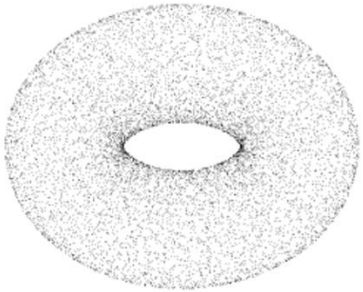
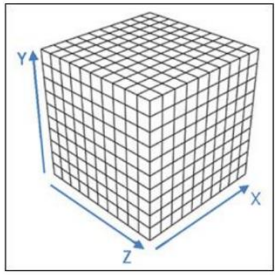
3D Deep Learning

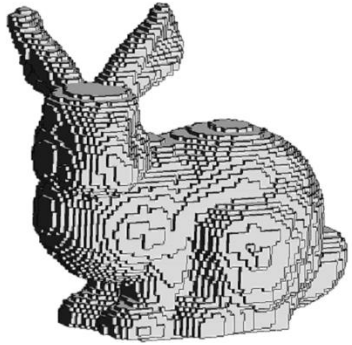
What is "3D"?

Point Clouds

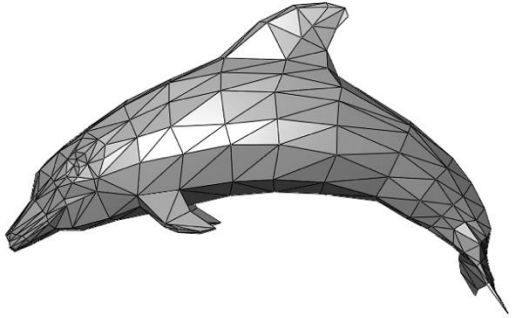


Voxels

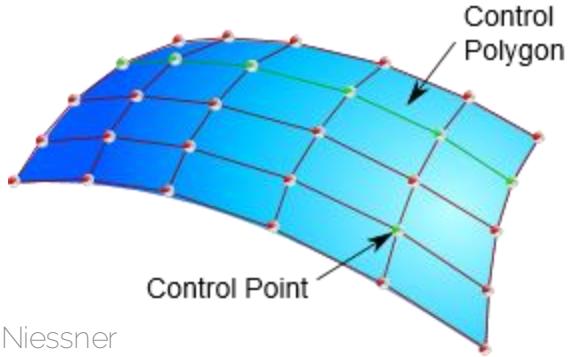




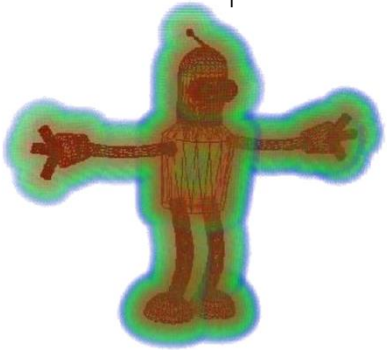
Polygonal Meshes

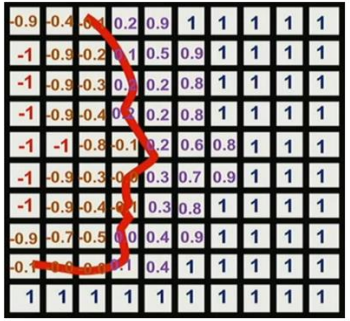


Parametric Surfaces



Implicit Surfaces

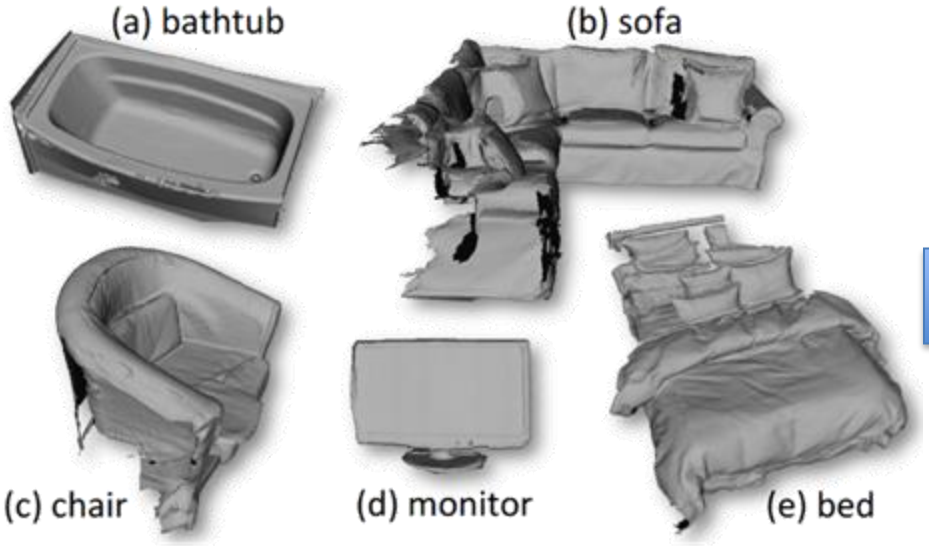


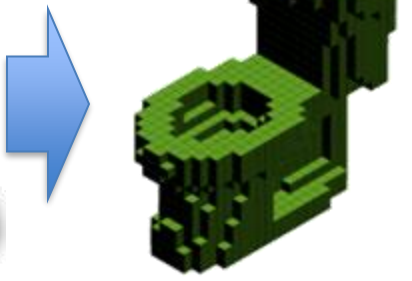


Truncated Signed Distance Field (TSDF)

3D Tasks

Tasks: 3D Classification

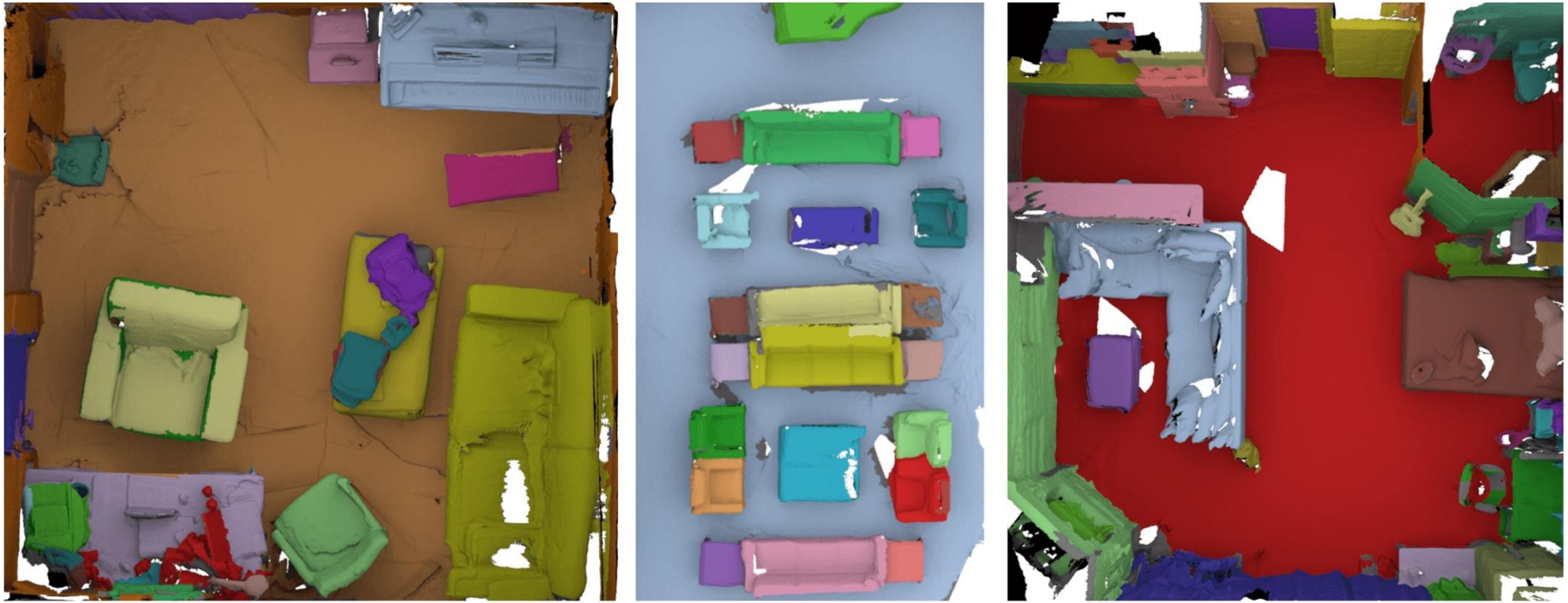




Class from 3D model (e.g., obtained with Kinect Scan)

[Maturana et al. 15] & [Qi et al. 16] 3D vs Multi-view

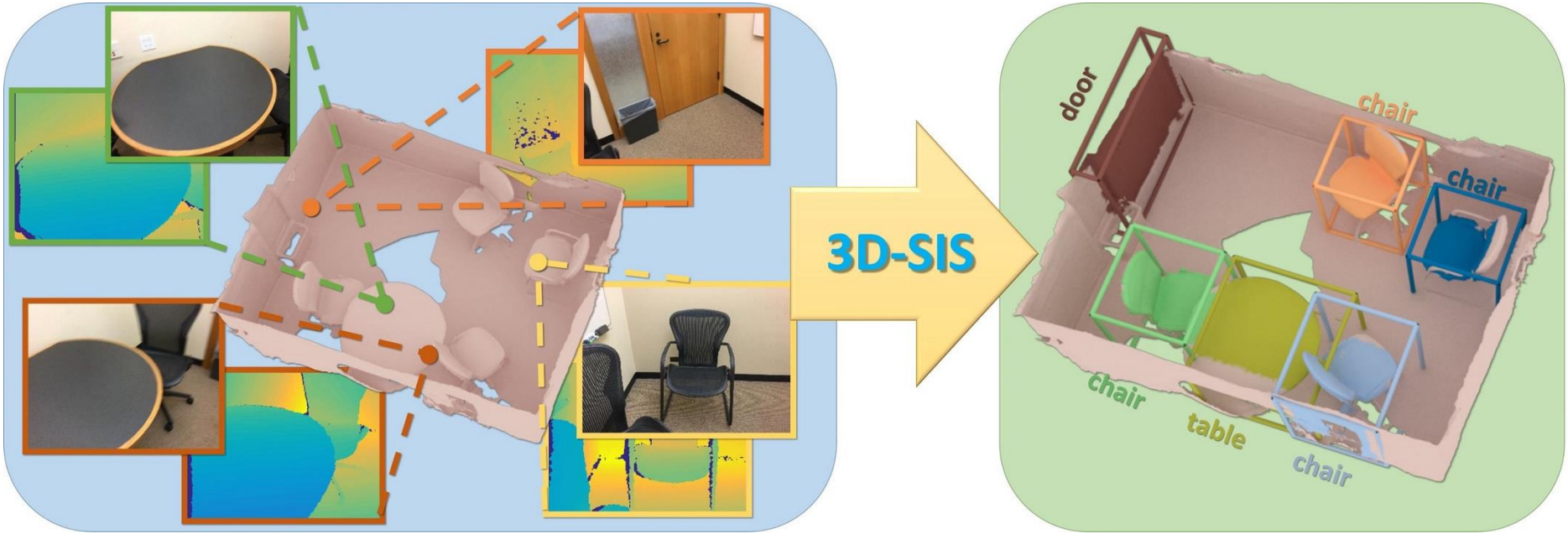
Tasks: 3D Semantic Segmentation



1500 densely annotated 3D scans; 2.5 mio RGB-D frames

[Dai et al. 17] ScanNet

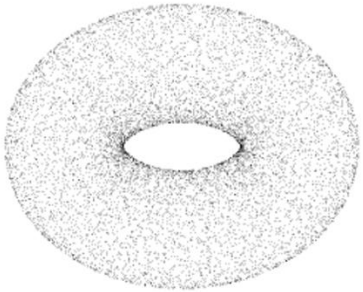
Tasks: 3D Detection / Instance Segmentation

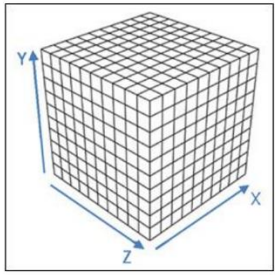


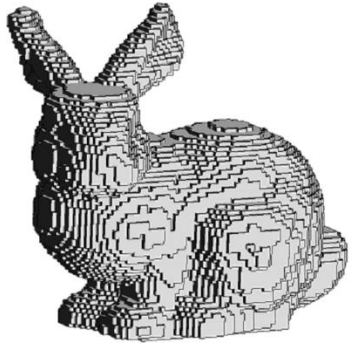
[Hou et al. 19] 3D-SIS

Generative 3D Tasks

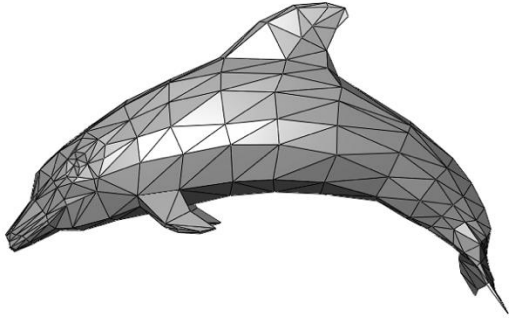
Point Clouds



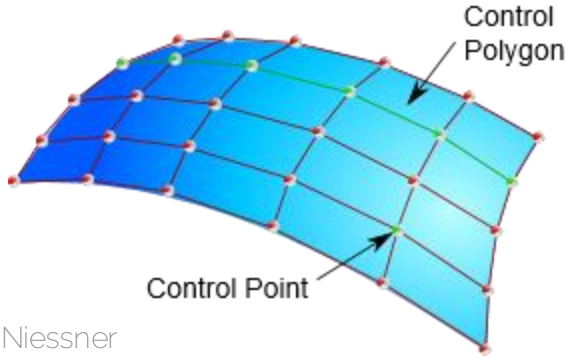




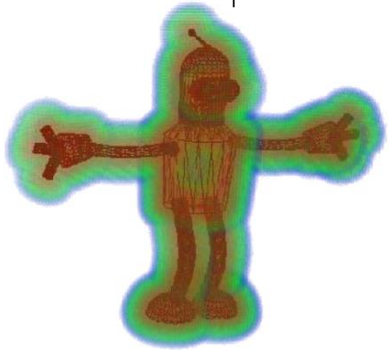
Polygonal Meshes

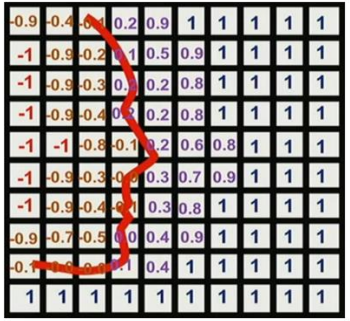


Parametric Surfaces



Implicit Surfaces



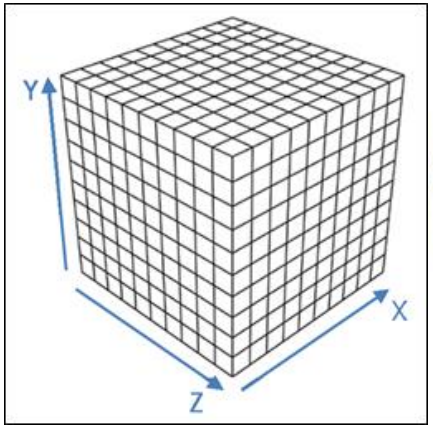


Volumetric Grids

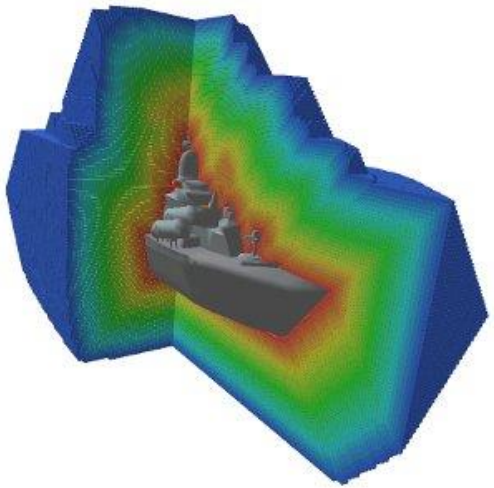
Volumetric Grids

Volumetric Data Structures

- Occupancy grids
- Ternary grids
- Distance Fields
- Signed Distance fields



(binary) Voxel Grid



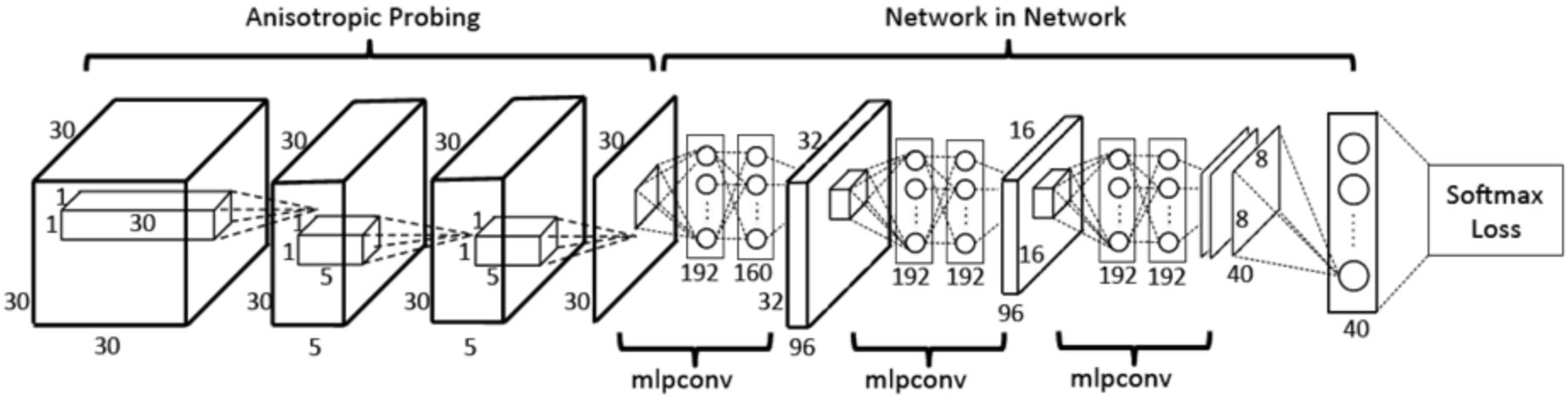
Volumetric Grids

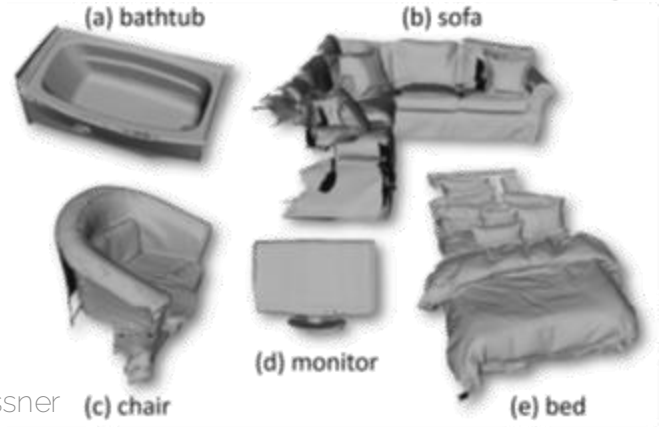
- 3D Convolutions
  - Direct extension of 2D convolutions

- Often: occupancy grid, signed/unsigned distance fields

- Operate on regular grid structures
  - Get neighborhood structures and spatial propagation
  - Well-defined pooling and unpooling for hierarchical processing
  - Cubic growth in memory

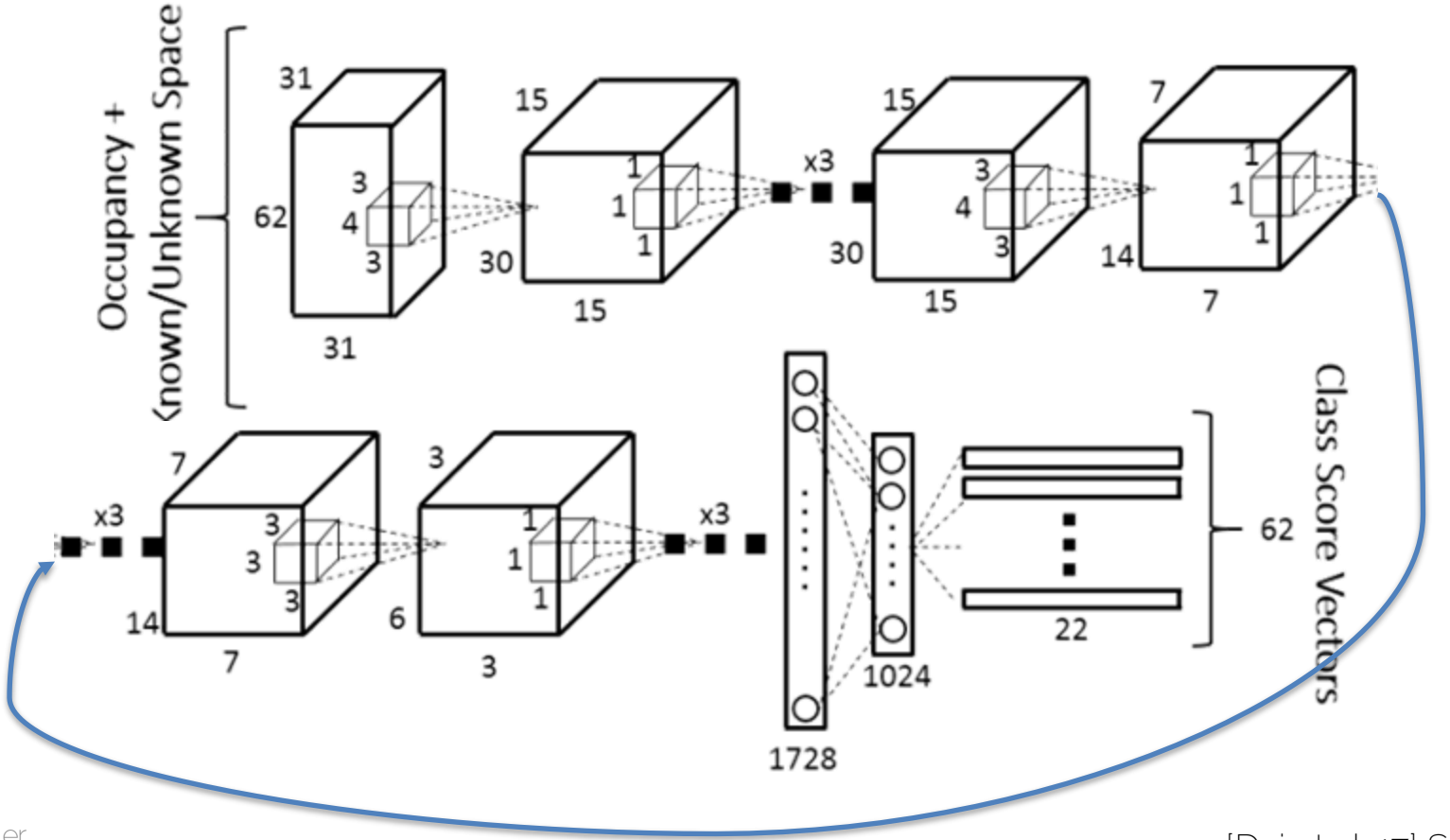
Object Classification /w 3DCNNs





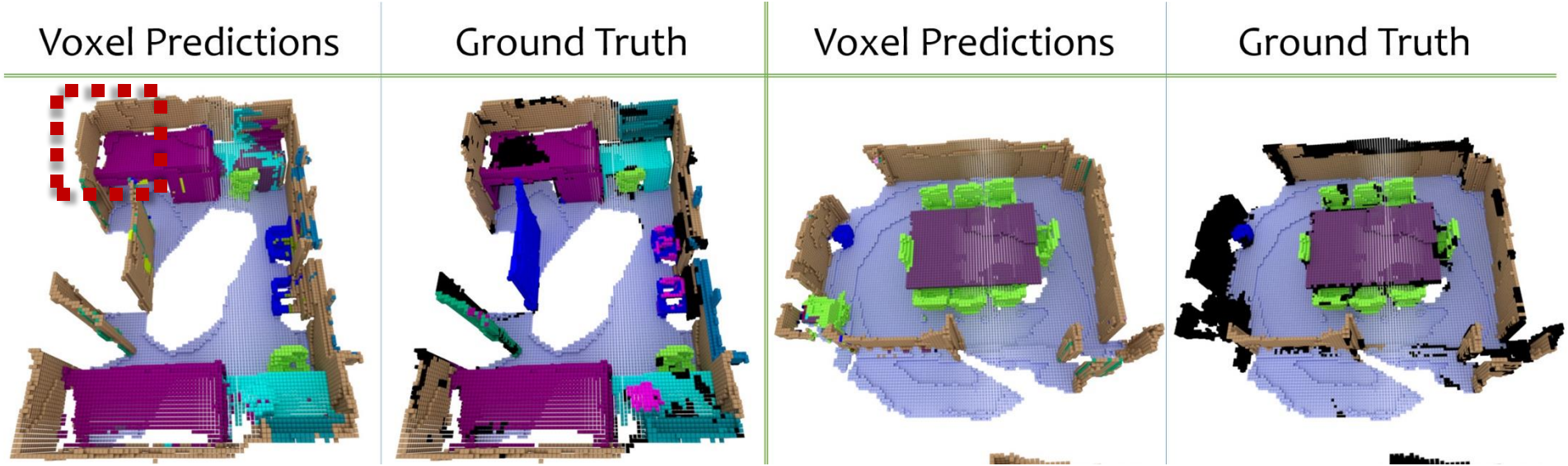
3D CNN Object Classification Architecture

ScanNet: Semantic Segmentation in 3D



[Dai et al. 17] ScanNet

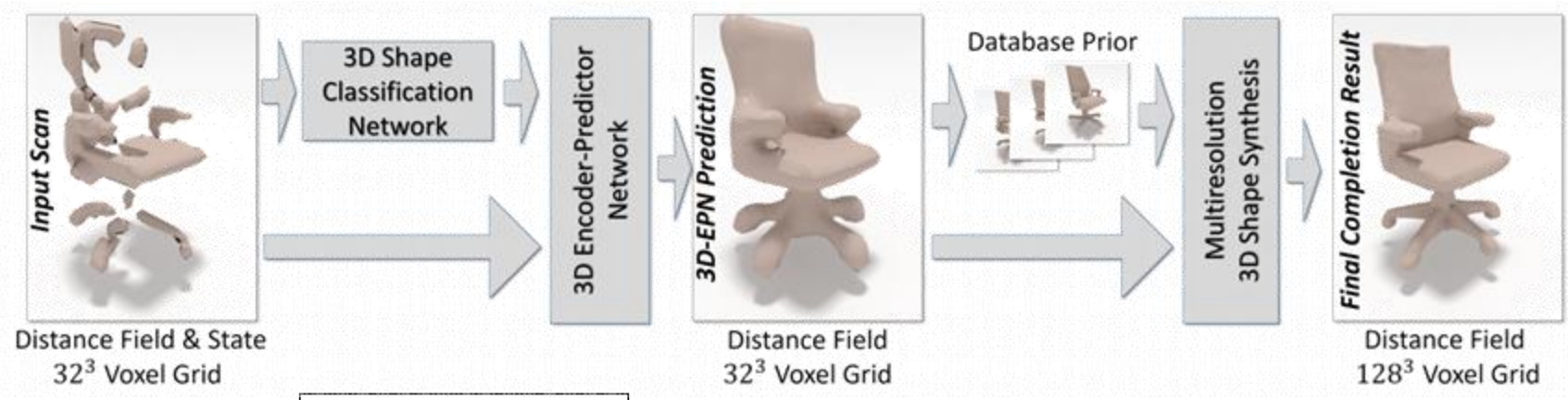
ScanNet: Sliding Window

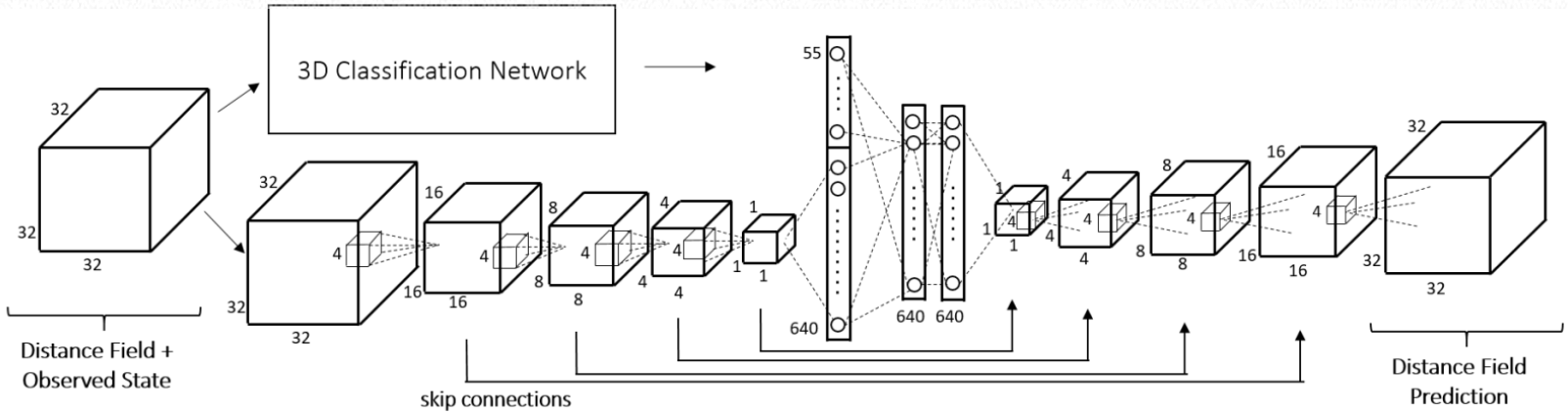


unannotated wall floor chair table desk bed bookshelf sofa sink bathtub toilet
curtain counter door window shower curtain refrigerator picture cabinet otherfurniture

[Dai et al. 17] ScanNet

Generative Tasks: 3D Shape Completion

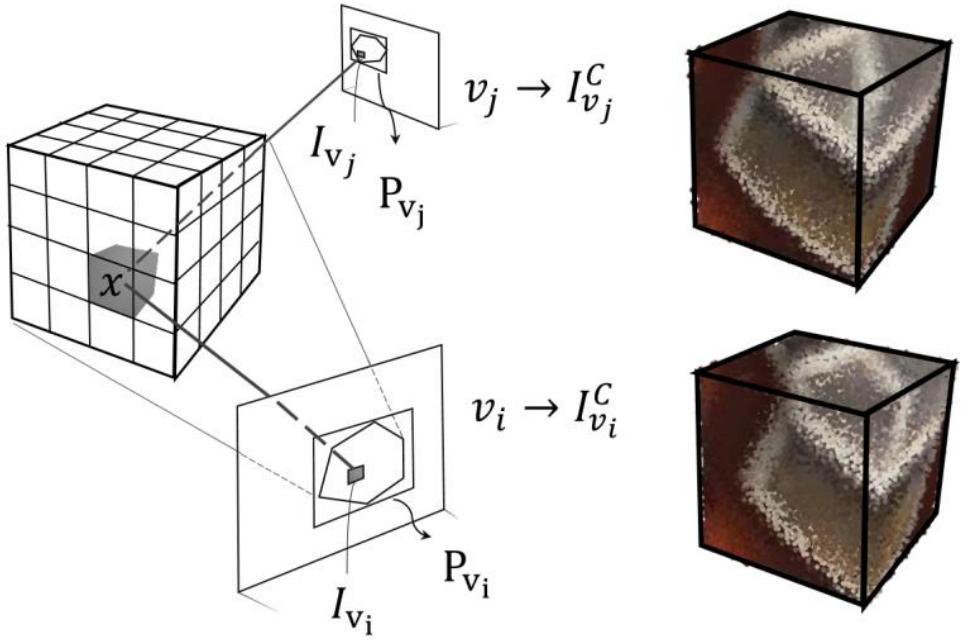


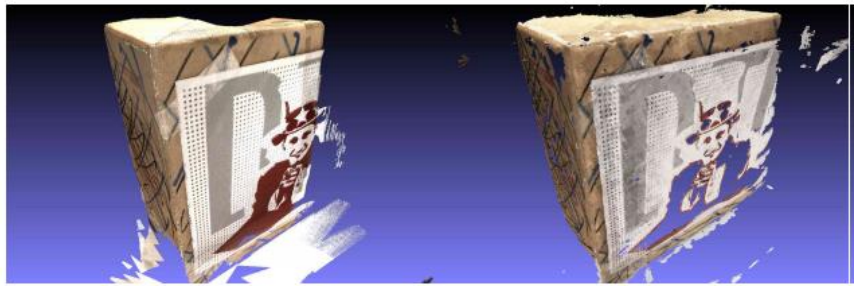


Works with 32 x 32 x 32 voxels…

[Dai et al. 17] CNNComplete

SurfaceNet: Stereo Reconstruction





(a) reference model (b) **SurfaceNet**

Run on 32 x 32 x 32 blocks -> takes forever…

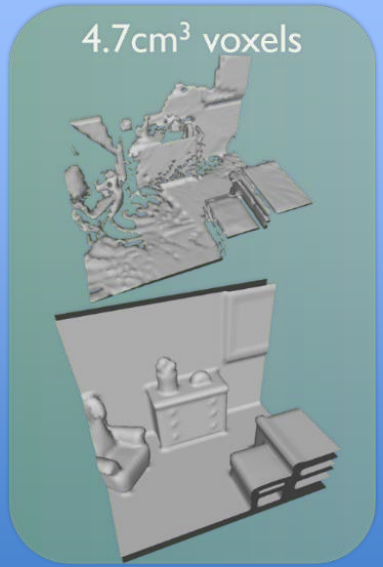
ScanComplete: Fully Convolutional

Train on crops of scenes

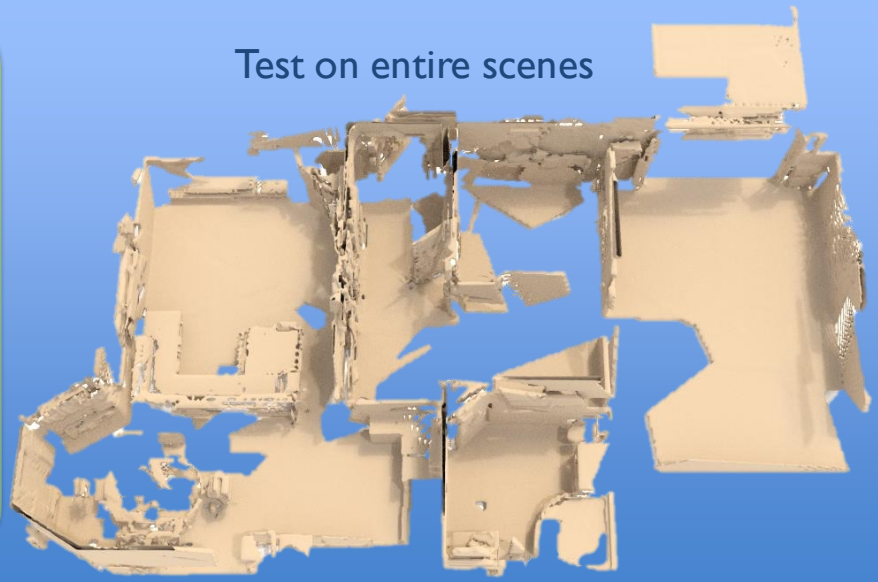
4.7cm^3 voxels

Train Block: Input Partial Scan

Train Block: Complete Target

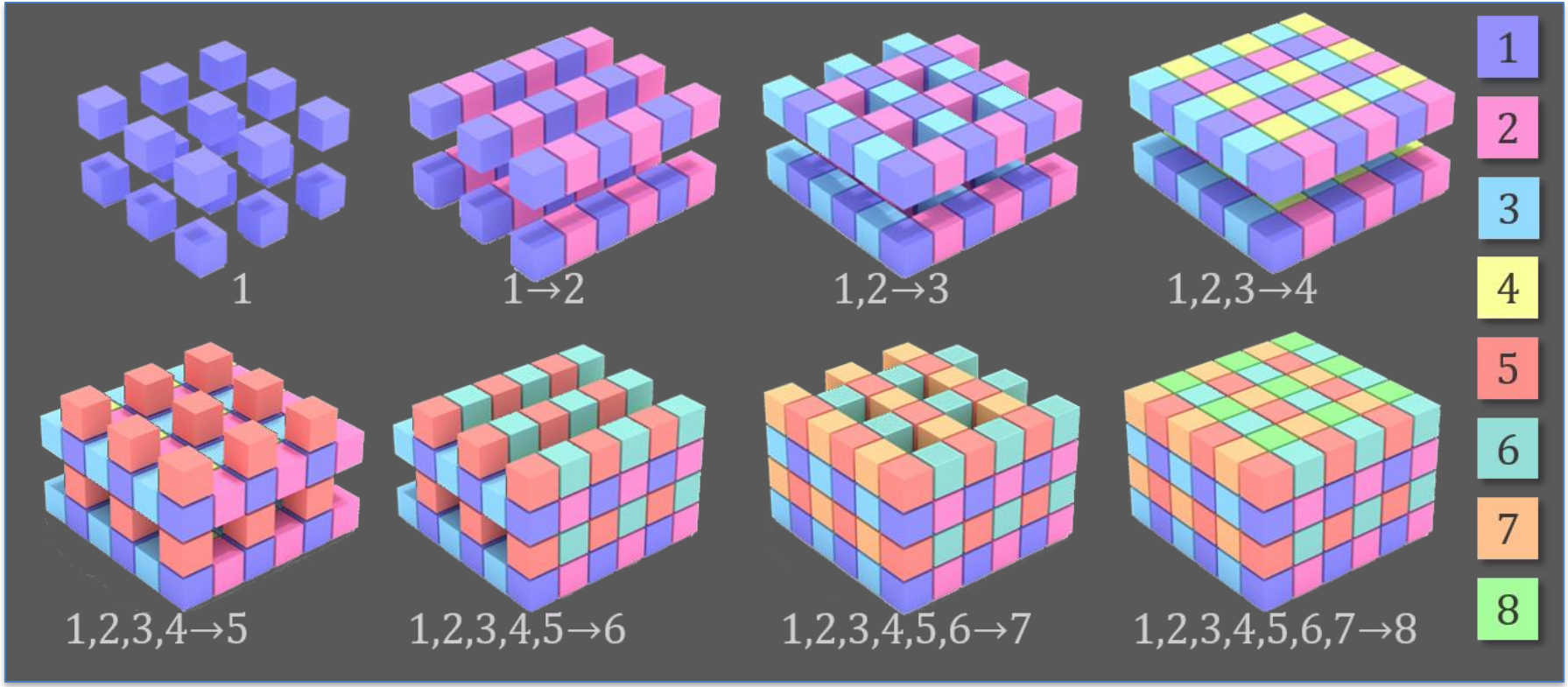


Test on entire scenes



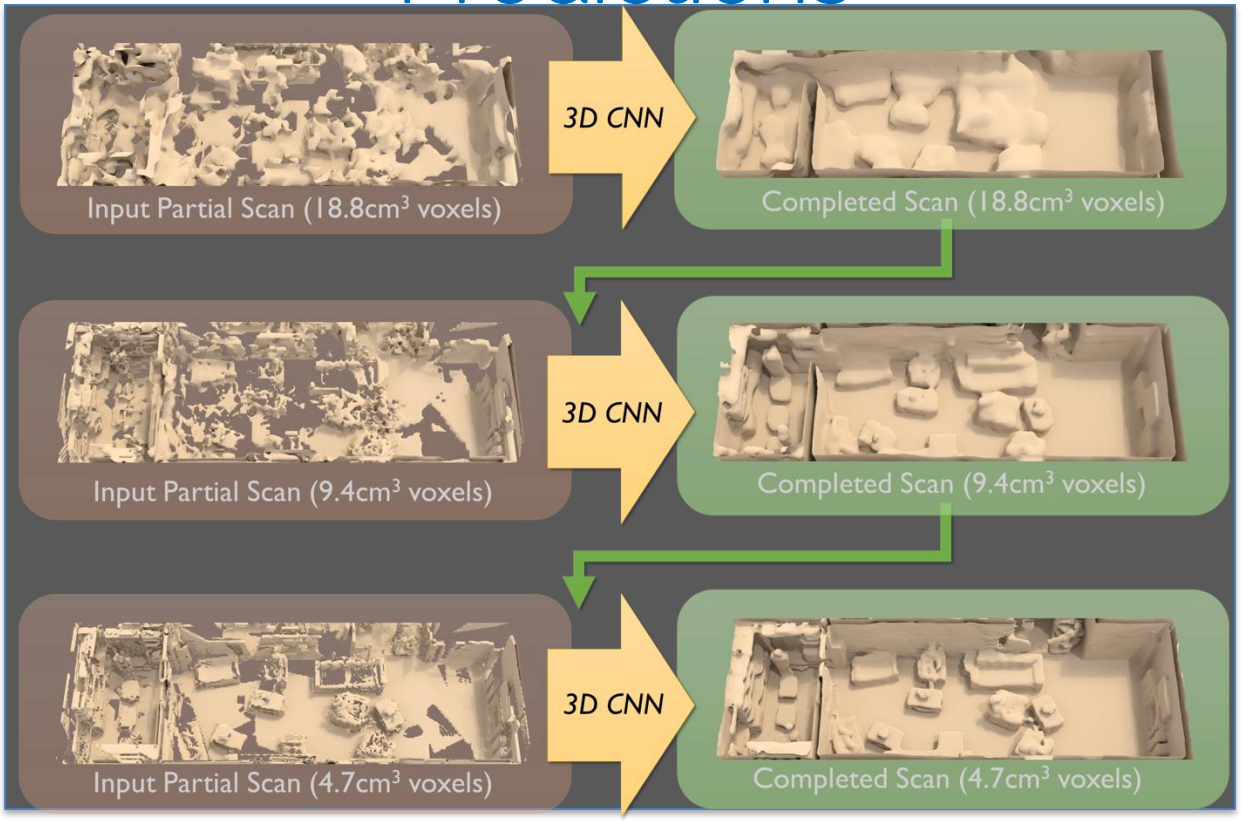
[Dai et al. 18] ScanComplete

Dependent Predictions: Autoregressive Neural Networks



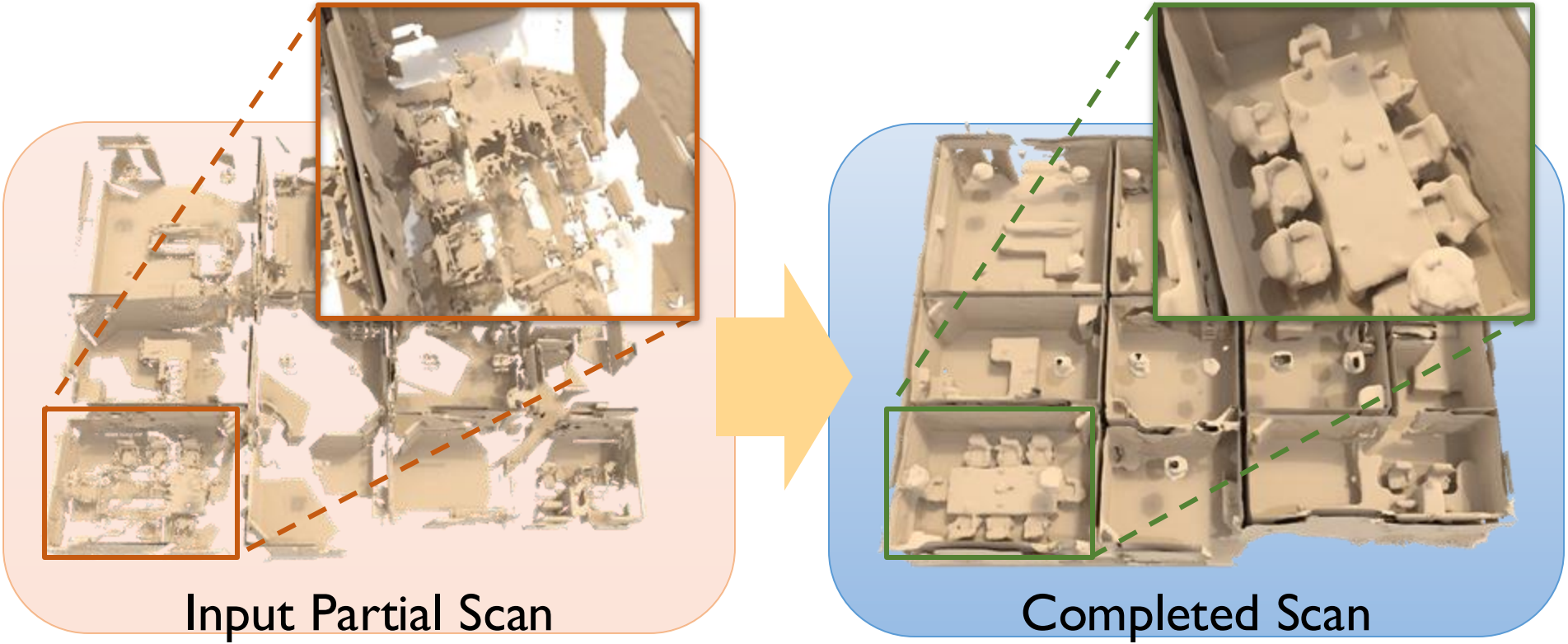
[Dai et al. 18] ScanComplete

Spatial Extent: Coarse-to-Fine Predictions



[Dai et al. 18] ScanComplete

ScanComplete: Fully Convolutional

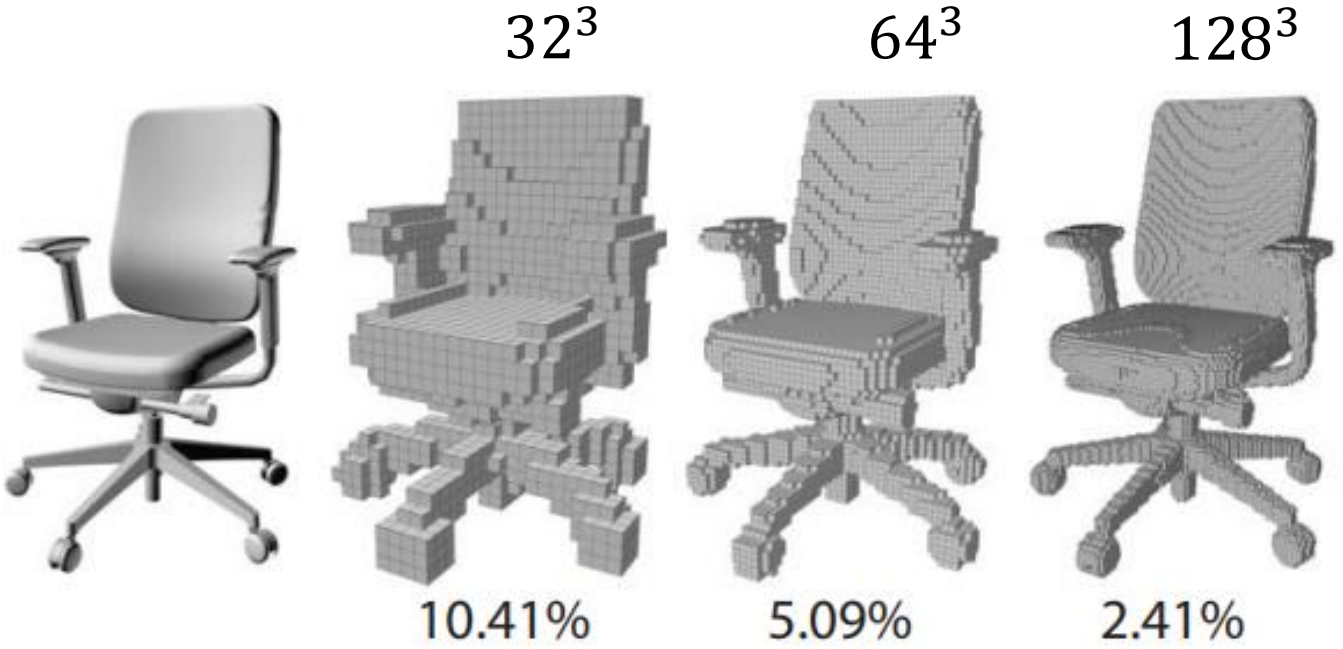


[Dai et al. 18] ScanComplete

Conclusion so far

- Volumetric grids are easy and naturally extend 2D CNNs and other concepts
  - Encode free space
  - Encode distance fields
  - Need a lot of memory
  - Need a lot of processing time
  - But can be used sliding window or fully-conv.

Conclusion so far



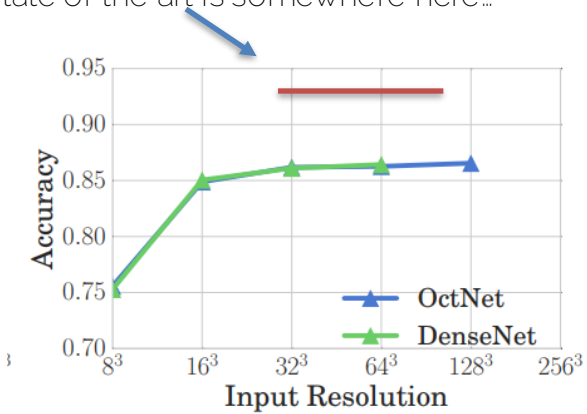
Surface occupancy gets smaller with higher resolutions

Volumetric Hierarchies

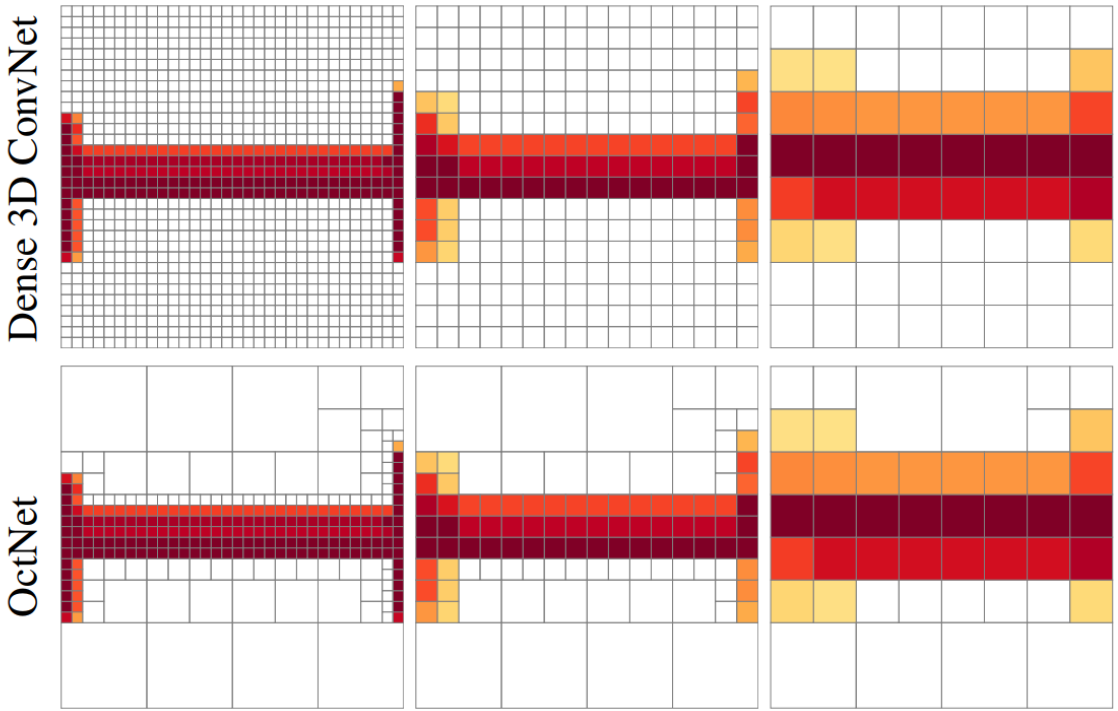
Discriminative Tasks

Structure is known in advance!

State of the art is somewhere here…



(b) Accuracy



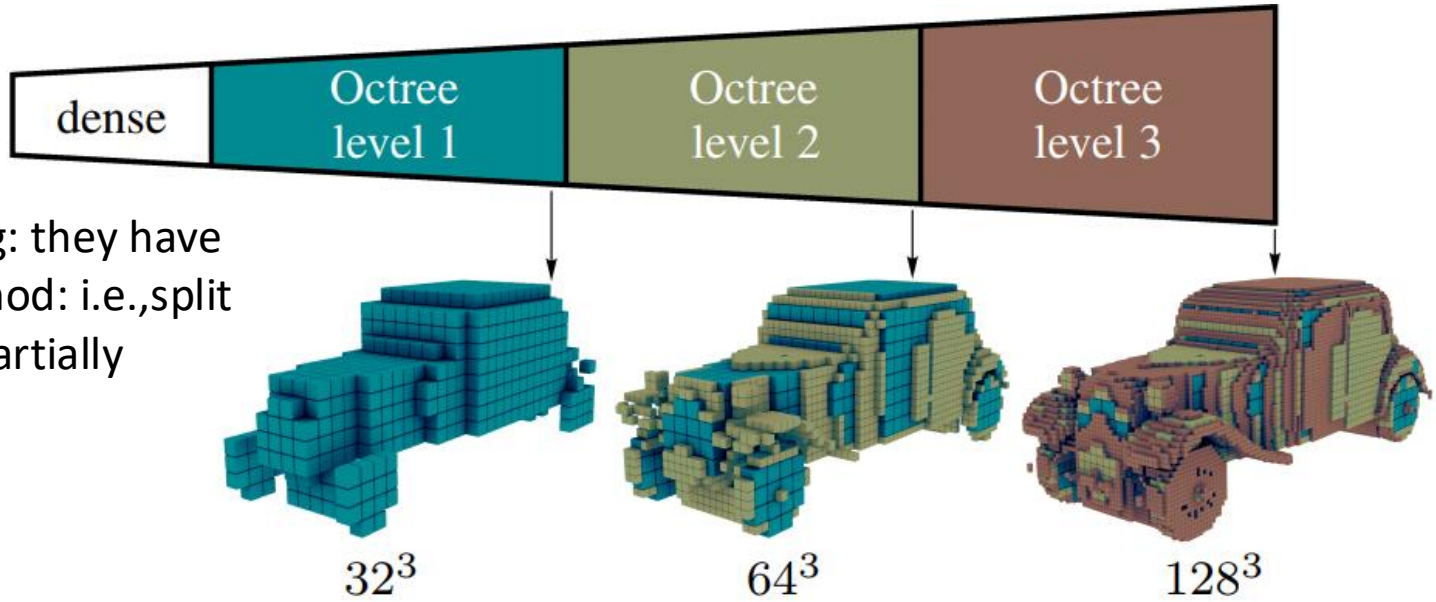
(a) Layer 1: 32^3 (b) Layer 2: 16^3 (c) Layer 3: 8^3

OctNet: Learning Deep 3D Representations at High Resolutions (CVPR 2017)

O-CNN: Octree-based Convolutional Neural Networks for 3D Shape Analysis (SIG17)

Generative Tasks

Need to infer structure!



Pretty interesting: they have end-to-end method: i.e.,split voxels that are partially occupied

Octree Generating Networks: Efficient Convolutional Architectures for High-resolution Outputs

OctNetFusion: Learning Depth Fusion from Data (that one not end to end)

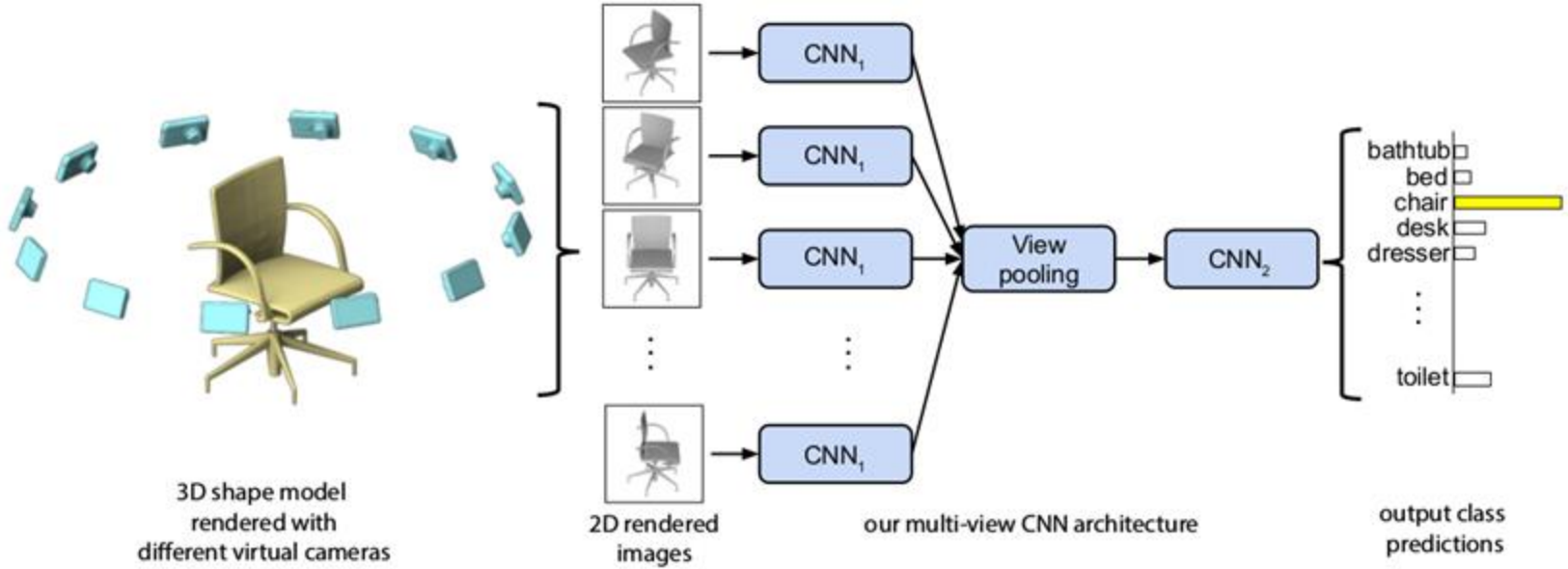
Conclusion so far

- Hierarchies
  - are great for reducing memory and runtime
  - Comes at a performance hit
  - Easier for discriminative tasks when structure is known

Multi-view

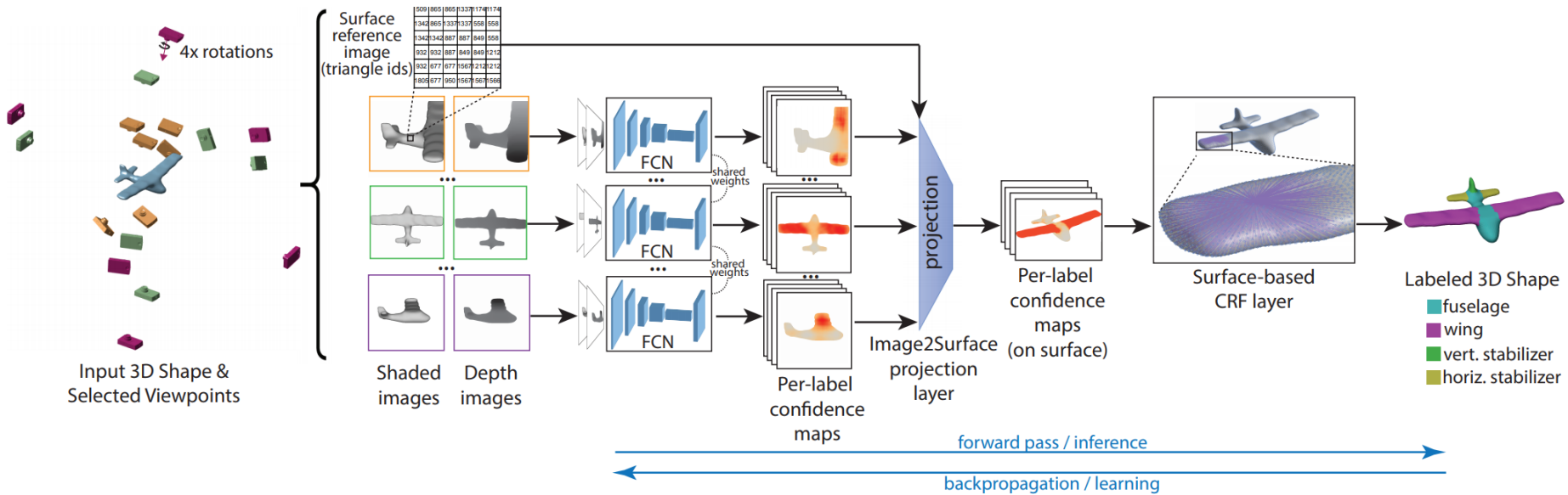
Multiple Views: Classification

- RGB images from fixed views around object:
  - view pooling for classification (only RGB; no spatial corr.)



Multi-view Convolutional Neural Networks for 3D Shape Recognition

Multiple Views: Segmentation



[3D Shape Segmentation with Projective Convolutional Networks]

This one is interesting in a sense that it does 3D shape segmentation (only on CAD models)

But it uses multi-view and has a spatial correlation on top of the mesh surface

Fun thing…

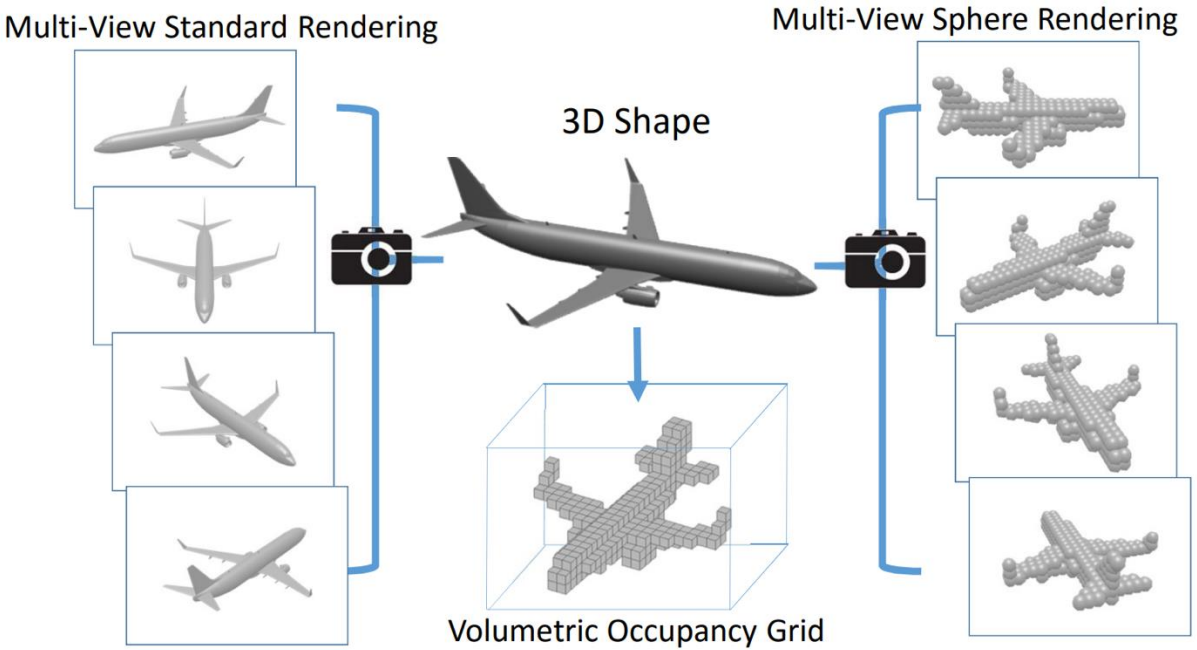


Figure 1. 3D shape representations.

Method	#Views	Accuracy (class)	Accuracy (instance)
SPH (reported by [33])	-	68.2	-
LFD (reported by [33])	-	75.5	-
FV (reported by [32])	12	84.8	-
Su-MVCNN [32]	80	90.1	-
PyramidHoG-LFD	20	87.2	90.5
Ours-MVCNN	20	89.7	92.0
Ours-MVCNN-MultiRes	20	**91.4**	**93.8**

Table 3. Comparison of multi-view based methods. Numbers reported are classification accuracy (class average and instance average) on ModelNet40.

Hybrid: Volumetric + Multi-view

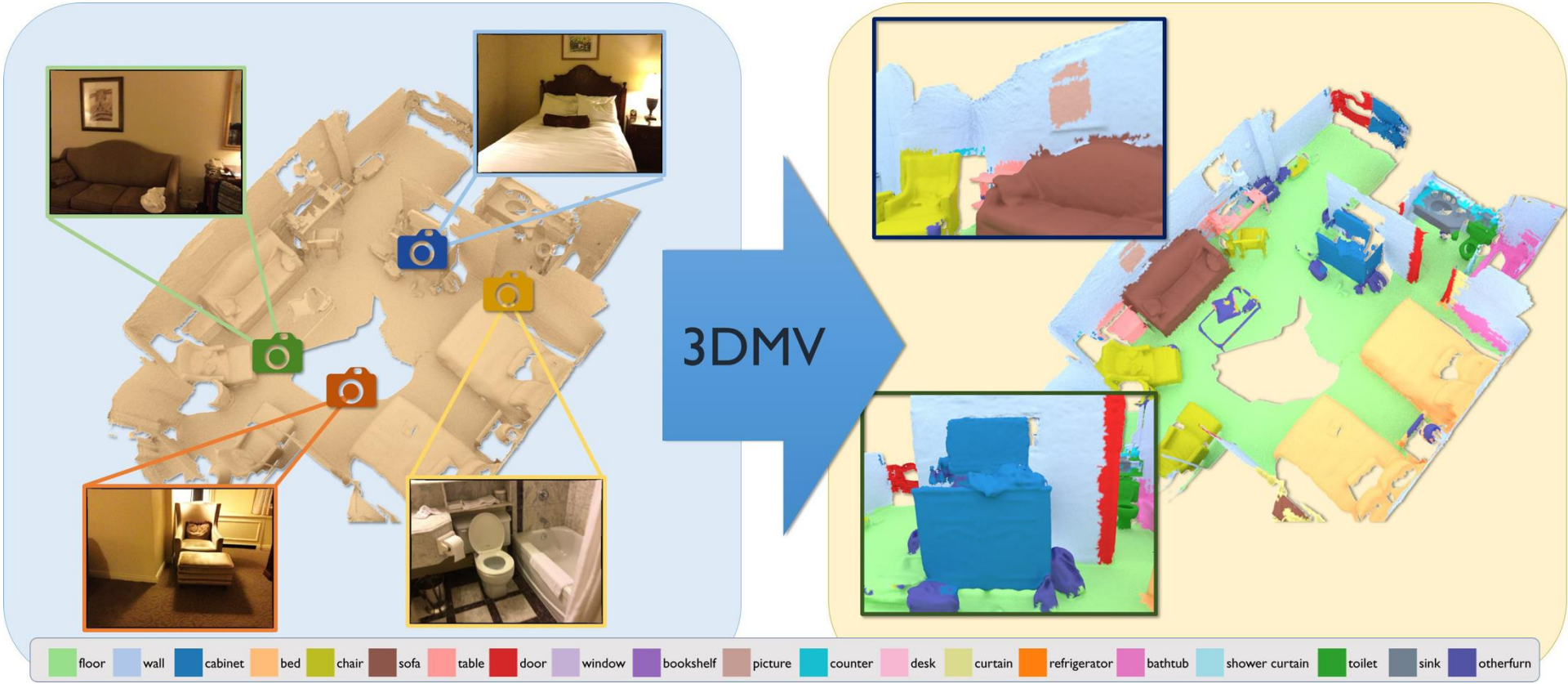
2D + 3D Semantic Segmentation

	avg class accuracy
geometry only	54.4
geometry + voxel colors	55.9

Resolution Mismatch!

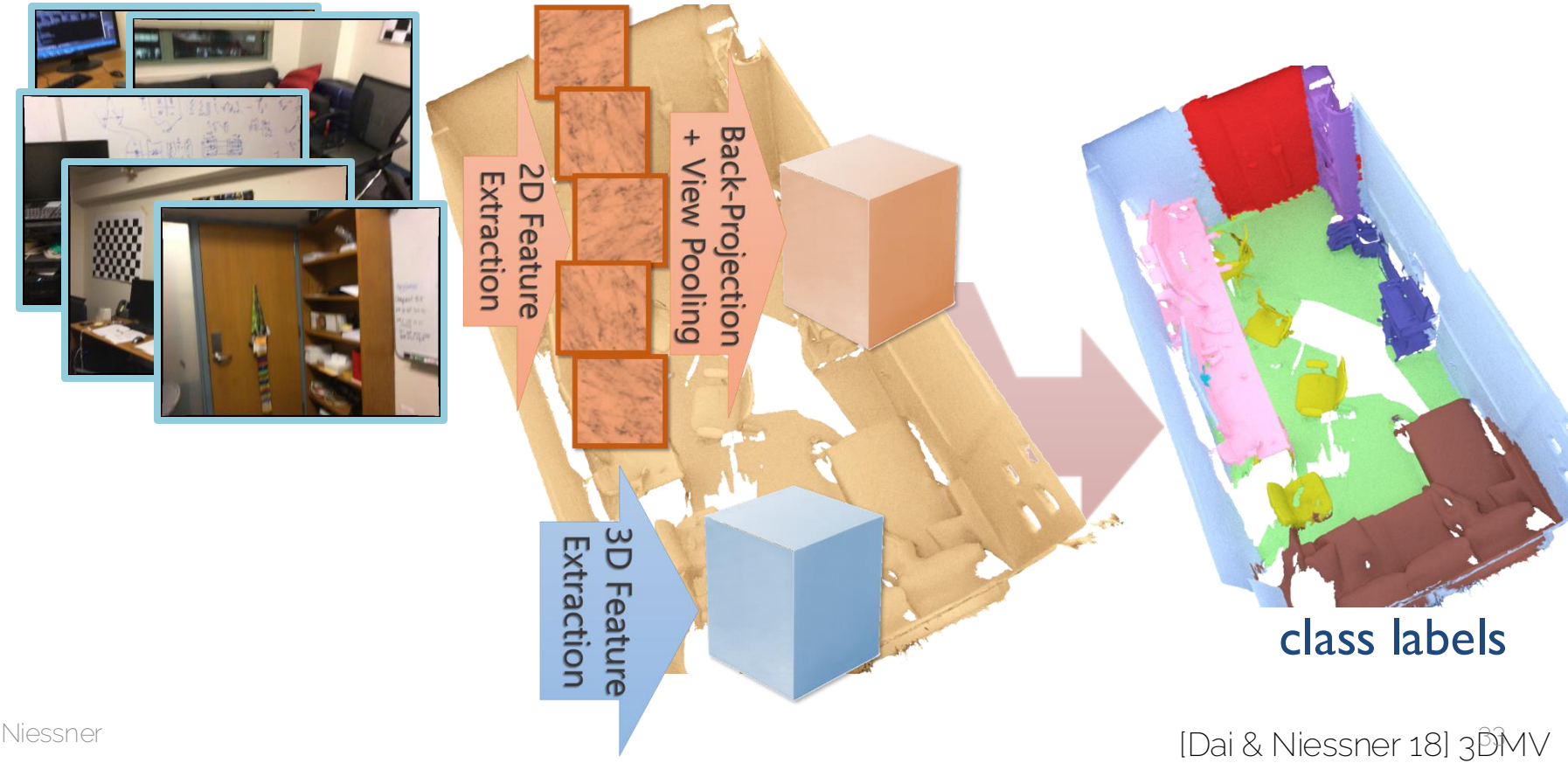
[Dai & Niessner 18] 3DMV

3D Volumetric + Multi-view



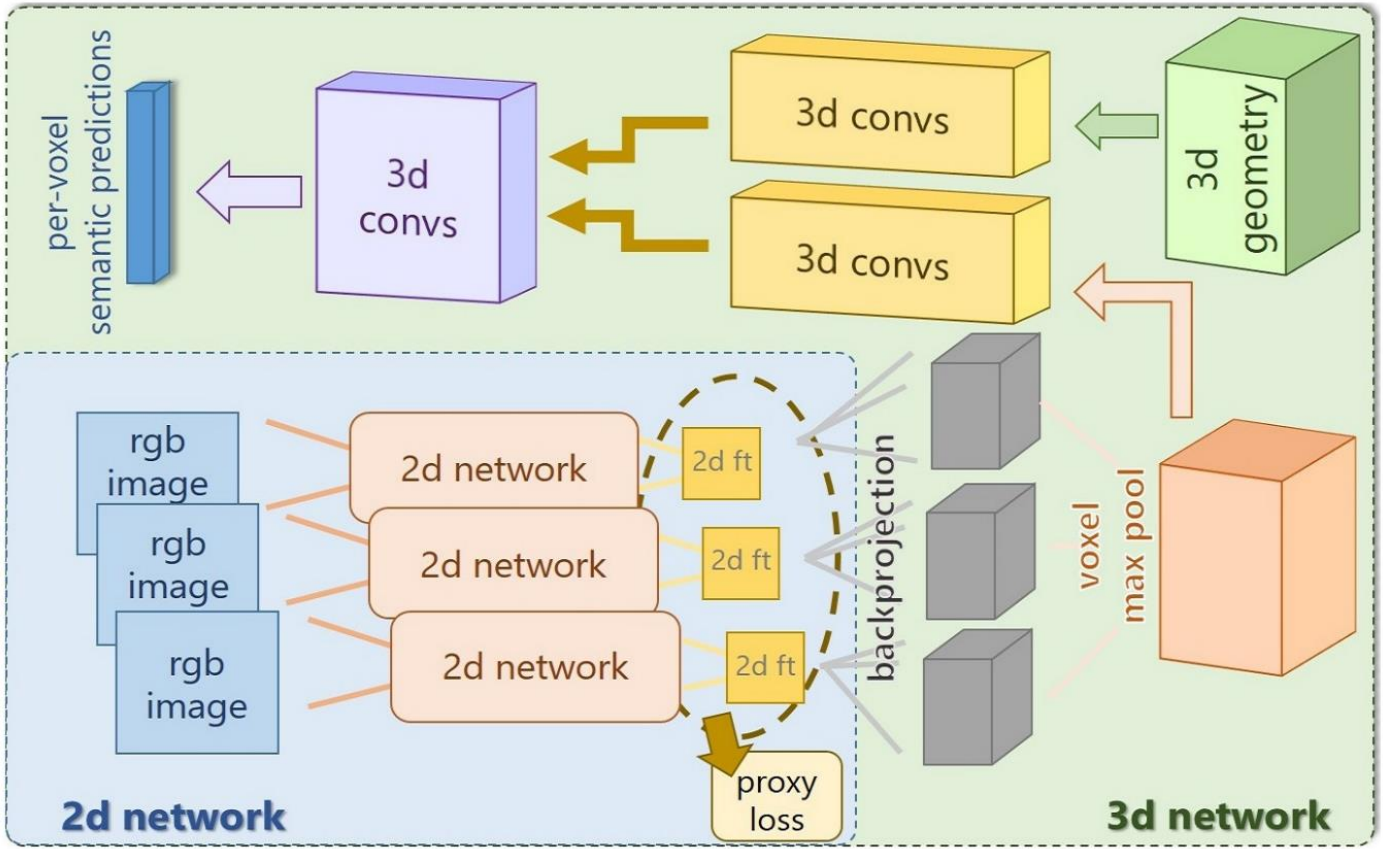
[Dai & Niessner 18] 3DMV

3D Volumetric + Multi-view



[Dai & Niessner 18] 3DMV

3D Volumetric + Multi-view



[Dai & Niessner 18] 3DMV

3D Volumetric + Multi-view

	avg class accuracy
color only	58.2
geometry only	54.4

[Dai & Niessner 18] 3DMV

3D Volumetric + Multi-view

	avg class accuracy
color only	58.2
geometry only	54.4
color+geometry	**75.0**

[Dai & Niessner 18] 3DMV

3D Volumetric + Multi-view

	avg class accuracy
geometry only	54.4
color+geometry (1 views)	70.1
color+geometry (3 views)	73.0
color+geometry (5 views)	**75.0**

[Dai & Niessner 18] 3DMV

3D Volumetric + Multi-view

	wall	floor	cab		bath	other	avg
2d only (1 view)	37.1	39.1	26.7		36.3	20.4	27.1
2d only (3 views)	58.6	62.5	40.8		61.5	34.3	44.2
Ours (no geo input)	76.2	92.9	59.3		80.8	9.3	58.2
Ours (3d geo only)	60.4	95.0	54.4		87.0	20.6	54.4
Ours (3d geo+voxel color)	58.8	94.7	55.5		85.4	20.5	55.9
Ours (1 view, fixed 2d)	77.3	96.8	**70.0**	...	87.0	58.5	69.1
Ours (1 view)	70.7	96.8	61.4		81.6	51.7	70.1
Ours (3 view, fixed 2d)	**81.1**	96.4	58.0		92.5	**60.7**	72.8
Ours (3 view)	75.2	**97.1**	66.4		89.9	57.2	73.0
Ours (5 view, fixed 2d)	77.3	95.7	68.9		93.5	59.6	74.5
Ours (5 view)	73.9	95.6	69.9		**94.7**	58.5	**75.0**

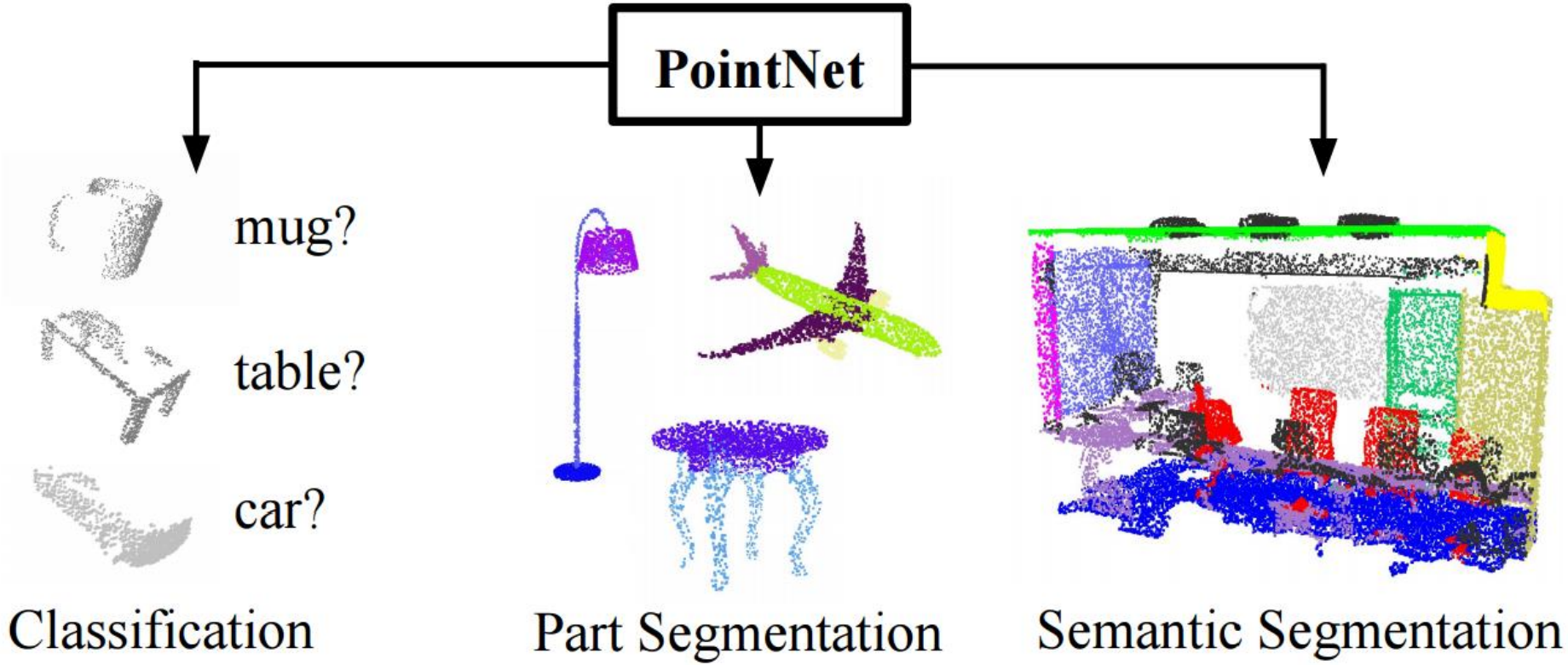
[Dai & Niessner 18] 3DMV

Conclusion so far

- Hybrid:
  - Nice way to combine color and geometry
  - Great performance (best so far for segmentation)
  - End-to-end helps less than we hoped for
  - Could be faster…

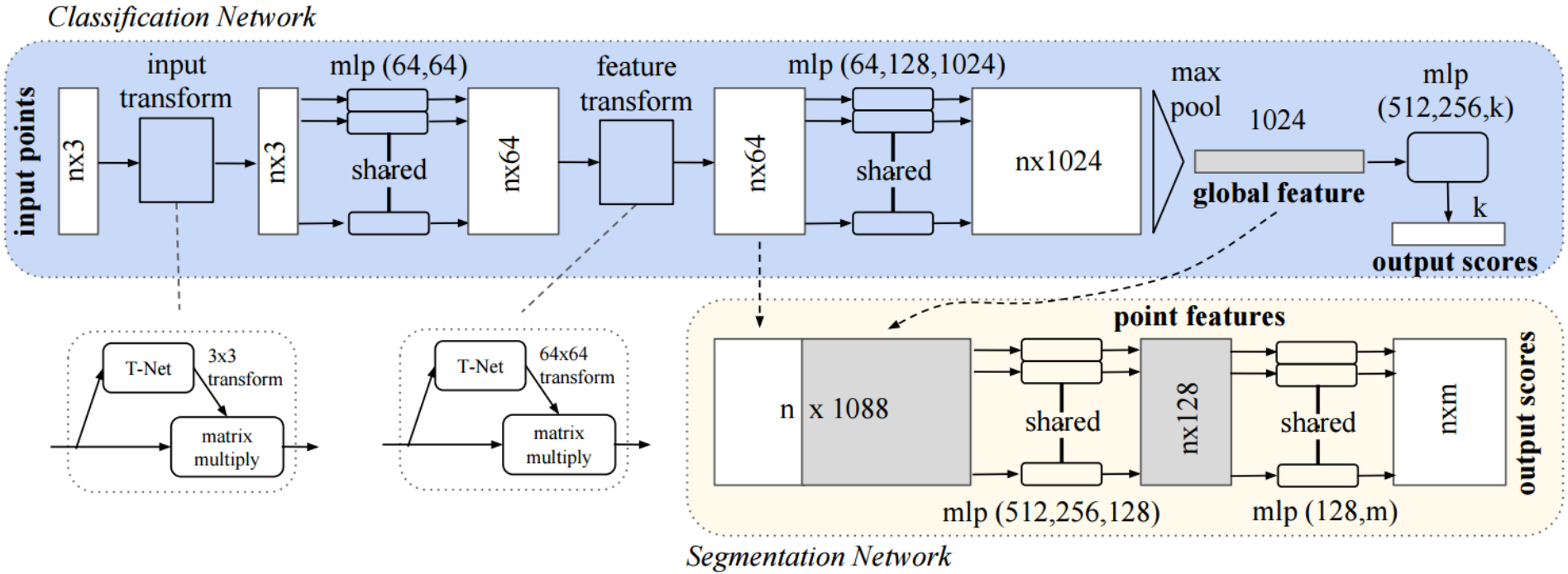
Point Clouds

Deep Learning on Point Clouds: PointNet



[Qi et al. 17] PointNet

Deep Learning on Point Clouds: PointNet



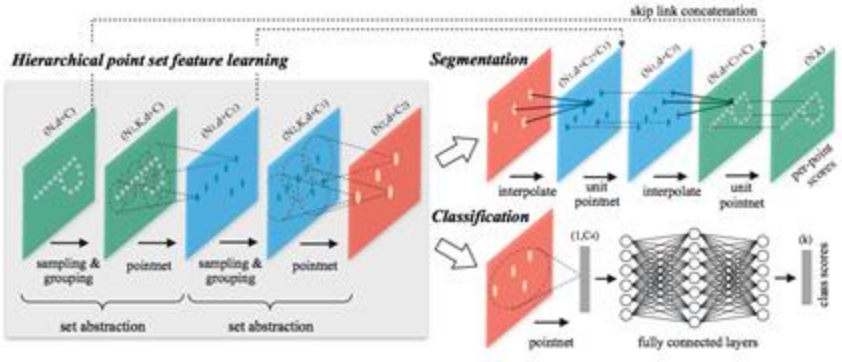
[Qi et al. 17] PointNet

PointNet++

Main idea

- Learn hierarchical representation of point cloud
- Apply multiple (simplified) PointNets at different locations and scales
- Each Scale: Furthest-Point Sampling -> Query Ball Grouping -> PointNet
- Multi-scale or Multi-resolution grouping for sampling density robustness

Evaluations: Classification, Part-Segmentation, Scene-Segmentation



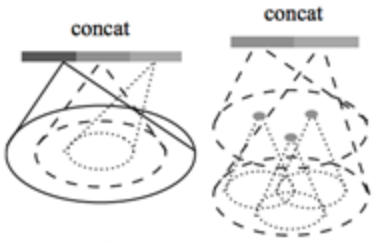


Figure 3: (a) Multi-scale grouping (MSG); (b) Multi-resolution grouping (MRG).

[Qi et al. 17] PointNet++

Point Convolutions

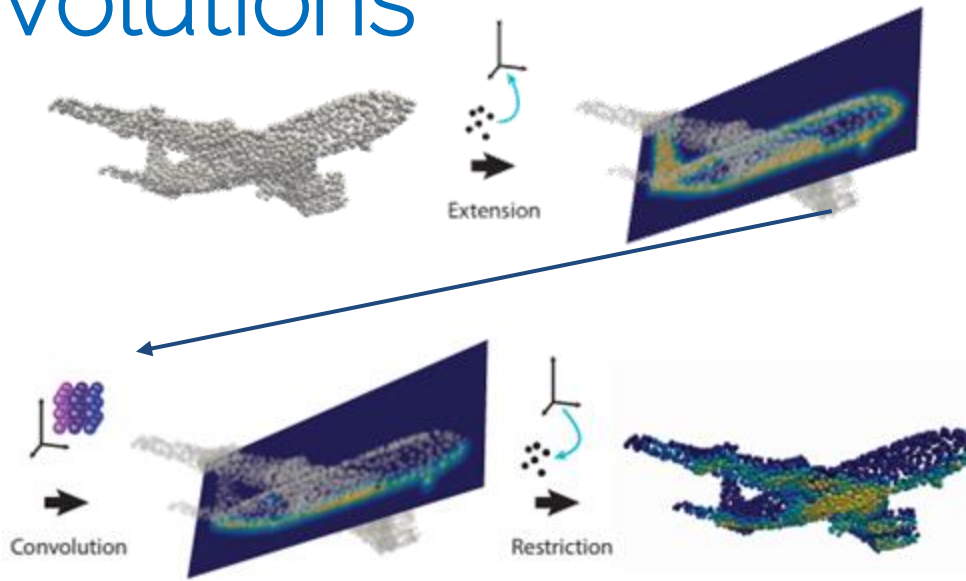
Main idea

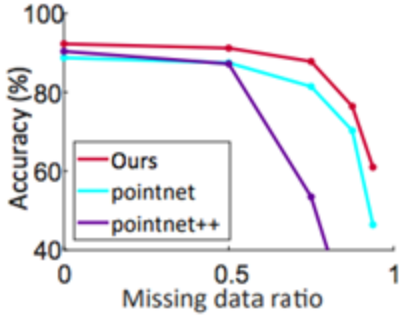
- Transform points to continuous R3 representation (RBFs)
- Convolve in R3
- Restrict results to points

Uses Gaussian RBF representation.

Boils down to computing fixed weights for convolution.

Don't use real data as far as I know!





Point Convolutional NN by Extension Operators
Matan Atzmon, Haggai Maron, Yaron Lipman (SIGGRAPH 2018)

Point Transformer

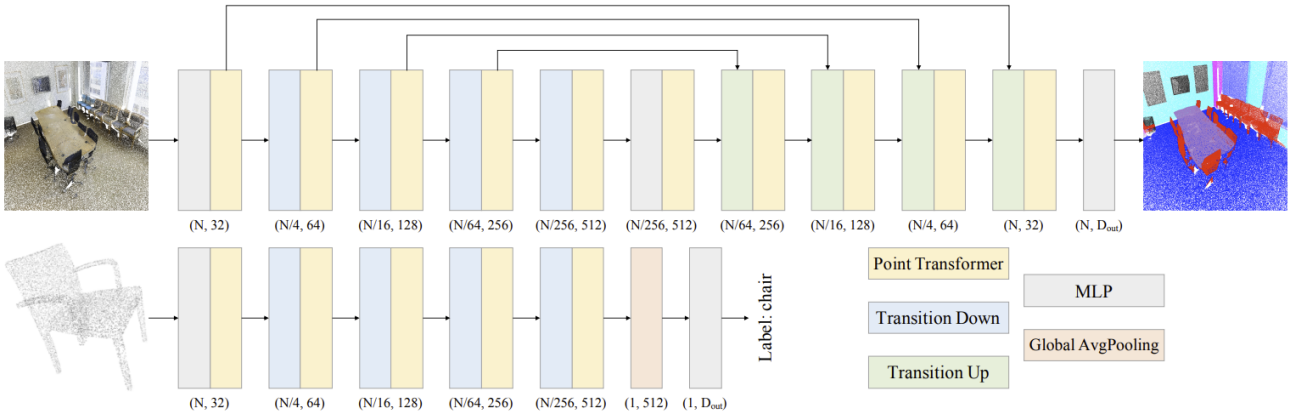
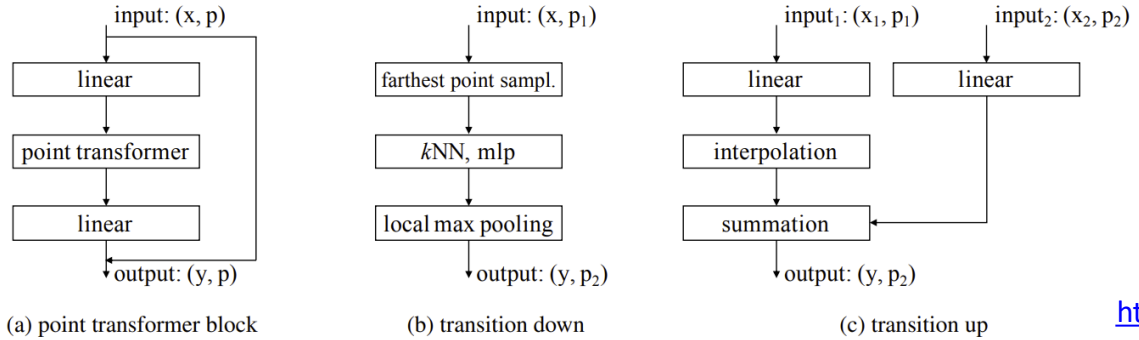


Figure 3. Point transformer networks for semantic segmentation (top) and classification (bottom).



(a) point transformer block (b) transition down (c) transition up

Figure 4. Detailed structure design for each module.

https://arxiv.org/pdf/2012.09164

[Zhao et al. 21] Point Transformer

Conclusions so far

- PointNet variants:
  - Train super fast (also testing)
  - Can cover large spaces in one shot
  - Cannot represent free space
  - Performance (mostly) worse than pure volumetric
  - Still lots of ongoing research!

Point Sets (local)

RBF

Point Convolutional NN by Extension Operators (SIGGRAPH 2018)
Tangent Convolutions for Dense Prediction in 3D (CVPR 2018)

Nearest point neighborhoods

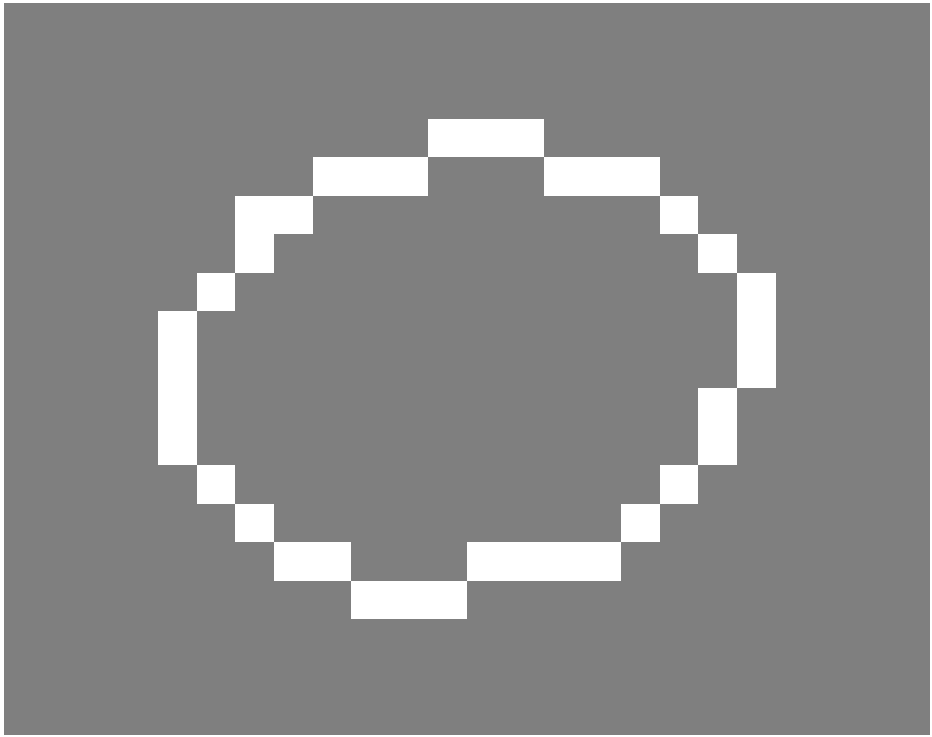
Dynamic edge-conditioned filters in convolutional neural networks on graphs (CVPR17)
3D Graph Neural Networks for RGBD Semantic Segmentation (ICCV17)
PPFNet: Global context aware local features for robust 3d point matching (CVPR18)
FeaStNet: Feature-Steered Graph Convolutions for 3D Shape Analysis (CVPR18)

Very interesting combination where convolutions are essentially over line segments in 3D, and where both locations and are being optimized
https://arxiv.org/abs/1605.06240
Idea is great, performance could be a bit better (probably hard to optimize)

Sparse Convolutions

Sparse Convolutional Networks

Regular, dense 3x3 Convolution
-> set of actives (non-zeros) grows rapidly
-> need a lot of memory
-> takes a long time for feature prop.

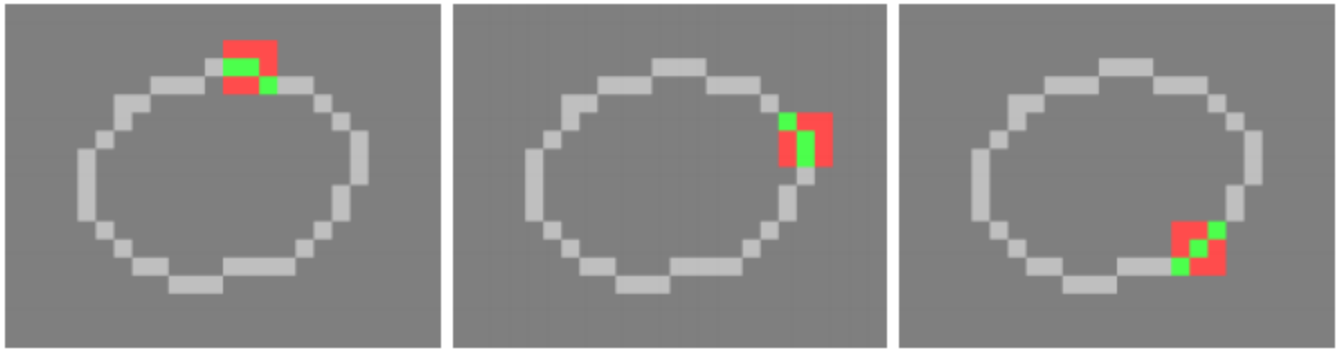


Submanifold Sparse Convolutional Networks, https://arxiv.org/abs/1706.01307 **[Graham et al. 18]**
3D Semantic Segmentation with Submanifold Sparse Convolutional Networks

Sparse Convolutional Networks

Dense

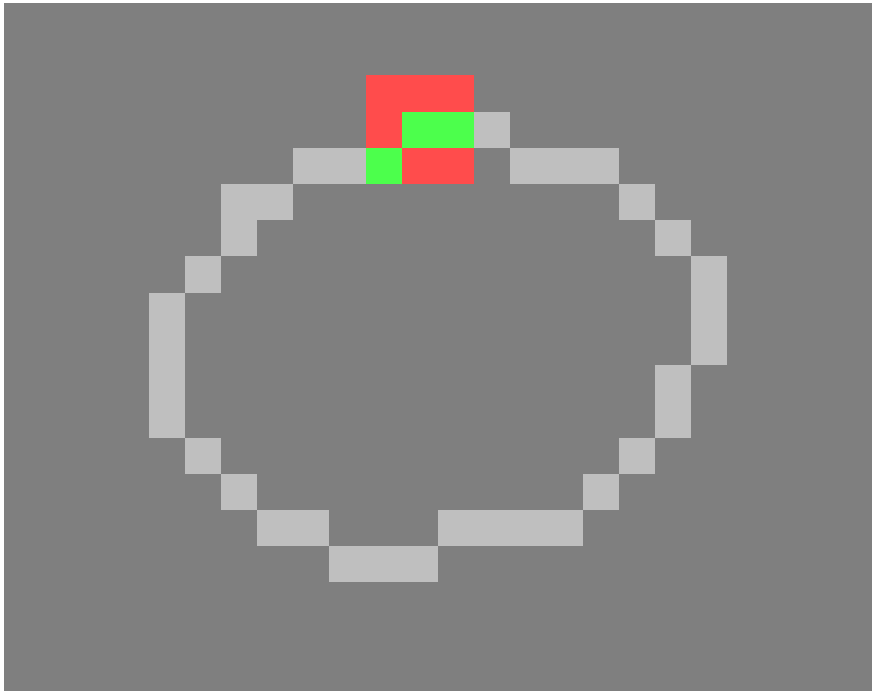
Sparse



Submanifold Sparse Convolutional Networks, https://arxiv.org/abs/1706.01307 **[Graham et al. 18]**
3D Semantic Segmentation with Submanifold Sparse Convolutional Networks

Sparse Convolutional Networks

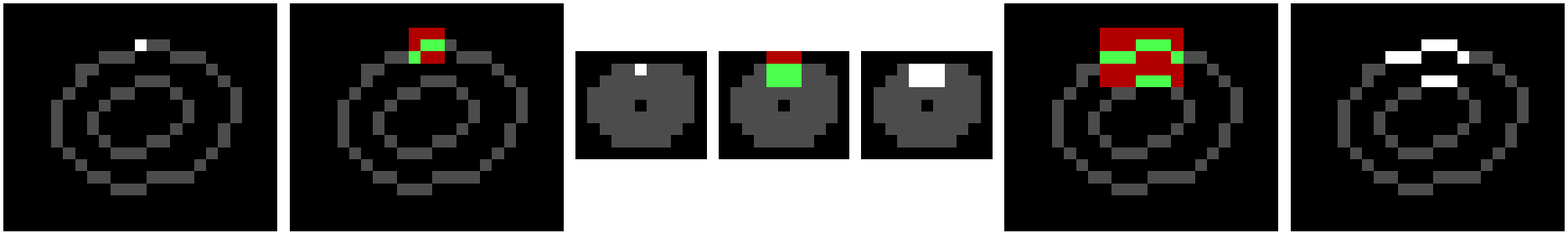
Submanifold Sparse Conv:
-> set of active sites is unchanged
-> active sites look at active neighbors (green)
-> non-active sites (red) have no overhead



Submanifold Sparse Convolutional Networks, https://arxiv.org/abs/1706.01307 **[Graham et al. 18]**
3D Semantic Segmentation with Submanifold Sparse Convolutional Networks

Sparse Convolutional Networks

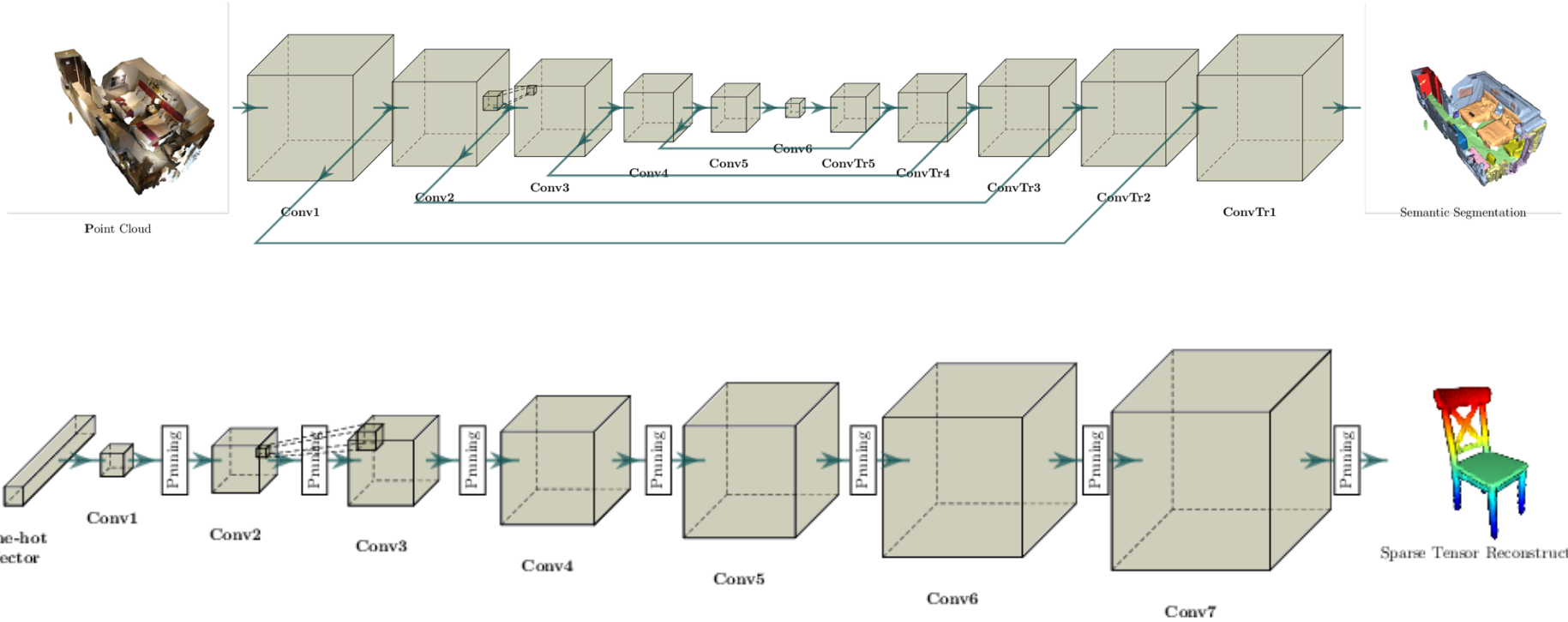
Submanifold Sparse Conv:
-> disconnected components do not communicate at first
-> although they will merge due to effect of stride, pooling, convs, etc.



from left: (i) an active point is highlighted; a convolution with stride 2 sees the green active sites (ii) and produces output (iii), 'children' of hightlighted active point from (i) are highlighted; a submanifold sparse convolution sees the green active sites (iv) and produces output (v); a deconvolution operation sees the green active sites (vi) and produces output (vii).

Submanifold Sparse Convolutional Networks, https://arxiv.org/abs/1706.01307 **[Graham et al. 18]**
3D Semantic Segmentation with Submanifold Sparse Convolutional Networks

Sparse Convolutional Networks



https://github.com/NVIDIA/MinkowskiEngine

Sparse Convolutional Networks

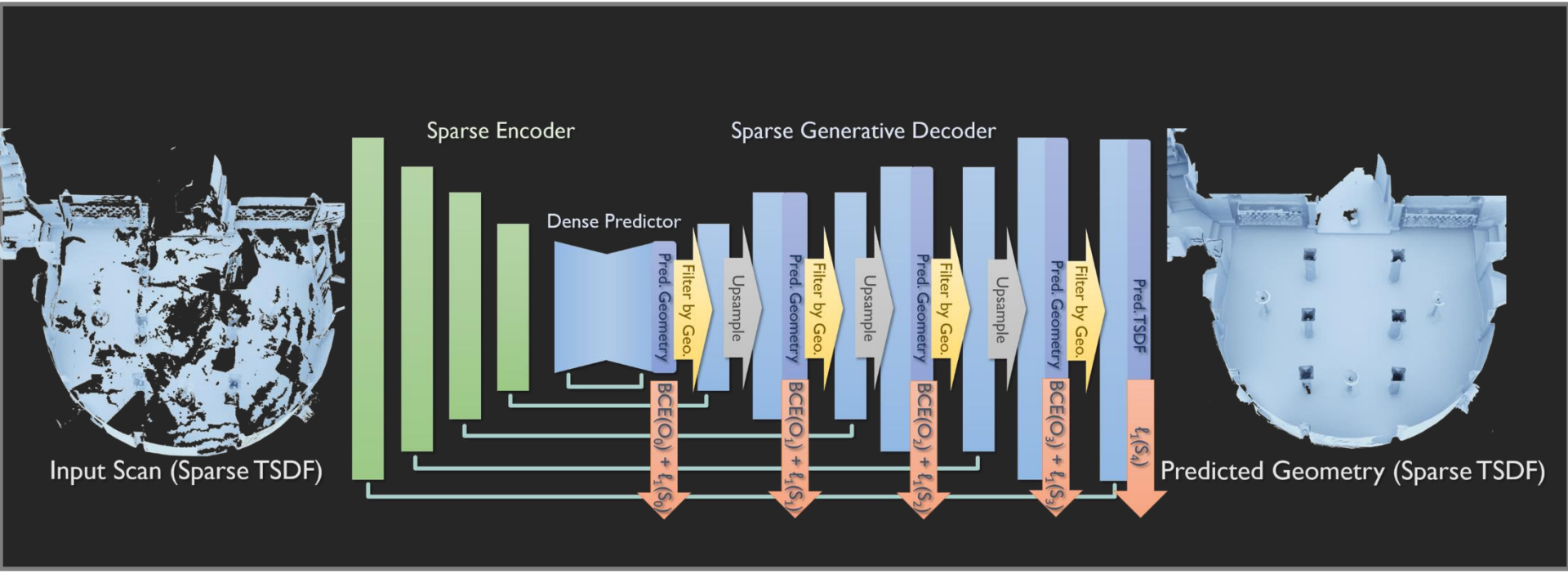
Dimension	Name in 'torch.nn'	Use cases
1	Conv1d	Text, audio
2	Conv2d	Lines in 2D space, e.g. handwriting
3	Conv3d	Lines and surfaces in 3D space or (2+1)D space-time
4	-	Lines, etc, in (3+1)D space-time

https://github.com/facebookresearch/SparseConvNet

Submanifold Sparse Convolutional Networks, https://arxiv.org/abs/1706.01307 **[Graham et al. 18]**
3D Semantic Segmentation with Submanifold Sparse Convolutional Networks

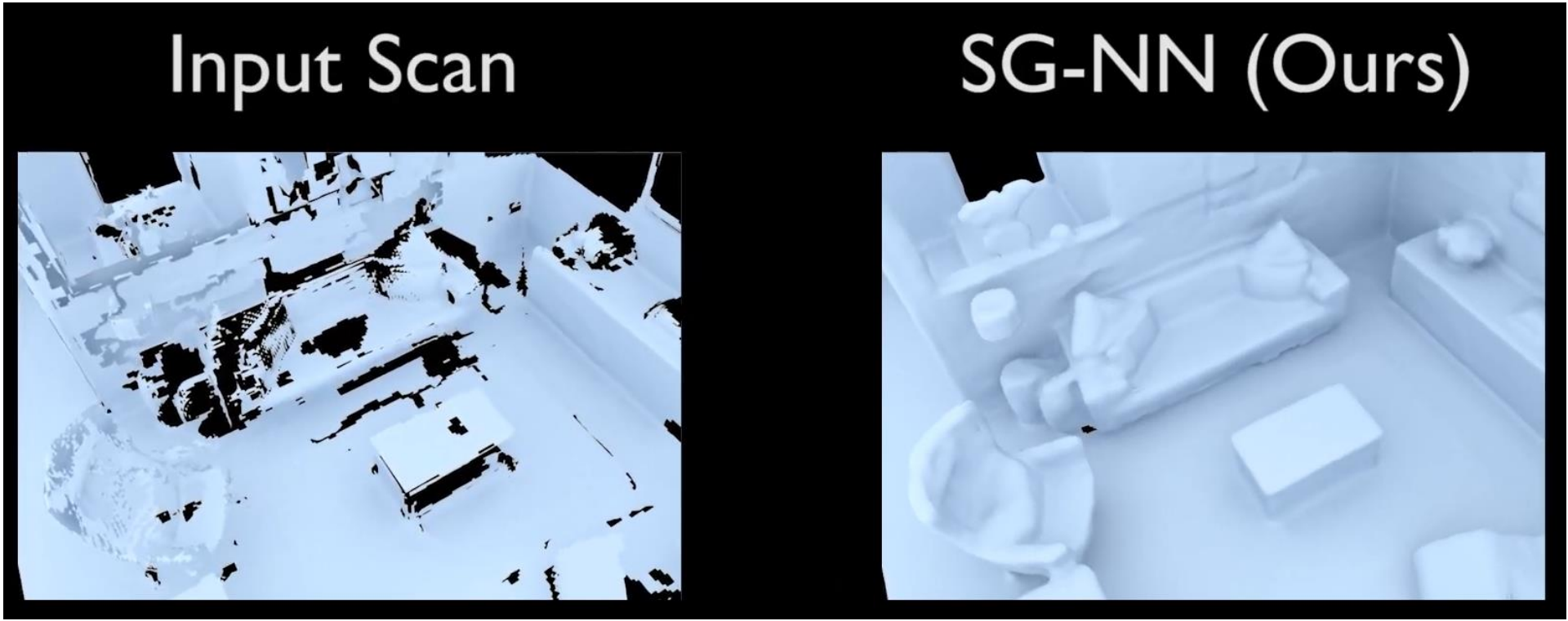
SparseConv Generators

SparseConv Generators



[Dai et al.]: SGNN (CVPR'20)

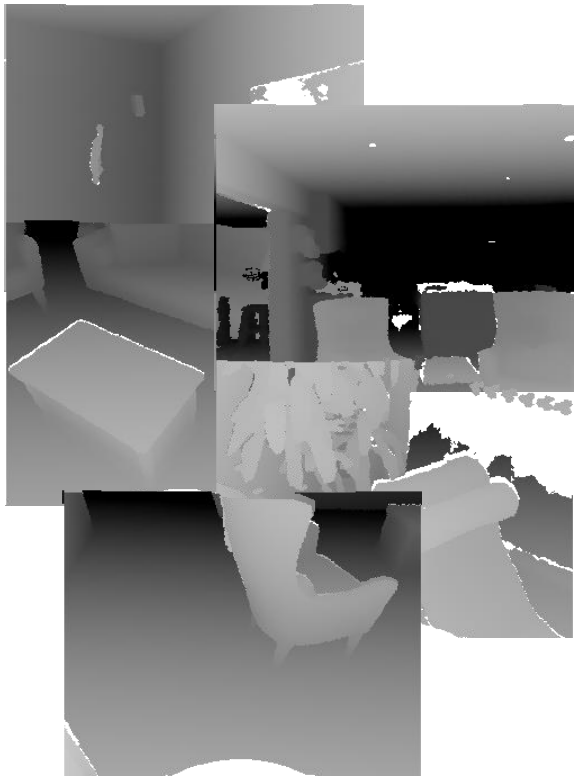
SG-NN: Sparse Convs + Self-Supervised



[Dai et al.]: SGNN (CVPR'20)

Self-Supervised Scan Completion

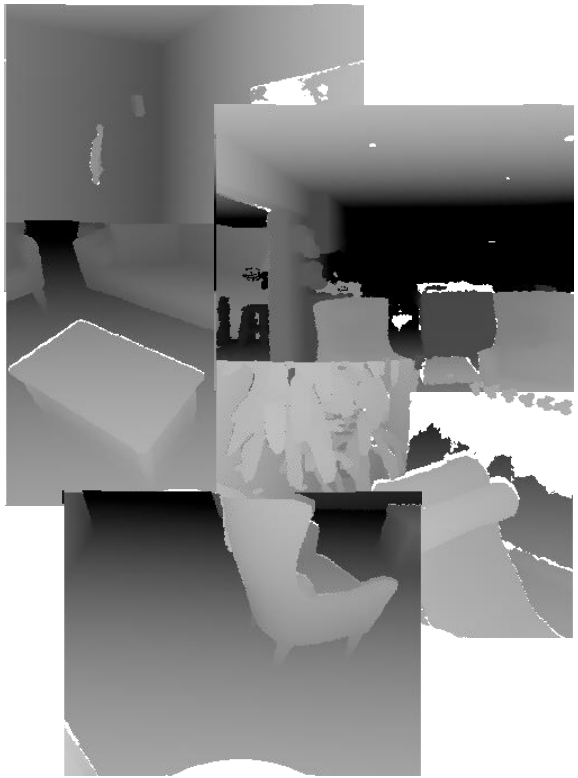
Depth Frames



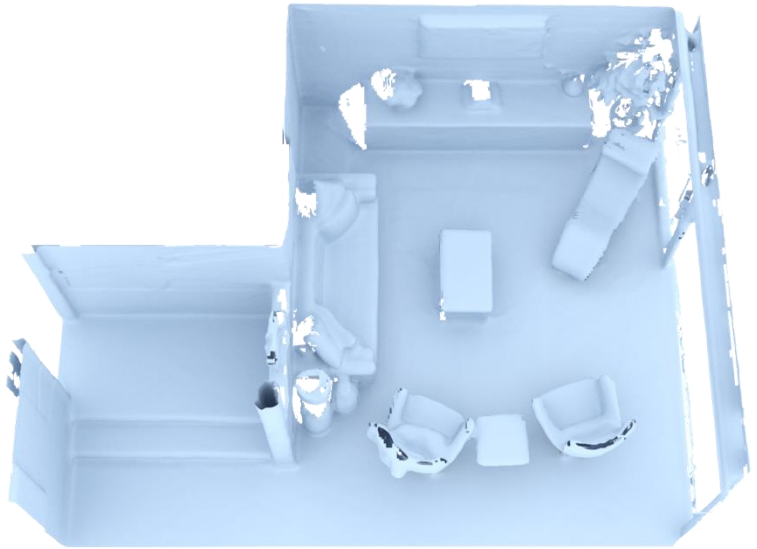
[Dai et al.]: SGNN (CVPR'20)

Self-Supervised Scan Completion

Depth Frames



Target Scan

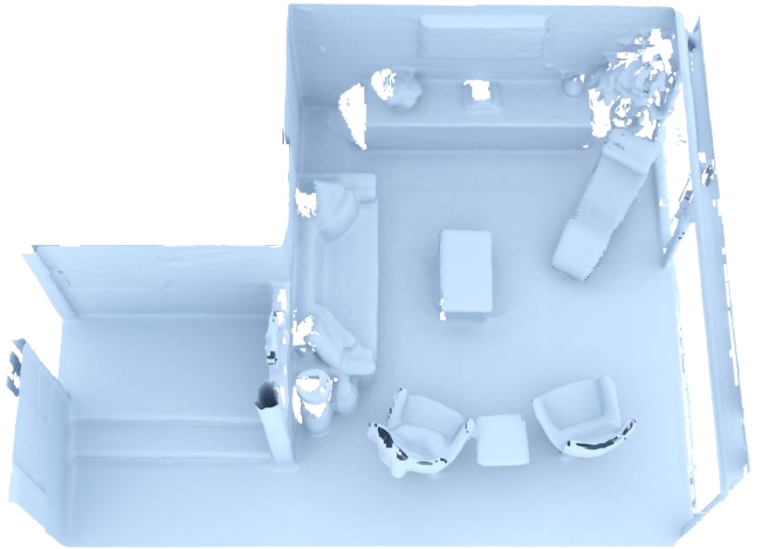


[Dai et al.]: SGNN (CVPR'20)

Self-Supervised Scan Completion

Depth Frames

Target Scan



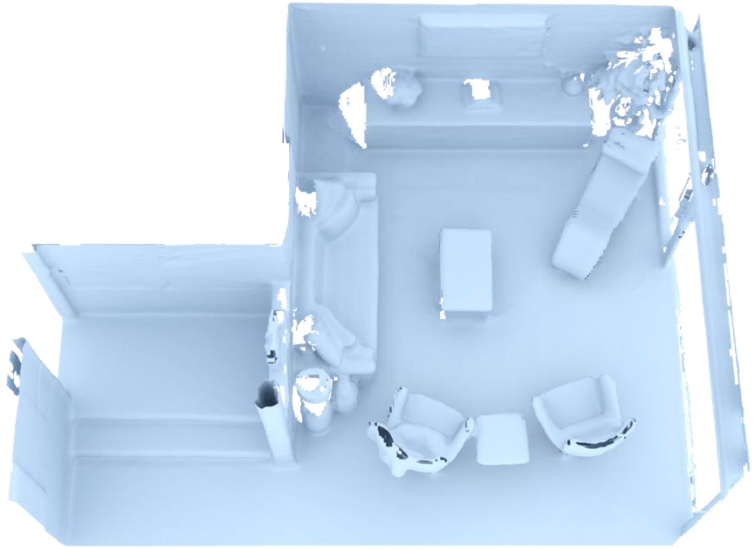
[Dai et al.]: SGNN (CVPR'20)

Self-Supervised Scan Completion

Depth Frames

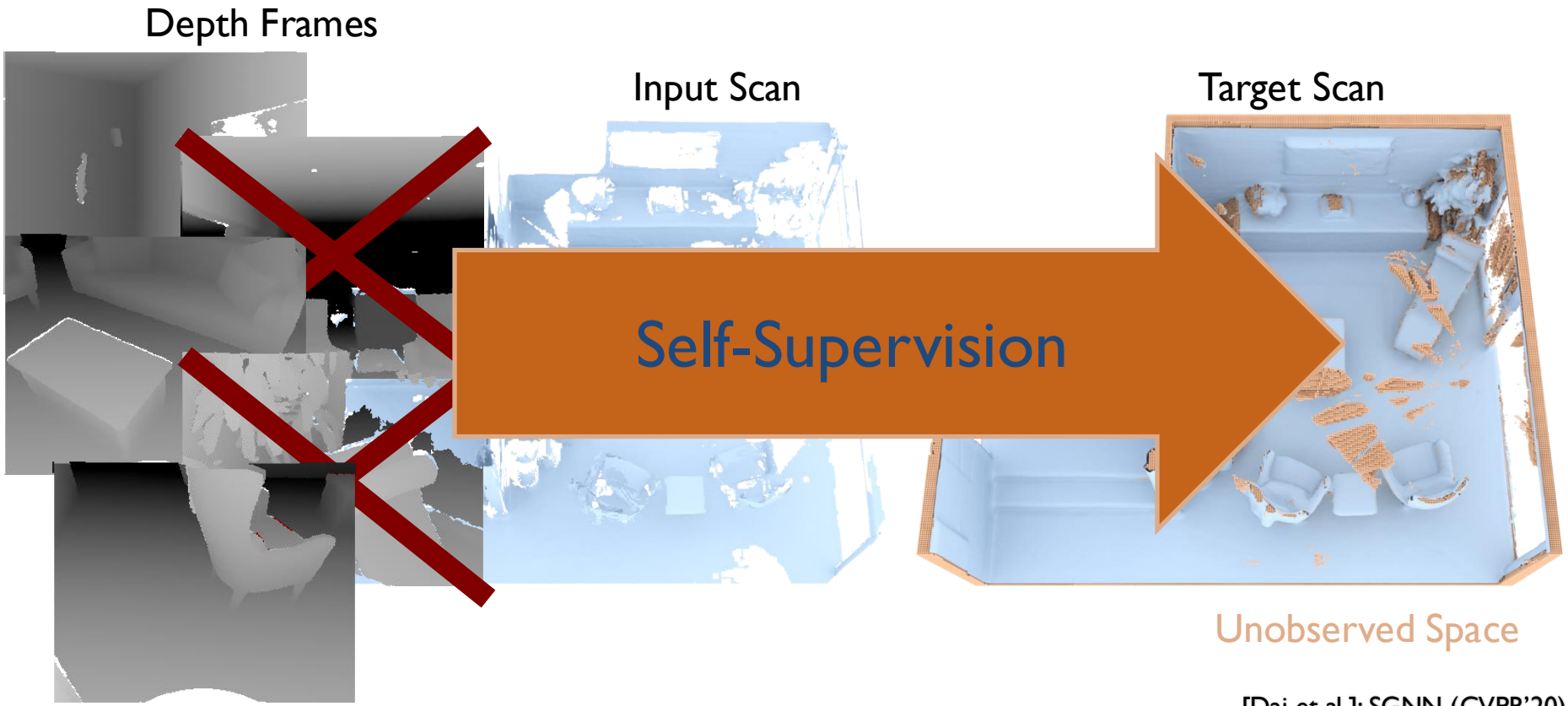
Input Scan

Target Scan



[Dai et al.]: SGNN (CVPR'20)

Self-Supervised Scan Completion

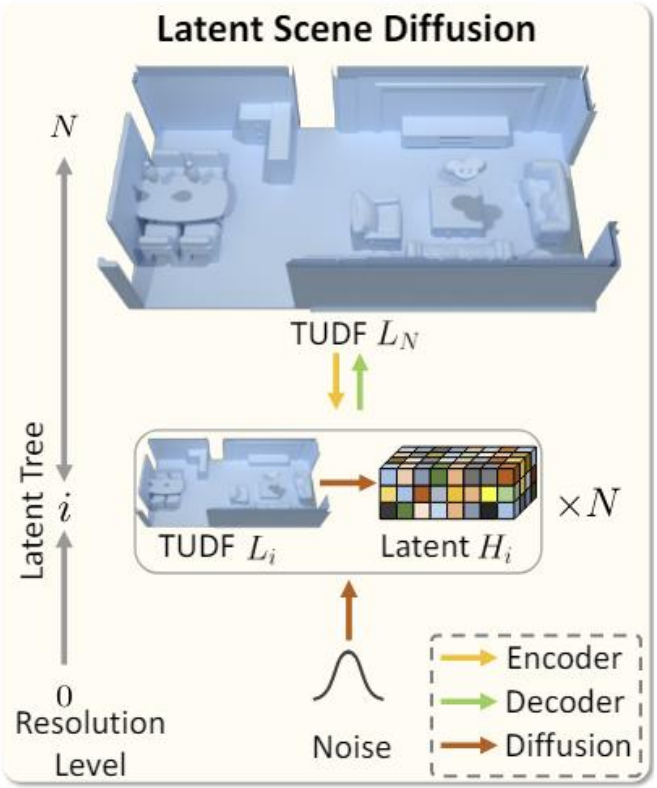


[Dai et al.]: SGNN (CVPR'20)

Latent 3D Scene Diffusion

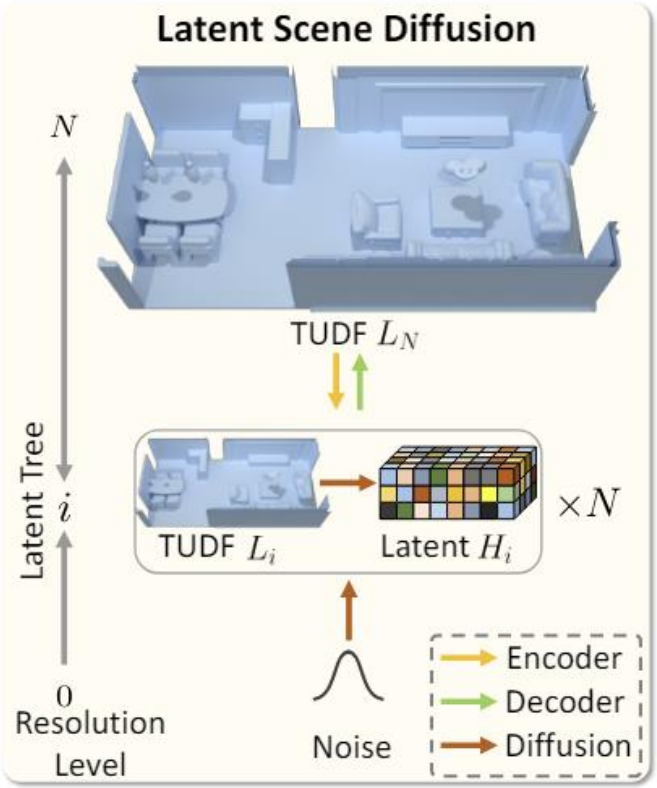
[Meng et al. 24]: Latent 3D Scene Diffusion

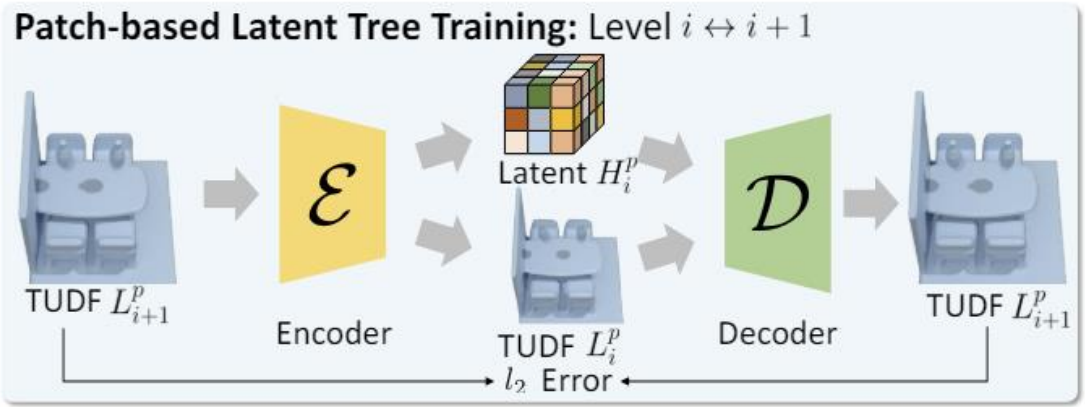
Latent 3D Scene Diffusion

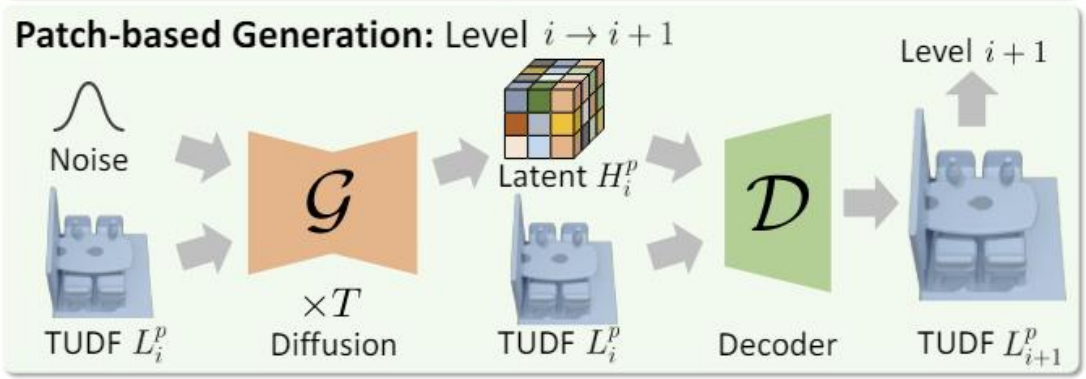


[Meng et al. 24]: Latent 3D Scene Diffusion

Latent 3D Scene Diffusion

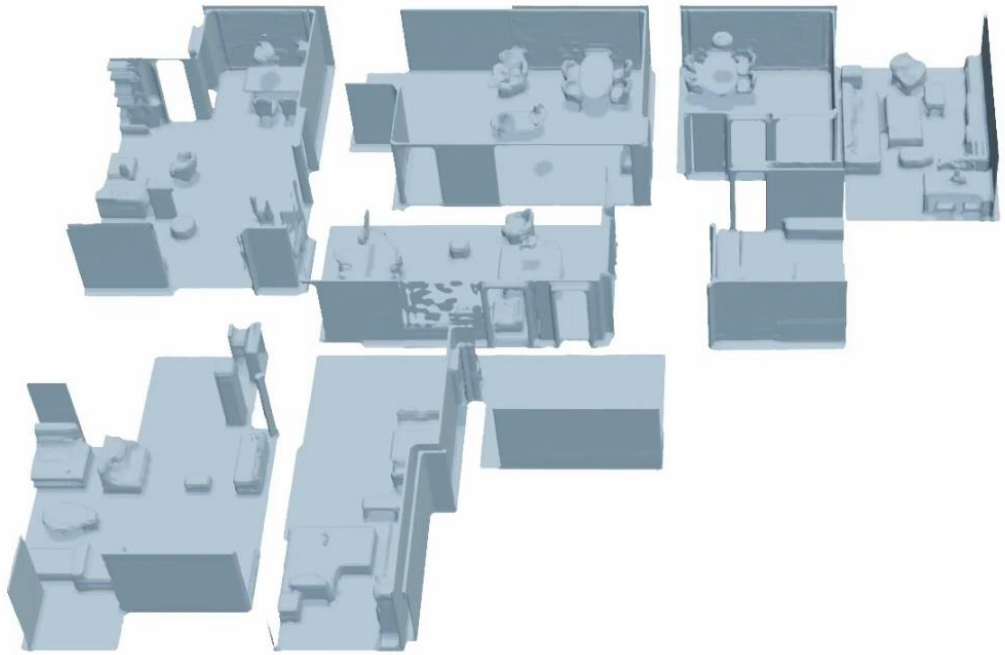






[Meng et al. 24]: Latent 3D Scene Diffusion

Latent 3D Scene Diffusion



[Meng et al. 24]: Latent 3D Scene Diffusion

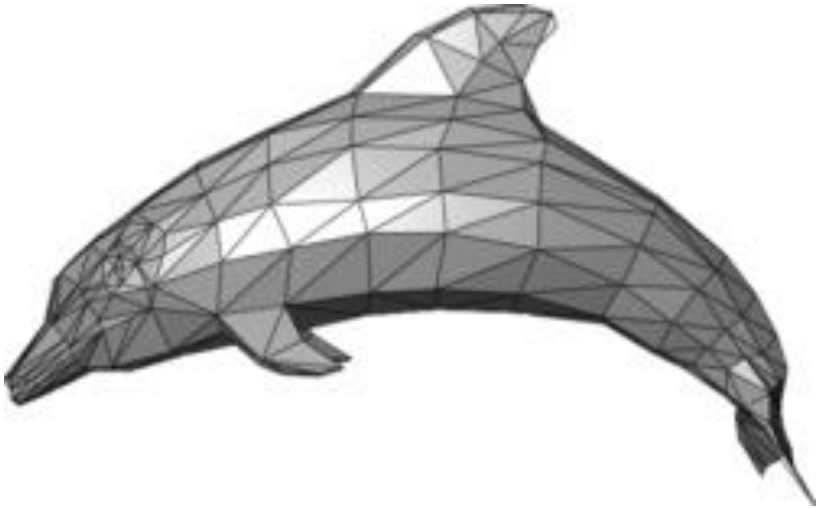
Conclusions so far

- Spares (volumetric) Convs:
  - Implemented with spatial hash function
  - Features only around “surface”
  - Require significantly less memory
  - Allow for much higher resolutions
  - It’s slower, but much higher accuracy

3D Meshes

3DMeshes

- Collection of vertices, edges, and faces
- Defines surface geometry



- Interpret meshes as graphs
- How to define a learnable operators / convs over a graph

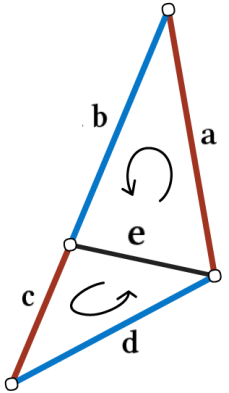
3DMeshes

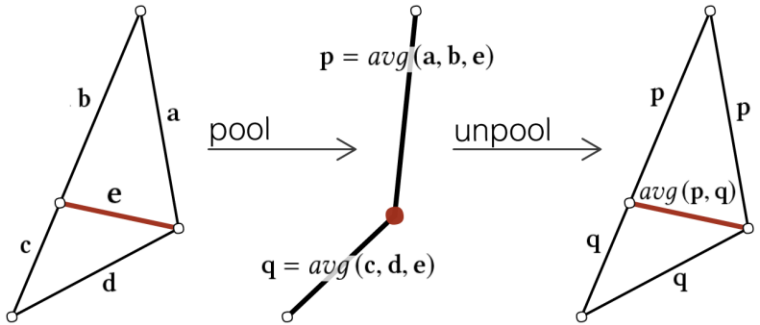
- Graph Networks
  - Message Passing
  - Graph Convolutions
  - Transformers?

- MeshCNN, Spectral Graph Convolutions, Geodesic CNNs

3D Meshes: MeshCNN

- A CNN for triangle meshes
- Define convolution and pooling operators on mesh edges
- Each edge has a feature
- Each edge has four edge neighbors (from 2 incident faces)
- Convolution filters applied to ea feature and its 4 neighbors
- Pooling by edge-collapse



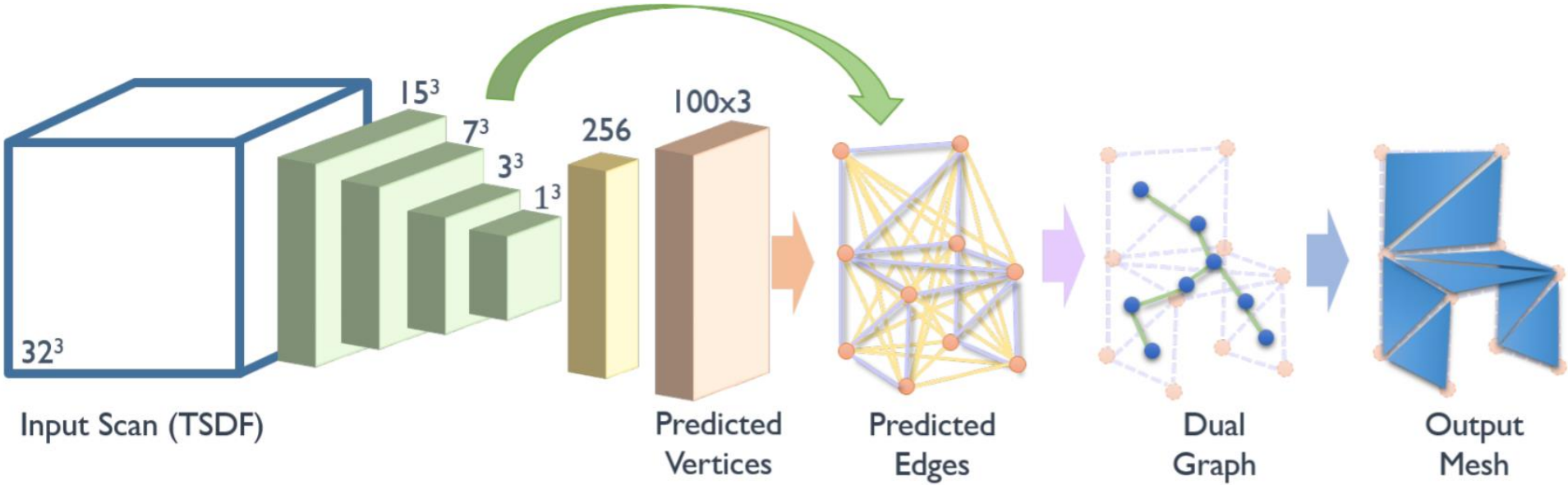


MeshCNN [Hanocka et al. '19]

Scan2Mesh: From Unstructured Range Scans to 3D Meshes

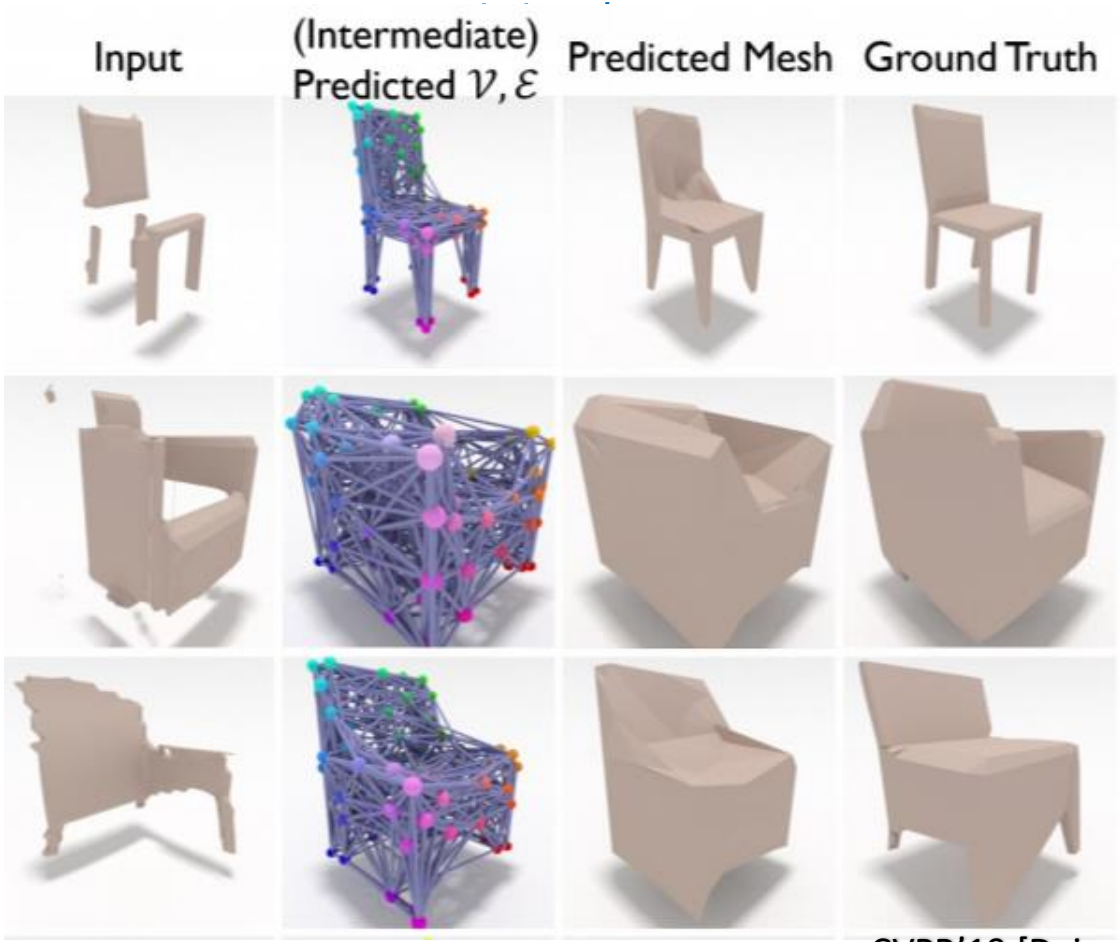
CVPR'19 [Dai and Niessner]: Scan2Mesh

Scan2Mesh: From Unstructured Range Scans to 3D Meshes



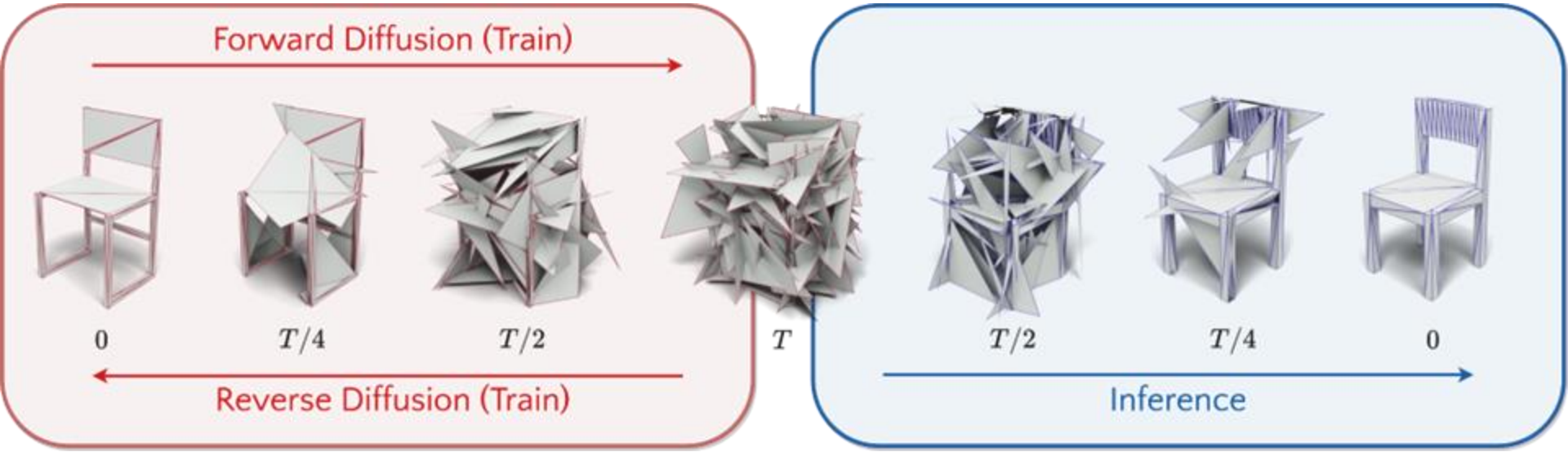
CVPR'19 [Dai and Niessner]: Scan2Mesh

Scan2Mesh: From Unstructured Range Scans to 3D



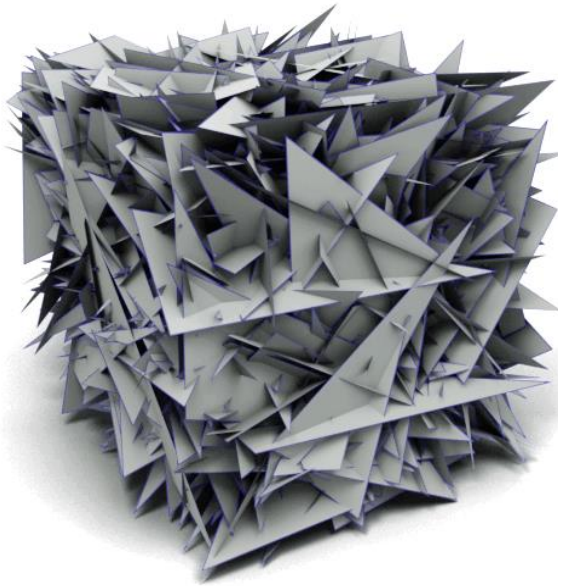
CVPR'19 [Dai and Niessner]: Scan2Mesh

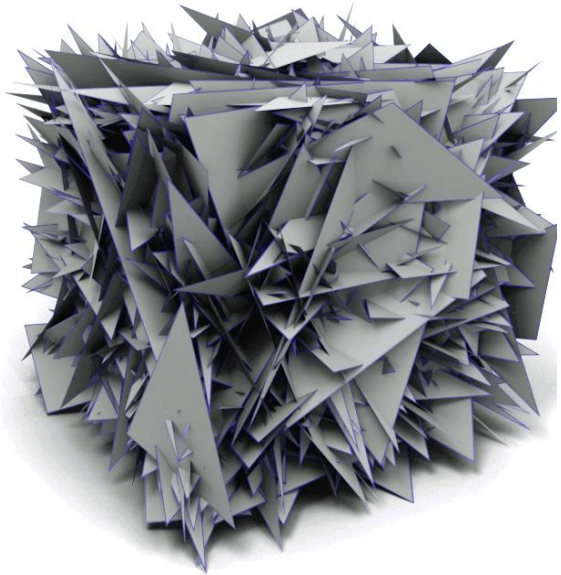
PolyDiff: Generating 3D Meshes with Diffusion

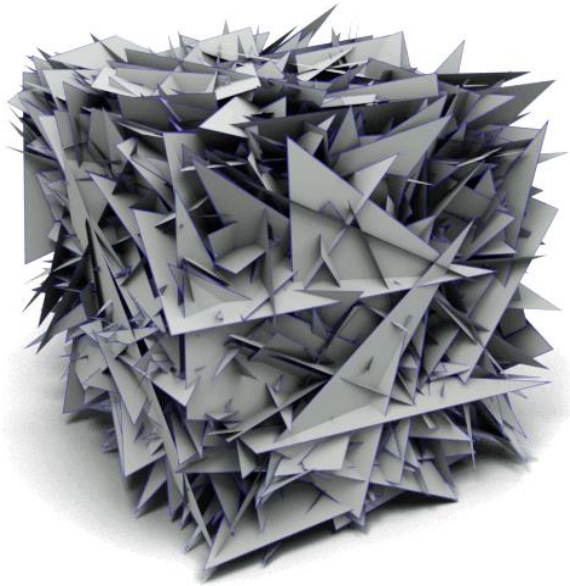


arXiv'23 [Alliegro et al.]: PolyDiff

PolyDiff: Generating 3D Meshes with Diffusion

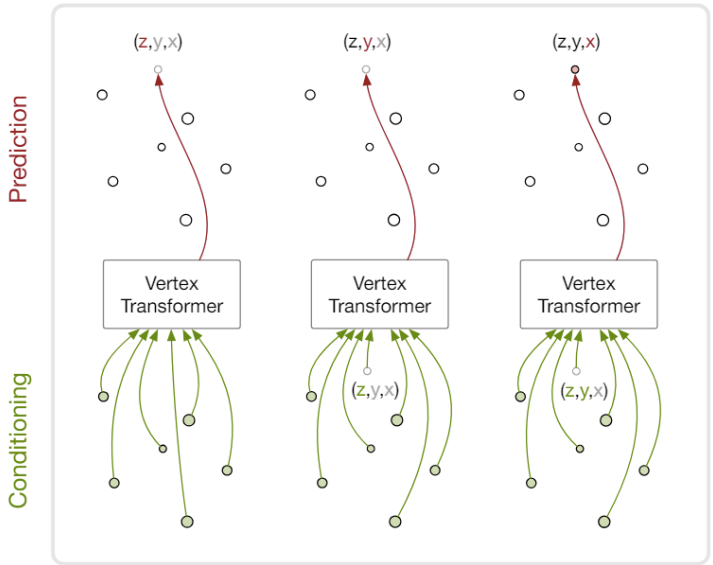




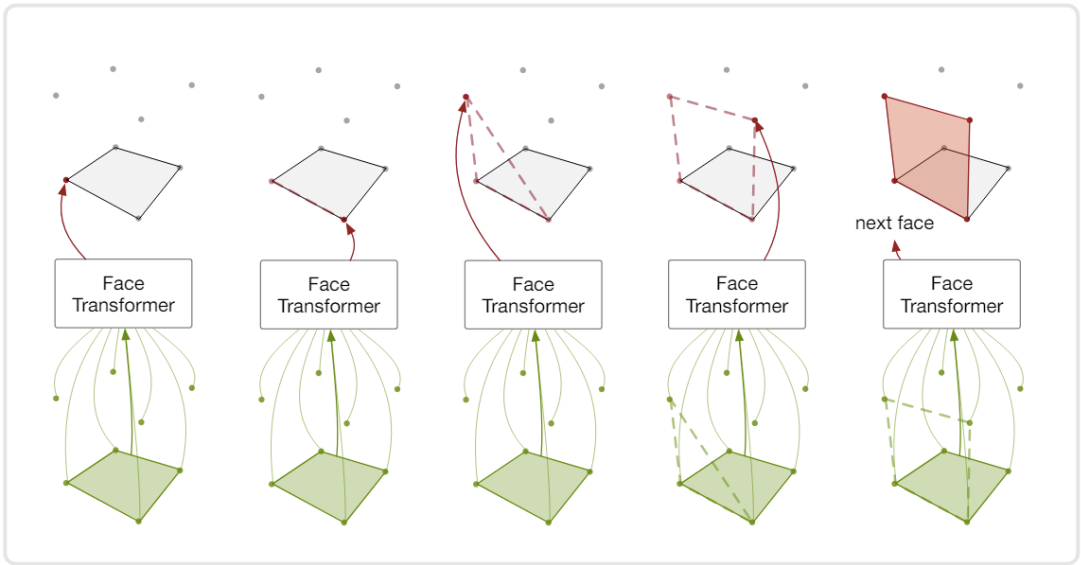


arXiv'23 [Alliegro et al.]: PolyDiff

Mesh Generation with Transformers



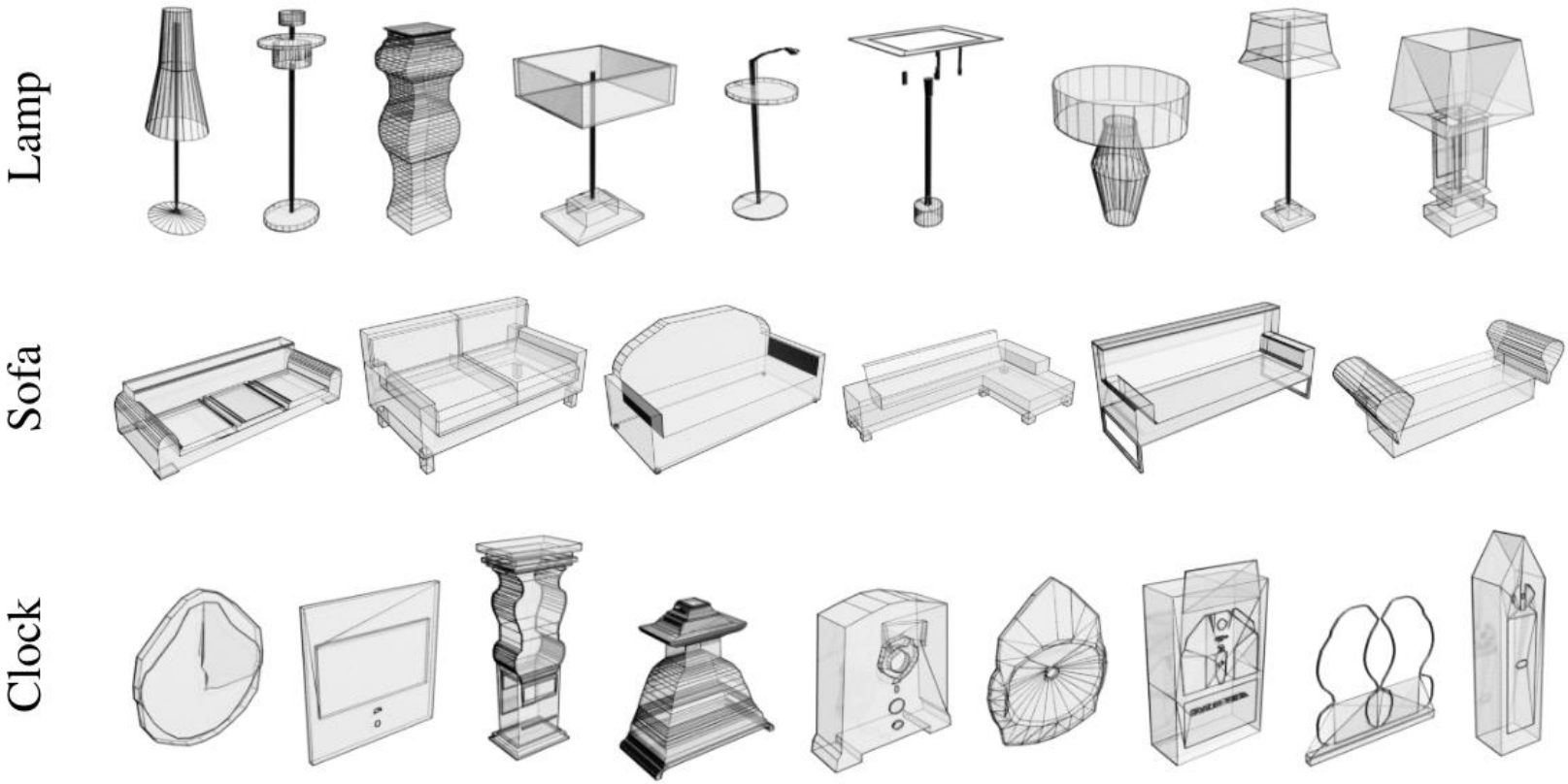
Vertex Model



Face Model

arXiv'20 [Nash et al.]: PolyGen: An Autoregressive Generative Model of 3D Meshes

Mesh Generation with Transformers



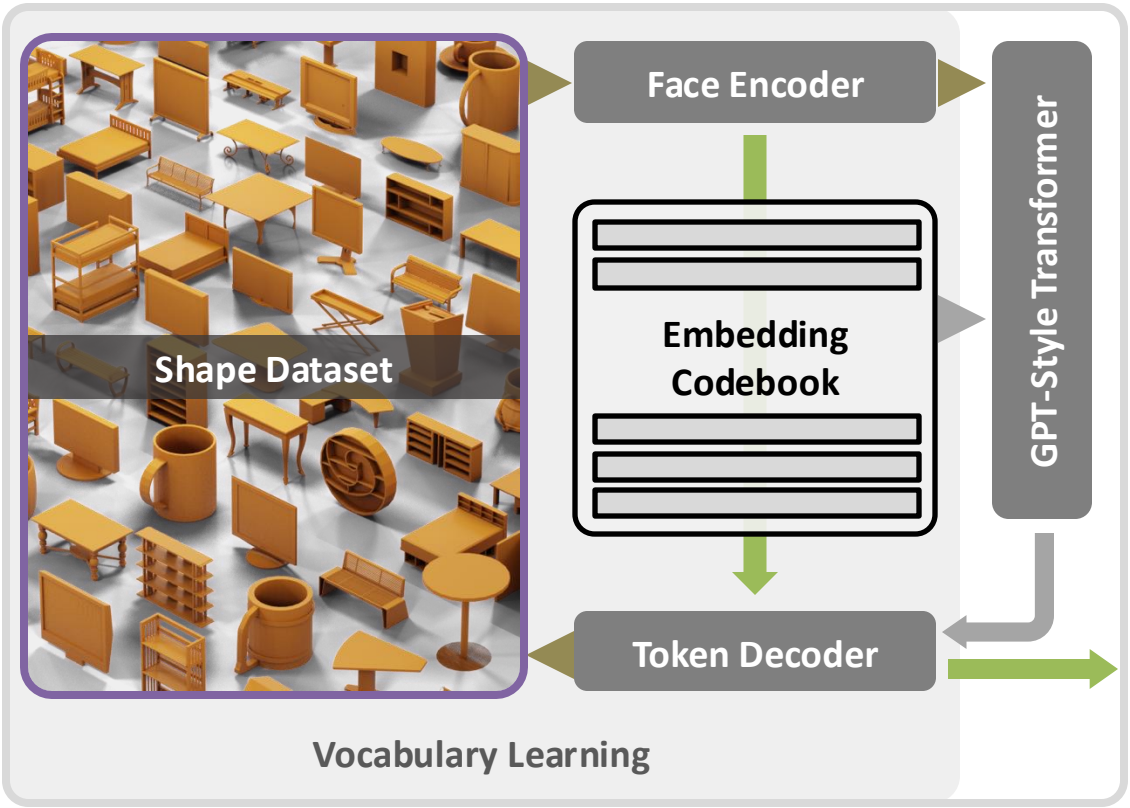
arXiv'20 [Nash et al.]: PolyGen: An Autoregressive Generative Model of 3D Meshes

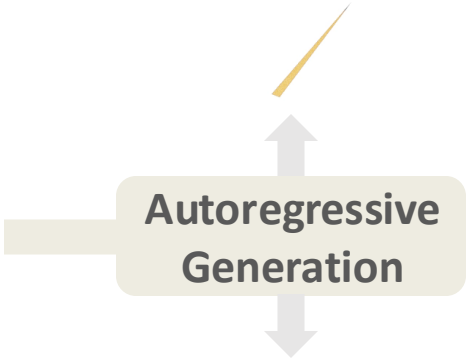
GPTs for Mesh Generation

MeshGPT

CVPR'24 [Siddiqui et al.]: MeshGPT

GPTs for Mesh Generation





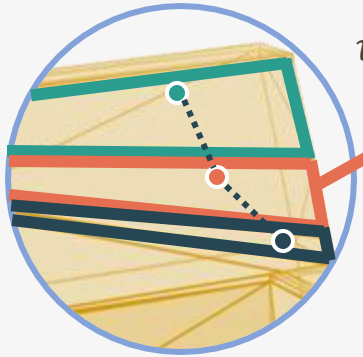
CVPR'24 [Siddiqui et al.]: MeshGPT

GPTs for Mesh Generation

Input Mesh

Face Graph + Input Features

Graph Convolutional Encoder



v_a v_c f v_b

Features on the faces

Feature Codebook

Residual Quantization

v_a v_c f' v_b

Face Embedding

CVPR'24 [Siddiqui et al.]: MeshGPT

MeshGPT: Inferencing the Transformer

Face Graph + Input Features

Graph Convolutional Encoder

Residual Quantization Module

Stacked Embeddings

Sequence & Flatten

< SOS >

< EOS >

GPT-Style Transformer

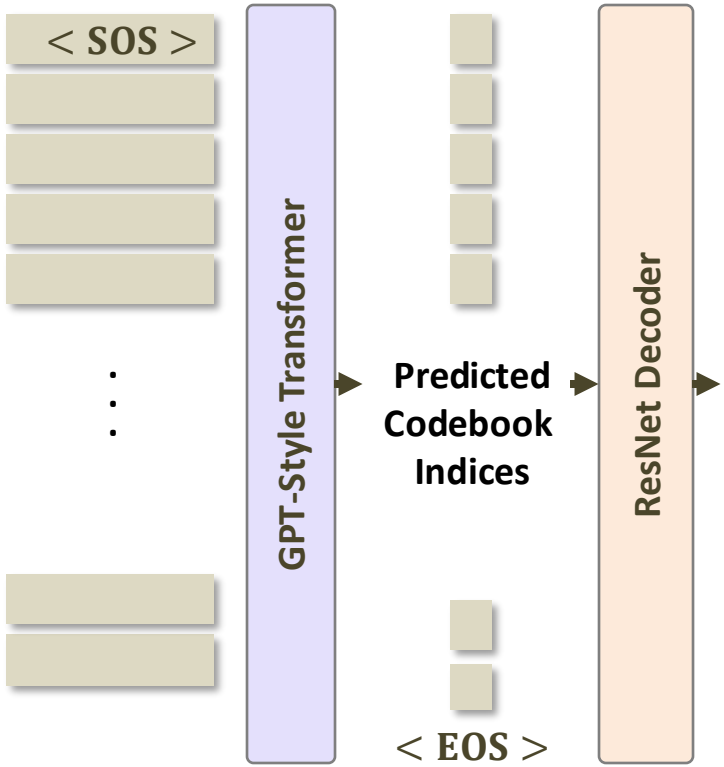
Predicted Codebook Indices

GT Codebook Indices

CE Loss

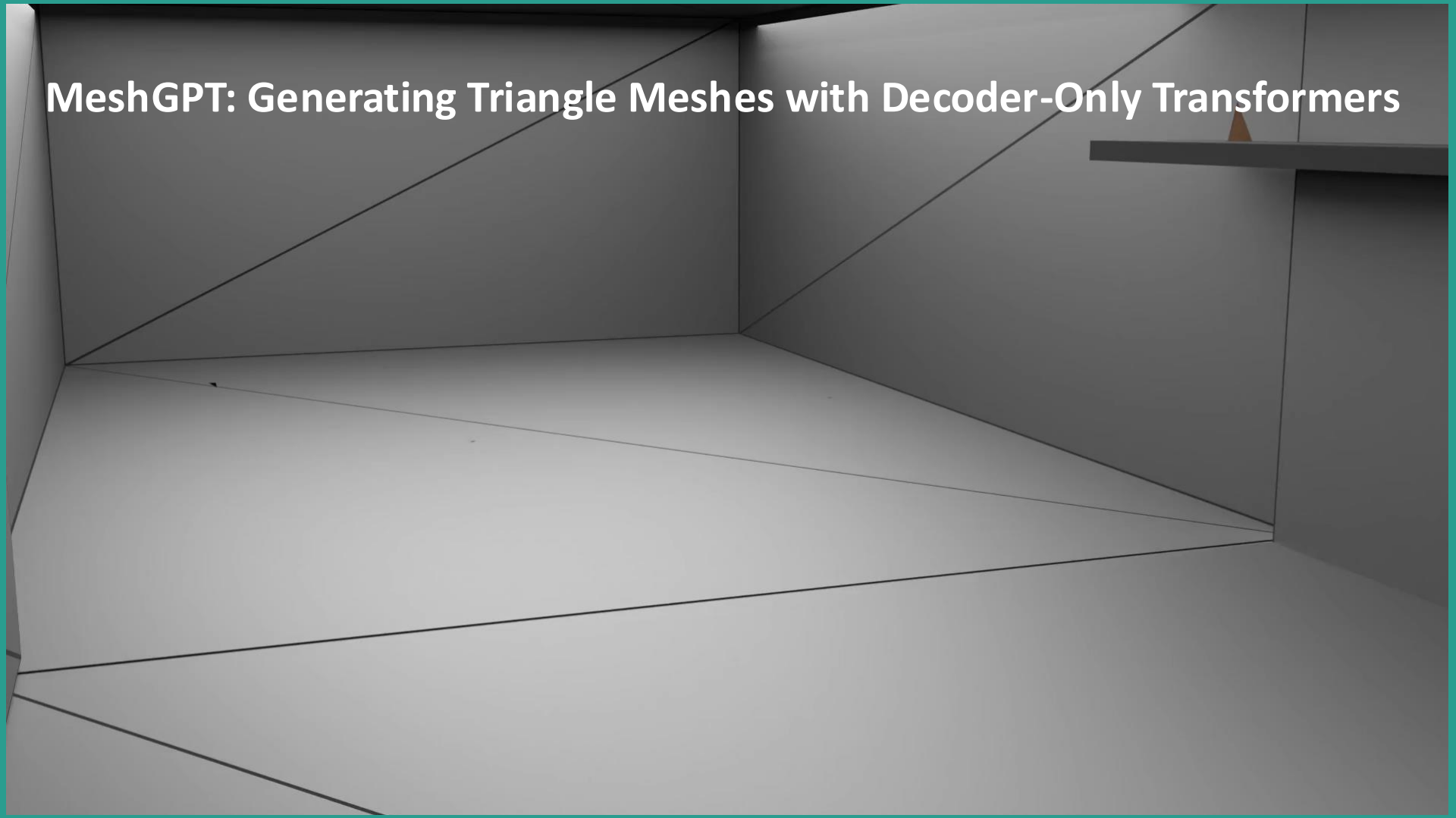
CVPR'24 [Siddiqui et al.]: MeshGPT

MeshGPT: Inferencing the Transformer



Generated Shape

CVPR'24 [Siddiqui et al.]: MeshGPT


MeshGPT: Generating Triangle Meshes with Decoder-Only Transformers

Generative 3D Mesh Approaches

- Scan2Mesh (Graph Network)
- PolyGen (Diffusion Model)
- PolyDiff (Transformer + PointerNetwork)
- MeshGPT (GPT-style Transformer)

3D Datasets

(mostly indoor)

3D Shapes (Synthetic)

- ShapeNet
  - Main dataset, everyone uses it
  - 55 classes (51.3k shapes); relatively uniform distribution but also mostly chairs; mediocre textures

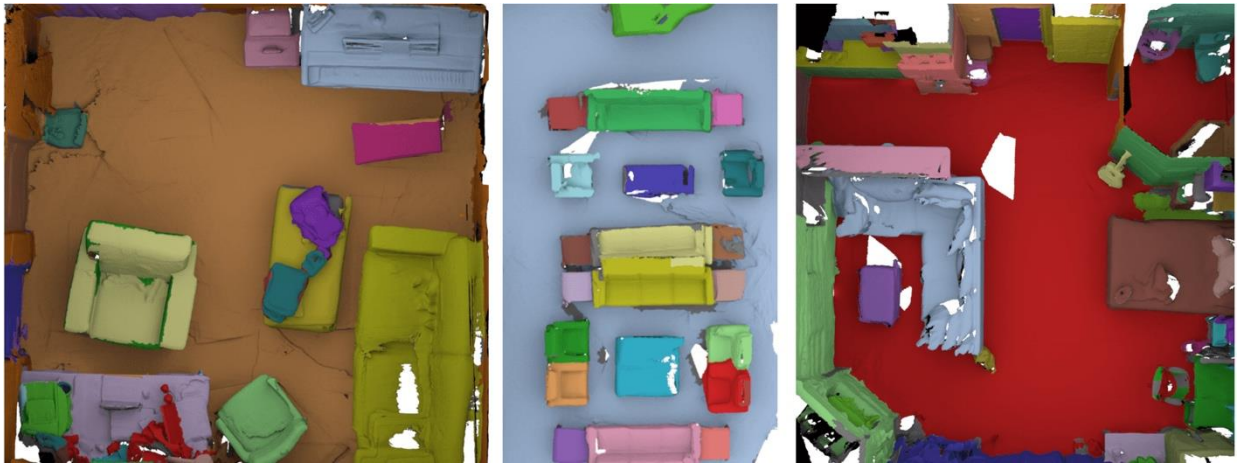
- Objaverse (links only)
  - 800k 3D shapes
- Objaverse-XL (links only)
  - 10mio 3D shapes (pretty heterogeneous)

3D Scenes (Synthetic)

- 3D-FRONT: 3D Furnished Rooms with Layouts and Semantics
  - 18,797 rooms (mostly procedural geometry)
  - 7,302 furniture objects

3D Scenes

- ScanNet
  - Kinect-style reconstructions; 1500 scenes; 2.5 mio views; semantic + instance annotations



3D Scenes

Evaluation on Hidden Test Set on ScanNet

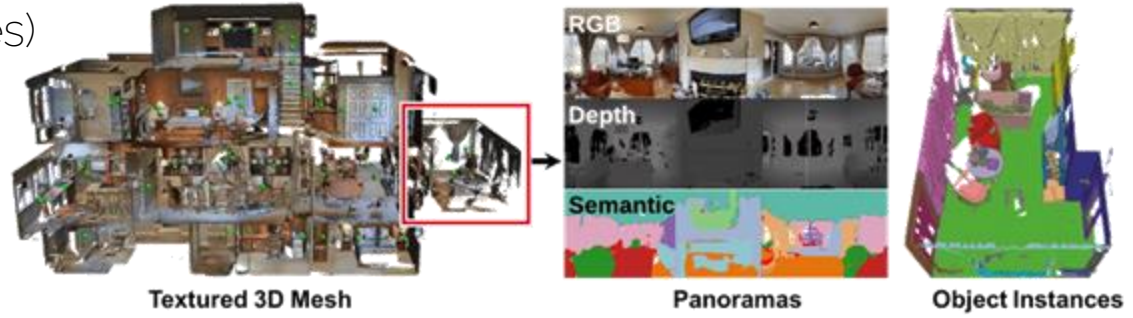
Method	Info	avg iou	bathtub	bed	bookshelf	cabinet	chair	counter	curtain	desk	door	floor	otherfurniture	picture	refrigerator	shower curtain
PTv3 ScanNet		0.794 1	0.941 3	0.813 17	0.851 7	0.782 5	0.890 2	0.597 1	0.916 2	0.696 7	0.713 3	0.979 1	0.635 1	0.384 2	0.793 2	0.907 7
Xiaoyang Wu, Li Jiang, Peng-Shuai Wang, Zhijian Liu, Xihui Liu, Yu Qiao, Wanli Ouyang, Tong He, Hengshuang Zhao: Point Transformer V3: Simpler, Faster, Stronger. CVPR 2024																
PonderV2		0.785 2	0.978 1	0.800 25	0.833 21	0.788 8	0.853 15	0.545 16	0.910 5	0.713 1	0.705 4	0.979 1	0.596 6	0.390 1	0.769 11	0.832 40
Haoyi Zhu, Honghui Yang, Xiaoyang Wu, Di Huang, Sha Zhang, Xianglong He, Tong He, Hengshuang Zhao, Chunhua Shen, Yu Qiao, Wanli Ouyang: PonderV2: Pave the Way for 3D Foundation Model with A Universal Pre-training Paradigm.																
Mix3D	P	0.781 3	0.964 2	0.855 1	0.843 15	0.781 6	0.858 11	0.575 6	0.831 31	0.685 13	0.714 2	0.979 1	0.594 7	0.310 26	0.801 1	0.892 15
Alexey Nekrasov, Jonas Schult, Or Litany, Bastian Leibe, Francis Engelmann: Mix3D: Out-of-Context Data Augmentation for 3D Scenes. 3DV 2021 (Oral)																
Swin3D	P	0.779 4	0.861 20	0.818 13	0.836 18	0.790 2	0.875 4	0.576 5	0.905 6	0.704 4	0.739 1	0.969 10	0.611 2	0.349 10	0.756 20	0.958 1
TTT-KD		0.773 5	0.646 89	0.818 13	0.809 33	0.774 8	0.878 3	0.581 2	0.943 1	0.687 11	0.704 5	0.978 4	0.607 5	0.336 15	0.775 8	0.912 5
Lisa Weijler, Muhammad Jehanzeb Mirza, Leon Sick, Can Ekkazan, Pedro Hermosilla: TTT-KD: Test-Time Training for 3D Semantic Segmentation through Knowledge Distillation from Foundation Models.																
ResLFE_HDS		0.772 6	0.939 4	0.824 6	0.854 6	0.771 9	0.840 29	0.564 10	0.900 8	0.686 12	0.677 11	0.961 16	0.537 29	0.348 11	0.769 11	0.903 9
OctFormer	P	0.766 7	0.925 7	0.808 21	0.849 9	0.786 4	0.846 25	0.566 9	0.876 14	0.690 9	0.674 13	0.960 17	0.576 16	0.226 65	0.753 22	0.904 8
Peng-Shuai Wang: OctFormer: Octree-based Transformers for 3D Point Clouds. SIGGRAPH 2023																
PPT-SpUNet-Joint		0.766 7	0.932 5	0.794 31	0.829 23	0.751 21	0.854 13	0.540 20	0.903 7	0.630 32	0.672 14	0.963 14	0.565 20	0.357 8	0.788 3	0.900 11
Xiaoyang Wu, Zhuotao Tian, Xin Wen, Bohao Peng, Xihui Liu, Kaicheng Yu, Hengshuang Zhao: Towards Large-scale 3D Representation Learning with Multi-dataset Point Prompt Training. CVPR 2024																
OccuSeg+Semantic		0.764 9	0.758 57	0.796 29	0.839 17	0.746 23	0.907 1	0.562 11	0.850 23	0.680 15	0.672 14	0.978 4	0.610 3	0.335 17	0.777 6	0.819 43
CU-Hybrid Net		0.764 9	0.924 8	0.819 11	0.840 16	0.757 16	0.853 15	0.580 3	0.848 24	0.709 3	0.643 22	0.958 20	0.587 11	0.295 32	0.753 22	0.884 19
O-CNN	P	0.762 11	0.924 8	0.823 7	0.844 14	0.770 10	0.852 17	0.577 4	0.847 26	0.711 2	0.640 26	0.958 20	0.592 8	0.217 71	0.762 16	0.888 16
Peng-Shuai Wang, Yang Liu, Yu-Xiao Guo, Chun-Yu Sun, Xin Tong: O-CNN: Octree-based Convolutional Neural Networks for 3D Shape Analysis. SIGGRAPH 2017																
OA-CNN-L_ScanNet20		0.756 12	0.783 43	0.826 5	0.858 4	0.776 7	0.837 32	0.548 15	0.896 11	0.649 24	0.675 12	0.962 15	0.586 12	0.335 17	0.771 10	0.902 47
ConDaFormer		0.755 13	0.927 6	0.822 8	0.836 18	0.801 1	0.849 20	0.516 30	0.864 20	0.651 23	0.680 10	0.958 20	0.584 14	0.282 40	0.759 18	0.855 30
Lunhao Duan, Shanshan Zhao, Nan Xue, Mingming Gong, Guisong Xia, Dacheng Tao: ConDaFormer : Disassembled Transformer with Local Structure Enhancement for 3D Point Cloud Understanding. Neurips, 2023																
PNE		0.755 13	0.786 41	0.835 4	0.834 20	0.758 14	0.849 20	0.570 8	0.836 30	0.648 25	0.668 16	0.978 4	0.581 15	0.367 6	0.683 33	0.856 28
P. Hermosilla. Point Neighborhood Embeddings.																
PointTransformerV2		0.752 15	0.742 65	0.809 20	0.872 1	0.758 14	0.860 10	0.552 13	0.891 12	0.610 39	0.687 6	0.960 17	[illegible]	[illegible]	[illegible]	[illegible]
Xiaoyang Wu, Yixing Lao, Li Jiang, Xihui Liu, Hengshuang Zhao: Point Transformer V2: Grouped Vector Attention and Partition-based Pooling. NeurIPS 2022																

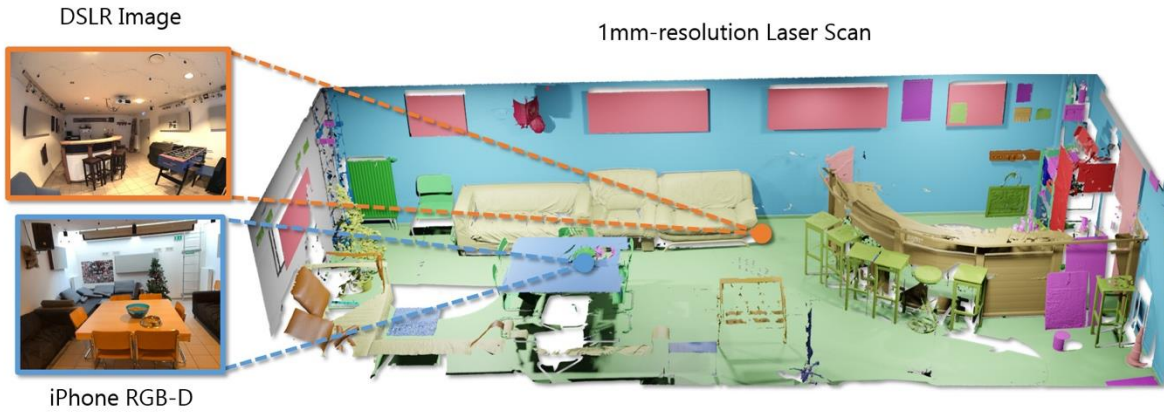
http://www.scan-net.org/

3D Scenes

- SceneNN
  - ScanNetStyle (100 scenes)
- Matterport3D
  - 90 buildings
- ARKitScenes
  - 1661 scenes
  - Faro + iPhone (no labels)
- ScanNet++
  - 450 3D indoor scenes
  - Faro Scans
  - DSLR + iPhone
  - Instance Labels

Many more….





Important Concepts

- Regular Structures vs Irregular structures
  - Convolutions, MLPs, Transformers, etc.
- Spatial data structures:
  - Octress, Hashes, Hierarchies
- Pooling:
  - n->1 mappings (points, views, etc.)
- Architecture vs Loss formulation
  - Transformer, Diffusion

Next Lectures

- Neural Scene Representations
- Neural Fields
- Neural Radiance Fields (NeRF)

Thanks and Questions