Creating and Understanding 3D Annotated Scene Meshes

Nov 18, 2019

Brisbane, Australia

SIGGRAPH ASIA 2019 BRISBANE

SA '19 Courses, November 17-20, 2019, Brisbane, QLD, Australia
ACM 978-1-4503-6941-1/19/11.
10.1145/3355047.3359426

Organizers

Duc Thanh Nguyen

Deakin University

Quang-Hieu Pham

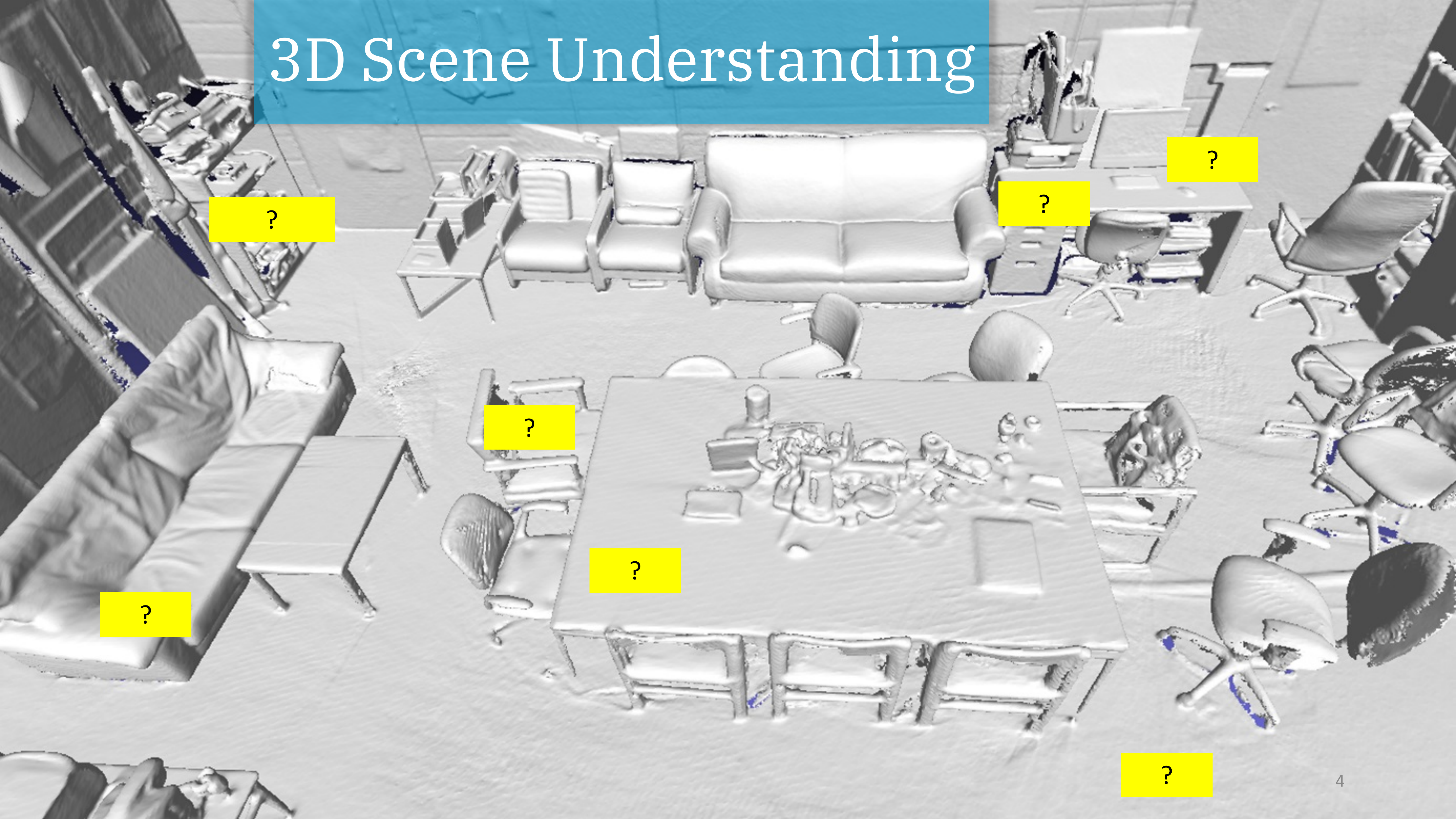
SUTD

Binh-Son Hua

The University of Tokyo

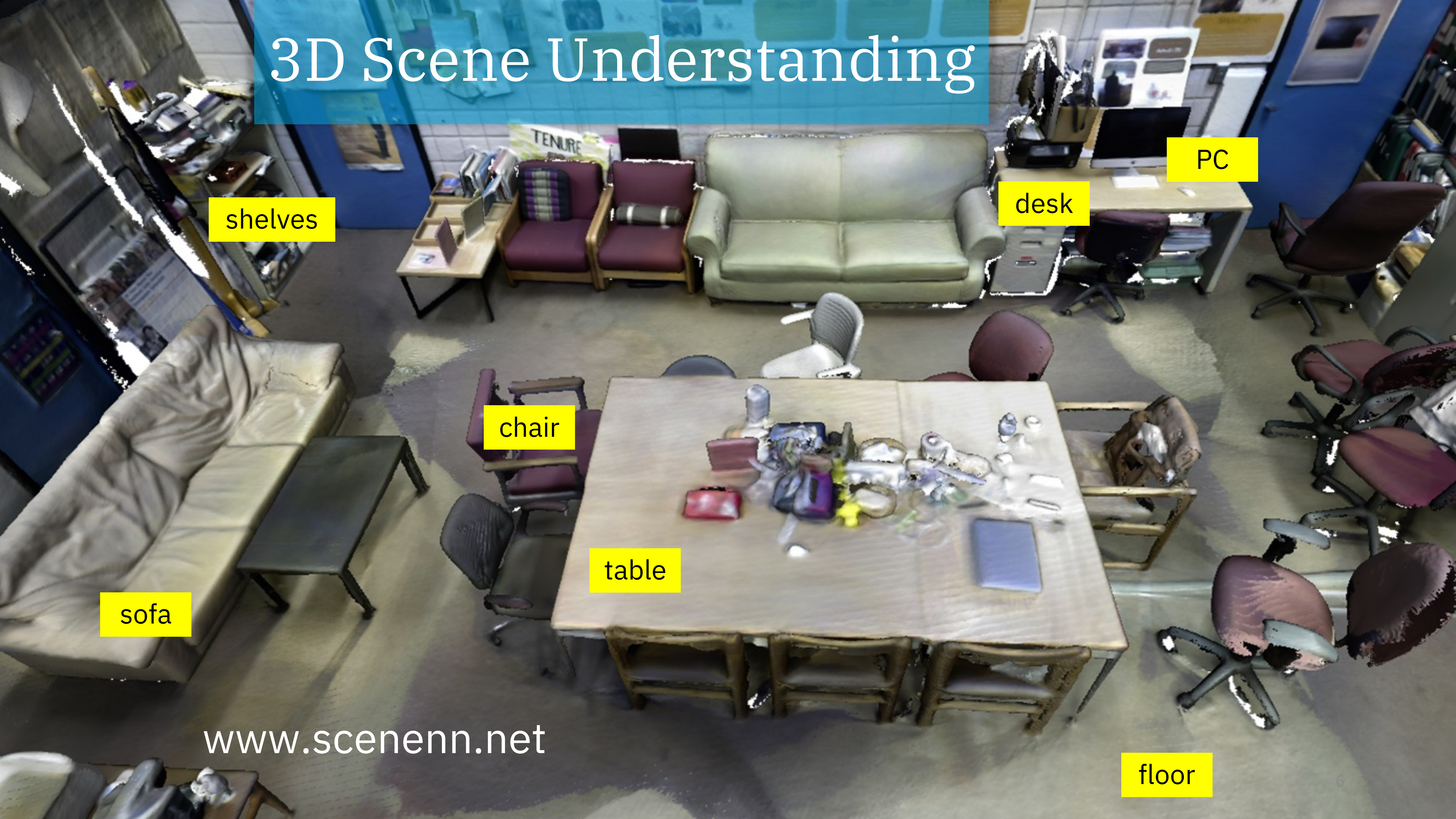
Scene Understanding

3D Scene Understanding



3D Scene Understanding

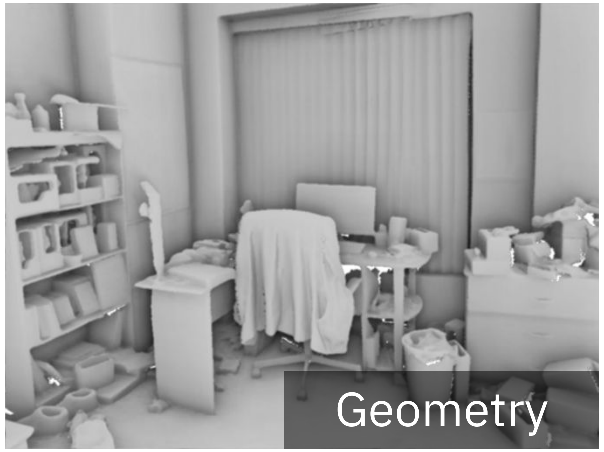
3D Scene Understanding

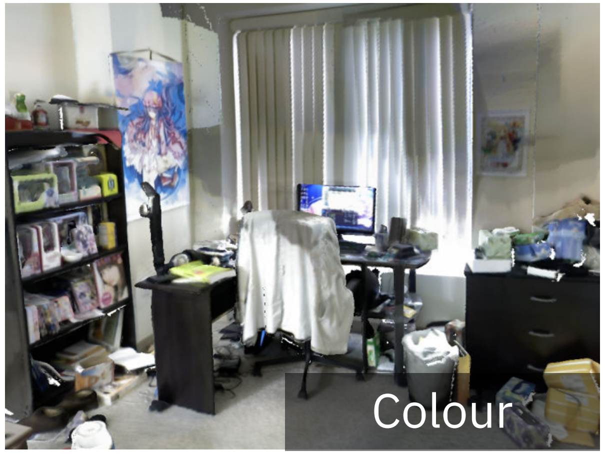


www.scenenn.net

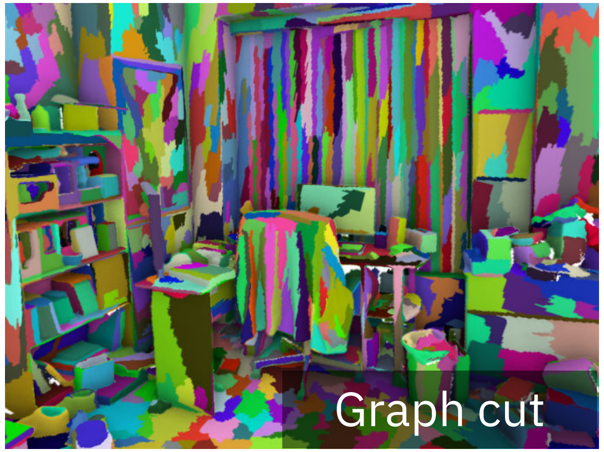
The Pipeline

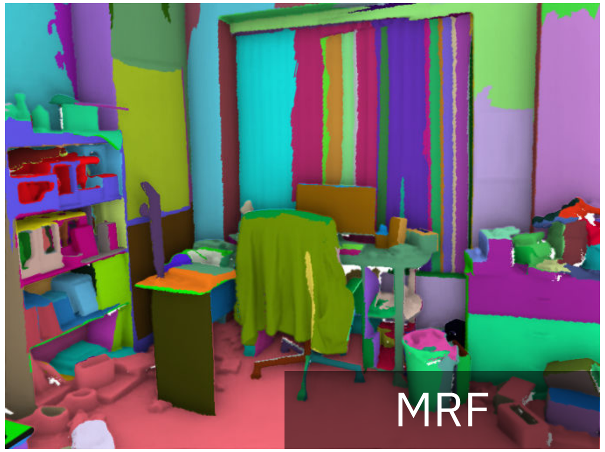
3D reconstruction

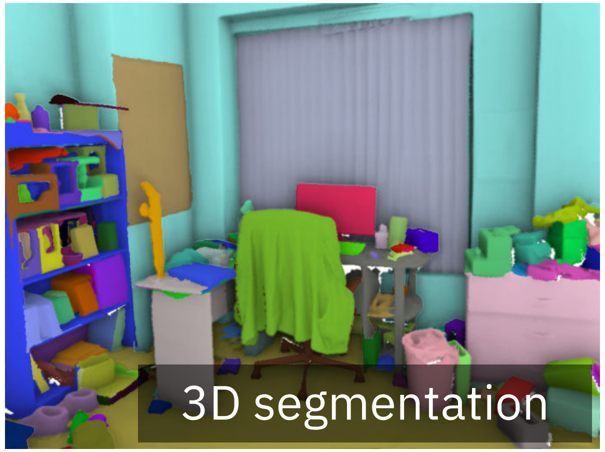


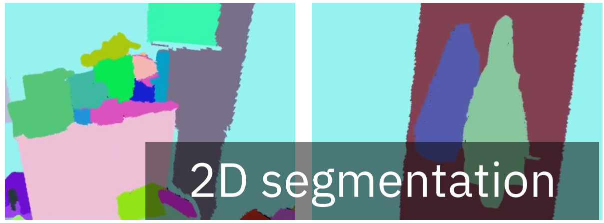


Automatic segmentation









User interaction

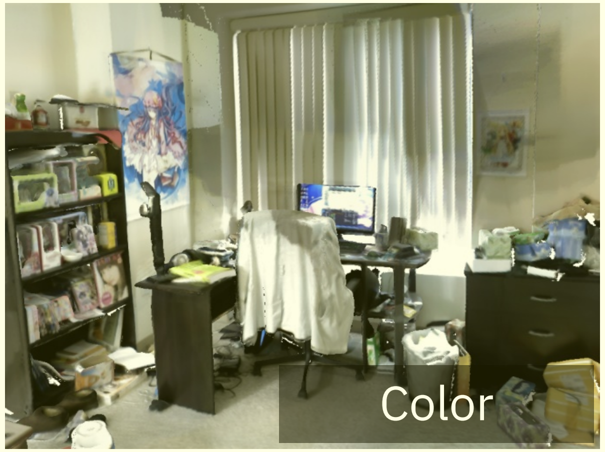
Fine-grained annotation

- 3D and 2D refinement
- Object annotation
- Object search

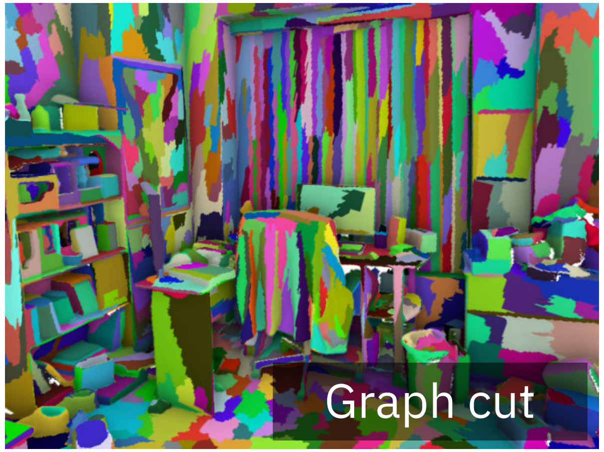
Part I:
Reconstructing 3D Scenes

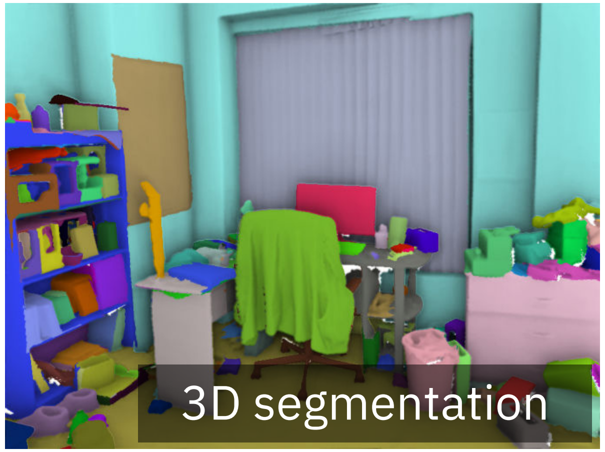
The Pipeline

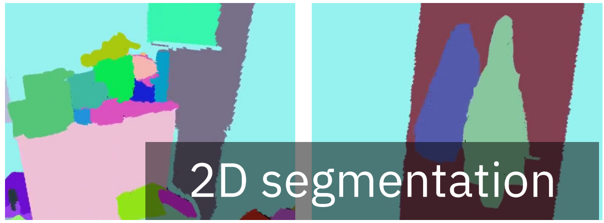
3D reconstruction



Automatic segmentation







User interaction

Fine-grained annotation

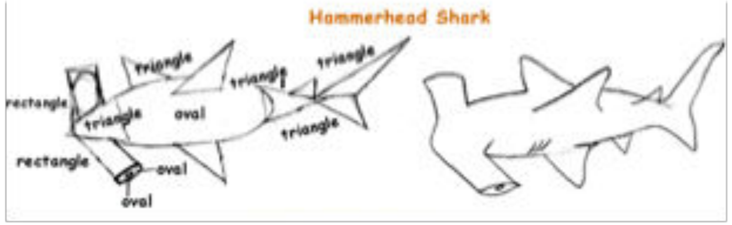
- 3D and 2D refinement
- Object annotation
- Object search

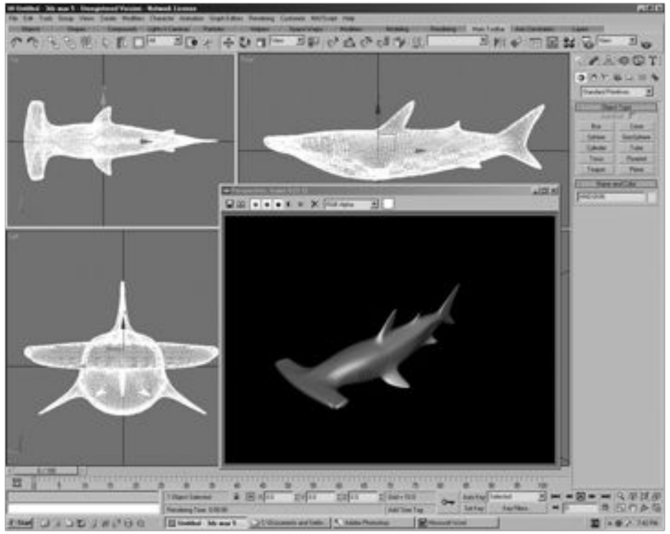
Motivation

- Deep learning requires availability of massive 3D data.
- How to acquire 3D scenes efficiently?

Digital Design and Manufacturing

Design Ideas → Digital Design: 3D Modeling → Manufacturing



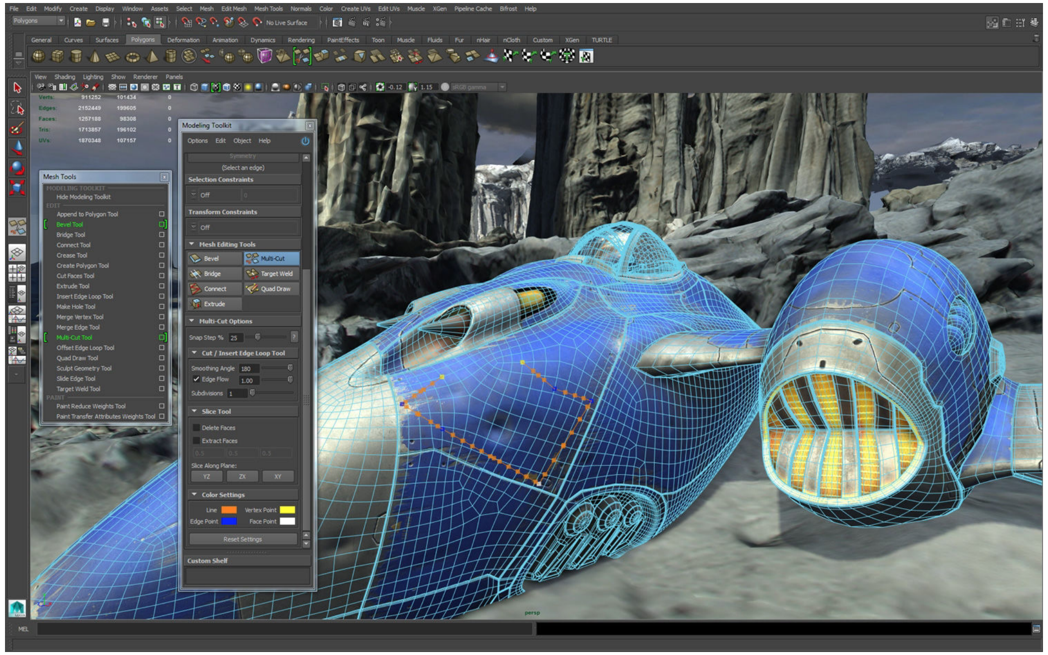


Digital Design Challenge: How to create 3D models?

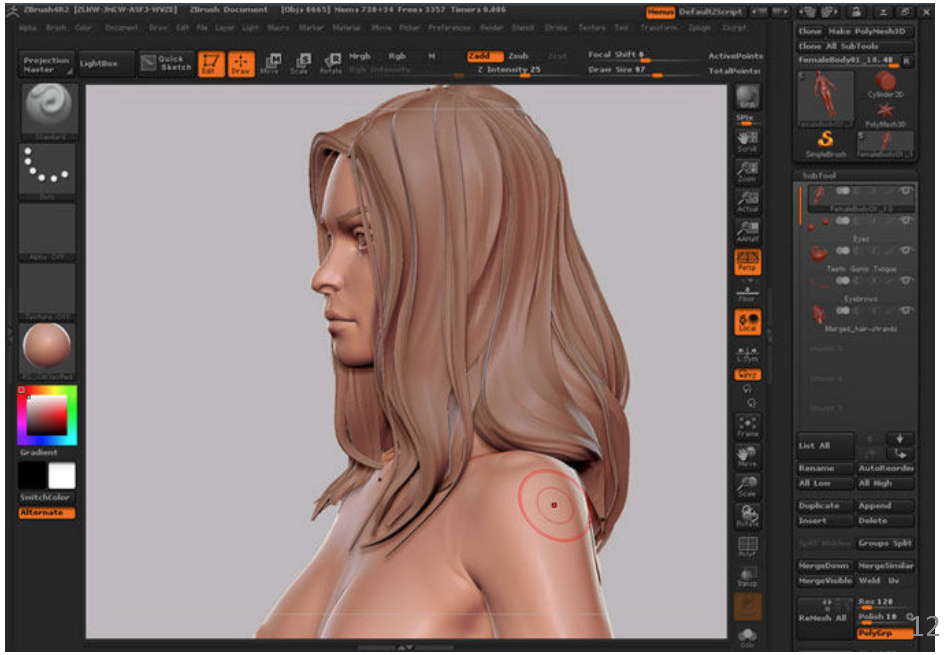
Common Approach - Manual Creation

- Polygonal Modeling: man-made objects
- Digital Sculpting: organic objects

Polygonal Modeling



Digital Sculpting



Manual Creation - Limitations

- Need modeling expertise
- Labor intensive & tedious
- Huge money & time investment
- Non-scalable

Experienced Artist: 7 days

Manual Creation - Limitations

Modeling time for an entire city?

Computer Vision:
3D Reconstruction from 2D images

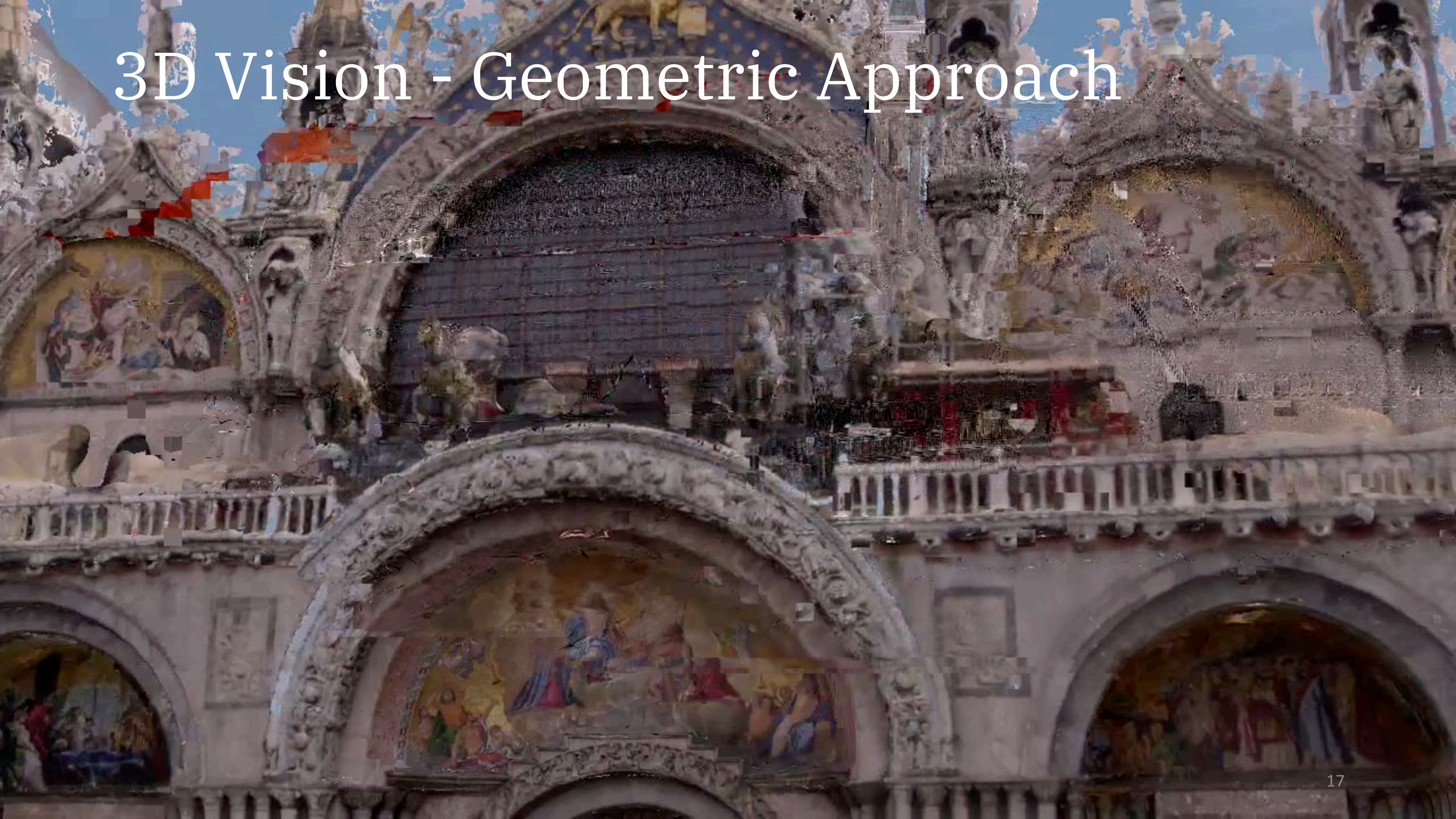
	Geometric approach, e.g. MVS	Photometric approach
Gross shape	○	✕
Detailed shape	✕	○

Colored points

Normal map

3D Vision - Photometric Approach

3D Vision - Geometric Approach



Multi-view Stereo (MVS)

Input

Feature Matching

[Lowe IJCV ’04]
[Rublee ICCV ’11]
[JW Bian CVPR ’17]

SfM

[Szeliski ICCV ’09]
[Wu C. 3DTV ’13]
[Schoenberger CVPR ’16]

Densification

[Kanade TPAMI ’93]
[Furukawa CVPR ’07]
[Goesele ICCV ’07]
[Langguth ECCV ’16]

Surface

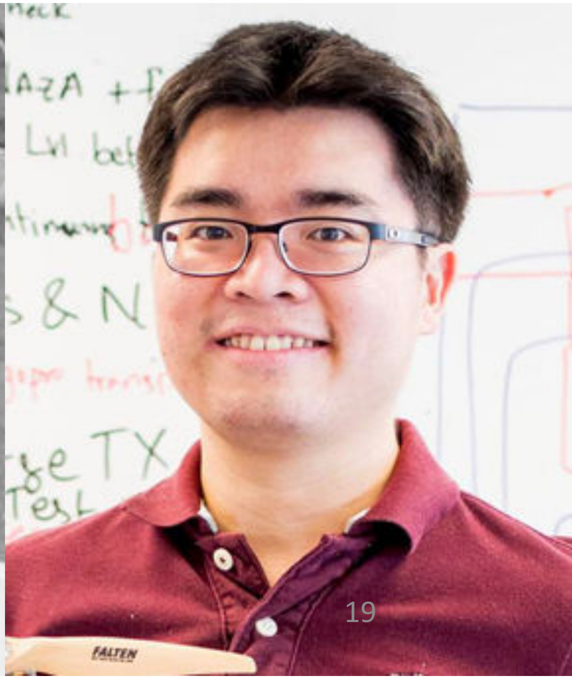
[Fuhrmann Siggraph ’14]
[Langguth ECCV ’16]
[Aroudj SiggraphAsia ’17]

Images courtesy of Furukawa, Yasutaka, and Jean Ponce. *TPAMI* (2010): 1362-1376.

Outdoor Reconstruction

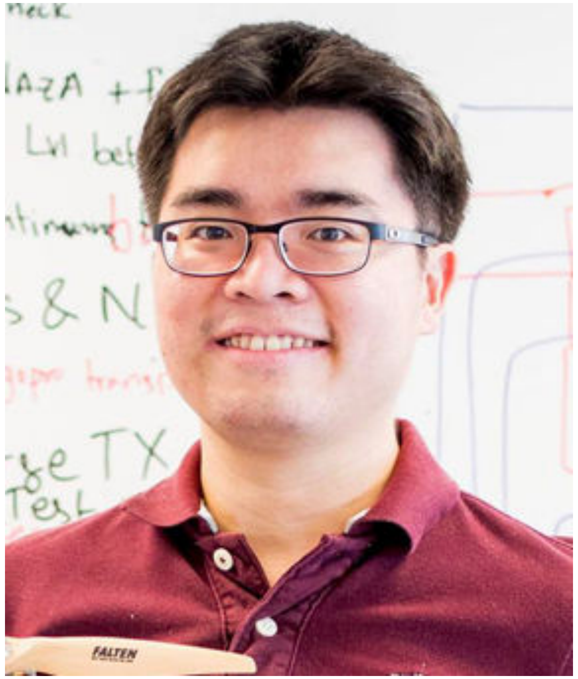
National Heritage Board

- Reconstructing tangible heritages
- Develop surface from point clouds algorithms
- VR viewer app



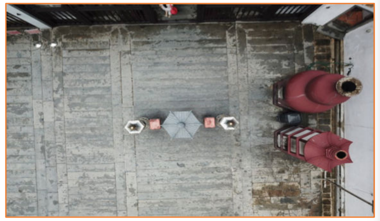
Outdoor Reconstruction

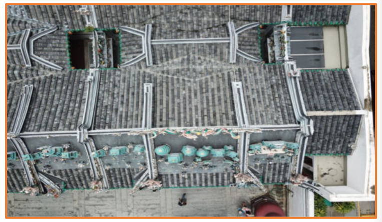
Drones to collect images

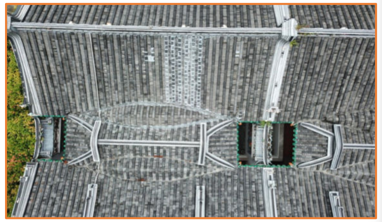


Multi-view Stereo (MVS)

Multi-view Stereo (MVS)

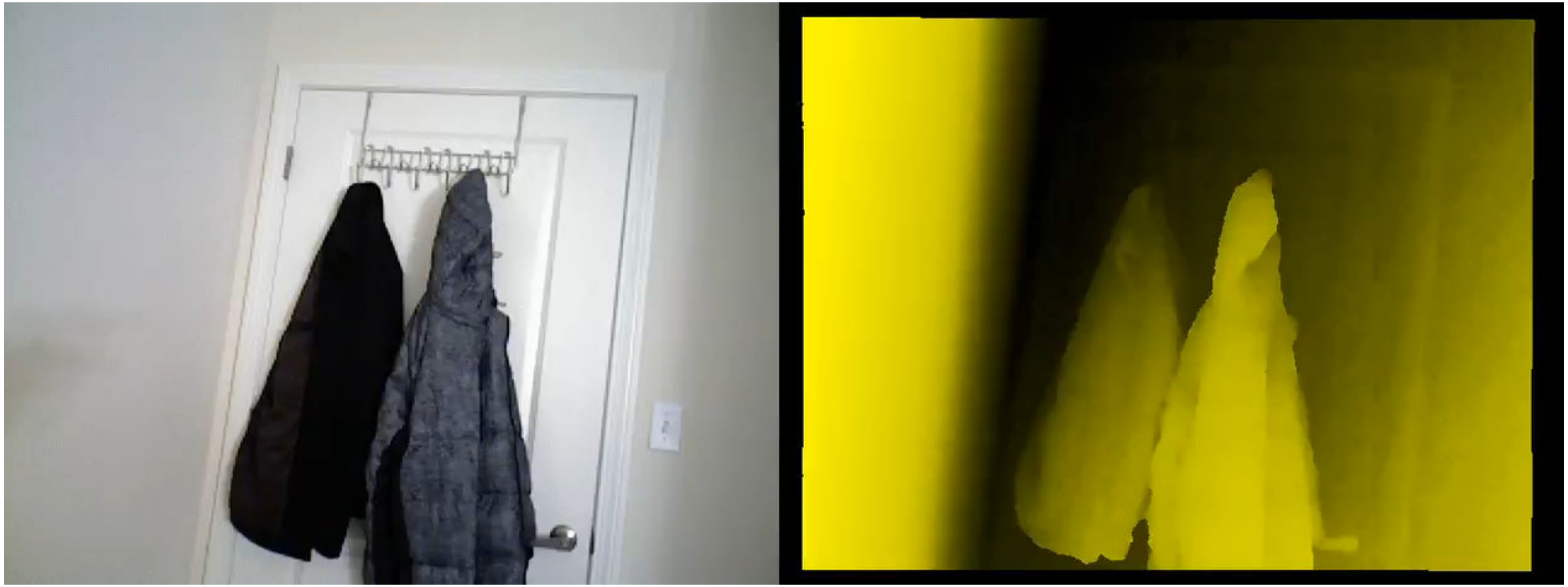




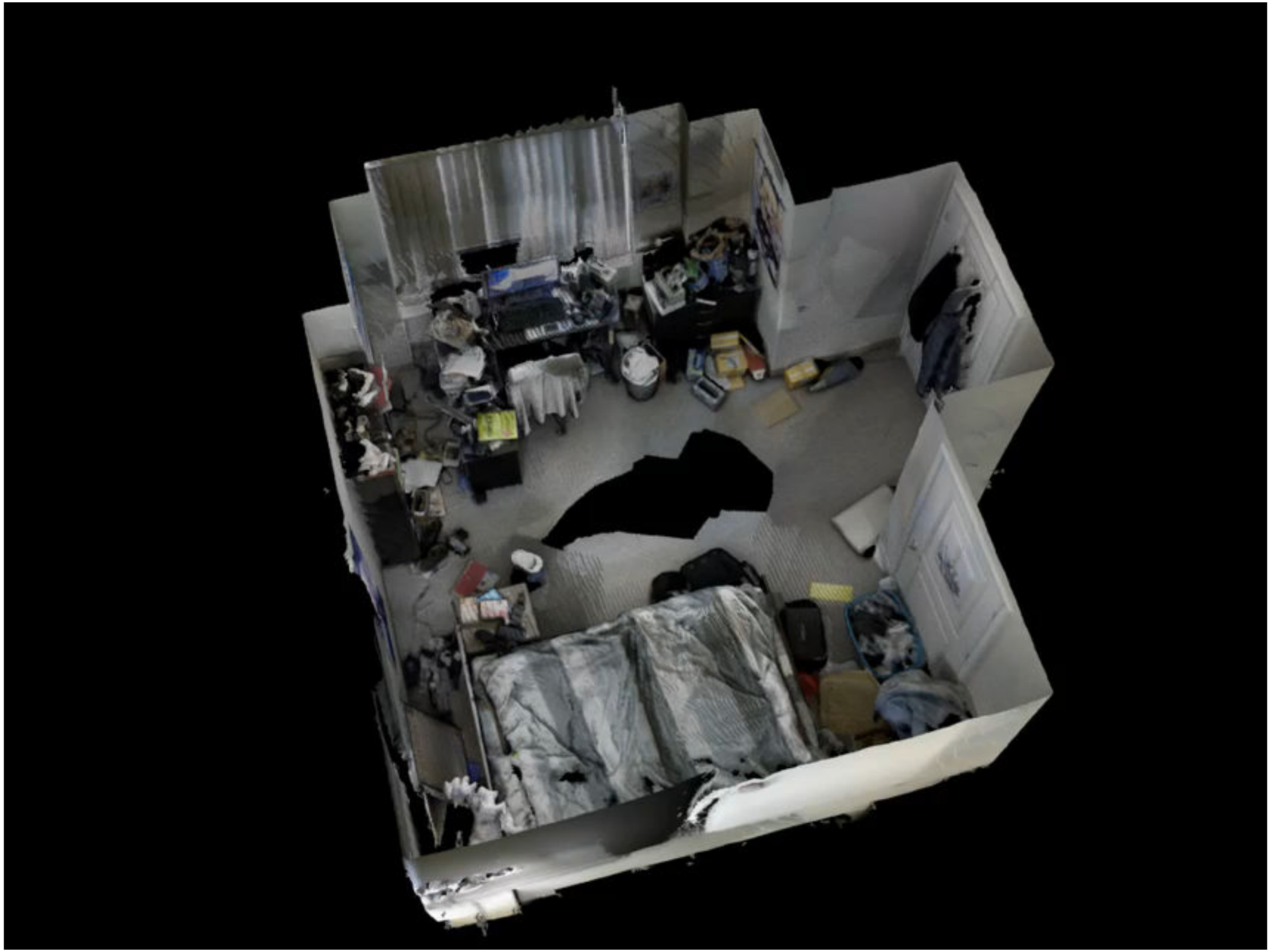


Indoor Reconstruction

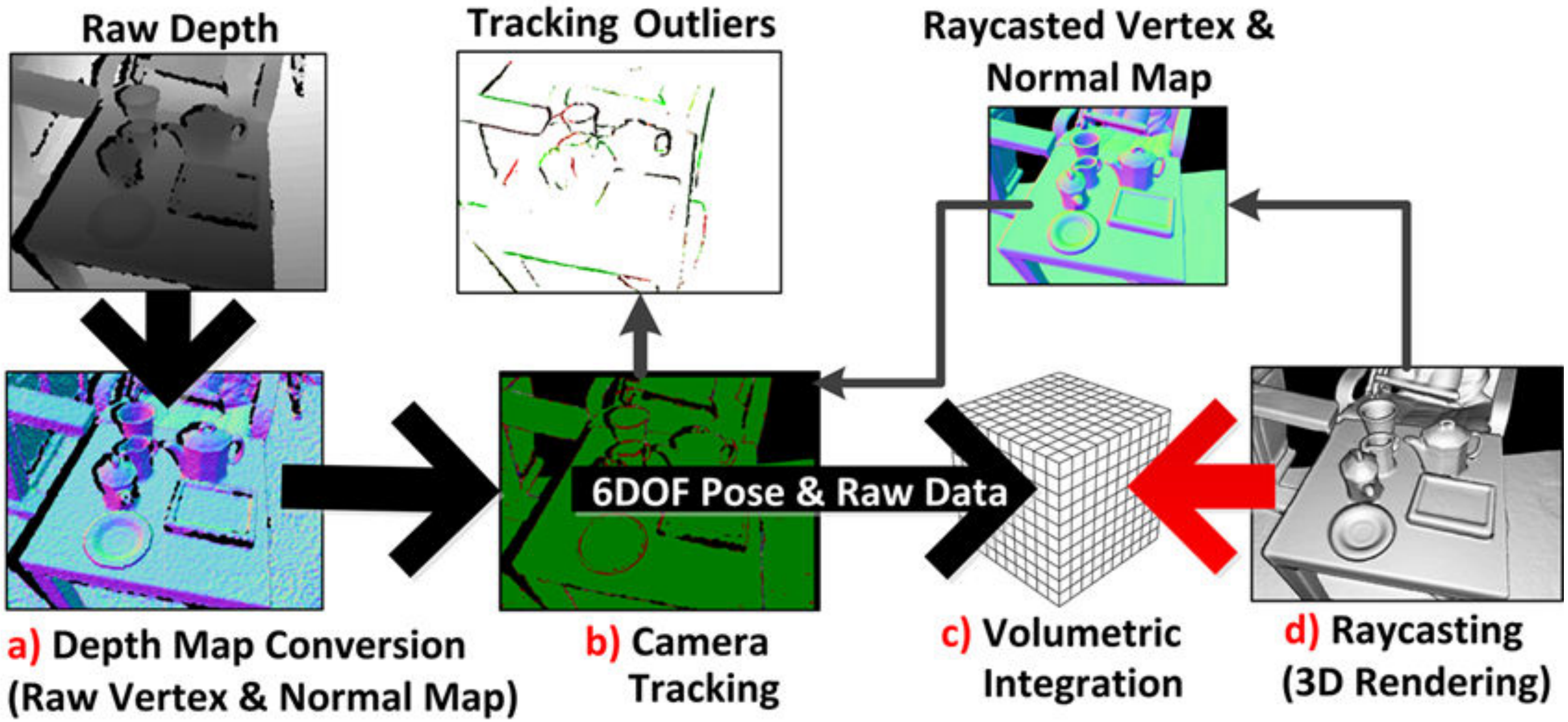
Scene reconstruction with RGB-D sensors



Indoor Reconstruction



KinectFusion

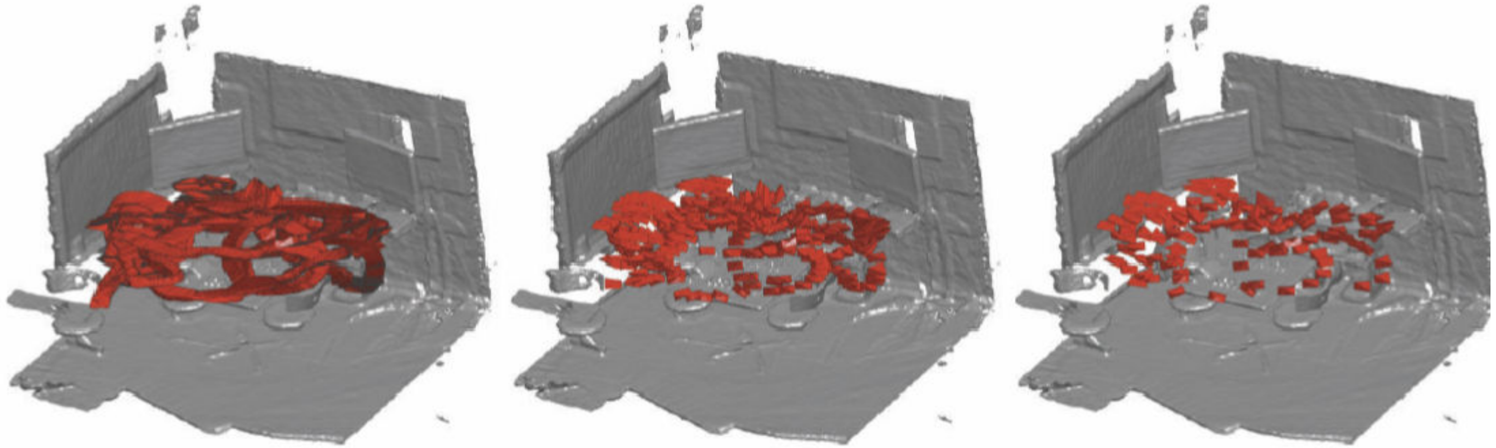


Images retrieved from https://msdn.microsoft.com/en-us/library/dn188670.aspx

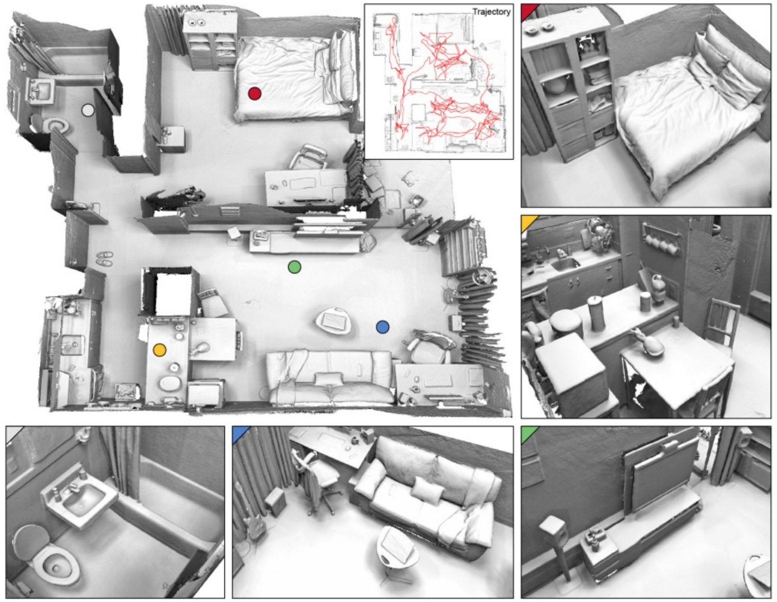
KinectFusion - Challenges

Large-scale reconstruction

Relocalization



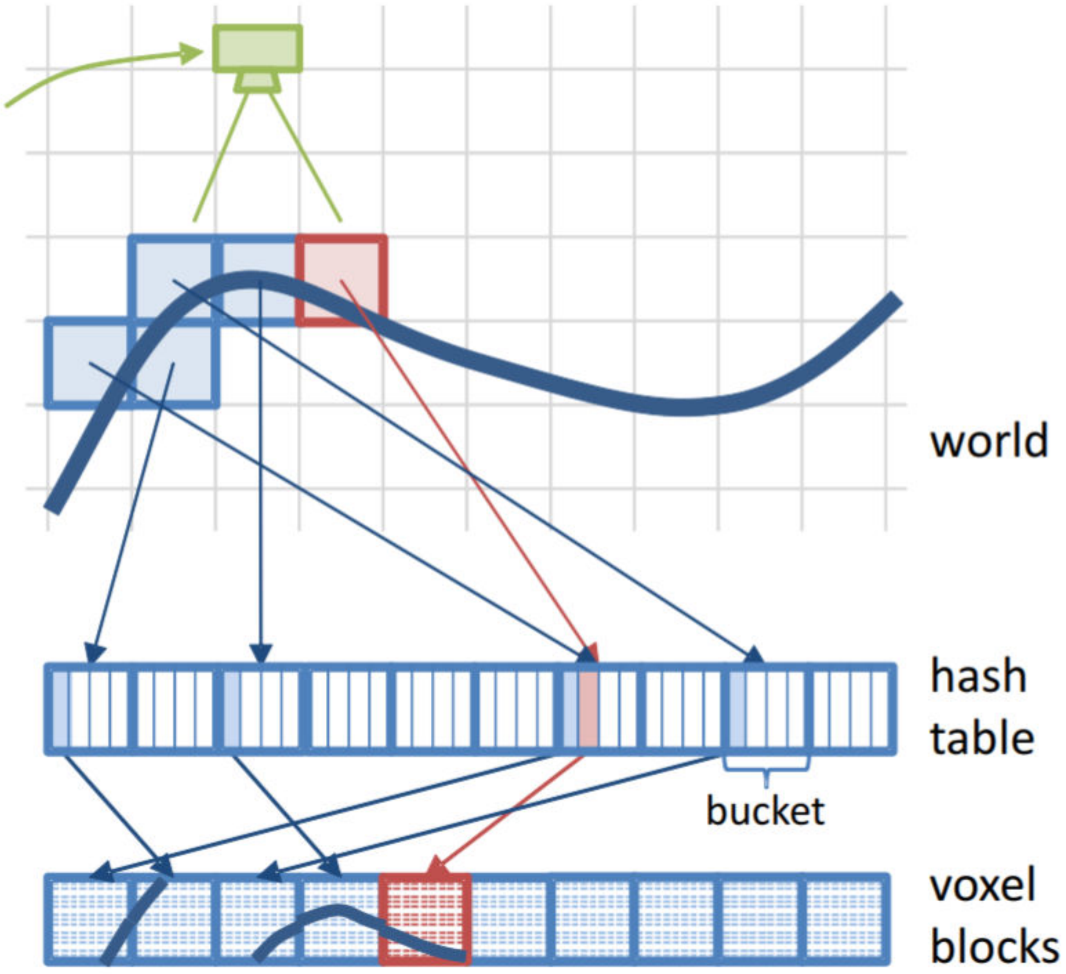
Global optimization



Sparse Voxel Hashing

Image courtesy of Niessner et al., 2013.

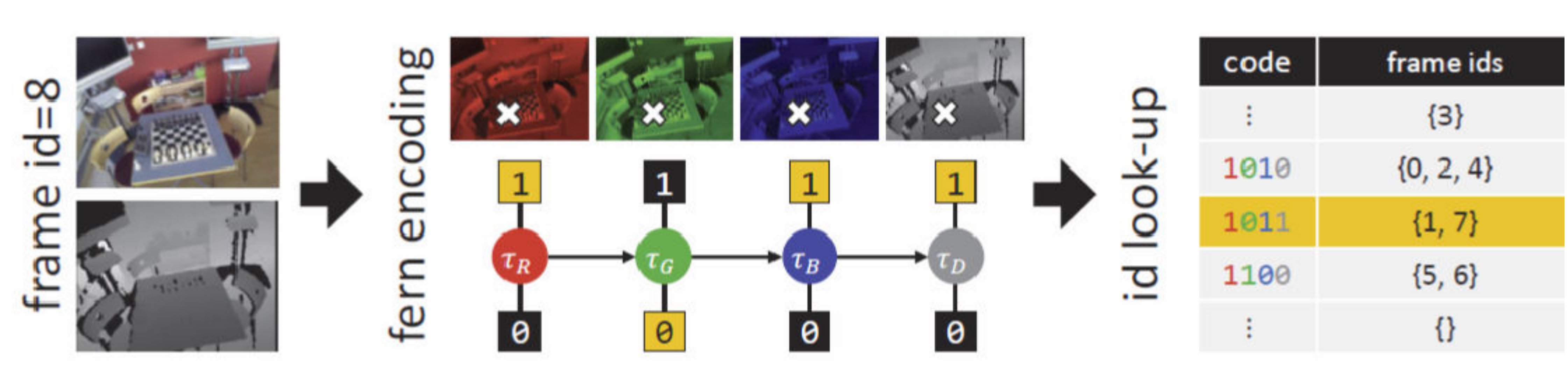
- Store only non-empty voxel block
- Hash table for book keeping
- Real-time but still no constraints for loop closure



Real-time 3D Reconstruction at Scale using Voxel Hashing, Niessner et al., TOG 2013

Relocalization

- Fern encoding on each keyframe
- Frame dissimilarity with block-wise Hamming distance
- Can recover from tracking failure

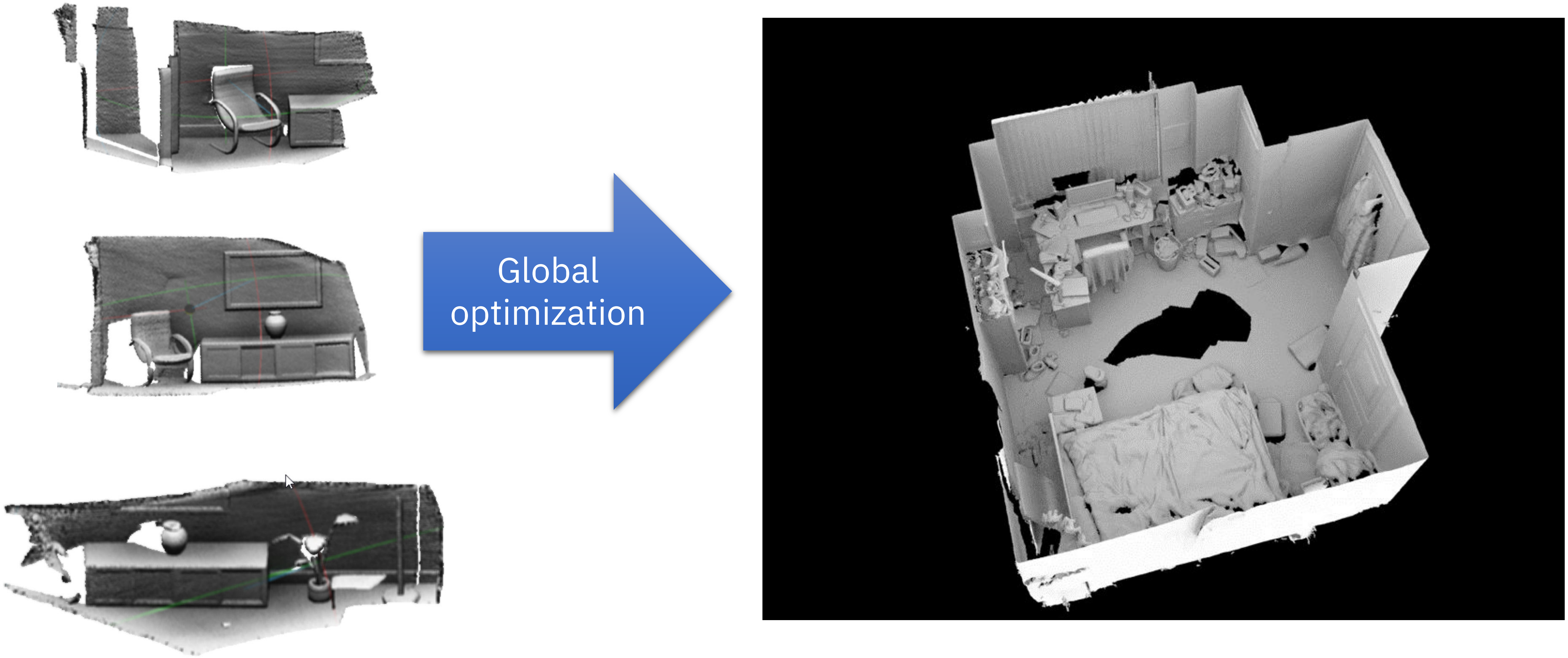


Real-time RGB-D Camera Relocalization, Glocker et al., ISMAR 2013

Relocalization

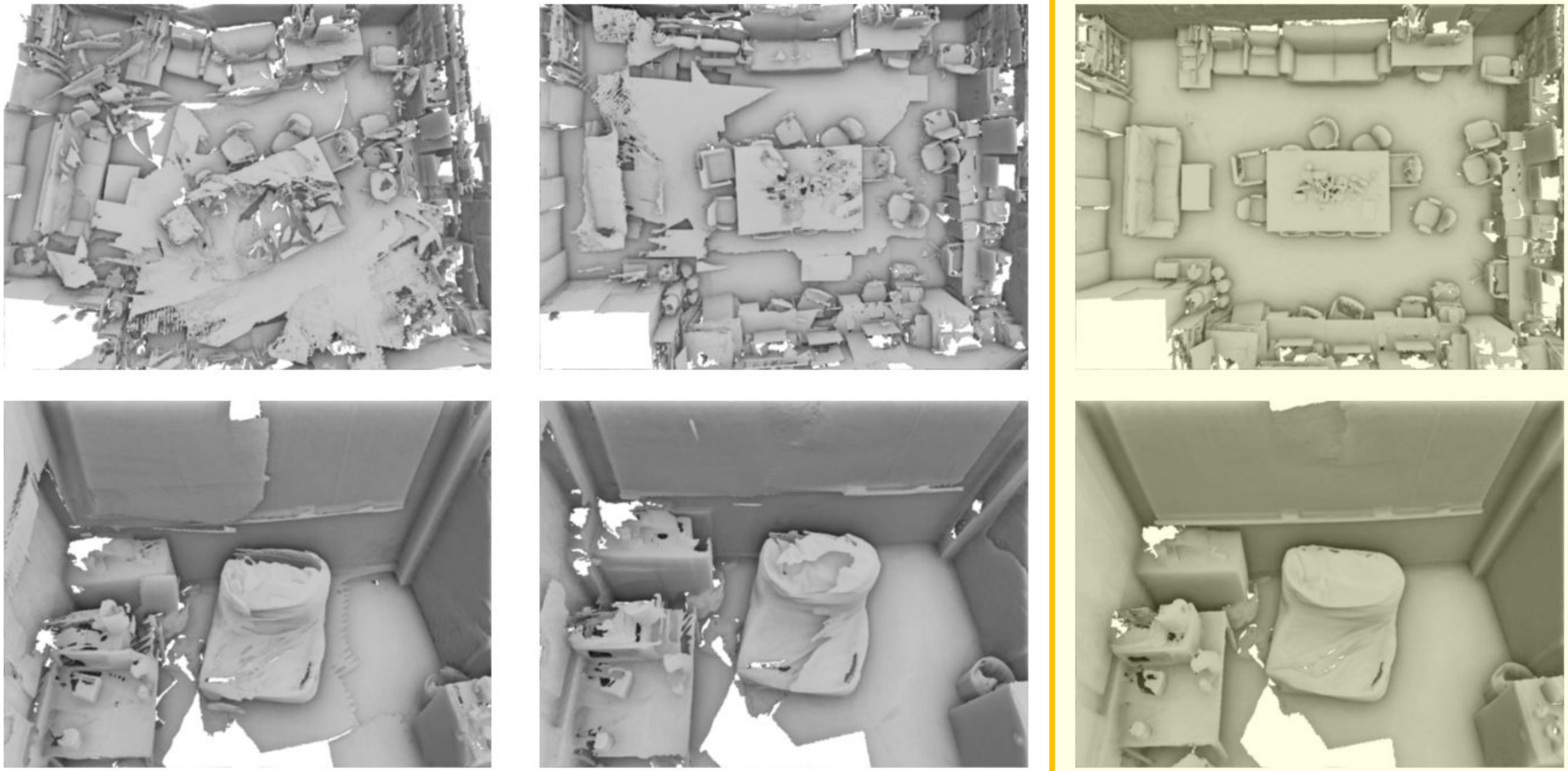
Real-time RGB-D Camera Relocalization, Glocker et al., ISMAR 2013

Global Optimization



Robust Reconstruction of Indoor Scenes, Sungjoon Choi, Qian-Yi Zhou, and Vladlen Koltun, CVPR 2015

RGBD reconstruction



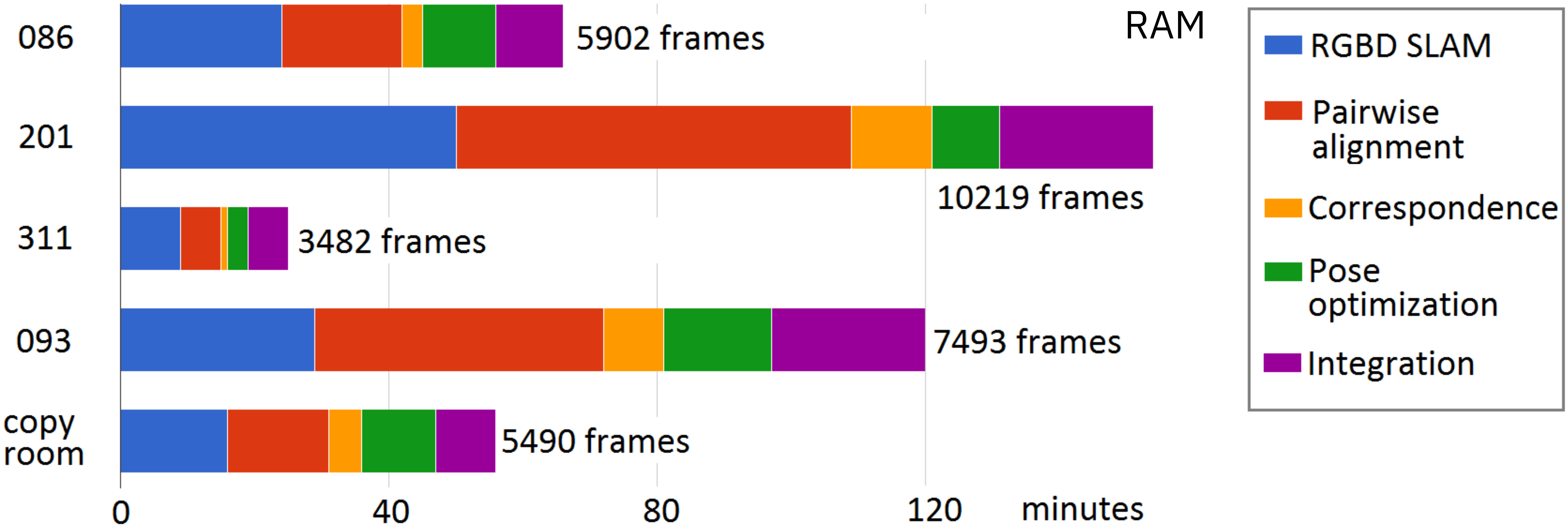
DVO SLAM
[Kerl et al., IROS 2013]

Elastic Fusion
[Whelan et al., RSS 2015]

Elastic Reconstruction
[Choi et al., CVPR 2015]

Reconstruction statistics

CPU Intel Core i7 5960X @3Ghz, 32GB RAM

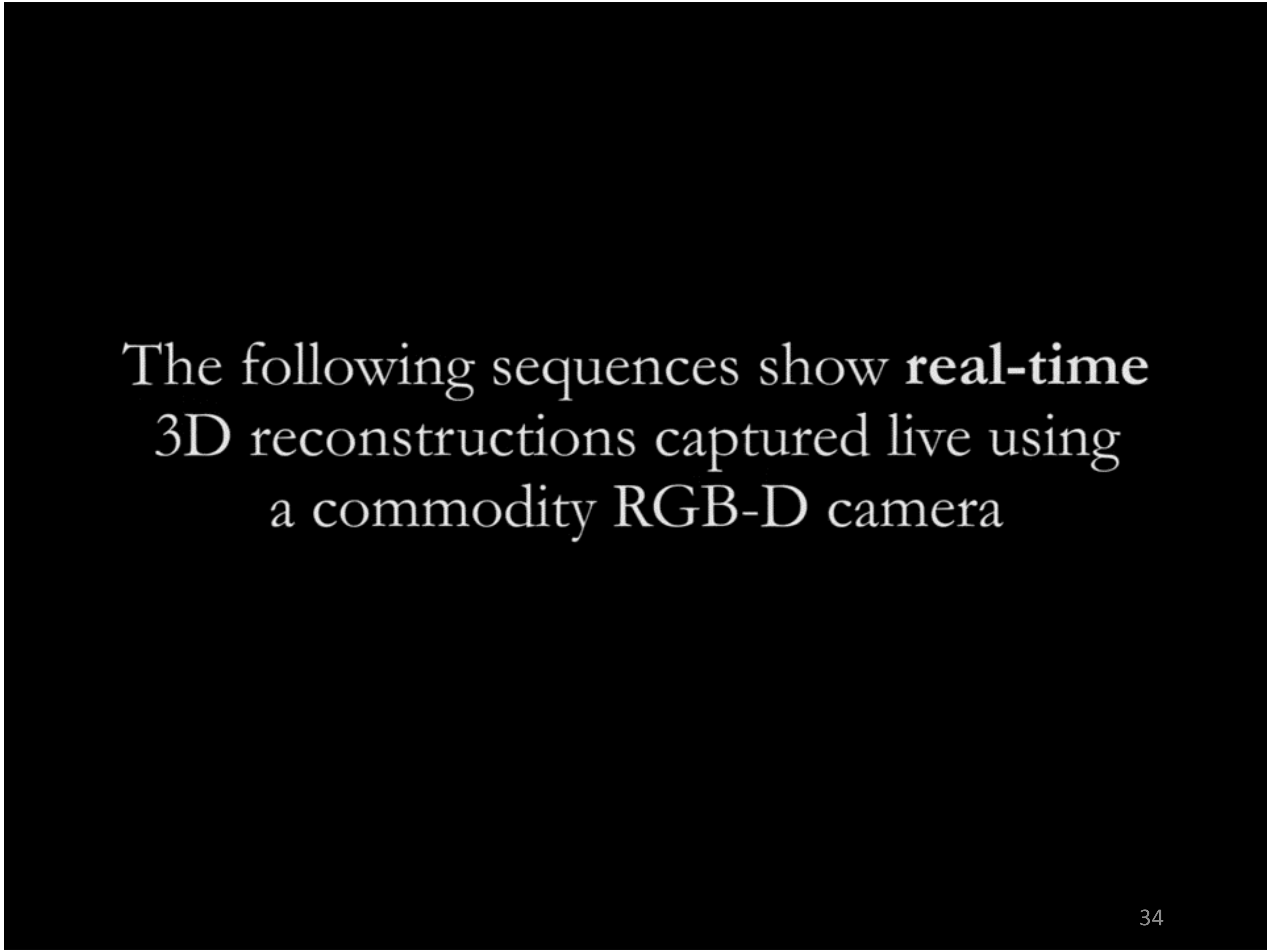


Robust Reconstruction of Indoor Scenes, Sungjoon Choi, Qian-Yi Zhou, and Vladlen Koltun, CVPR 2015

Real-time Reconstruction with BundleFusion

- Sparse-to-dense matching
- Local-to-global optimization
- On-the-fly model update

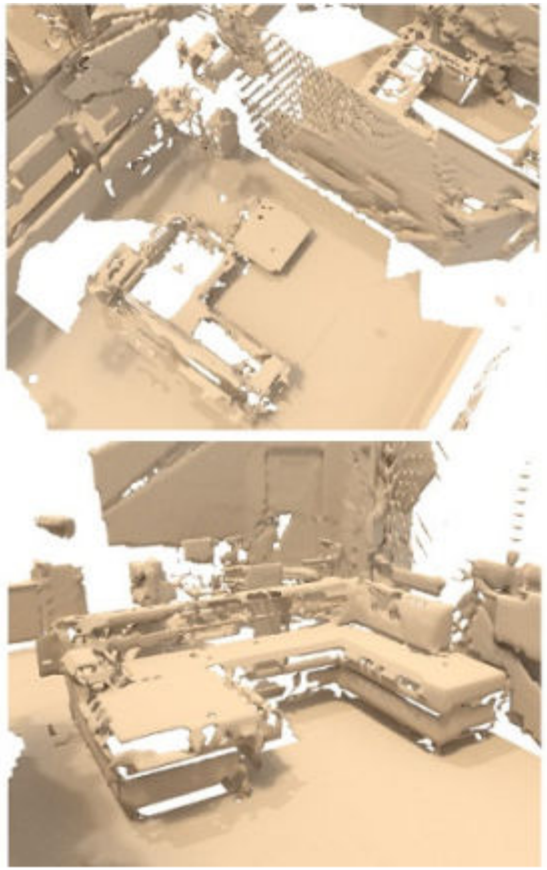
BundleFusion: Real-time Globally Consistent 3D Reconstruction using On-the-fly Surface Re-integration, Dai et al, TOG 2017



3D Reconstruction - Limitations

Reflective surfaces

Incomplete scans



Low-quality textures

Scene Representation

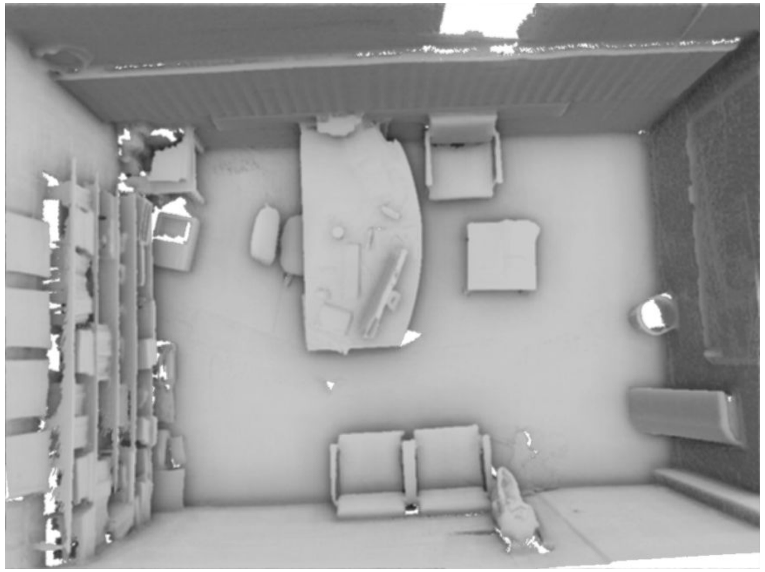
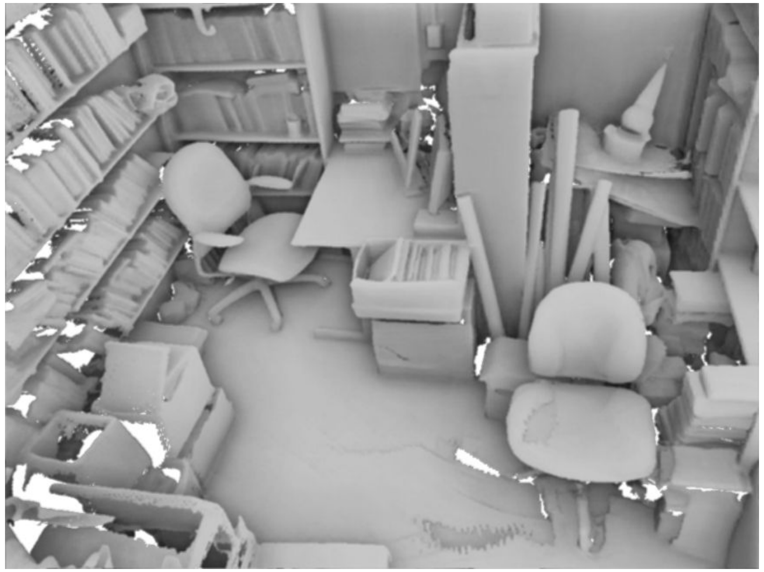
	Point Cloud	Triangle Mesh	Volume	Images
Storage	Efficient	Efficient	Sparse representation	Efficient
Learning	On-going research	Few previous works	Octree, KD-tree	Multiple 2D views
Rendering	Splatting	Rasterization, ray tracing	Ray marching	View interpolation

Part II: Designing a Robust Interactive Tool for 3D Scene Segmentation

SIGGRAPH ASIA 2019 BRISBANE

Motivation

High-quality 3D scenes using RGB-D cameras are widely available.

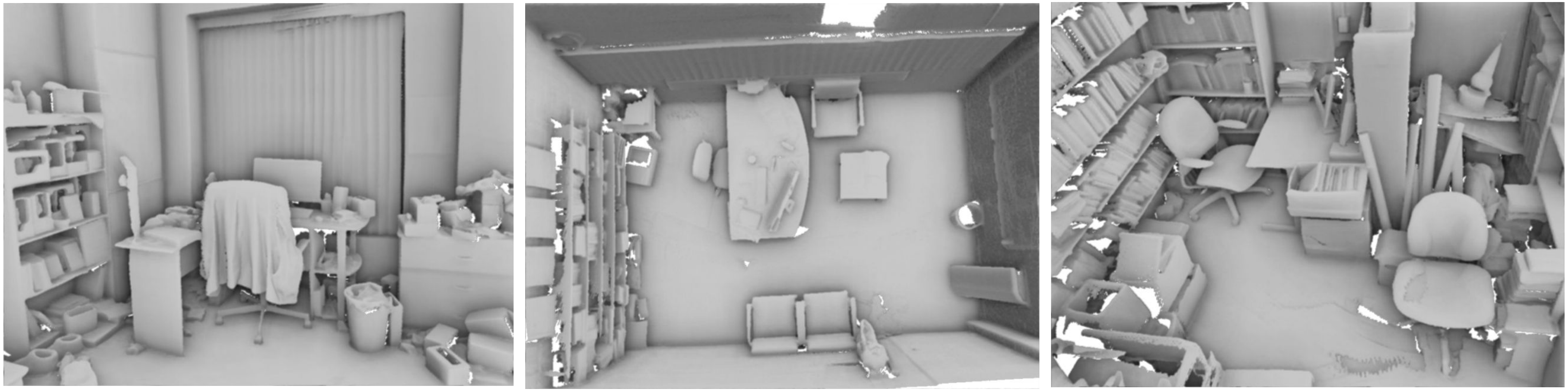




What to do with reconstructed data?


poster
curtain
wall
monitor
lamp
clothes
cabinet
desk
chair
bin
shelves
floor

Motivation

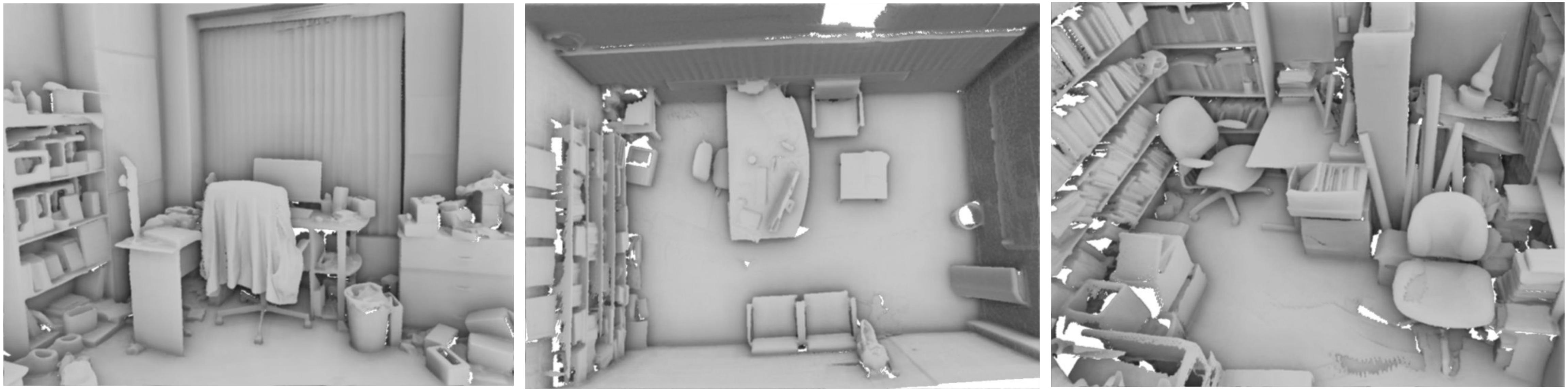
Semantic segmentation, object detection, pose estimation are still challenging to solve for 3D scenes.



Motivation

Deep learning needs massive ground truth data for training.

How to annotate 3D scenes effectively?

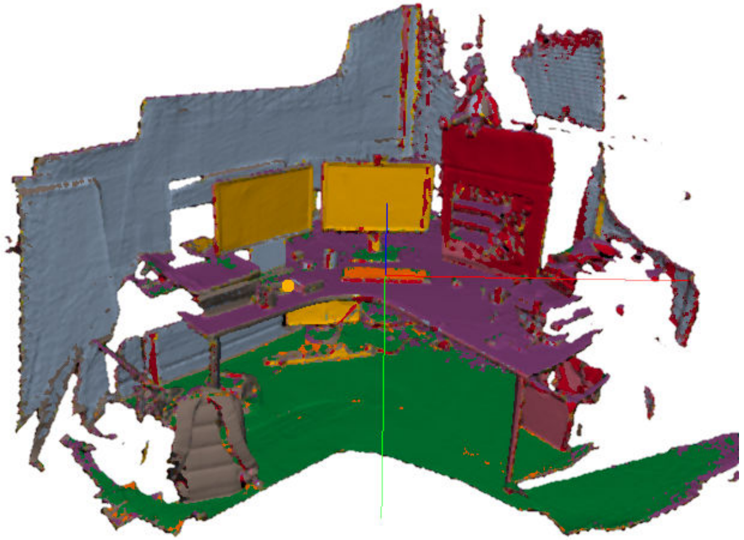


Problem statement

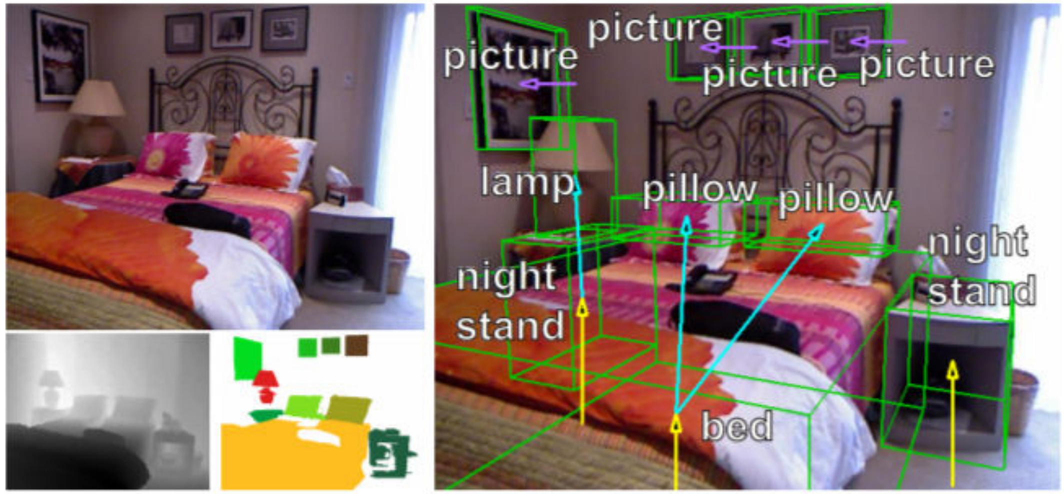
An **interactive** tool for scene segmentation and annotation

- Annotate a 3D scene and many RGB-D images in a single system.
- No world assumption. Capable to annotate any scenes.
- Dense annotation: per vertex and per pixel label.
- Fine-grained annotation: object poses, bounding boxes.

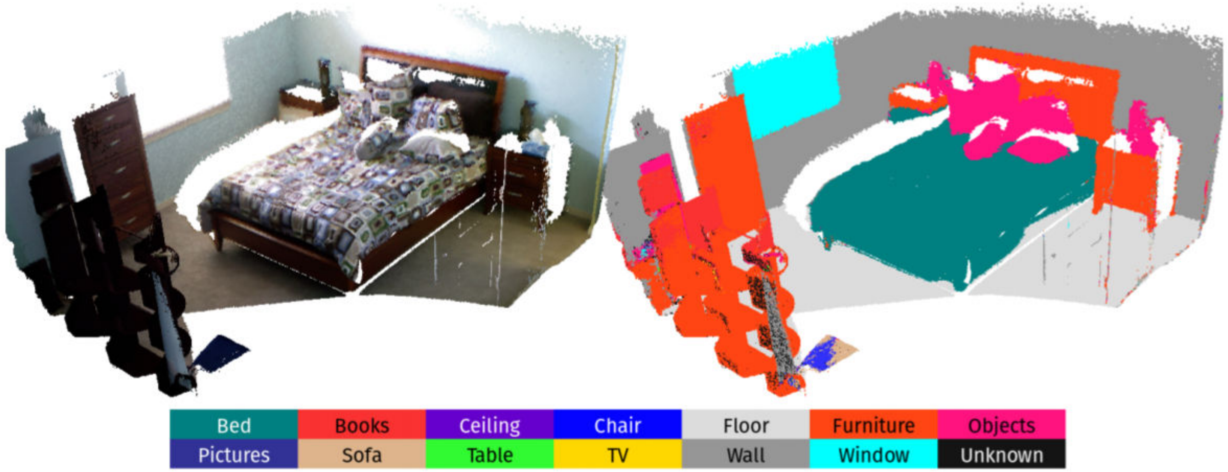
Related works



SemanticPaint,
Valentin et al., TOG 2015



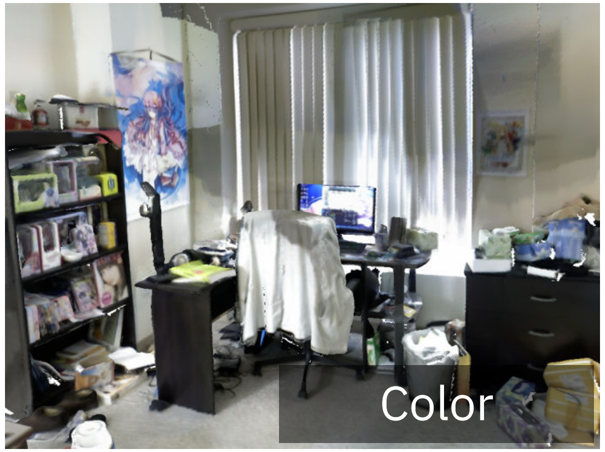
SmartAnnotator, Wong et al., Eurographics 2015



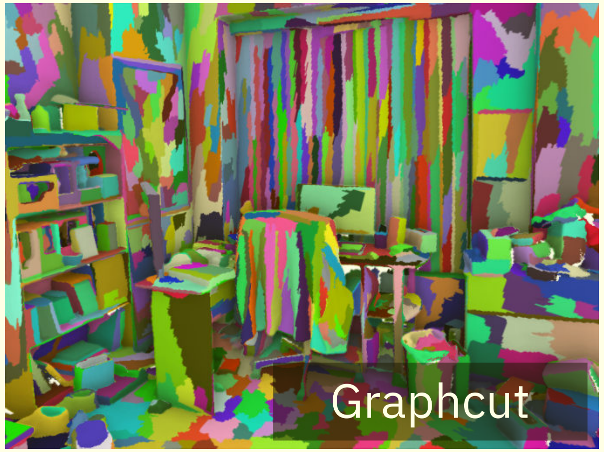
SemanticFusion, McCormac et al., ICRA 2017

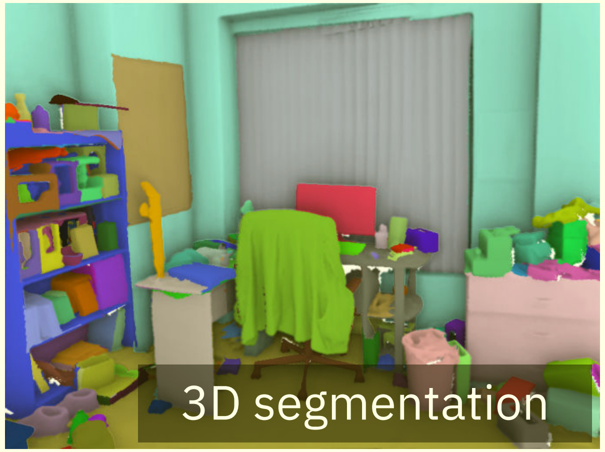
The Pipeline

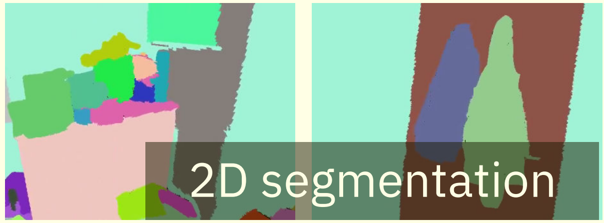
3D reconstruction



Automatic segmentation





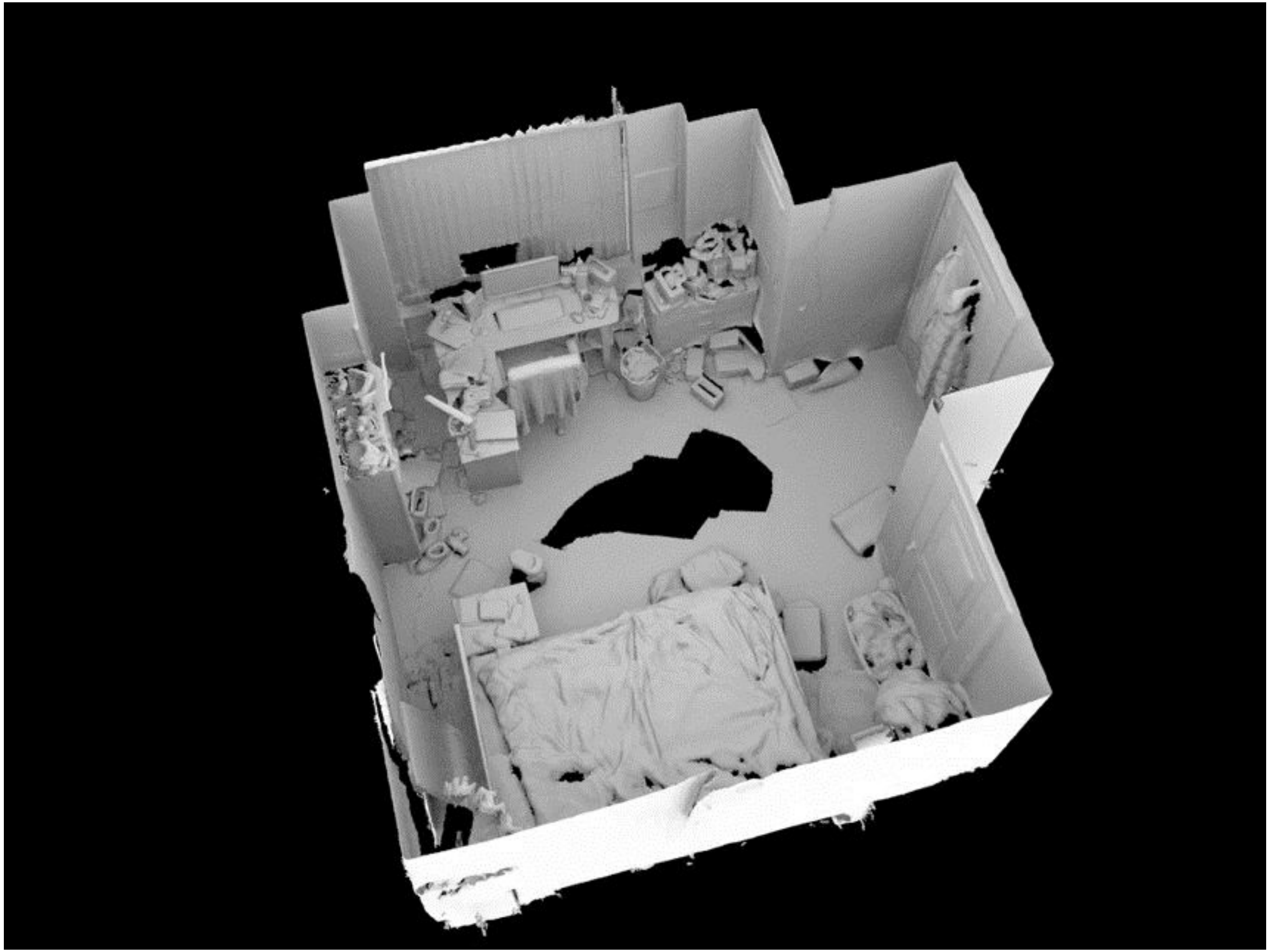


User interaction

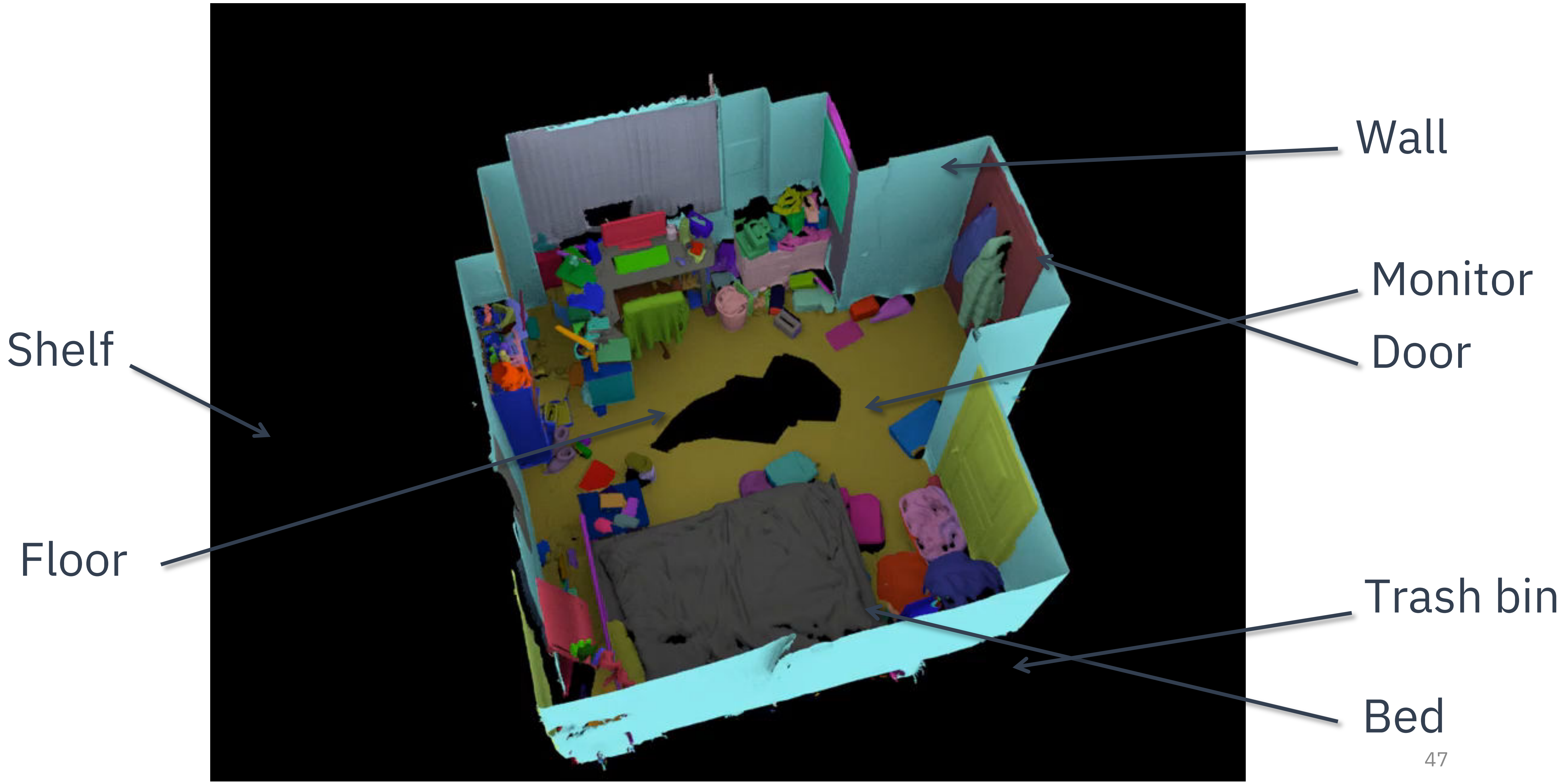
Fine-grained annotation

- 3D and 2D refinement
- Object annotation
- Object search

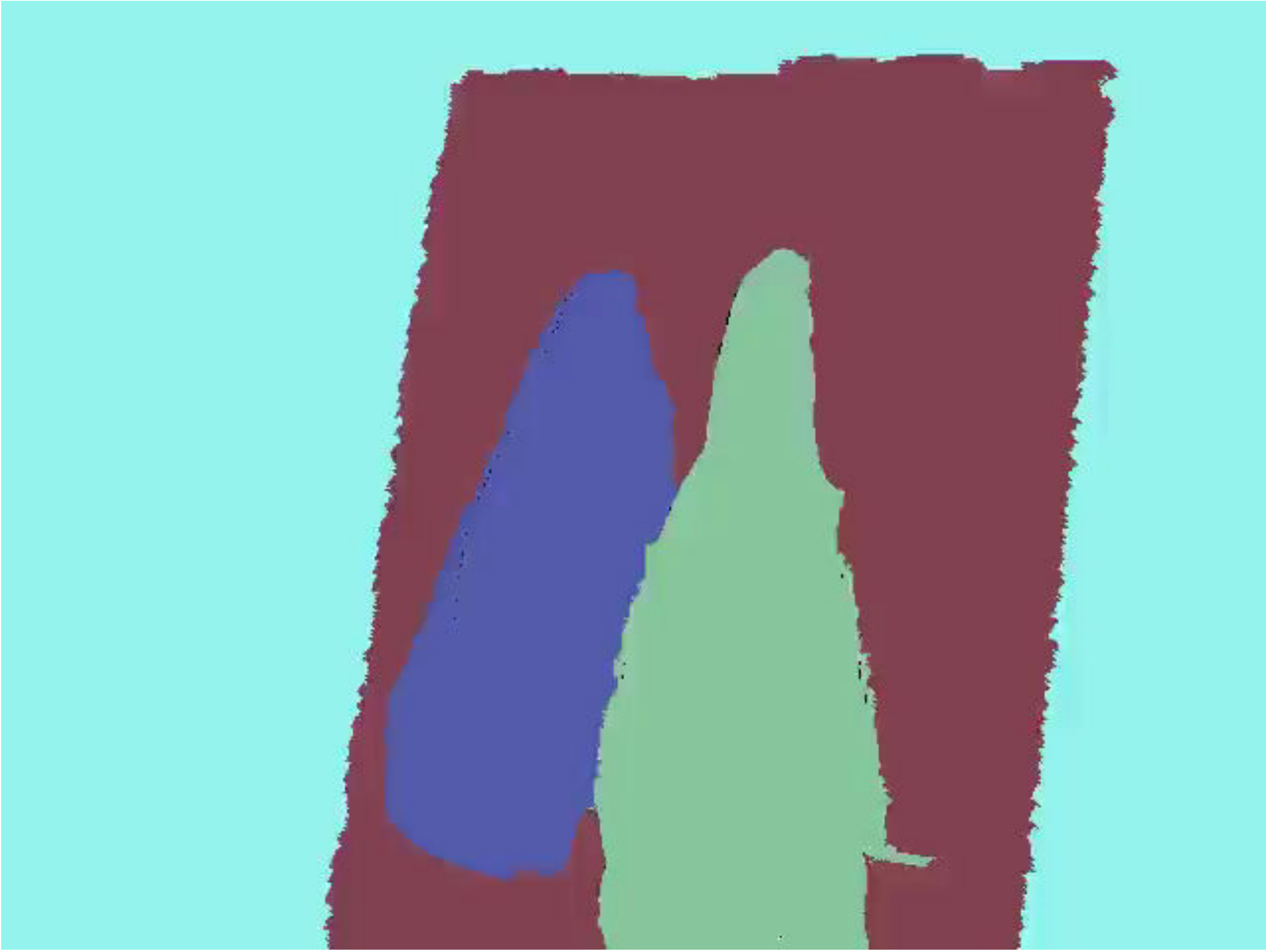
3D reconstruction



Output: 3D segmentation and annotation



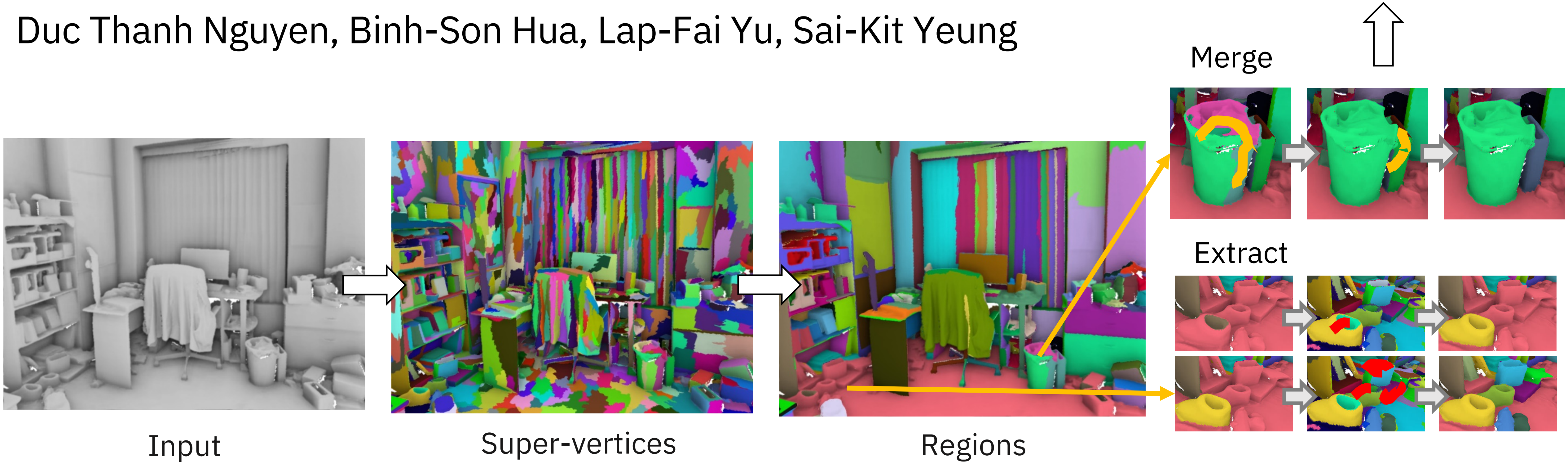
Output: 2D segmentation and annotation



3D segmentation

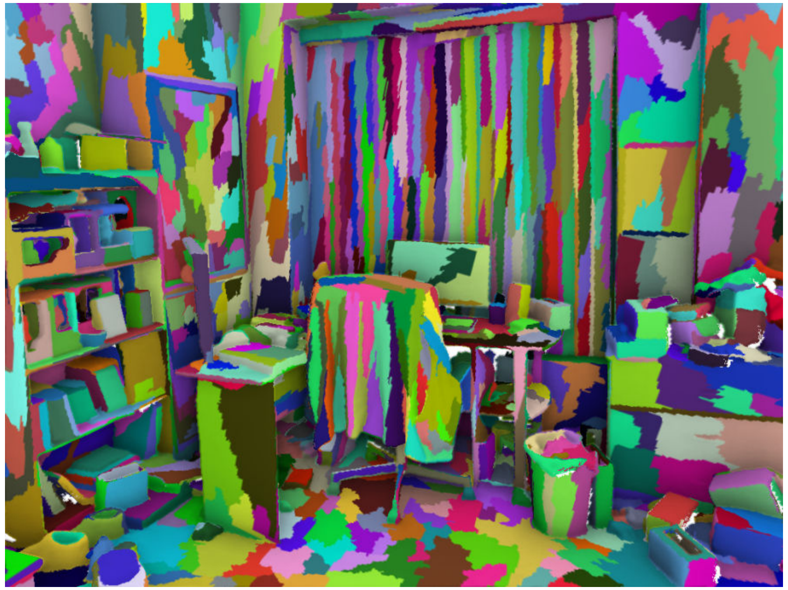
Semi-automatic Scene Annotation

A Robust 3D-2D Interactive Tool for Scene Segmentation and Annotation
TVCG 2017
Duc Thanh Nguyen, Binh-Son Hua, Lap-Fai Yu, Sai-Kit Yeung



http://scenenn.net/webgl/index.html

Bottom-up segmentation

Automatic		Interactive
		
Super-vertices	Regions	Objects

Graph-based segmentation

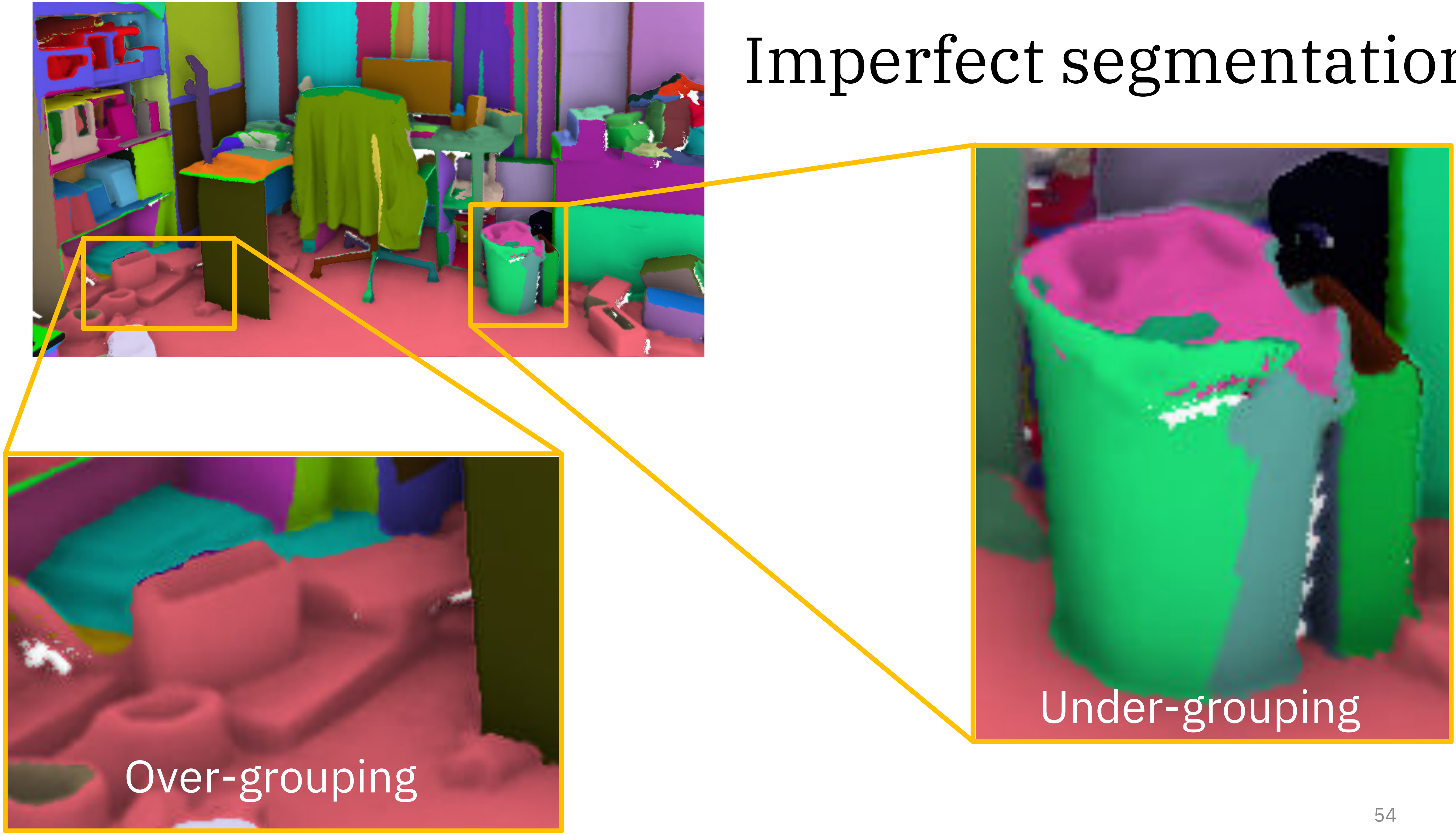
- Geometric segmentation on mesh vertices.
- Super-vertex is the smallest geometric unit to manipulate.
- Each scene has ~5000 super-vertices.

P. Felzenszwalb and D. P. Huttenlocher, "Efficient graph-based image segmentation," *IJCV 2004.*

Markov random field

- Geometric + colour segmentation on mesh vertices.
- Attempt to group similar super-vertices together.
- Each scene has ~500 regions.

Imperfect segmentation

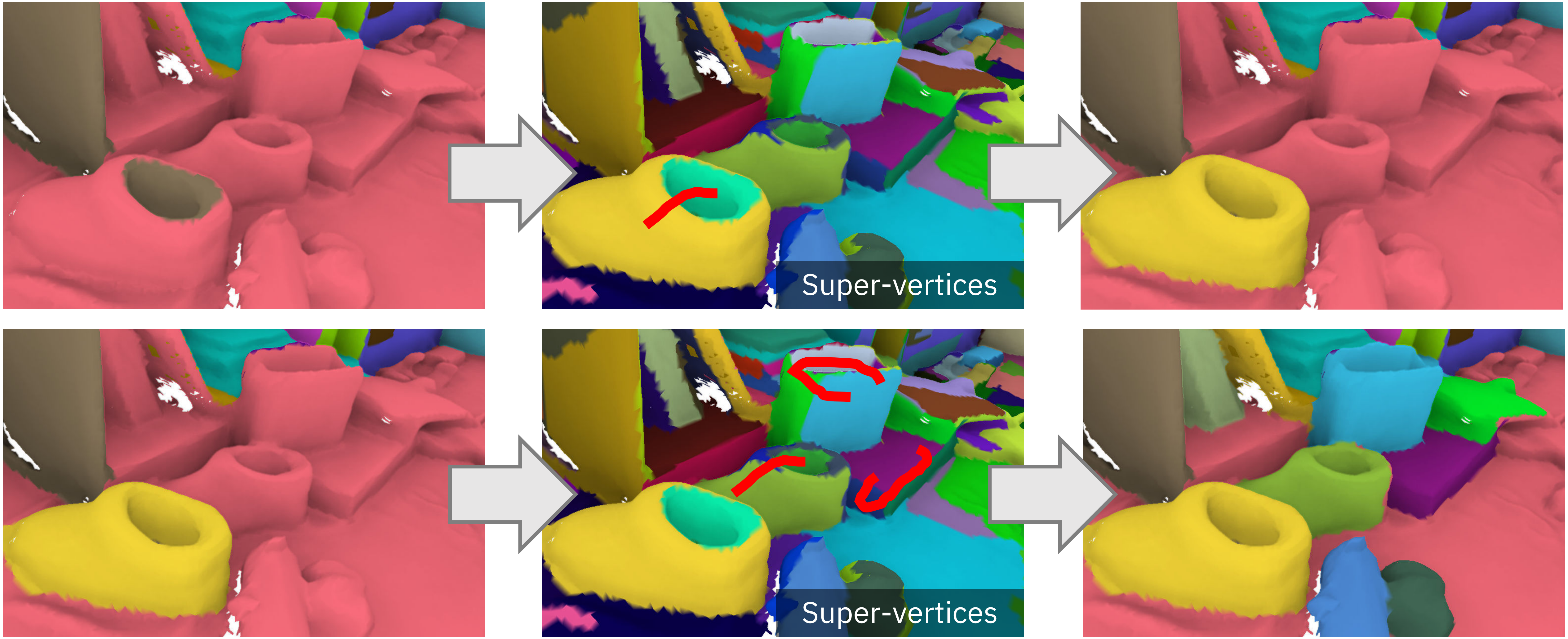


Merge

Before

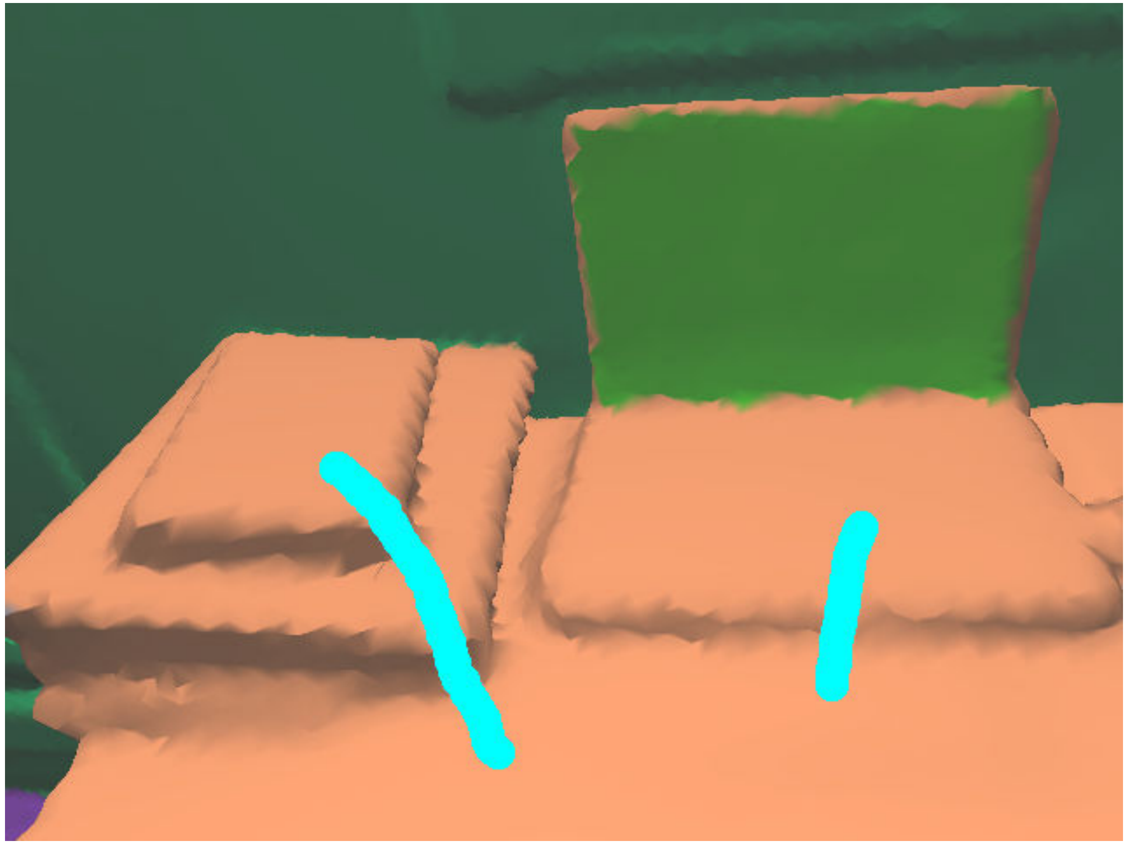
After

Extract

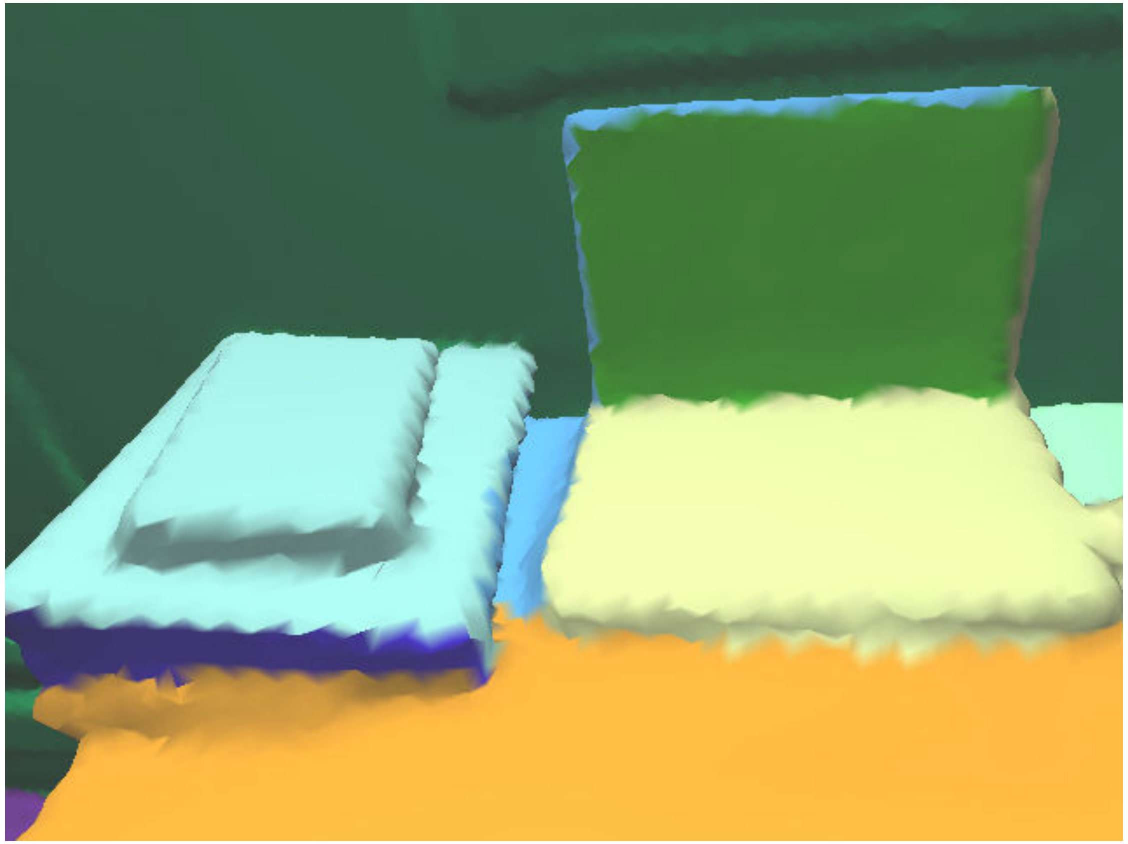


Before　　Group graph cut results　　After

Split



Before



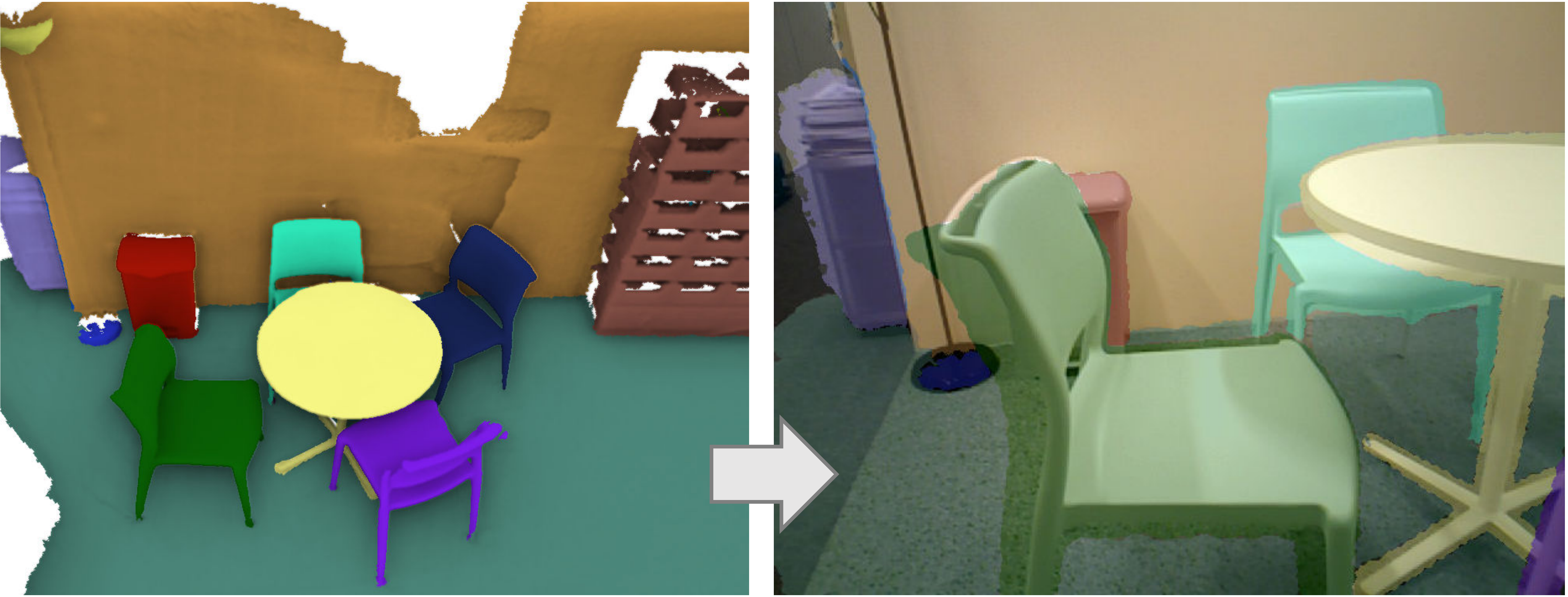
After

User interaction

- Simple operations: merge, extract, split, undo.
- Cache graph optimization results for fast switch between current regions and super-vertices.
- ~15 – 30 mins for a typical 16sqm room.

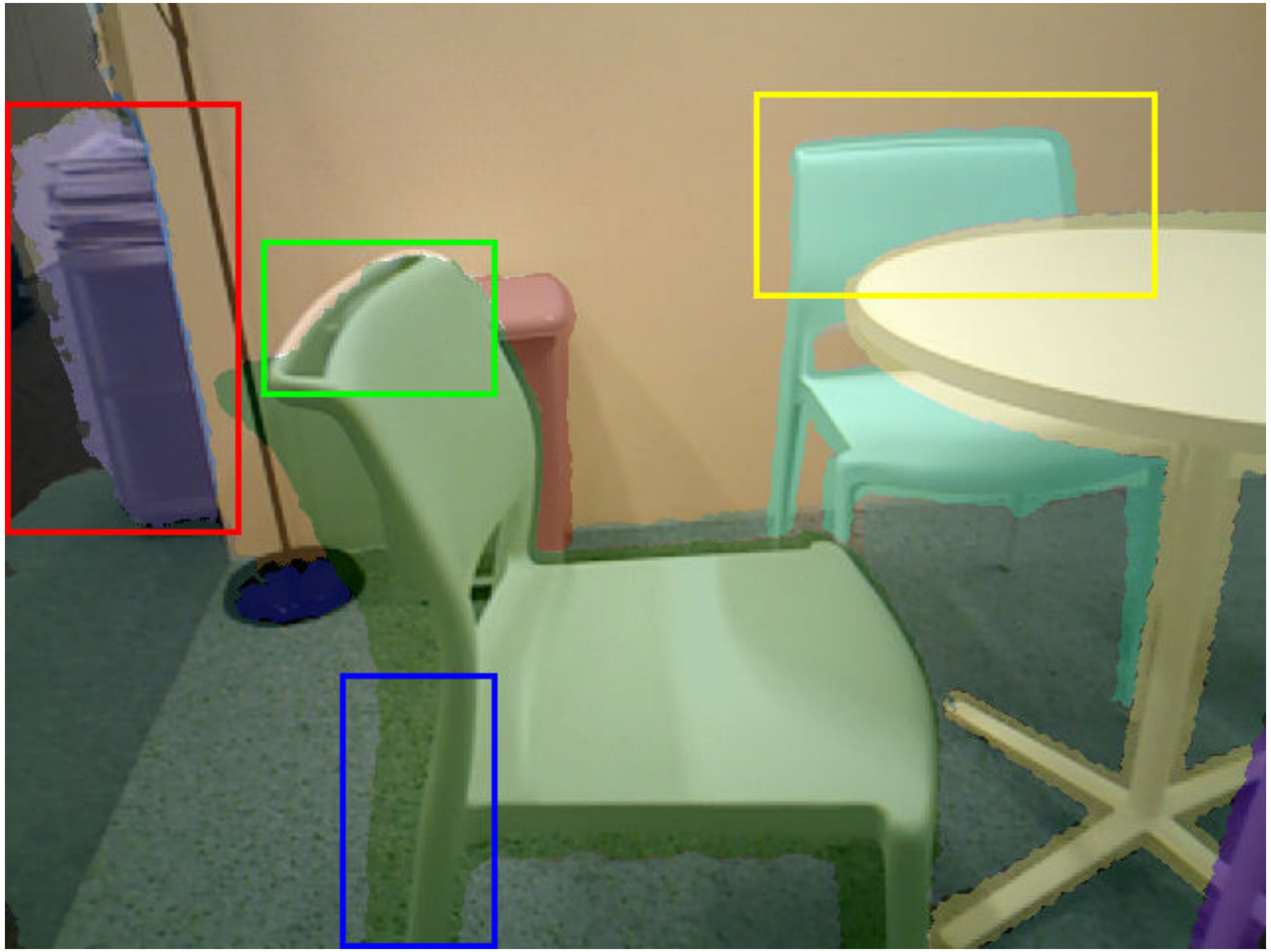
Advanced features

2D segmentation

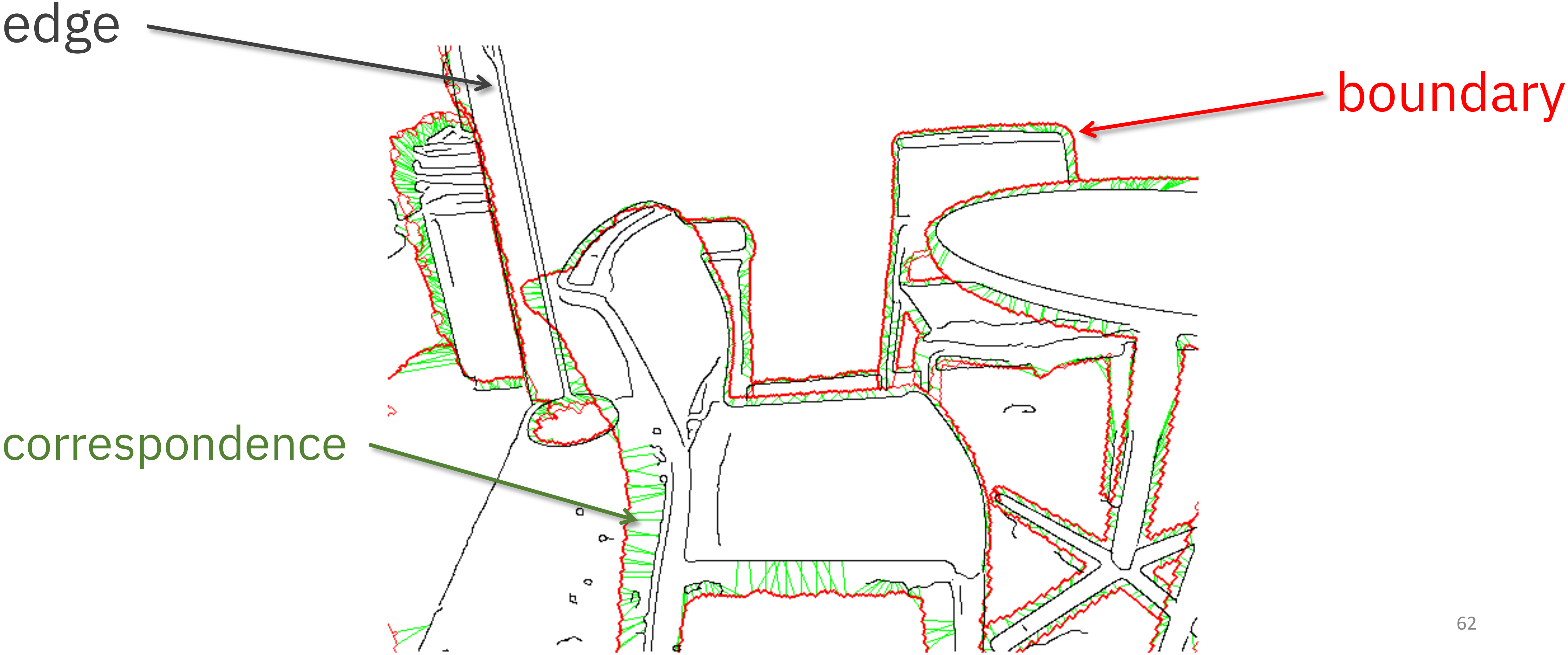


3D segmentation | 2D projection

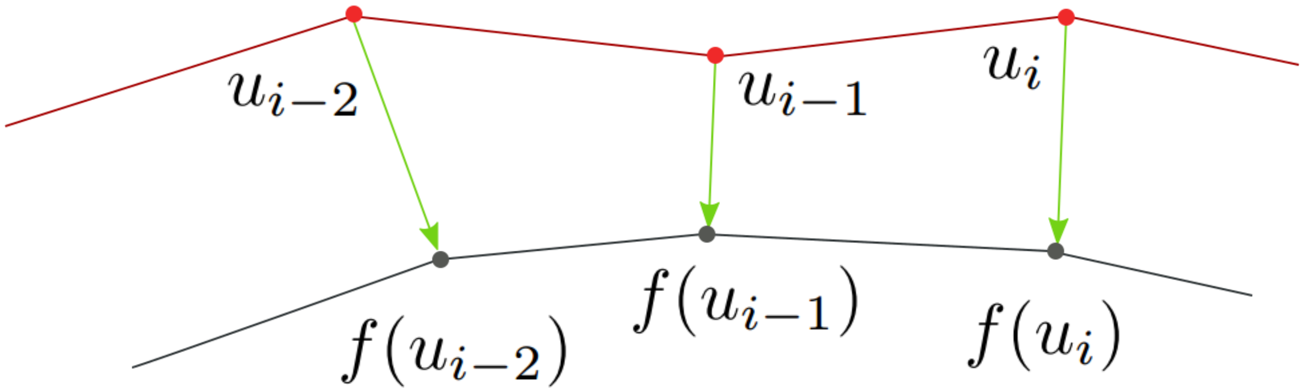
Boundary misalignment



Boundary snapping



Boundary snapping



- Find correspondences such that

$$\underset{f}{\text{minimize}}\left[\sum_{i=1}^{|U|}\chi^2(h_{u_i}, h_{f(u_i)})\right.$$

Difference of histogram of orientations

Continuity prior

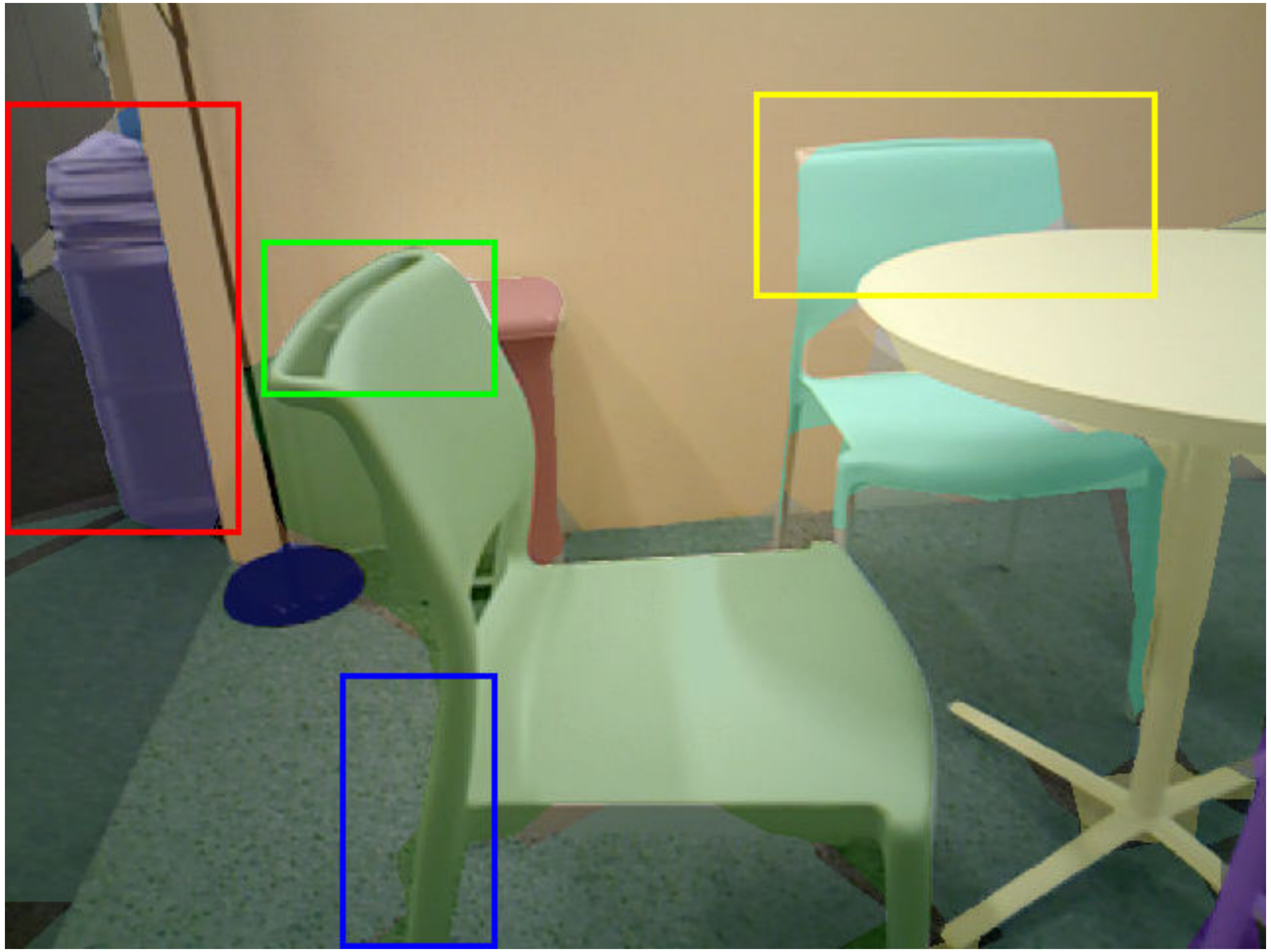
$$+\,\kappa_1\sum_{i=2}^{|U|}\|f(u_i)-f(u_{i-1})\|$$

Smoothness prior

$$\left.+\,\kappa_2\sum_{i=3}^{|U|}\cos(f(u_i)-f(u_{i-1}), f(u_{i-2})-f(u_{i-1}))\right]$$

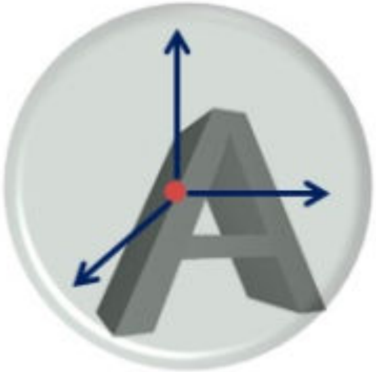
Optimization details in the paper

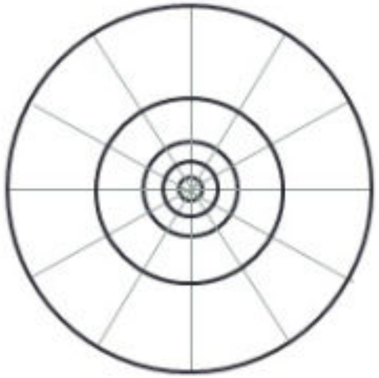
Snapping result



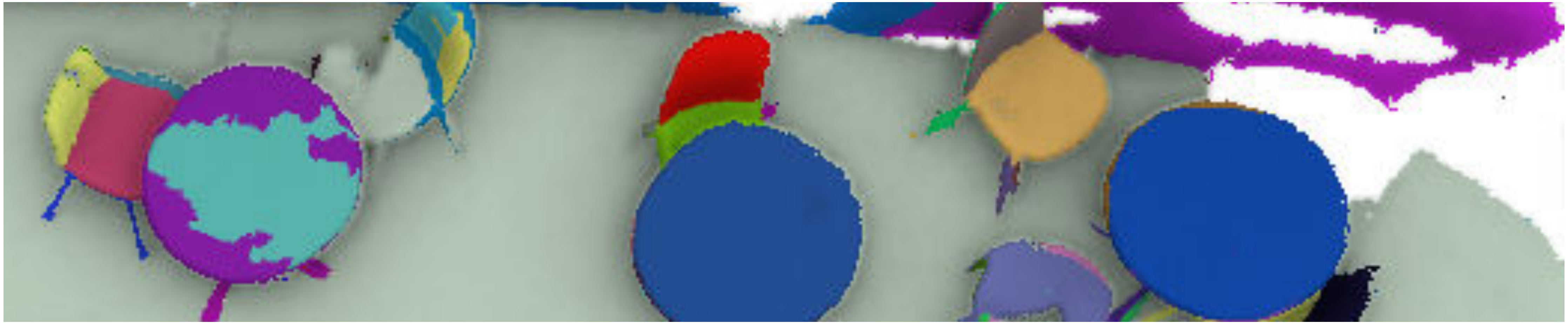
Object search

Template-based object search

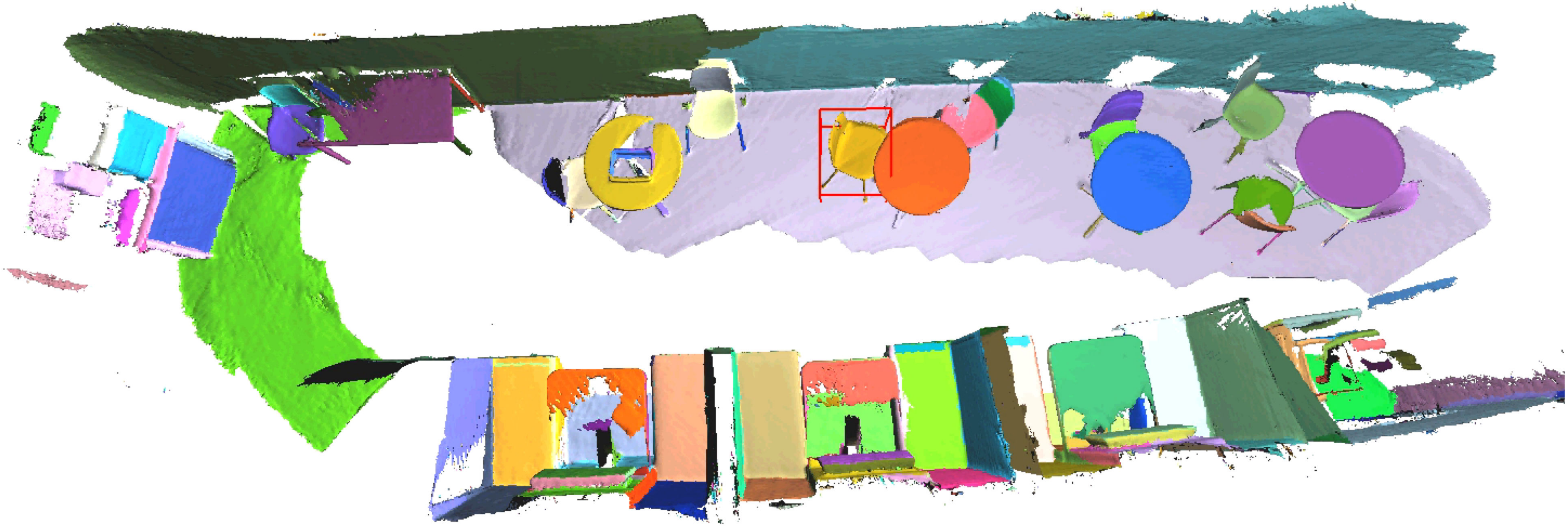




- 3D shape context descriptor
- Sliding window search
- For each candidate region:
  Apply a greedy grow-shrink procedure to find the best combination of labels



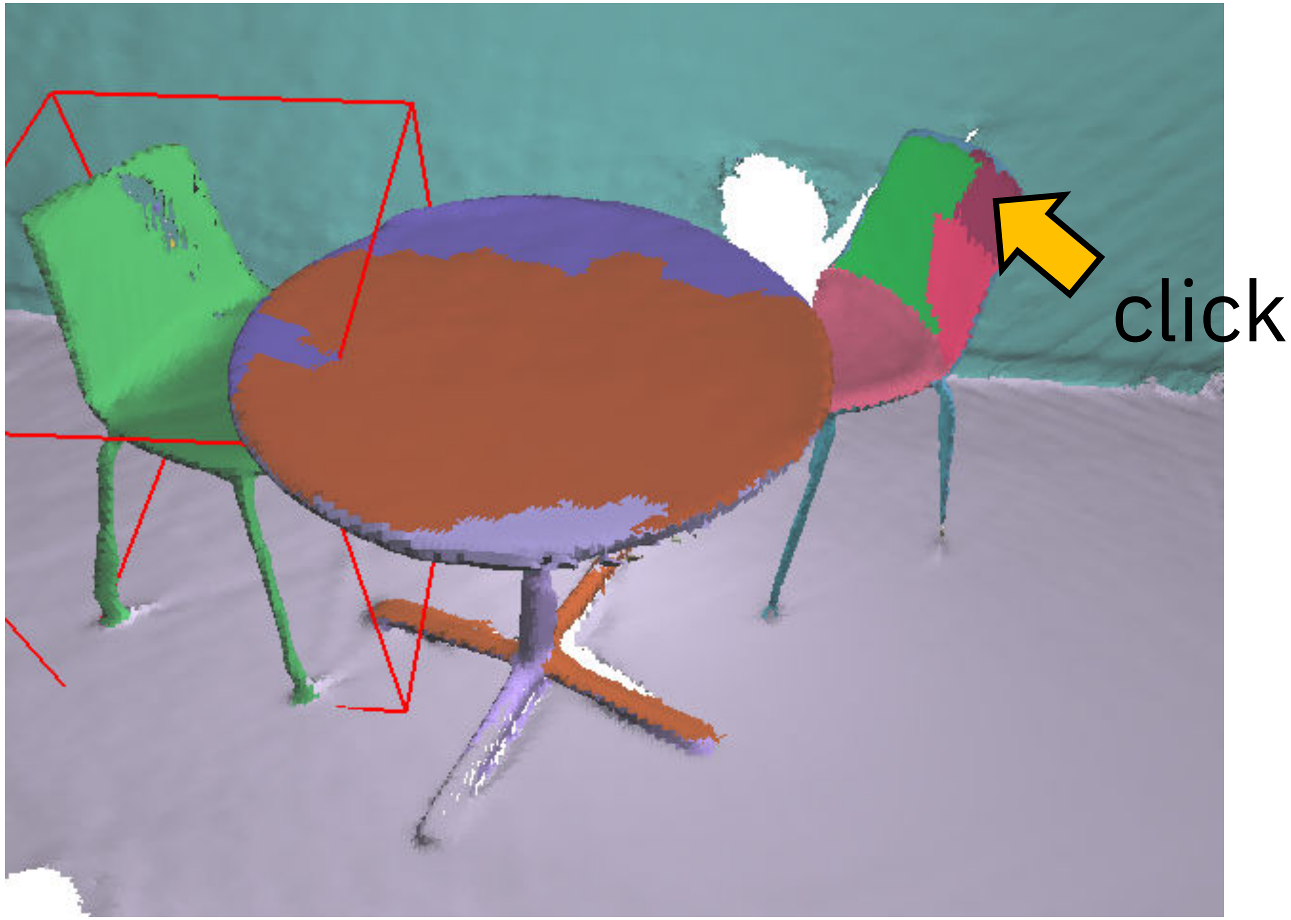
Template-based object search



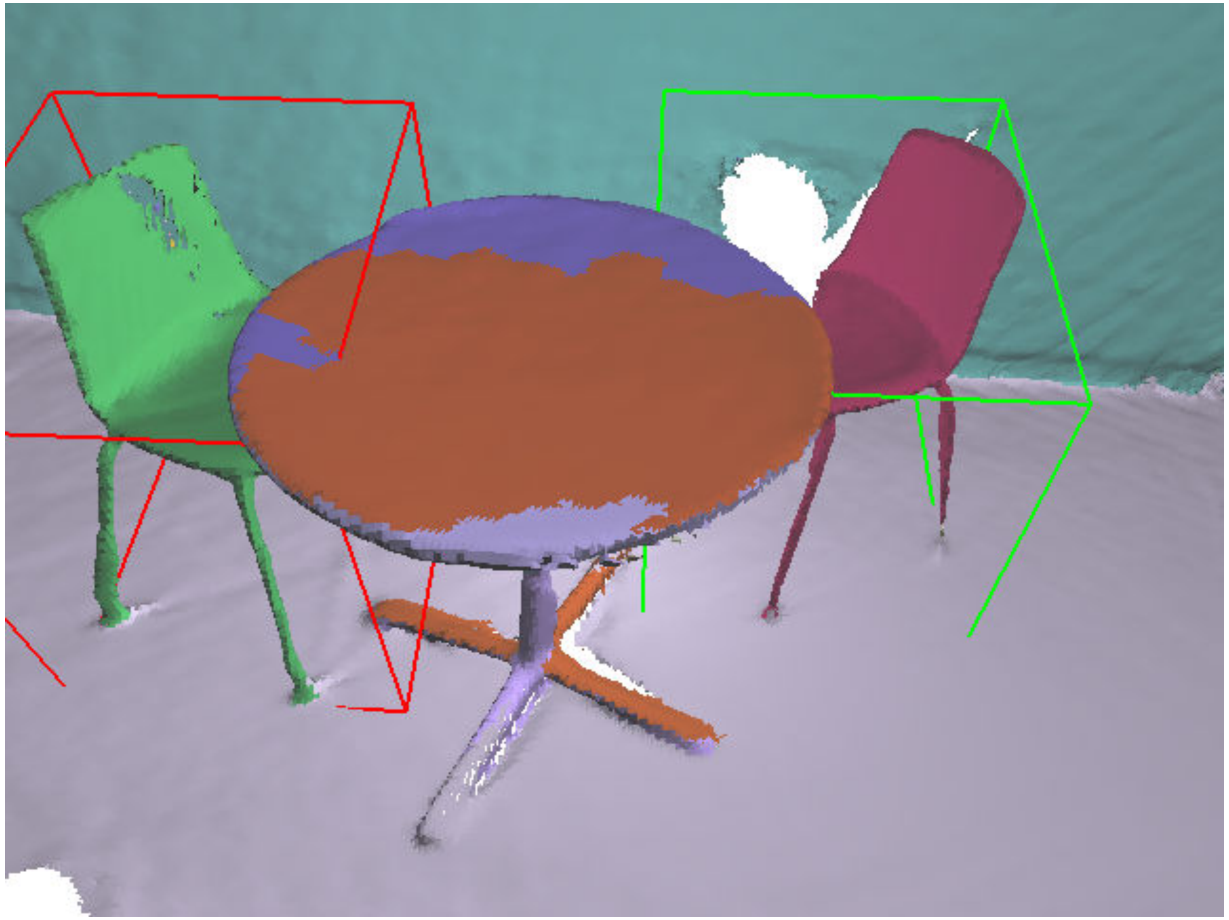
Guided merge

Apply grow-shrink after user interaction

Guided merge

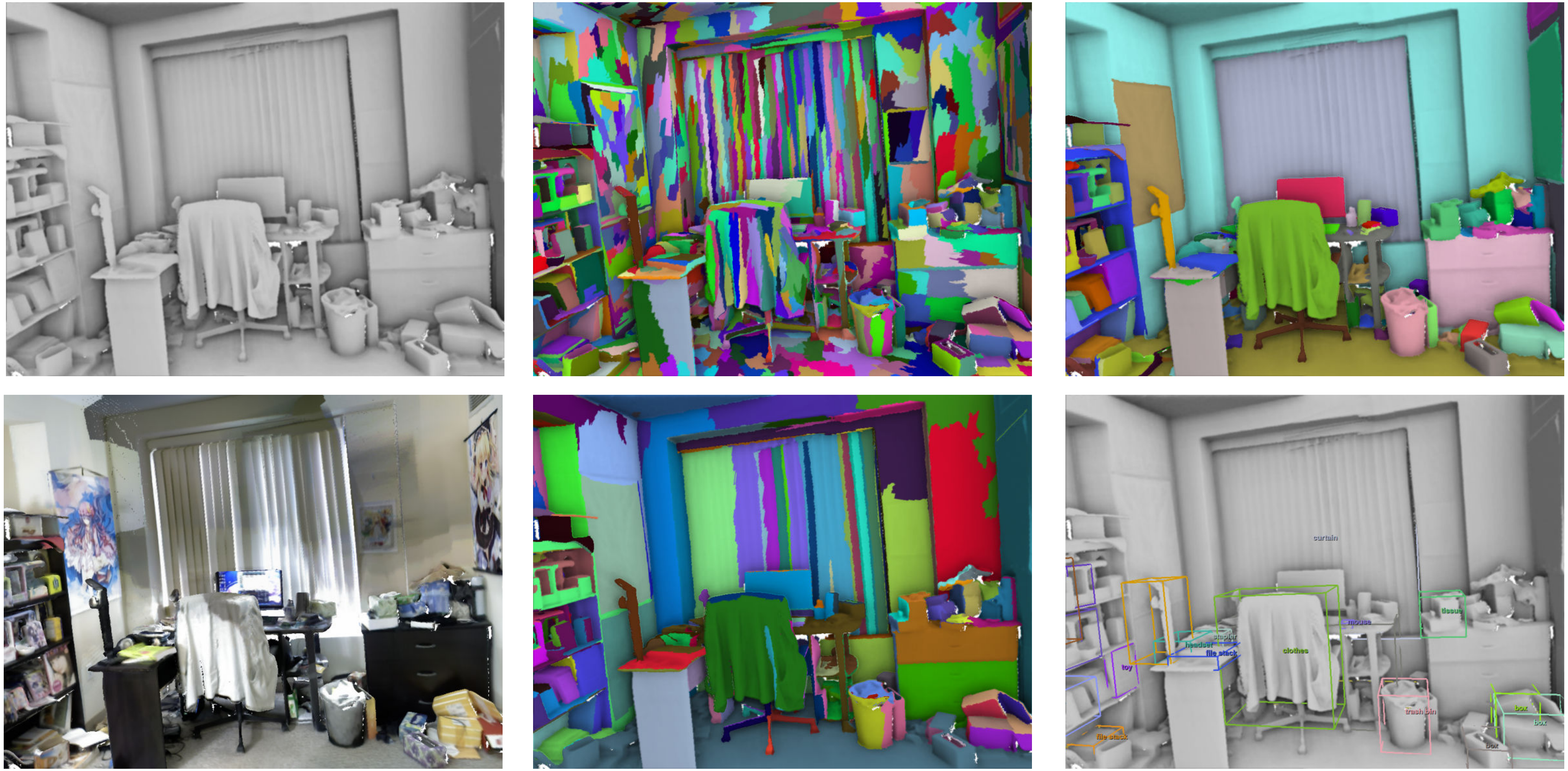


Guided merge



Experiments

SceneNN dataset annotation

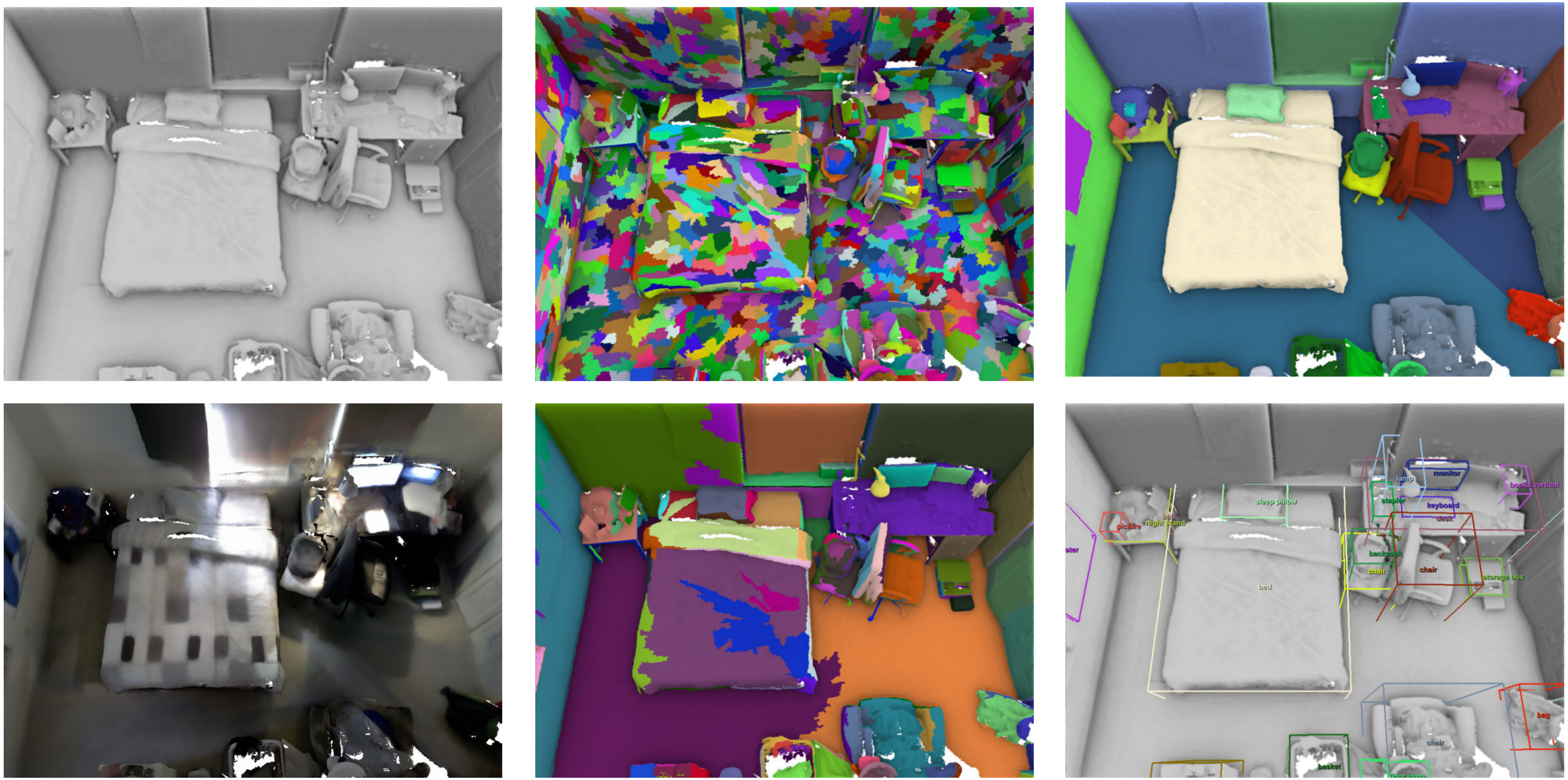


Reconstruction

Automatic segmentation

Refined segmentation

SceneNN dataset annotation

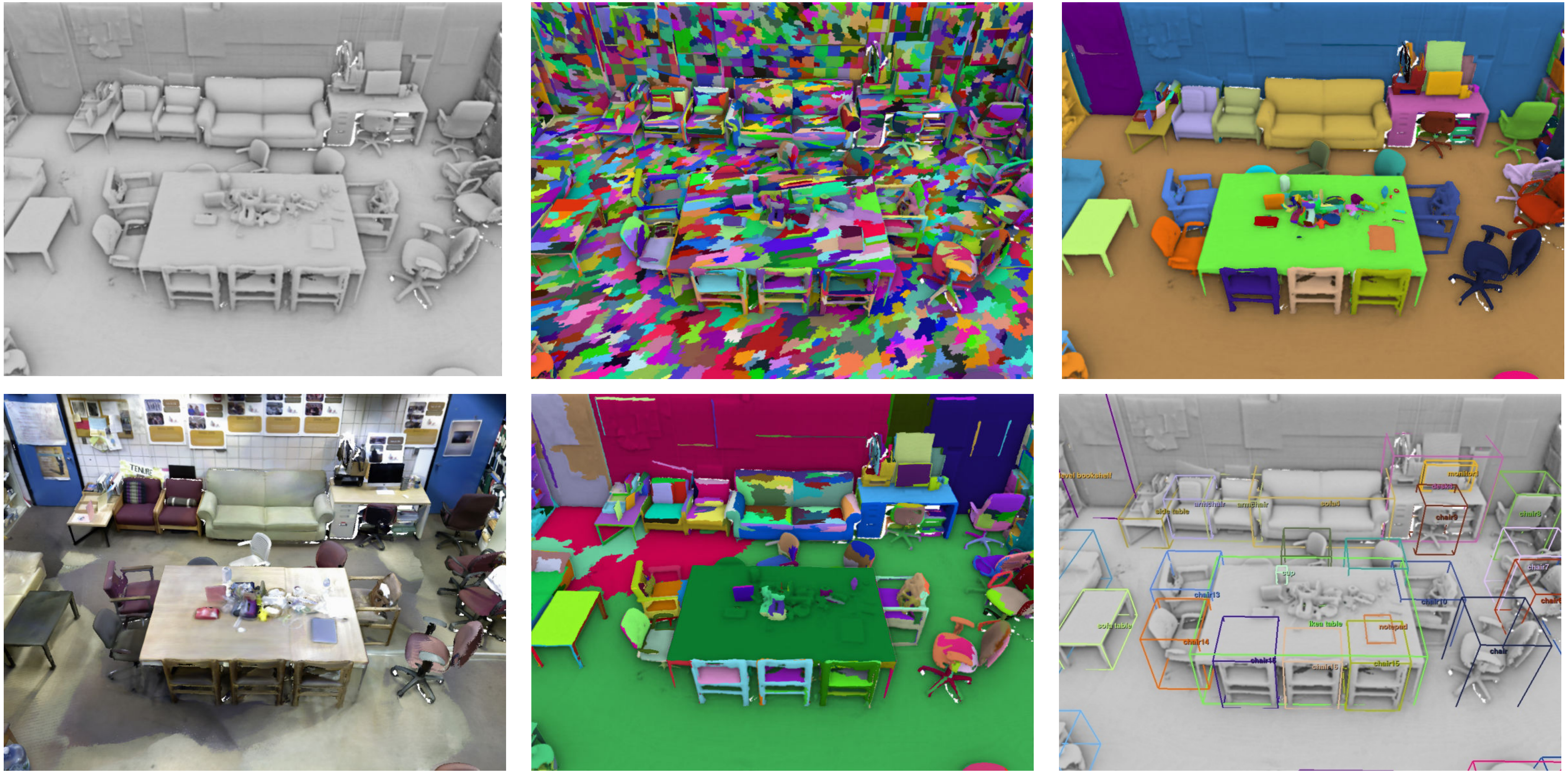


Reconstruction

Automatic segmentation

Refined segmentation

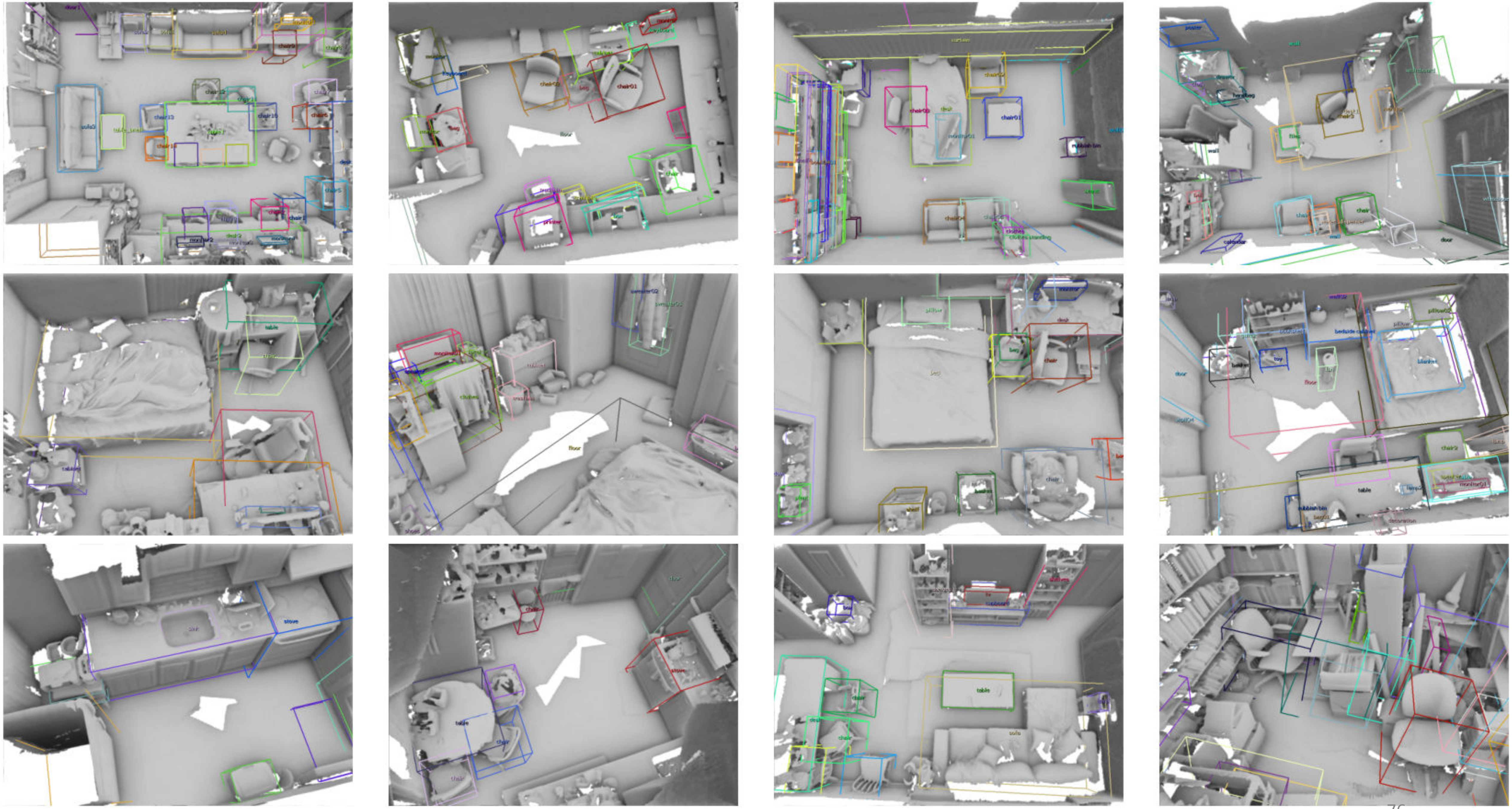
SceneNN dataset annotation

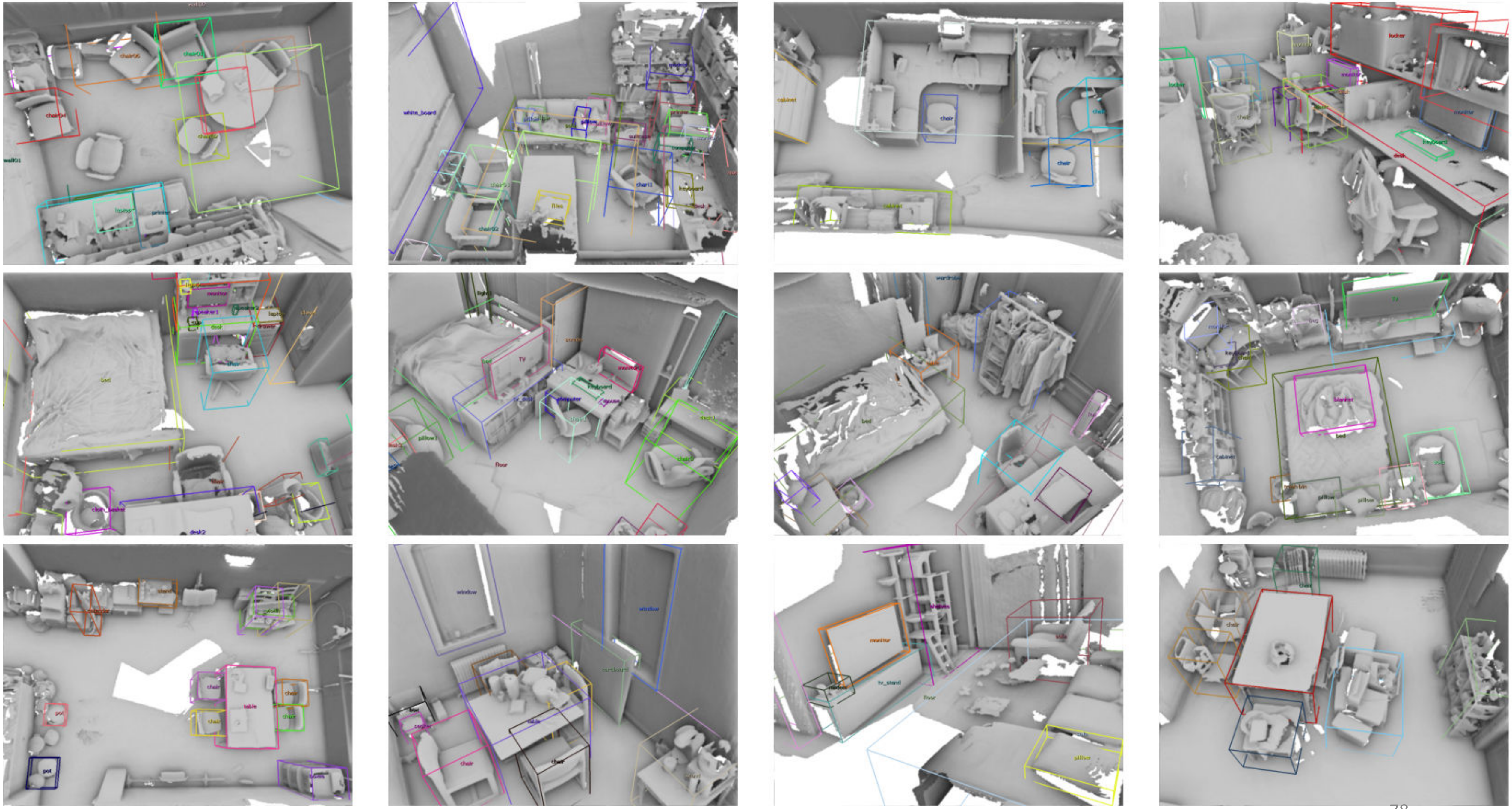


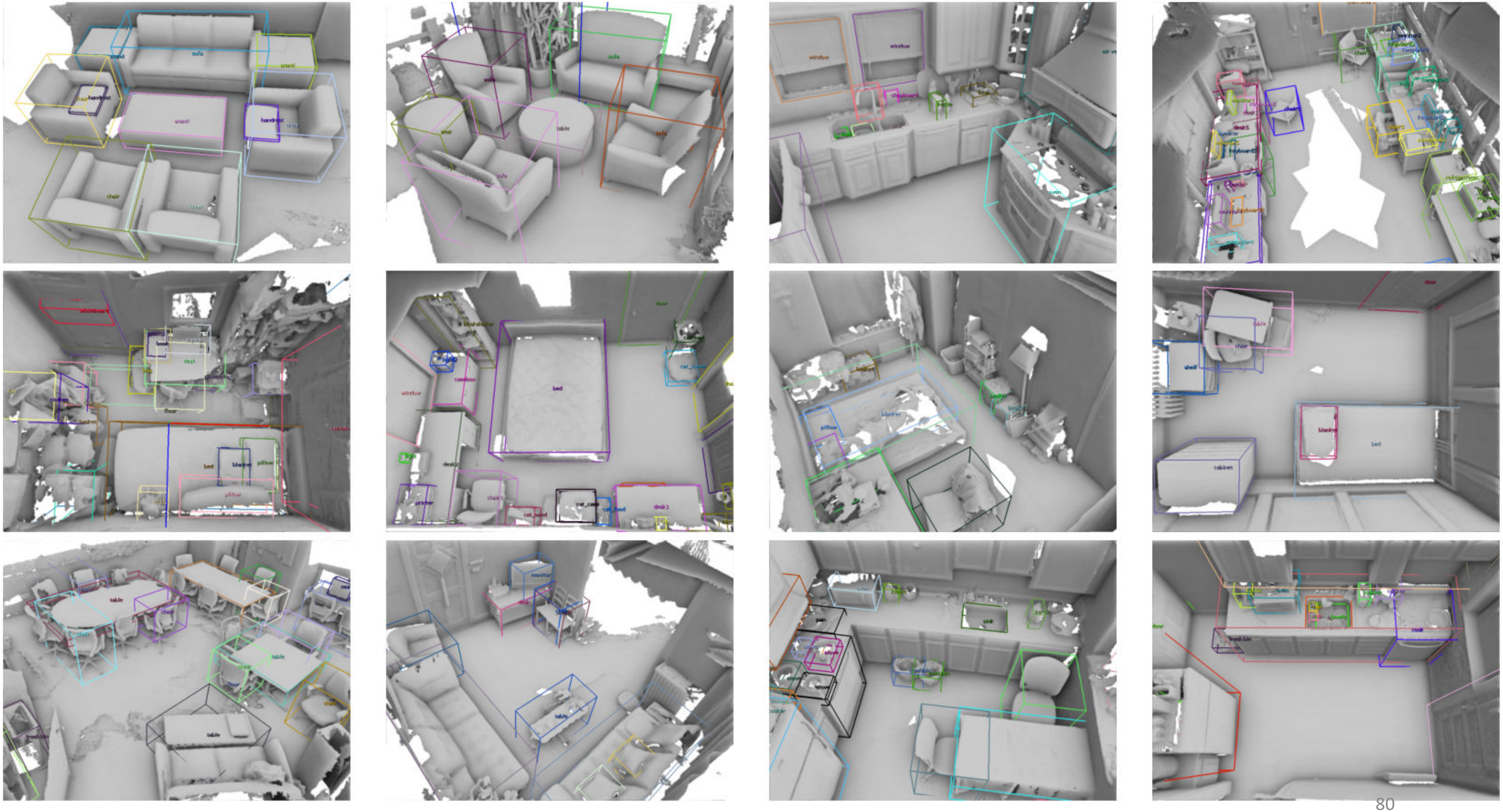
Reconstruction

Automatic segmentation

Refined segmentation







Outdoor Scene Annotation

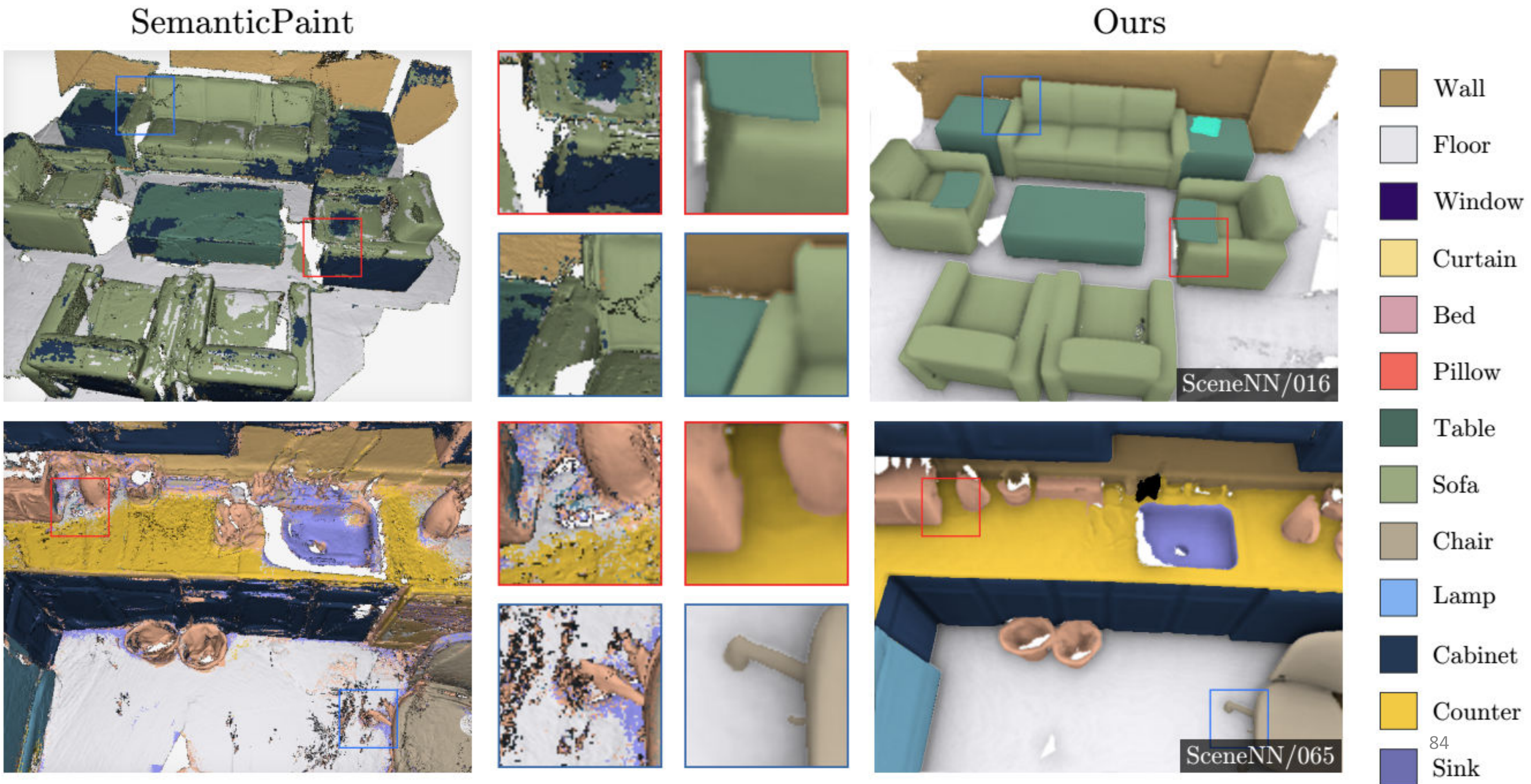
Automatic segmentation statistics

		Graph-based			MRF-based		
Scene	#Vertices	#Supervertices	OCE	Time (seconds)	#Regions	OCE	Time (seconds)
copyroom	1,309,421	1,996	0.92	1.0	347	0.73	10.9
lounge	1,597,553	2,554	0.97	1.1	506	0.93	7.3
hotel	3,572,776	13,839	0.98	2.7	1433	0.88	17.8
dorm	1,823,483	3,276	0.97	1.2	363	0.78	7.8
kitchen	2,557,593	4,640	0.97	1.8	470	0.85	12.2
office	2,349,679	4,026	0.97	1.7	422	0.84	10.9
Our scenes	1,450,748	2,498	0.93	1.4	481	0.77	12.1

User interaction statistics

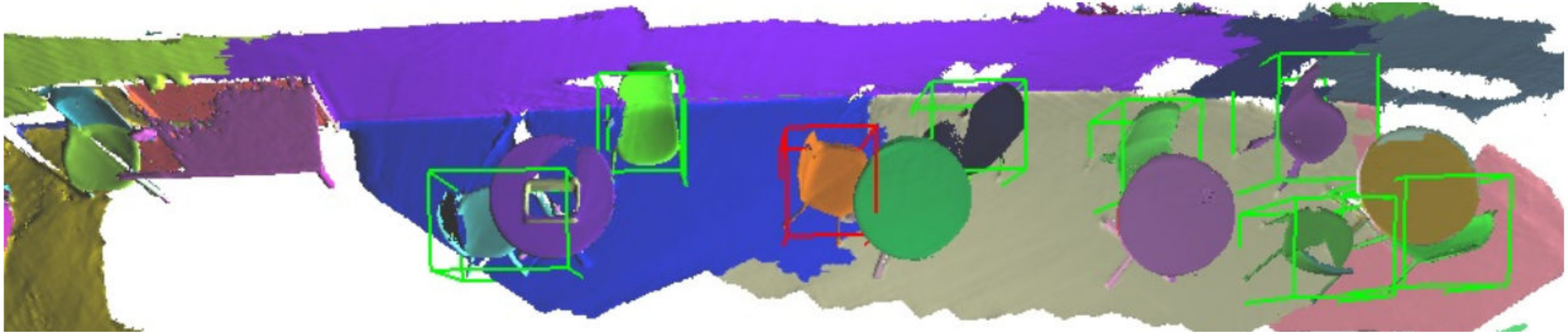
Scene	#Vertices	User refined #Labels	#Objects	Interactive time (minutes)
copyroom	1,309,421	157	15	19
lounge	1,597,553	53	12	16
hotel	3,572,776	96	21	27
dorm	1,823,483	75	10	15
kitchen	2,557,593	75	24	23
office	2,349,679	69	19	24
Our scenes	1,450,748	179	19	30

Comparison to SemanticPaint



Object search evaluation

- 45 objects in two categories, chair and table.
- Each object is used as a template.
- 69% precision and 70% recall.
- Template represented by 150 points.
- Search completes within 15 seconds.

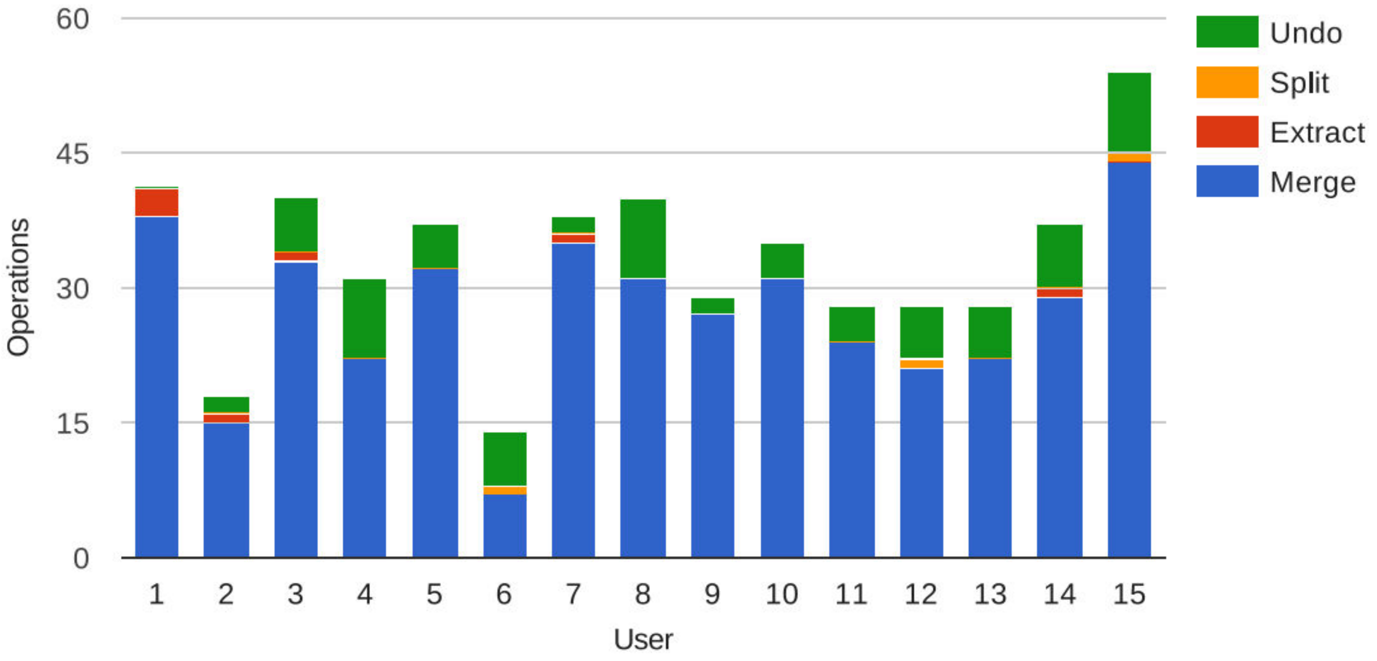


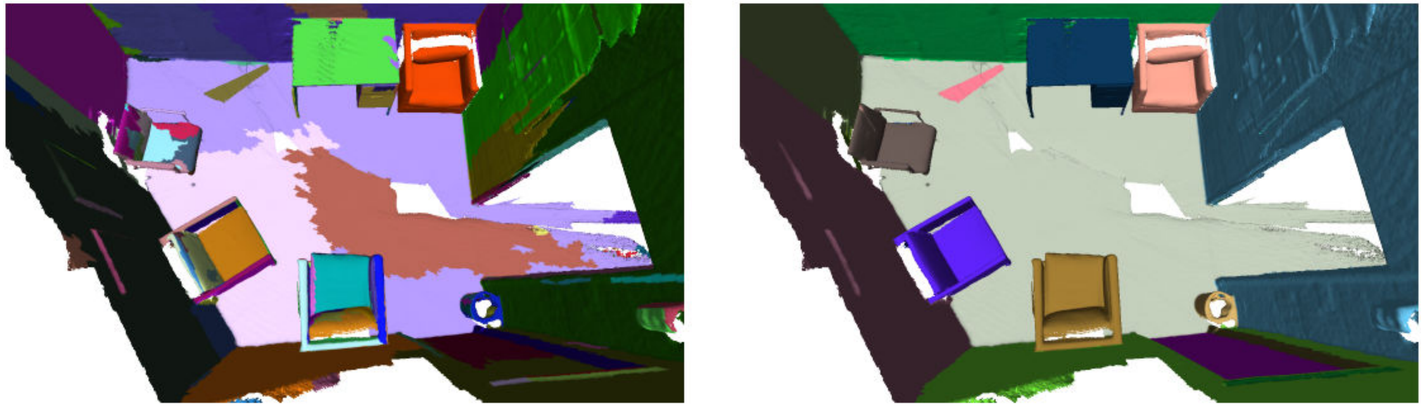
Boundary snapping evaluation

Segmentation method	OCE
Projection	0.57
Local shape	0.60
Local shape + Continuity	0.55
Local shape + Smoothness	0.55
Local shape + Continuity + Smoothness	**0.54**

User study

- To measure how merge and extract are used in practice.
- Task A: Simple scene, 2 minutes.
- Merge is dominant.



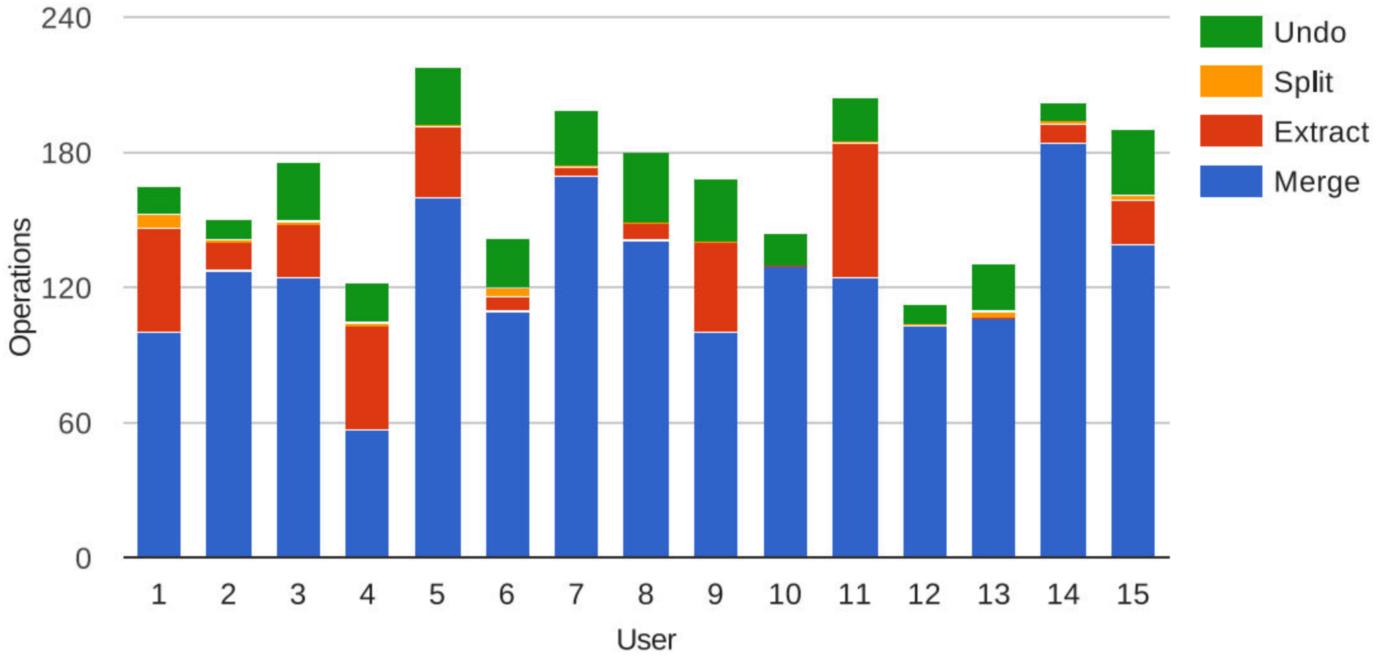


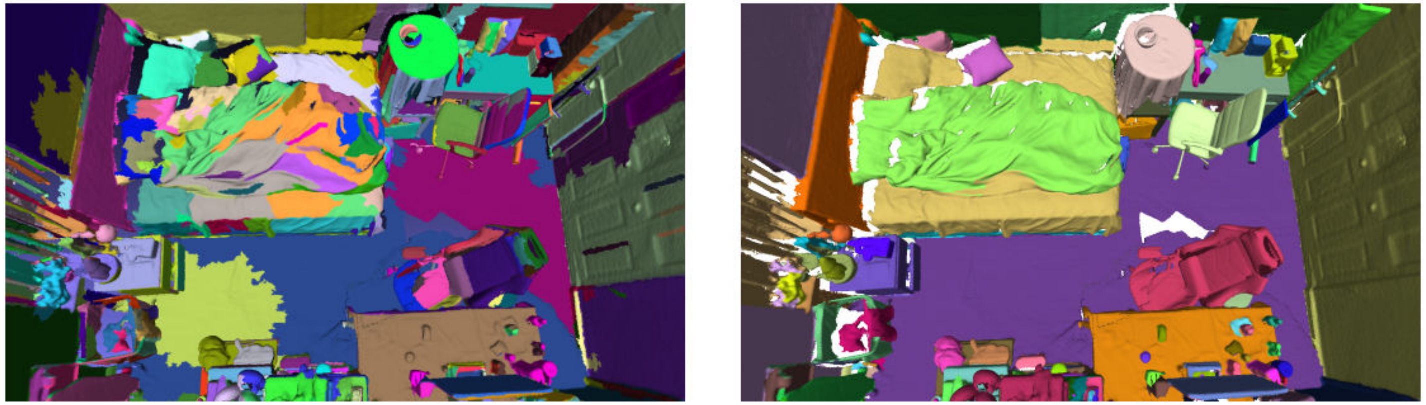
(a) Task A (two minutes)

User study

- Task B:
  Complex scene,
  10 minutes.

- More extracts used in complex scenes.





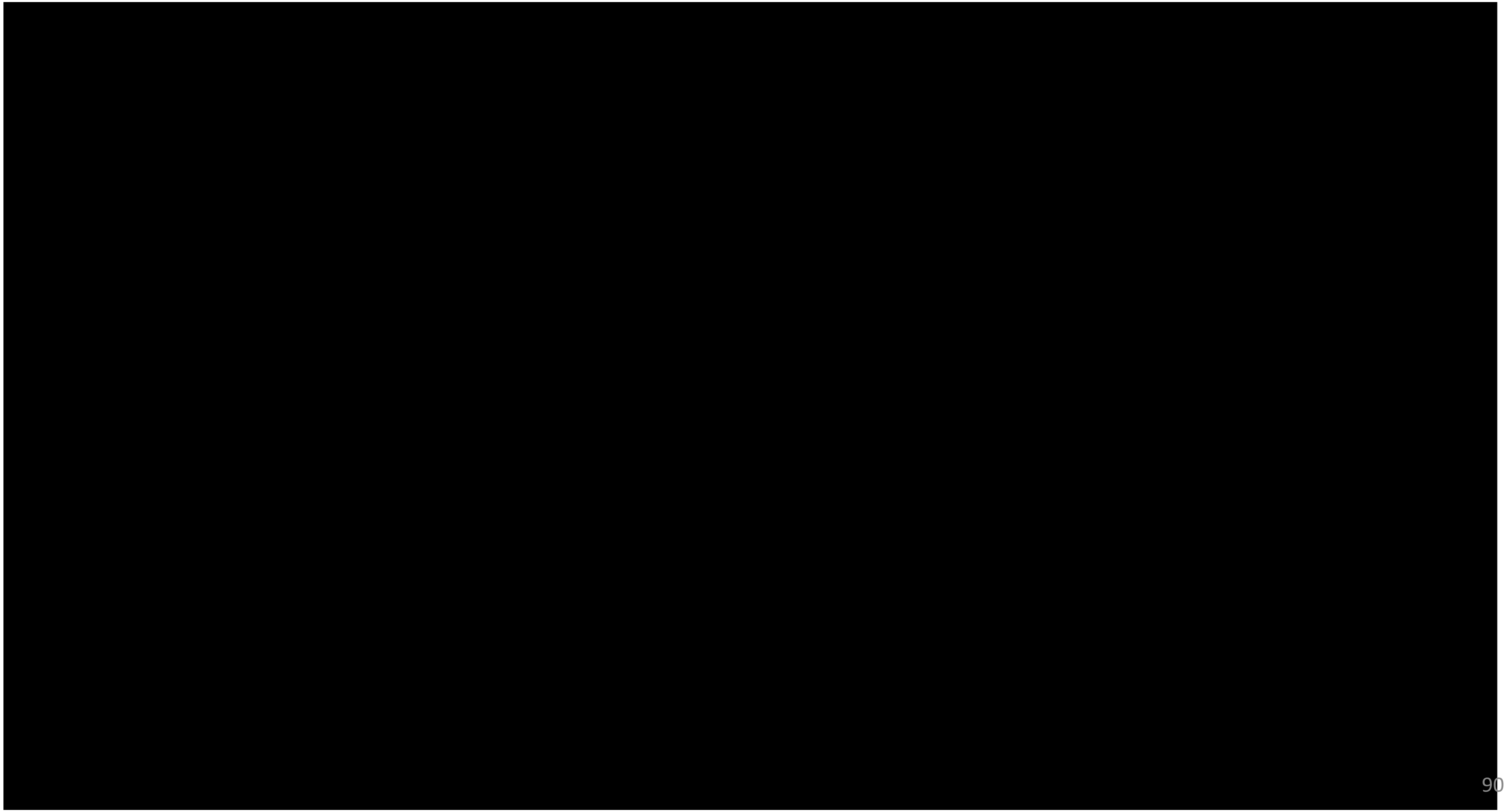
(b) Task B (ten minutes)

WebGL annotation tool http://scenenn.net/webgl/index.html

- Reimplementation of our original C++ annotation tool.
- Graph cut, MRF with merge and extract.
- Open source.

WebGL annotation tool

http://scenenn.net/webgl/index.html



Future work

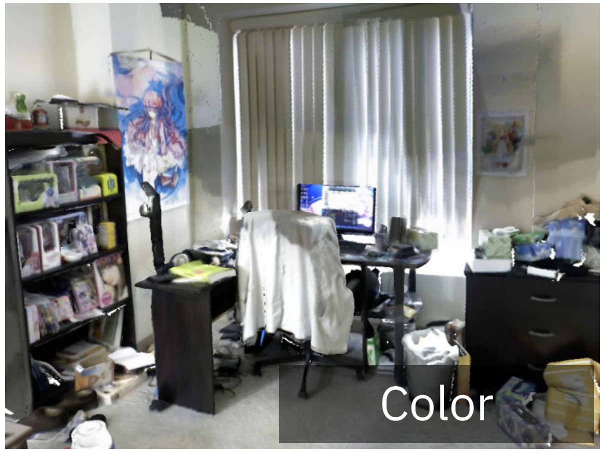
- Less time, more quality.
  Initial segmentation powered by a deep network.
- Better user experience.
  Online learning of user operations to reduce undo.
- Annotate more scenes for 3D deep learning.

Part III:
Dataset and Applications

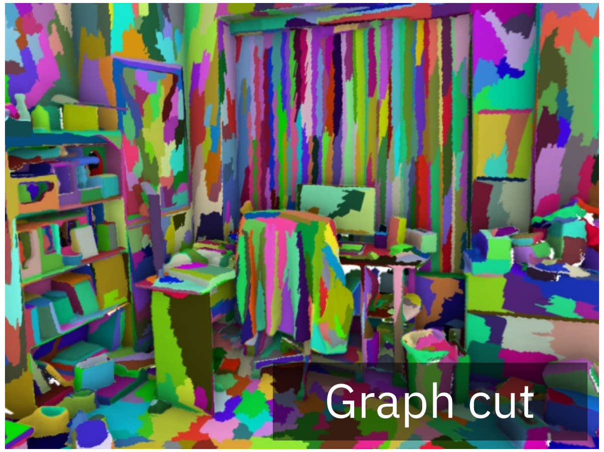
SIGGRAPH
ASIA 2019
BRISBANE

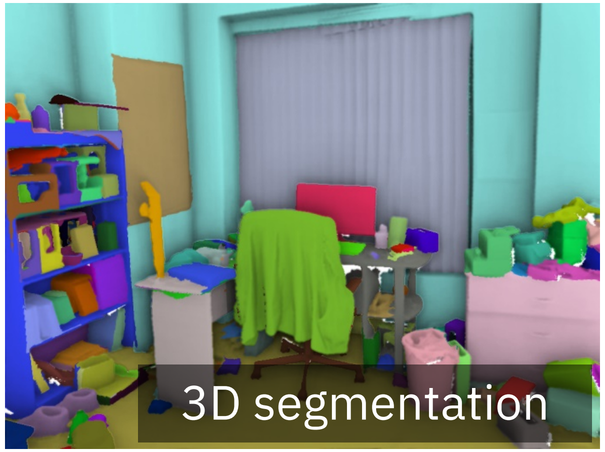
System overview

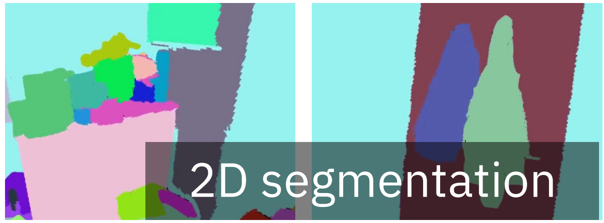
3D reconstruction



Automatic segmentation







User interaction

Fine-grained annotation

- 3D and 2D refinement
- Object annotation
- Object search

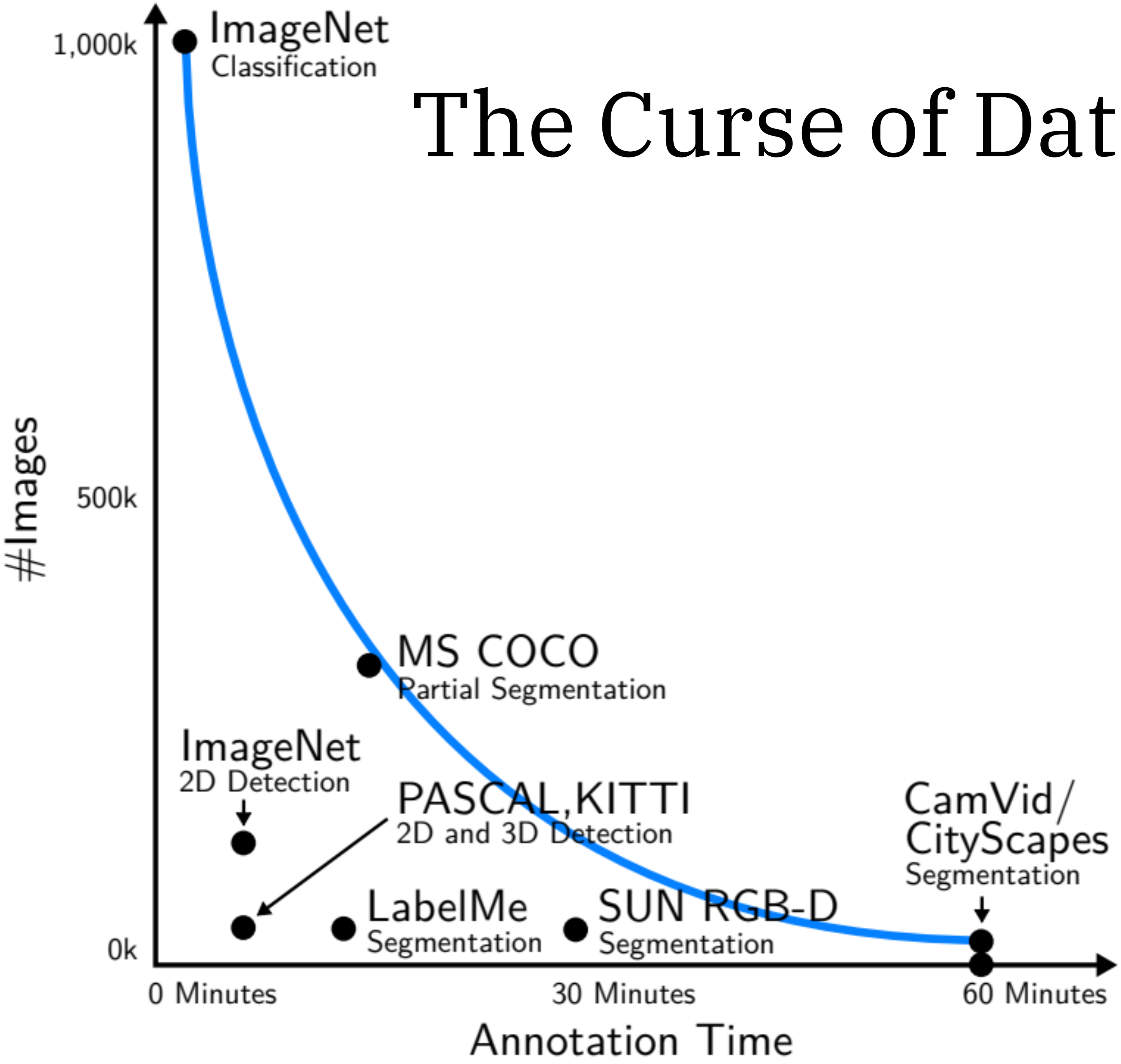
Building an Effective Pipeline

- Define input and output
- Define components to process the input and generate output
- Determine the state-of-the-art techniques for each component
- The selected techniques should balance between quality, speed, and scalability.

Logistics for an Effective Pipeline

- Know your team and manpower.
- Estimate time to annotate 1 sample.
- Dry run, feedback, improve the pipeline.
- Annotate and validate.
- Scale to mass annotation.
- Re-annotate.

The Curse of Dataset Annotation



Courtesy of Xie et al.,
Semantic Instance Annotation of Street Scenes by 3D to 2D Label Transfer, CVPR 2016

Scene and Object Datasets since 2012

				Redwood	
		ICL-NUIM		ObjectNet3D	
		RGB-D v2	ShapeNet	DROT	Matterport3D
		BigBIRD	3D ShapeNets	GMU Kitchen	SunCG
	SUN3D	PASCAL3D+	SUN RGB-D	SceneNet	Semantic3D
NYU	KITTI	COCO	ViDRILO	CoRBS	S3DIS
TUM	IKEA	MV-RED	YCB	Rutgers APC	ScanNet
2012	**2013**	**2014**	**2015**	**2016**	**2017**

Scene datasets

Dataset	Quantity	Annotation	Format	Pose
NYU v2	1449 frames	All	Image	N
SUN RGB-D	10K frames	All	Image	N
RGB-D v2	17 scenes	All	Cloud	Y
TUM	47 scenes	N.A.	Image	Y
SUN3D	254 scenes	8 scenes	Cloud	Y
Ours	100 scenes	All	Mesh	Y

SceneNN: A Scene Meshes Dataset with aNNotations

3DV 2016

Binh-Son Hua, Quang-Hieu Pham, Duc Thanh Nguyen,
Minh-Khoi Tran, Lap-Fai Yu, Sai-Kit Yeung

Best paper honorable mention

- 100+ scene meshes (offices, dorms, classrooms, bedrooms, kitchens)
- Captured from UMass Boston, SUTD

SceneNN: A Scene Meshes Dataset with aNNotations

Binh-Son Hua[1], Quang-Hieu Pham[1], Duc Thanh Nguyen[2], Minh-Khoi Tran[1], Lap-Fai Yu[3], and Sai-Kit Yeung[1]

[1]Singapore University of Technology and Design [2]Deakin University [3]University of Massachusetts Boston

We introduce an RGB-D scene dataset consisting of more than 100 indoor scenes. Our scenes are captured at various places, e.g., offices, dormitory, classrooms, pantry, etc., from University of Massachusetts Boston and Singapore University of Technology and Design.
All scenes are reconstructed into triangle meshes and have per-vertex and per-pixel annotation. We further enriched the dataset with fine-grained information such as axis-aligned bounding boxes, oriented bounding boxes, and object poses.

News

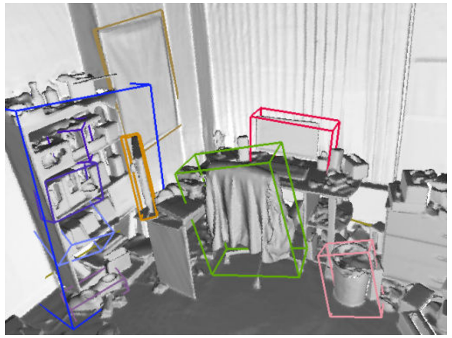
- **Oct 13, 2017**: Experimental WebGL annotation tool is here.
- **Sep 06, 2017**: An alternative data server (SUTD server) is up.
- **Aug 04, 2017**: The SceneNN annotation tool paper is accepted to TVCG! We will present it at Pacific Graphics 2017. See here.
- **Jan 27, 2017**: Data for SHREC 2017 track released! Please see the description here.

www.scenenn.net

Capture

Reconstruct

Annotate in 3D



- Triangle mesh
- Camera poses

monitor
keyboard
poster
cabinet
bookshelf
trash bin

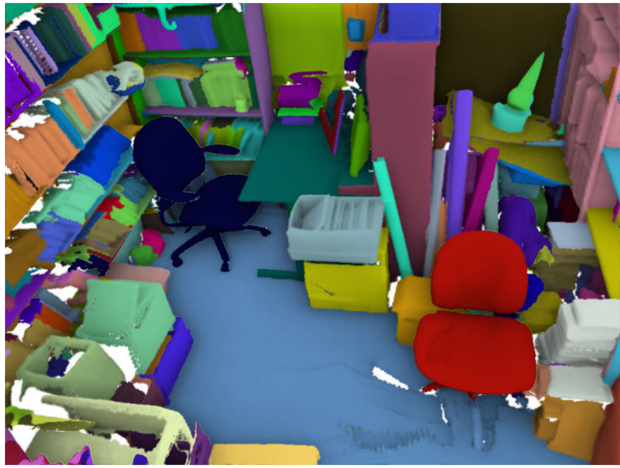
- Per-vertex and per-pixel labels
- Bounding boxes, object poses

SceneNN dataset

http://www.scenenn.net

- **100+** RGBD indoor scenes
- Raw videos from 2,000 to more than 10,000 frames
- Reconstructed triangle meshes in PLY format
- Per-frame camera poses
- Per-vertex and per-pixel labelling
- Annotated bounding boxes, object poses



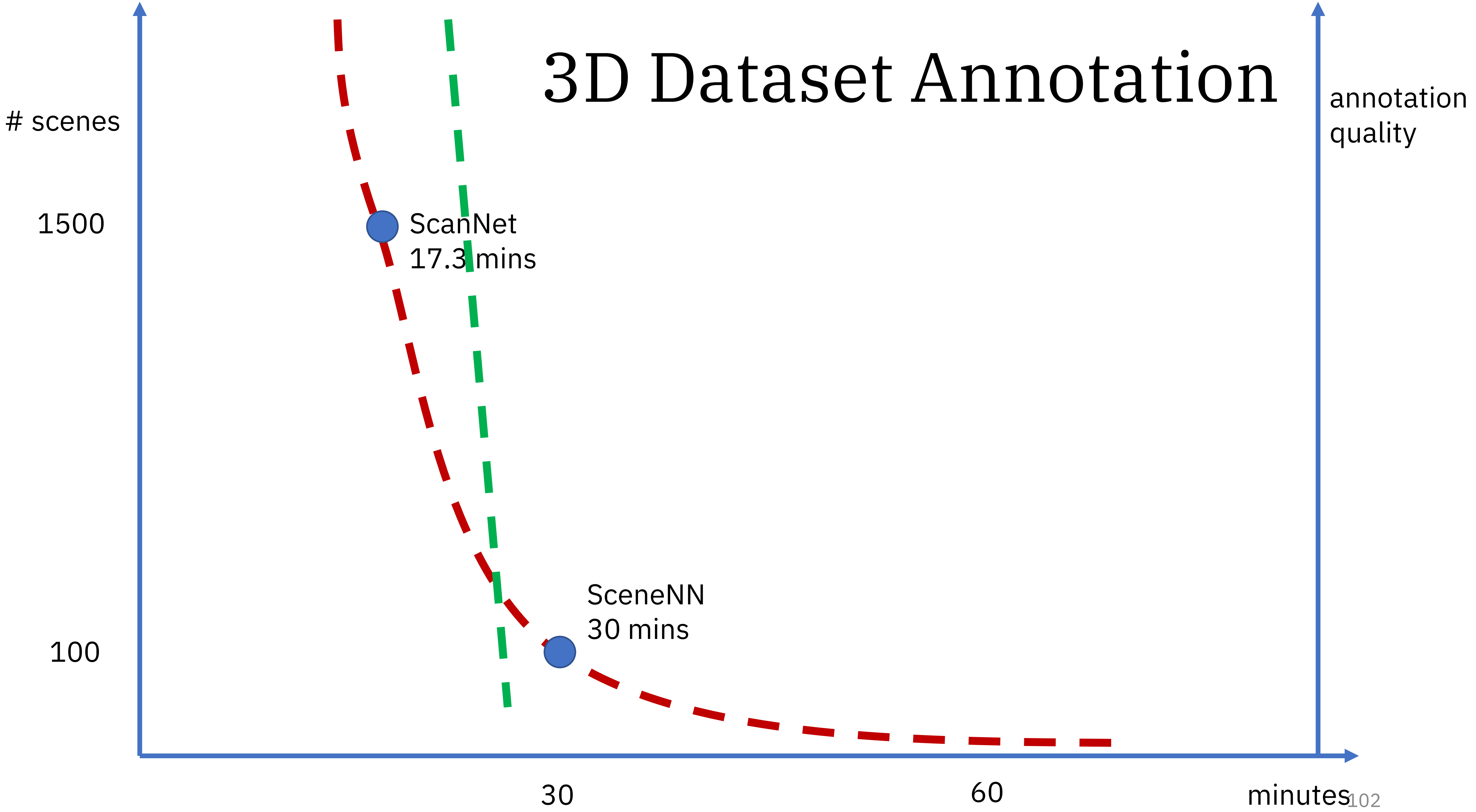



ScanNet

- 1500+ indoor scenes
- Per-vertex
  instance segmentation
- Crowdsourcing annotation:
  massive scale vs.
  quality control.
- Voxel labelling

Dai et al., ScanNet: Richly-annotated 3D Reconstructions of Indoor Scenes, CVPR 2017

3D Dataset Annotation



ShapeNet

https://www.shapenet.org

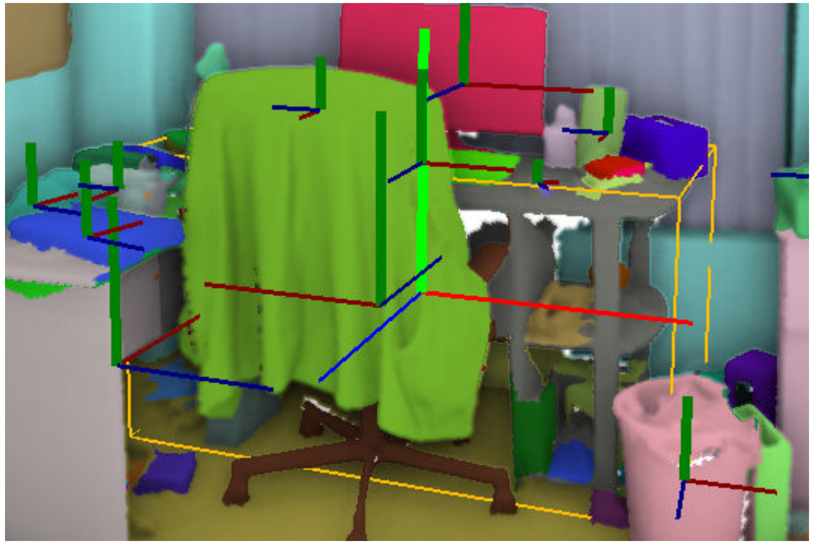
12,000 CAD models

270 categories

SceneNN-CAD

Bounding box

Object pose



Position CAD models

Application: RGB-D to CAD retrieval

Query: RGB-D object

- Colour and depth images
- Triangle mesh

Target: CAD model

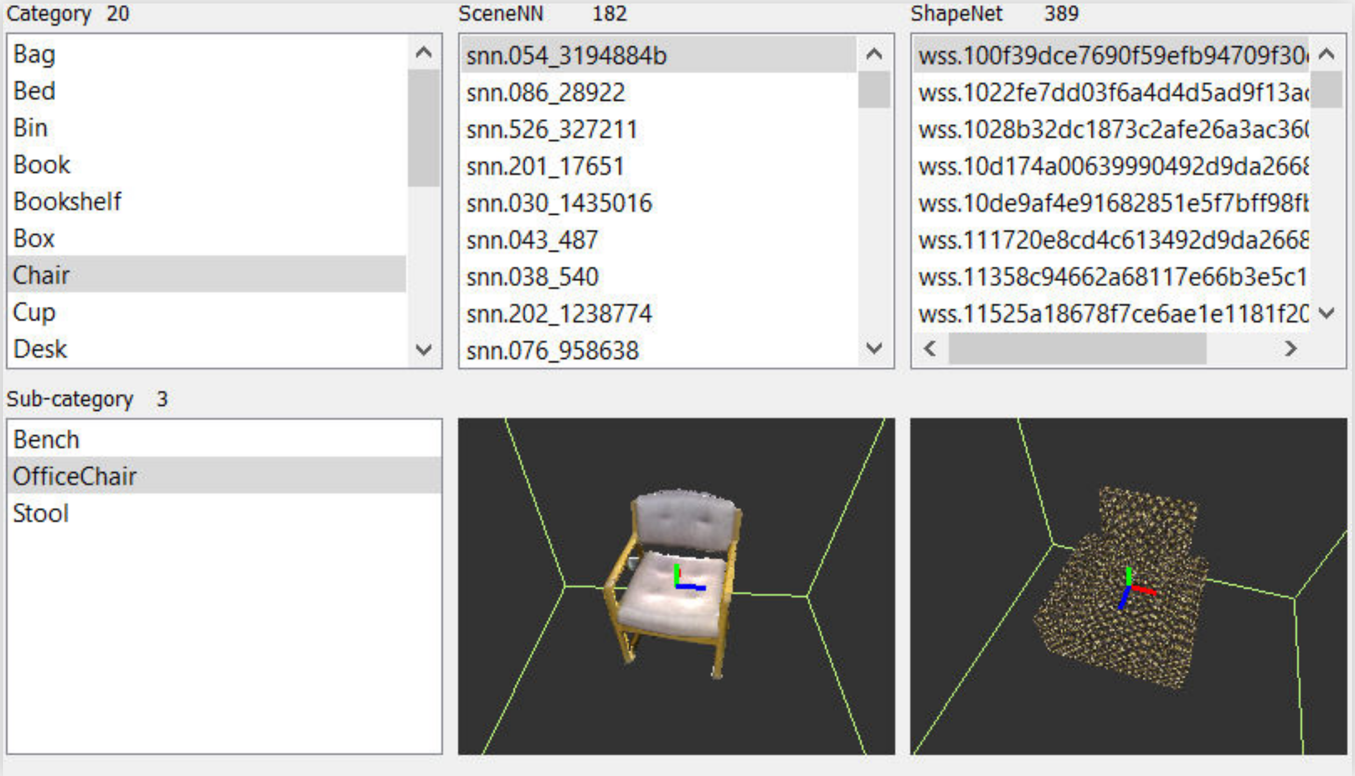
- Triangle mesh

SceneNN

ShapeNet

For each query, return a ranked list of retrieved CAD models

Objects from SceneNN

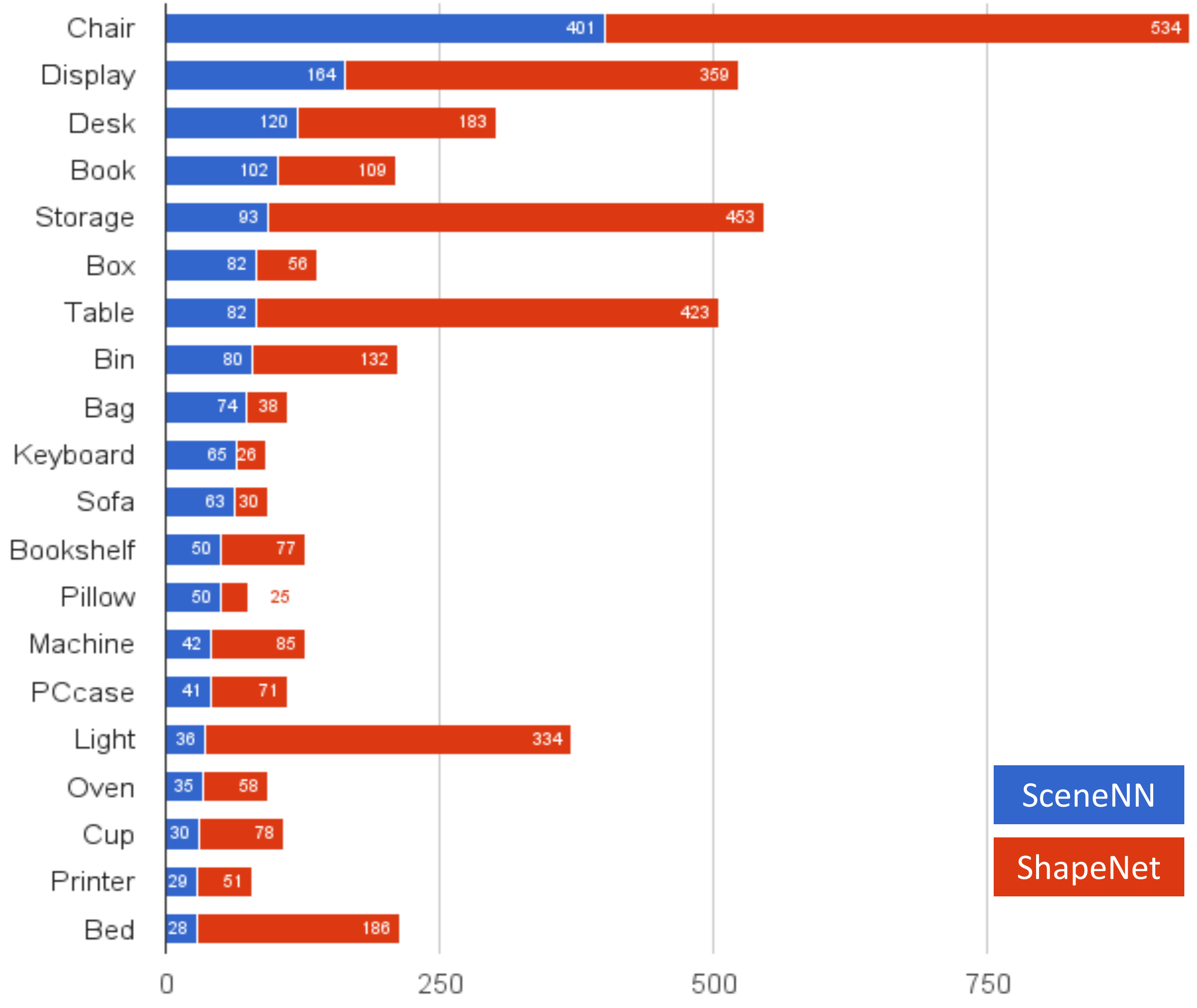


Objects from SceneNN

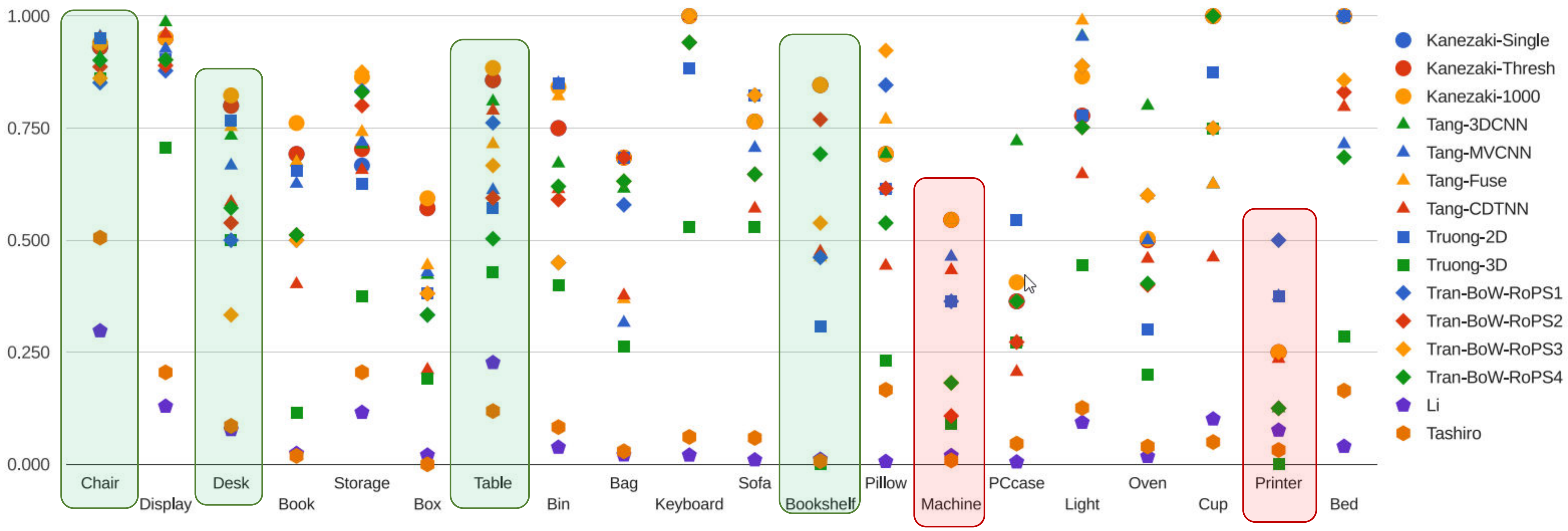
20 categories

1667 objects from SceneNN

3308 objects from ShapeNet



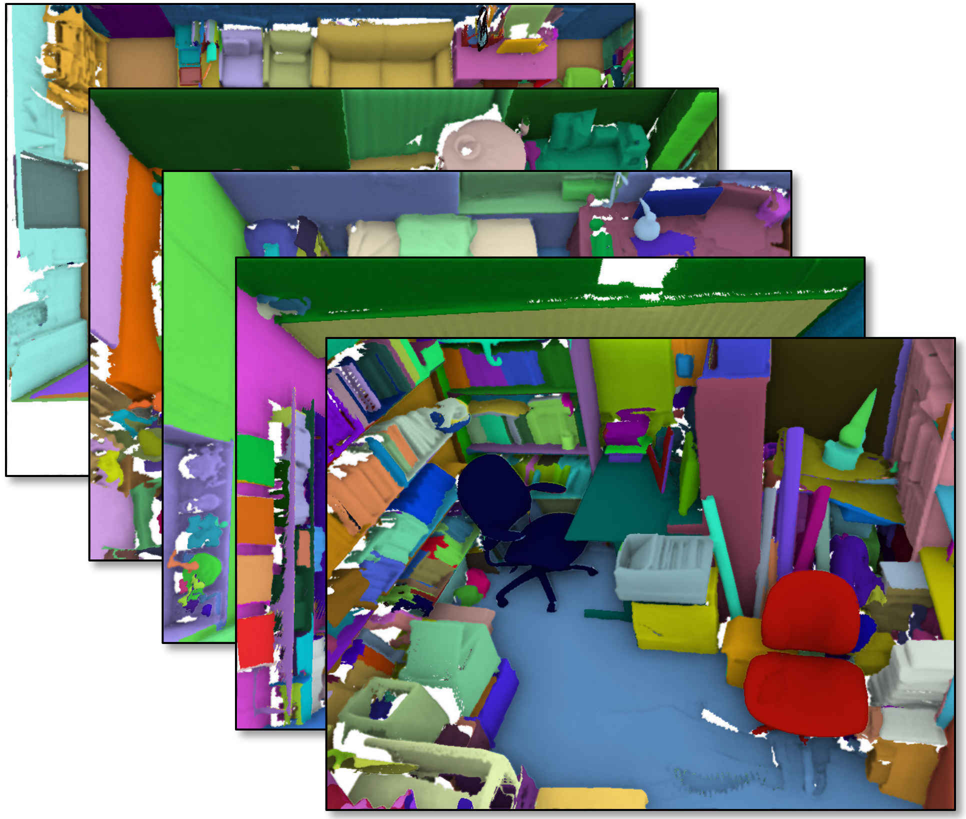
Object Classification



More details in our SHREC’17 and SHREC’18 workshop paper.

ScanObjectNN

https://hkust-vgd.github.io/scanobjectnn/



100+ scenes, very cluttered

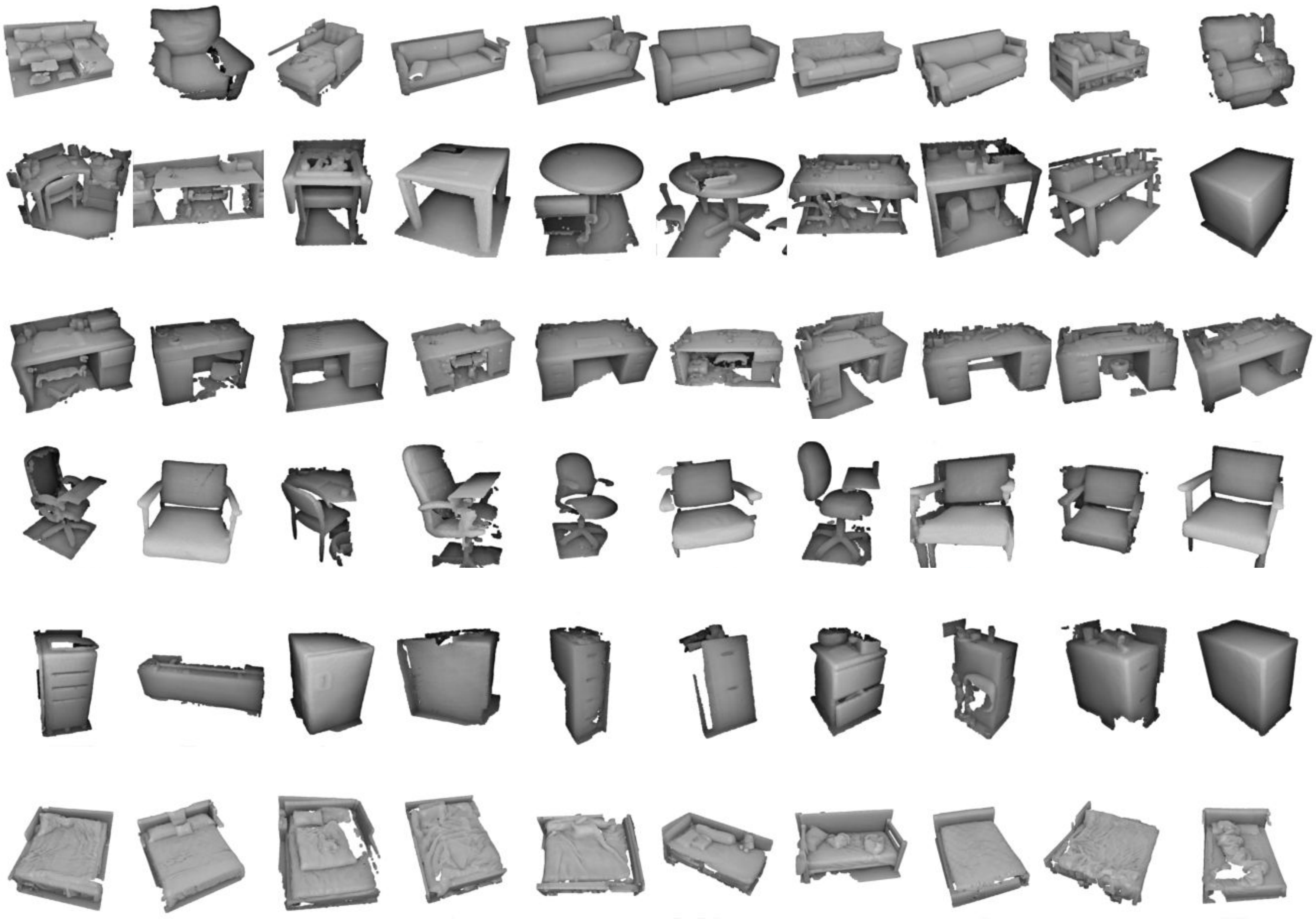
SceneNN [Hua et al., 2016]

1500+ scenes, large-scale scans

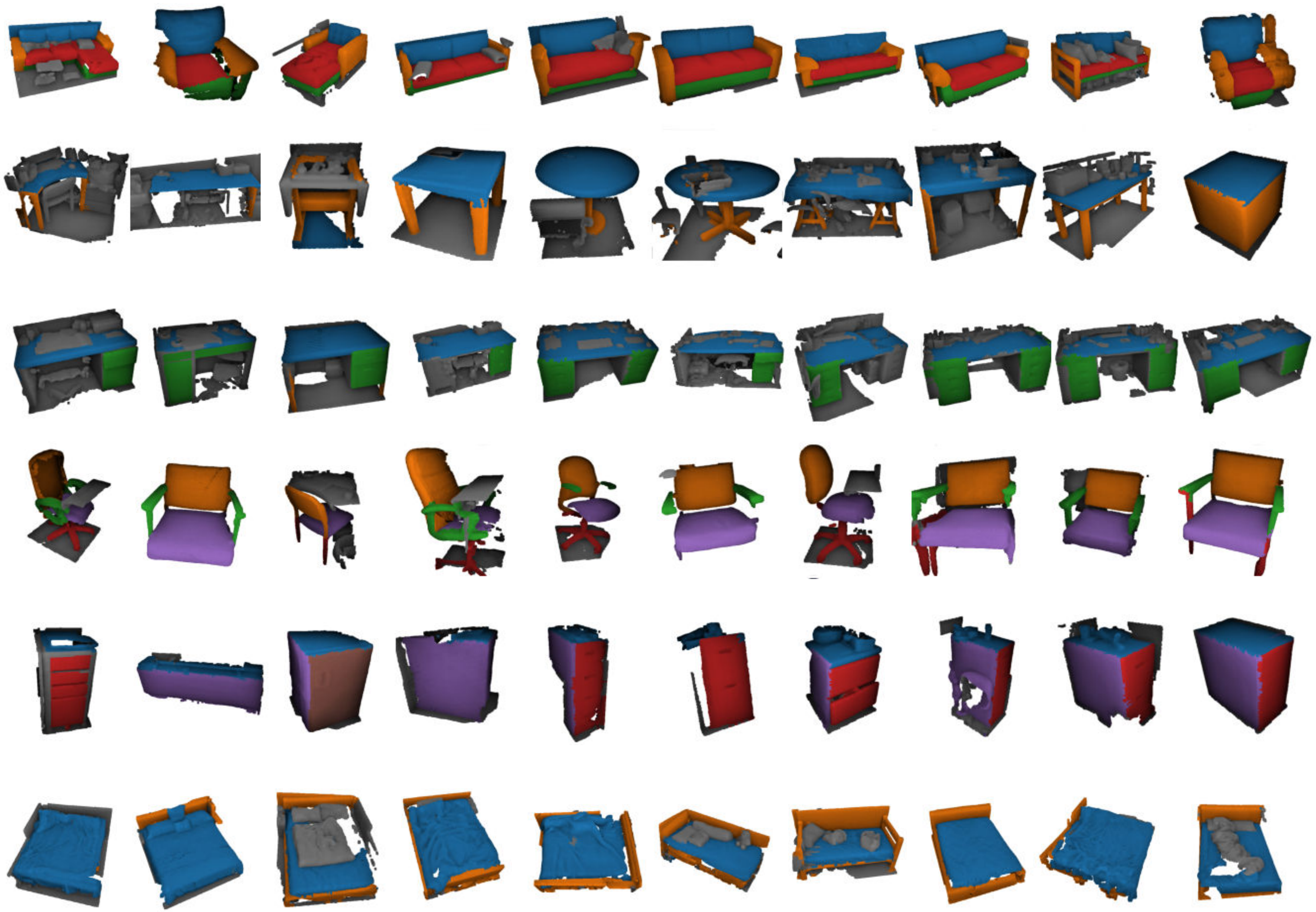
ScanNet [Dai et al., 2017]

Uy et al., Revisiting Point Cloud Classification: A New Benchmark Dataset..., ICCV 2019

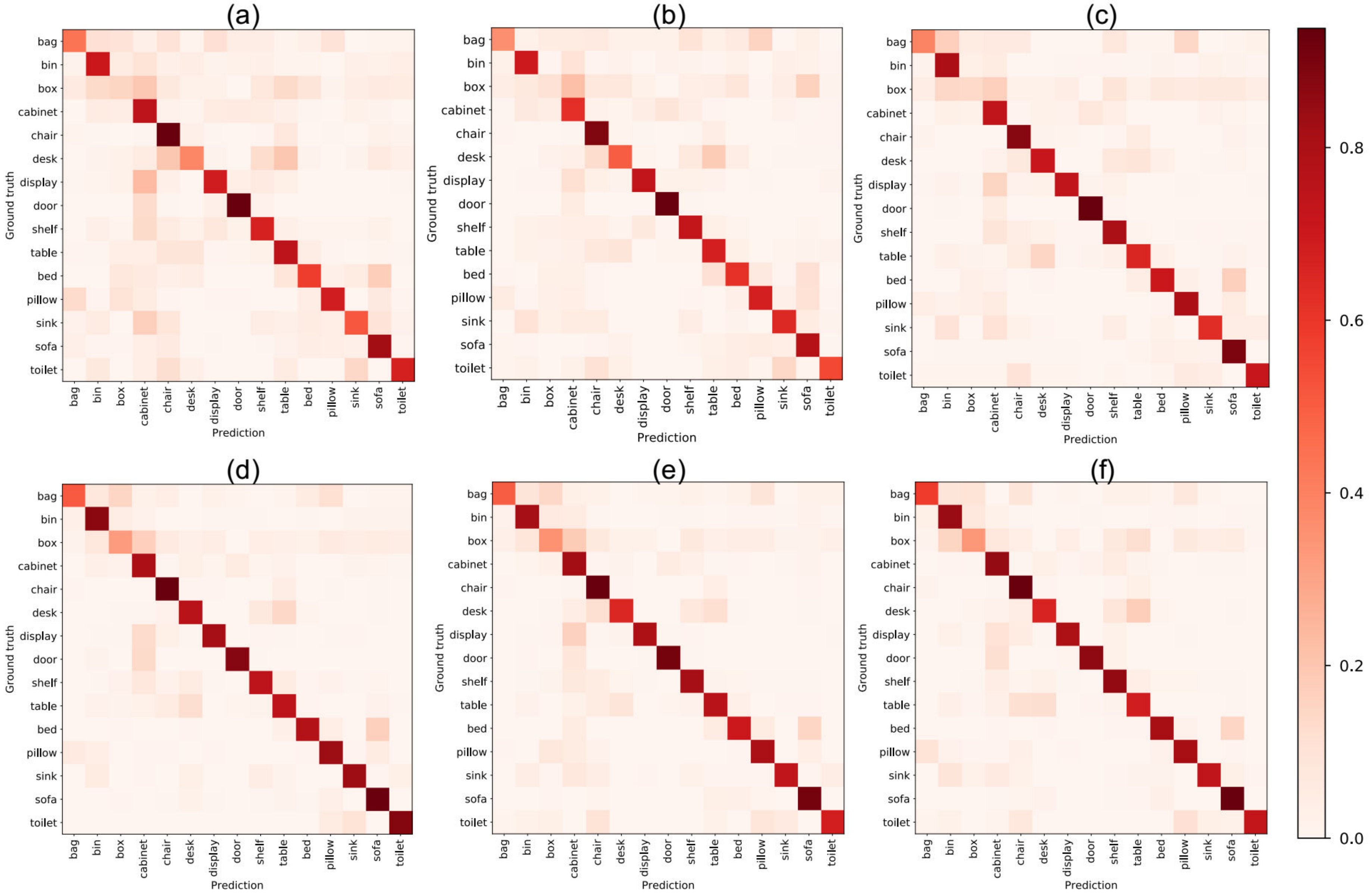
A new object dataset from **real-world scans** for **point cloud classification**



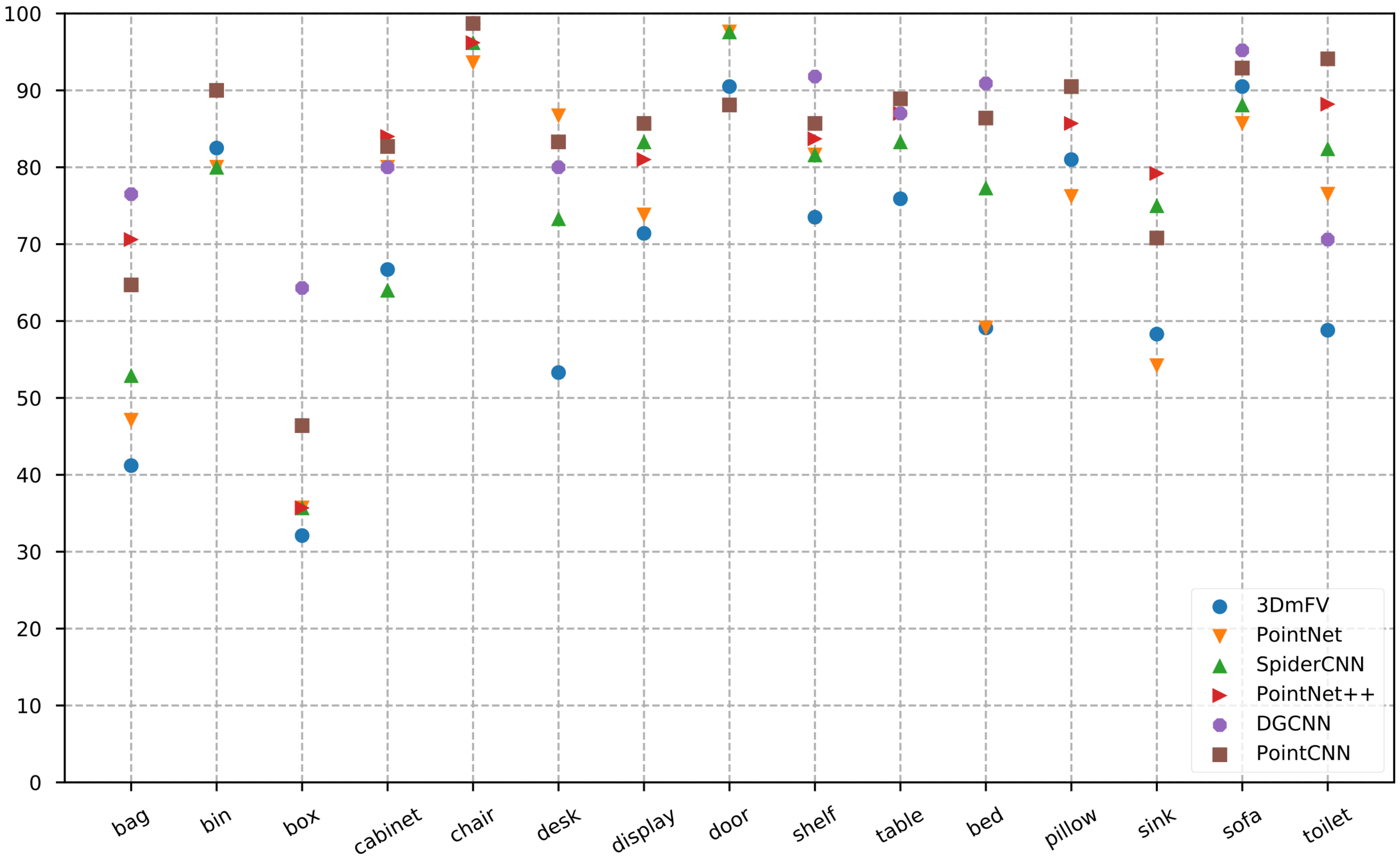
Our dataset: 15,000 objects, 6 variants, 5 train/test splits



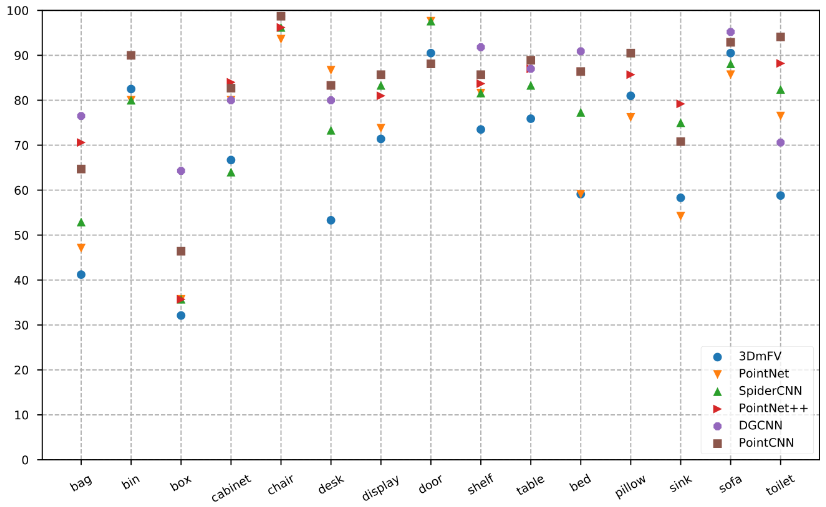
We also support **part annotation** for real-world scans



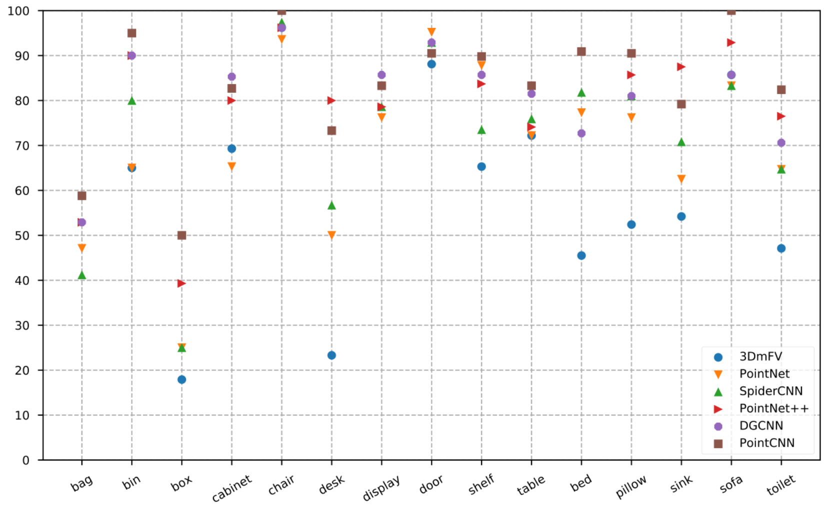
A comprehensive evaluation of existing point cloud classification methods



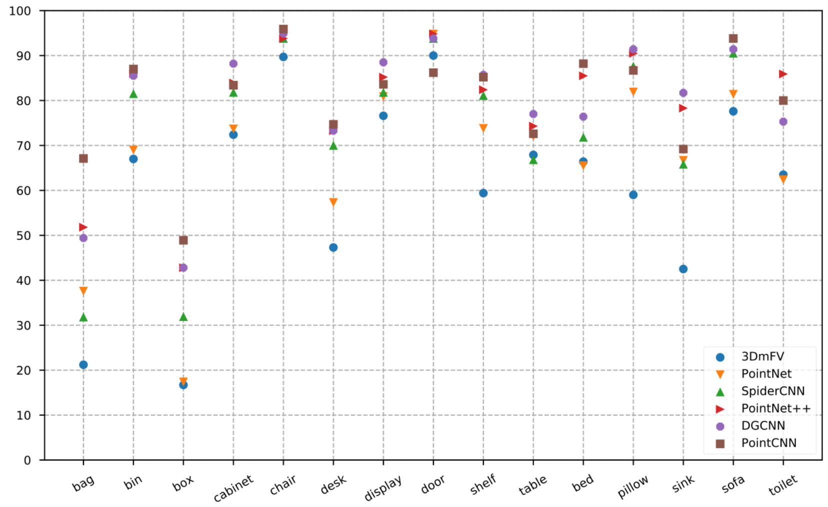
With detailed comparisons of existing point cloud classification methods



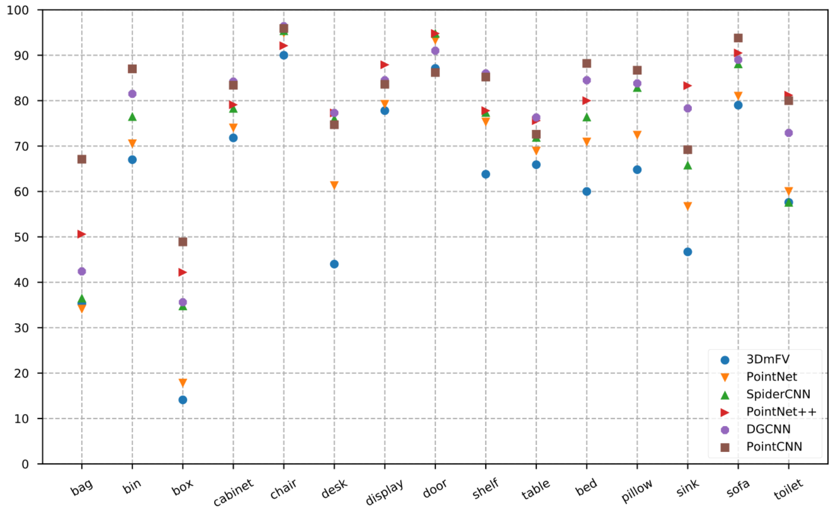
OBJ_ONLY



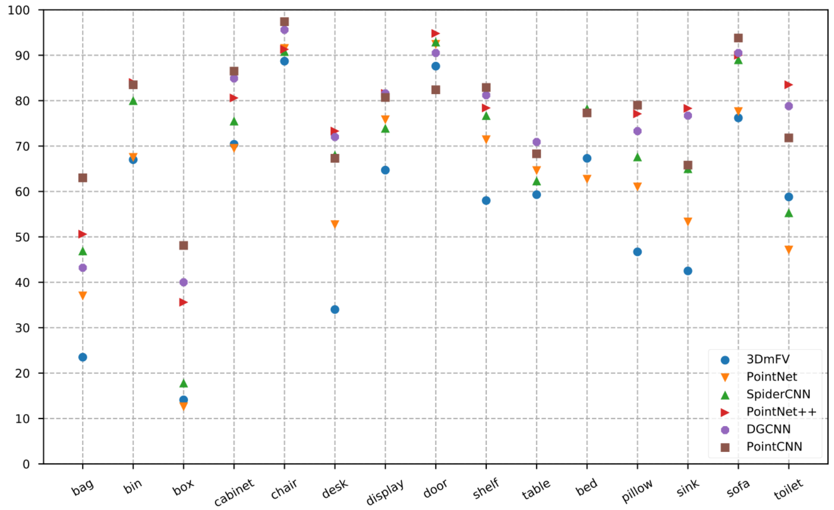
OBJ_BG



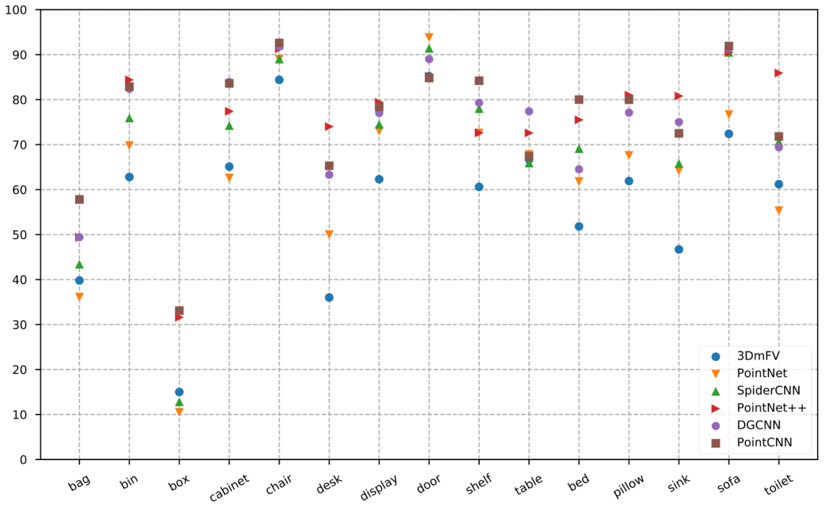
PB_T25



PB_T25_R



PB_T50_R



PB_T50_RS (hardest)

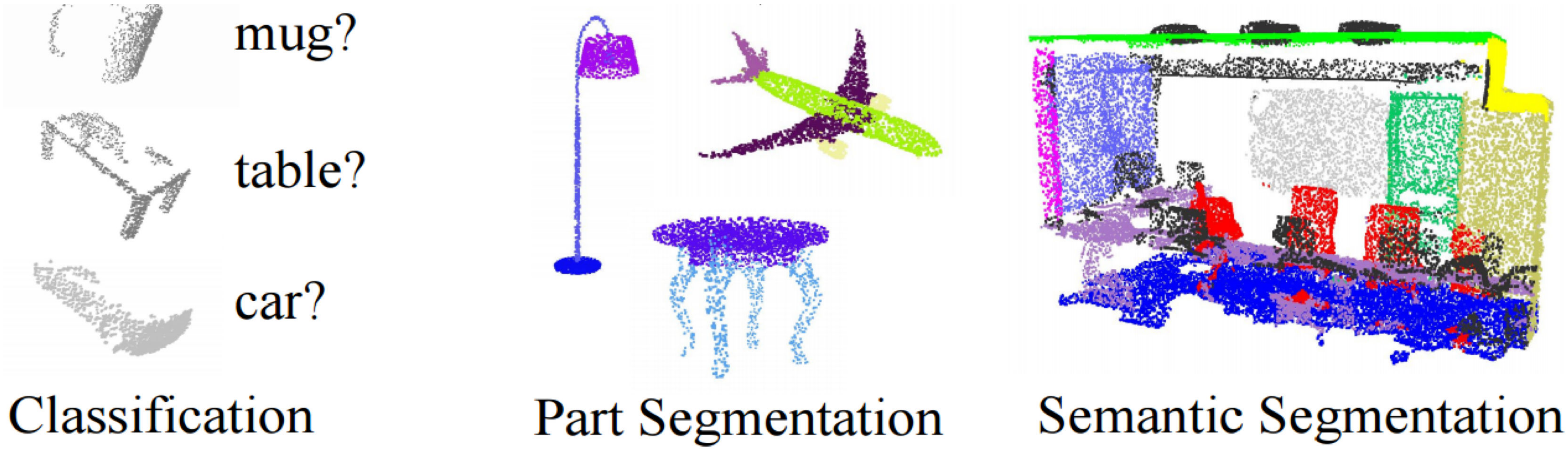
With detailed comparisons of existing point cloud classification methods

Summary

- Creating large real world 3D datasets is challenging.
- Acquisition and annotation are both time consuming.
- How to scale further, e.g., to tens of thousands scenes?

Part IV:
3D Deep Learning

Robot Vision with Point Cloud



Qi et al, PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation, CVPR 2017

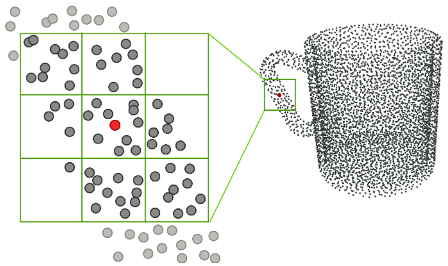
Challenges in Deep Learning with Point Cloud

- Points are unordered
  - Sorting
  - Mapping to order invariance
  - Recurrent neural network
- Convolution with a point cloud?
- Down-sampling and up-sampling

Pointwise Convolutional Neural Networks

Pointwise Convolution

- Convolution at every point of the cloud



- **On-the-fly** uniform grid for nearest neighbour search
- Forward convolution

$$x_i^\ell = \sum_k w_k \frac{1}{|\Omega_i(k)|} \sum_{p_j \in \Omega_i(k)} x_j^{\ell-1},$$

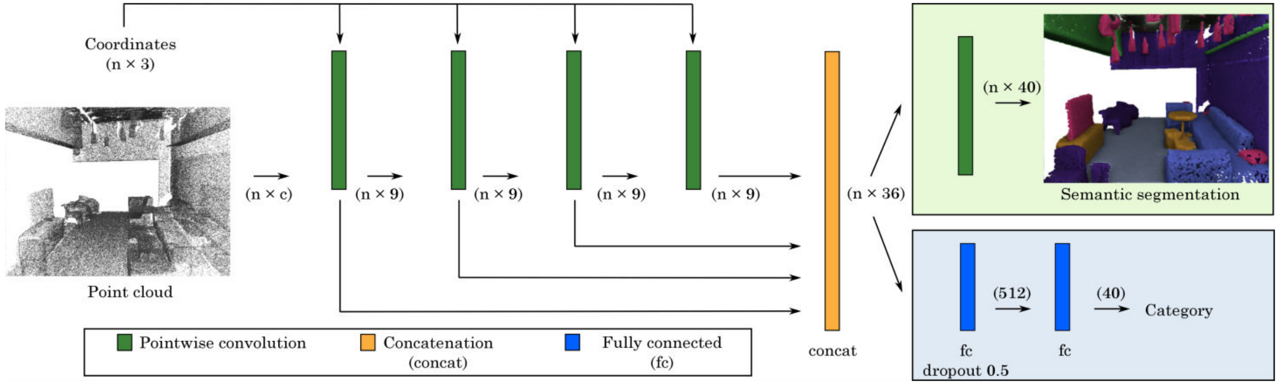
- Backward propagation

$$\frac{\partial L}{\partial x_j^{\ell-1}} = \sum_{i \in \Omega_j} \frac{\partial L}{\partial x_i^\ell} \frac{\partial x_i^\ell}{\partial x_j^{\ell-1}} \qquad \frac{\partial x_i^\ell}{\partial x_j^{\ell-1}} = \sum_k w_k \frac{1}{|\Omega_i(k)|} \sum_{p_j \in \Omega_i(k)} 1$$

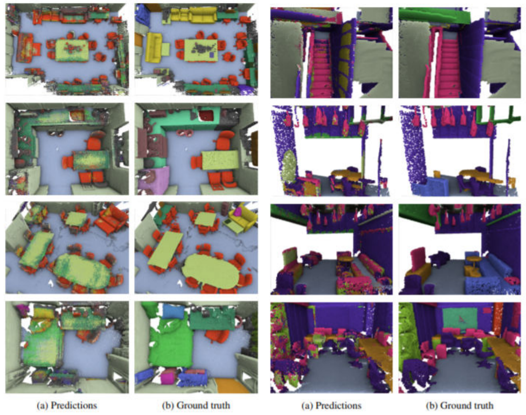
$$\frac{\partial L}{\partial w_k} = \sum_i \frac{\partial L}{\partial x_i^\ell} \frac{\partial x_i^\ell}{\partial w_k} \qquad \frac{\partial x_i^\ell}{\partial w_k} = \frac{1}{|\Omega_i(k)|} \sum_{p_j \in \Omega_i(k)} x_j^{\ell-1}$$

- À-trous convolution
- Self-normalizing activation function (SeLU)
- CUDA and multi-GPU implementation

Neural Network



Semantic Segmentation

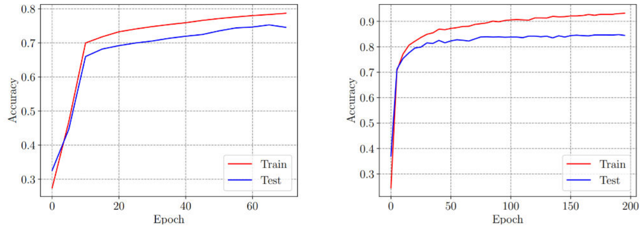


SceneNN S3DIS

Object Recognition

Base	Concat.	À-trous	SELU	Dropout	Accuracy
✓					78.6
✓	✓				78.0
✓		✓			75.0
✓	✓	✓			82.5
✓			✓		81.7
✓	✓		✓		81.9
✓	✓		✓	✓	85.2
✓	✓	✓	✓	✓	86.1

Convergence



(a) Scene segmentation (b) Object recognition

Future Works

- Adapt neural network design from 2D to 3D with pointwise convolution.
- Global feature learning.
- Applications: denoising, up-sampling, colorization.

www.scenenn.net

- Source code available!

SINGAPORE UNIVERSITY OF TECHNOLOGY AND DESIGN

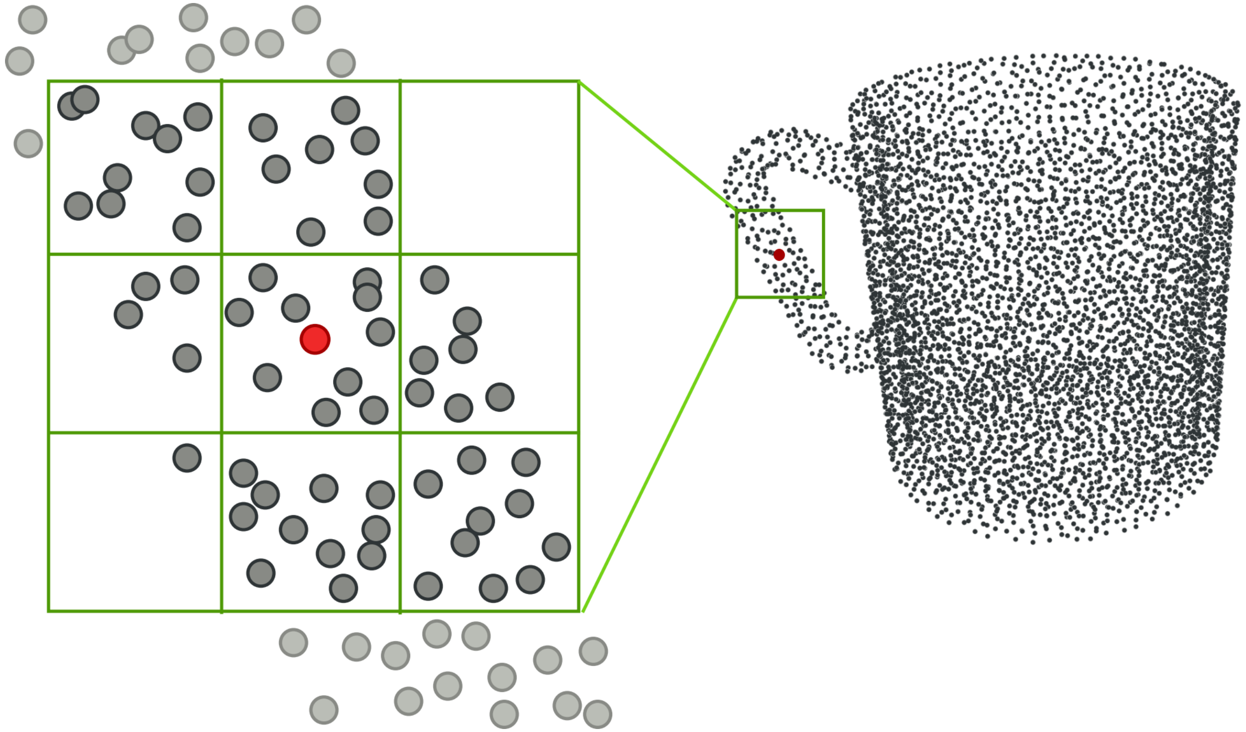
DManD

Courtesy of Hua et al., Pointwise Convolutional Neural Networks, CVPR 2018

THE UNIVERSITY OF TOKYO

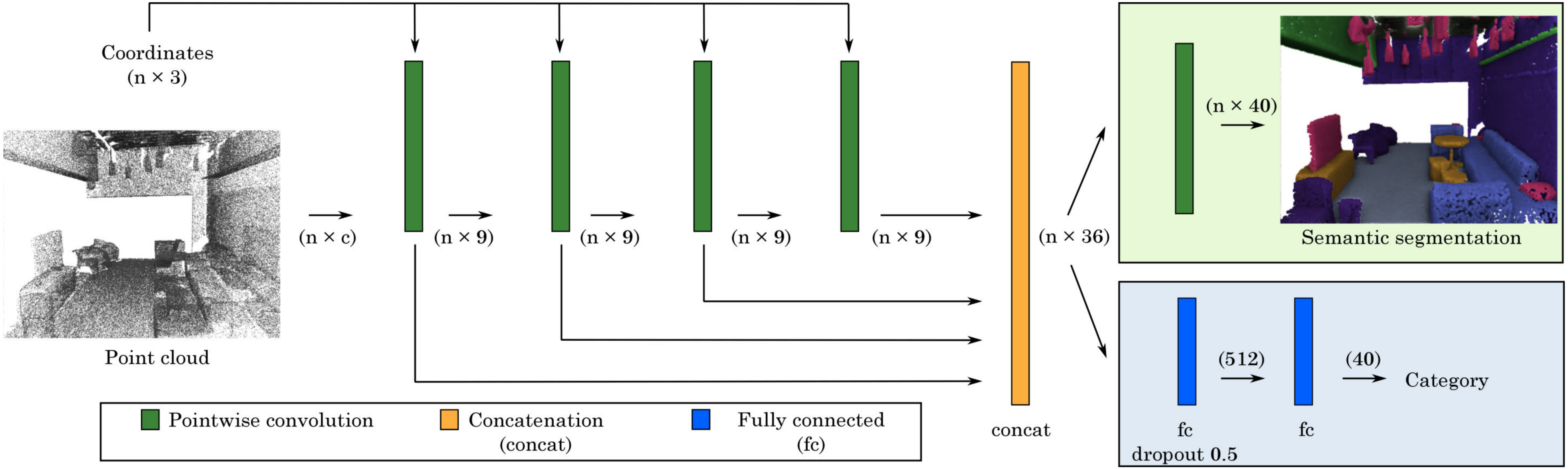
Pointwise Convolution

- At each point, centre a grid
- Take points in the grid for convolution
- Points each cell have the same weight
- Nearest neighbour query on the fly



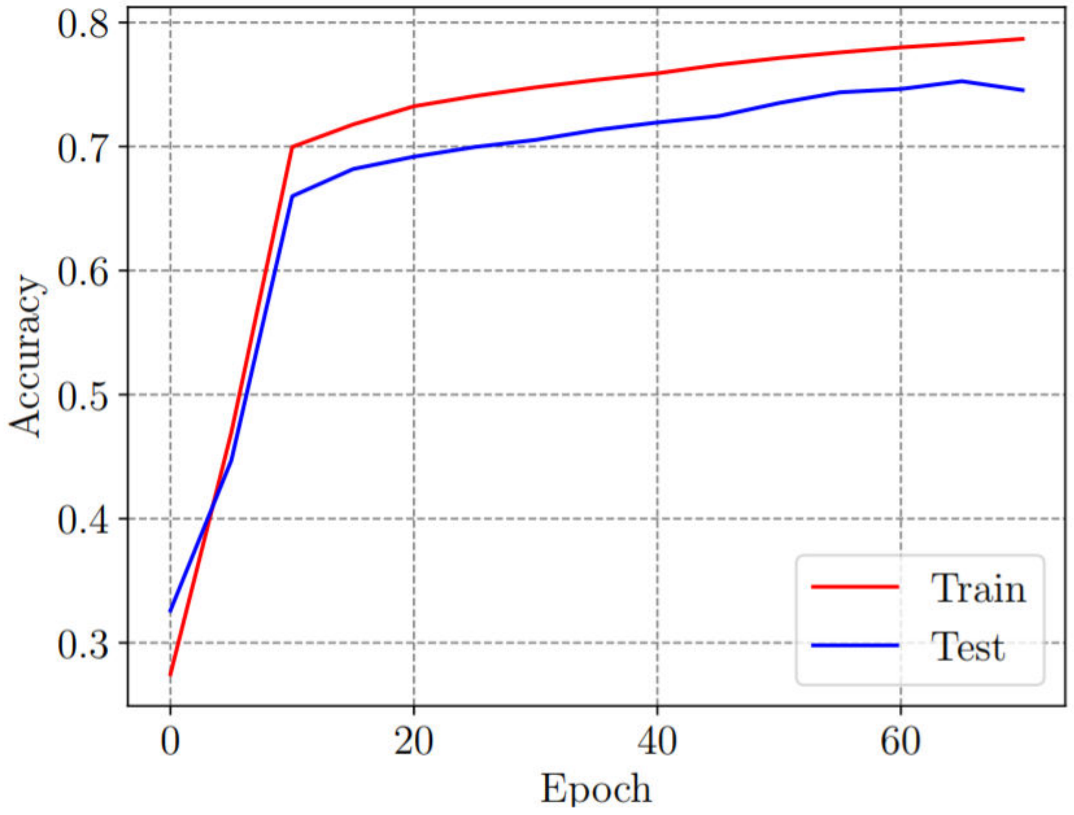
Courtesy of Hua et al., Pointwise Convolutional Neural Networks, CVPR 2018

Pointwise Convolutional Neural Network

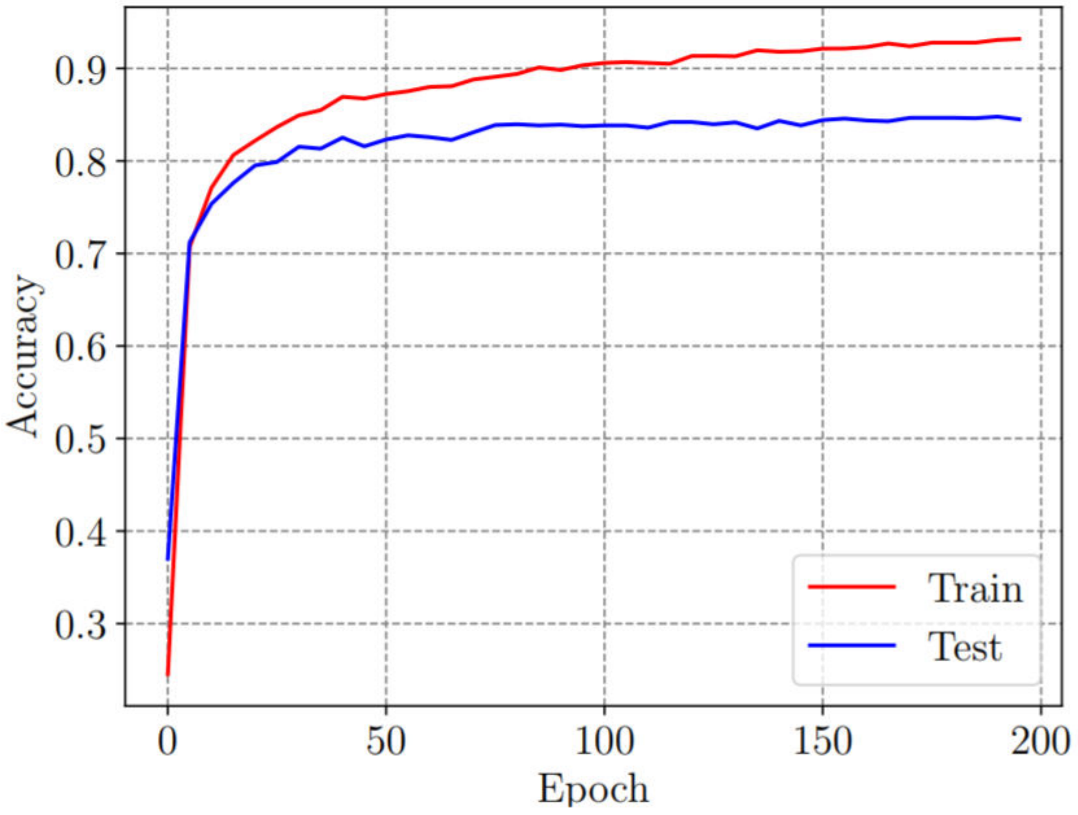


Courtesy of Hua et al., Pointwise Convolutional Neural Networks, CVPR 2018

Training and Testing



(a) Scene segmentation



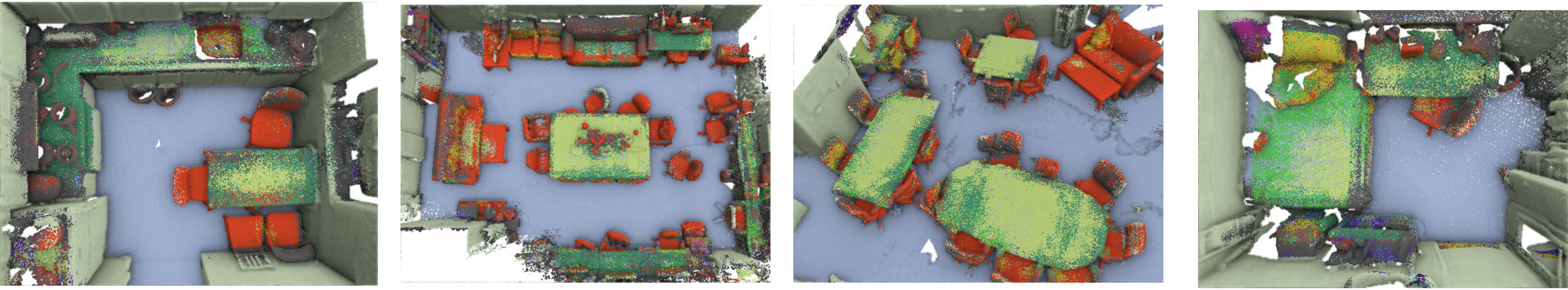
(b) Object recognition

Object Classification

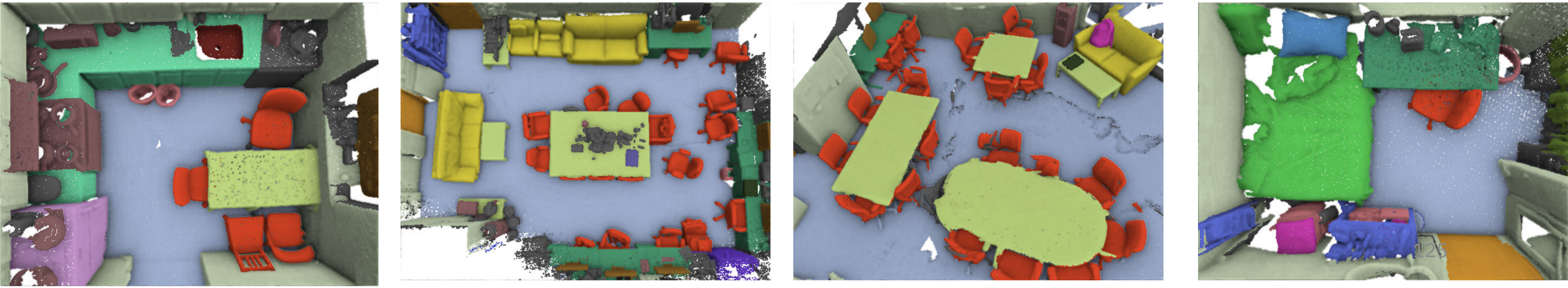
Base	Concat.	À-trous	SELU	Dropout	Accuracy
✓					78.6
✓	✓				78.0
✓		✓			75.0
✓	✓	✓			82.5
✓			✓		81.7
✓	✓		✓		81.9
✓	✓		✓	✓	85.2
✓	✓	✓	✓	✓	86.1

Semantic Segmentation

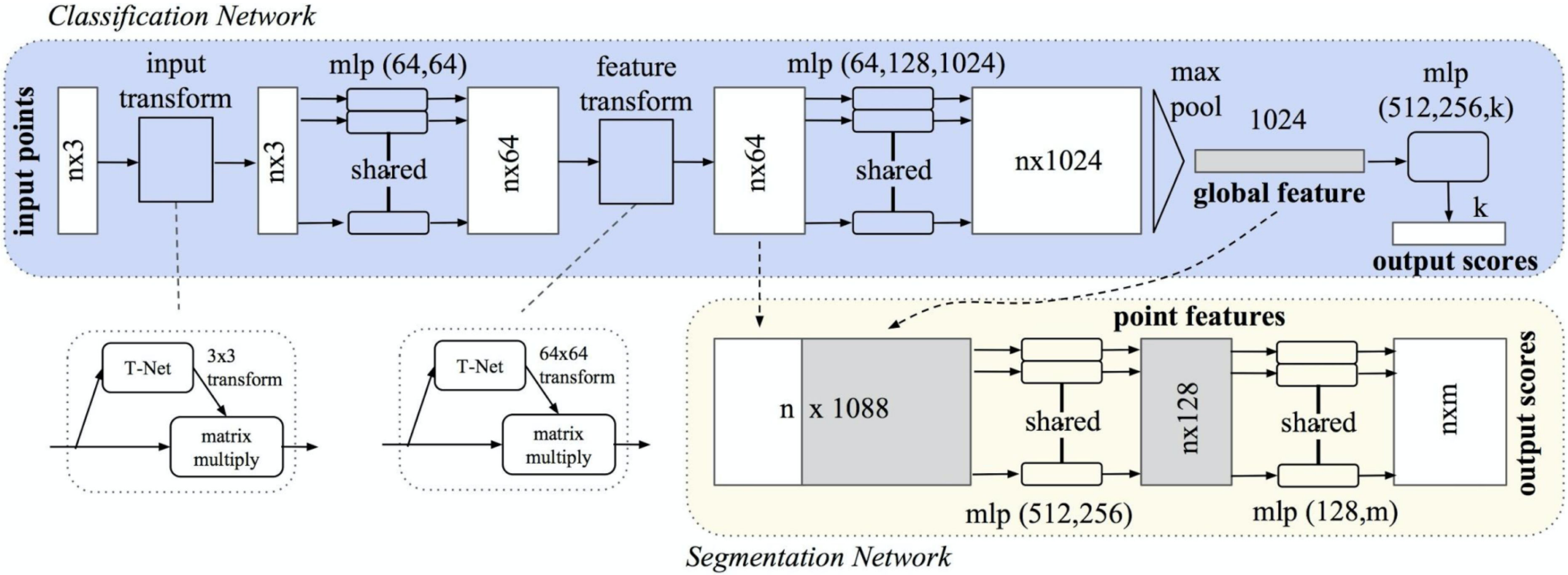
Prediction



Ground truth

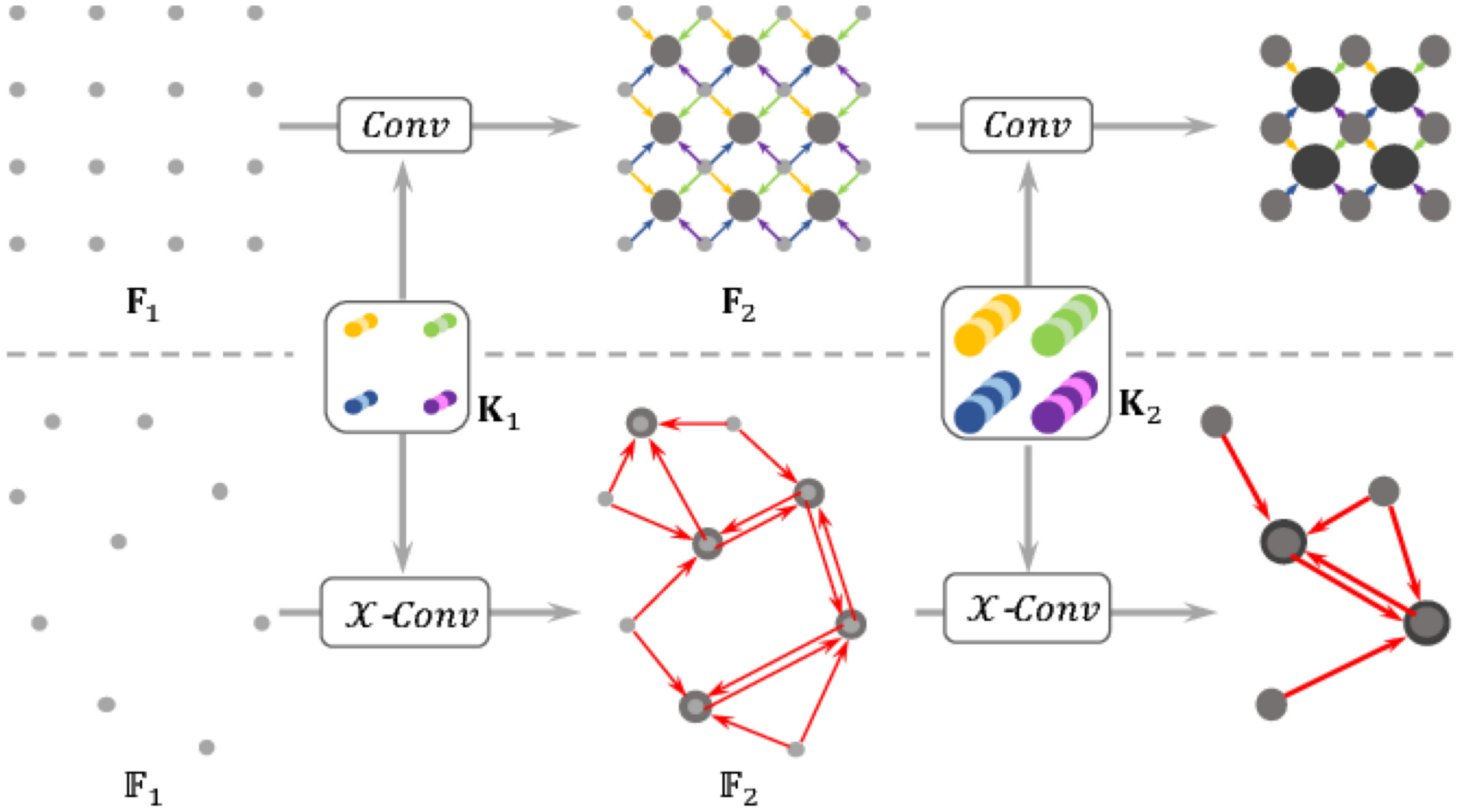


PointNet



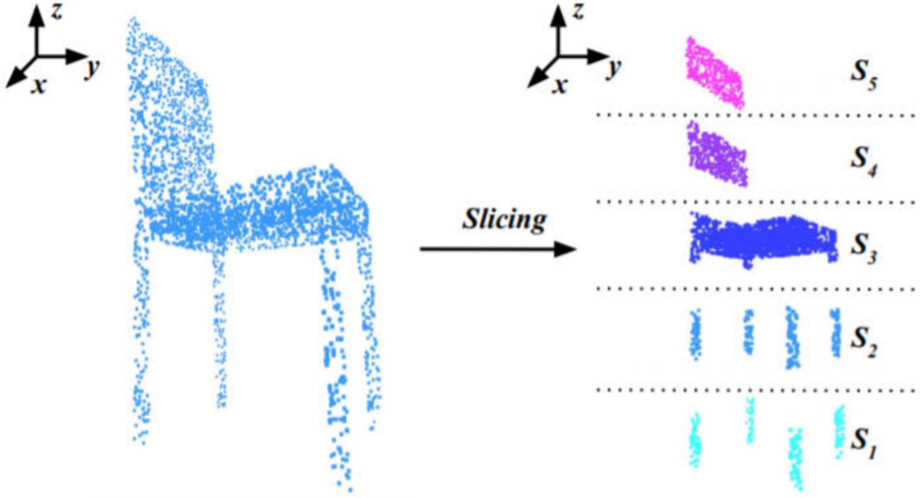
Qi et al., PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation, CVPR 2017

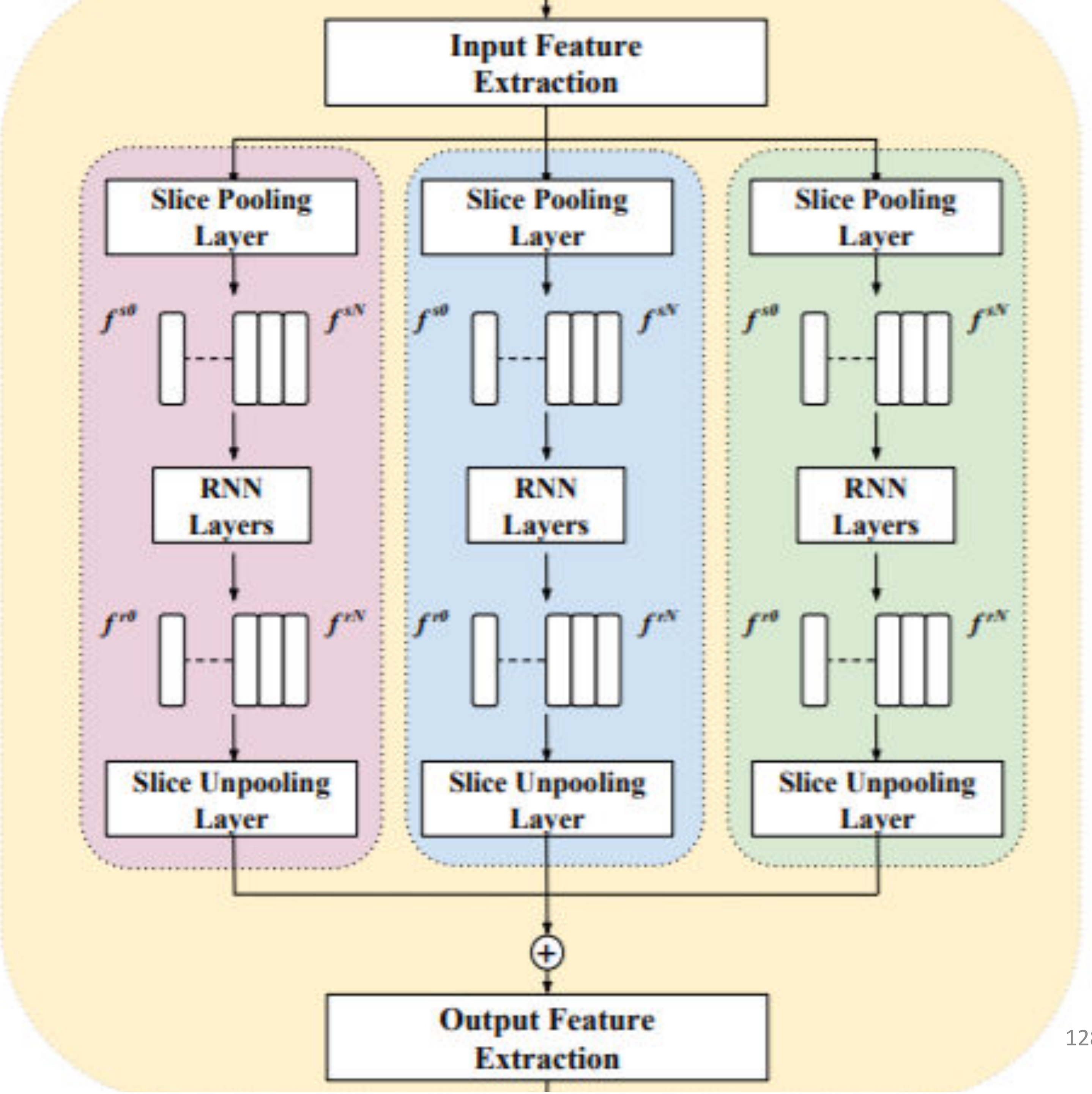
PointCNN



Li et al., PointCNN, arXiv 2018

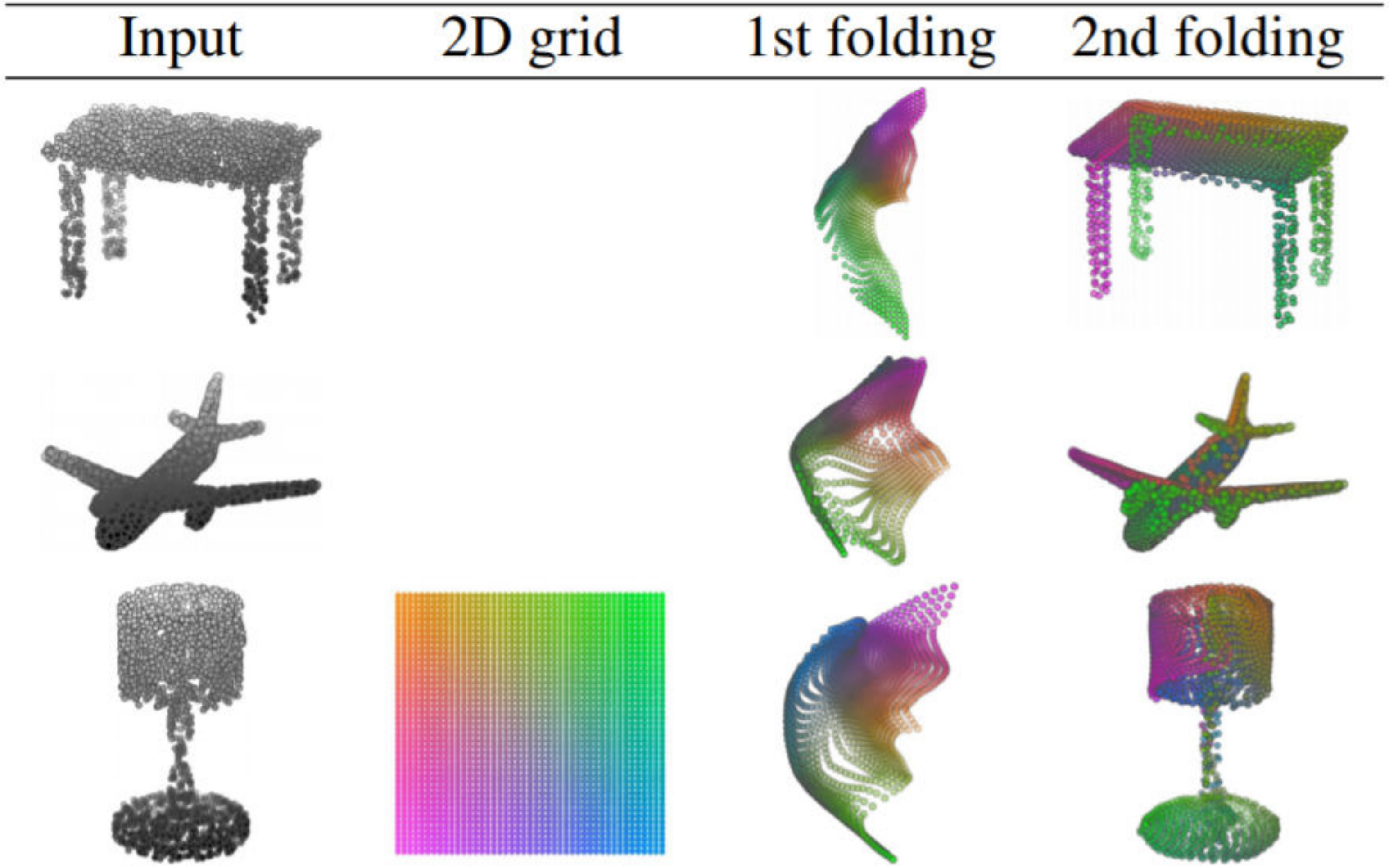
Recurrent Slice Networks (RSNet)





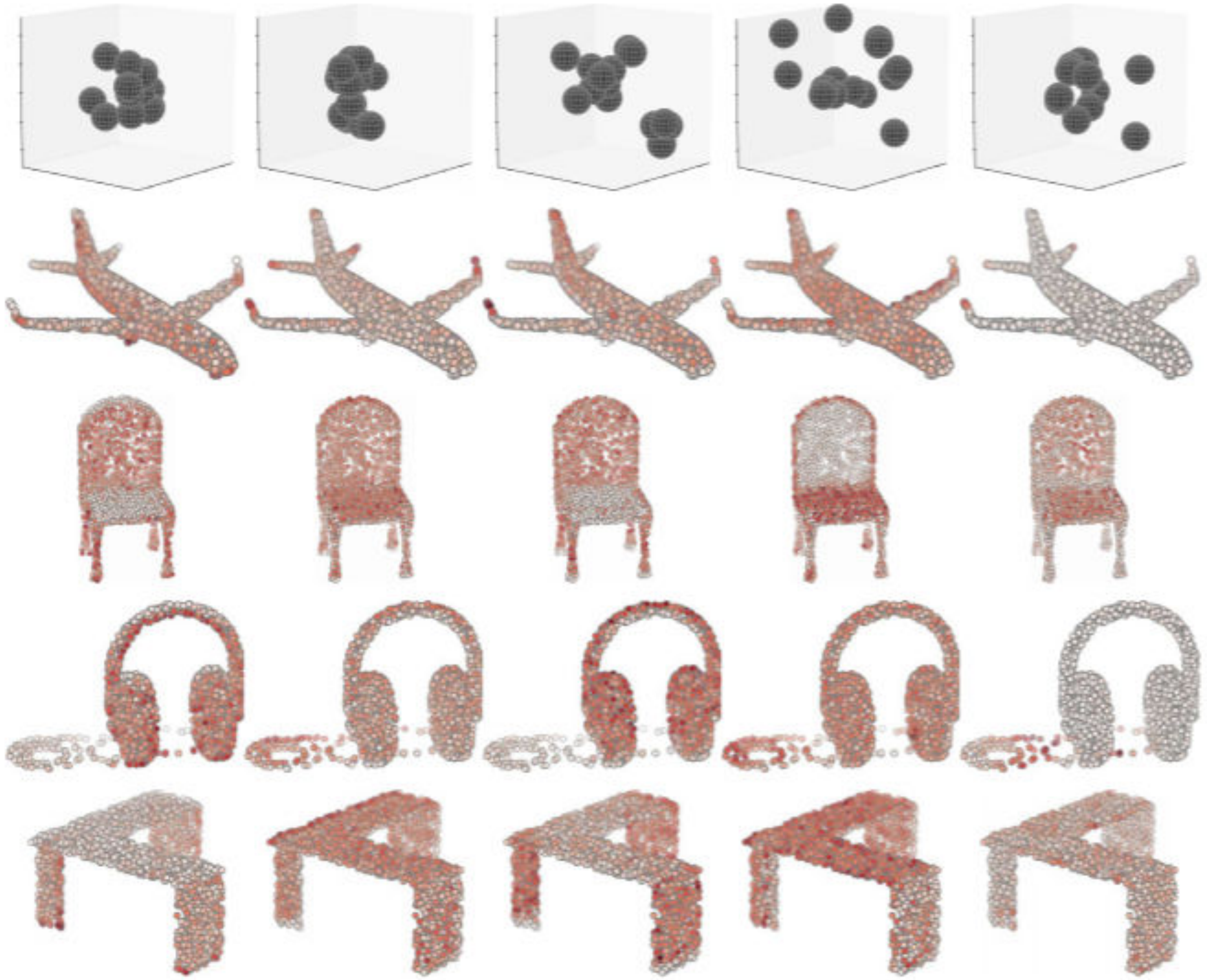
Huang et al., Recurrent Slice Networks for 3D Segmentation of Point Clouds, CVPR 2018

FoldingNet



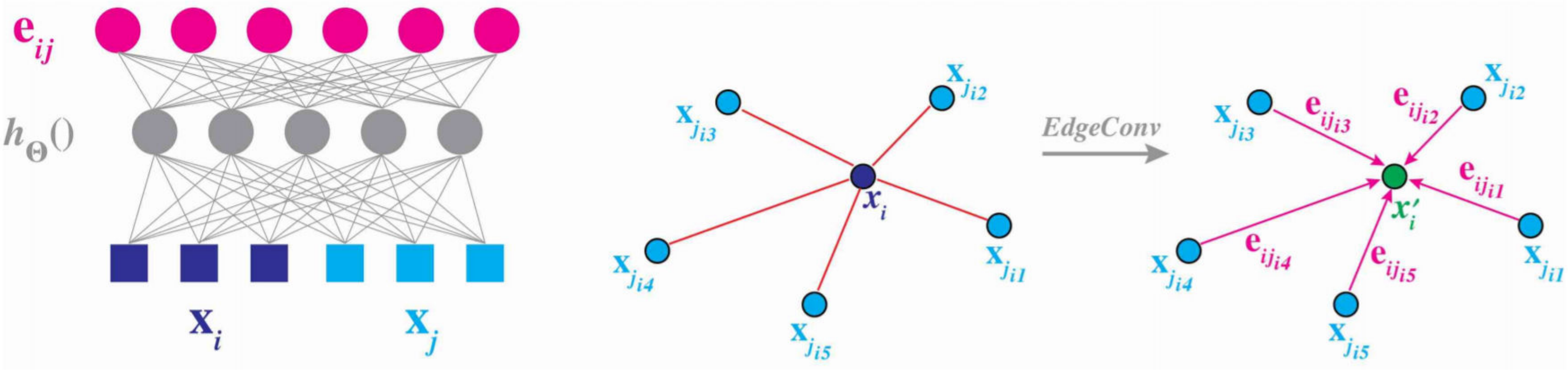
Yang et al., FoldingNet: Point Cloud Auto-encoder via Deep Grid Deformation, CVPR 2018

Point Cloud Local Structures by Kernel Correlation



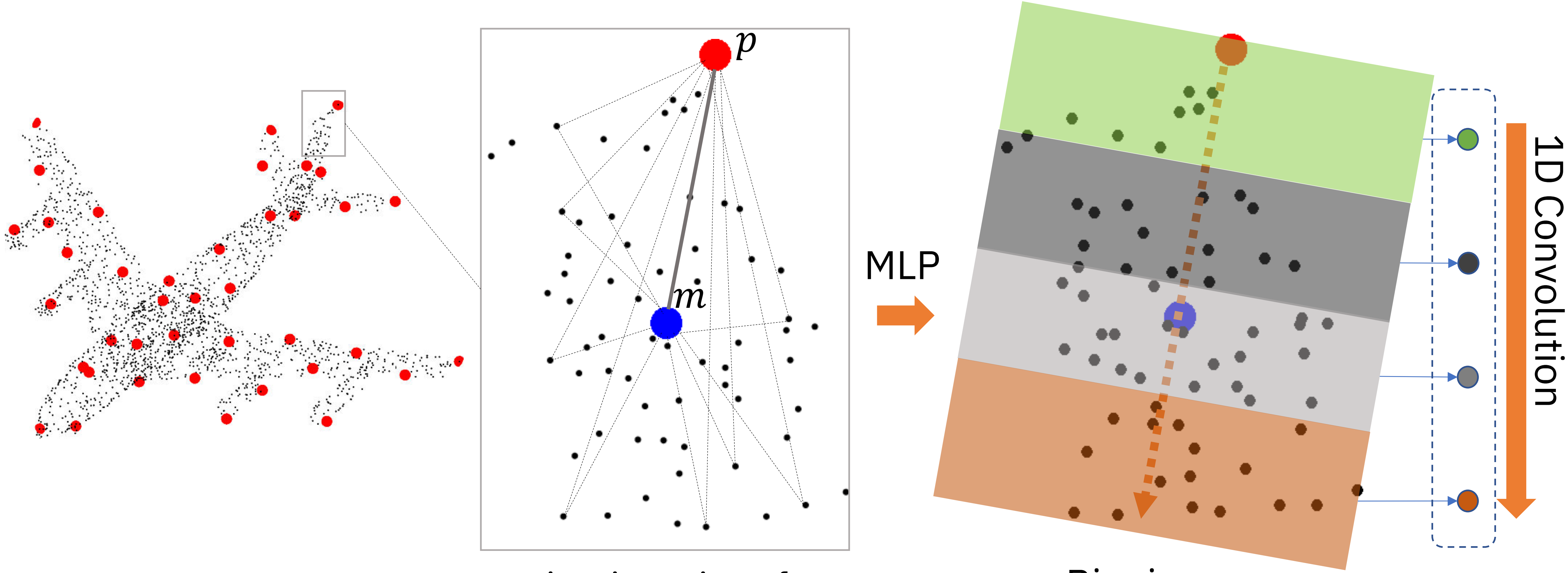
Shen et al., Mining Point Cloud Local Structures by Kernel Correlation and Graph Pooling, CVPR 2018

EdgeConv



Wang et al., Dynamic Graph CNN for Learning on Point Clouds, arXiv 2018

ROTATION INVARIANT CONVOLUTION



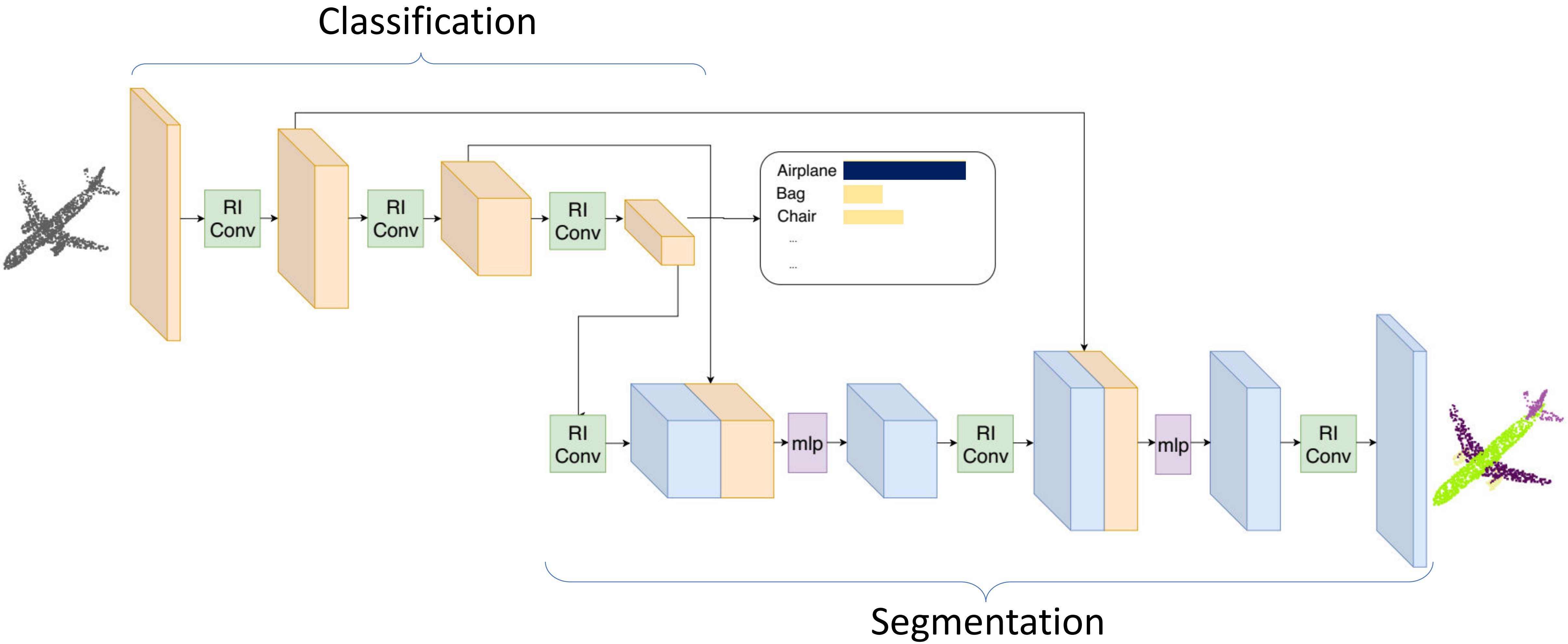
Rotation invariant features

Binning

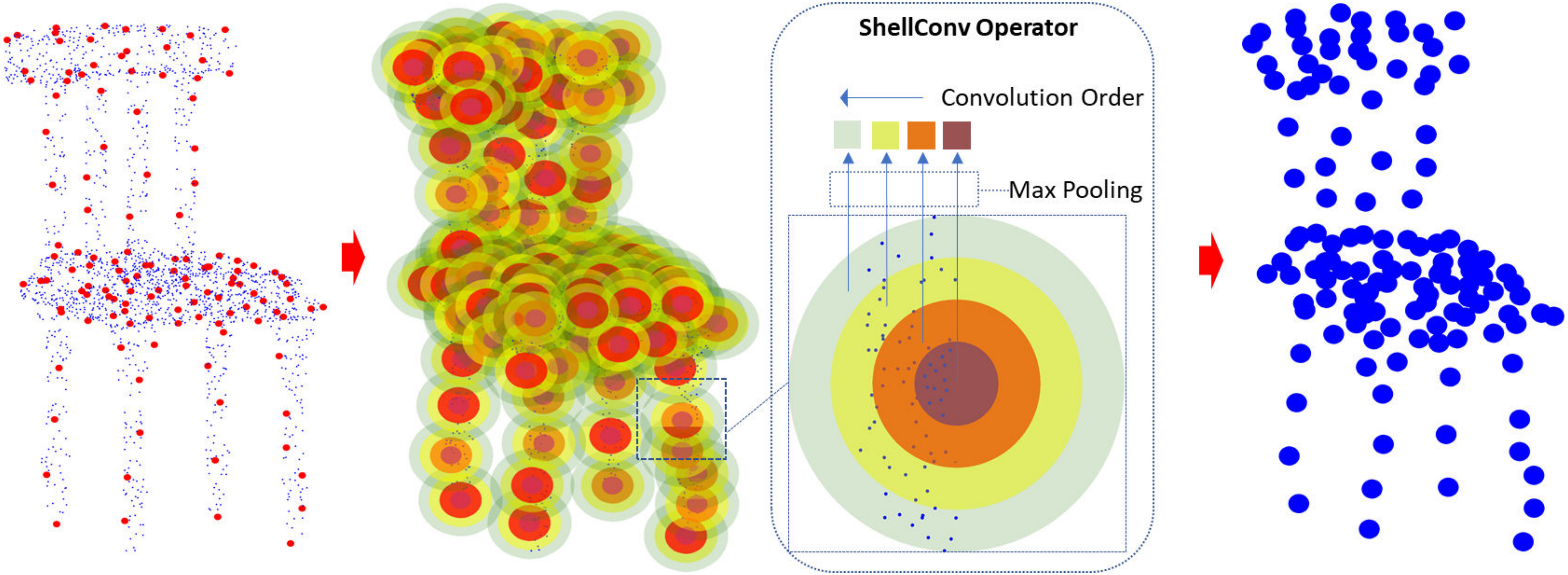
Maxpool

Zhang et al., Rotation Invariant Convolutions for 3D Point Clouds Deep Learning, 3DV 2019

ROTATION INVARIANT NEURAL NETWORK

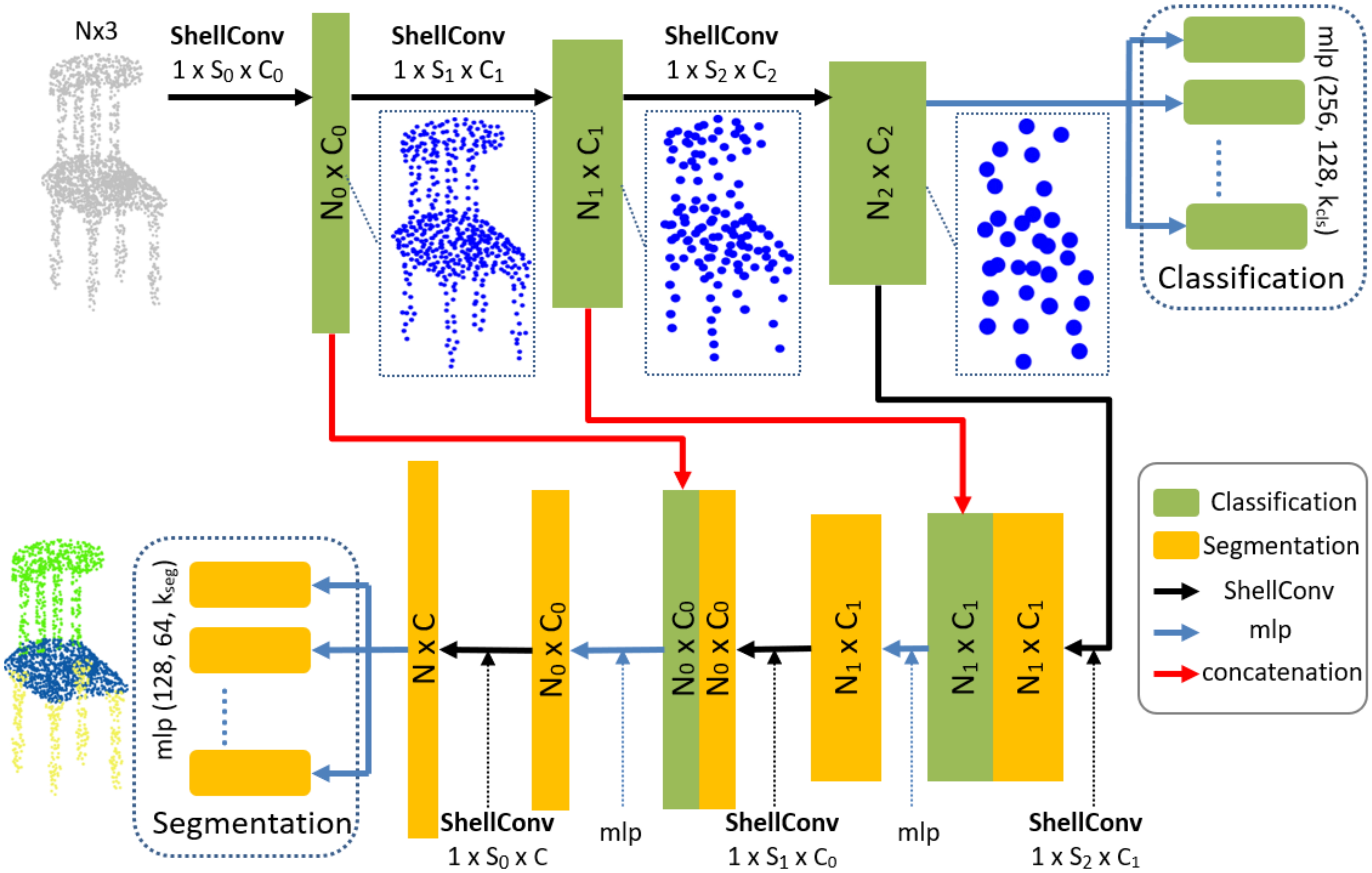


SHELL-BASED CONVOLUTION



Zhang et al., ShellNet: Efficient Point Cloud Convolutional Neural Networks, ICCV 2019

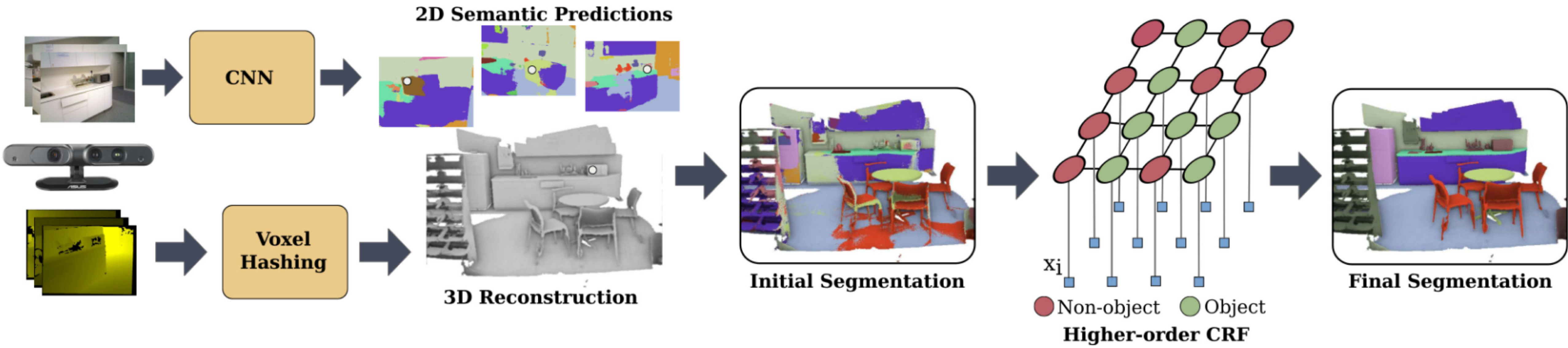
CLASSIFICATION AND SEGMENTATION NETWORK



SUMMARY	Order	Convolution	Applications	O MN40	S S3DIS
Pointwise	Sort	3D, with nearest neighbor search	O S	86.1	-
PointNet	Symmetric function	Per-point using multi-layer perceptron	O S P	89.2	47.6
PointNet++	Symmetric function	Split into groups, each group has a PointNet	O S P	90.7	-
PointCNN	X-transform	Transformation and point downsampling	O S	91.7	62.7
RSNet	Recurrent network	Recurrent network	S P	-	51.9
FoldingNet	Symmetric function	Mapping 3D points onto 2D grid	O (Unsupervised)	88.4	-
Shen et al.	Symmetric function	Kernel correlation to learn corners, planarity	O P	91.0	-
Wang et al.	Symmetric function	EdgeConv: weight between a point and its neighbors	O S P	92.2	56.1

Real-time Semantic Segmentation

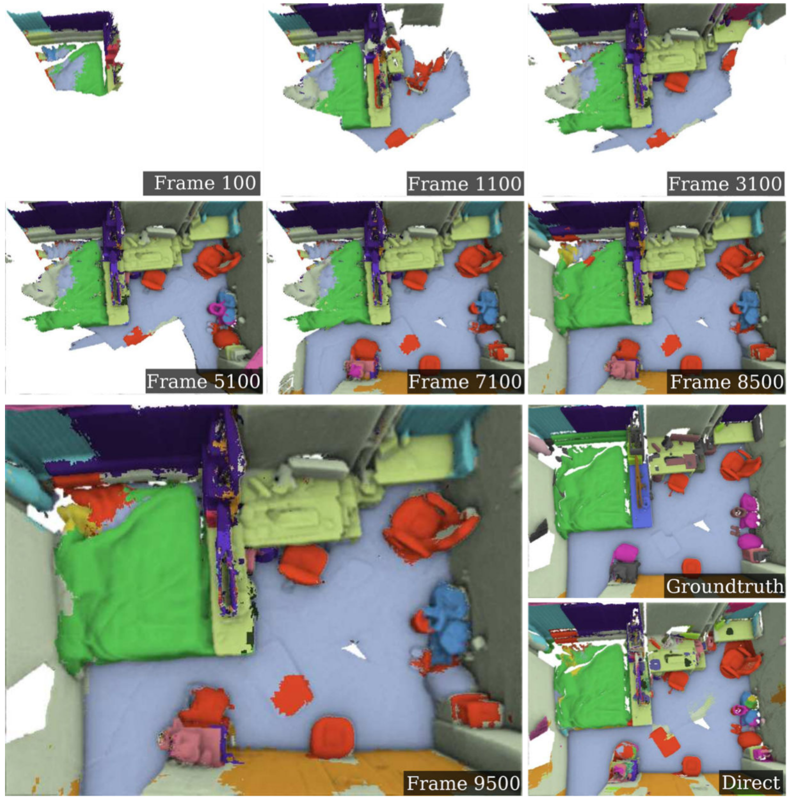
- On-the-fly 3D segmentation and reconstruction
- Using higher-order constraints from structures and objects

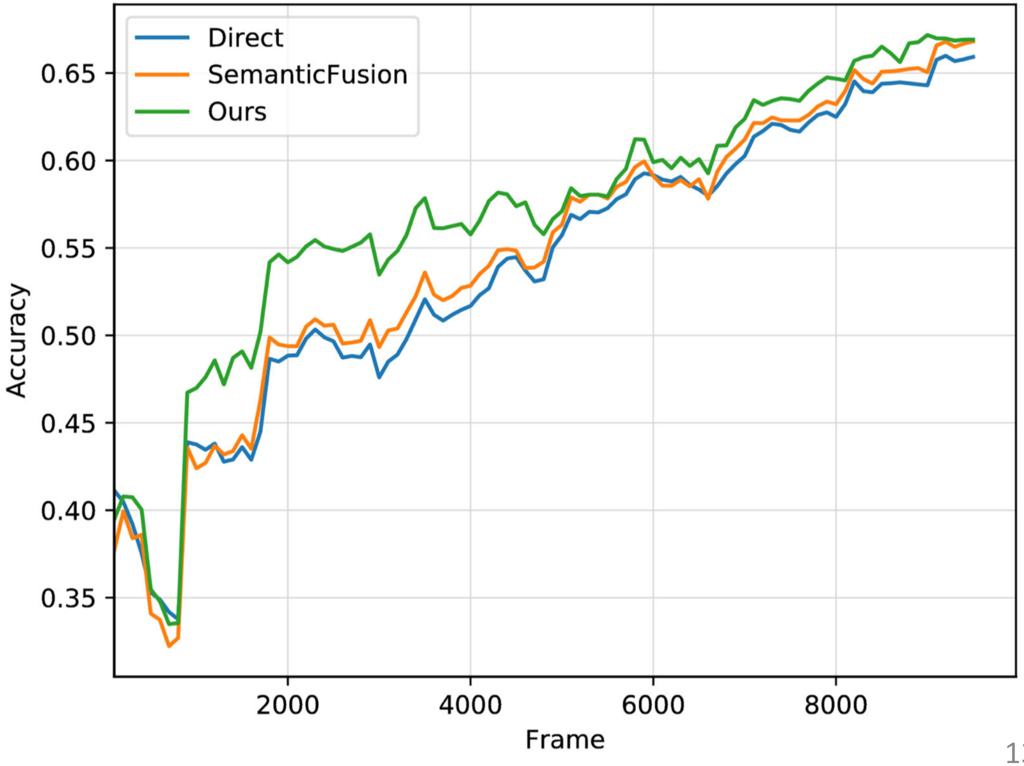


Pham et al., Real-time Progressive 3D Semantic Segmentation for Indoor Scenes, arXiv 2018

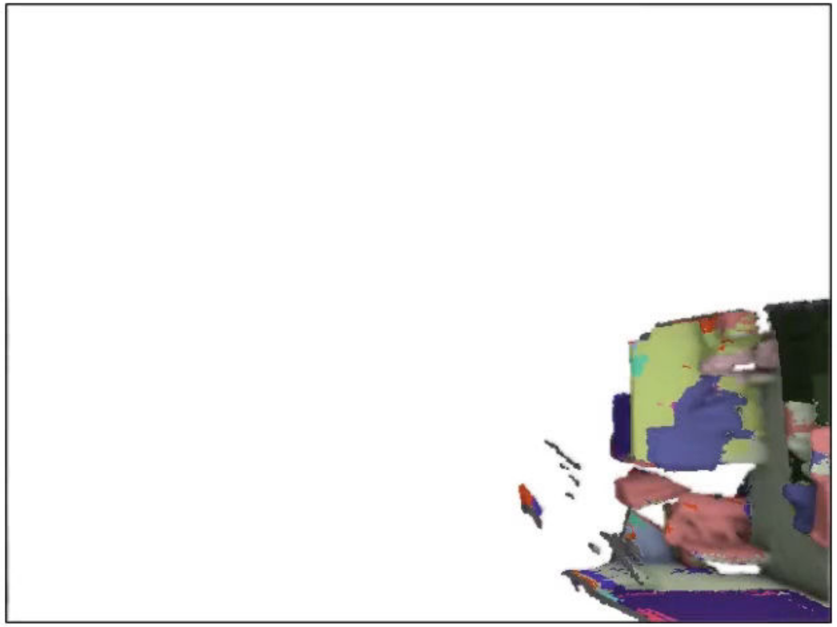
Real-time Semantic Segmentation

- Resolving semantic segmentation error while scanning
- Extensive evaluation on large-scale indoor scenes



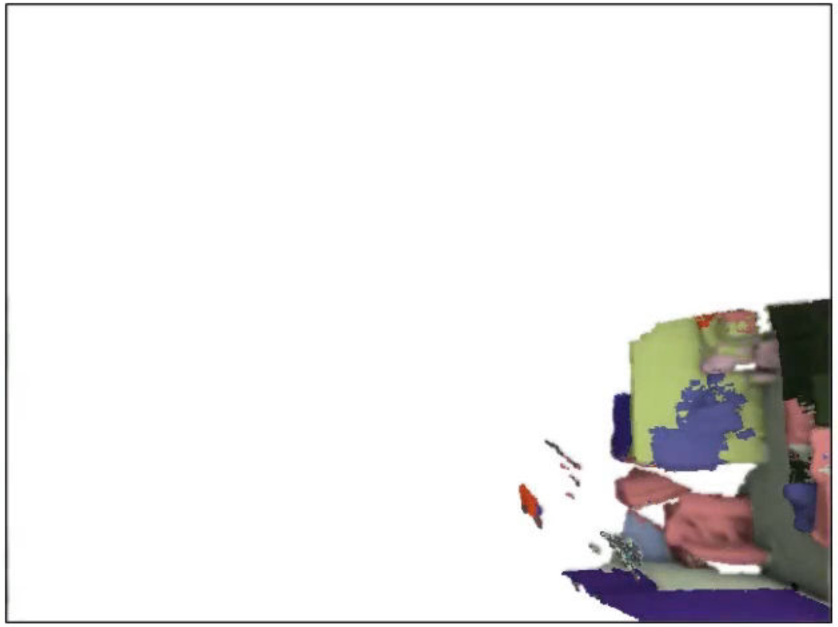


SceneNN/030: Progressive Segmentation



(a) Direct

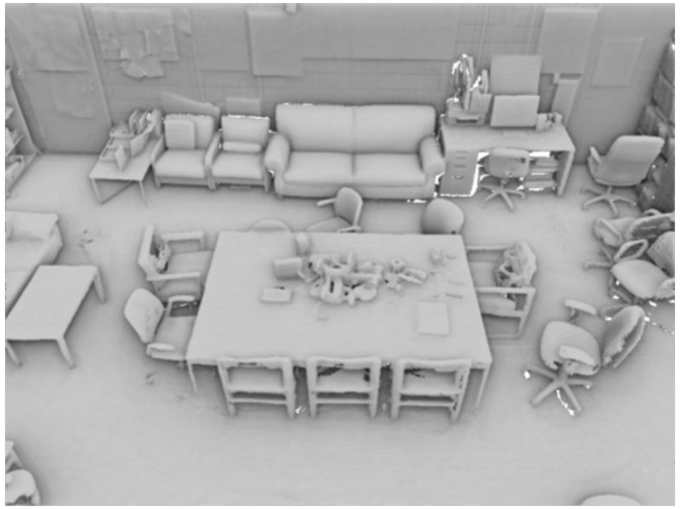
(b) SemanticFusion



(c) Ours

JOINT SEMANTIC-INSTANCE SEGMENTATION

Input



Instance

Semantic

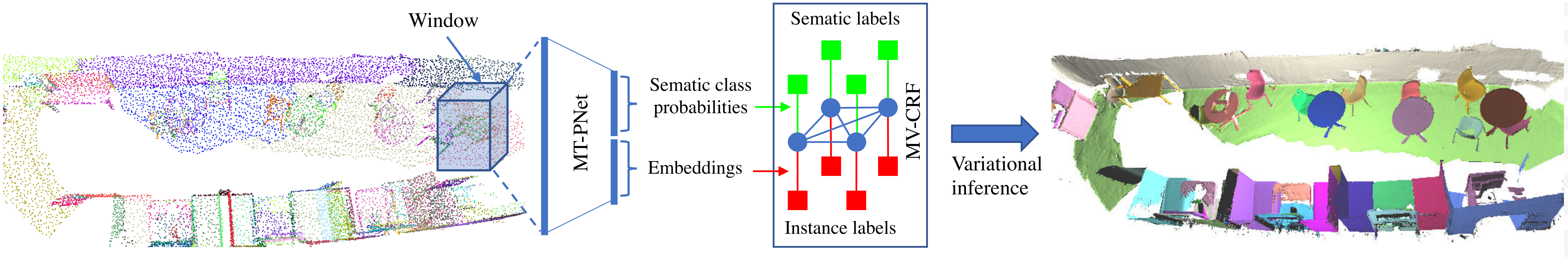
Output

Joint semantic-instance

table
sofa 1
sofa 2
chair 1
chair 2

Pham et al., JSIS3D: Joint Semantic-Instance Segmentation of 3D Point Clouds, CVPR 2019

PROPOSED METHOD

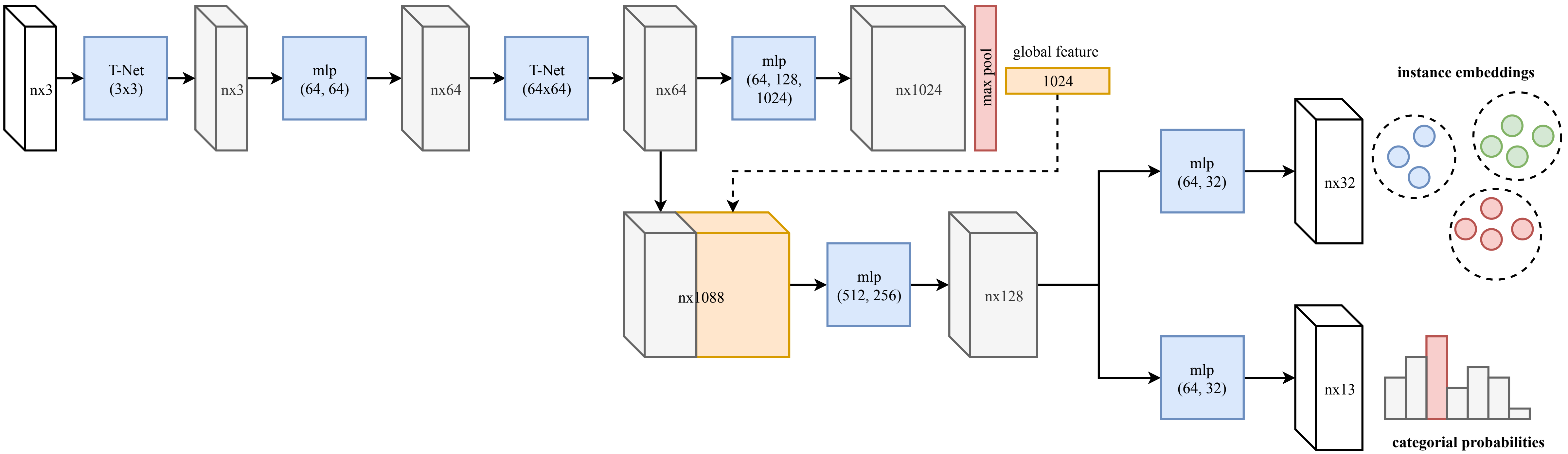


<u>Contributions</u>: **joint semantic-instance segmentation on 3D point clouds**.

- ❑ A multi-task pointwise network architecture (MT-PNet)
- ❑ Joint optimisation with a novel multi-value conditional random field model (MV-CRF)
- ❑ Extensive experiments on different indoor datasets

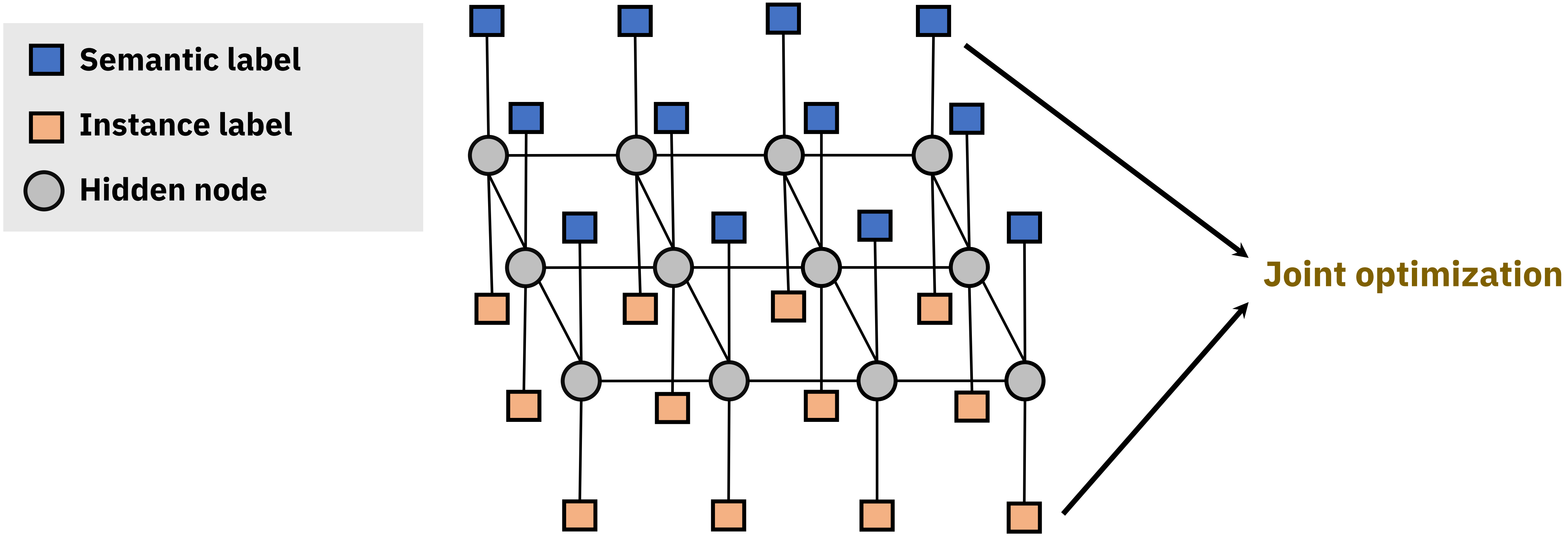
Achieve **state-of-the-art** semantic segmentation performance.

MULTI-TASK NETWORK

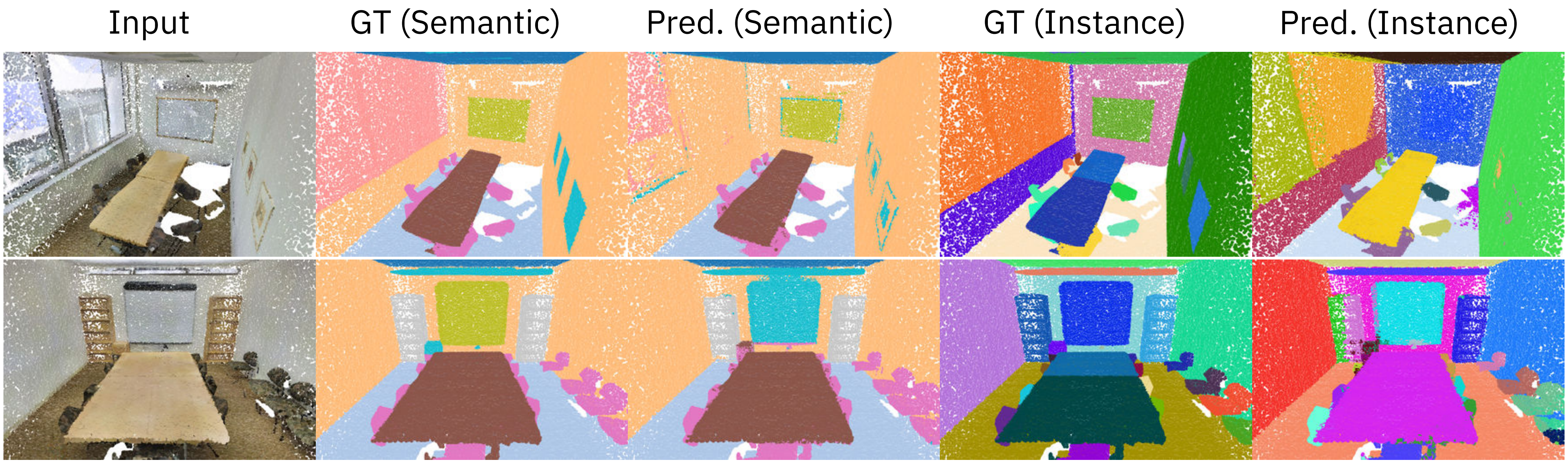


- Two branches for semantic classification and instance embedding
- Architecture based on PointNet
- The loss function is the sum of two losses: $\mathcal{L} = \mathcal{L}_{prediction} + \mathcal{L}_{embedding}$

MULTI-VALUE CRF

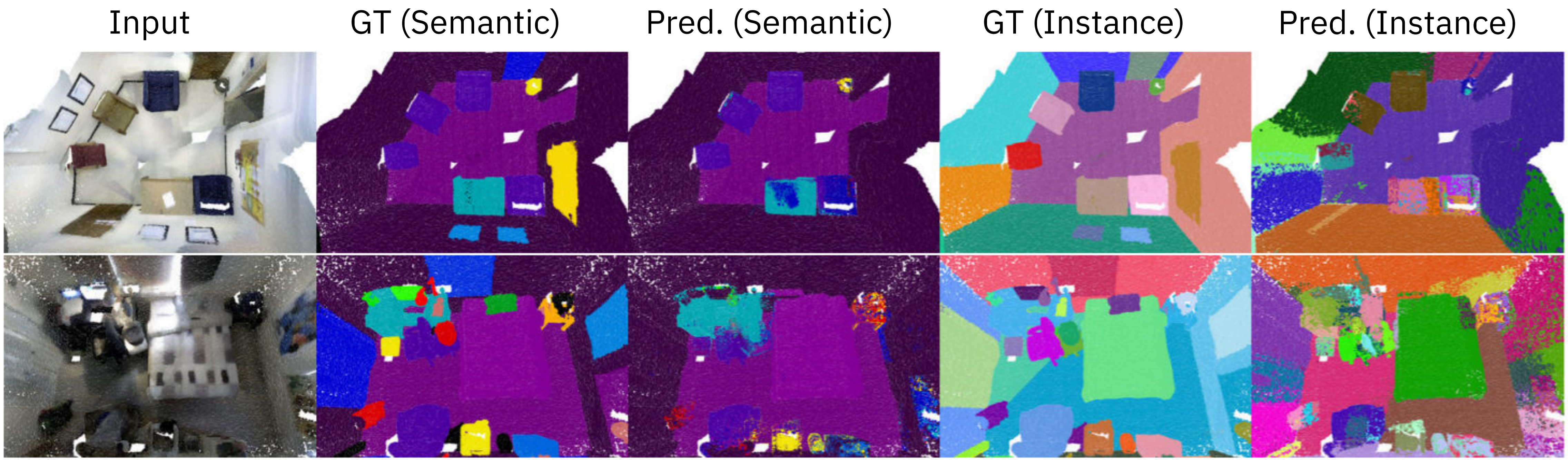


EVALUATION



S3DIS Dataset

EVALUATION



SceneNN Dataset

EVALUATION

Semantic Segmentation (accuracy)

Method	mAcc	ceiling	floor	wall	window	door	table	chair	sofa	bookcase	board	clutter
PointNet [32]	78.6	88.8	97.3	69.8	46.3	10.8	52.6	58.9	40.3	5.9	26.4	33.2
Pointwise [16]	81.5	97.9	99.3	92.7	49.6	**50.6**	74.1	58.2	0	39.3	0	61.1
SEGCloud [40]	80.8	90.1	96.1	69.9	38.4	23.1	75.9	70.4	**58.4**	40.9	13	41.6
Ours (MT-PNet)	86.7	97.4	99.6	92.7	**60.1**	26.4	**80.8**	83.7	23.7	61.1	**55.2**	**70.6**
Ours (MV-CRF)	**87.4**	**98.4**	**99.6**	**94.4**	59.7	24.9	80.6	**84.9**	30	**63.0**	52.5	70.5

Instance Segmentation (mAP)

Method	mAP	ceiling	floor	wall	window	door	table	chair	sofa	bookcase	board	clutter
Armeni *et al.* [1]	-	71.6	88.7	72.9	25.9	54.1	46	16.2	6.8	54.7	3.9	-
SGPN [44]	54.4	**79.4**	66.3	**88.8**	**66.6**	**56.8**	**46.9**	**40.8**	6.4	**47.6**	11.1	-
Ours (MT-PNet)	24.9	71.5	78.4	28.3	24.4	3.5	12.1	36.2	10	12.6	34.5	12.8
Ours (MV-CRF)	36.3	76.9	**83.6**	32.2	51.4	7.2	16.3	23.6	**16.7**	21.8	**52.1**	**13.4**

Future Works

- Additional cues for point cloud deep learning: edge, triangle
- 3D point cloud networks for real-time semantic predictions
- Apply point cloud learning to object pose estimation, instance segmentation

Q&A

Acknowledgement

- Tan-Sang Ha, Quang-Trung Truong, Fangyu Lin, Guoxuan Zhang for helping with data capture and tool development.
- Minh-Khoi Tran, Tian Feng, Zhipeng Mo, Benjamin Kang, Daniel Teo, Xuequan Lu, William Lai for helping with user study.

National Research Foundation
Prime Minister's Office
Singapore

Singapore University of Technology and Design

UMass Boston

Deakin University

Ministry of Education
Singapore

香港科技大學
The Hong Kong University of Science and Technology

DManD
SUTD Digital Manufacturing and Design Centre

National Heritage Board

NVIDIA

NSF

MIT

東京大学
The University of Tokyo