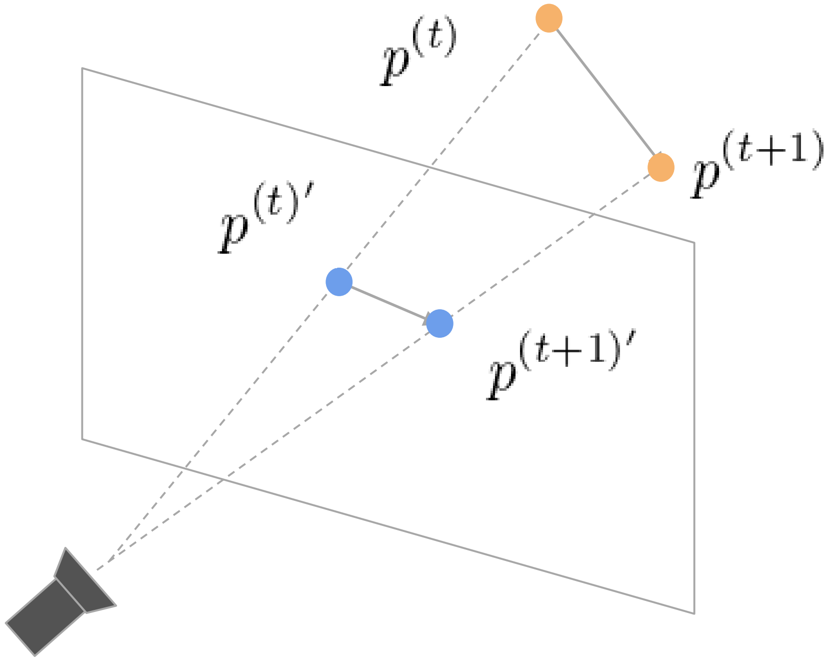
3D Scene Flow



source: Wedel et al.

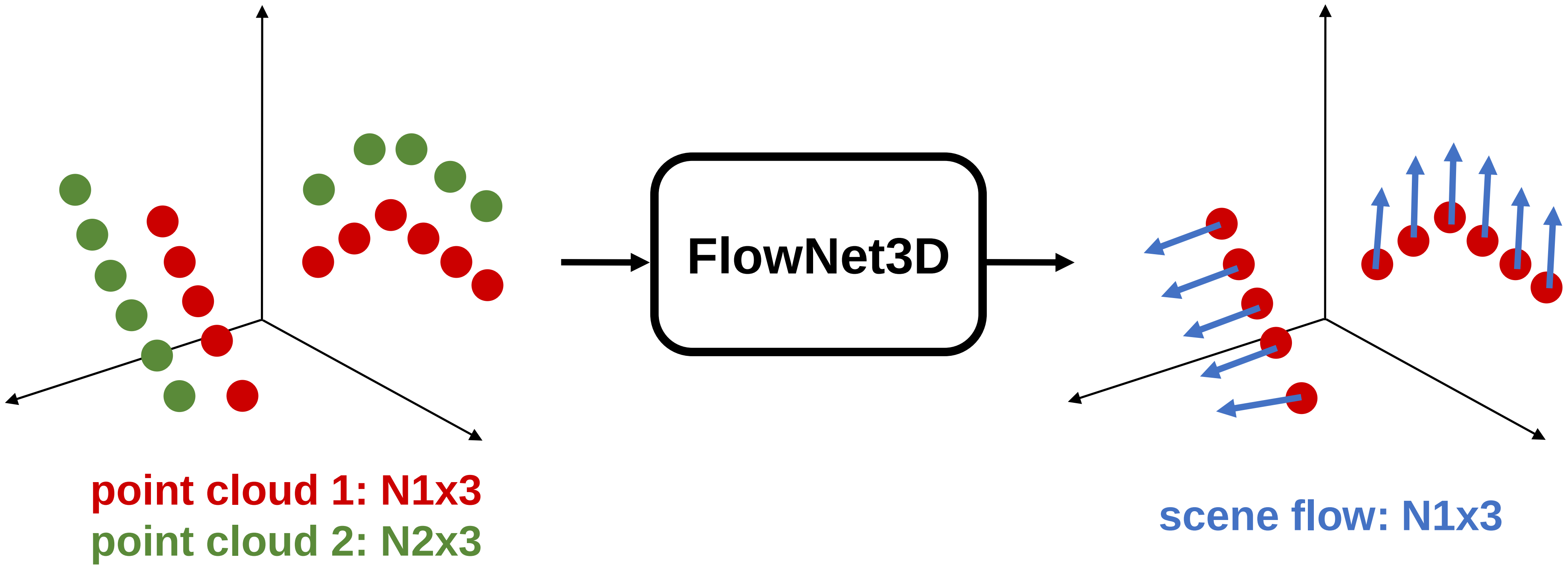
Scene Flow [Vedula et al. 1999]

- Scene flow: 3D motion field of points
- Optical flow is its projection to 2D image plane.
- Low-level understanding of a dynamic environment



Our Approach: FlowNet3D

- Directly learning scene flow in 3D point clouds, with 3D deep learning architectures.



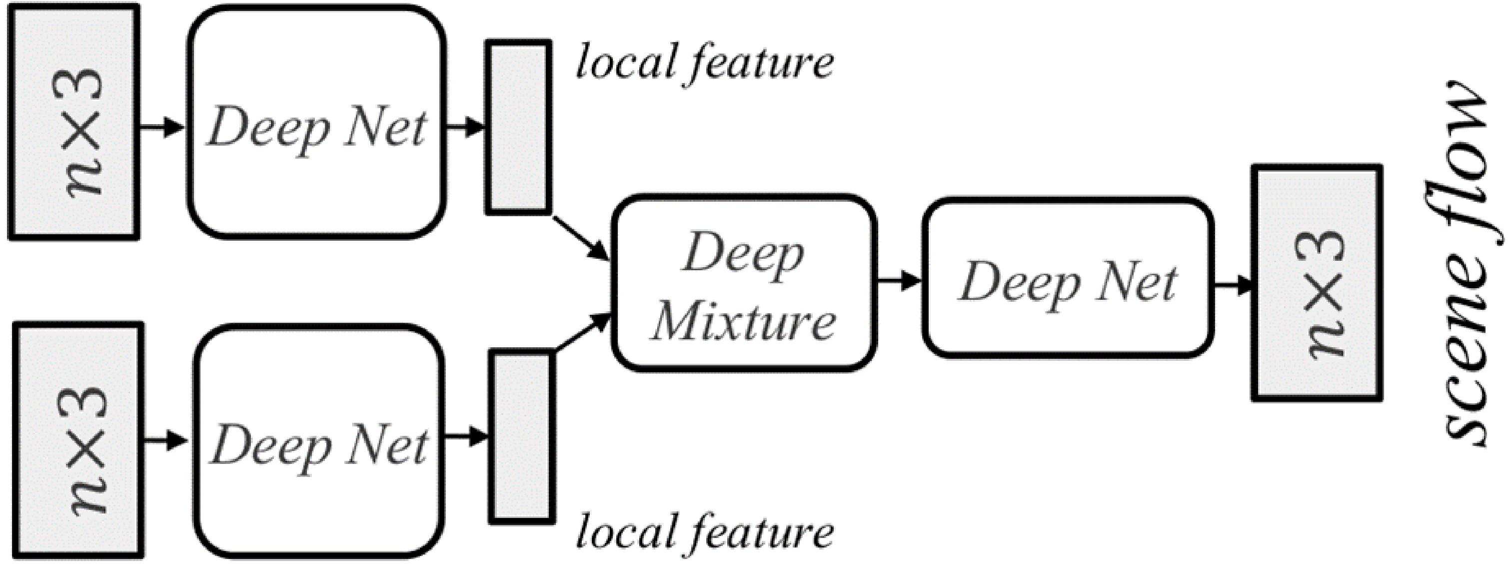
[4] *Xingyu Liu*, Charles R. Qi*, Leonidas Guibas. Learning Scene Flow in 3D Point Clouds, arXiv preprint.*

Deep Net Architecture

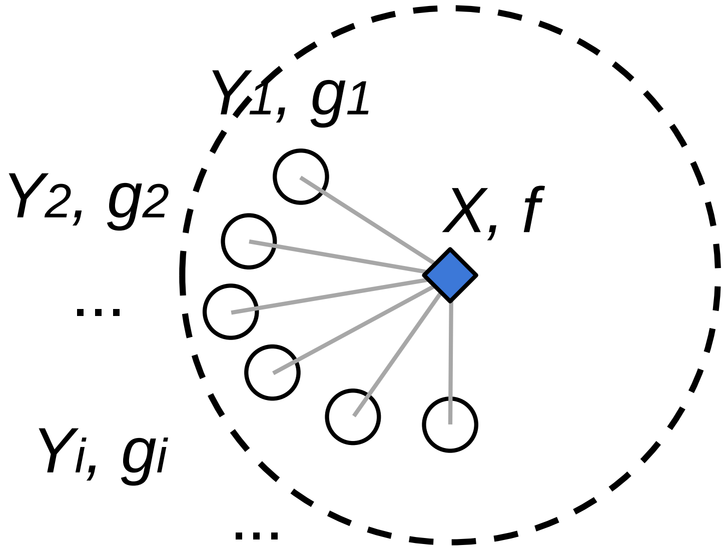
- How to learn point cloud features?
- Where in the network architecture to mix point features from consecutive frames?
- How to mix them?

Intermediate level

Middle-level
deep mixture



Point Attributes



dist(g_1, f), Y_1-X

dist(g_2, f), Y_2-X

⋮

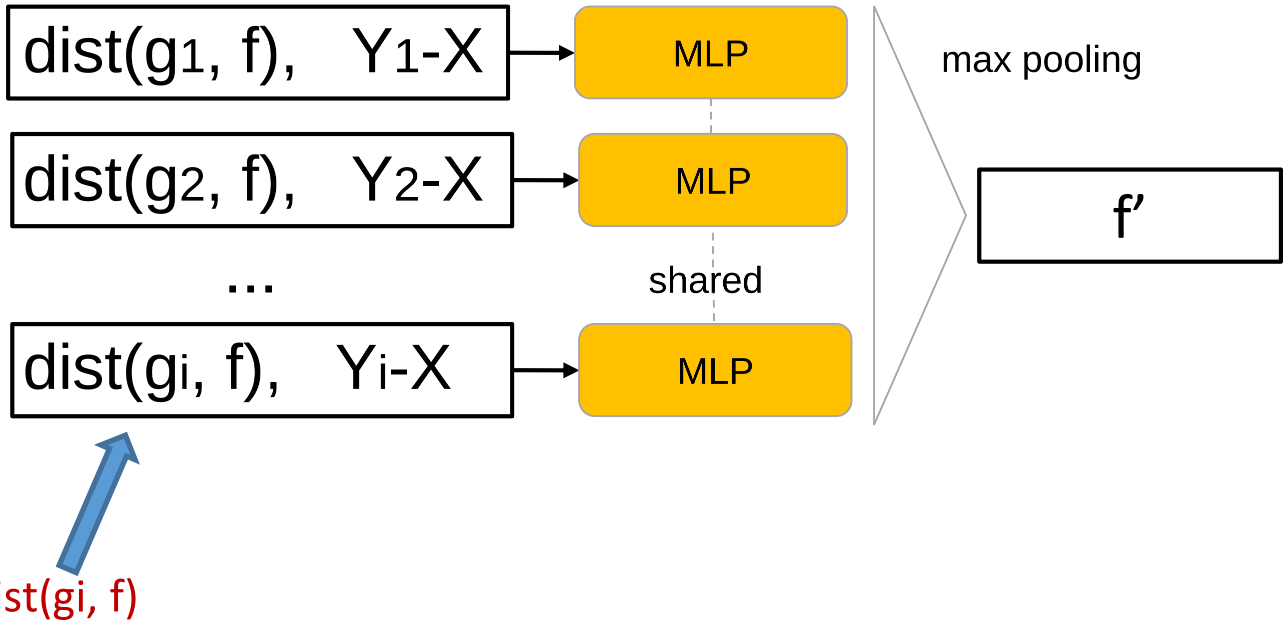
dist(g_i, f), Y_i-X

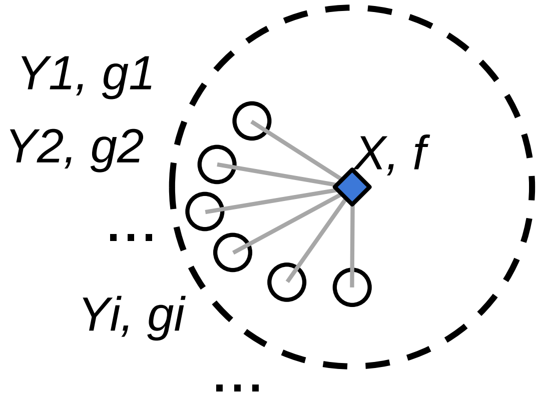
⋮

Naive approach: concatenation

dist(g_1, f), Y_1-X	dist(g_2, f), Y_2-X	...

A More Structured Approach





"Distance" functions:

Euclidean distance (scalar)

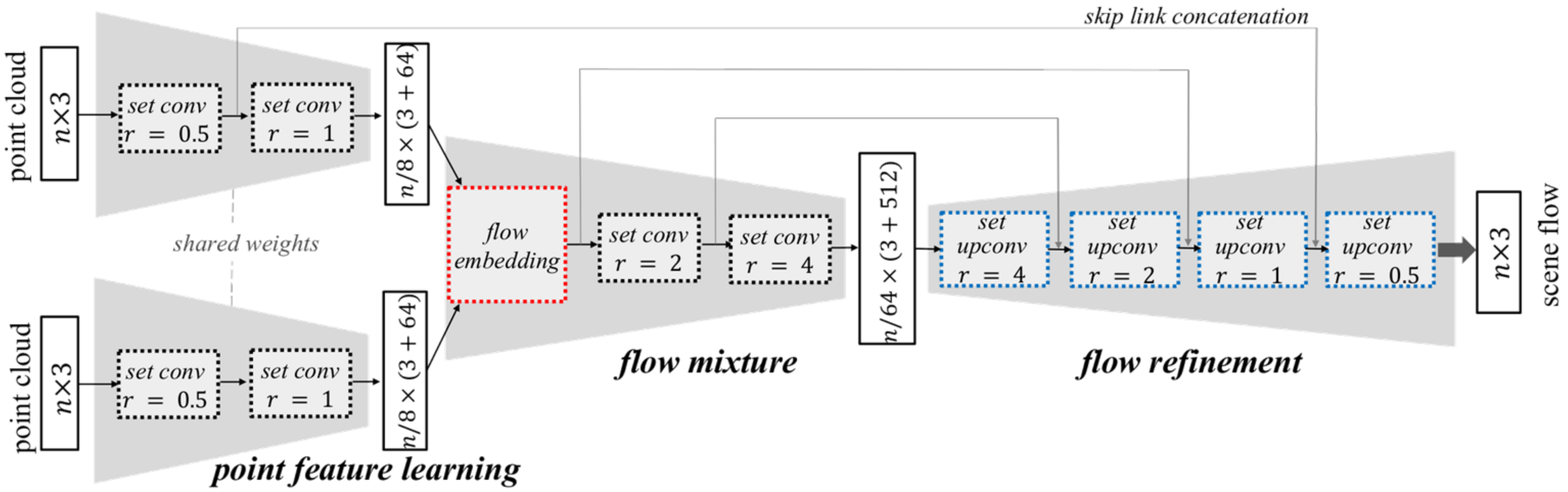
Cosine distance (scalar)

Element-wise product (vector)

Simple concatenation — let the network learn the distance function (vector)

…

FlowNet3D



Composed of many many mini-pointnet++ modules ...

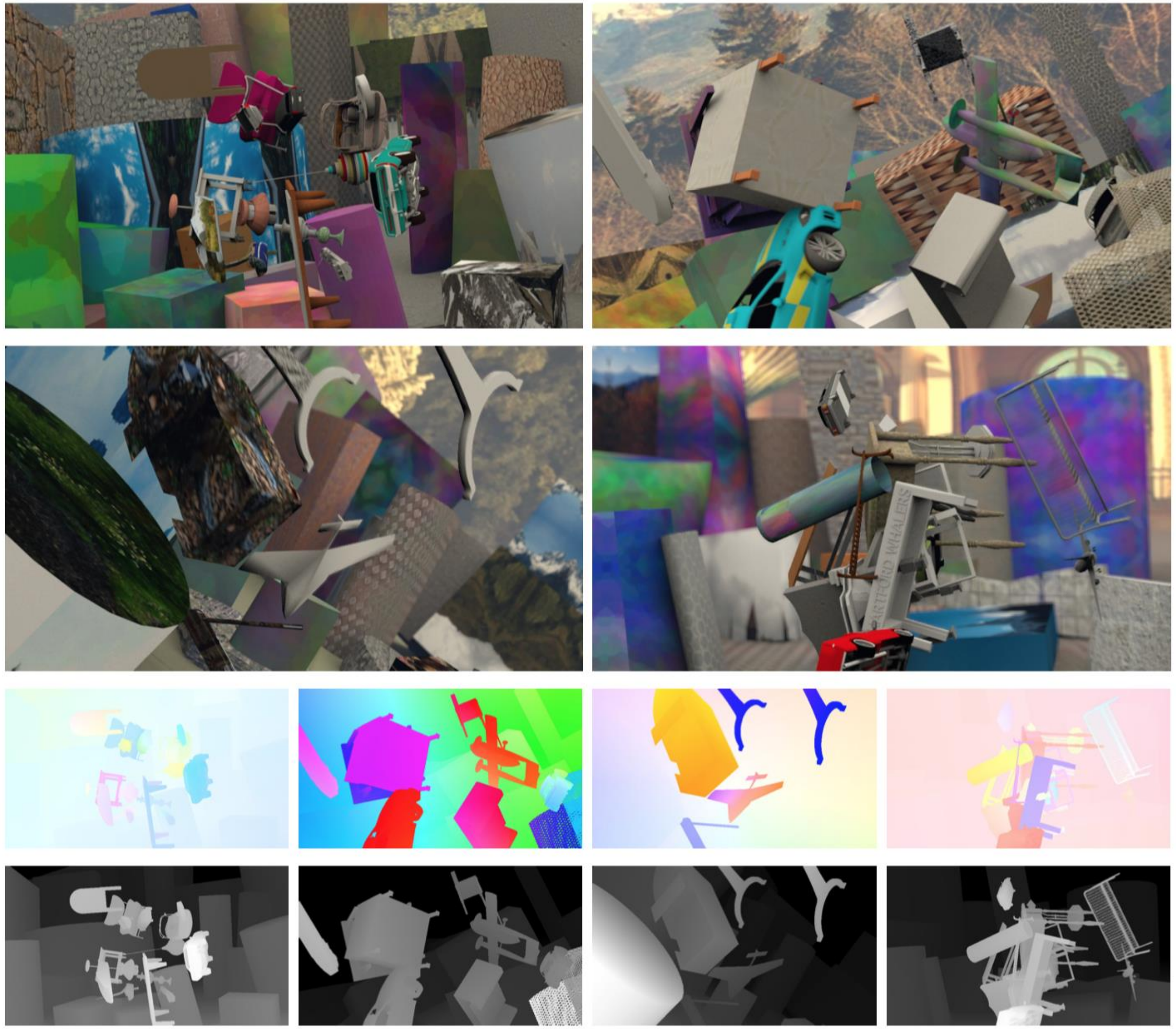
Pointnet++

Training on Synthetic Data

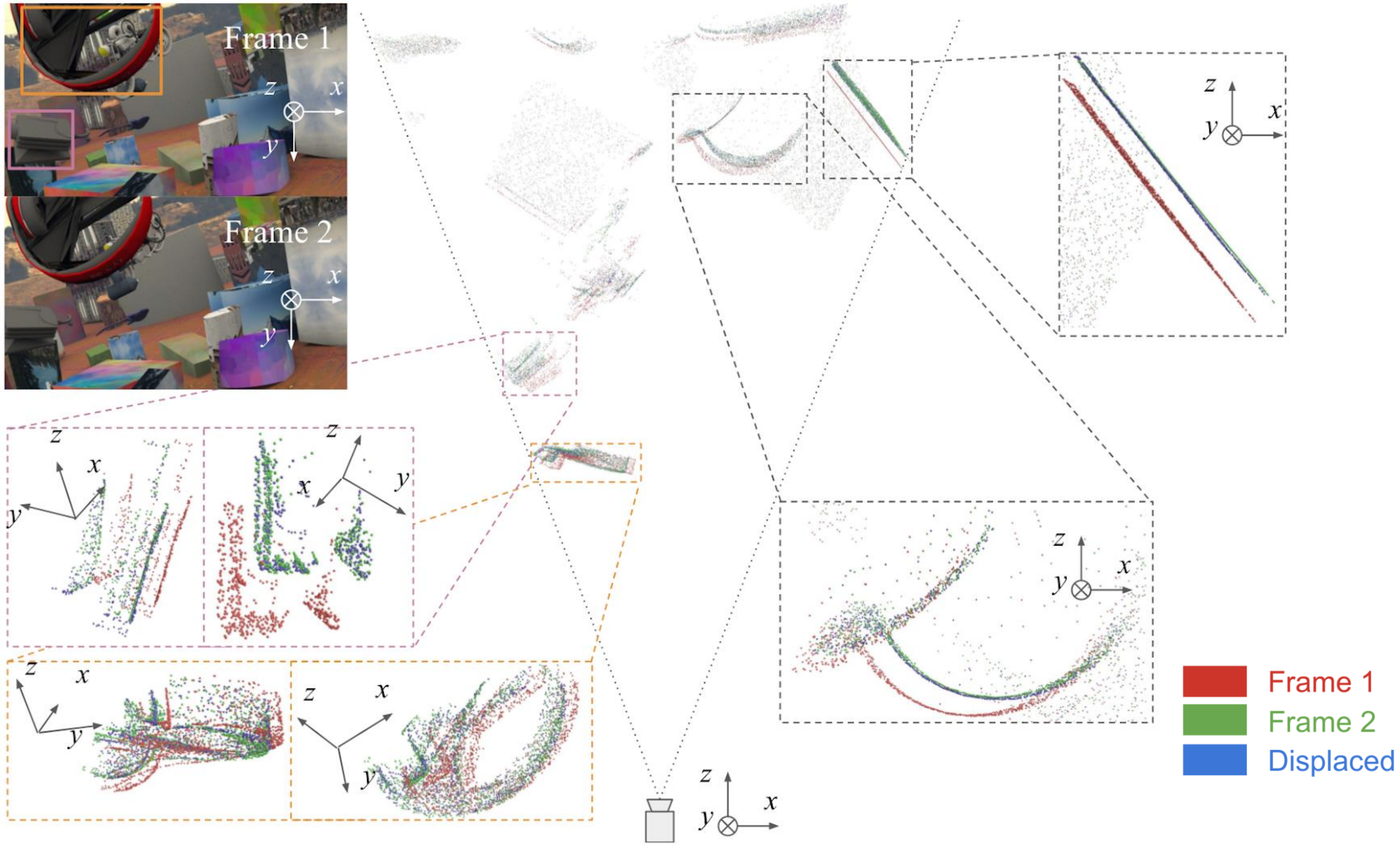
FlyingThings3D [Mayer et al. 2016] dataset from MPI

Random ShapeNet objects

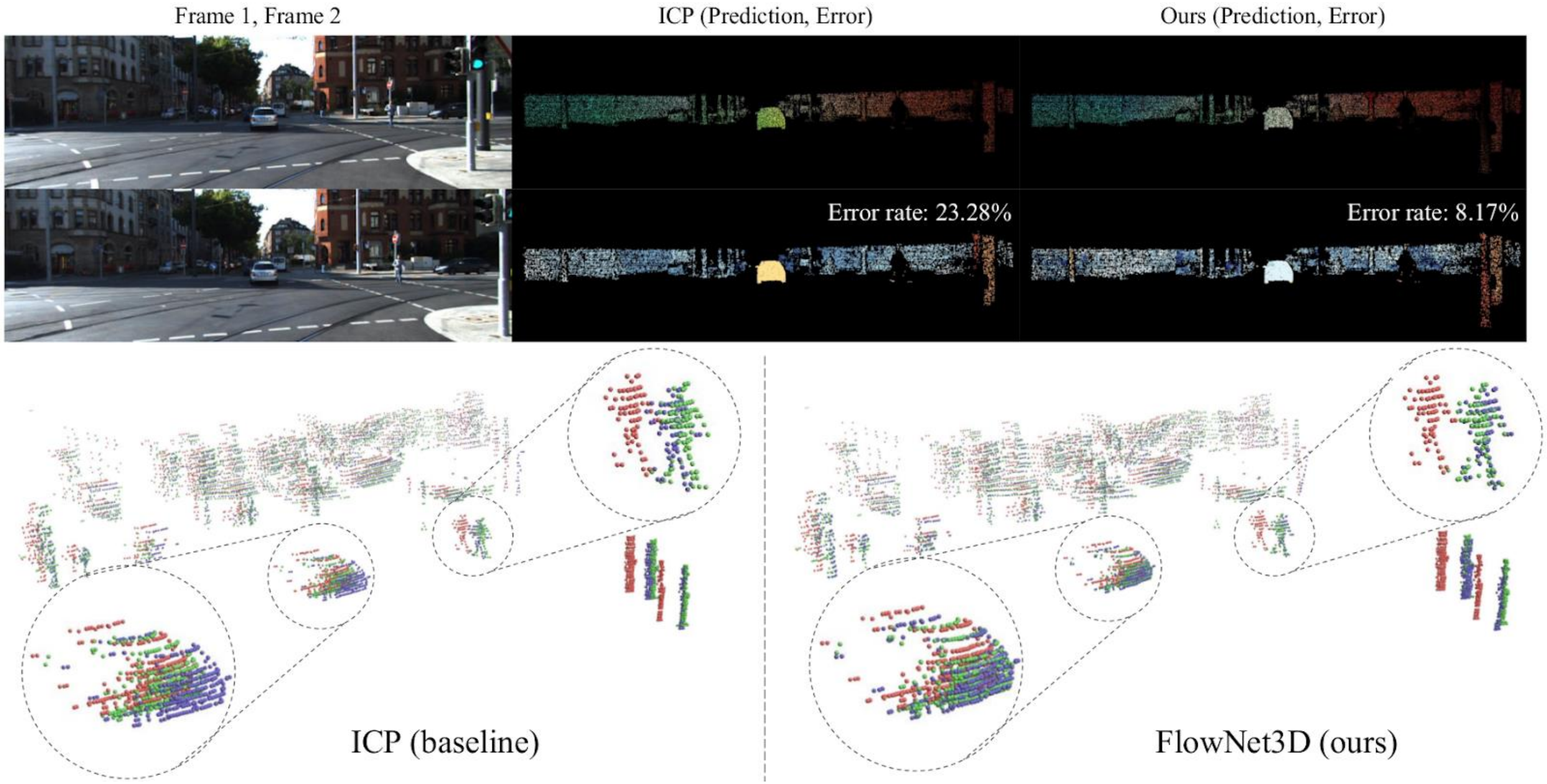
Very challenging dataset with strong occlusions and large motions.



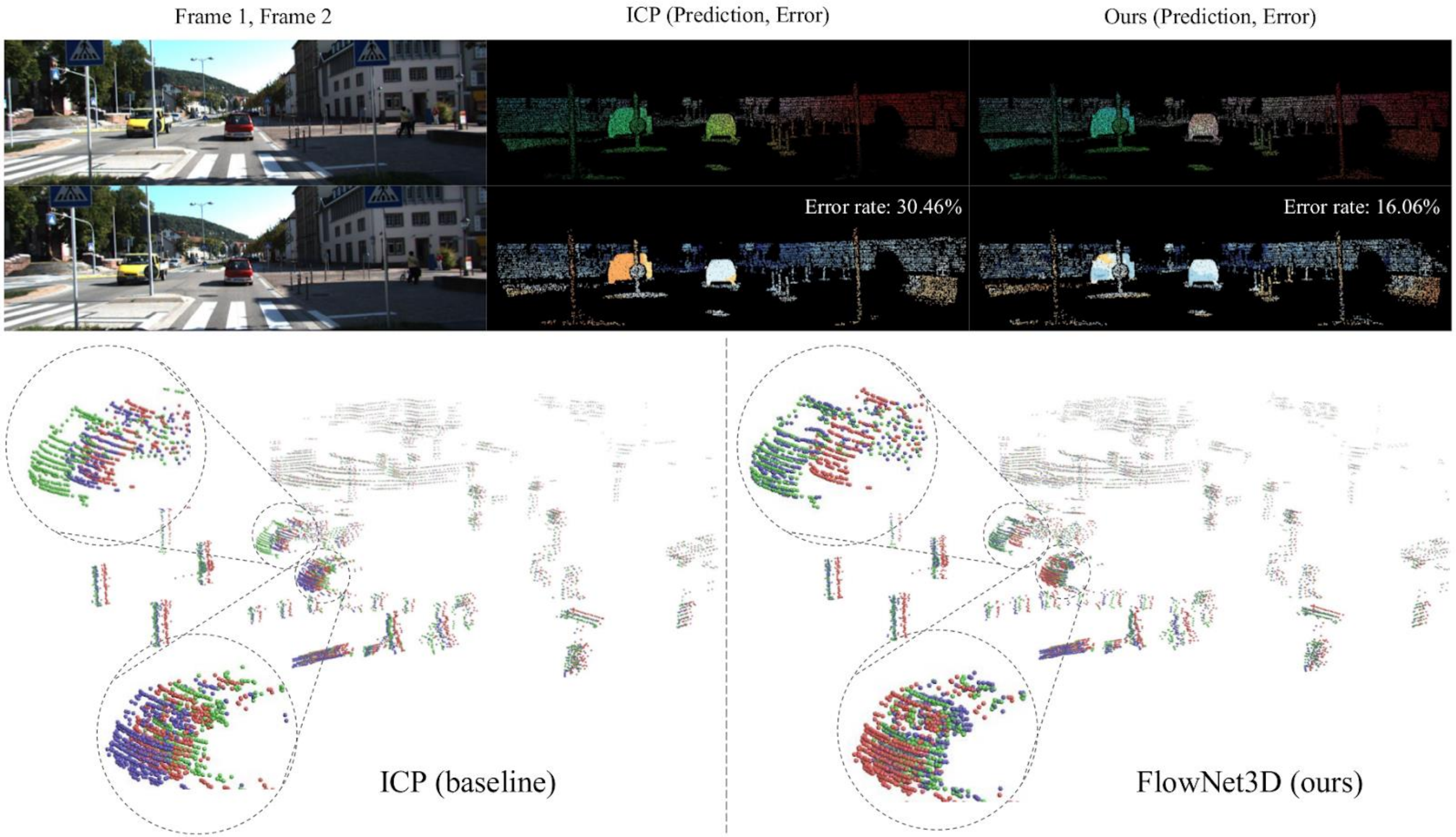
FlyingThings3D Results



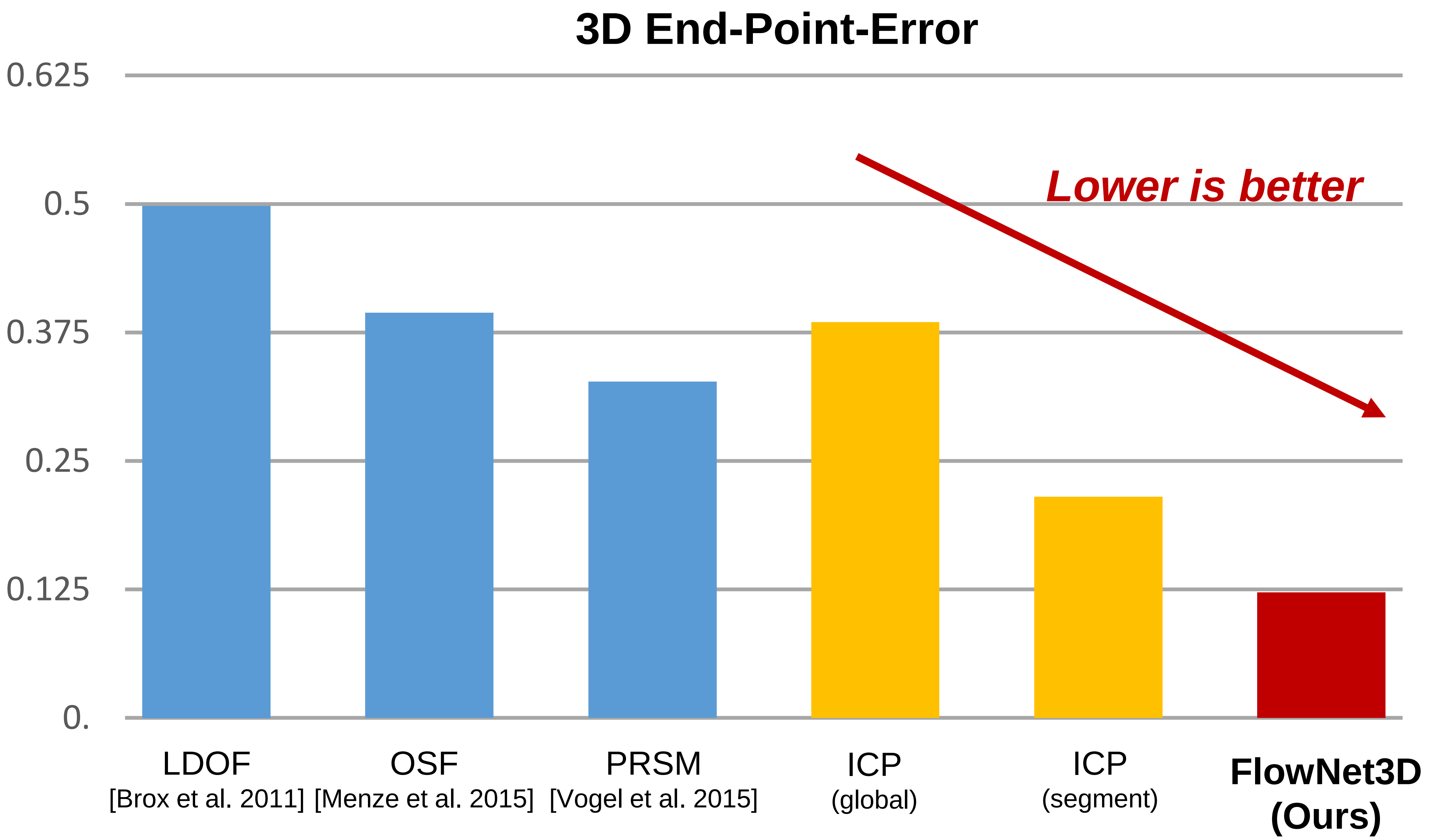
KITTI Results



KITTI Results

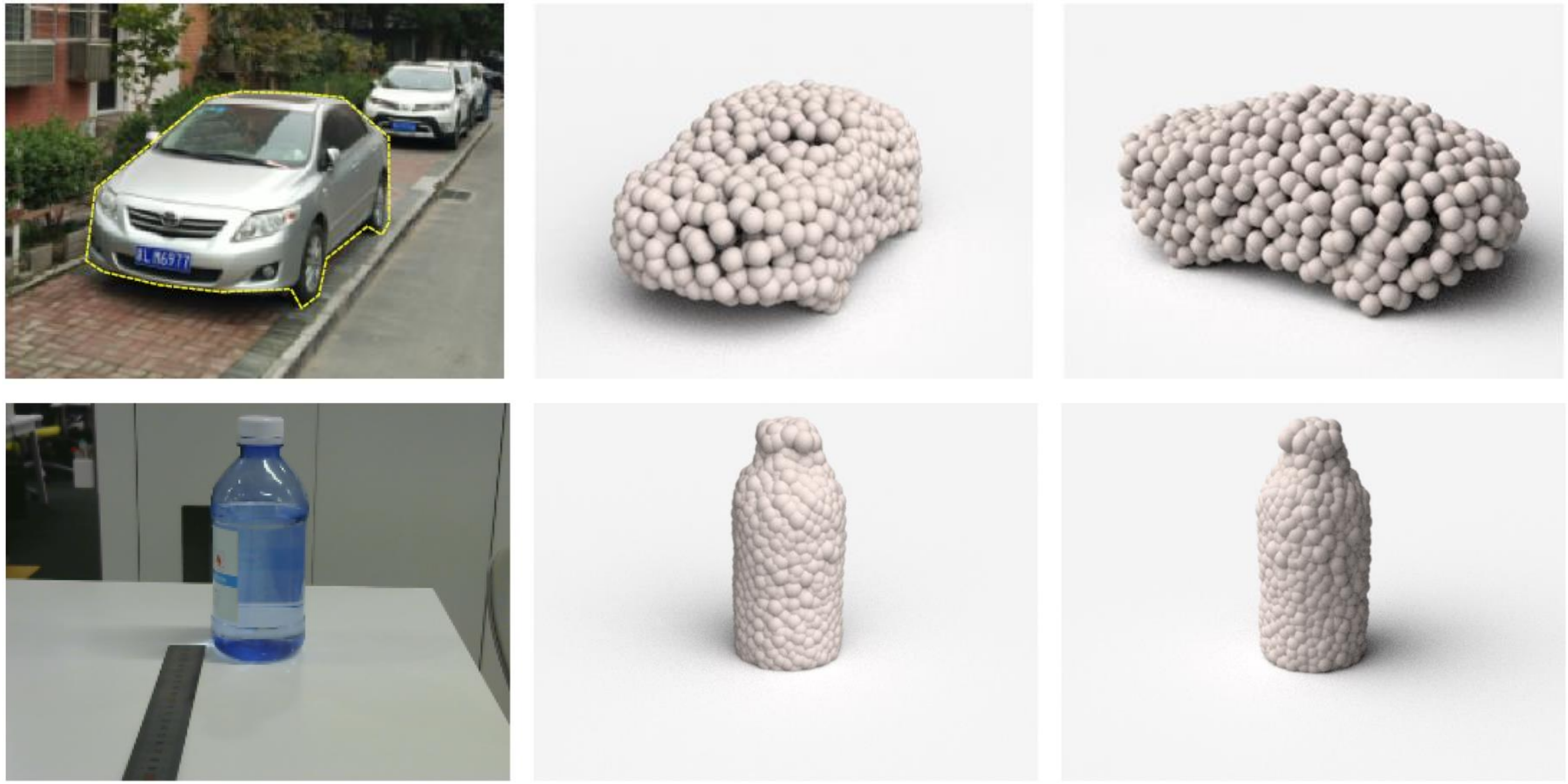


Generalizing to KITTI: Quantitative



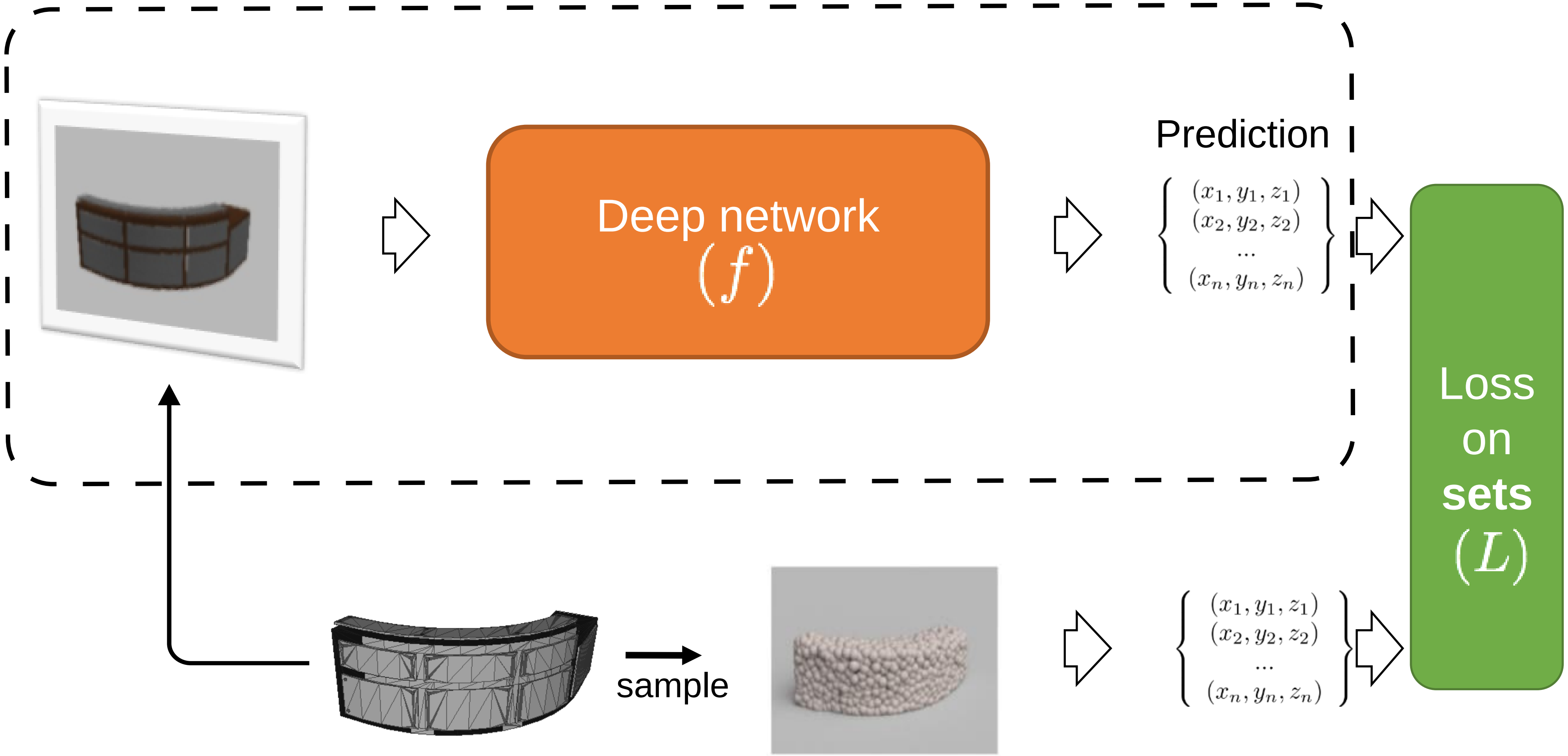
Point Cloud Synthesis

Point Cloud Synthesis from a Single Image

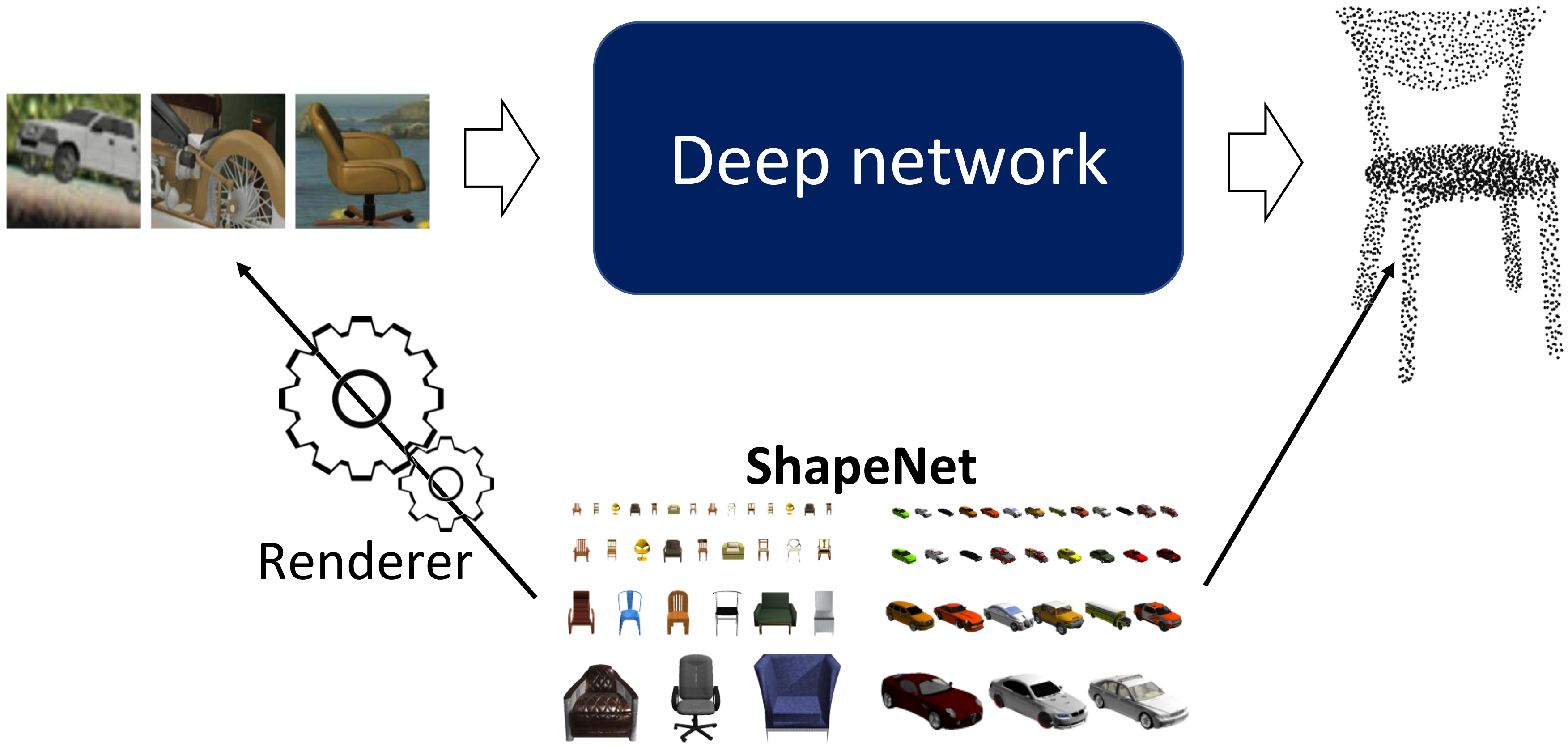


Hao Su, Haoqiang Fan, Leonidas Guibas
Learning Shape Abstractions by Assembling Volumetric Primitives
CVPR 2017

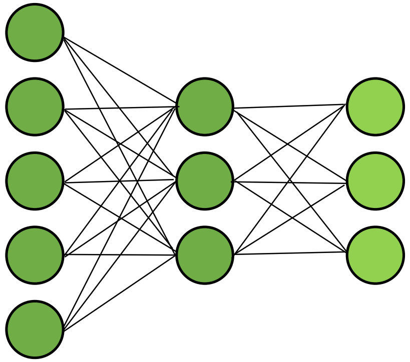
End-to-End Learning



Synthesize for Learning



Supervision from “Synthesize for Learning”



- **200K shapes from 2K categories**
- **10M images with ground truth**

Point Cloud Distance Metrics

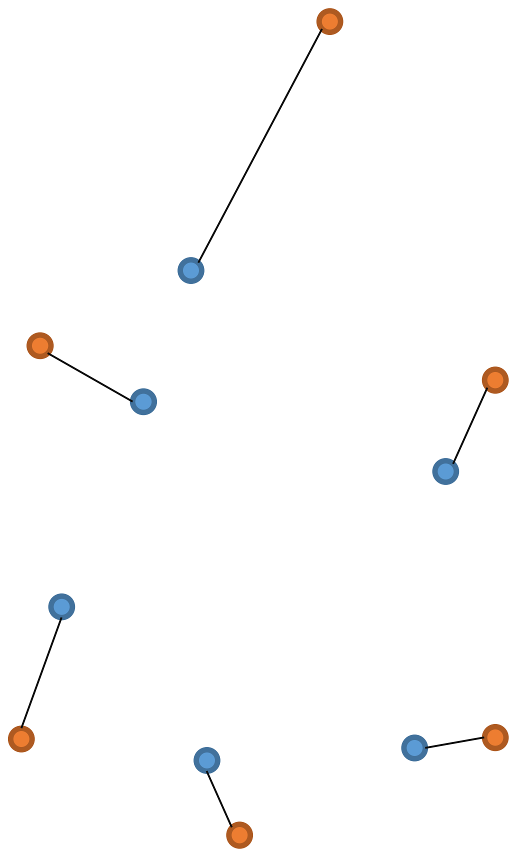
Worst case: Hausdorff distance (HD)

Average case: Chamfer distance (CD)

Optimal case: Earth Mover's distance (EMD)

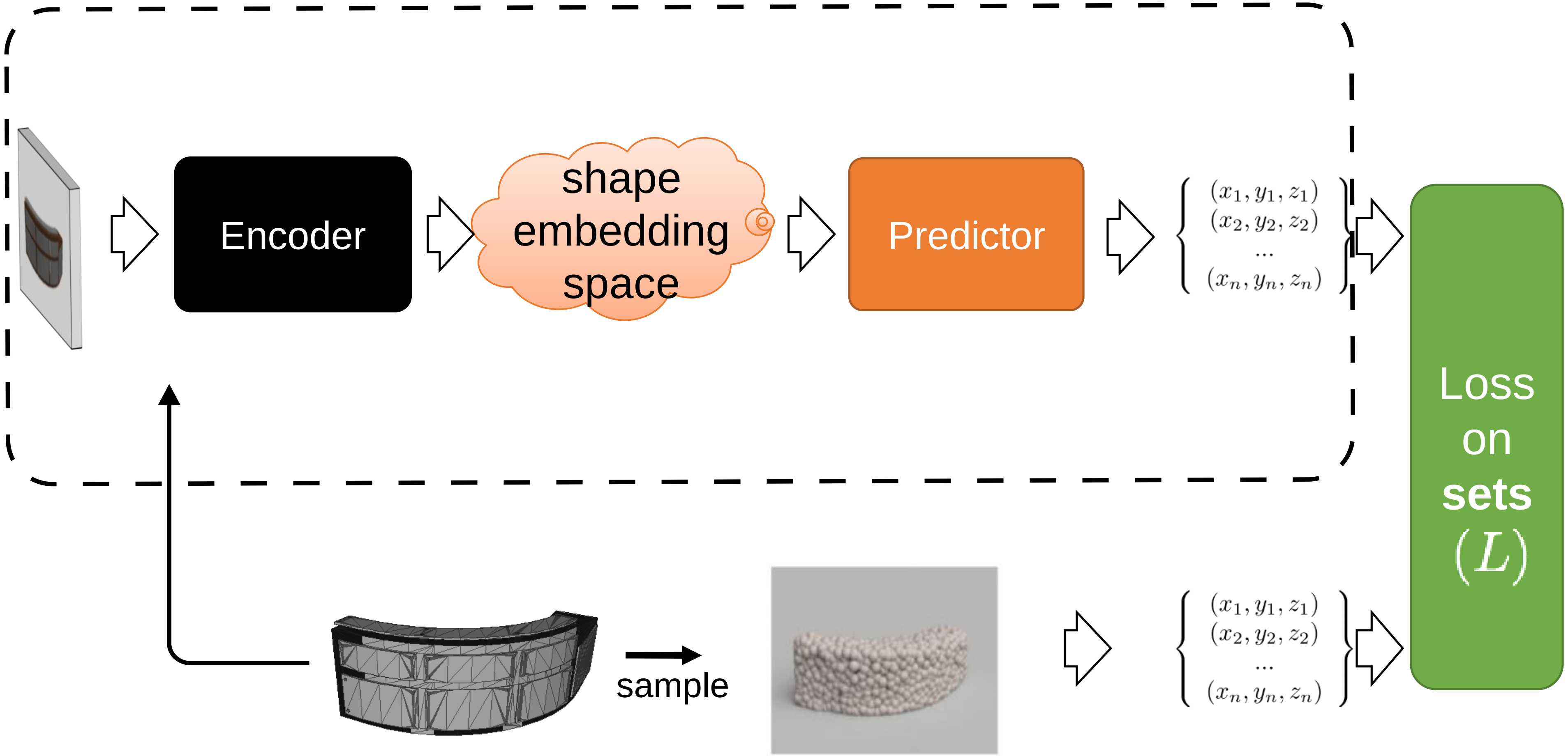
$$d_{EMD}(S_1, S_2) = \min_{\phi: S_1 \to S_2} \sum_{x \in S_1} \|x - \phi(x)\|_2$$

where $\phi : S_1 \to S_2$ is a bijection.

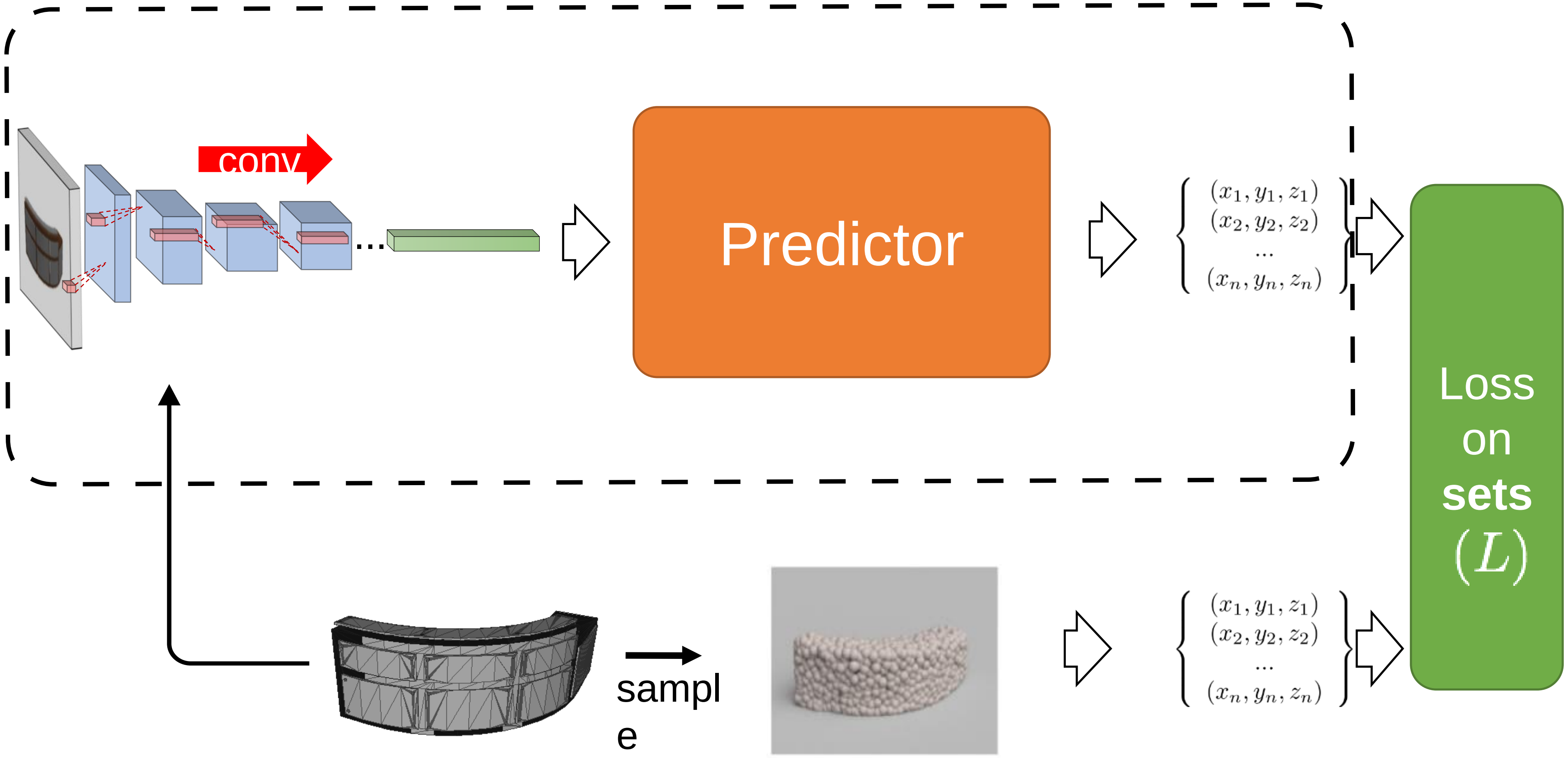


Solves the optimal transportation (bipartite matching) problem!

End-to-End Learning



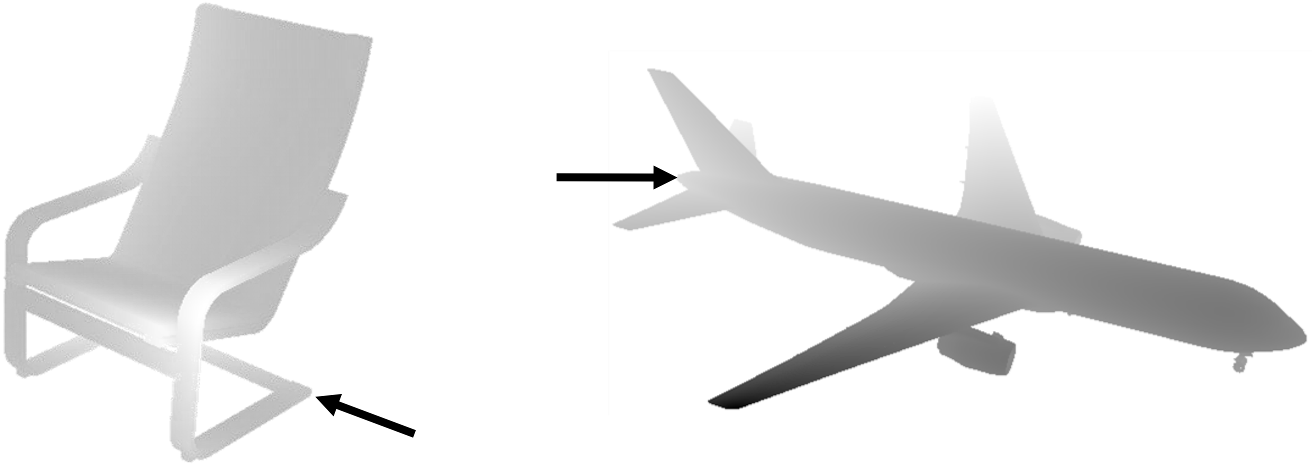
End-to-End Learning



Natural Statistics of Object Geometry

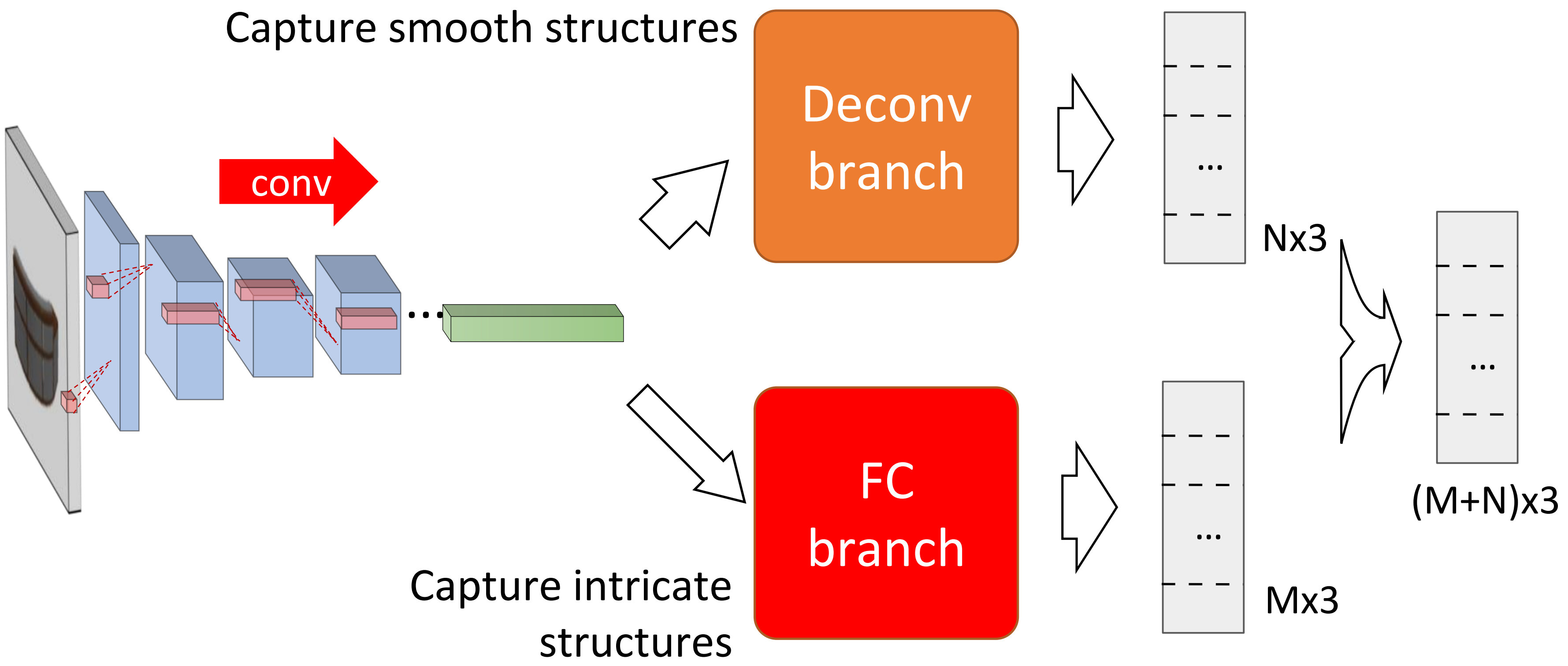
- Many local smooth structures are common
  - e.g., planar patches, cylindrical patches
  - **strong local correlation** among point coordinates

Natural Statistics of Object Geometry



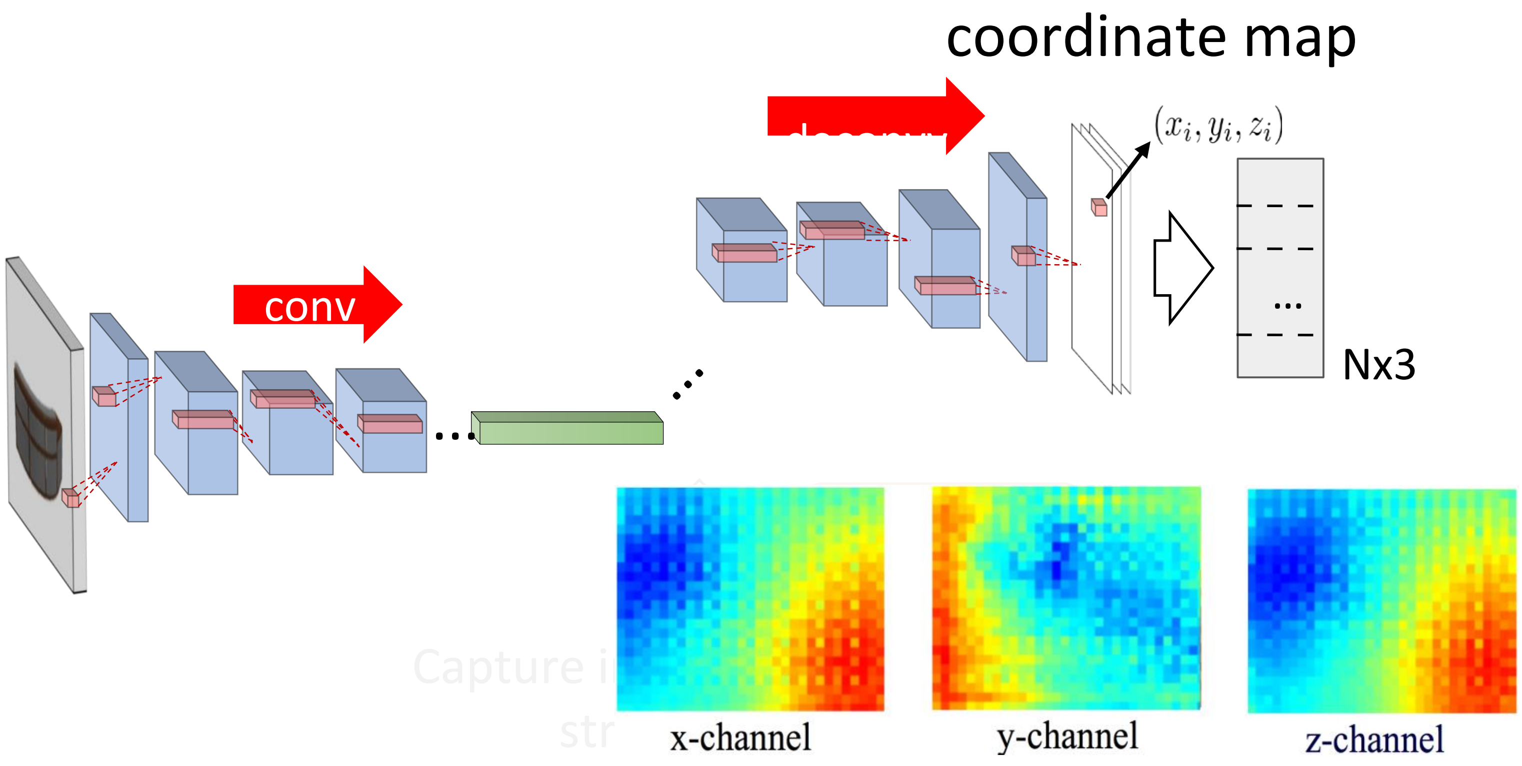
- But also some sharp/intricate local structures
  - **some points have high variability neighborhoods**

Two-Branch Architecture



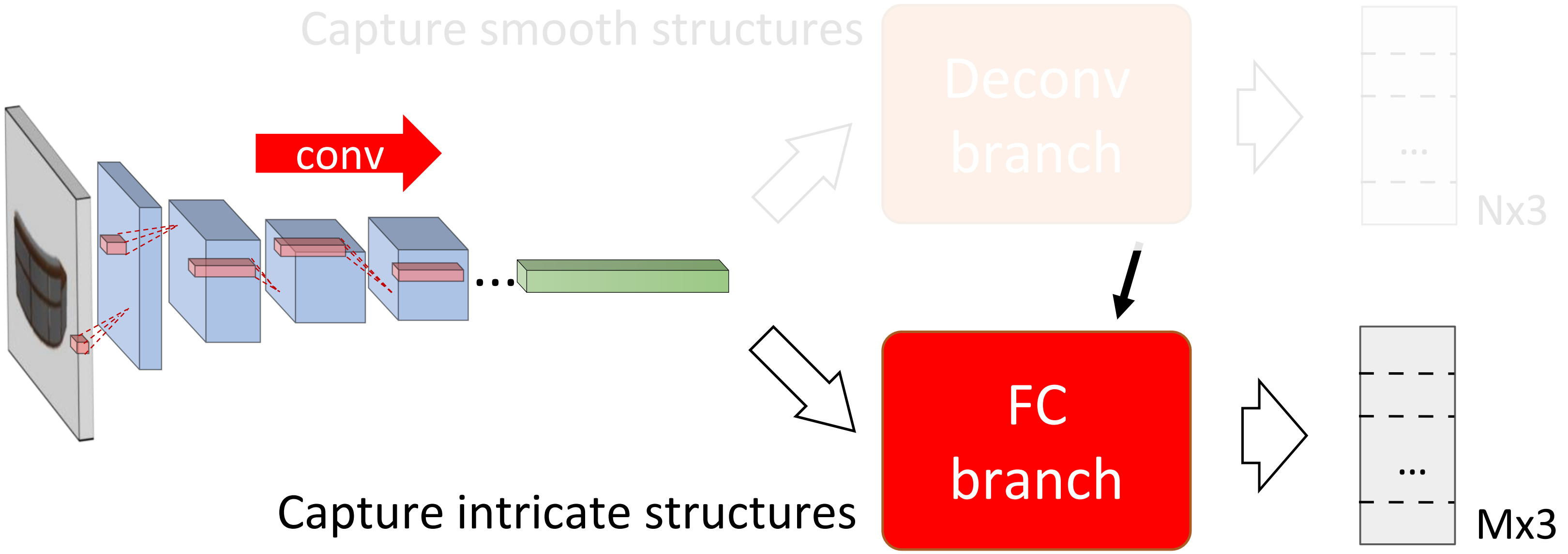
Set union by array concatenation

Deconvolution Branch



- Deconvolution induces a smooth coordinate map
- Geometrically, it learns a smooth parameterization

Fully-Connected Branch

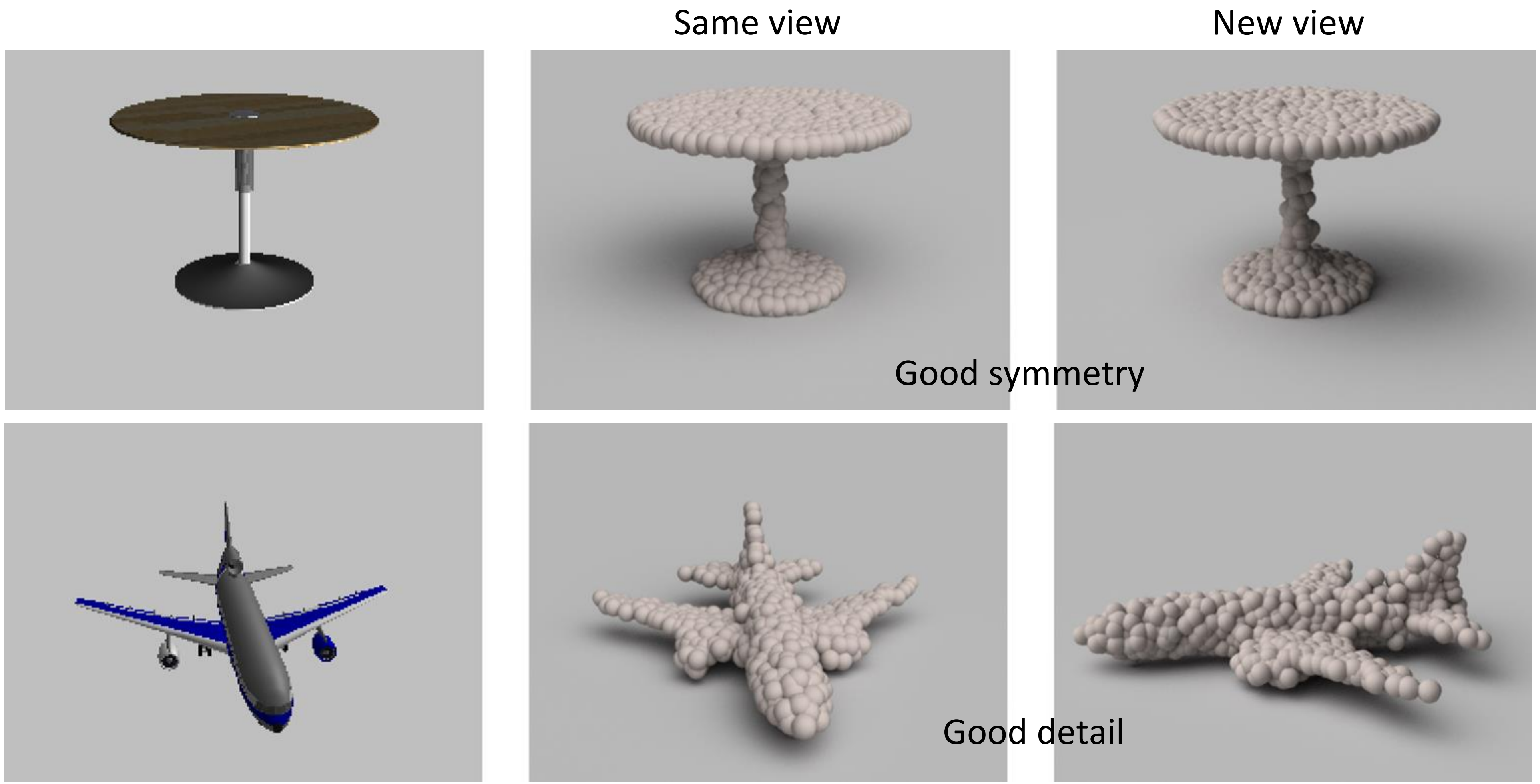


The Two Branches

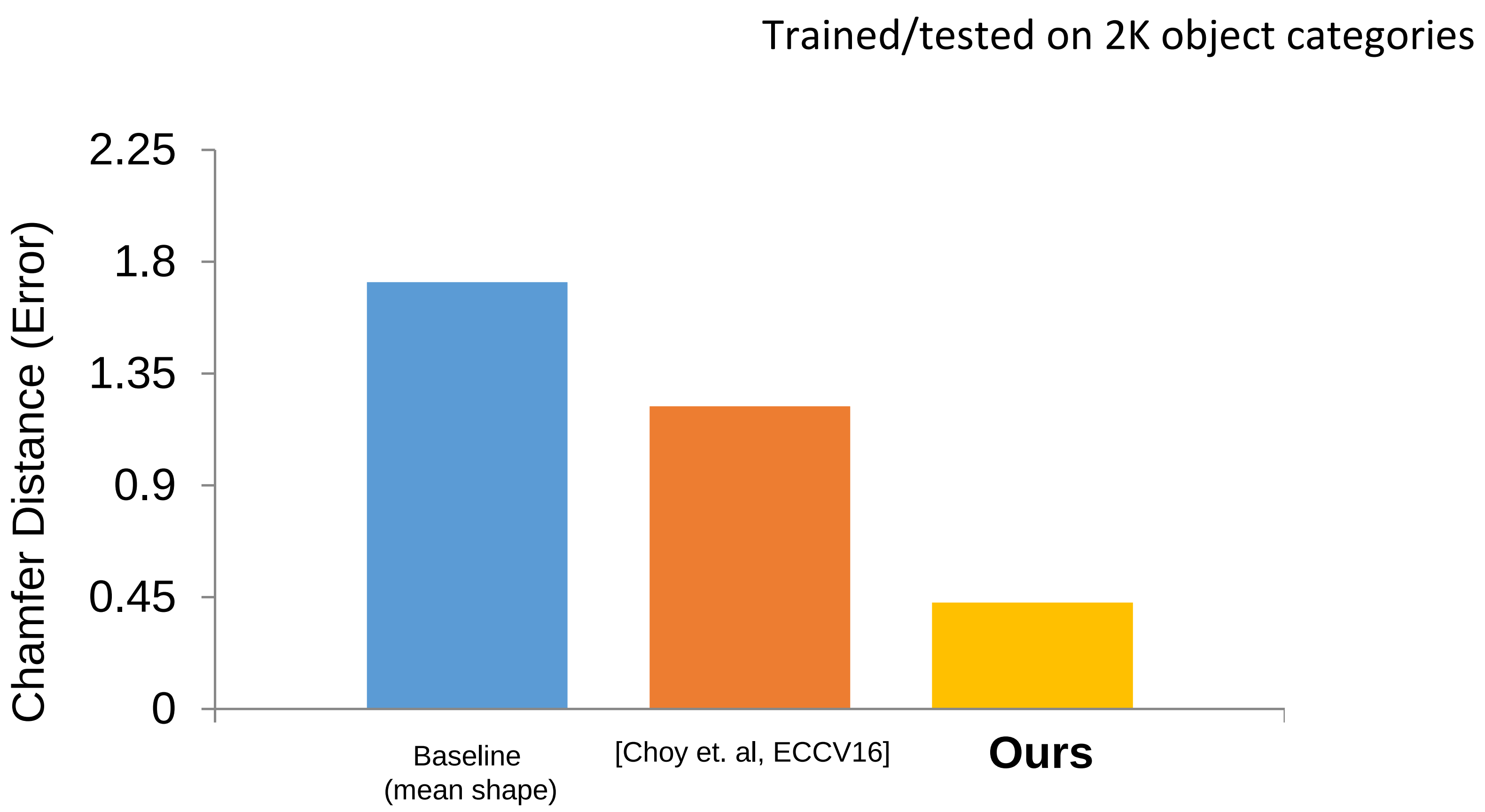
blue: deconv branch – **large, consistent, smooth** structures

red: fully-connected branch – **more intricate** structures

Example Results



Comparison to State-of-the-Art

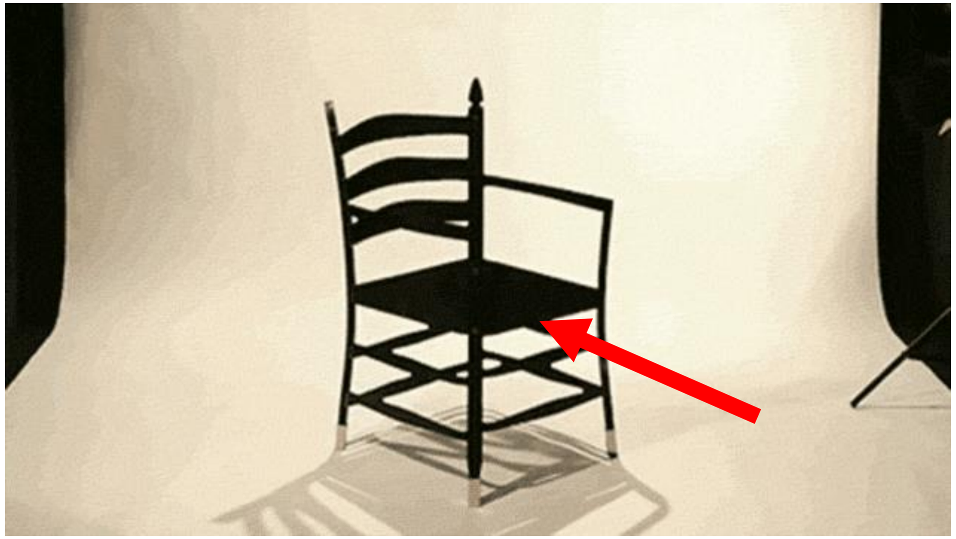


Influence of Distance Metrics

A fundamental issue: inherent ambiguity in prediction

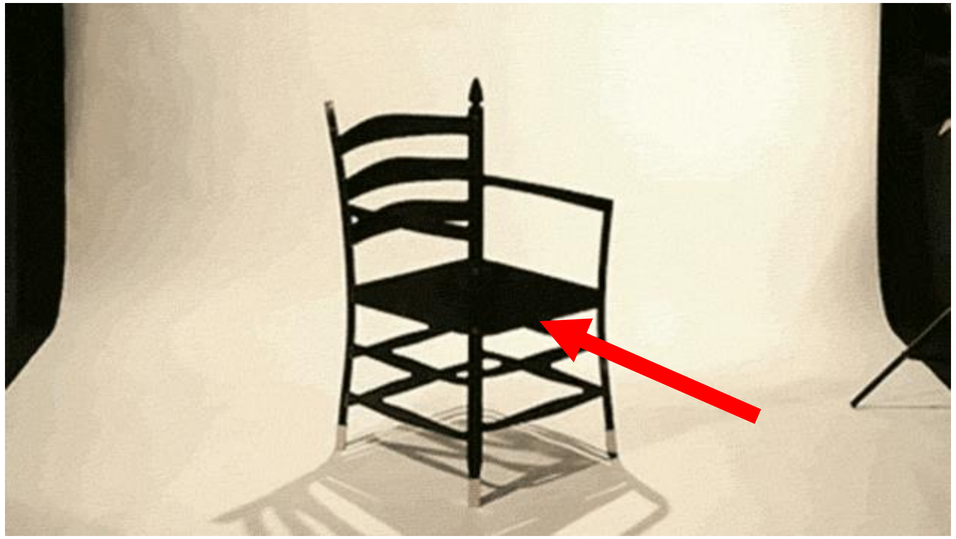
Influence of Distance Metrics

A fundamental issue: inherent ambiguity in prediction



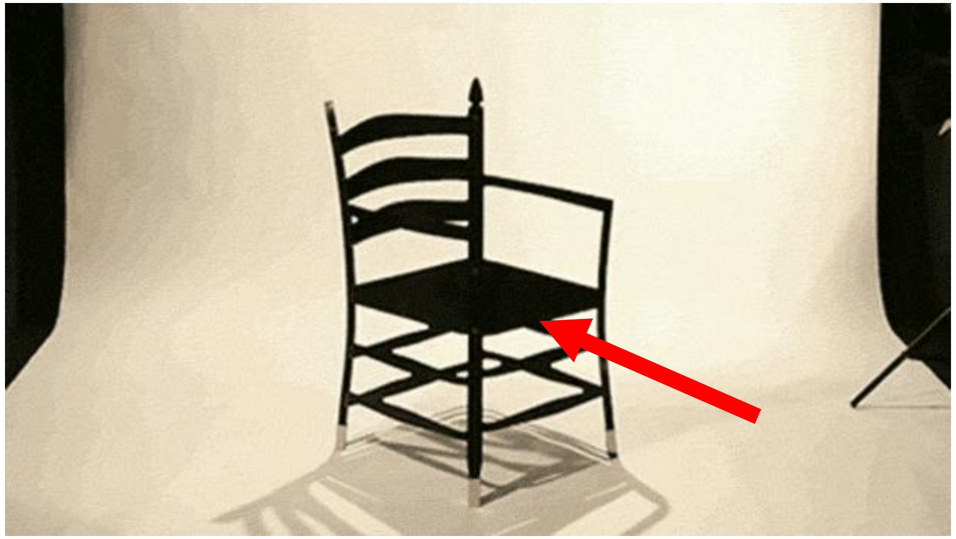
Influence of Distance Metrics

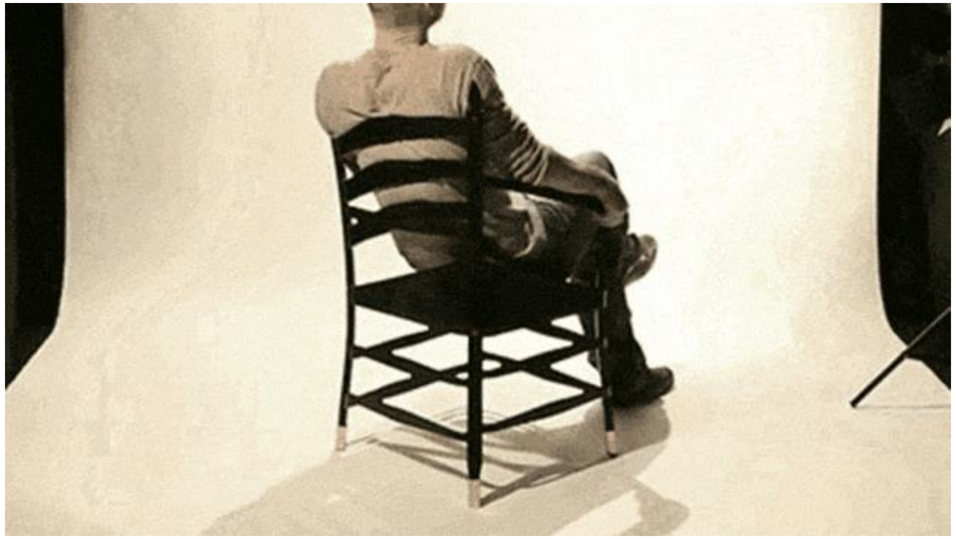
A fundamental issue: inherent ambiguity in prediction



Influence of Distance Metrics

A fundamental issue: inherent ambiguity in prediction





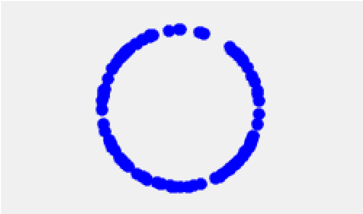
- By loss minimization, the network tends to predict a "**mean shape**" that **averages out** uncertainty

Distance Metrics Affect Mean Shapes

The mean shape carries characteristics of the distance metric

$$\bar{x} = \underset{x}{\operatorname{argmin}}\ \mathbb{E}_{s\sim\mathbb{S}}[d(x,s)]$$

continuous
hidden variable
(radius)



Input

EMD mean

Chamfer mean

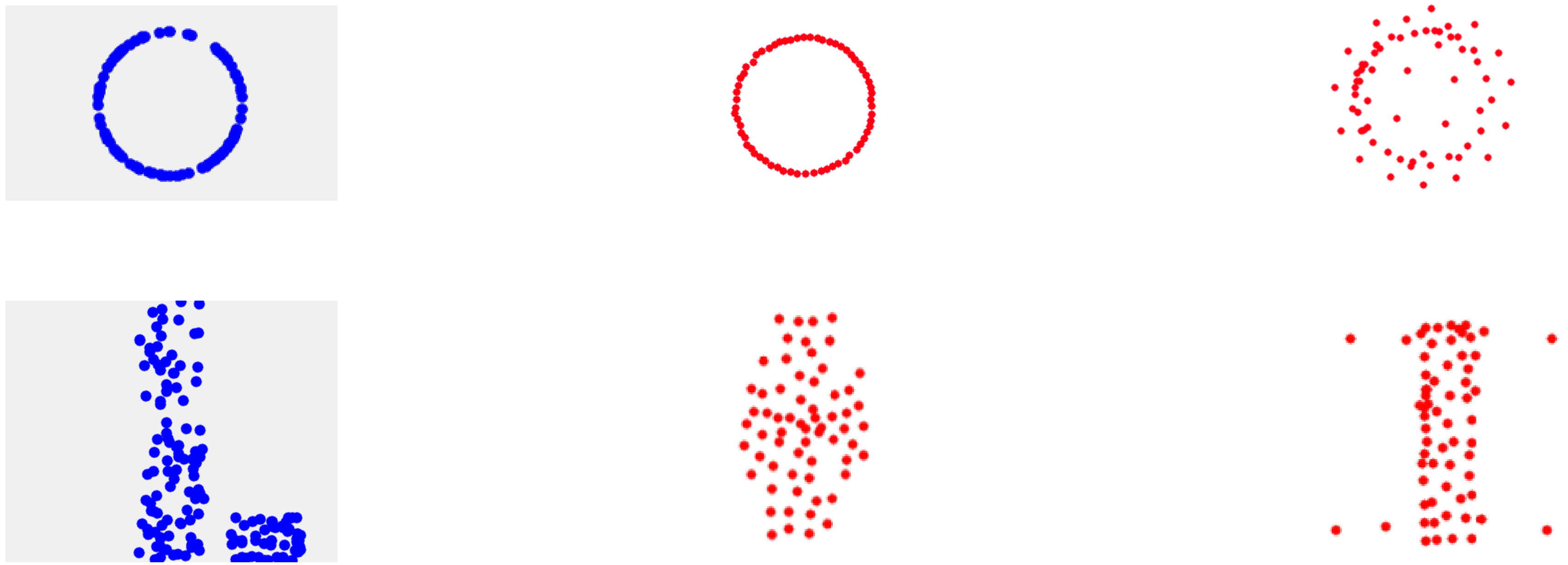
Distance Metrics Affect Mean Shapes

The mean shape carries characteristics of the distance metric

$$\bar{x} = \operatorname*{argmin}_{x} \mathbb{E}_{s \sim \mathbb{S}}[d(x, s)]$$

continuous
hidden variable
(radius)

discrete
hidden variable
(add-on location)



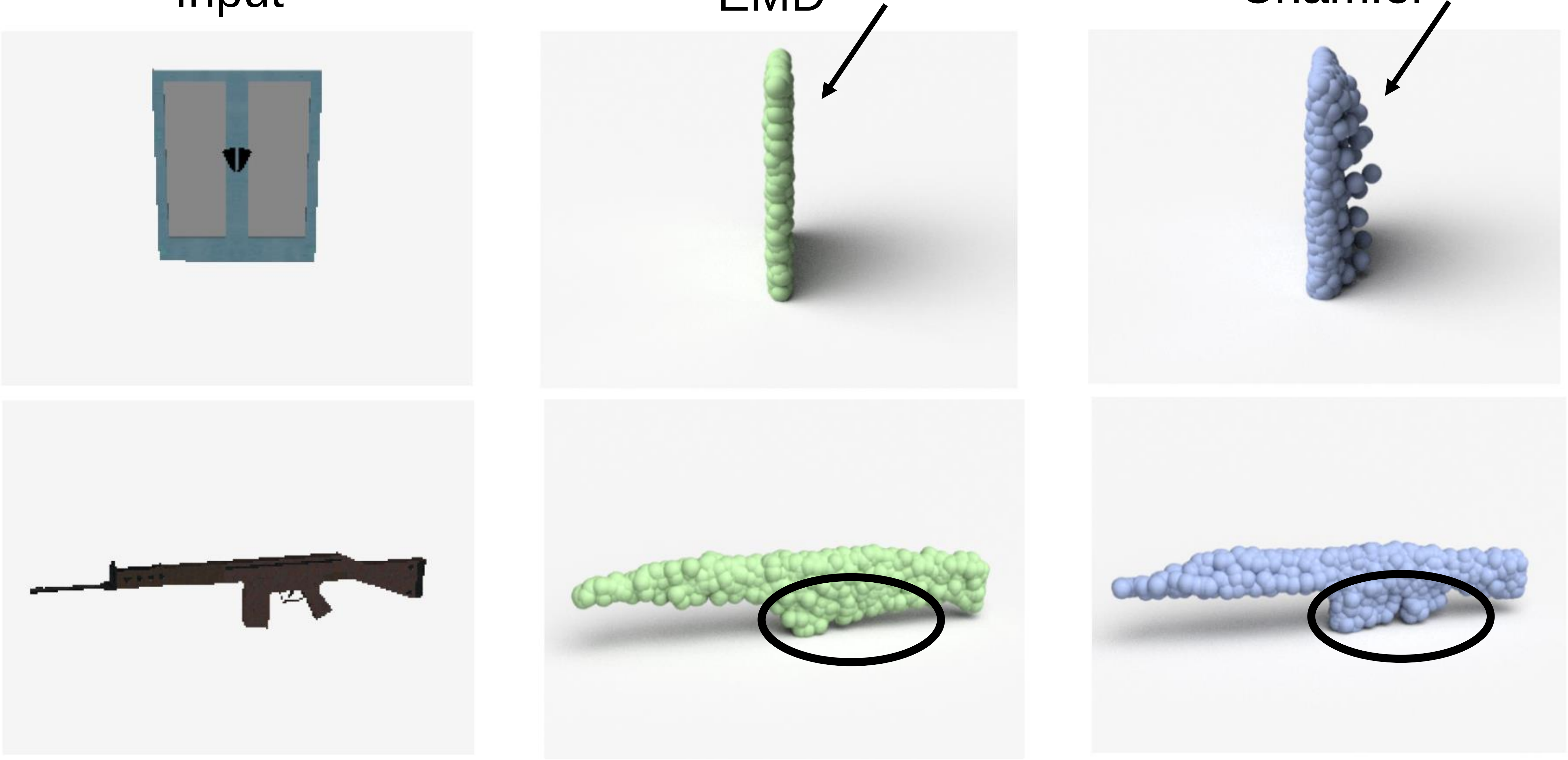
Input

EMD mean

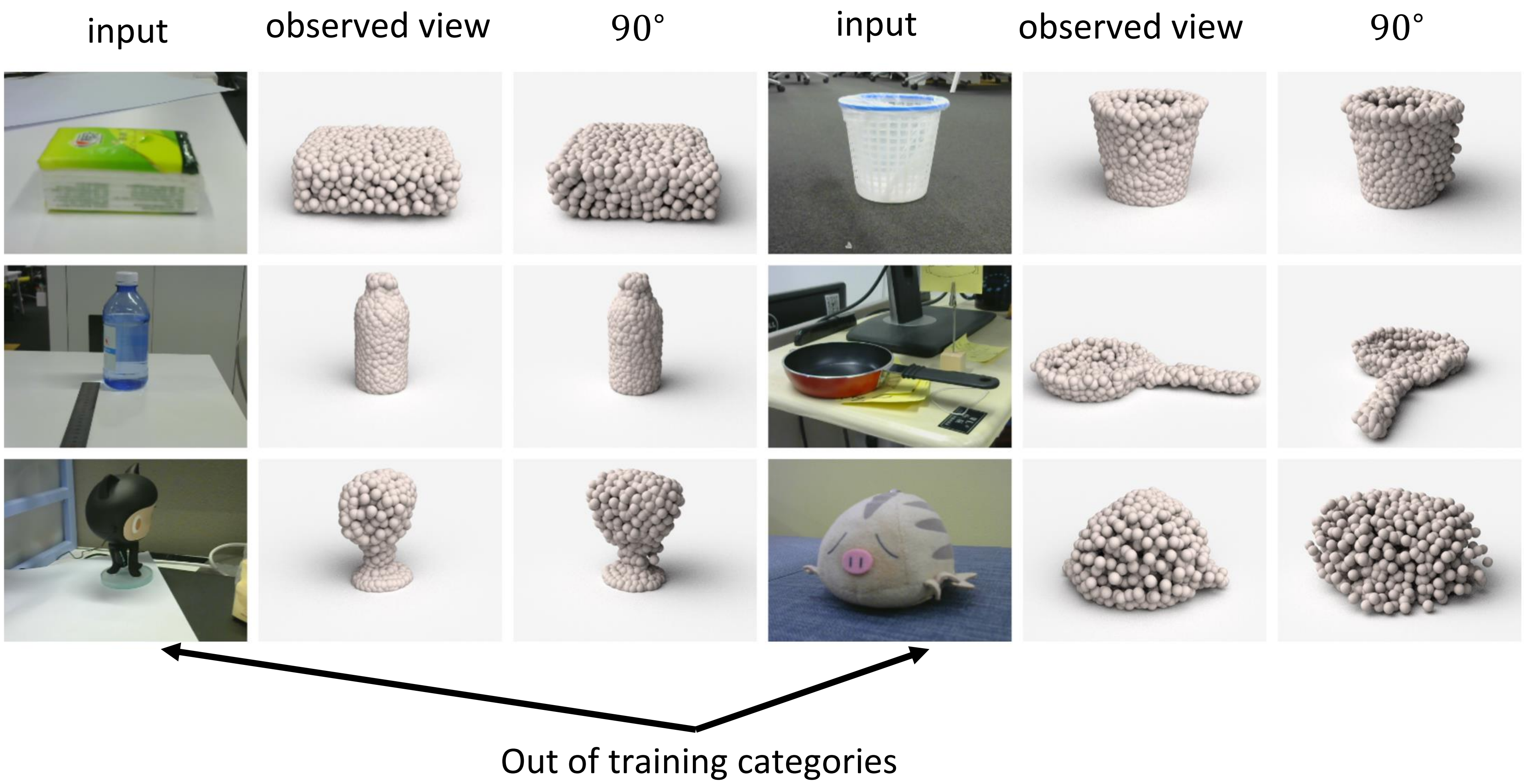
Chamfer mean

Comparison of Predictions by EMD versus CD

Input EMD Chamfer



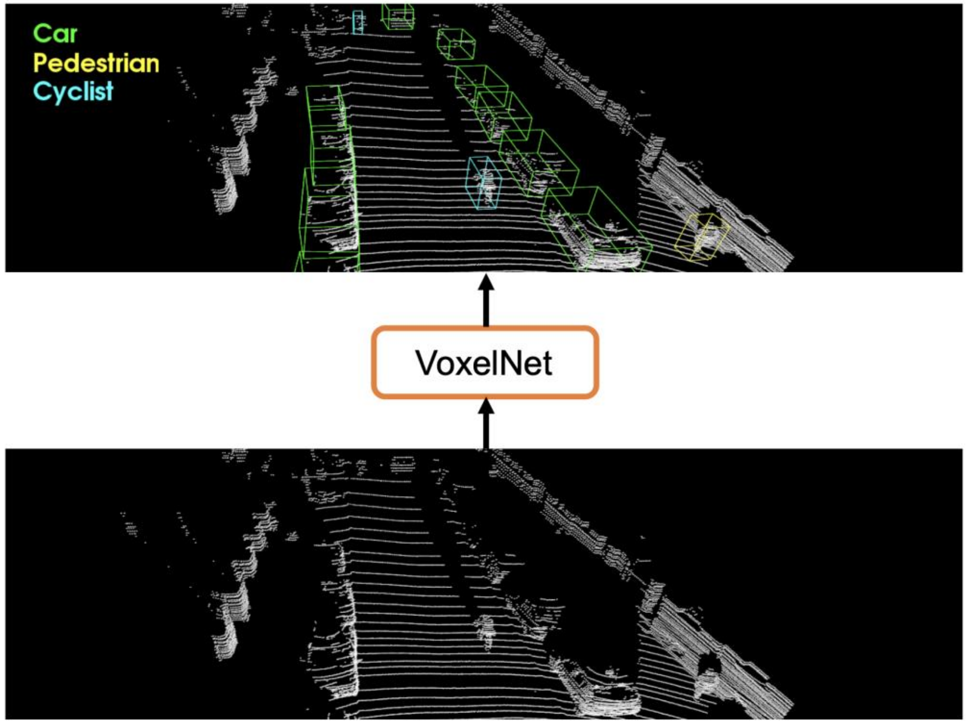
From Real Images



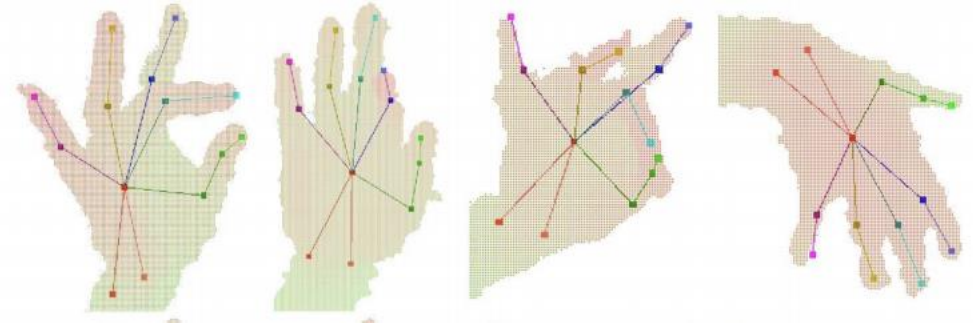
More Applications of Point Cloud Deep Learning

Applications of Point Cloud Deep Learning

- **3D object & scene understanding**



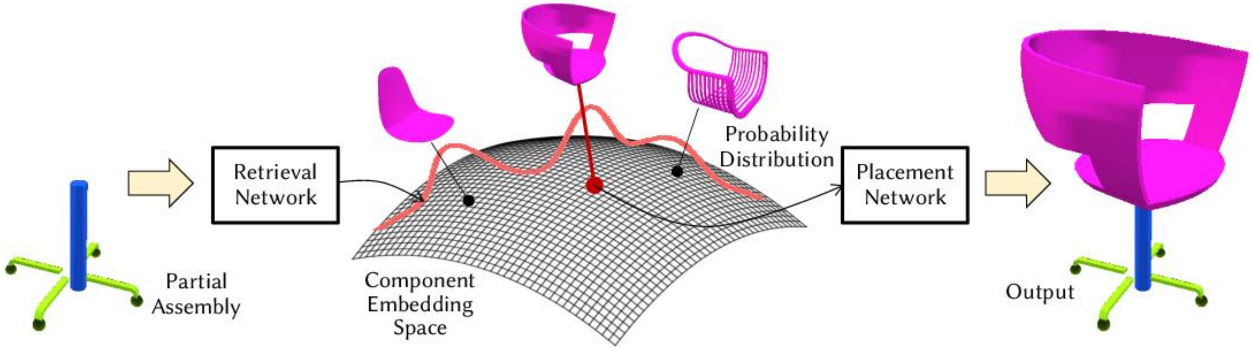
3D Object Detection [VoxelNet by Yin et al.]



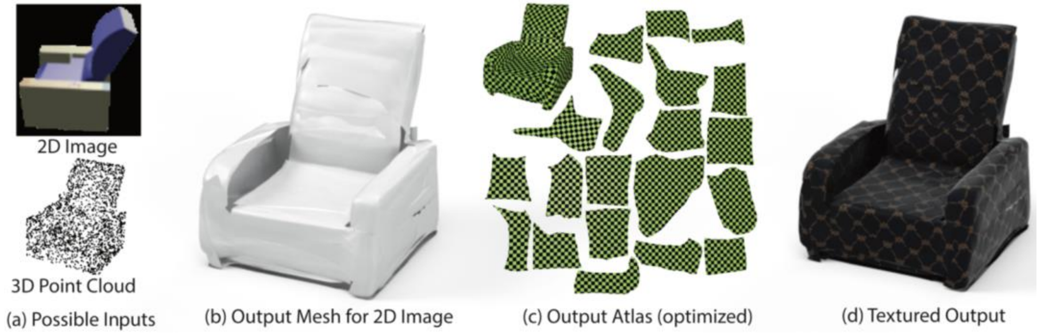
Hand Pose Estimation [Hand PointNet by Ge et al.]

Applications of Point Cloud Deep Learning

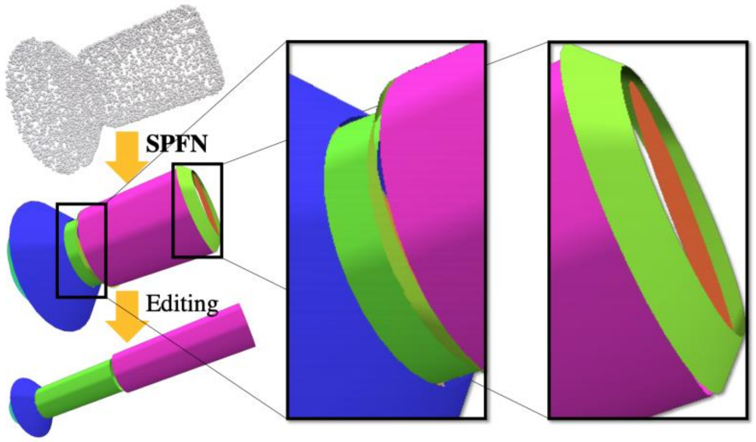
- 3D object & scene understanding
- **AI-assisted shape design**



ComplementMe [Sung et al. 2017]



AtlasNet [Groueix et al. 2018]

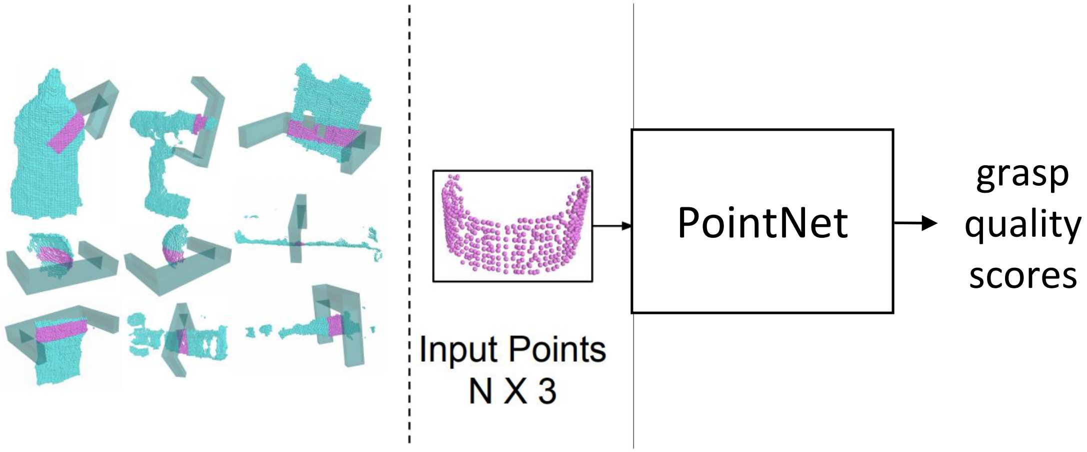


Primitive fitting [Li et al. CVPR'19]

Applications of Point Cloud Deep Learning

- 3D object & scene understanding
- AI-assisted shape design
- **Robotics: grasping, manipulation and simulation**

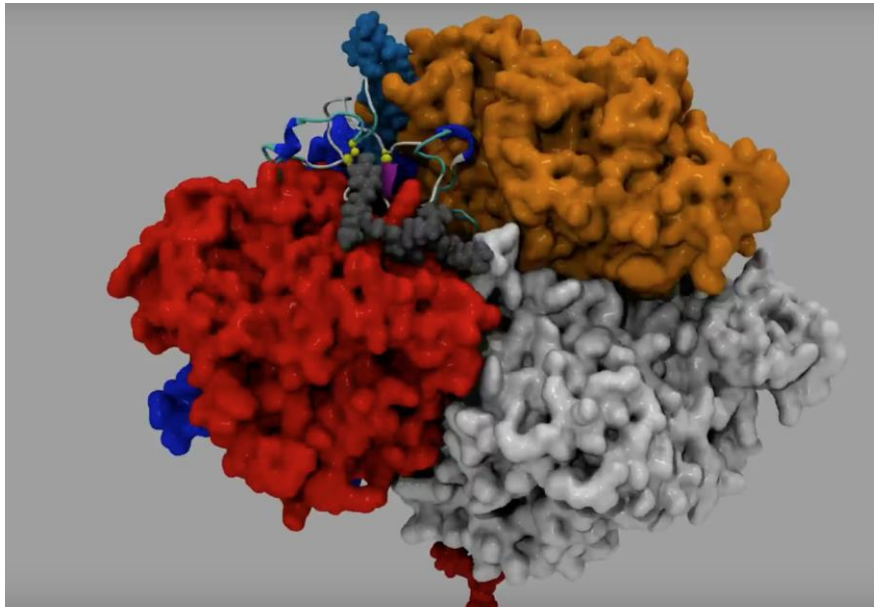
source: Ludovic Righetti



PointNetGPD by Liang et al. ICRA19

Applications of Point Cloud Deep Learning

- 3D object & scene understanding
- AI-assisted shape design
- Robotics: grasping, manipulation and simulation
- **Molecular biology: from structure to function**



source: BPC@University Greifswald

Future Directions for Point Cloud Deep Learning

Future Directions

- **Scalability**

How to scale up from processing 100k points to 1M or even 10M points?

(1024 x 1024 image ~= 1M pixels)

Trade-offs in neighborhood sampling

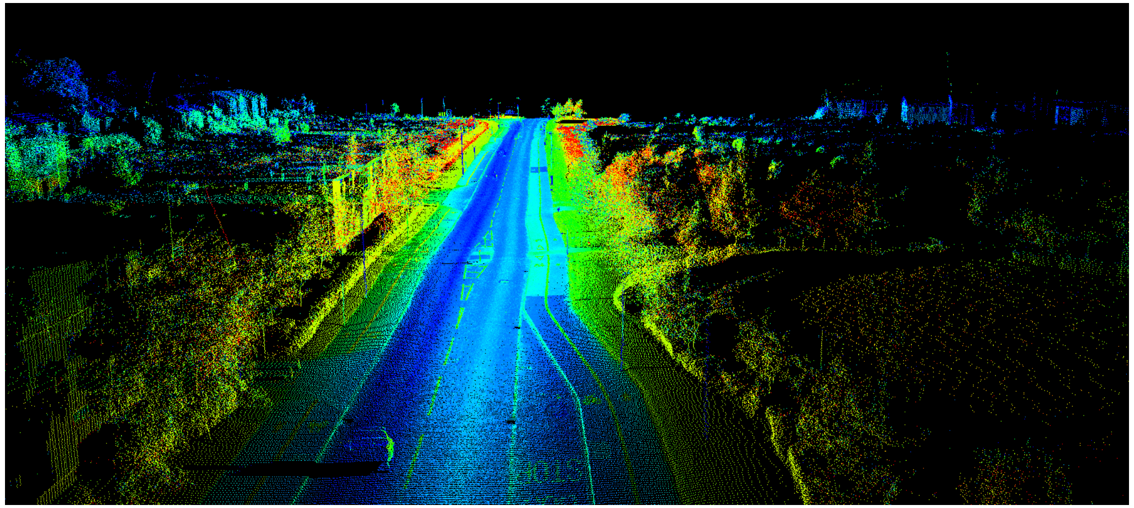
More memory efficient operators

Future Directions

- Scalability
- **Multi-modality**

RGB images

High resolution
Rich textures

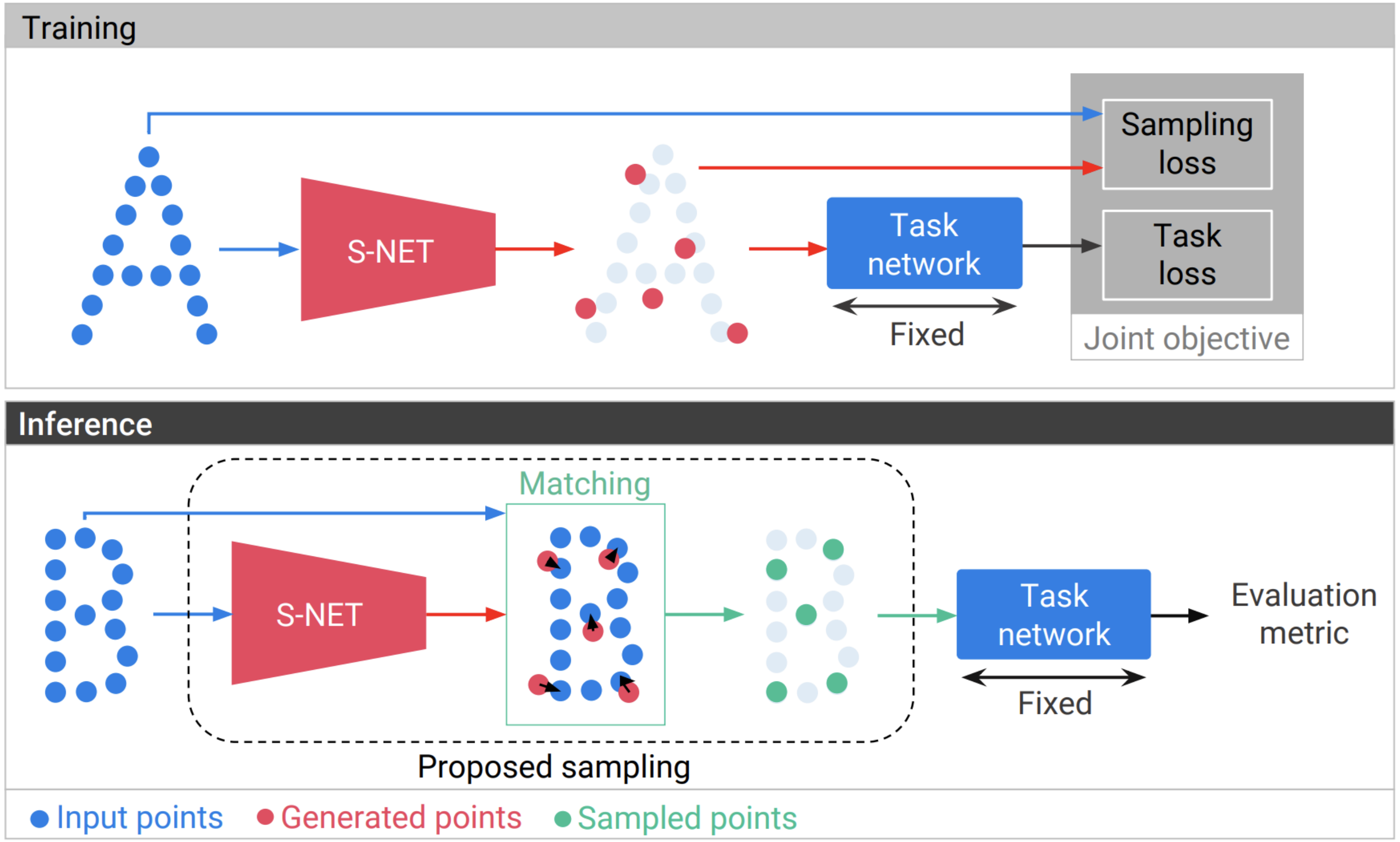


Lidar point clouds

Accurate depth
Accurate 3D geometry

Future Directions

- Scalability
- Multi-modality
- **Sampling**



Learning to sample [Dovrat et al.]

General Set / Graph Processors

- Scalability
- Multi-modality
- Sampling
- **Set processing**

Future Directions

- Scalability
- Multi-modality
- Sampling
- Set processing
- **Geometry generation**

How to generate?
How to measure quality?

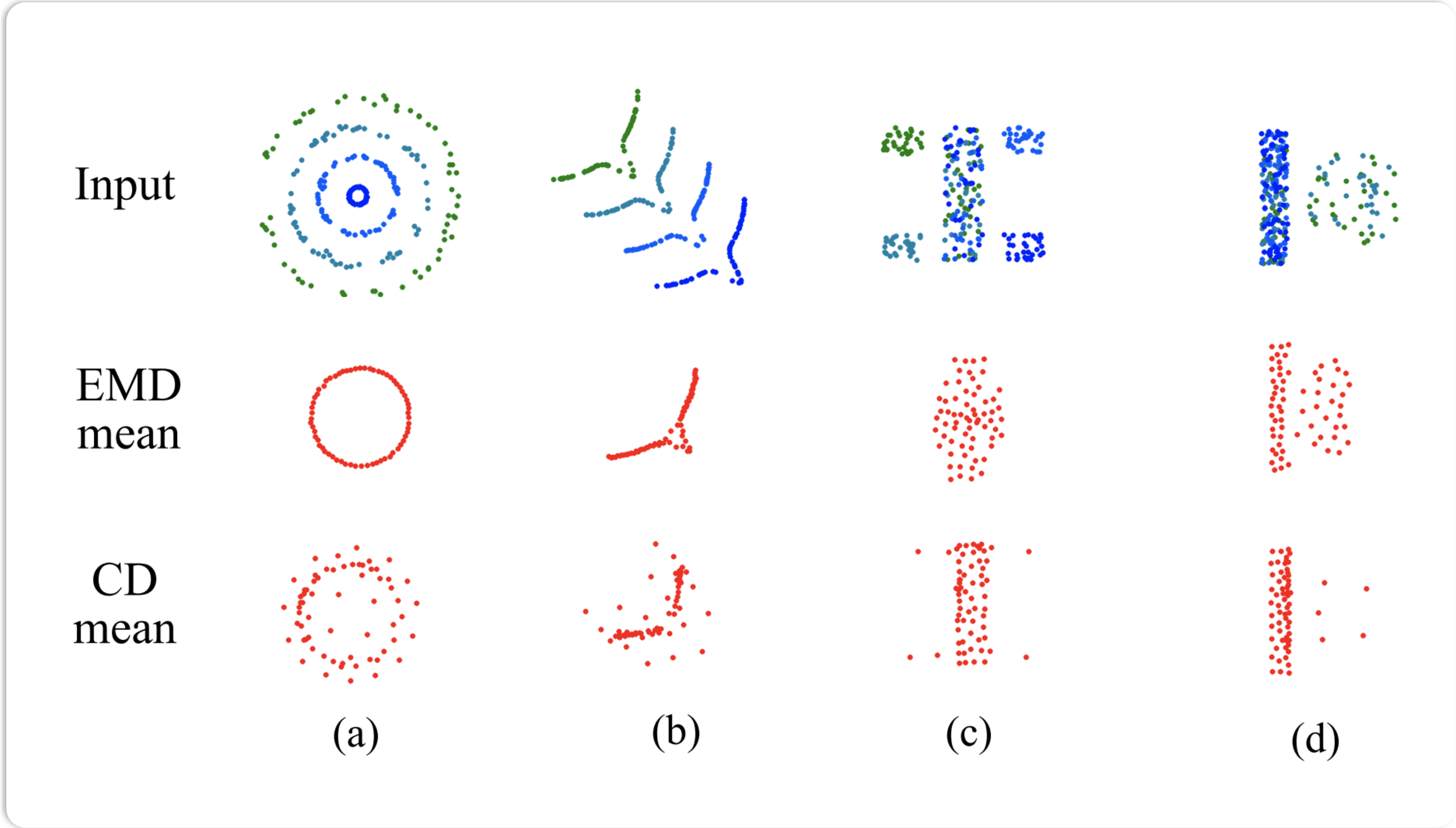
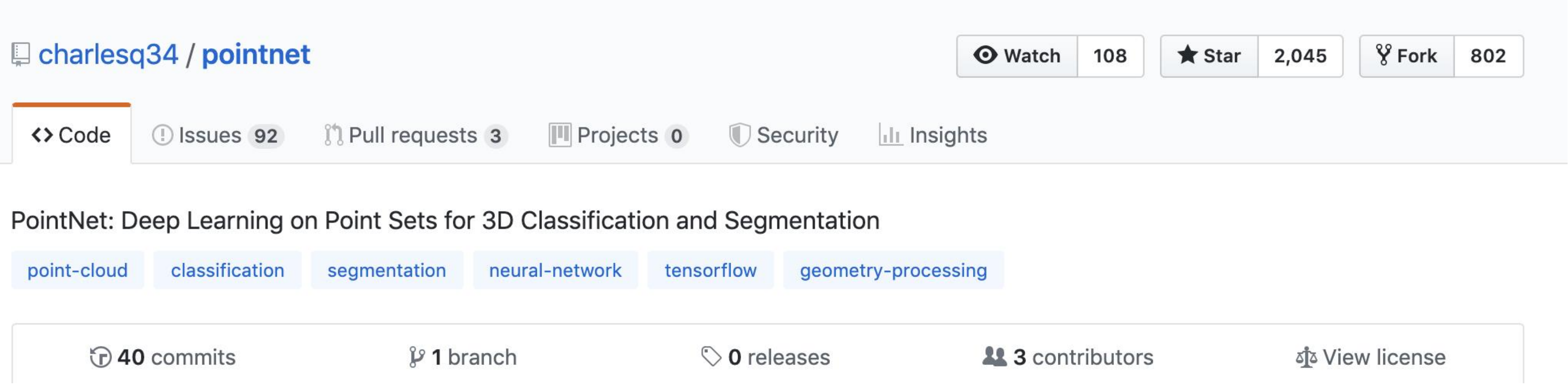


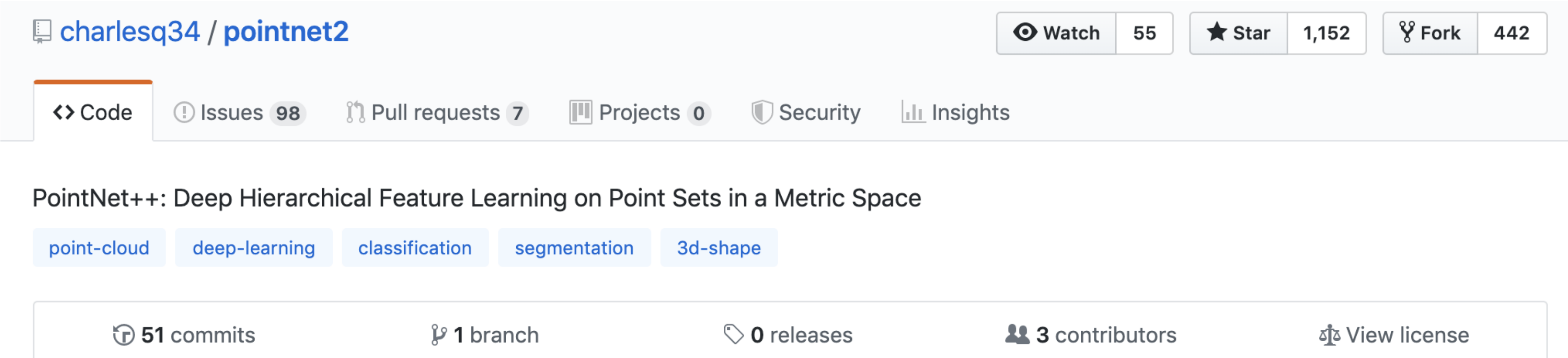
Figure from [Fan et al. CVPR 2017]

Code for PointNet, PointNet++ on GitHub

- https://github.com/charlesq34/pointnet



- https://github.com/charlesq34/pointnet2



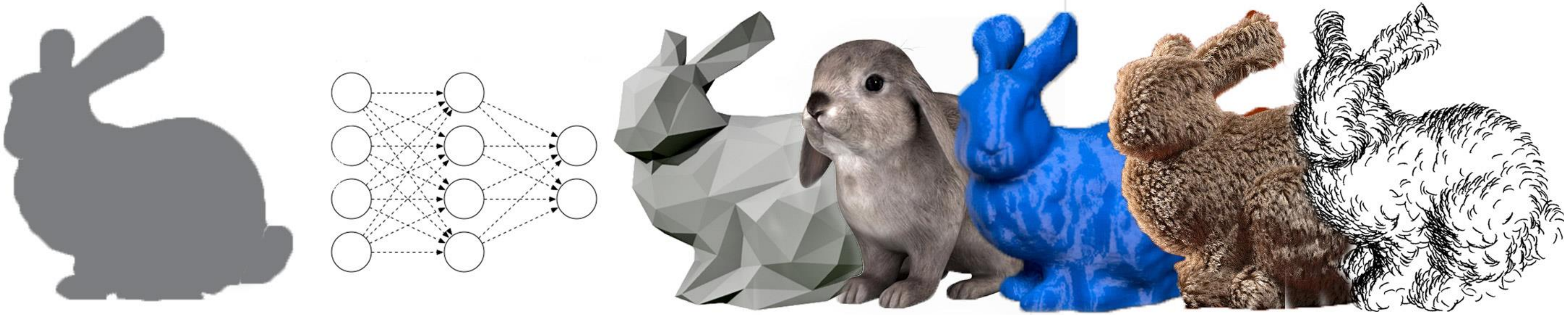
Acknowledgements

- Collaborators:
  - Current/past students: Xingyu Liu, Kaichun Mo, Charles Qi, Hao Su, Minhyuk Sung, Eric Yi
  - Current/past postdocs: Or Litany
  - Senior: Kaiming He

Charles Qi

Hao Su

Course Information (slides/code/comments)



http://geometry.cs.ucl.ac.uk/creativeai/