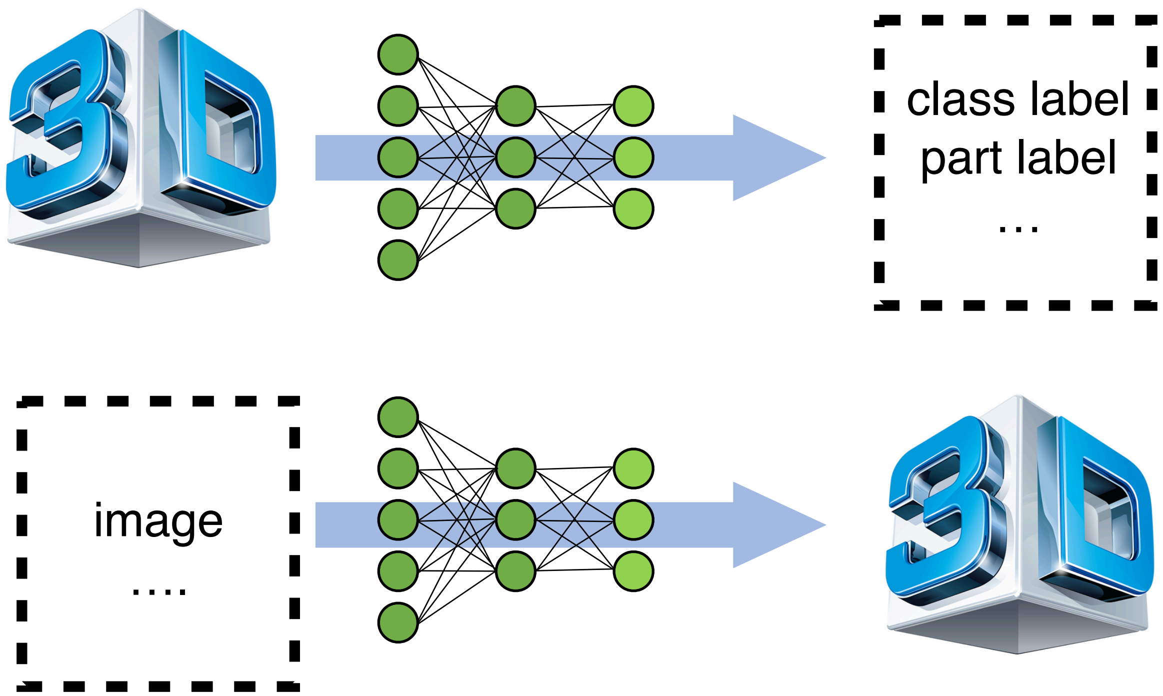
CS468: 3D Deep Learning on Point Cloud Data



Hao Su

Stanford University

May 10, 2017

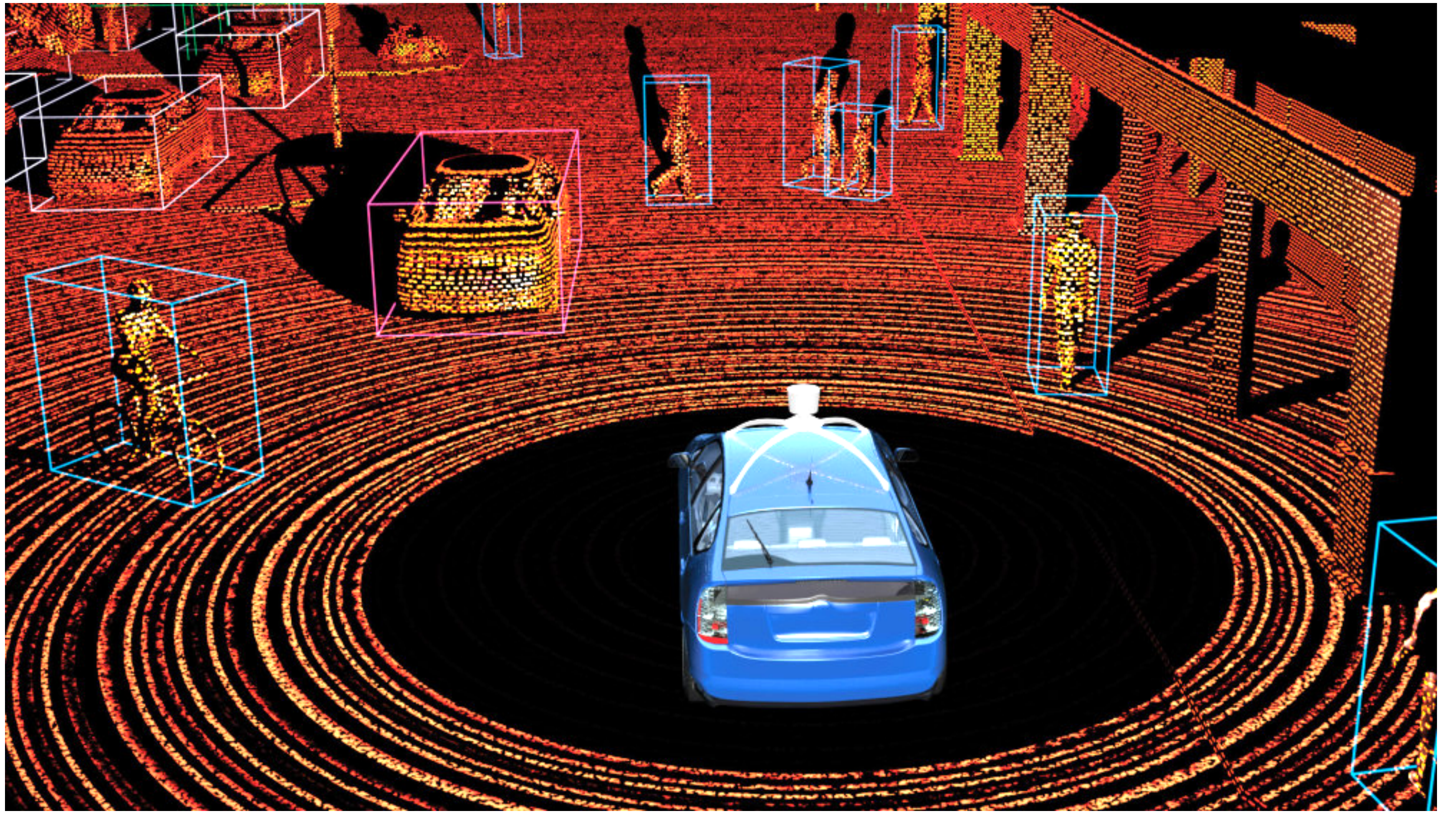
Agenda

- **Point cloud analysis**
  - PointNet
  - PointNet++
- Joint embedding learning for cross-modality image-shape retrieval

Applications of Point Set Learning

- **Robot Perception**

What and where are the objects in a LiDAR scanned scene?

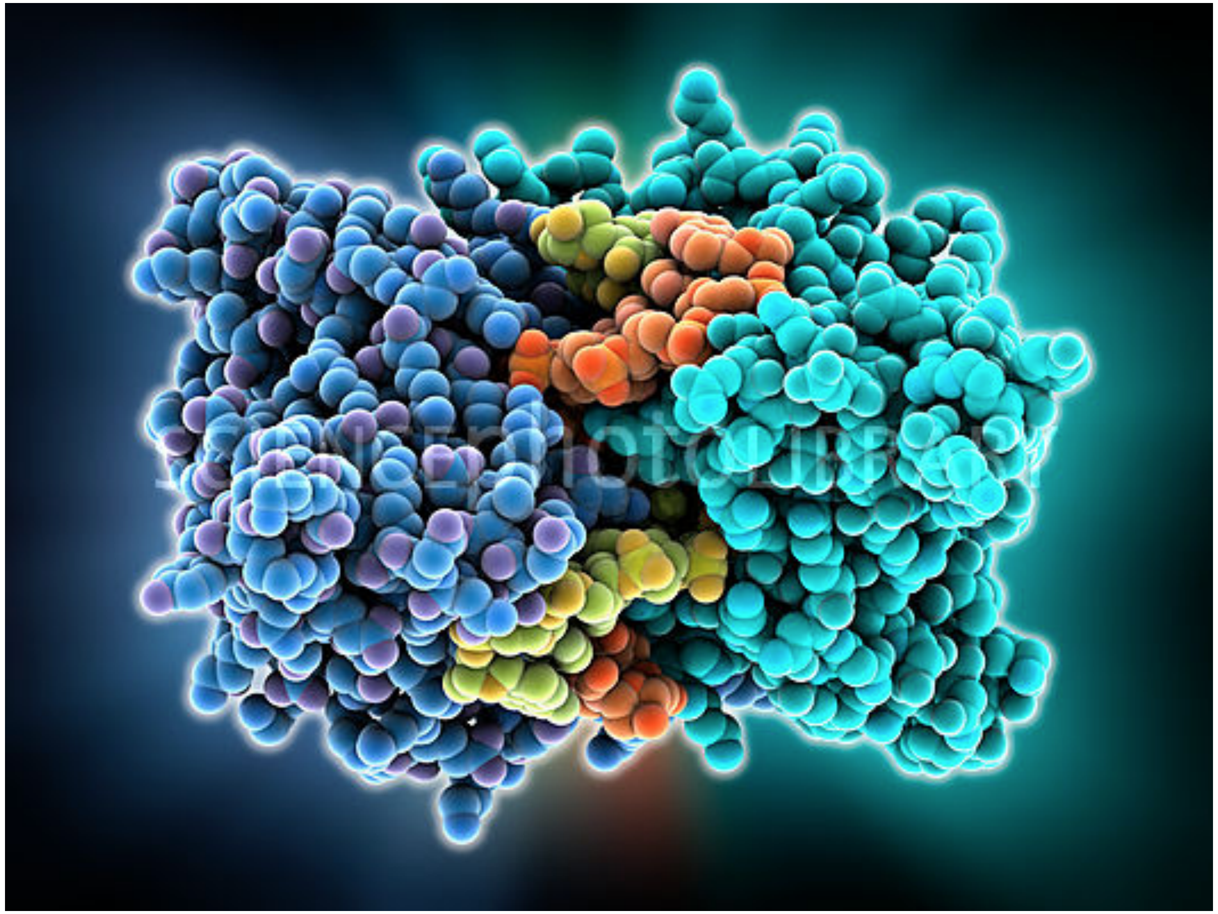


https://3dprint.com/116569/self-driving-cars-privacy/

Applications of Point Set Learning

- **Molecular Biology**

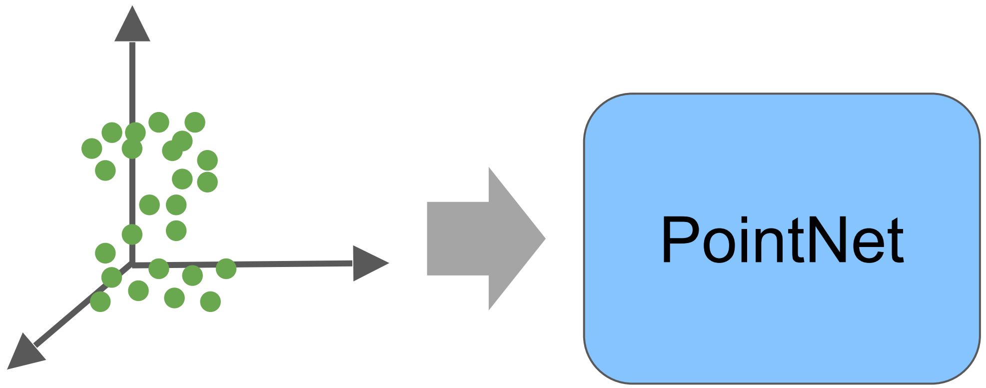
Can we infer an enzyme's category (reactions they catalyze) from its structure?



EcoRV restriction enzyme molecule, LAGUNA DESIGN/SCIENCE PHOTO LIBRARY

Directly process point cloud data

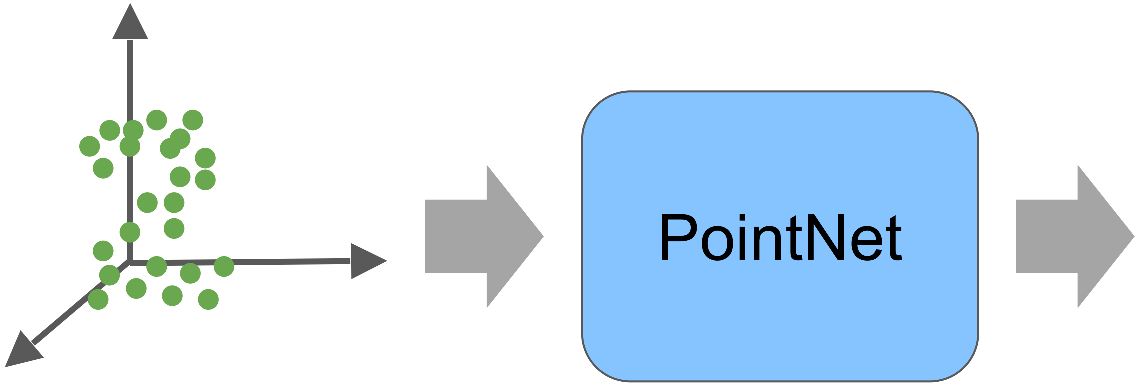
End-to-end learning for **unstructured, unordered** point data



Directly process point cloud data

End-to-end learning for **unstructured, unordered** point data

Unified framework for various tasks



Object Classification

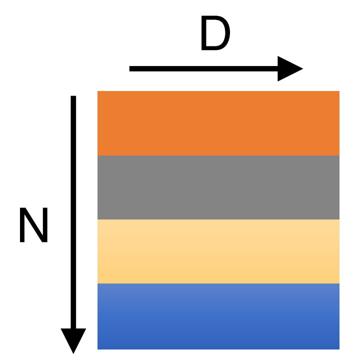
Part Segmentation

Scene Parsing

...

Properties of a desired neural network on point clouds

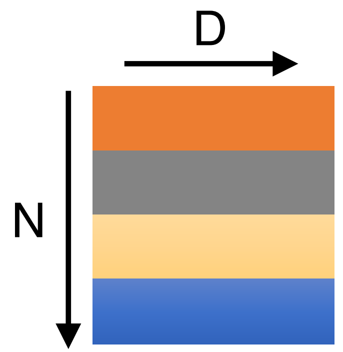
Point cloud: N **orderless** points, each represented by a D dim coordinate



2D array representation

Properties of a desired neural network on point clouds

Point cloud: N **orderless** points, each represented by a D dim coordinate



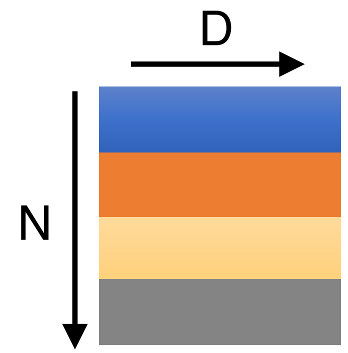
2D array representation

Permutation invariance

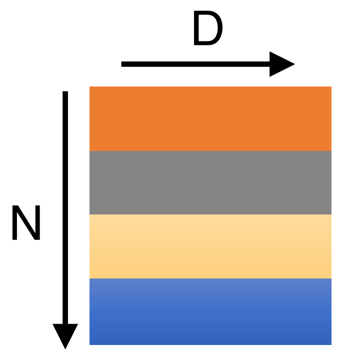
Transformation invariance

Properties of a desired neural network on point clouds

Point cloud: N **orderless** points, each represented by a D dim coordinate



represents the same **set** as



2D array representation

Permutation invariance

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:

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

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

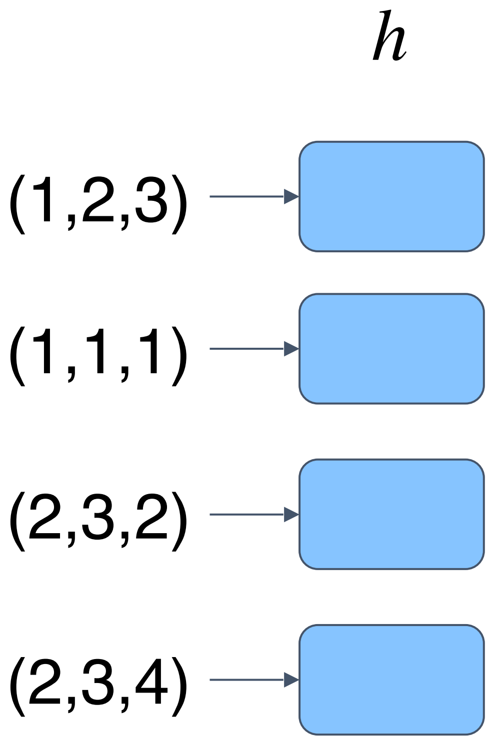
…

Construct symmetric function family

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

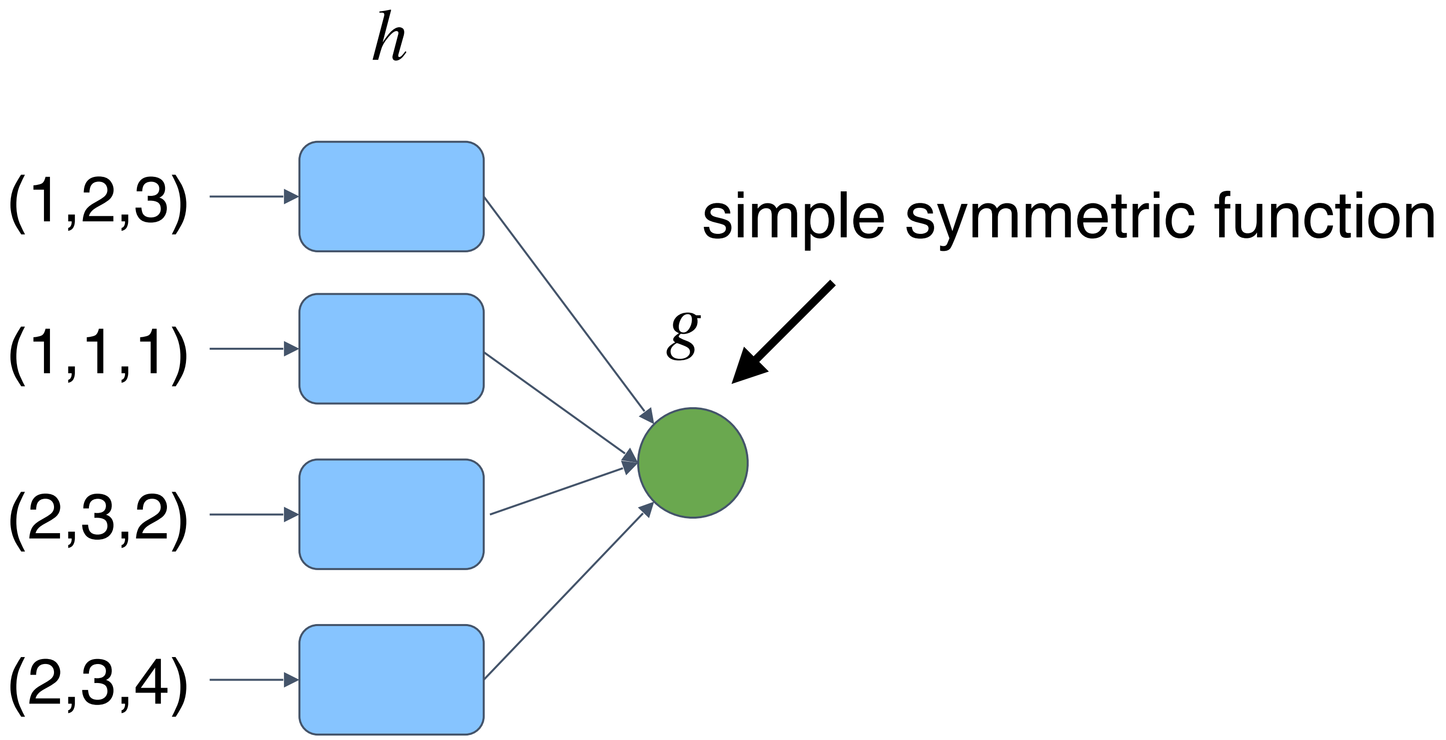
Construct symmetric function family

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



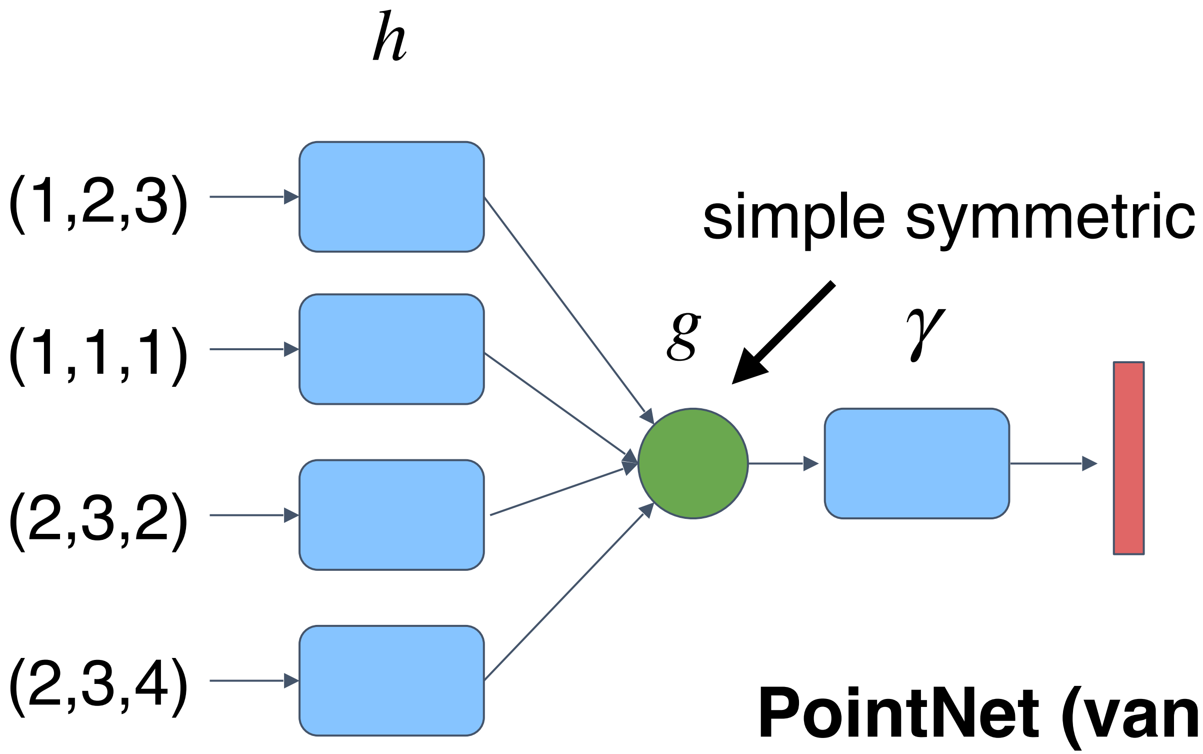
Construct symmetric function family

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



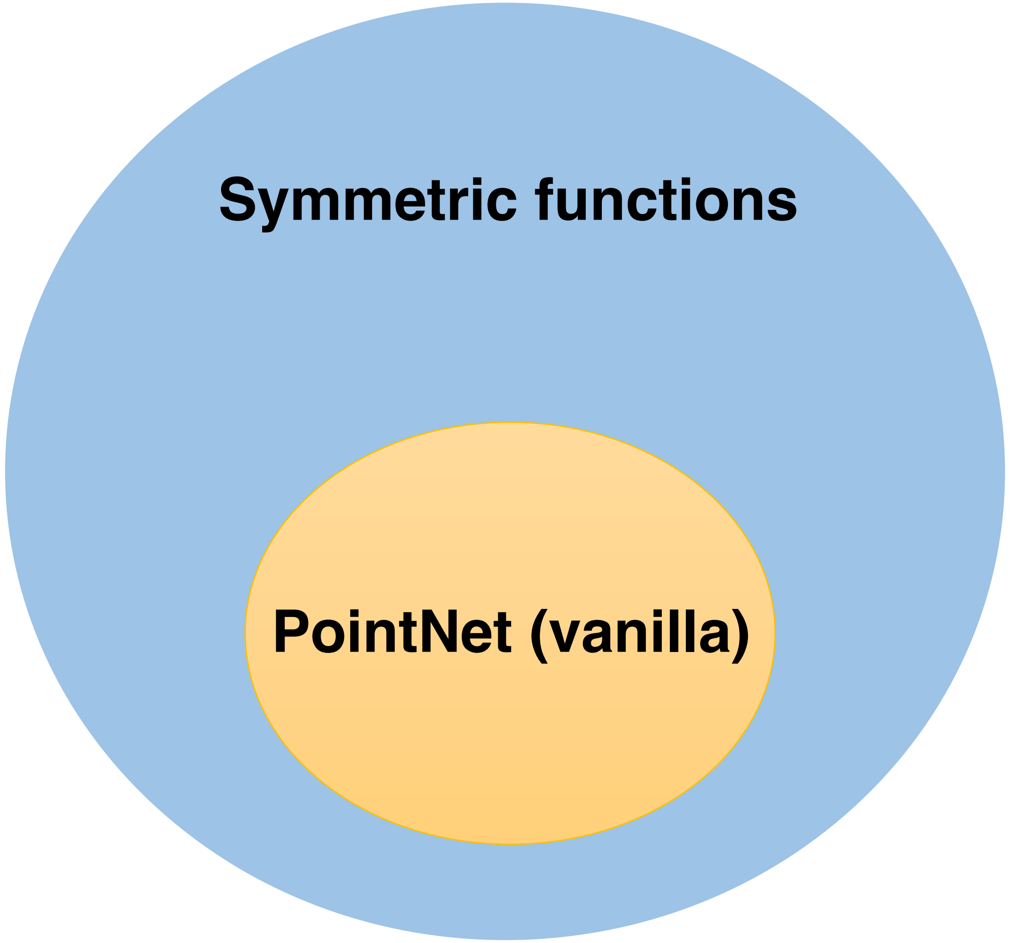
Construct symmetric function family

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



PointNet (vanilla)

Q: What symmetric functions can be constructed by PointNet?



A: Universal approximation to **continuous** symmetric functions

Theorem:

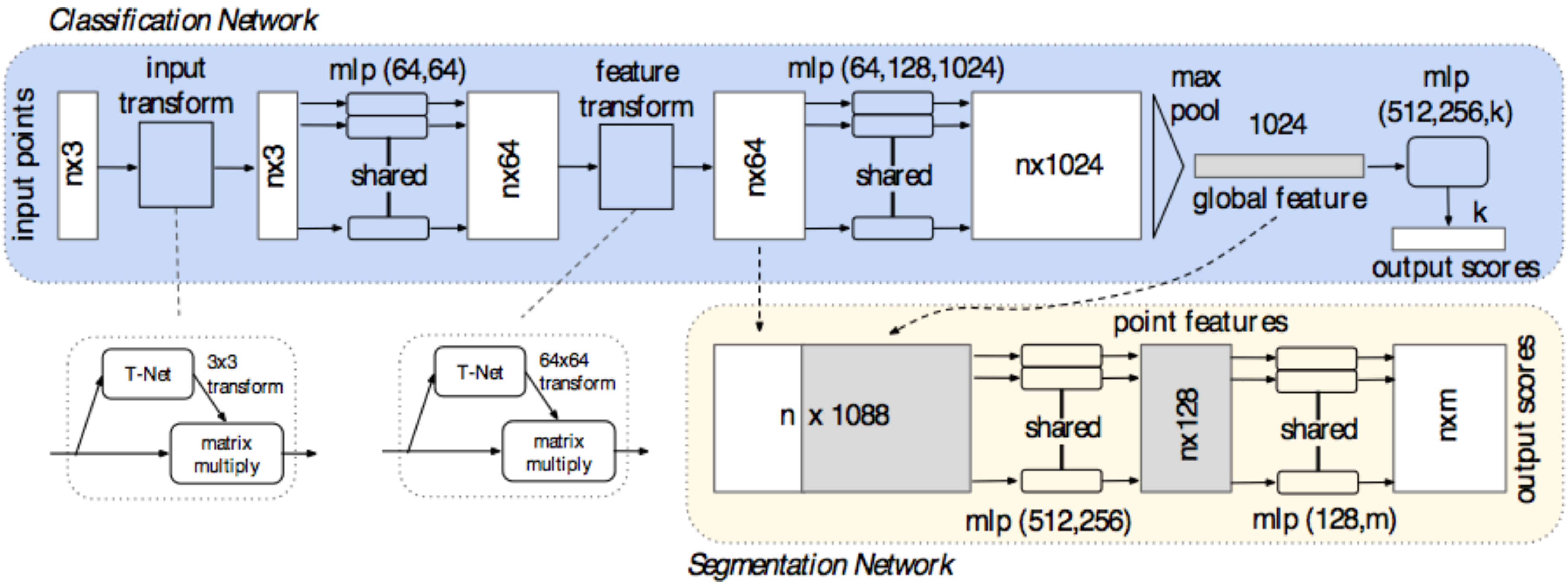
A Hausdorff continuous symmetric function $f : 2^{\mathcal{X}} \rightarrow \mathbb{R}$ can be arbitrarily approximated by PointNet.

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

$S \subseteq \mathbb{R}^d,$

PointNet (vanilla)

PointNet Architecture

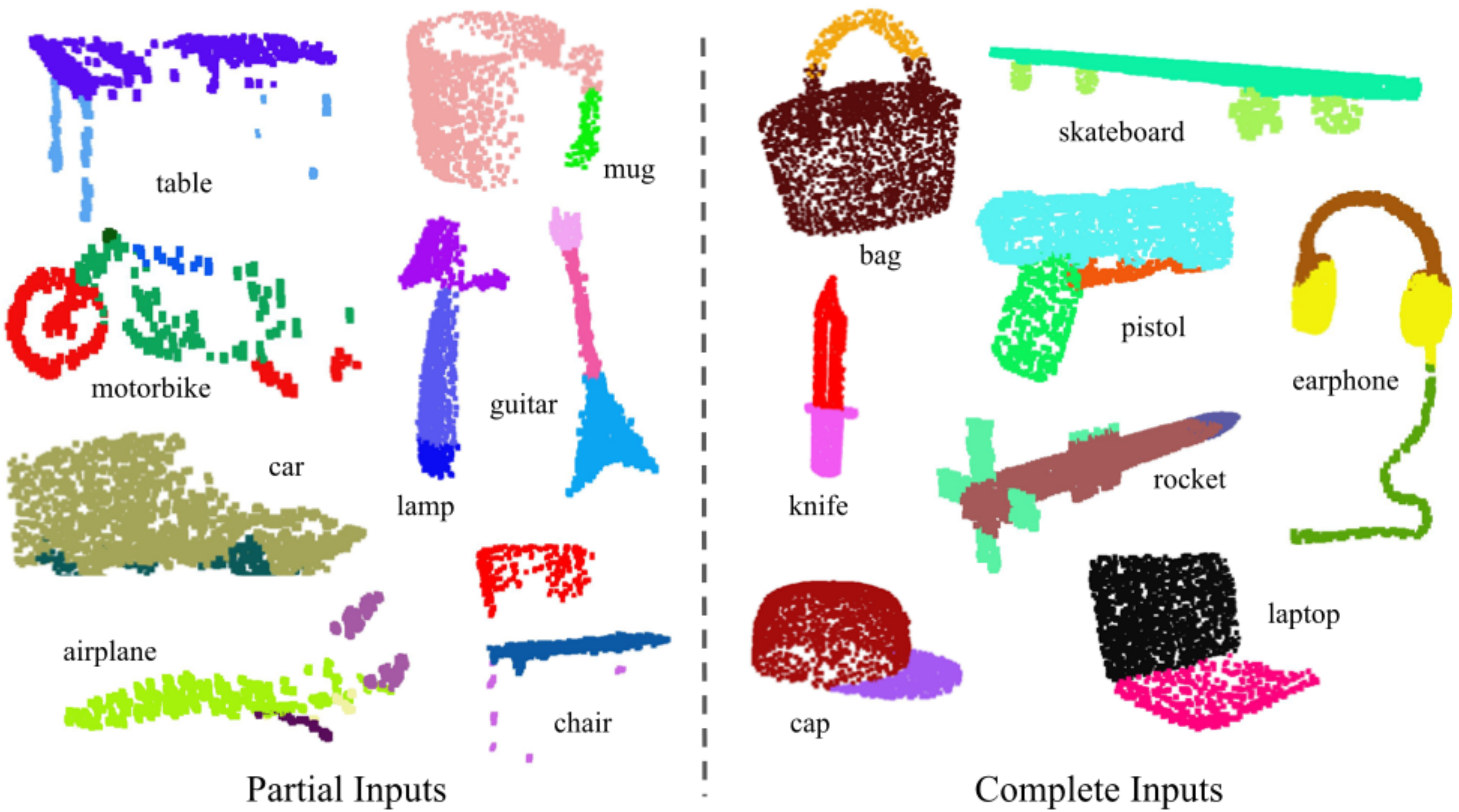


Results on Object Classification

Object Classification Accuracy on ModelNet40

	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**

Results on Object Part Segmentation



Results on Object Part Segmentation

Part Segmentation mIoU on ShapeNet Part Dataset

	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**

Results on Semantic Scene Segmentation

Results on Semantic Scene Parsing

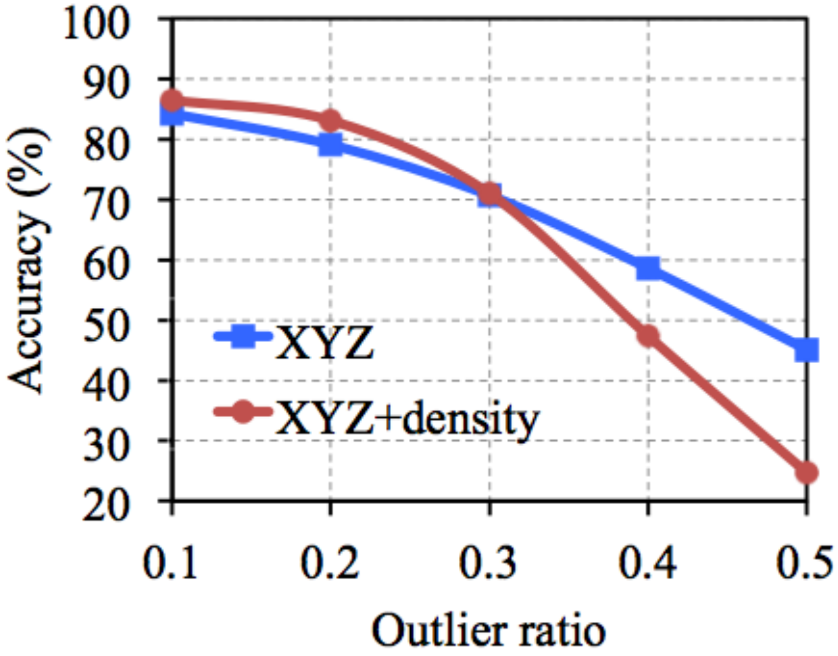
Semantic Segmentation (point based)
on Stanford Semantic Parsing dataset

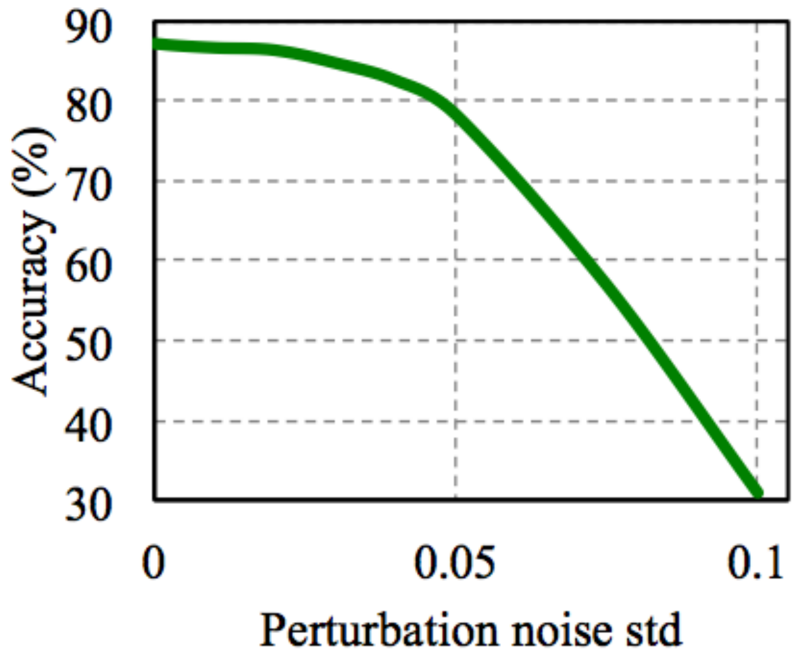
	mean IoU	overall accuracy
Ours baseline	20.12	53.19
Ours PointNet	**47.71**	**78.62**

3D Object Detection (bounding box based)

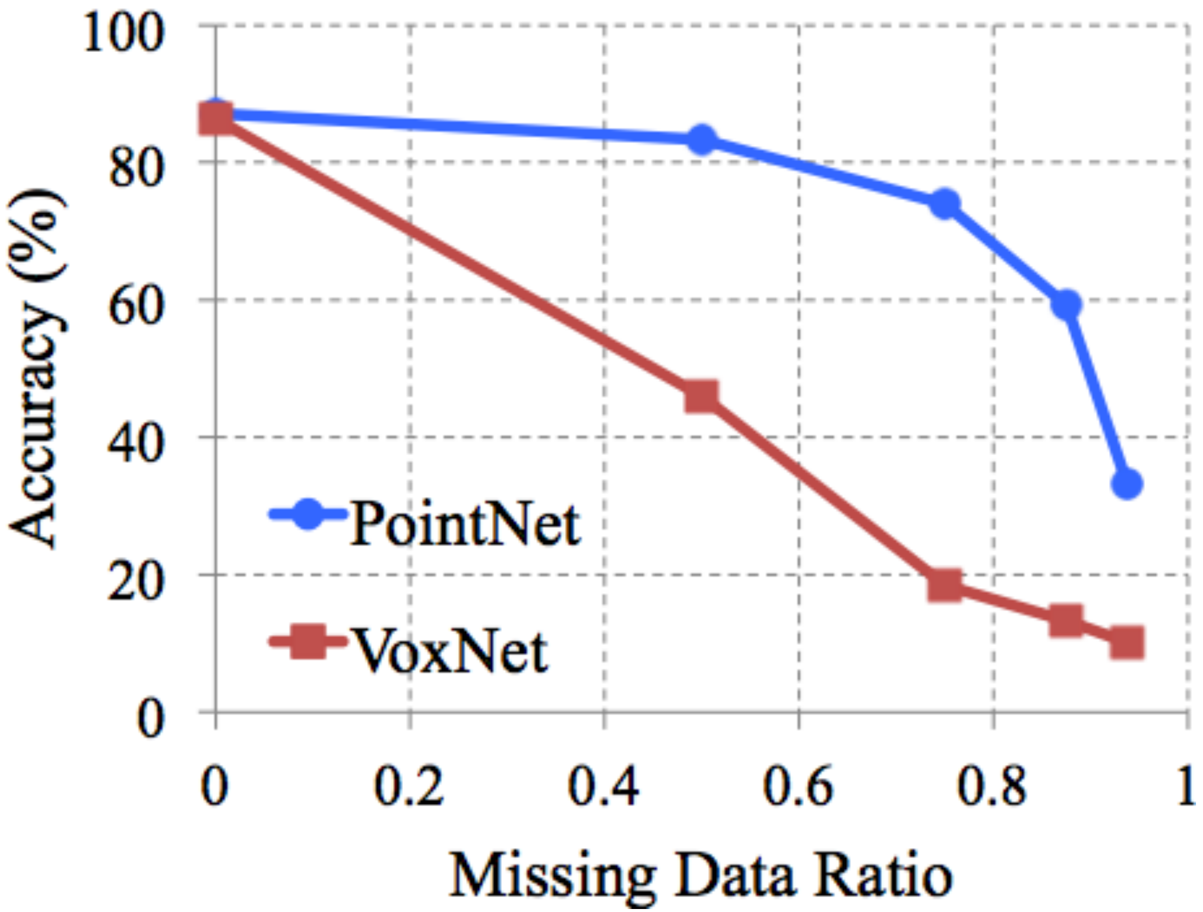
	table	chair	sofa	board	mean
# instance	455	1363	55	137	
Armeni et al. [2]	46.02	16.15	**6.78**	3.91	18.22
Ours	**46.67**	**33.80**	4.76	**11.72**	**24.24**

Robustness to Data Corruption

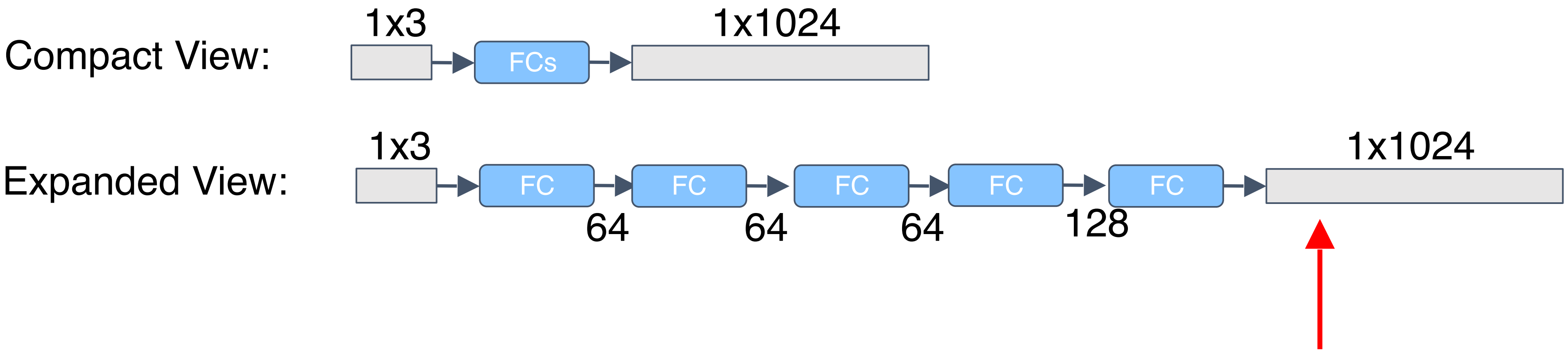




Robustness to Data Corruption



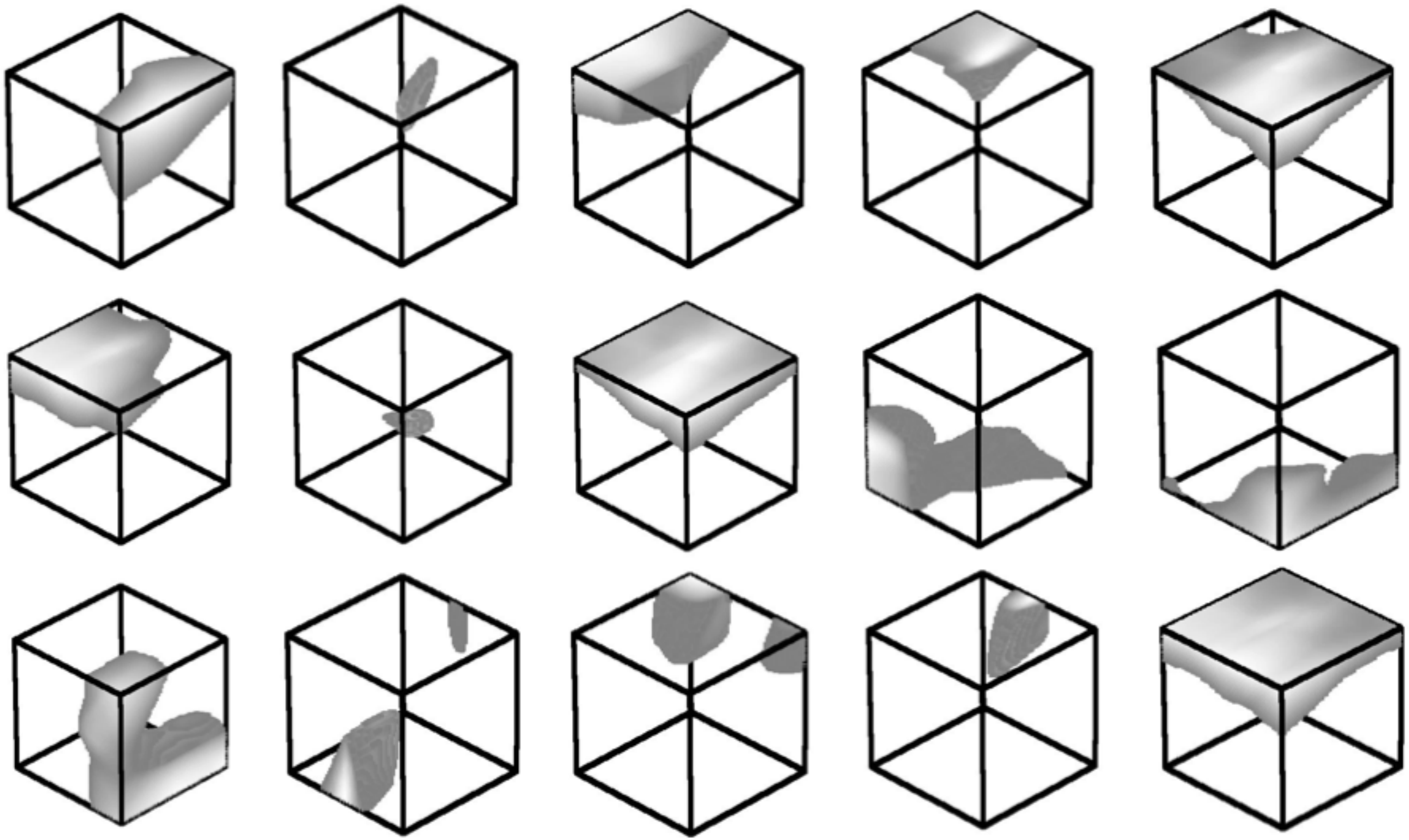
Visualizing Point Functions



Which input point will activate neuron j?

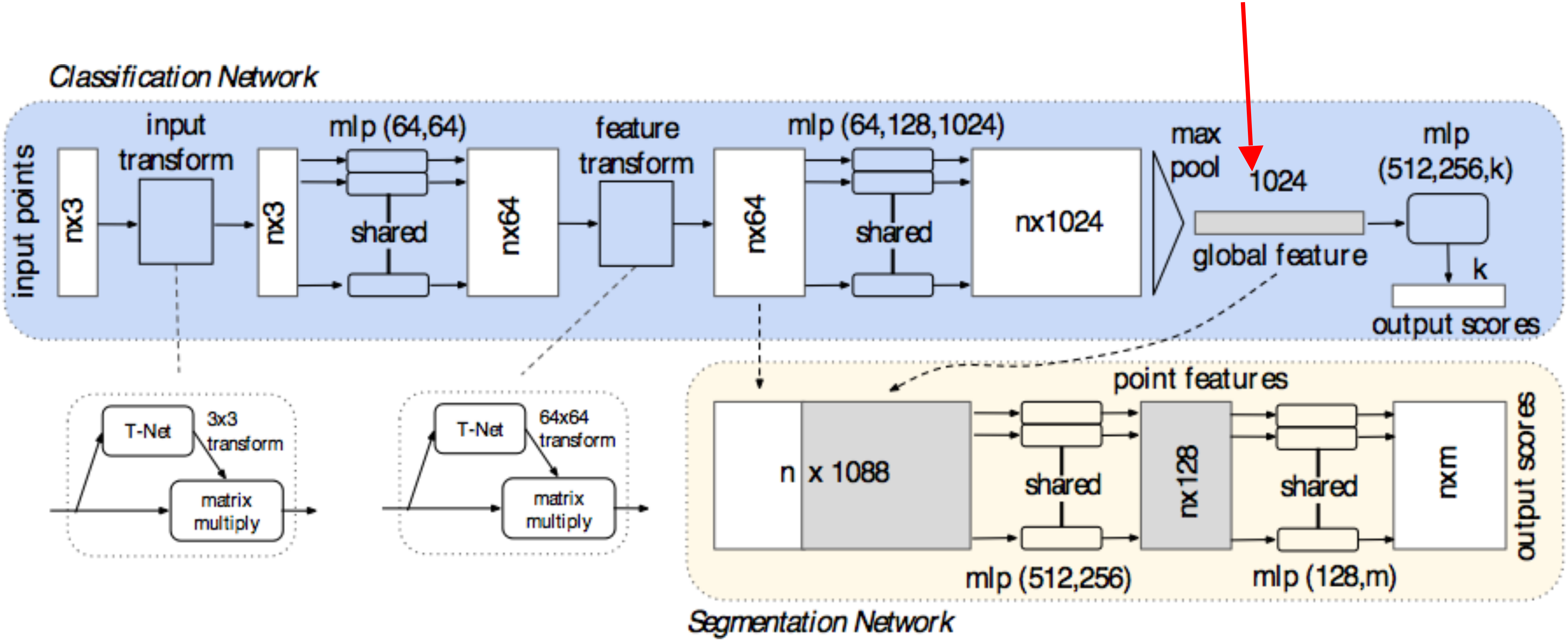
Find the top-K points in a dense volumetric grid that activates neuron j.

Visualizing Point Functions

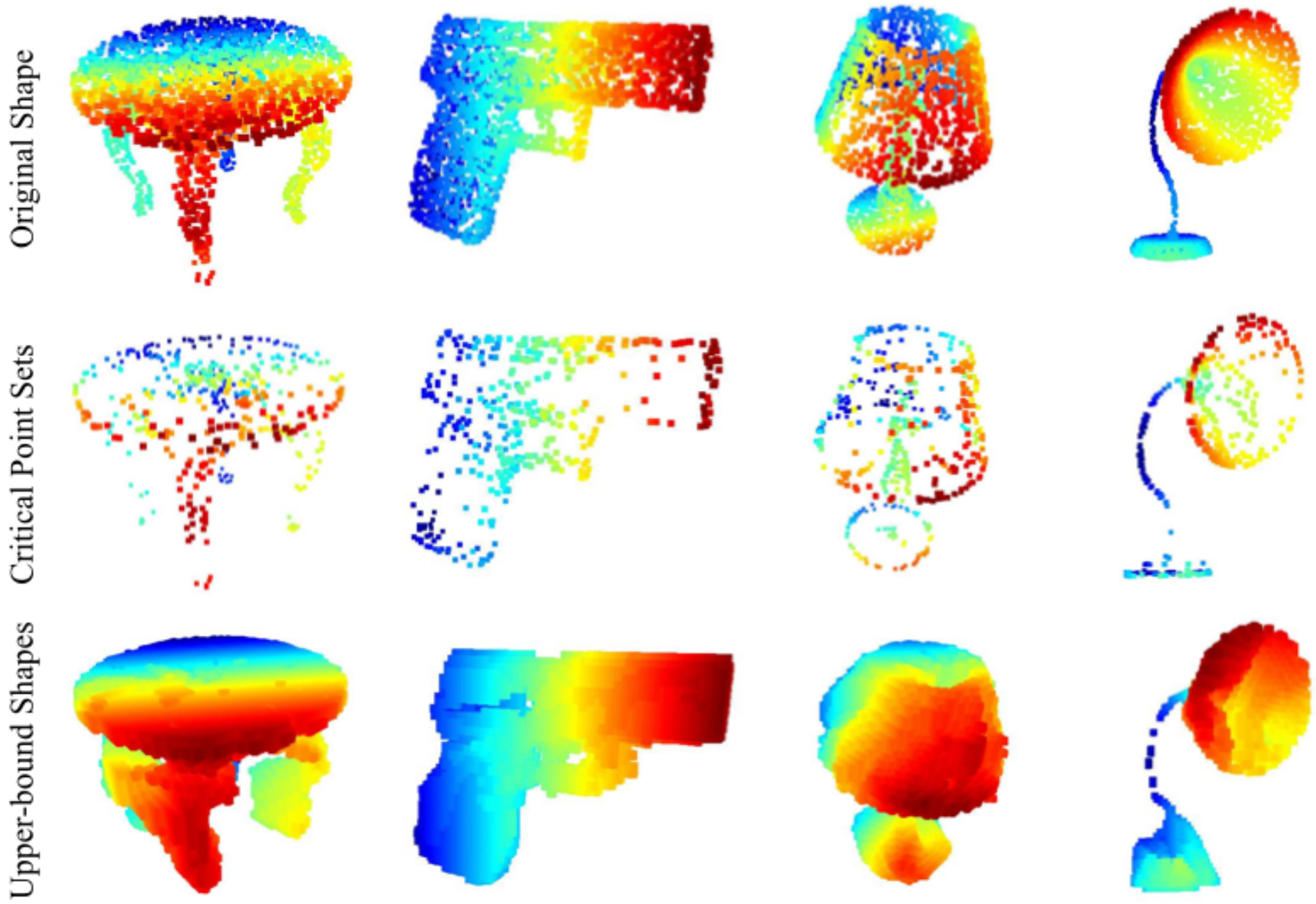


Visualizing Global Point Cloud Features

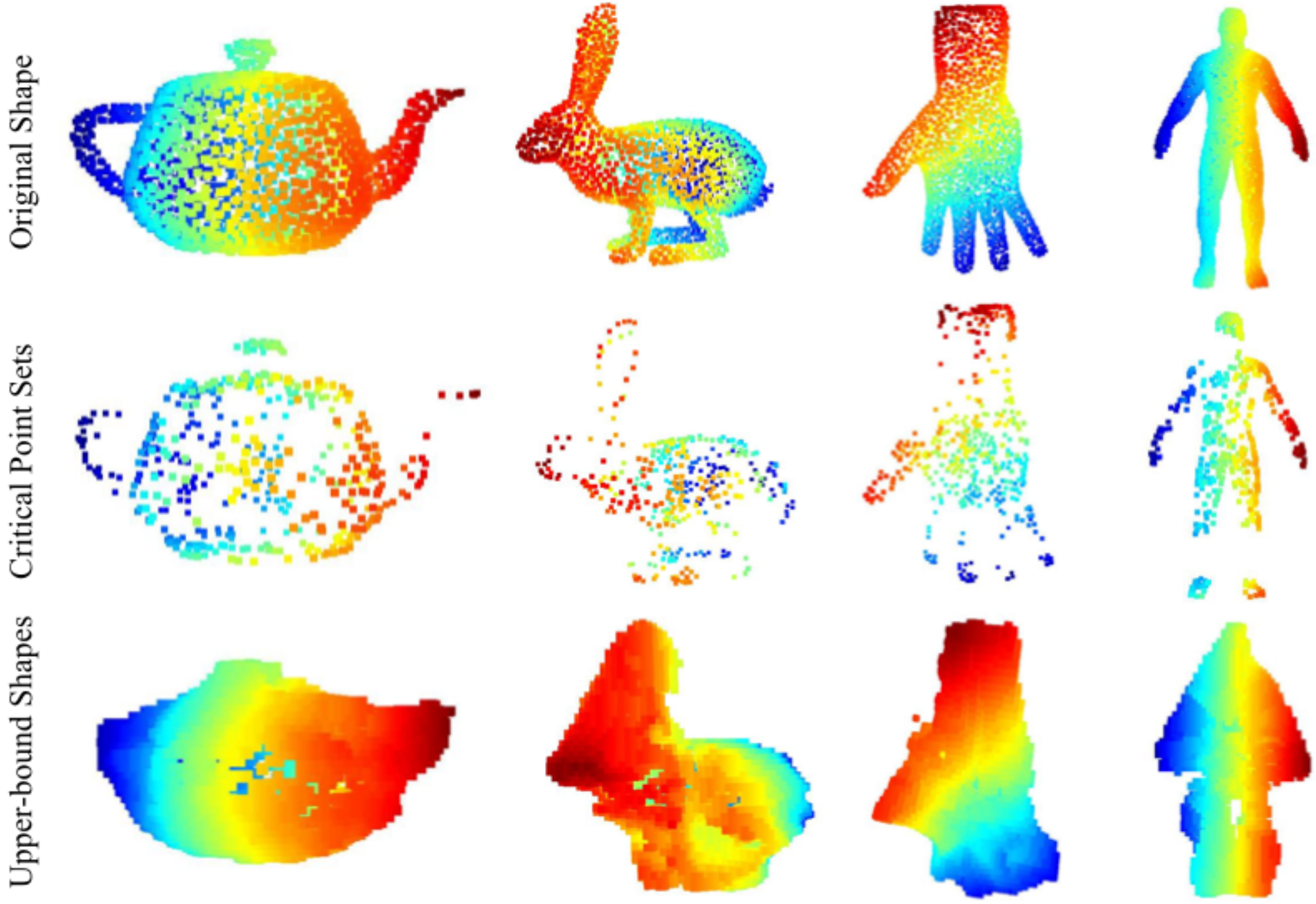
What's captured and left out here?



Visualizing Global Point Cloud Features

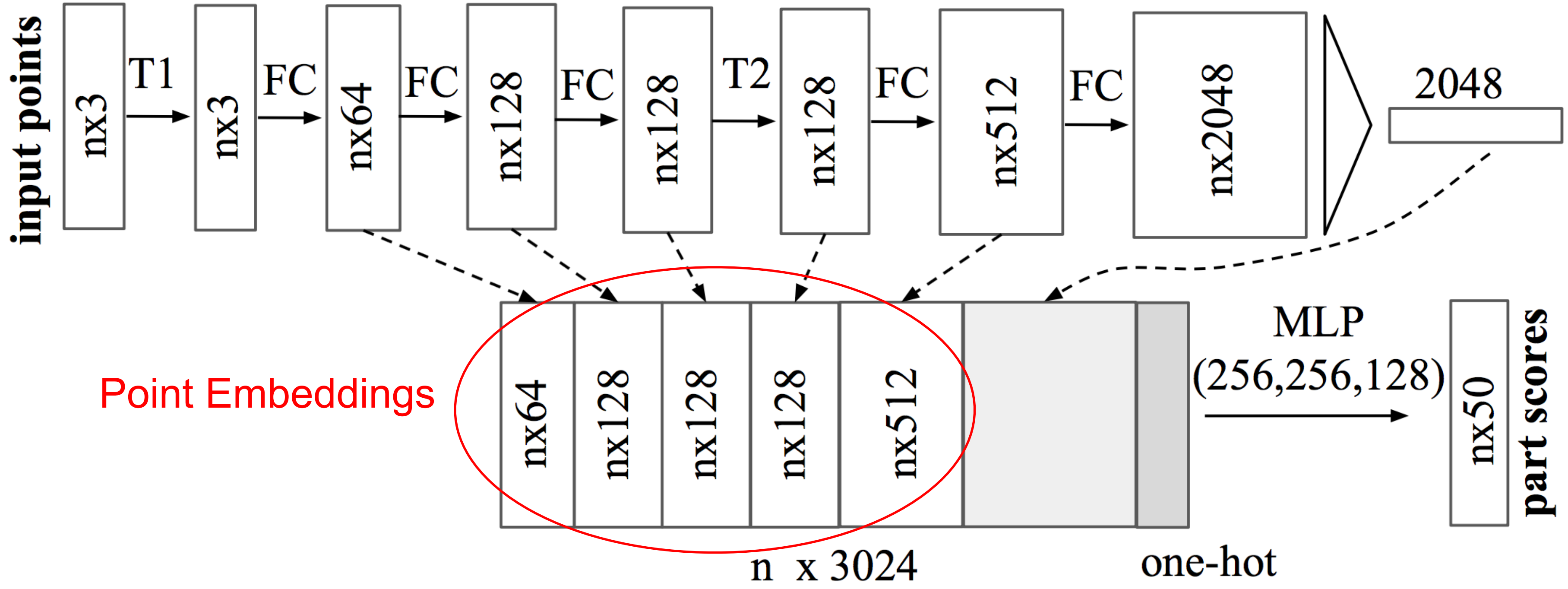


Visualizing Global Point Cloud Features (OOS)



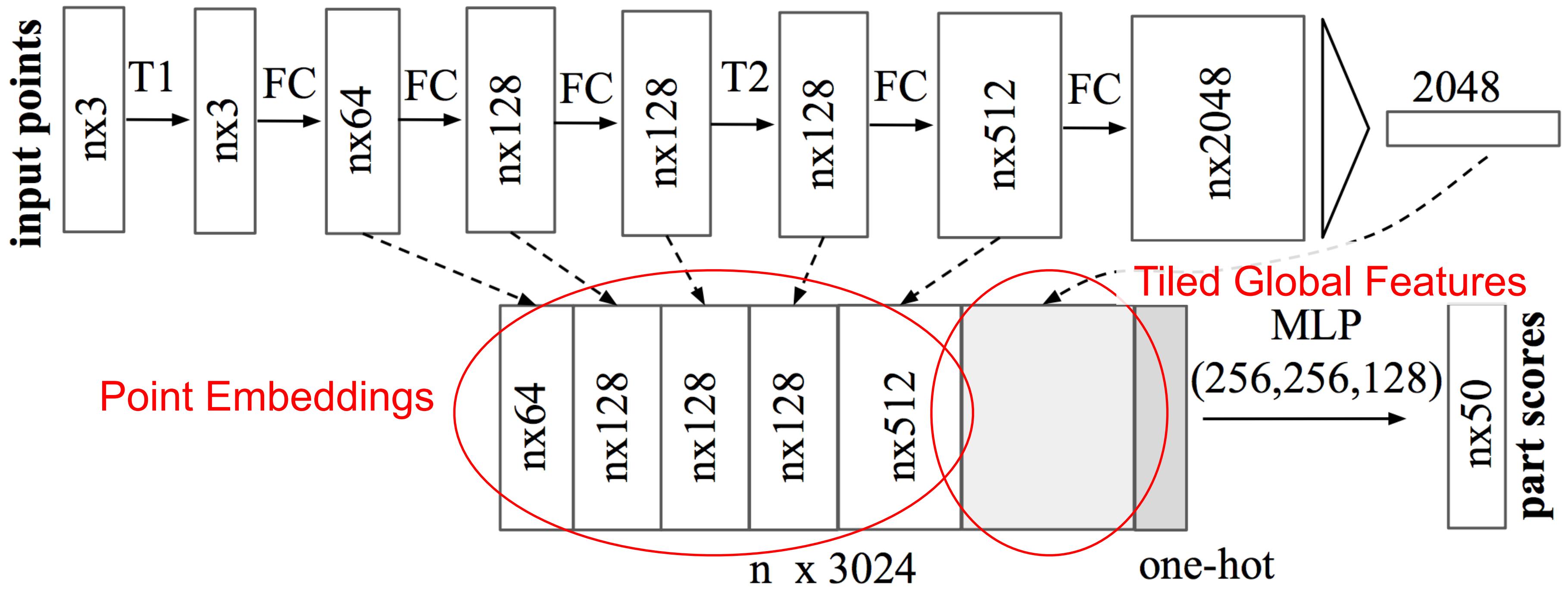
Previous Work: PointNet v1.0

Segmentation Network



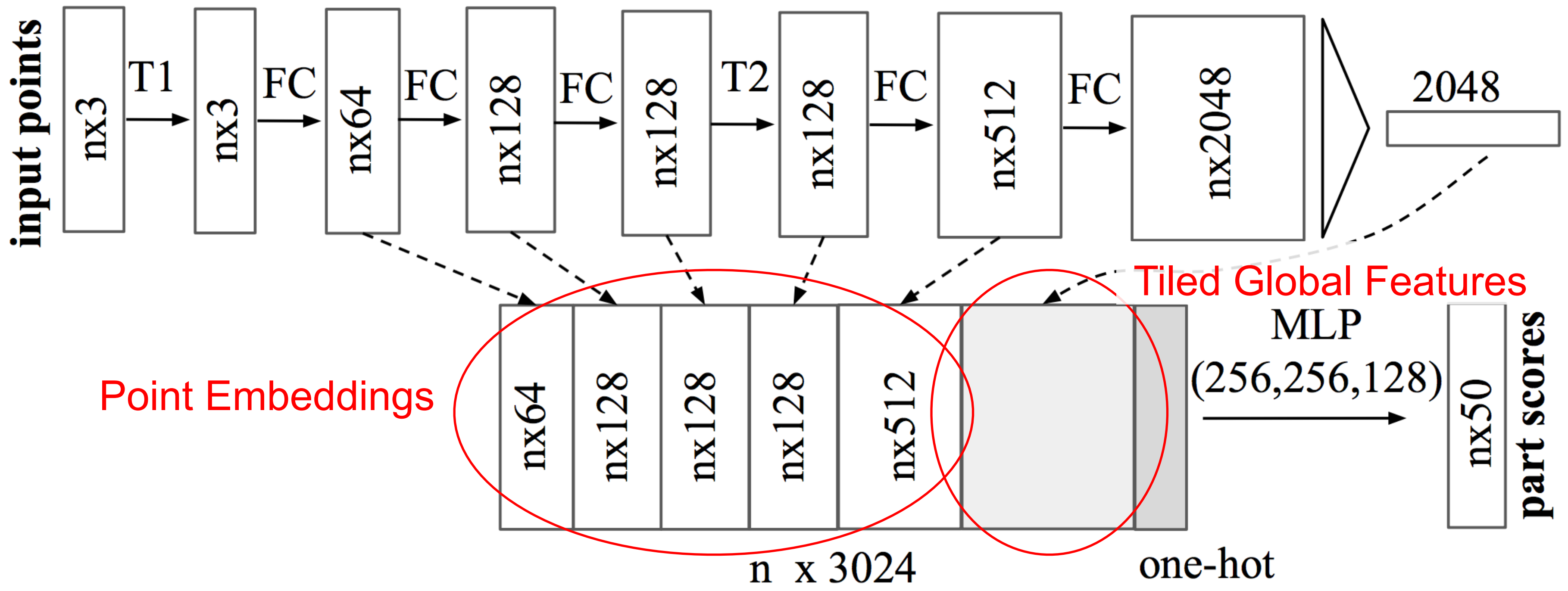
Previous Work: PointNet v1.0

Segmentation Network



Previous Work: PointNet v1.0

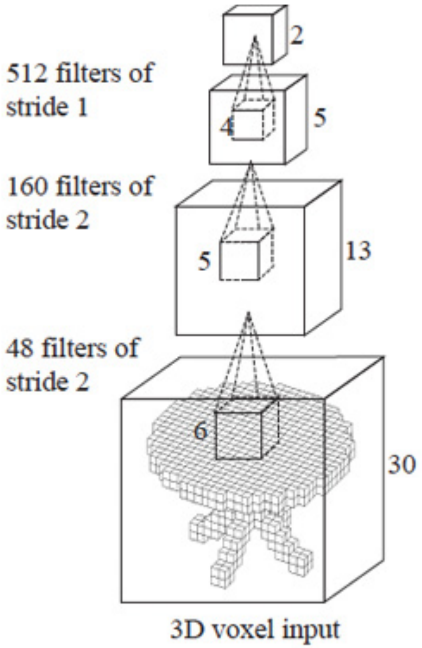
Segmentation Network



- No local context for each point!

Limitations of PointNet v1.0

- Hierarchical Feature Learning
- Increasing receptive field

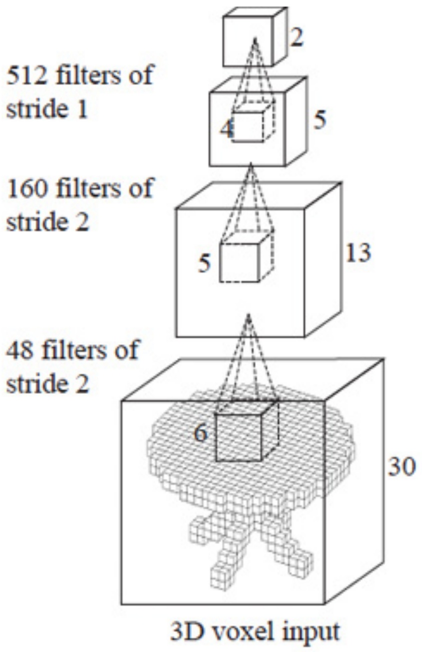


3D CNN (Wu et al.)

Limitations of PointNet v1.0

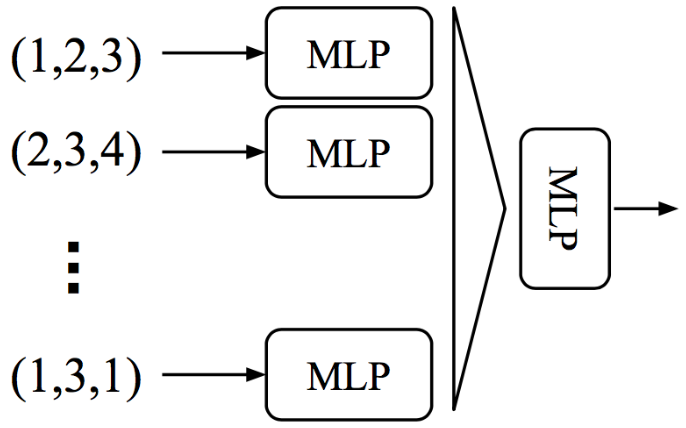
- Hierarchical Feature Learning
- Increasing receptive field

Global Feature Learning
Receptive field:
one point OR all points



3D CNN (Wu et al.)

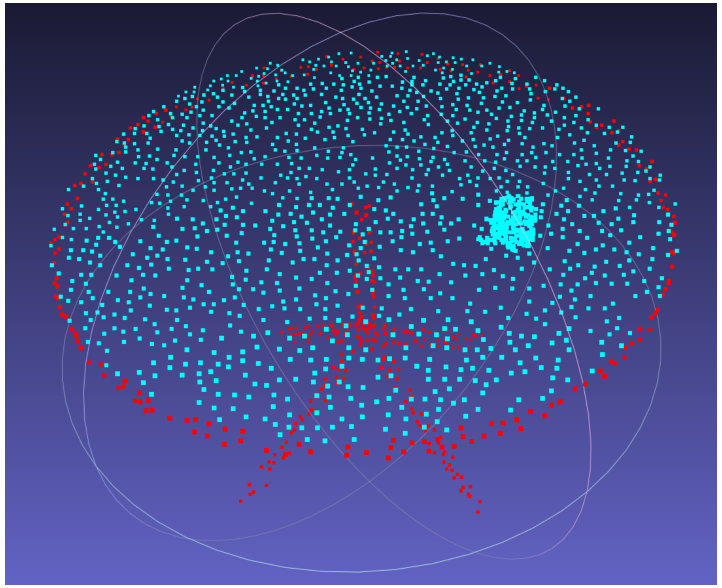
v.s.



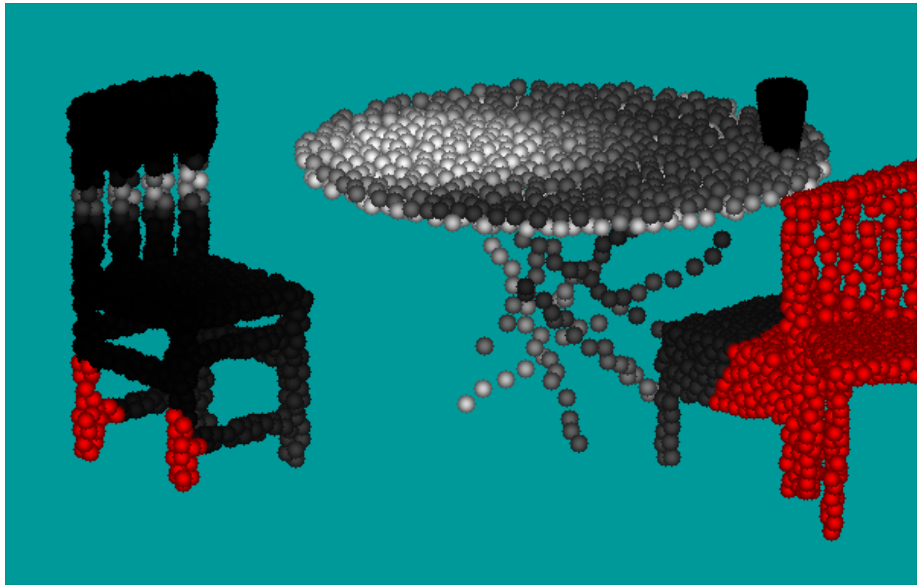
PointNet (vanilla) (Qi et al.)

Limitations of PointNet v1.0

Artifacts in segmentation tasks:



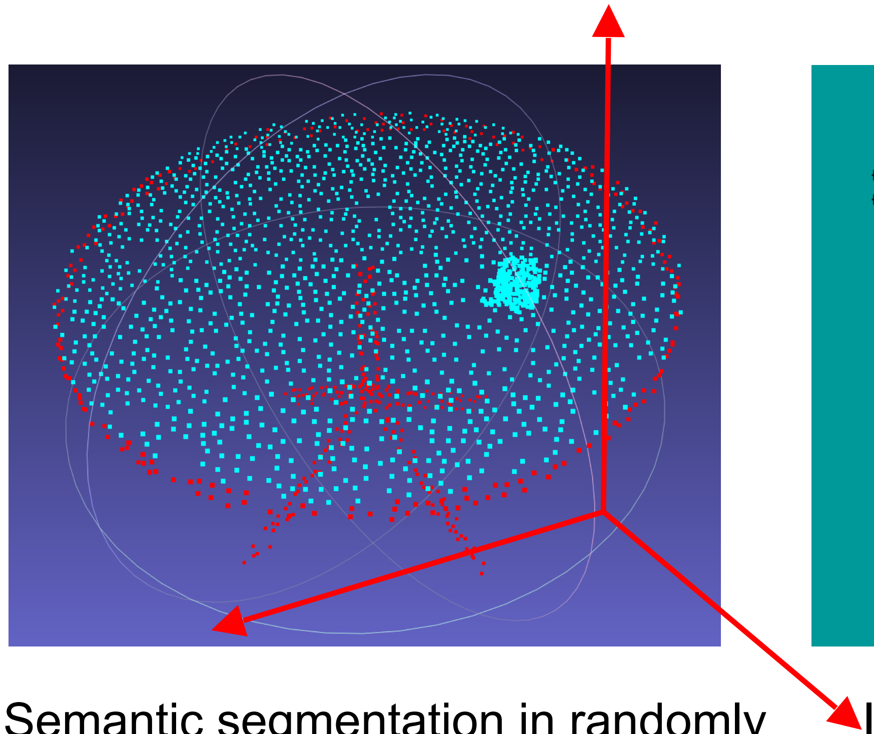
Semantic segmentation in randomly translated table-cup scene.



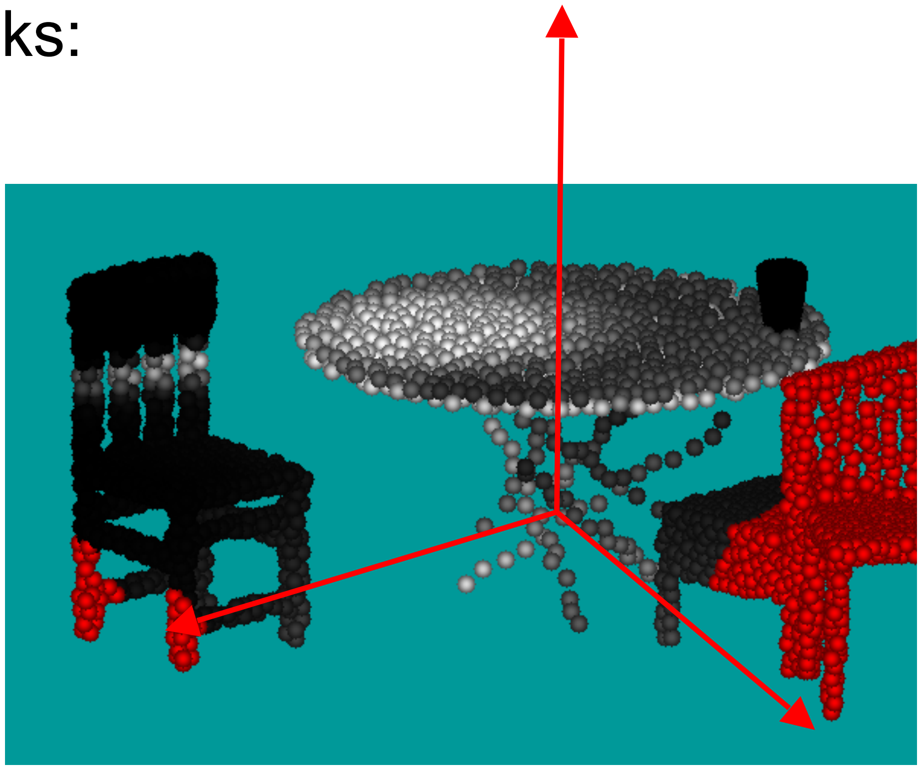
Instance segmentation in table-chair-cup scene

Limitations of PointNet v1.0

Artifacts in segmentation tasks:



Semantic segmentation in randomly translated table-cup scene.



Instance segmentation in table-chair-cup scene

- Global feature depends on absolute XYZ!
- Hard to generalize to unseen point configurations

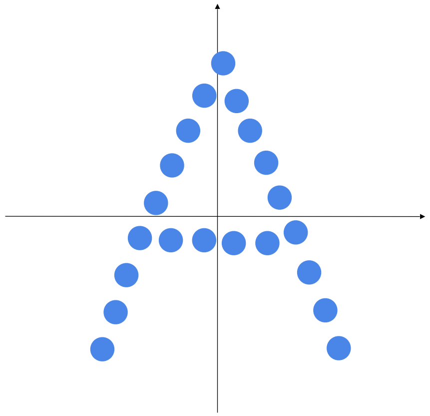
Question

- How to learn local context feature for points?

- Use PointNet in local regions, aggregate local region features by PointNet again..
-
- Hierarchical feature learning!

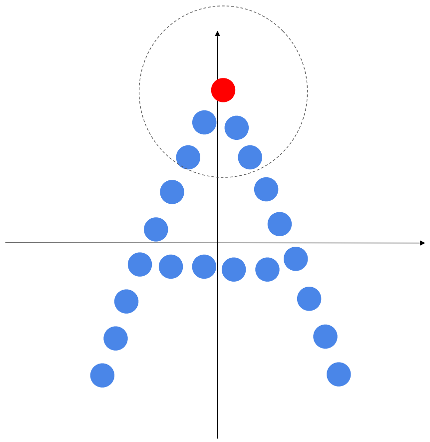
Multi-Scale PointNet for Hierarchical Feature Learning

PointNet v2.0: Multi-Scale PointNet



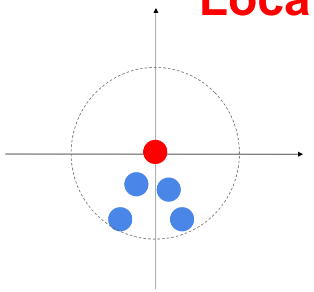
N points in (x,y)

PointNet v2.0: Multi-Scale PointNet



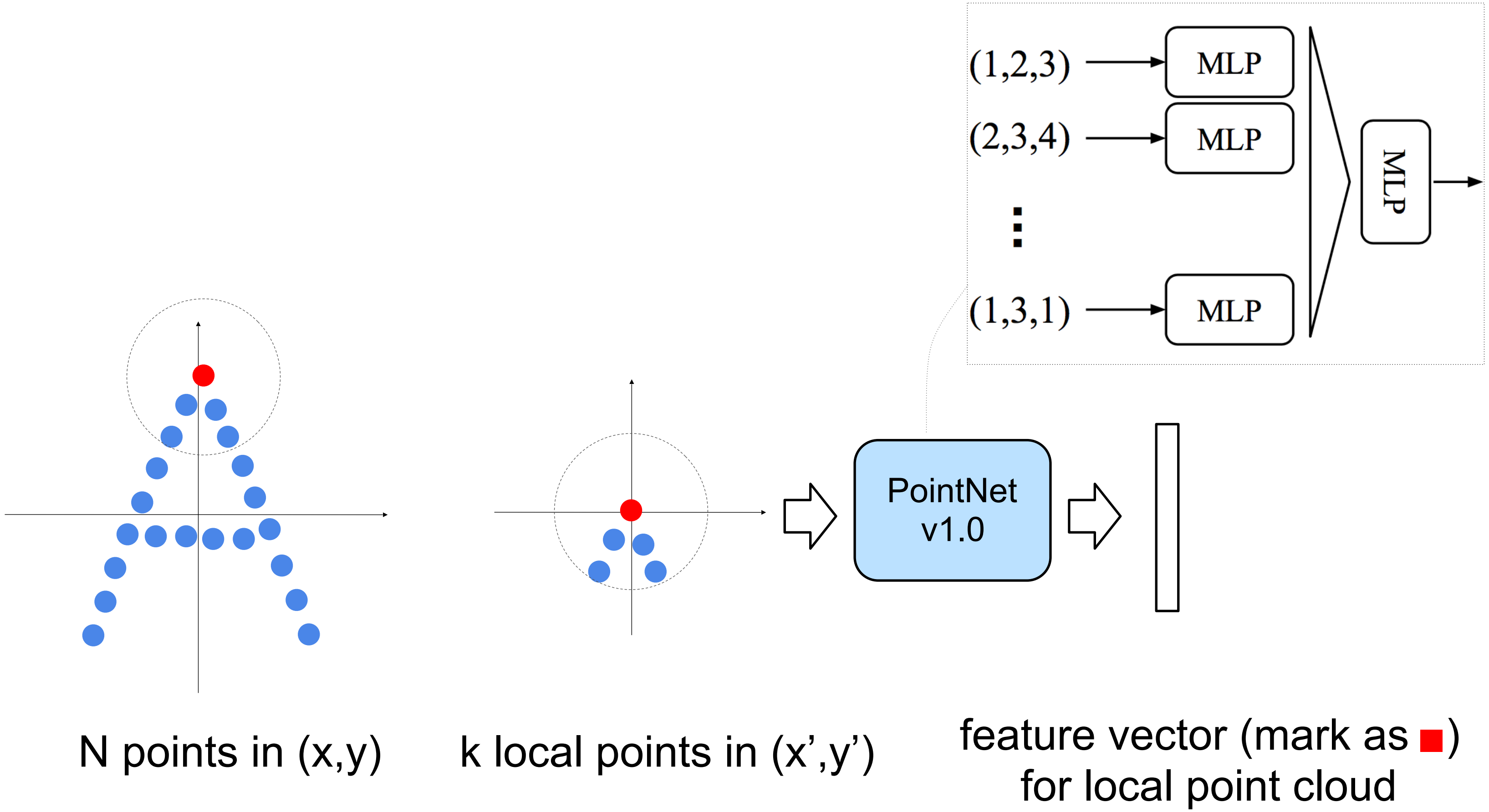
N points in (x,y)

Local Coordinates!



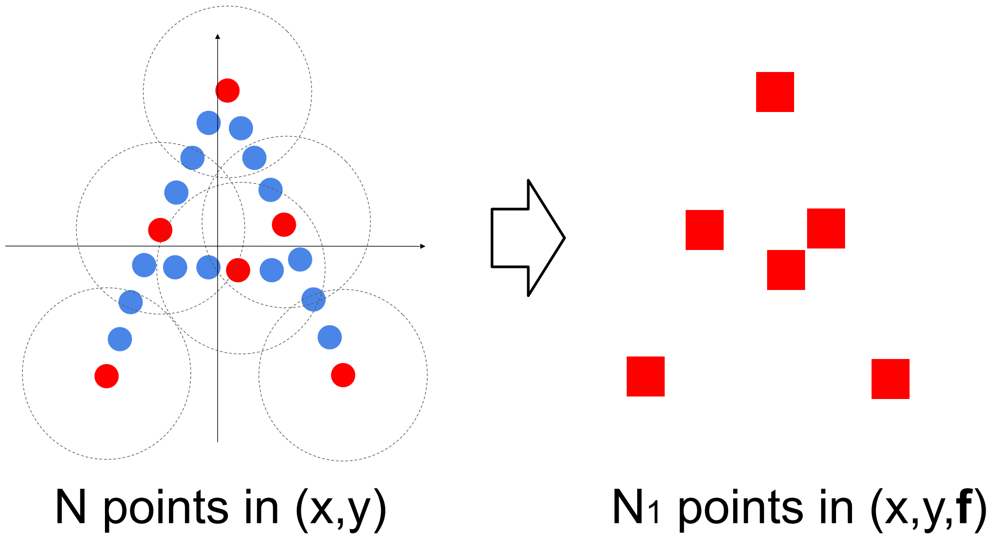
k local points in (x’,y’)

PointNet v2.0: Multi-Scale PointNet

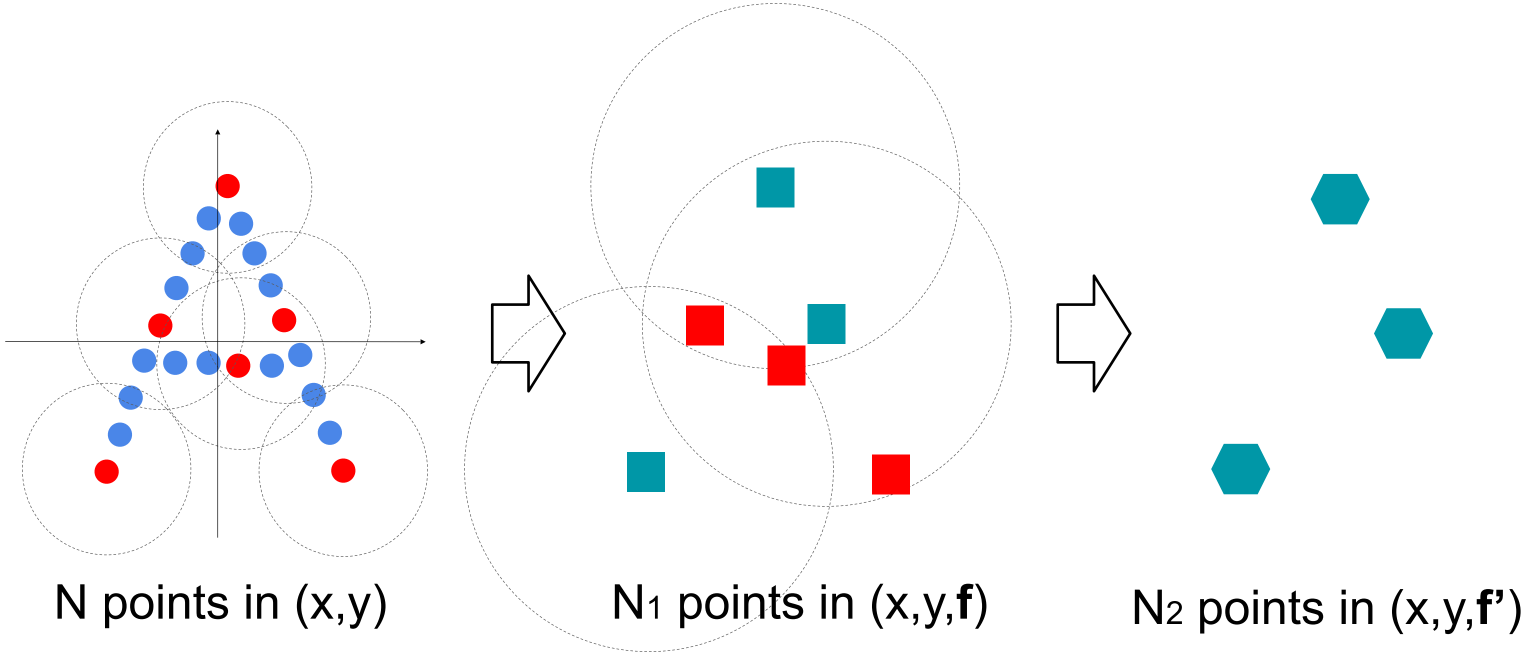


PointNet v2.0: Multi-Scale PointNet

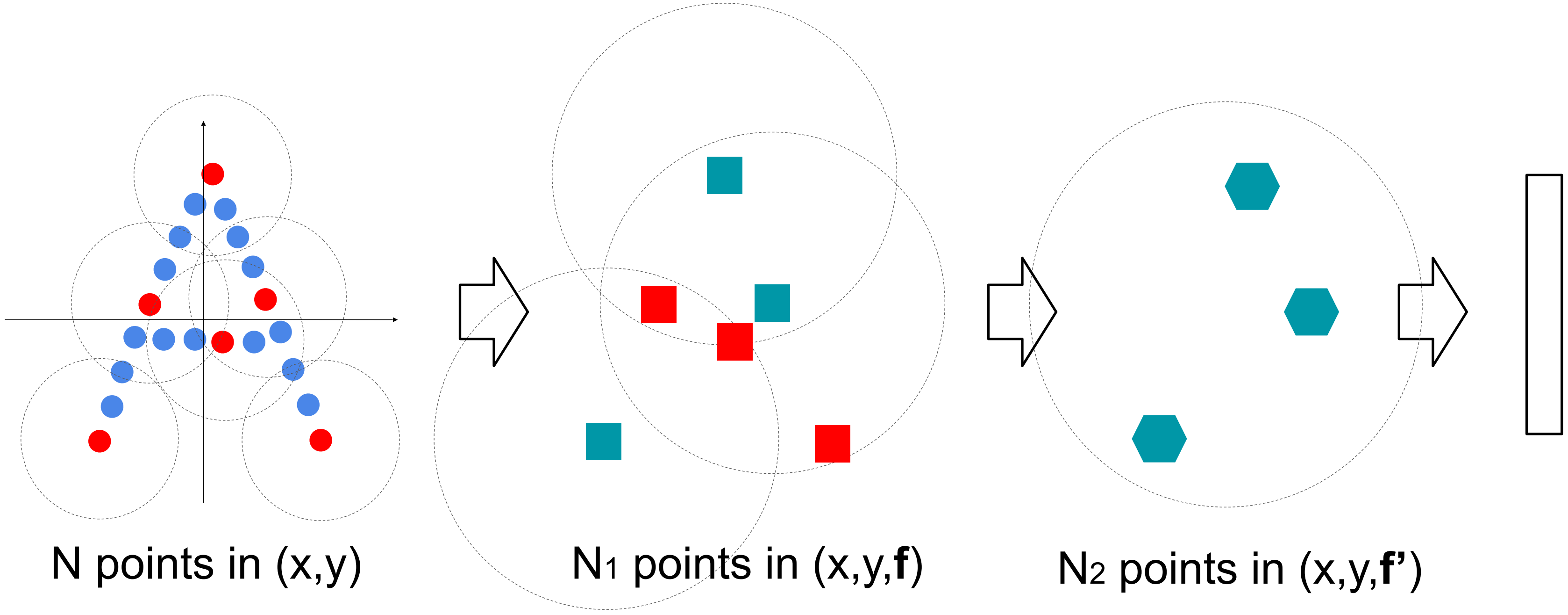
PointNet Module/Layer: Farthest Point Sampling + Grouping + PointNet v1.0



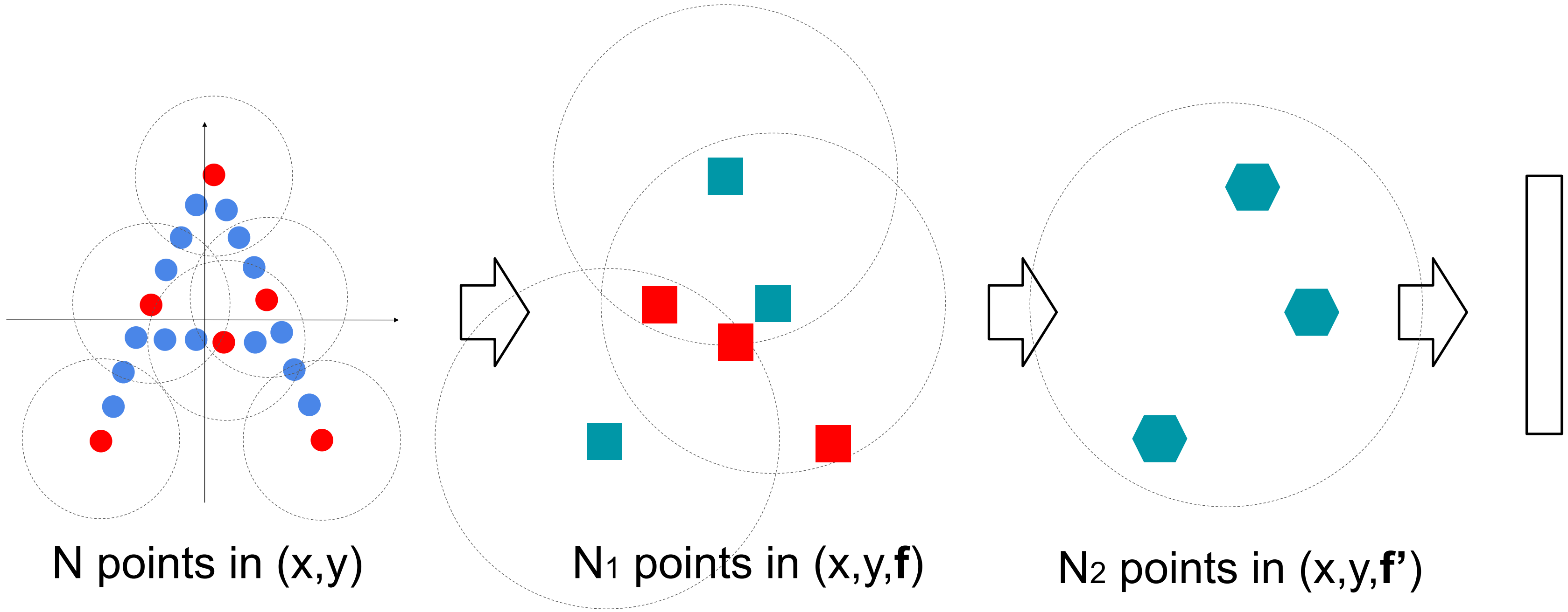
PointNet v2.0: Multi-Scale PointNet



PointNet v2.0: Multi-Scale PointNet



PointNet v2.0: Multi-Scale PointNet

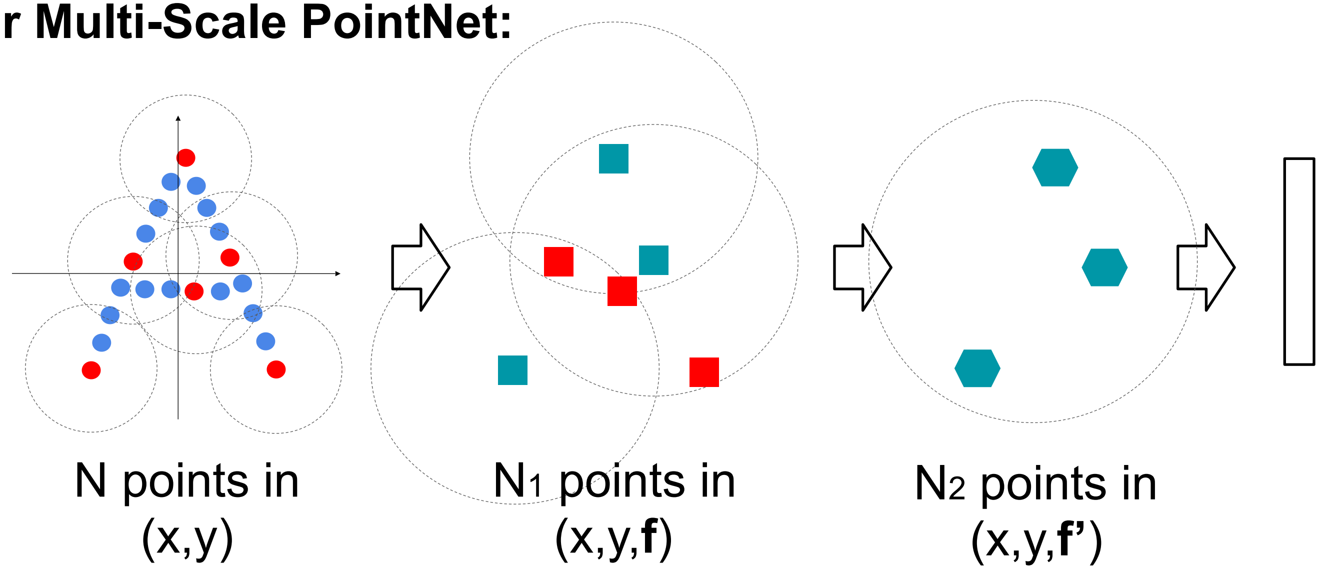


1. Larger receptive field in higher layers ✓
2. Less points in higher layers (more scalable) ✓
3. Weight sharing ✓
4. Translation invariance (local coordinates in local regions) ✓

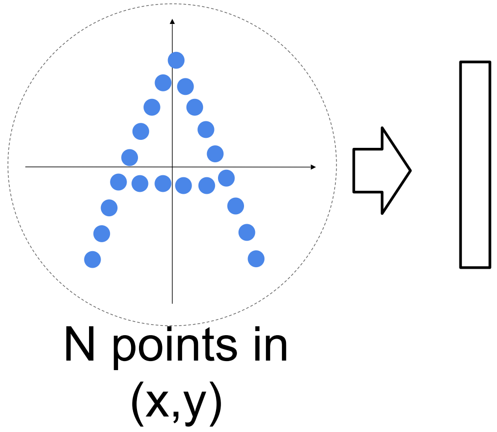
Discussions on Multi-Scale PointNet

Multi-Scale PointNet v.s. PointNet v1.0

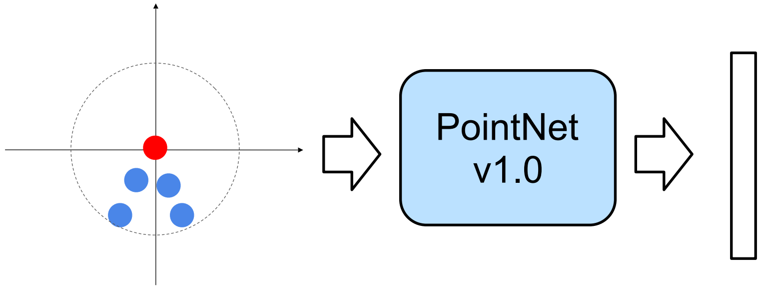
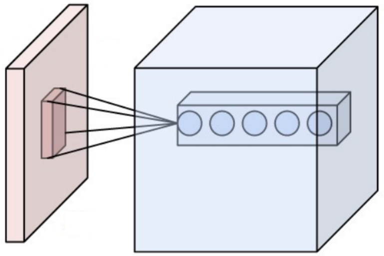
Three-layer Multi-Scale PointNet:



One-layer Multi-Scale PointNet <=> PointNet v1.0

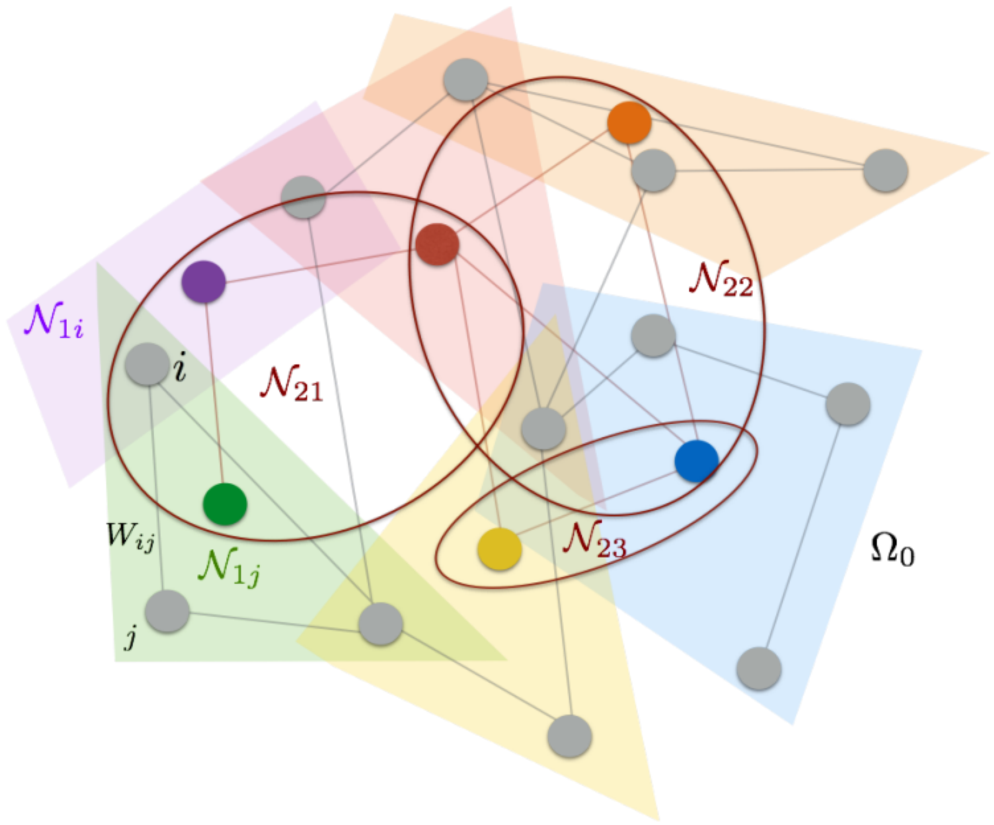


PointNet Layer v.s. Convolution Layer

	PointNet Layer	Convolution Layer
		
Input:	Point set	Dense array
Operation:	MLP + max pooling	Multiply and add
Neighbor-hood:	Distance query	Array index

Multi-Scale PointNet v.s. Graph CNN

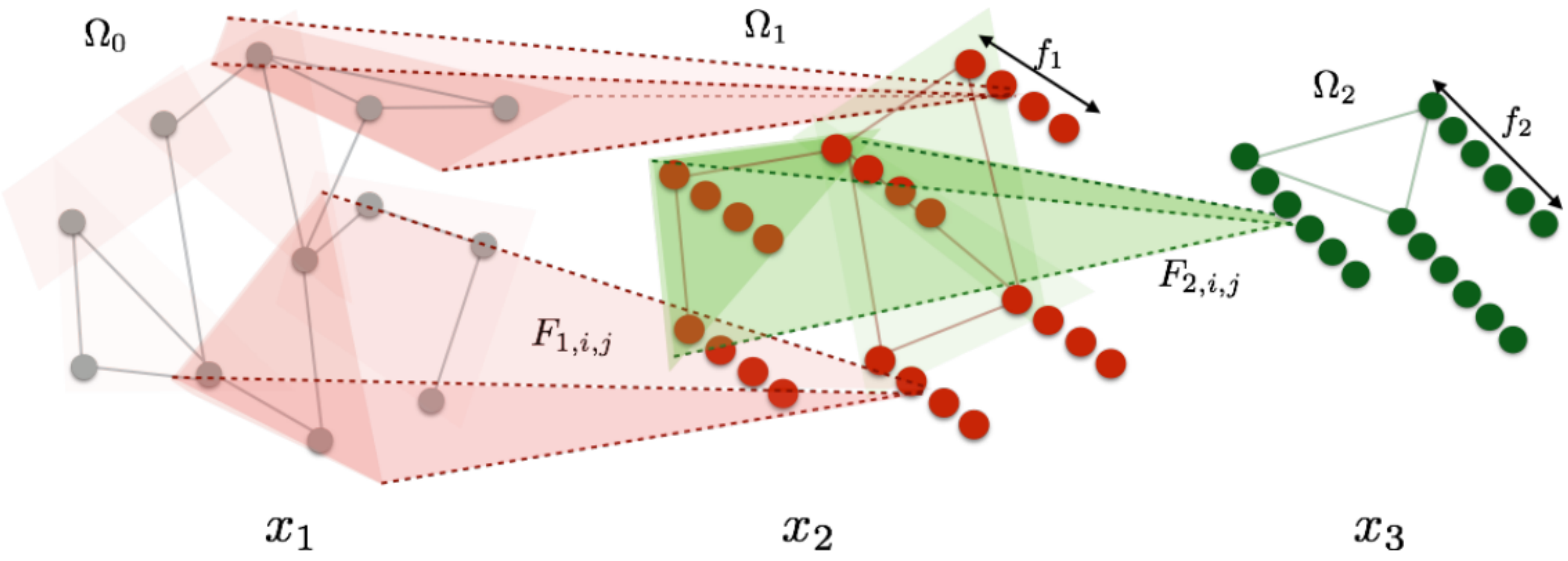
- Unexpectedly strong relation with Graph CNN:



Joan Bruna et al. Spectral Networks and Deep Locally Connected Networks on Graphs. ICLR 2014

Multi-Scale PointNet v.s. Graph CNN

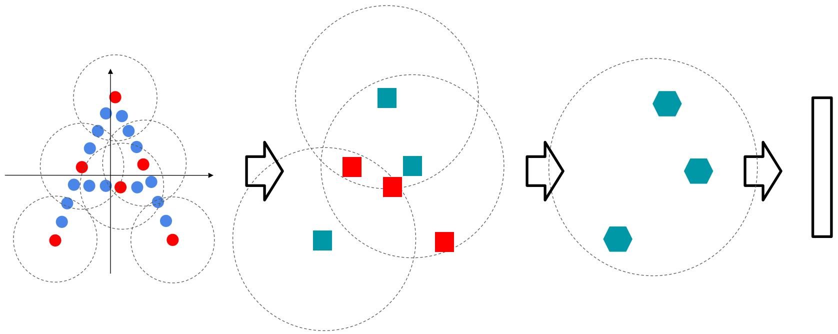
- Local feature extraction, graph coarsening, repeat..



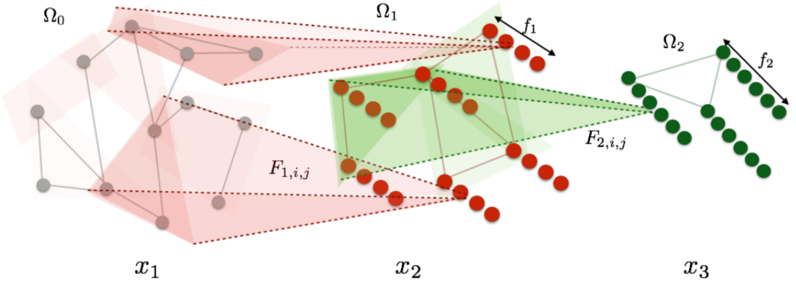
Joan Bruna et al. Spectral Networks and Deep Locally Connected Networks on Graphs. ICLR 2014

Multi-Scale PointNet v.s. Graph CNN

- In Graph CNN's perspective:
- Multi-Scale PointNet defines
    1. Graph connectivity through Euclidean distance
    2. Graph coarsening by farthest point sampling
    3. Local feature extraction with PointNet (v1.0)



Multi-scale PointNet

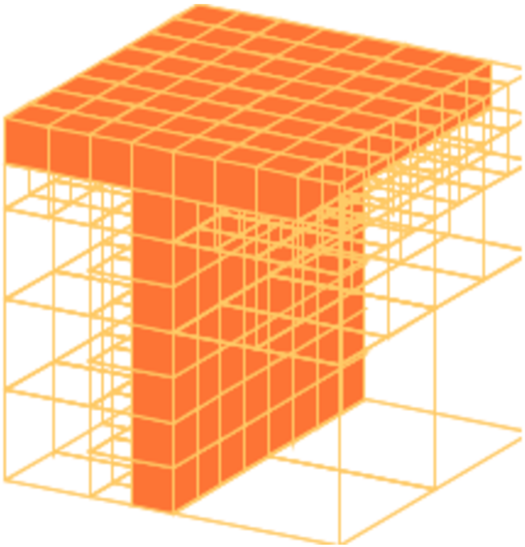


Graph CNN

Relations to OctNet (Octree based 3D CNN)

OctNet in Graph CNN's perspective:

1. Both connectivity and graph coarsening are defined by the Octree.
2. Local feature extraction by convolution layer.



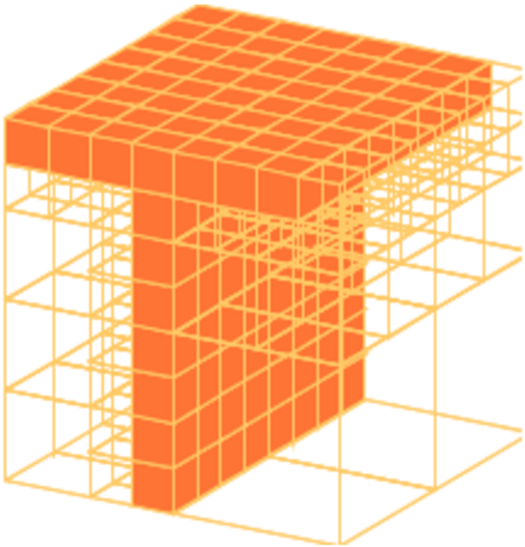
OctNet: Learning Deep 3D Representations at High Resolutions
Gernot Riegler, Ali Osman Ulusoy and Andreas Geiger

Relations to OctNet (Octree based 3D CNN)

OctNet in Graph CNN's perspective:

In Multi-Scale PointNet

1. Both connectivity and graph coarsening are defined by the Octree. By ground distance
2. Local feature extraction by convolution layer. By PointNet (v1.0)

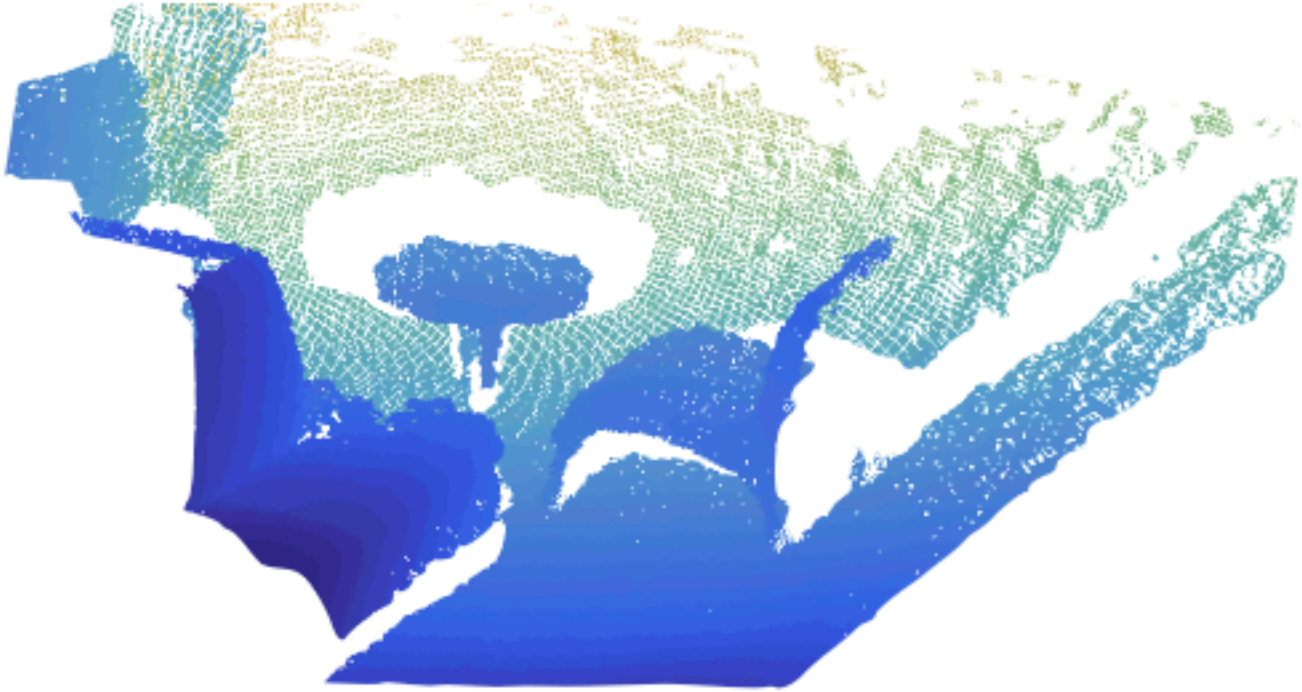


OctNet: Learning Deep 3D Representations at High Resolutions
Gernot Riegler, Ali Osman Ulusoy and Andreas Geiger

PointNet++: Addressing density variation of point cloud

Density variation is a common issue of 3D point cloud

- perspective effects, radial density variation, motion, etc



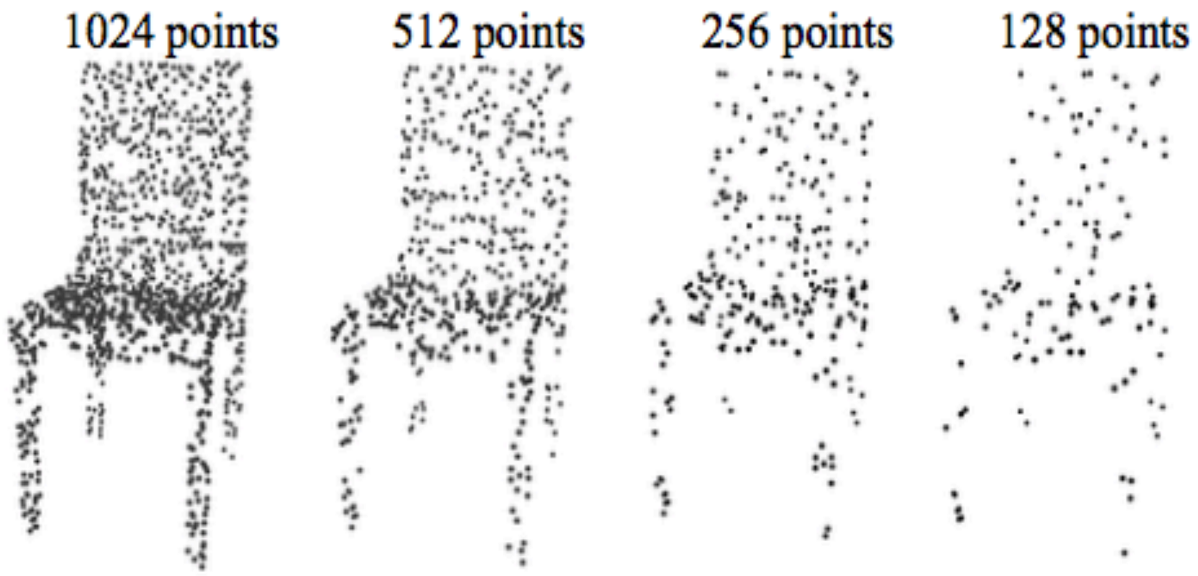
Density variation affects hierarchy

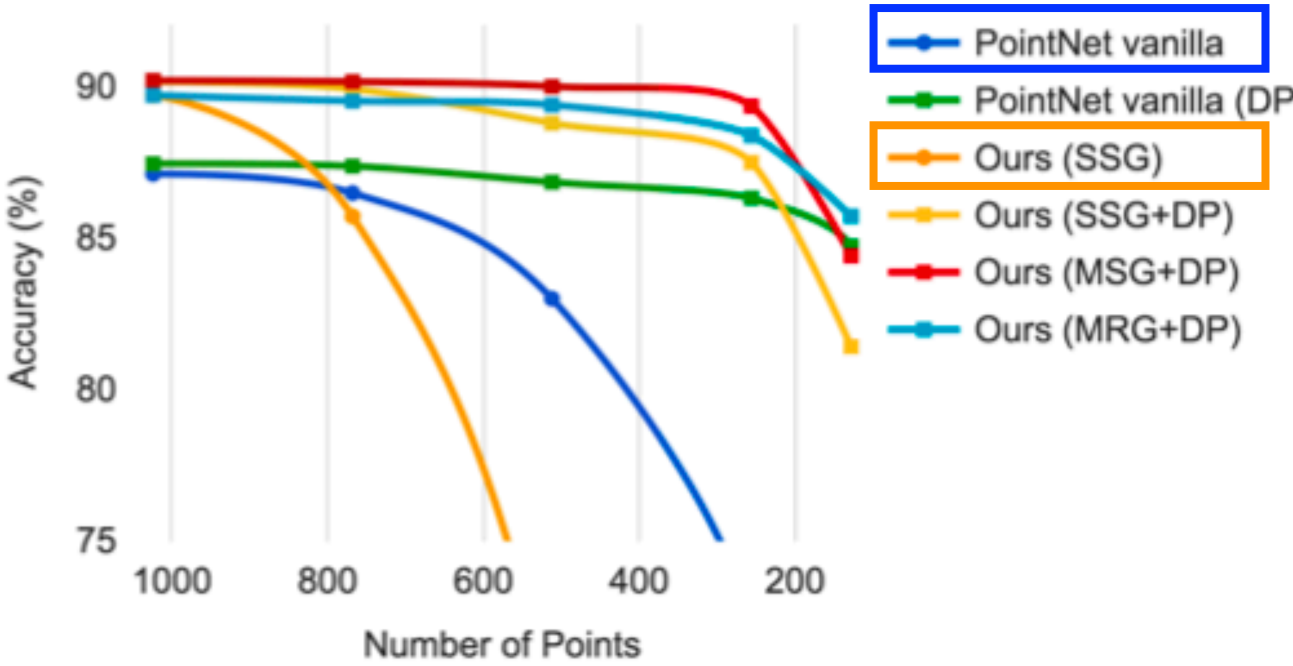
- In CNN, small kernels are “always” better

Karen Simonyan & Andrew Zisserman, VERY DEEP CONVOLUTIONAL NETWORKS FOR LARGE-SCALE IMAGE RECOGNITION, ICLR2015

- Is it also true for point cloud learning?

Density variation affects hierarchy



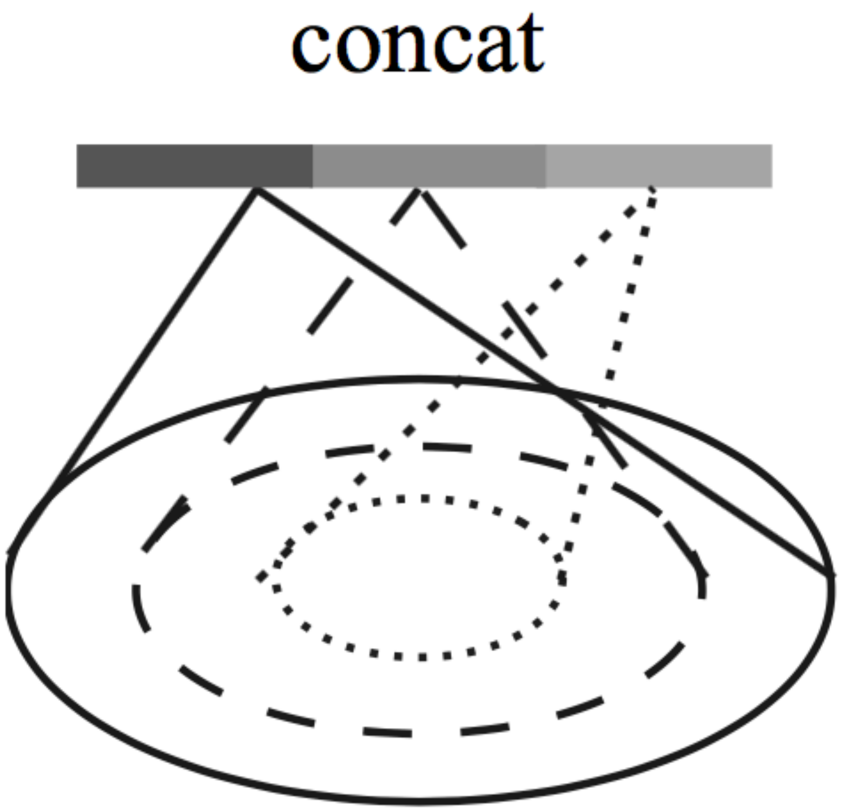


Intuition

- At high density area, we should look more closely
- At low density area, we should look more broadly
- However, parameters at different scales cannot be shared

Idea 1: Multi-scale grouping (MSG)

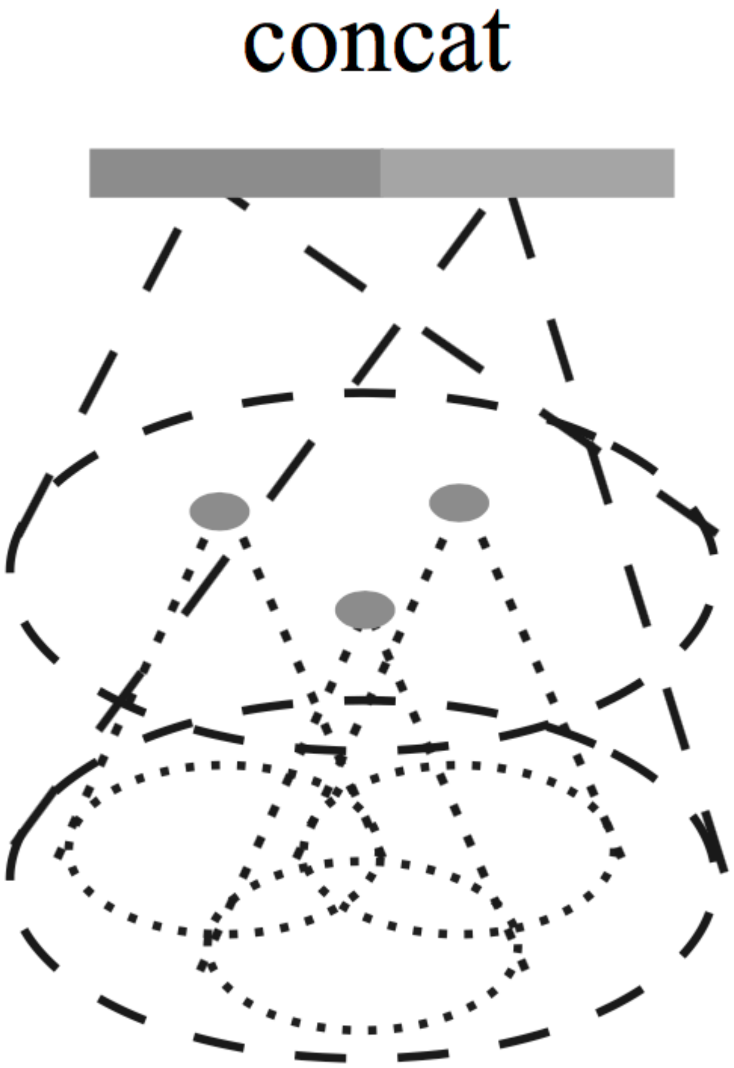
- Extract features at multiple scales and combine them
- Add random dropout to input point cloud to simulate scanning deficiency
- Dropout ratio is sampled uniformly in [0, 1]



Idea 2: Multi-resolution grouping (MRG)

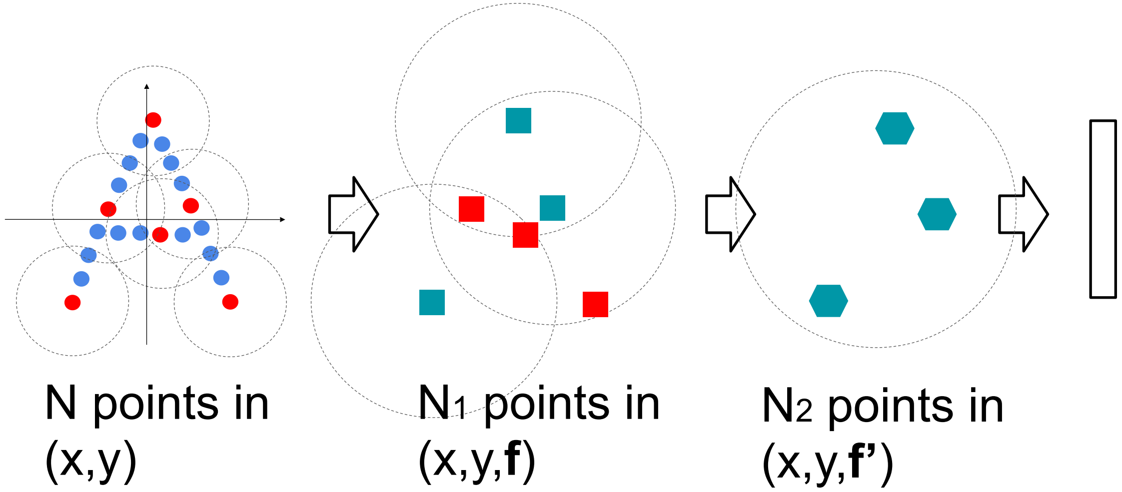
- Drawback of MSG: expensive
  - Need to run PointNet on many neighborhoods

- Multi-resolution grouping: reuse the computation from different levels

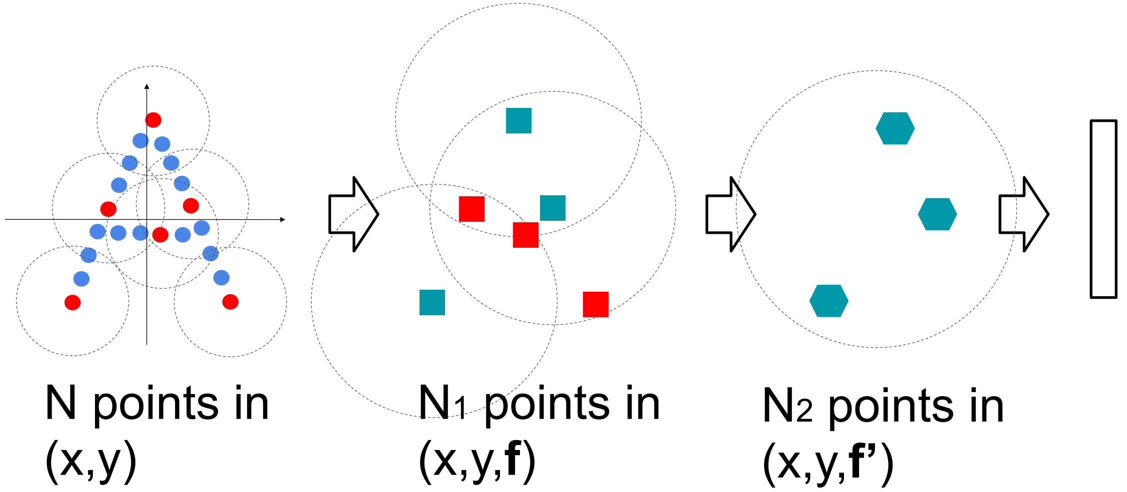


Multi-Scale PointNet for Segmentation with “Up-convolution” Module

“Up-convolution” in Multi-Scale PointNet

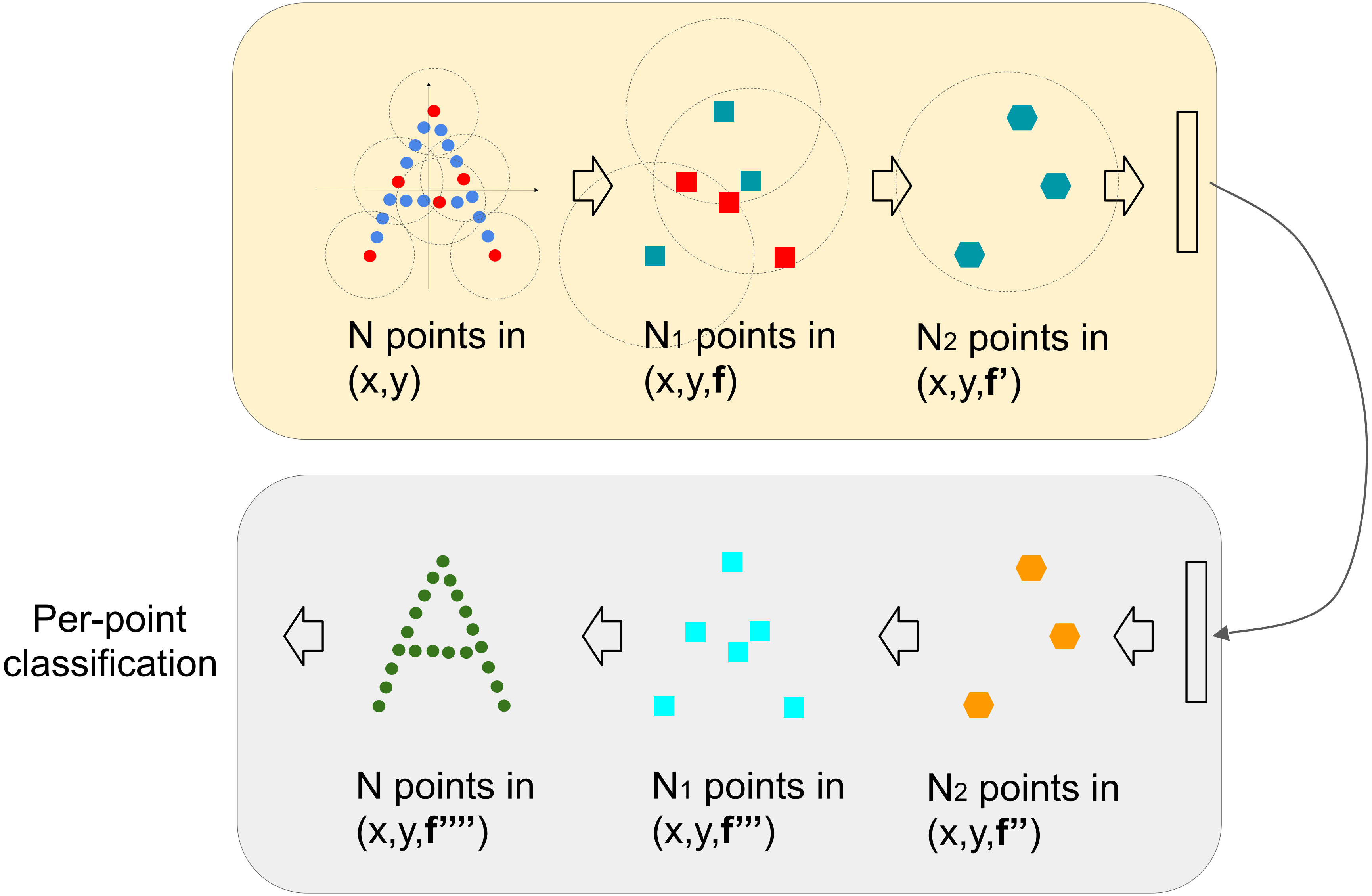


“Up-convolution” in Multi-Scale PointNet



How to achieve segmentation?

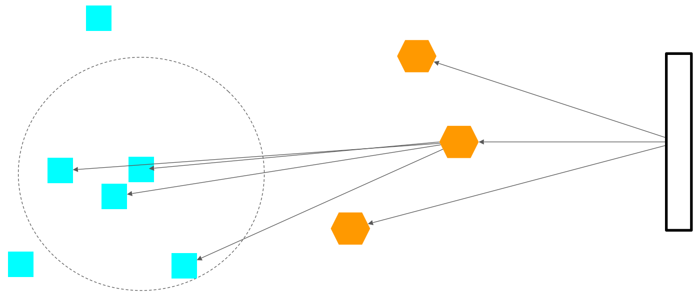
“Up-convolution” in Multi-Scale PointNet



“Up-convolution” Module

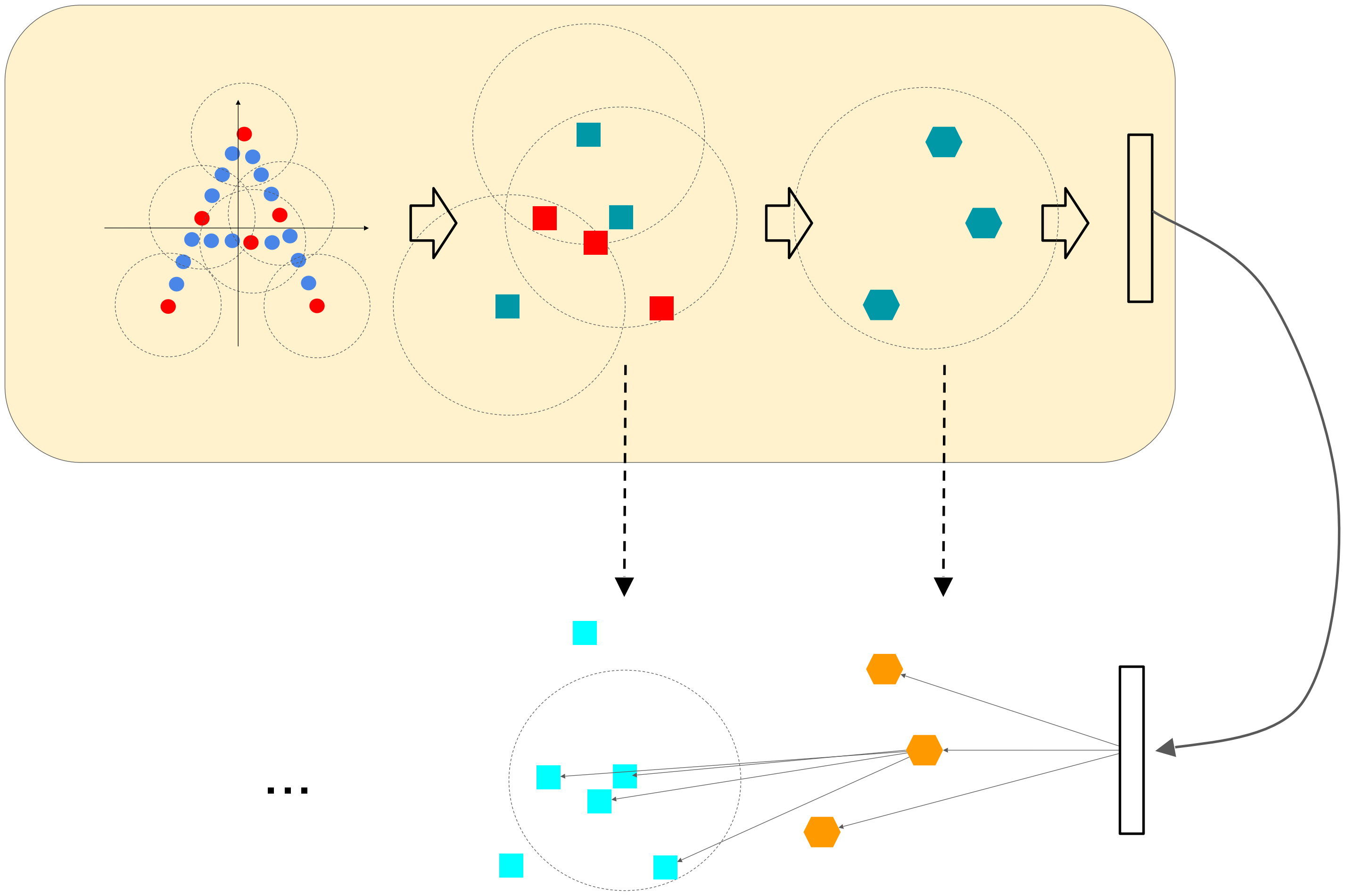
Naive solution: Broadcasting

...



“Up-convolution” Module

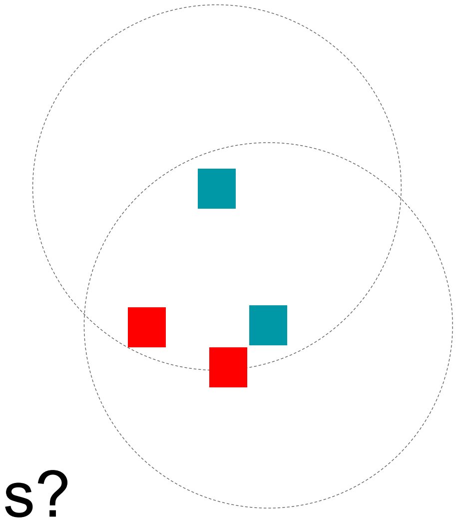
Naive solution Plus: Broadcasting + Skip links



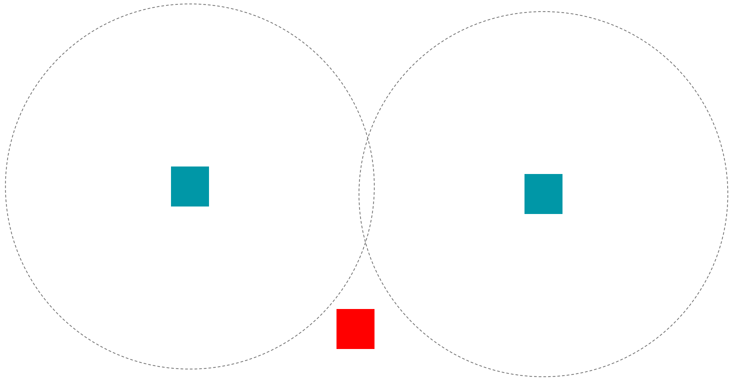
“Up-convolution” Module

Naive solution (broadcasting) Problems:

1. How to deal with points that belong to multiple regions?



2. What if some point belongs to no regions?



"Up-convolution" Module with 3D Interpolation

Instead of broadcasting, use 3D interpolation.

- Nearest Neighbor
- Inverse distance weighting
- Using delaunay triangulation

...

“Up-convolution” Module with 3D Interpolation

Instead of broadcasting, use 3D interpolation.

- Nearest Neighbor
- <u>**Inverse distance weighting**</u>
- Using delaunay triangulation

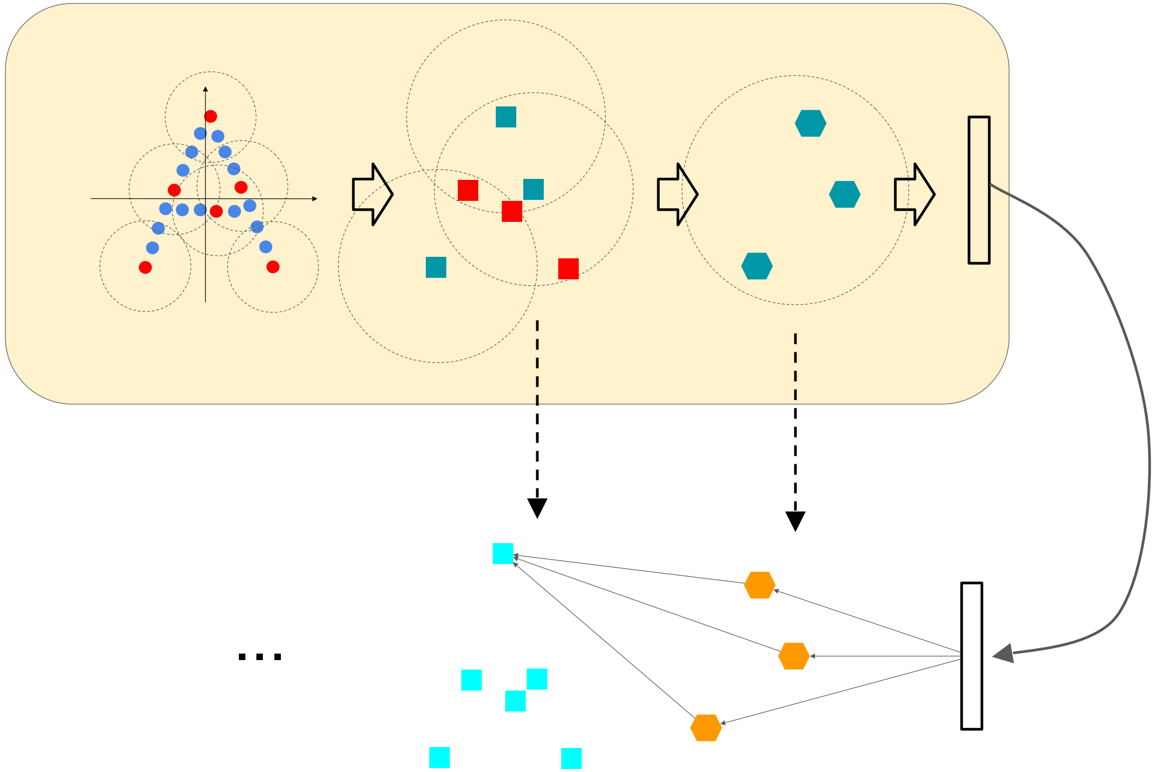
...

$$u(\mathbf{x}) = \begin{cases} \dfrac{\sum_{i=1}^{N} w_i(\mathbf{x}) u_i}{\sum_{i=1}^{N} w_i(\mathbf{x})}, & \text{if } d(\mathbf{x}, \mathbf{x}_i) \neq 0 \text{ for all } i \\ u_i, & \text{if } d(\mathbf{x}, \mathbf{x}_i) = 0 \text{ for some } i \end{cases}$$

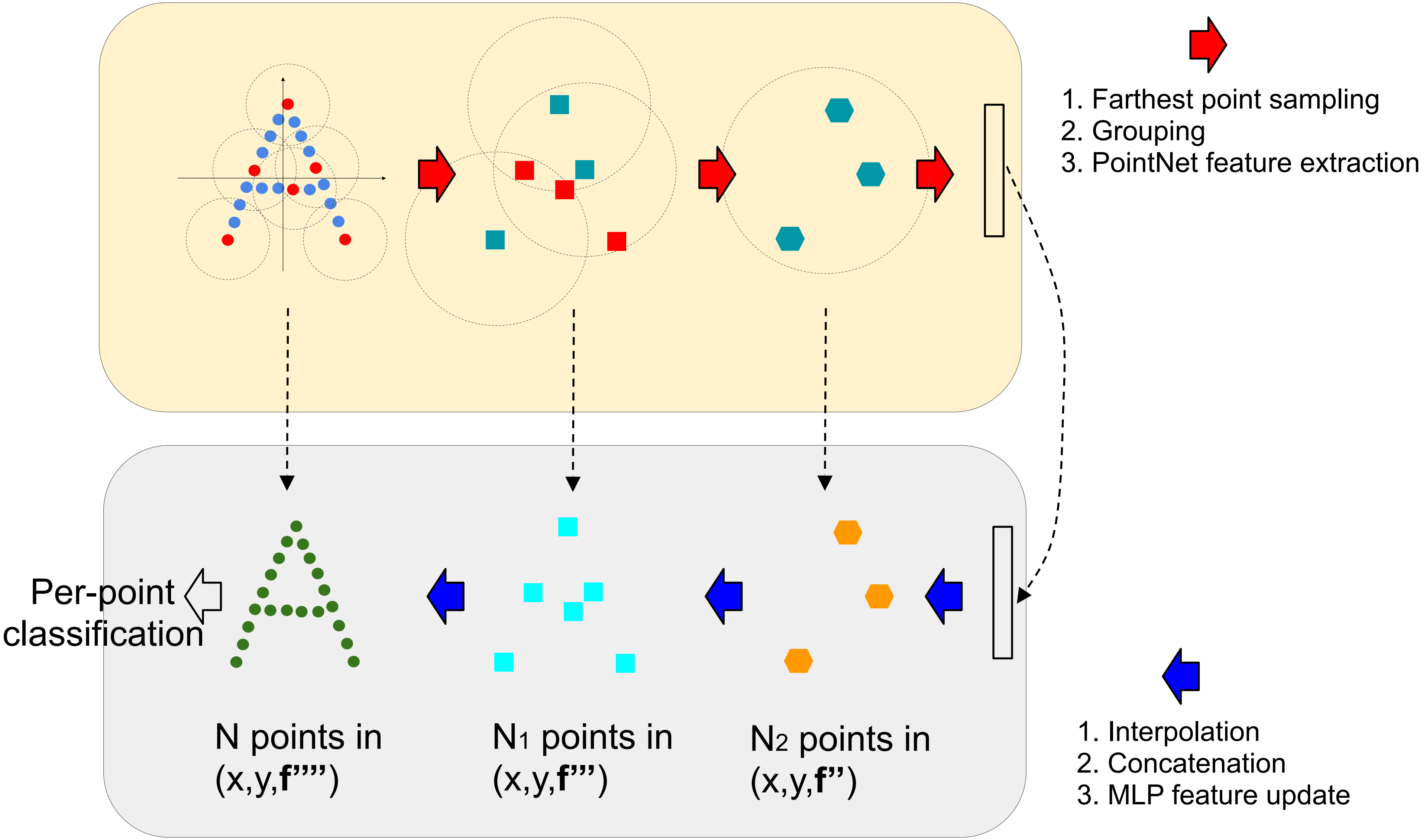
$$w_i(\mathbf{x}) = \frac{1}{d(\mathbf{x}, \mathbf{x}_i)^p}$$

“Up-convolution” Module

1. Feature Interpolation based on Euclidean distances to kNN
2. Skip link feature aggregation
3. MLP on aggregated feature for feature update and compression



Multi-Scale PointNet: Segmentation Network



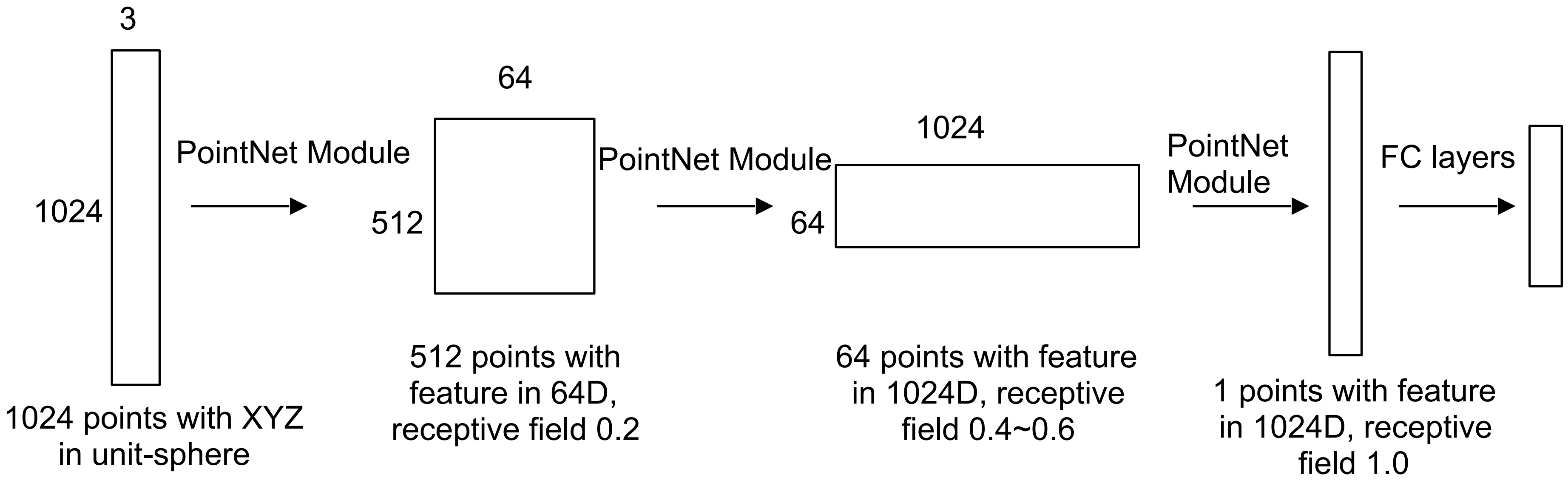
Experimental Results (preliminary)

ModelNet40 Classification Benchmark

	Accuracy
PointNet (vanilla)	87.2%
PointNet	89.2%
MultiScale PointNet	90.1%
MultiScale PointNet (voting)	**90.7%**

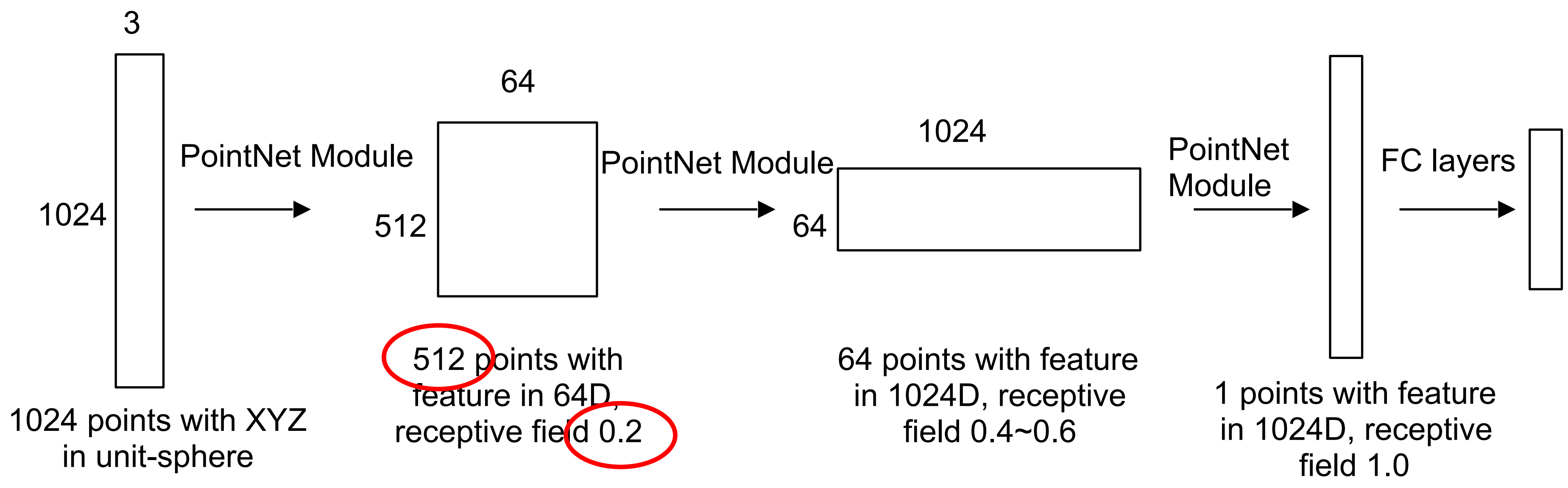
ModelNet40 Classification Benchmark

	Accuracy
PointNet (vanilla)	87.2%
PointNet	89.2%
MultiScale PointNet	90.1%
MultiScale PointNet (voting)	**90.7%**



ModelNet40 Classification Benchmark

	Accuracy
PointNet (vanilla)	87.2%
PointNet	89.2%
MultiScale PointNet	90.1%
MultiScale PointNet (voting)	**90.7%**



ShapeNet Part Segmentation Benchmark

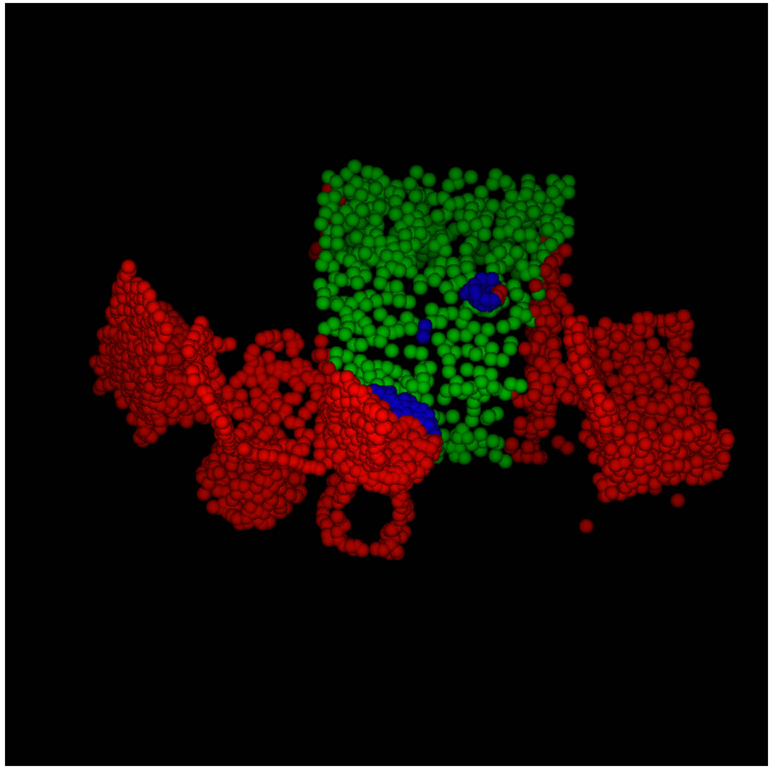
- First try...

	mIoU
PointNet	80.7%
PointNet + one-hot vector	82.7%
PointNet + one-hot vector + skip links etc.	83.7%
MultiScale PointNet	**83.8%**

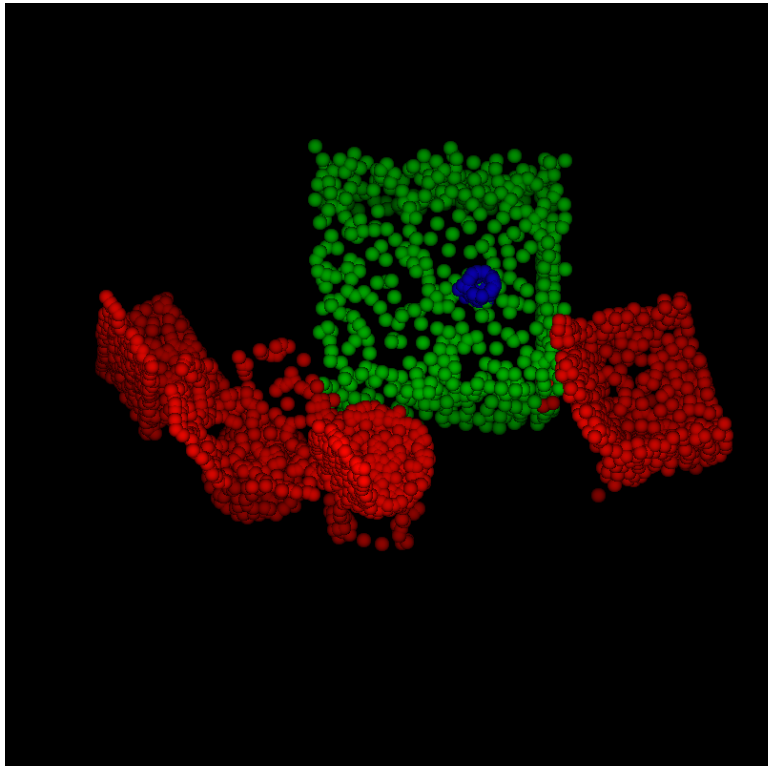
Semantic Segmentation in Scenes

	mIoU
PointNet	75.5%
MultiScale PointNet	**94.6%**

PointNet v1.0



PointNet v2.0: Multi-Scale PointNet

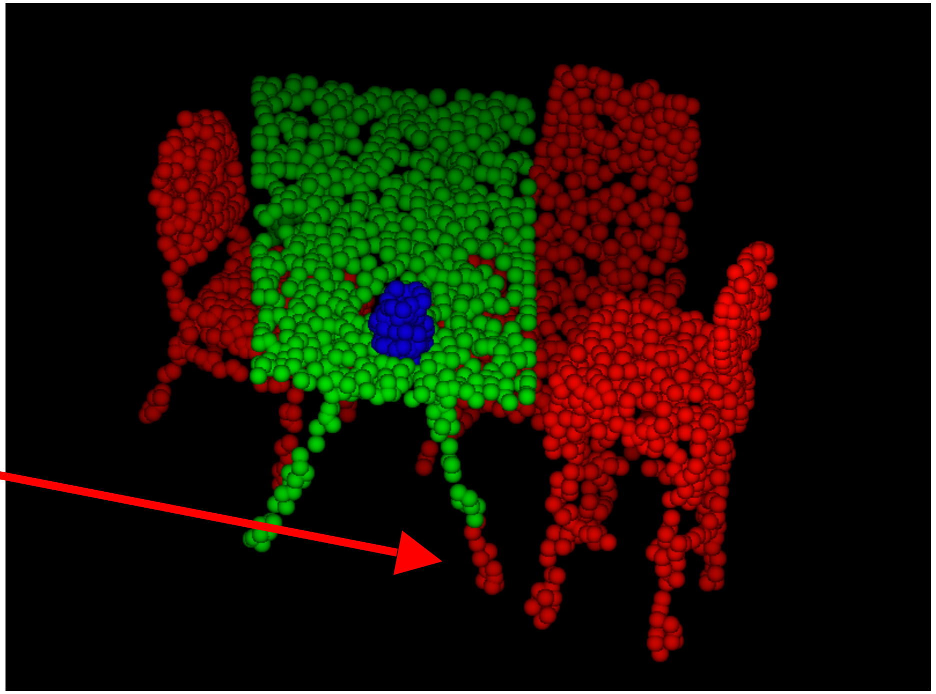


Semantic Segmentation in Scenes

	mIoU
PointNet	75.5%
MultiScale PointNet	**94.6%**

PointNet v2.0: Multi-Scale PointNet

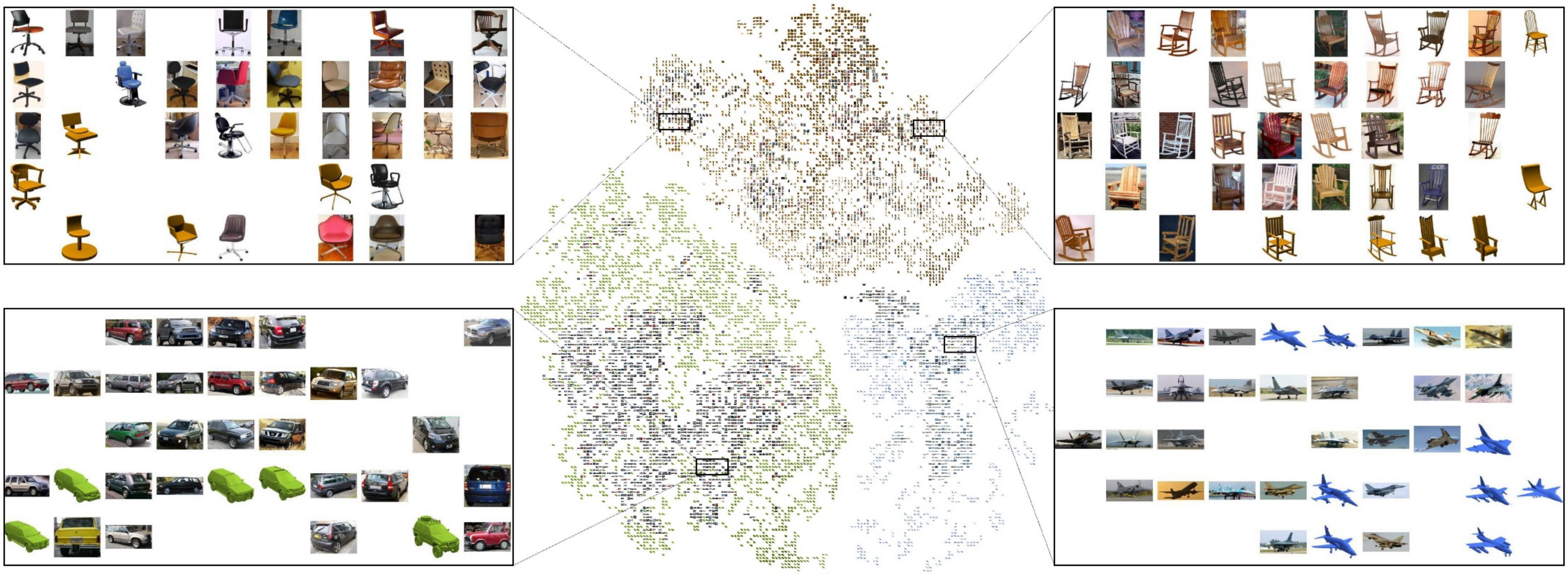
Misclassified table leg



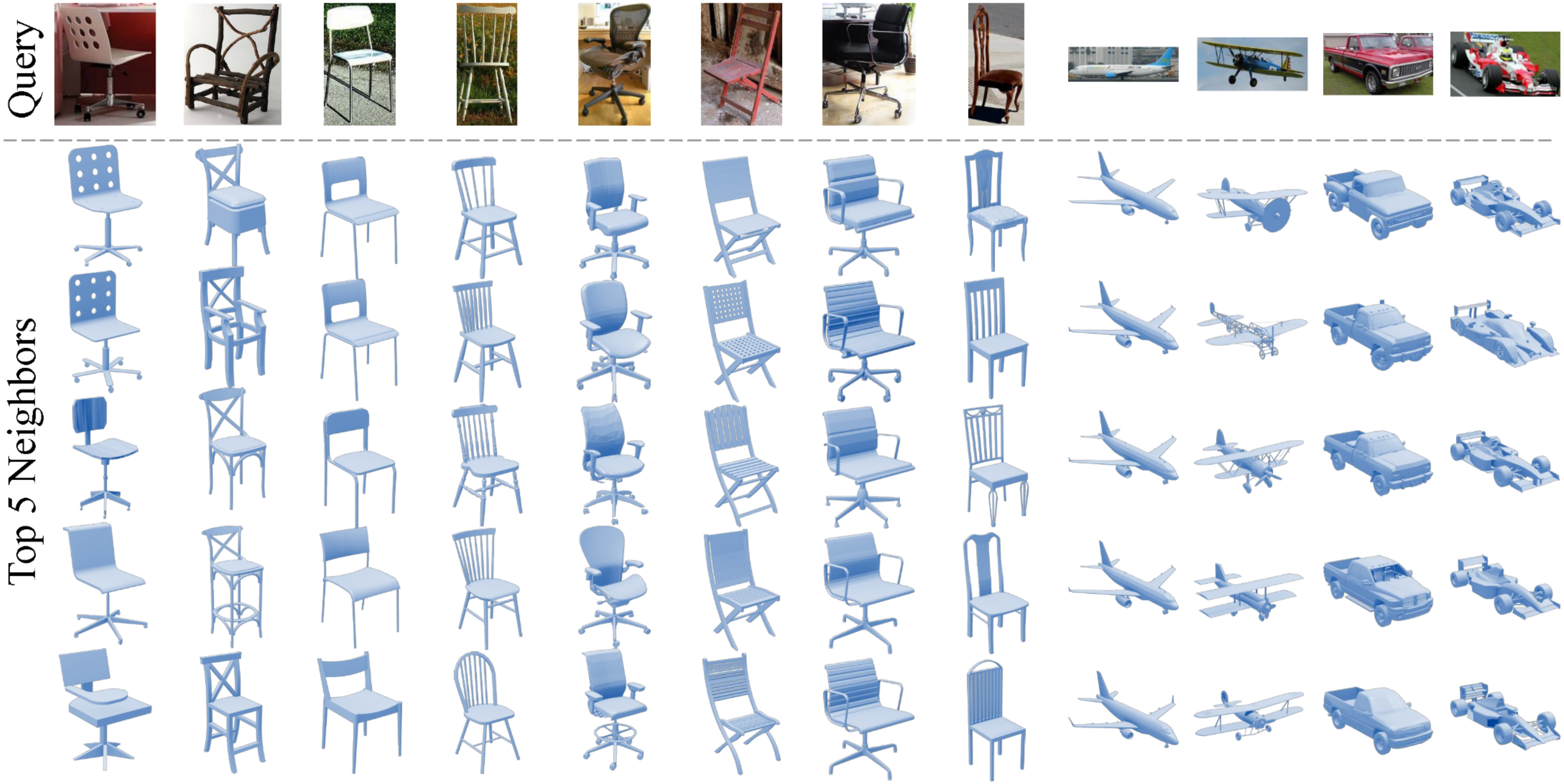
Geodesic distance based weights may help!

Joint Embedding of 3D shapes and 2D images

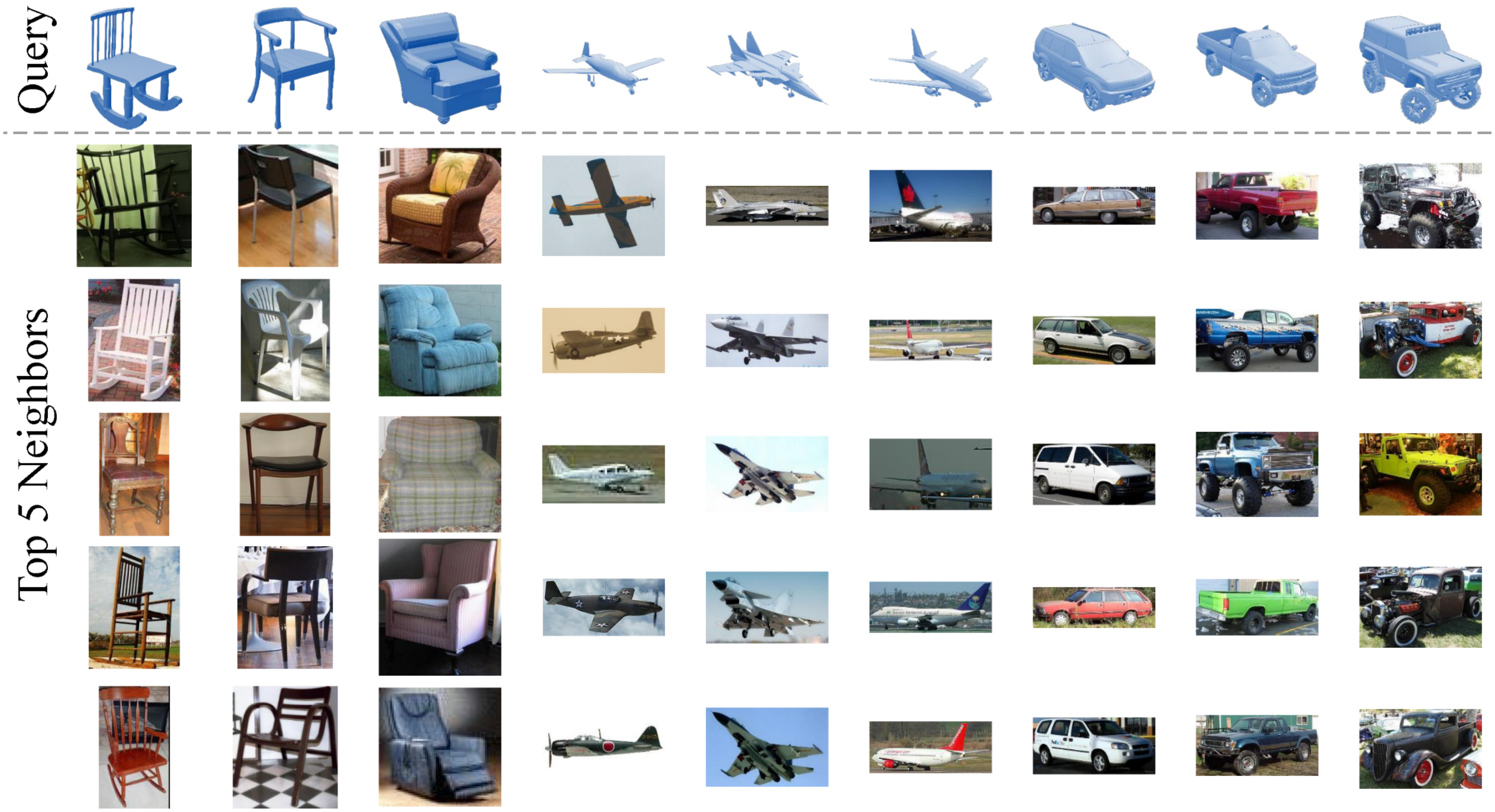
What is Joint Embedding of Shapes and Images?



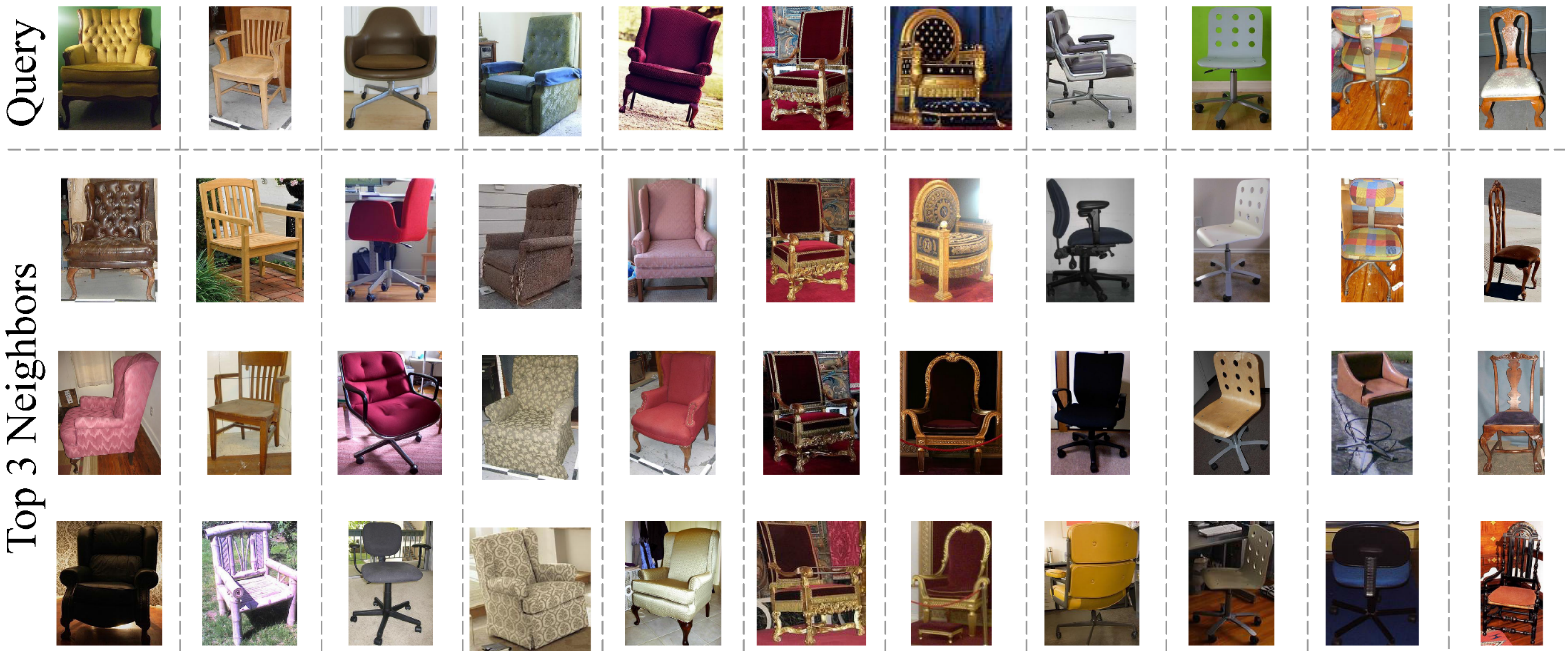
Application: Image-based Shape Retrieval



Application: Shape-based Image Retrieval

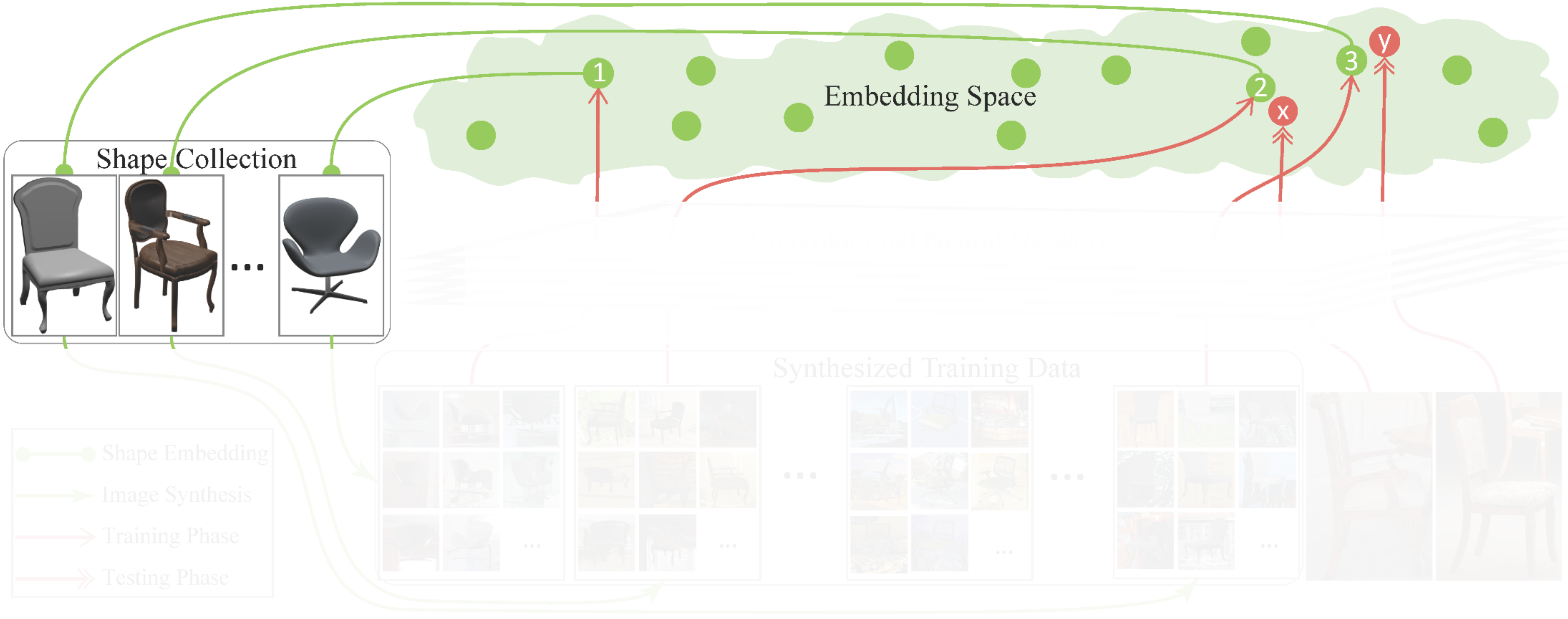


Application: Cross-View Image Retrieval



How to construct the joint embedding space?

Step 1: Construct Shape Embedding Space



Why not start from images?

- Object pose
- Lighting condition
- Texture variance
- Background clutterness
- …

I_1

I_2

I_3

Similarity between Shapes can be Easily Factored

Step 1: Construct Shape Embedding Space

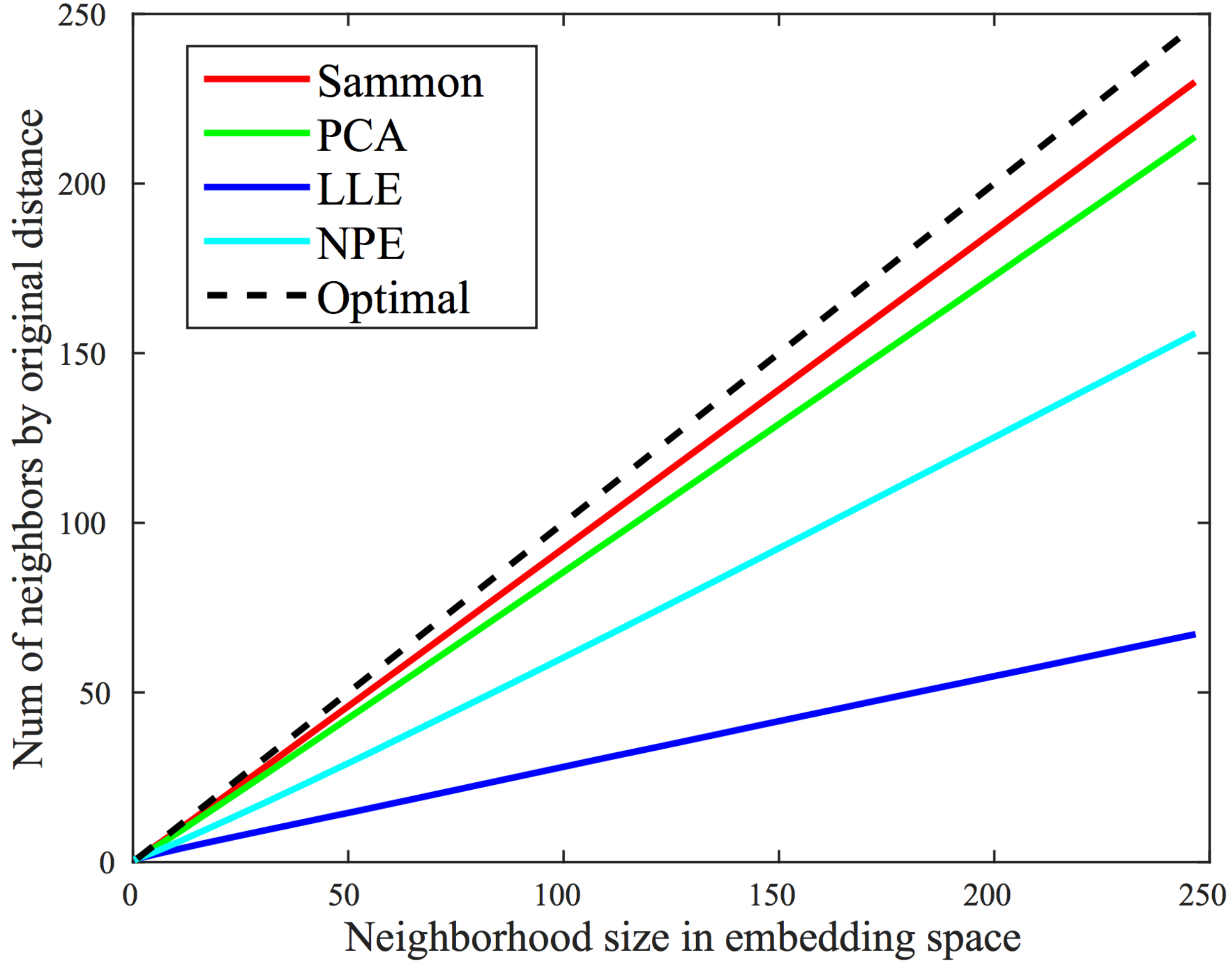
- Shape signature: Light Field Descriptor (with HoG feature)

Pairwise distance matrix of 3D shapes

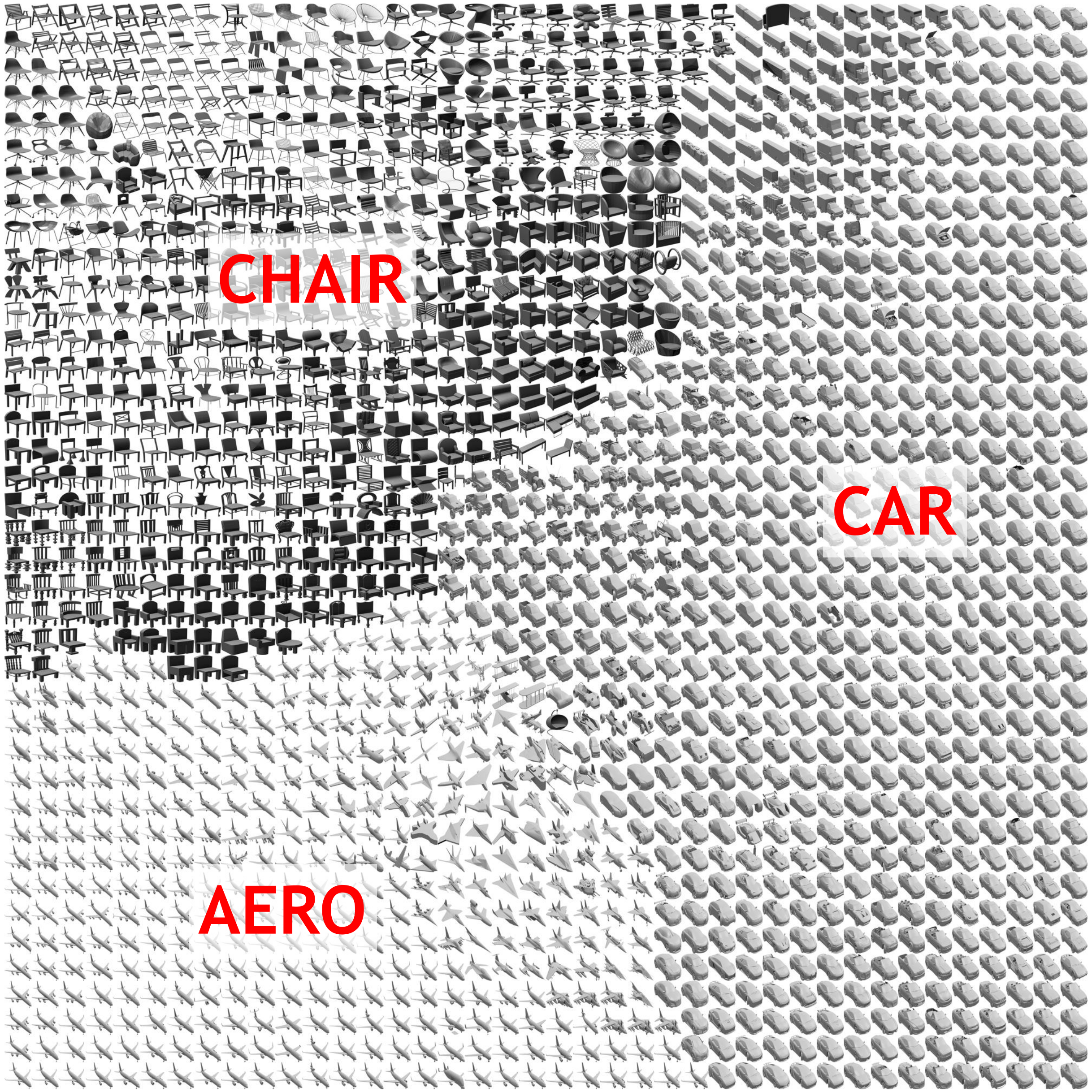
Step 1: Construct Shape Embedding Space

- Dimension reduction (MDS with Sammon mapping)

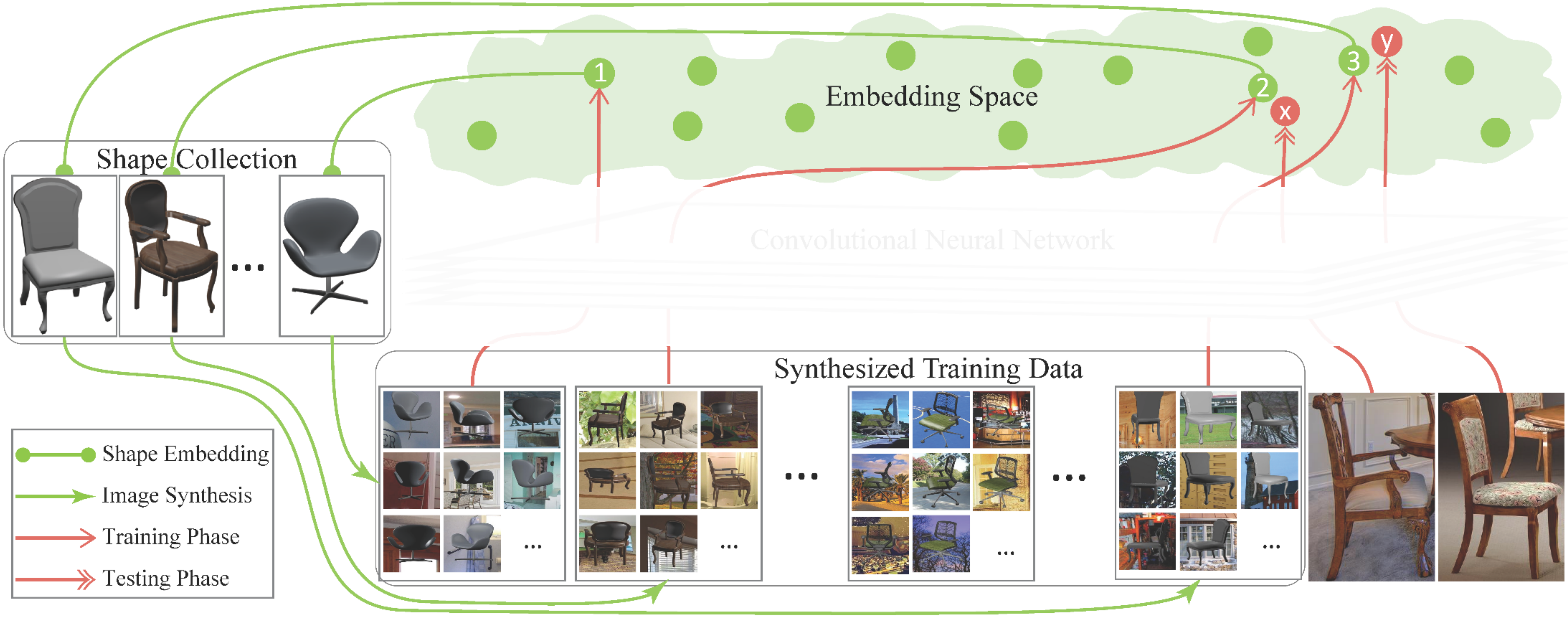
Key idea:
We care more about **structure of local neighborhood,** instead of distances between very dissimilar objects.



2D Visualization of Shape Embeddings



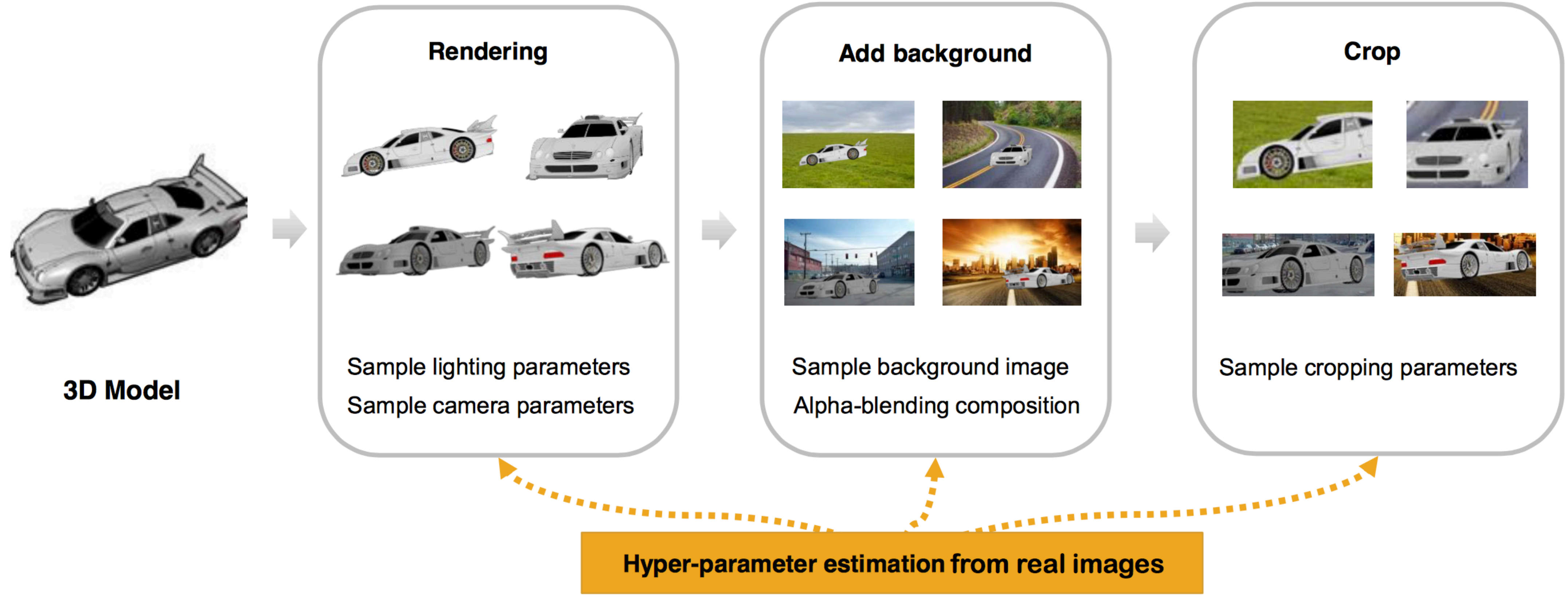
How to construct the joint embedding space?



Step 2: Projecting Images to the Joint Embedding space:
Prepare training data.

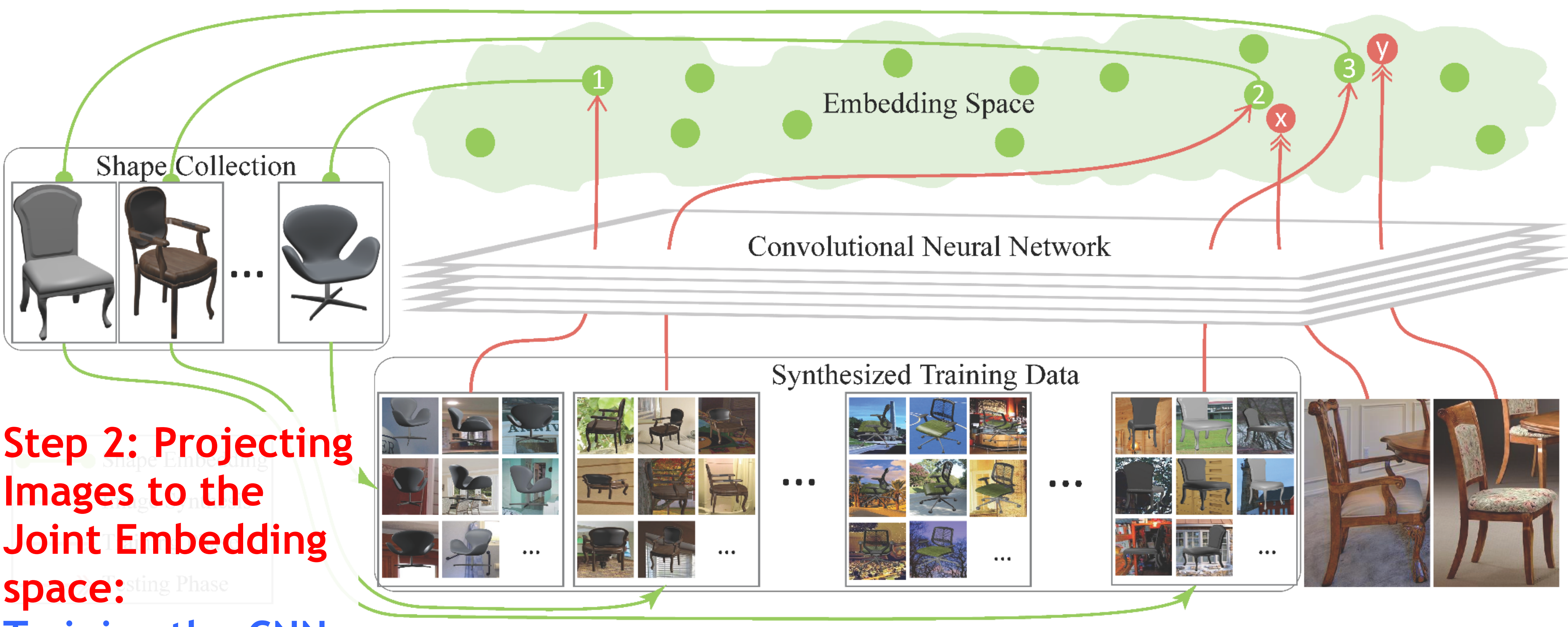
Step 2: Projecting Images to Joint Embedding Space

- Render for CNN Pipeline



Hao Su, Charles Qi, Yangyan Li, Leonidas Guibas, Render for CNN: Viewpoint Estimation in Images Using CNNs Trained with Rendered 3D Model Views, *ICCV 2015 Oral Presentation*

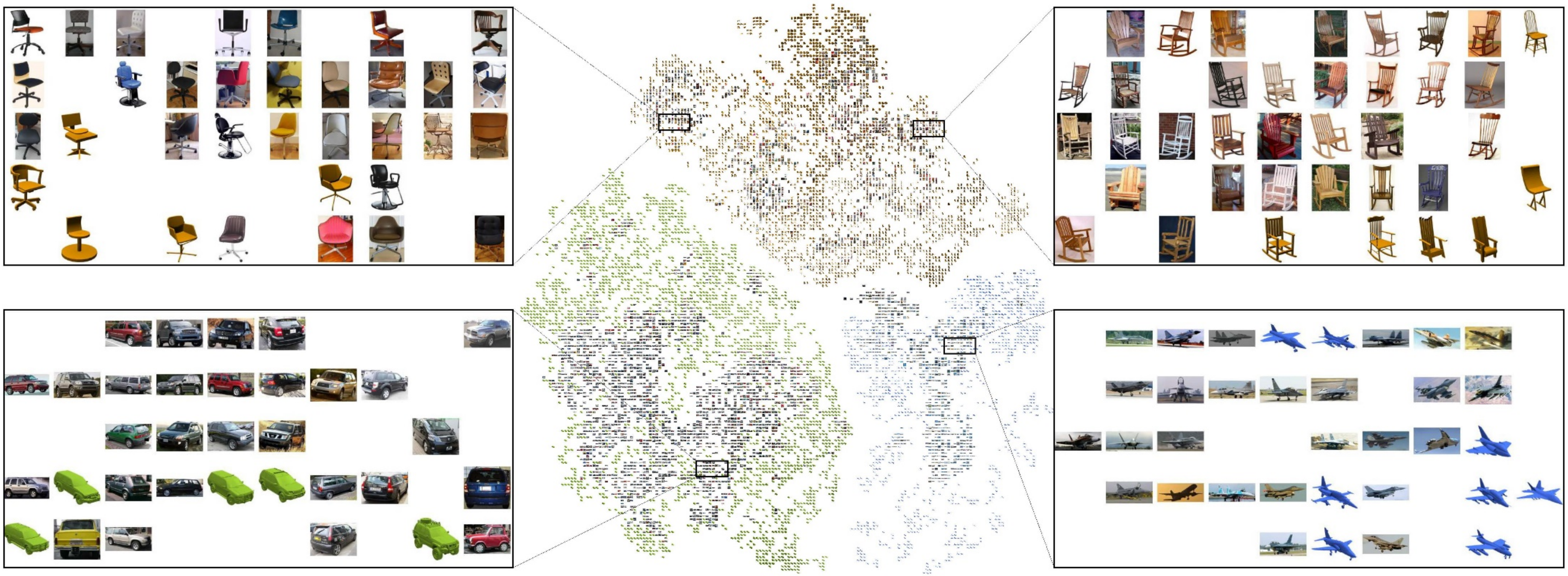
How to construct the joint embedding space?



Step 2: Projecting Images to the Joint Embedding space:

Training the CNN.

Joint Embedding Space of Shapes and Images



t-SNE visualization. Embeddings are projected into 2D

Results

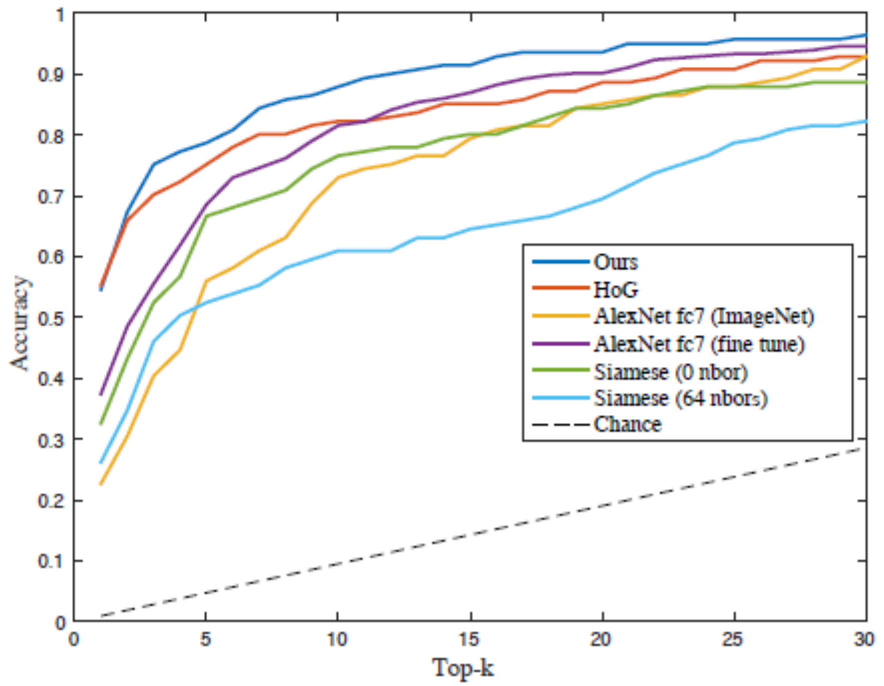


Figure 8: Comparison of top-k accuracy on image-based same-instance shape retrieval.

Median rank of	HoG	AlexNet fc7 (ImageNet)	AlexNet fc7 (fine tune)	Siamese (64 nbors)	Siamese (0 nbor)	Ours
first matched	**1**	7	5	3	3	**1**
last matched	32	84	71	94	49	**5**

Comparison of performance on shape-based same instance imag
retrieval

Other approaches to build 3D shape space: GANs

Learning a Probabilistic Latent Space of Object Shapes via 3D Generative-Adversarial Modeling

NIPS 2016





Jiajun Wu* Chengkai Zhang* Tianfan Xue Bill Freeman Josh Tenenbaum

MIT CSAIL Google Research

* indicates equal contribution