3D Scene Understanding from RGB-D Images

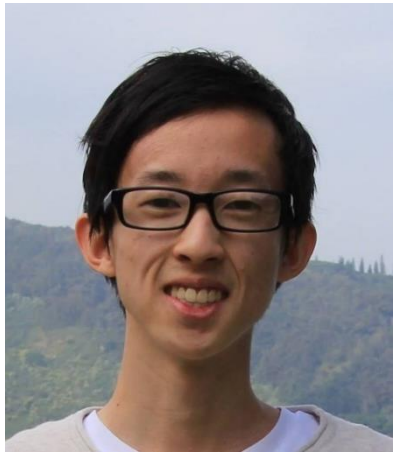
Thomas Funkhouser

Disclaimer: I am talking about the work of these people …

Current Ph.D. Students

Shuran Song

Yinda Zhang



Andy Zeng

Maciej Halber

Kyle Genova

Recent Ph.D. Student

Fisher Yu

Current Postdocs

Manolis Savva

Angel Chang

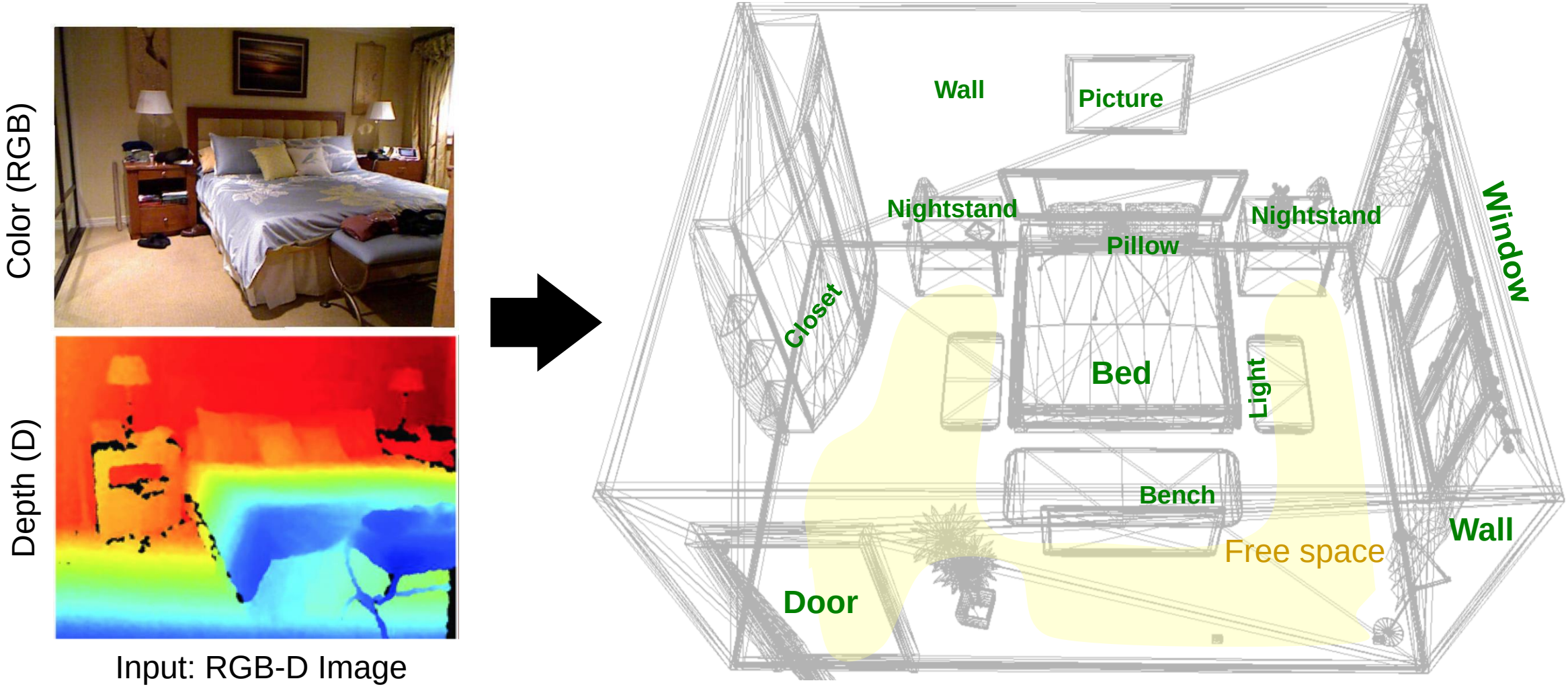
Motivation

Help devices with RGB-D cameras understand their 3D environments

- Robot manipulation
- Augmented reality
- Virtual reality
- Personal assistance
- Surveillance
- Navigation
- Mapping
- Games
- etc.

Goal

Given a RGB-D image, infer a complete, annotated 3D representation

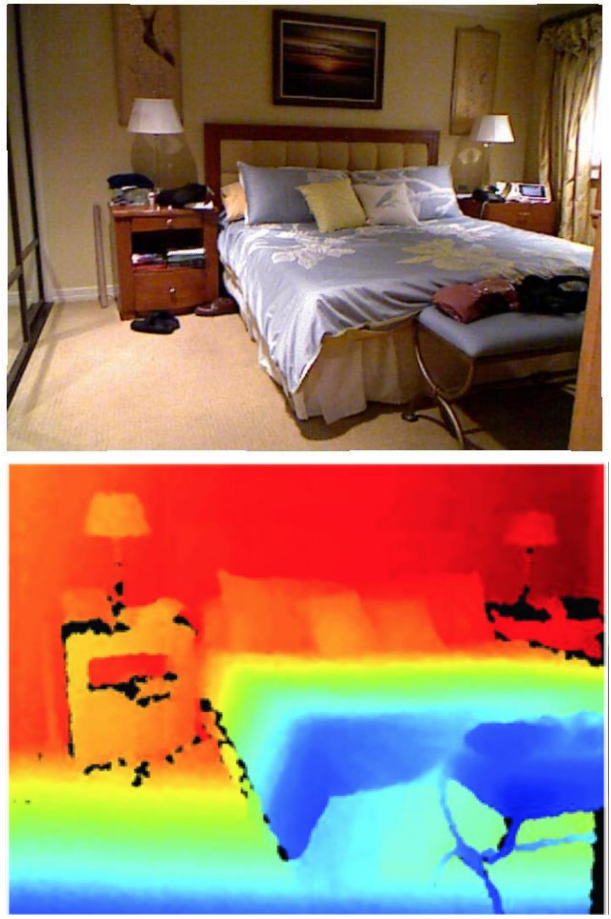


Input: RGB-D Image

Output: complete, annotated 3D representation

Problem

Challenge: get only partial observation of scene, must infer the rest

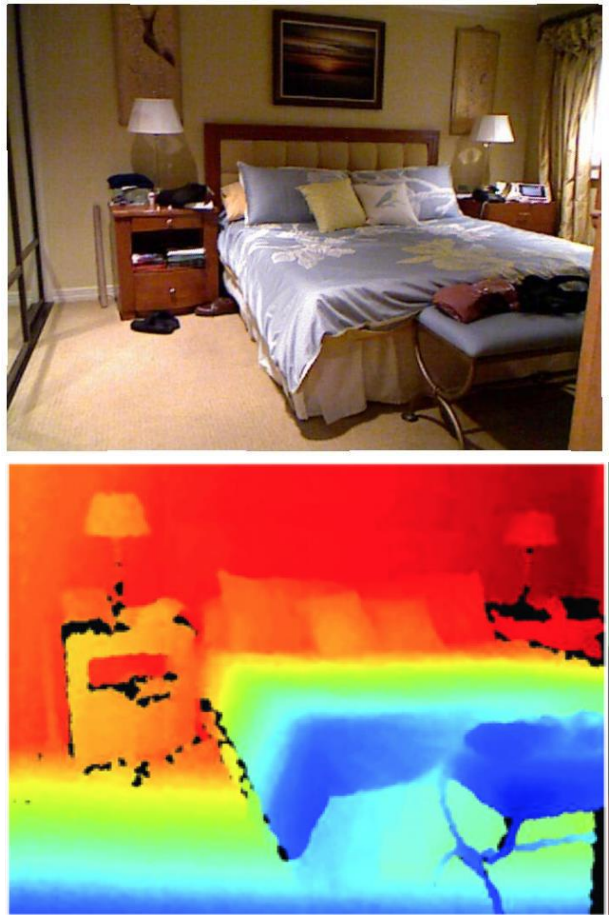


Input: RGB-D Image

Side view

Problem

Challenge: get only partial observation of scene, must infer the rest

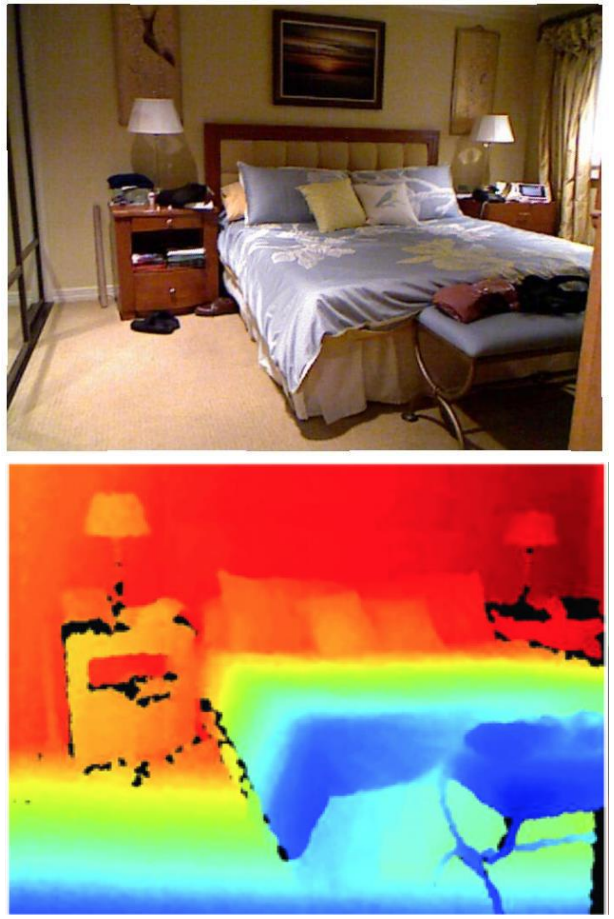


Input: RGB-D Image

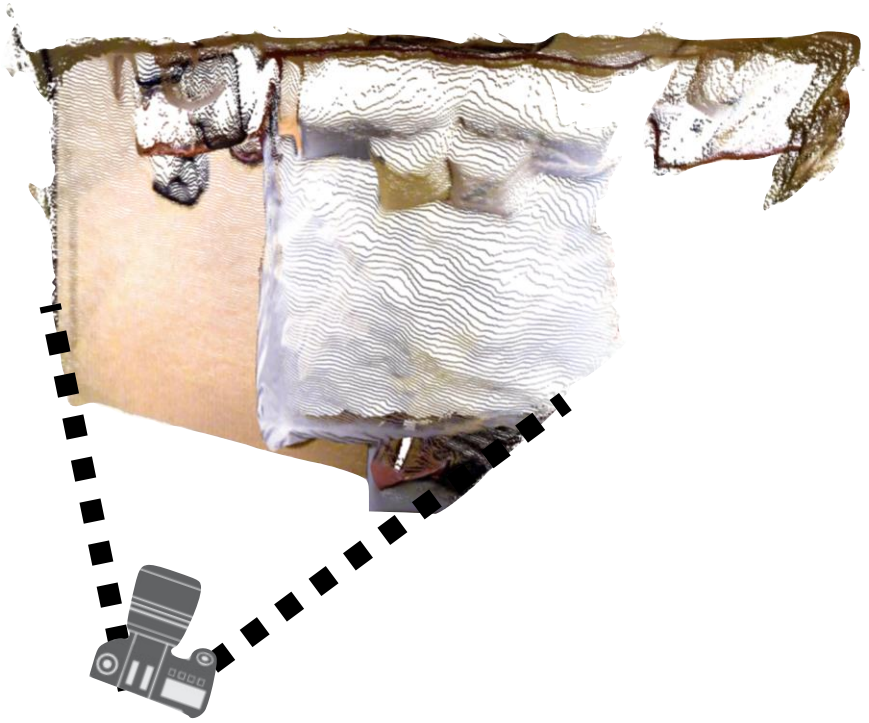
Rotating side view

Problem

Challenge: get only partial observation of scene, must infer the rest



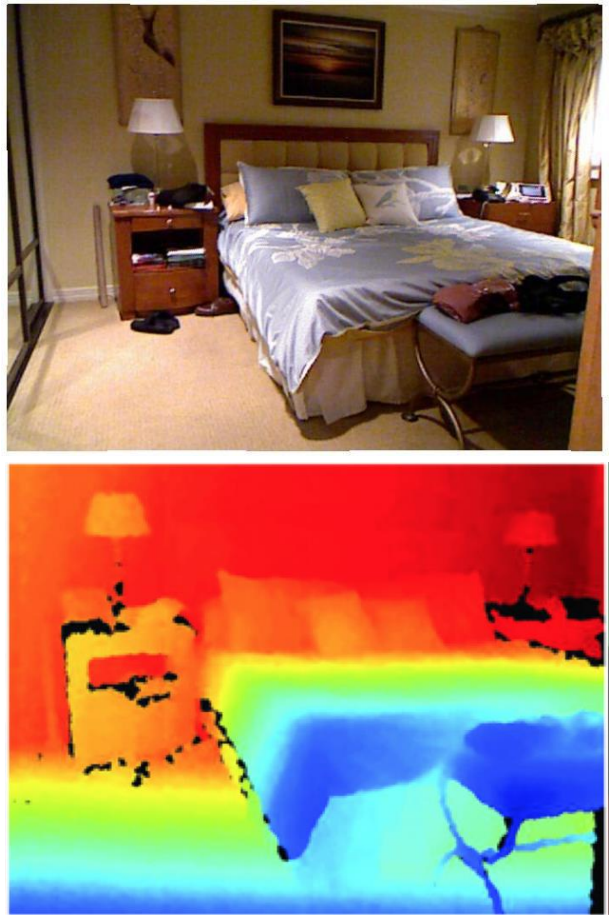
Input: RGB-D Image



Top view

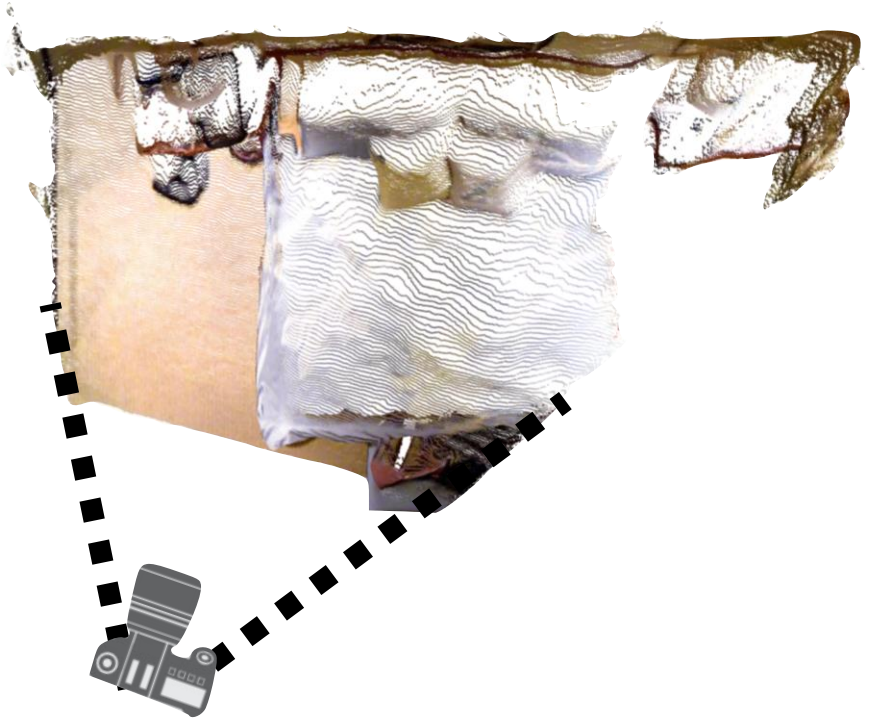
Problem

Challenge: get only partial observation of scene, must infer the rest



Input: RGB-D Image

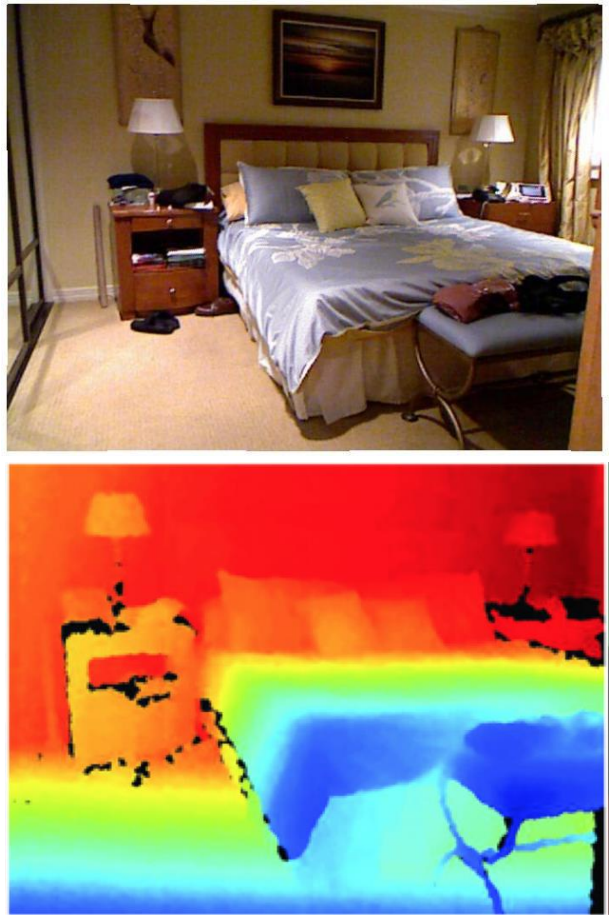
Beyond
Field of View



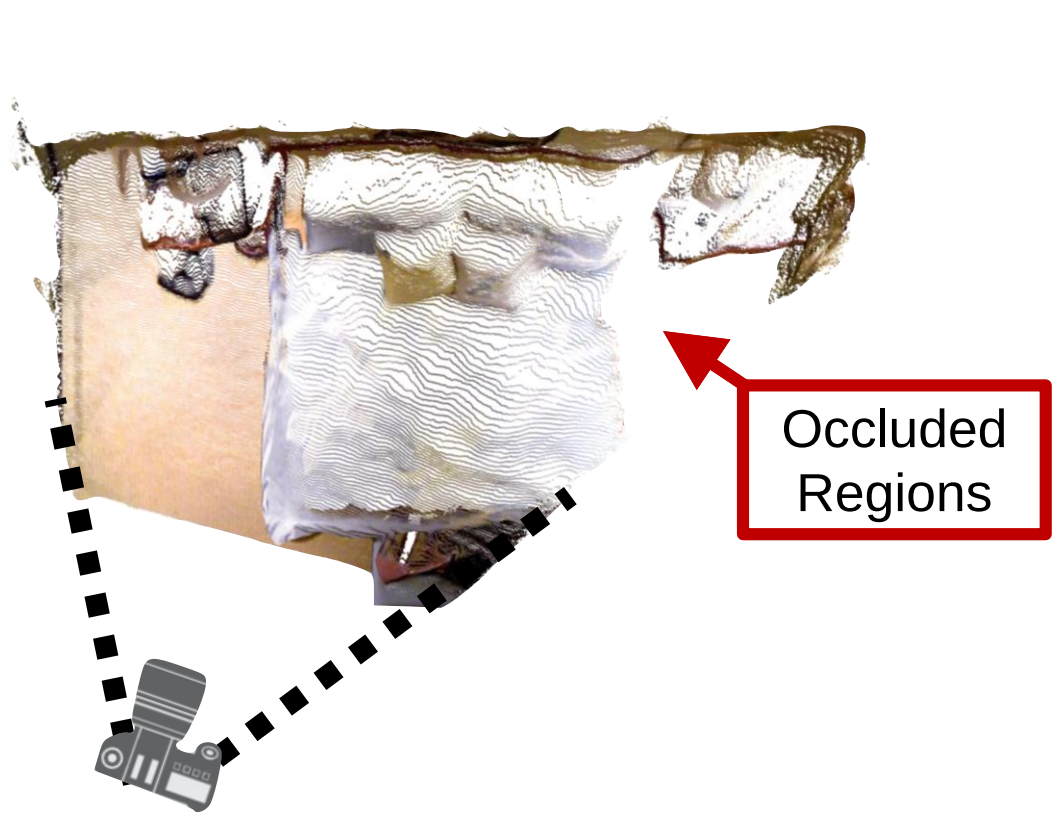
Top view

Problem

Challenge: get only partial observation of scene, must infer the rest



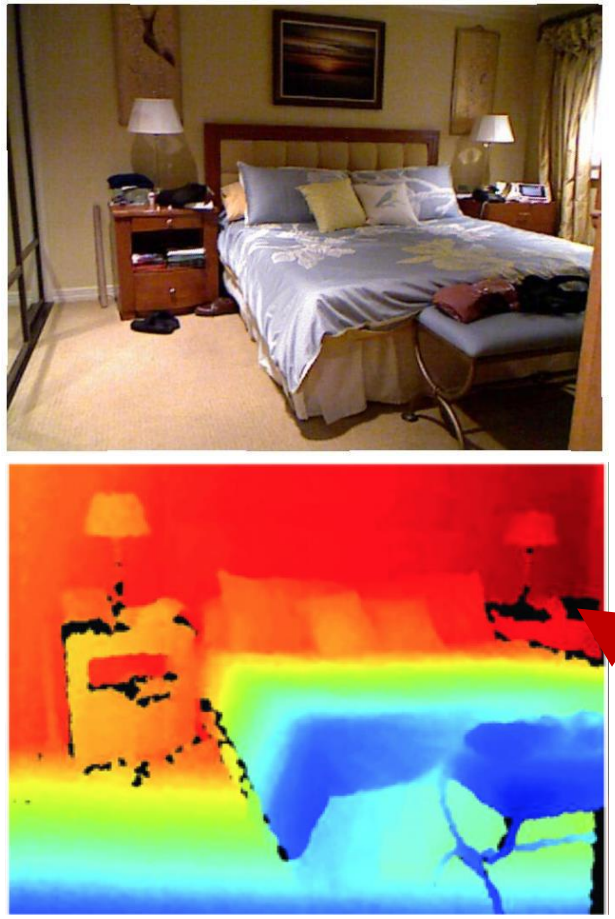
Input: RGB-D Image



Top view

Problem

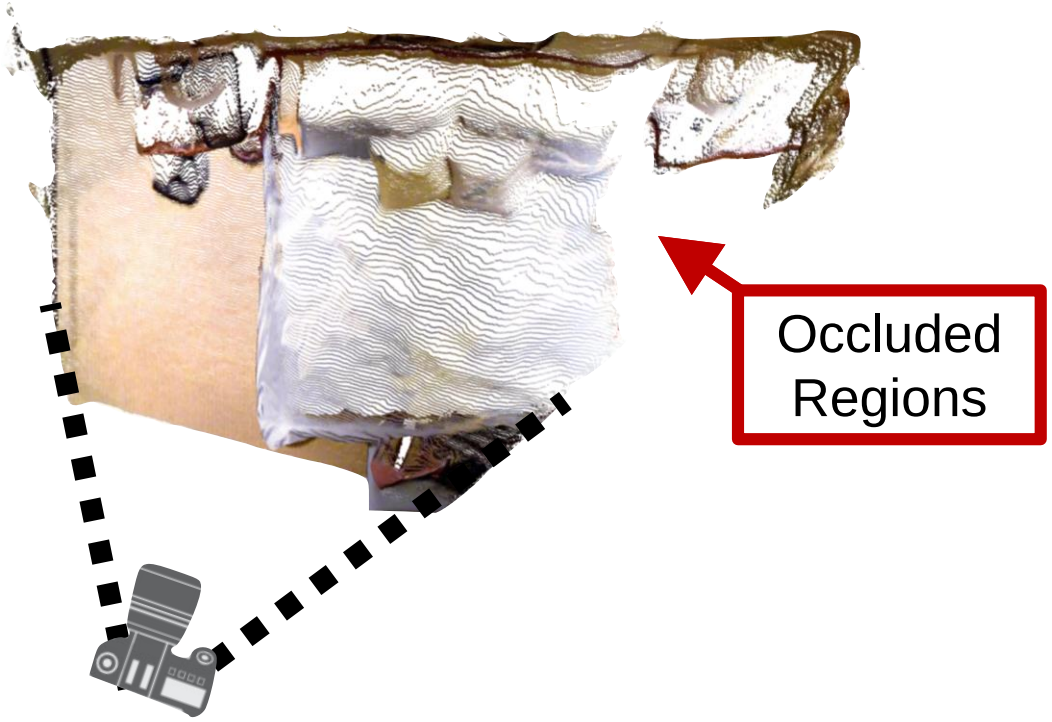
Challenge: get only partial observation of scene, must infer the rest



Input: RGB-D Image

Beyond Field of View

Missing Depths

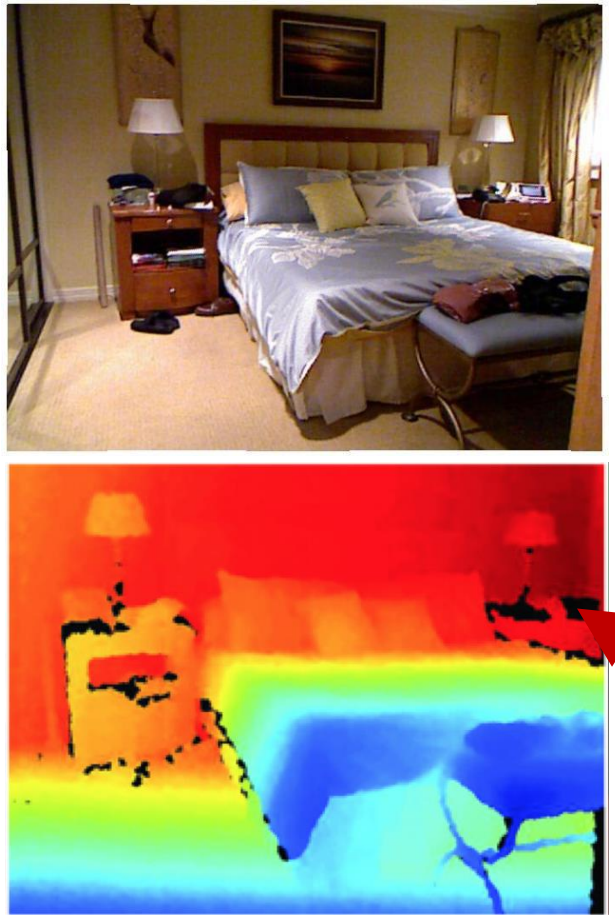


Occluded Regions

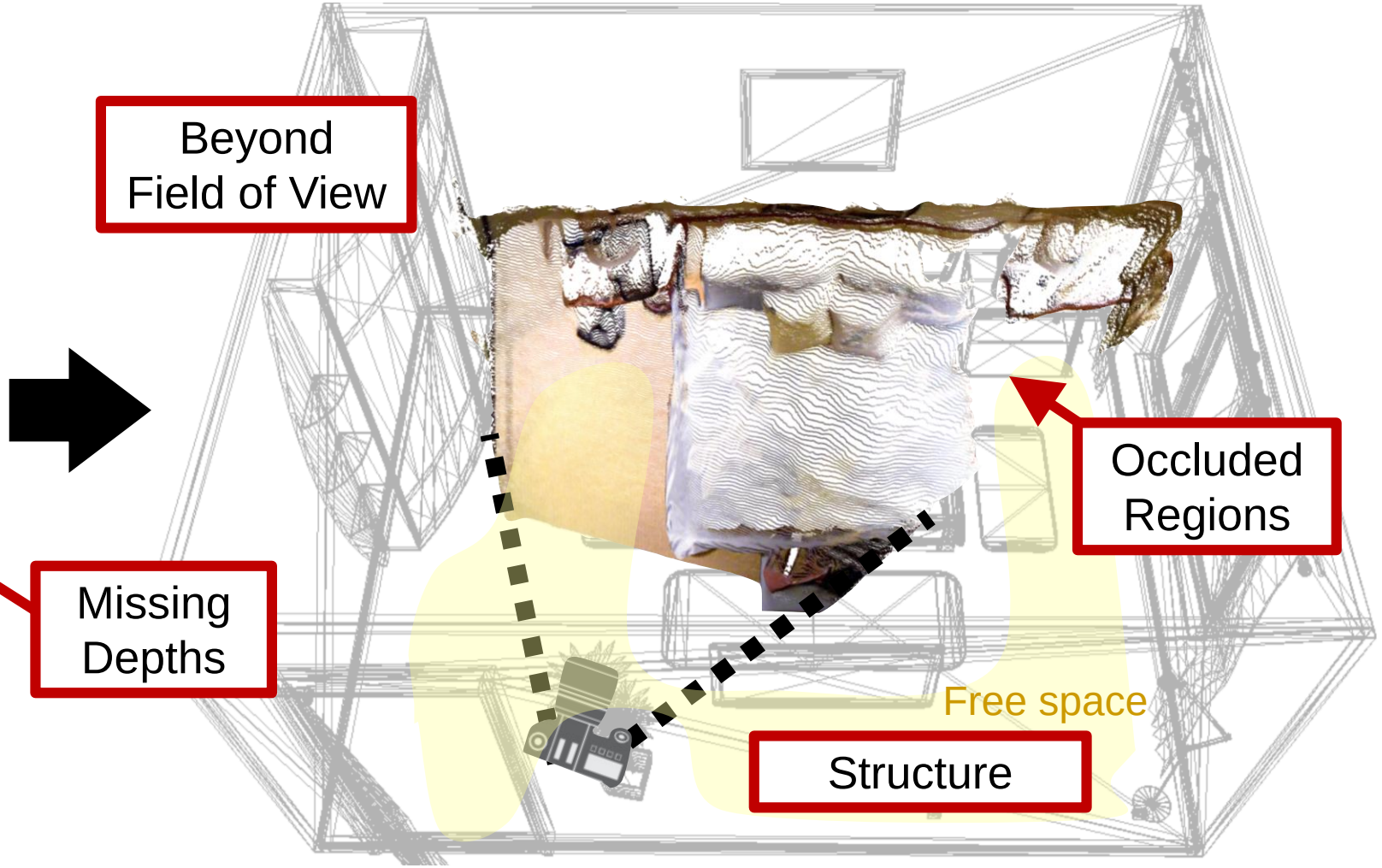
Top view

Problem

Challenge: get only partial observation of scene, must infer the rest



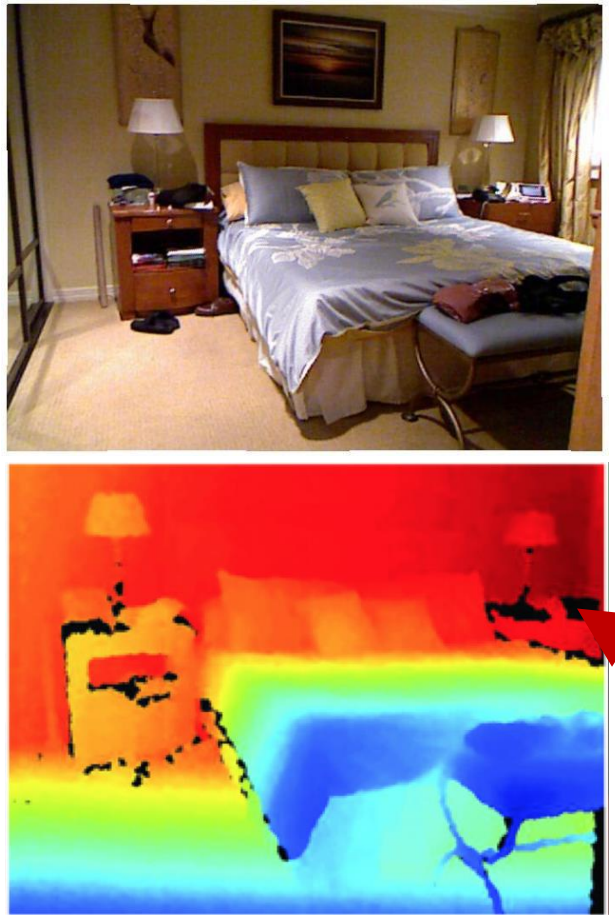
Input: RGB-D Image



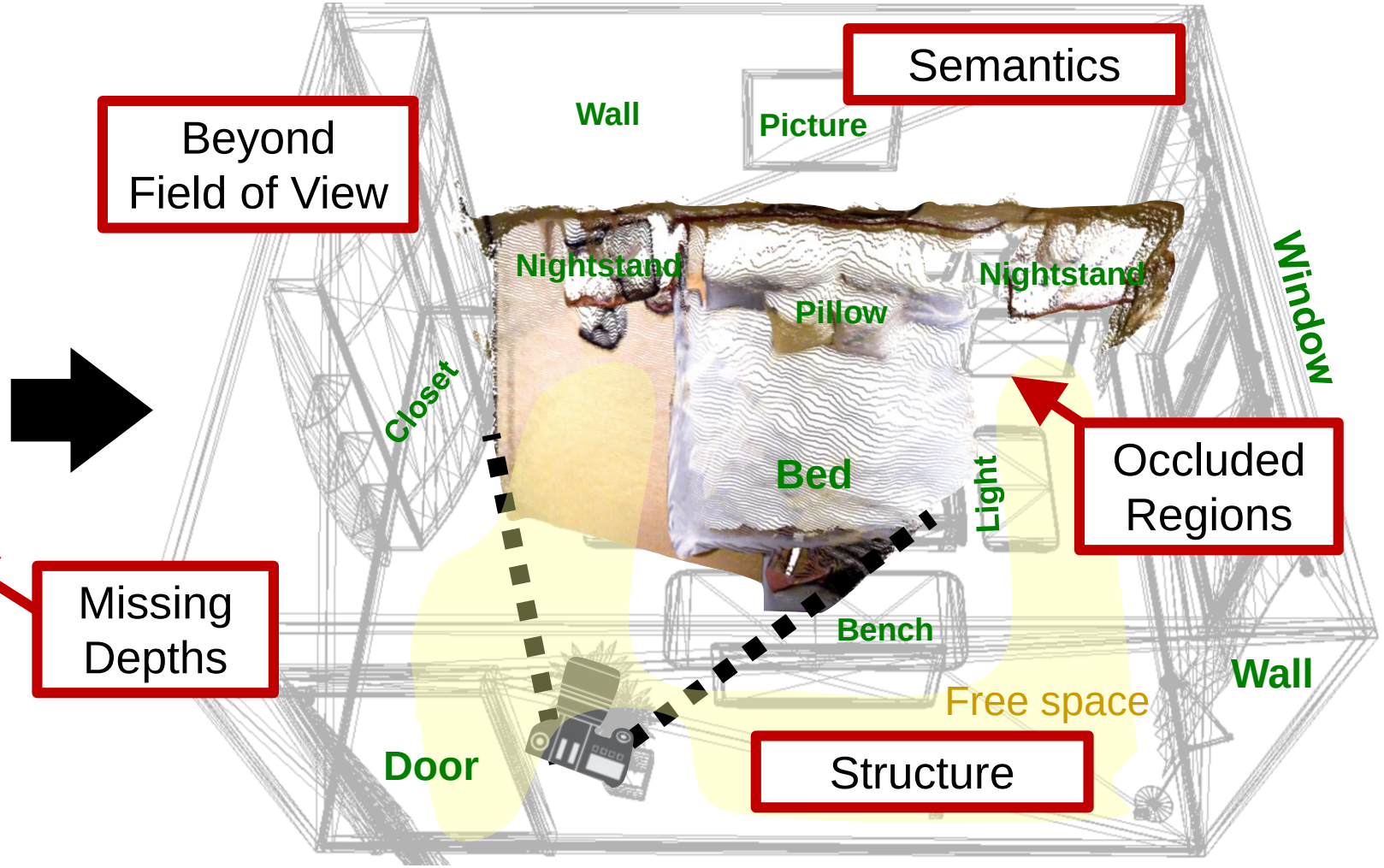
Top view

Problem

Challenge: get only partial observation of scene, must infer the rest



Input: RGB-D Image



Top view

Talk Outline

Introduction

Three recent projects

- Deep depth completion [CVPR 2018]
- Semantic scene completion [CVPR 2017]
- Semantic view extrapolation [CVPR 2018]

Common themes

Future work

Talk Outline (Part 1)

Introduction

Three recent projects

→ Deep depth completion [CVPR 2018]

- Semantic scene completion [CVPR 2017]
- Semantic view extrapolation [CVPR 2018]

Common themes

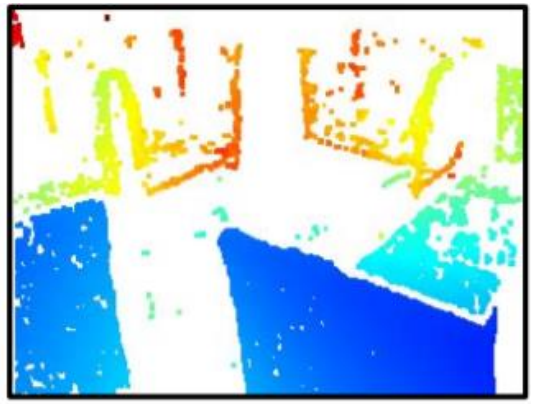
Future work

Yinda Zhang and Thomas Funkhouser,
"Deep Depth Completion of a Single RGB-D Image,"
CVPR 2018 (spotlight on Tuesday)

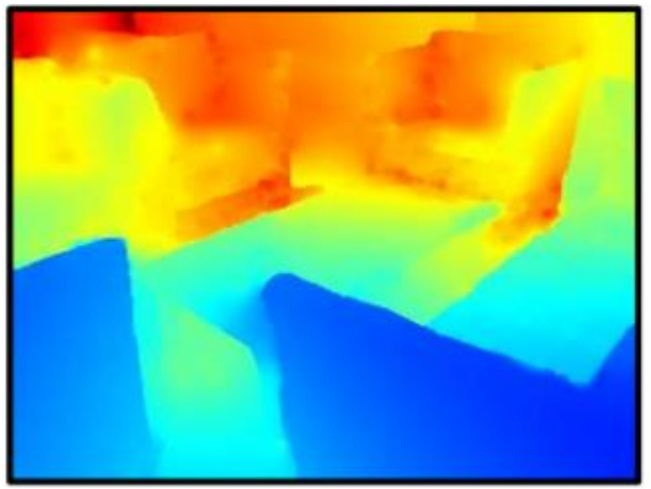
Deep Depth Completion

Goal: estimate depths missing from an RGB-D image

Color (RGB)



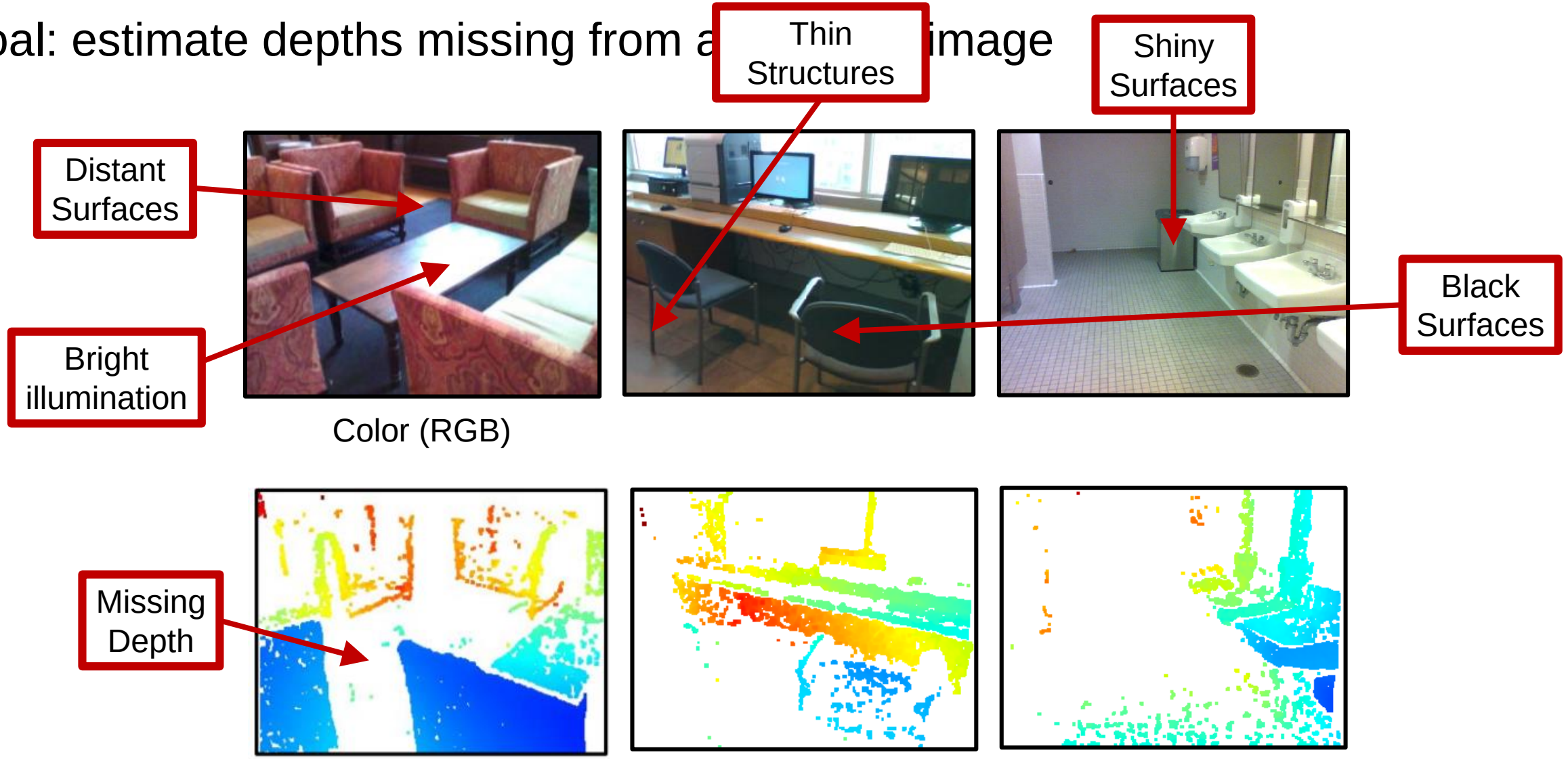
Raw Depth (D)



Output Depth (D)

Deep Depth Completion

Goal: estimate depths missing from a image



Deep Depth Completion

Motivation: help upstream applications “understand” 3D environment

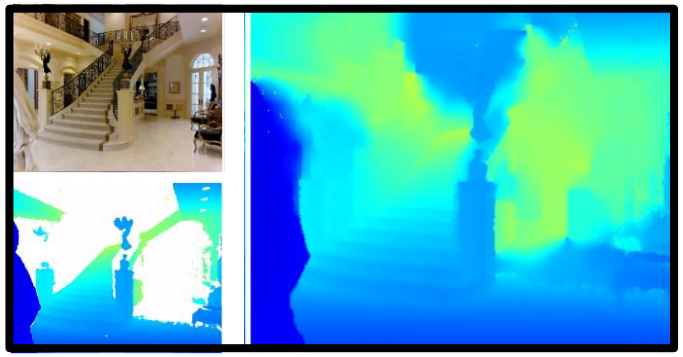
Raw Depth

Output Depth

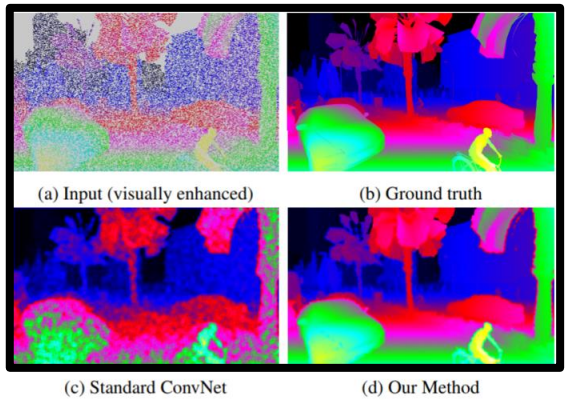
RGB-D images shown as colored 3D point clouds

Deep Depth Completion

Previous work on depth completion (from RGB-D):

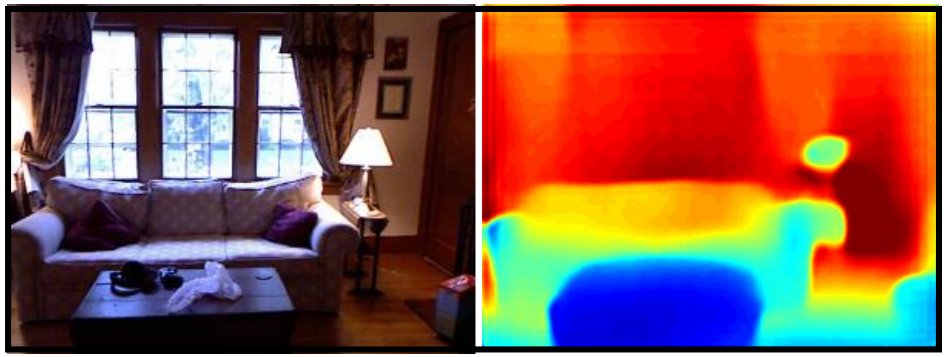


Joint Bilateral Filter
[Silberman, 2012]

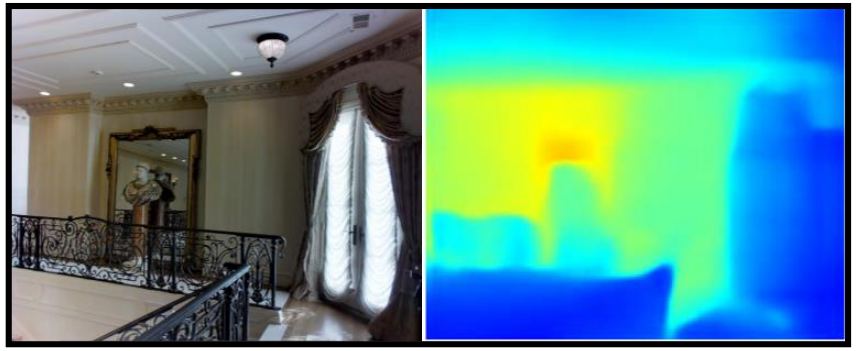


Sparsity Invariant CNNs
[Uhrig, 2017]

Previous work on depth estimation (from RGB):



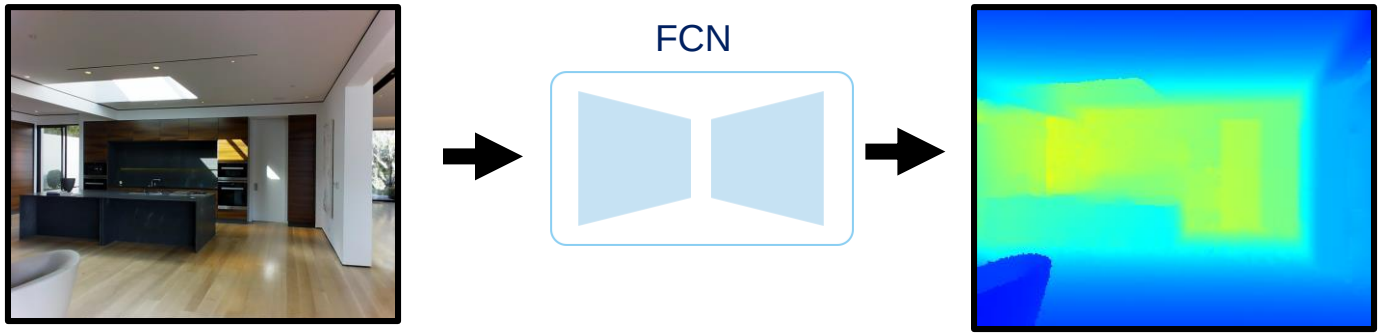
Deeper Depth Prediction
[Laina, 2016]



Harmonizing Overcomplete Predictions
[Chakrabarti, 2016]

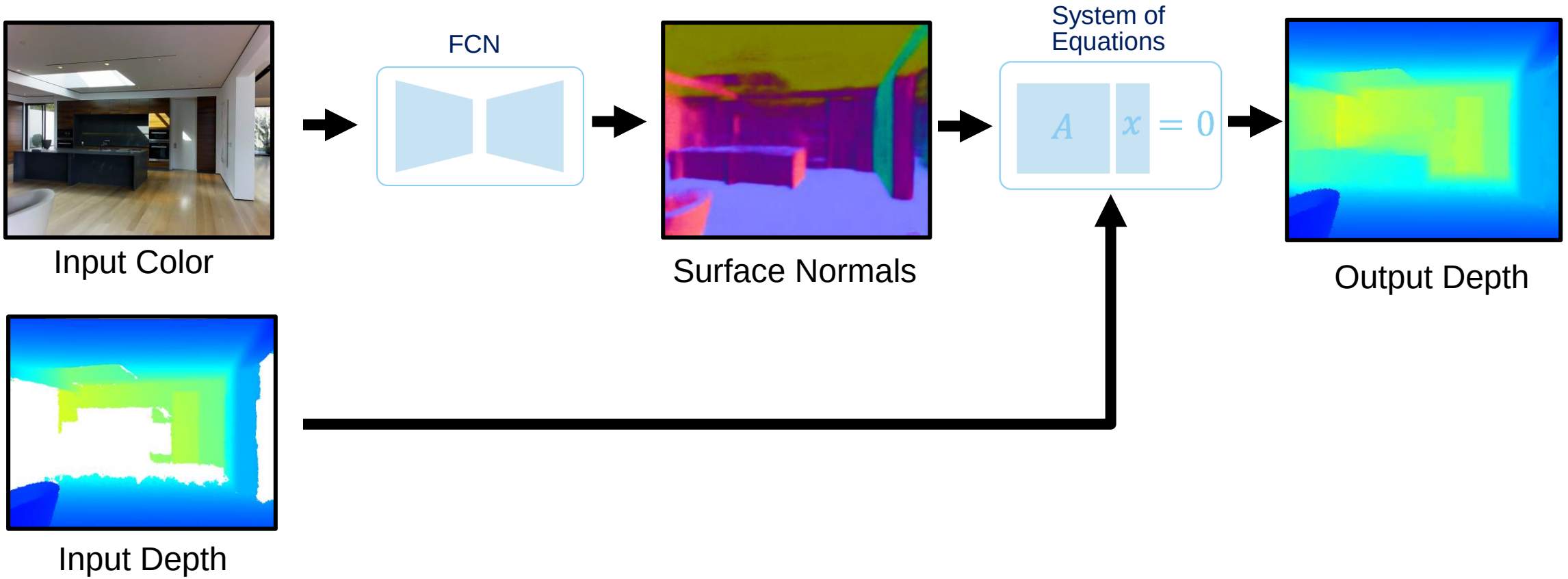
Deep Depth Completion

Problem: estimating depth from color requires global scene understanding



Deep Depth Completion

Approach: estimate local surface normals from color,
and then solve for depths globally with system of equations

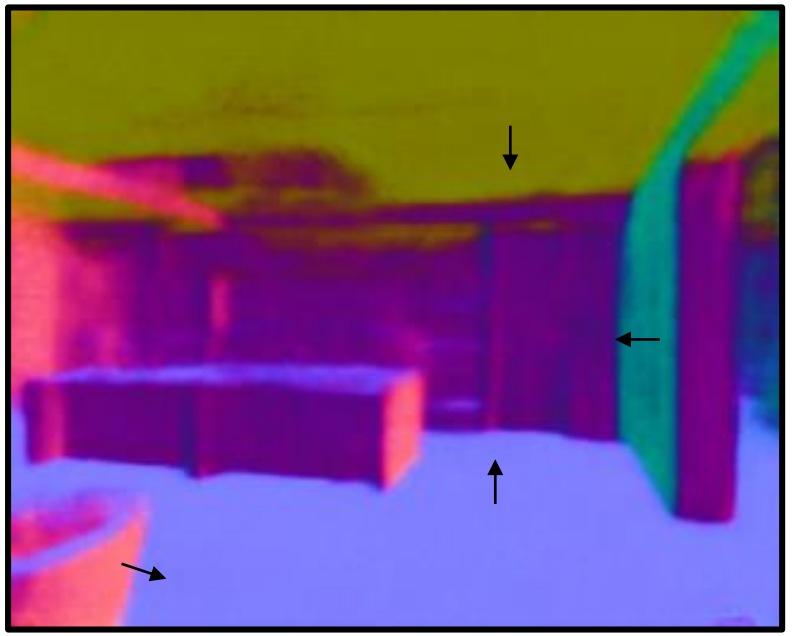


Deep Depth Completion

Rationale 1: estimating surface normals is easier than estimating depths

- Constant within planar regions
- Determined by local shading (for diffuse surfaces)
- Often associated with specific textures

Color



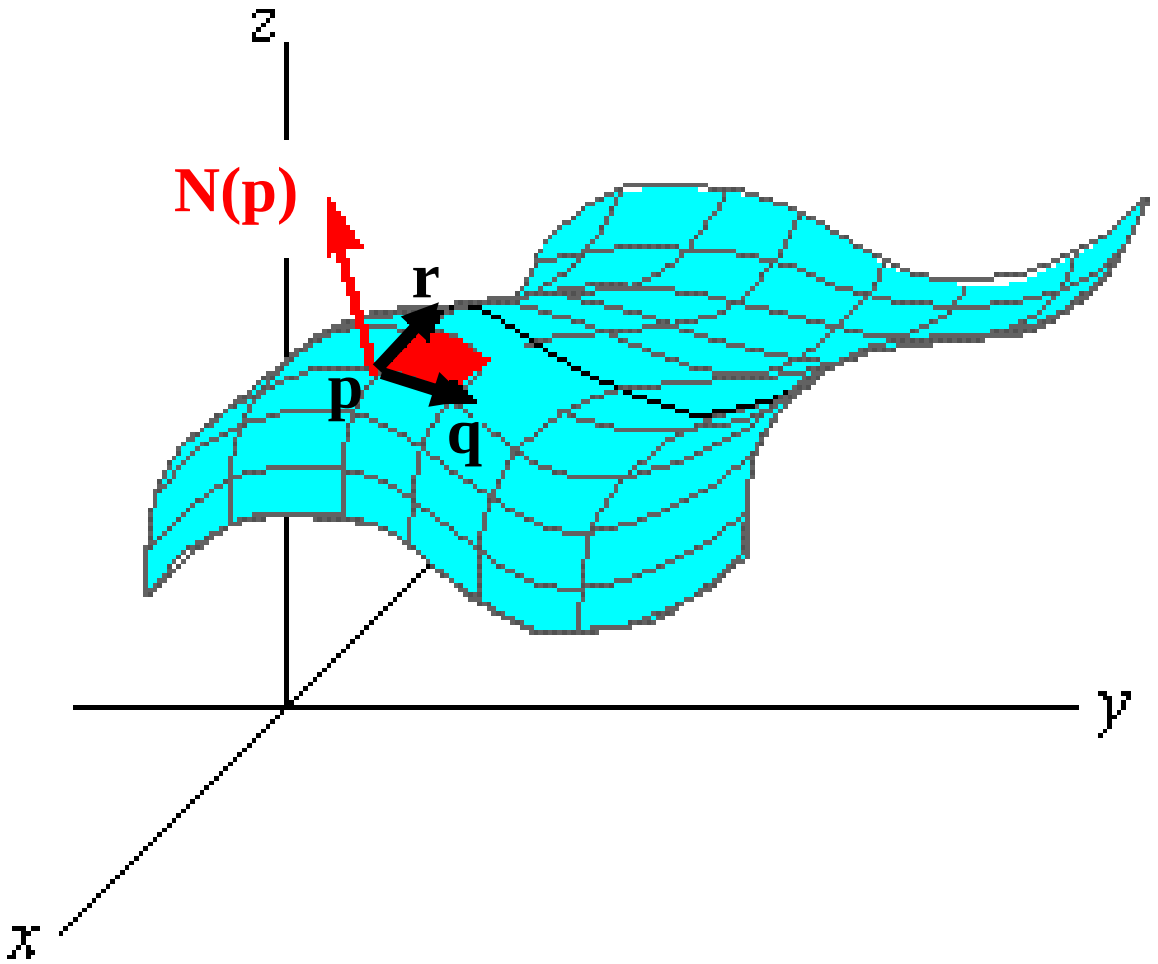
Estimated Surface Normals

Y. Zhang, S. Song, E. Yumer, M. Savva, J.-Y. Lee, H. Jin, T. Funkhouser, "Physically-Based Rendering for Indoor Scene Understanding Using Convolutional Neural Networks," CVPR 2017

Deep Depth Completion

Rationale 2: depths can be estimated robustly from normals

- Solution is unique for each continuously connected component (up to scale)



Non-linear system of equations:

$$N(p) = (v(p,q) \times v(p,r))/\|(v(p,q) \times v(p,r))\|$$

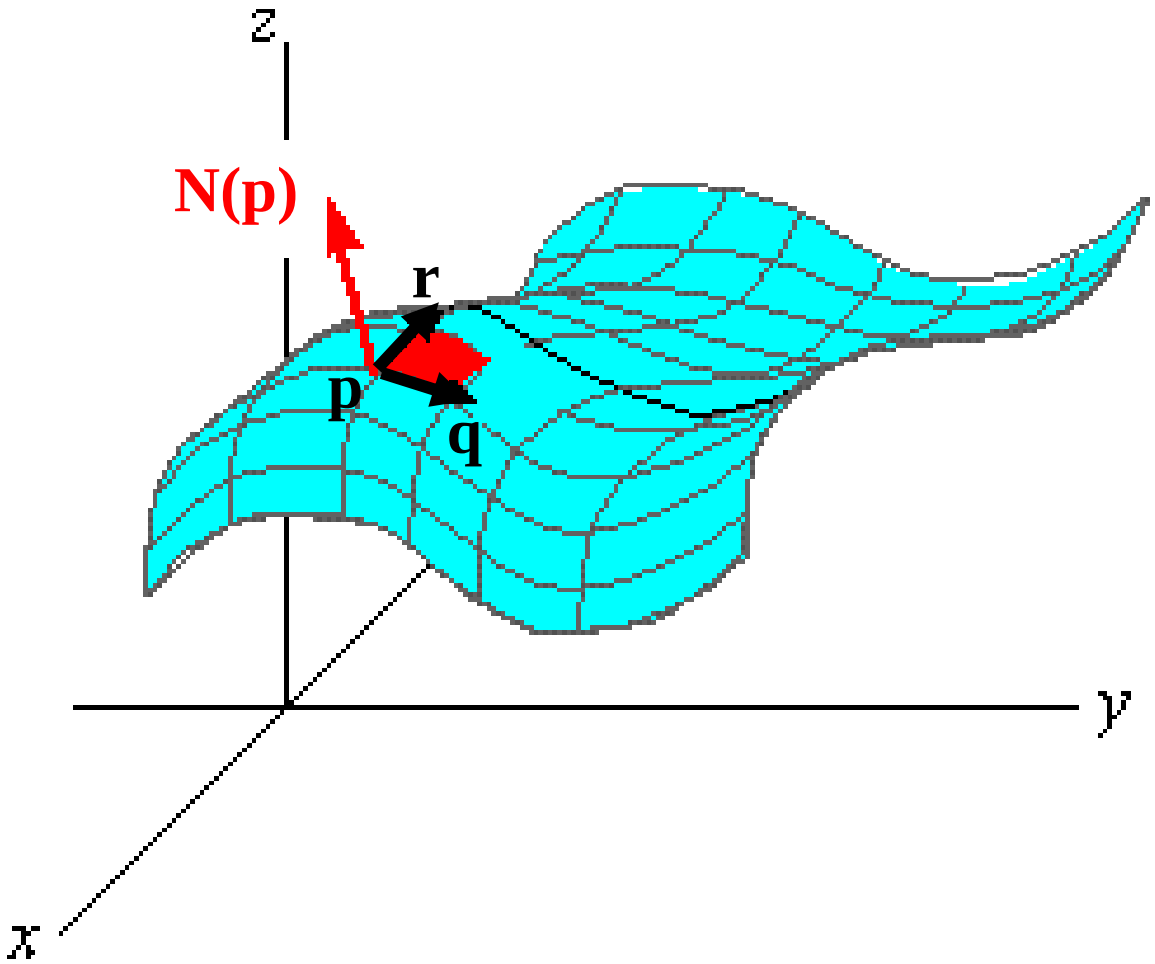
Linear approximation:

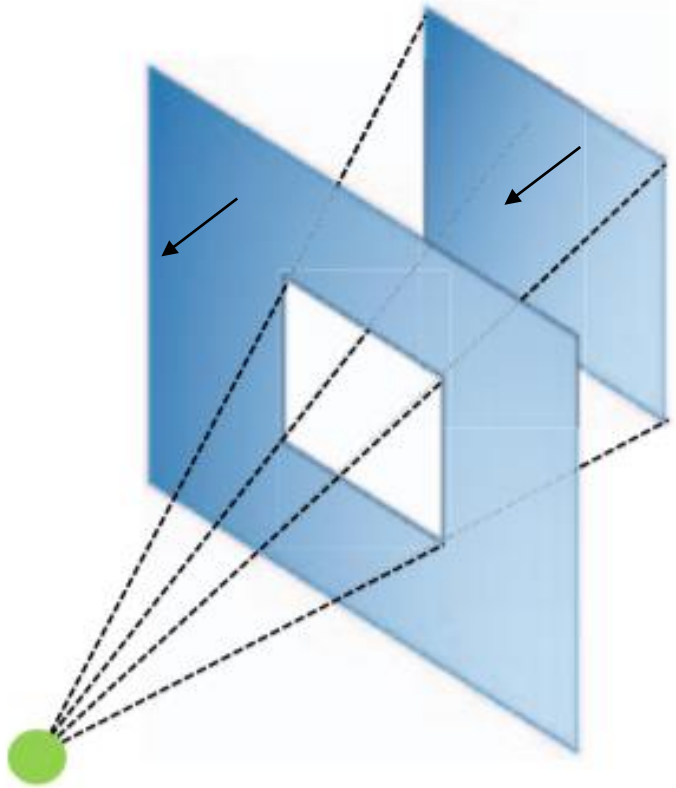
$$N(p) \cdot v(p,q) = 0$$
$$N(p) \cdot v(p,r) = 0$$

Deep Depth Completion

Rationale 2: depths can be estimated robustly from normals

- Solution is unique for each continuously connected component (up to scale)





Deep Depth Completion

Rationale 2: depths can be estimated robustly from normals

- Real-world scenes generally have few (one) continuously connected components

Deep Depth Completion

Rationale 2: depths can be estimated robustly from normals

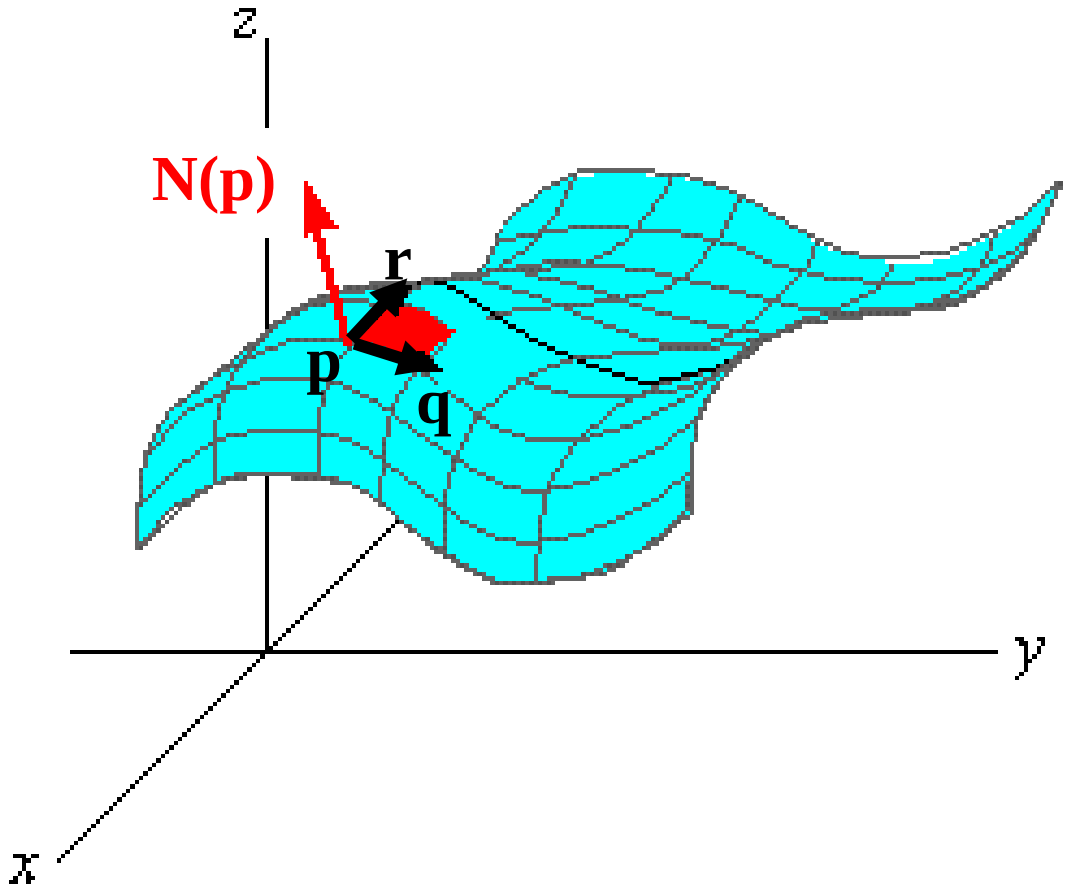
- We use observed depths and smoothness constraints to guarantee a solution

$$E = \lambda_D E_D + \lambda_S E_S + \lambda_N E_N B$$

$$E_D = \sum_{p \in T_{obs}} ||D(p) - D_0(p)||^2$$

$$E_N = \sum_{p,q \in N} || < v(p,q), N(p) > ||^2$$

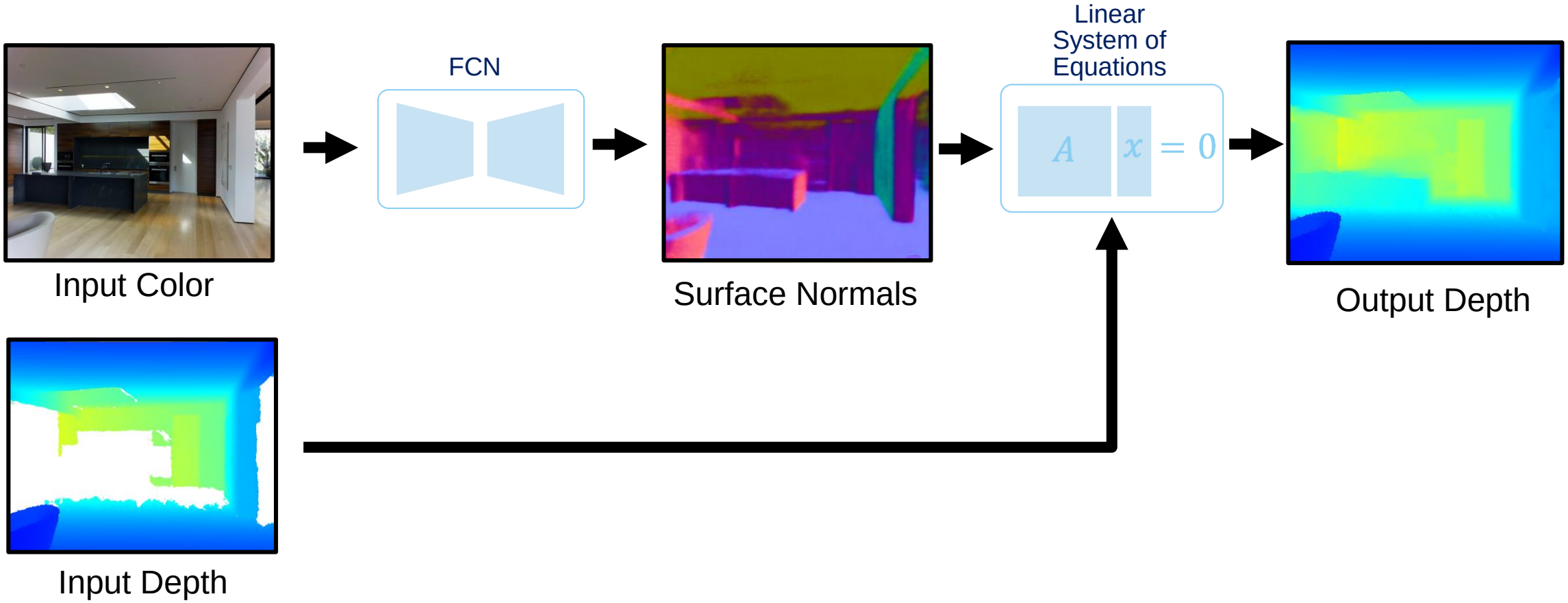
$$E_S = \sum_{p,q \in N} ||D(p) - D(q))|^2$$



Deep Depth Completion

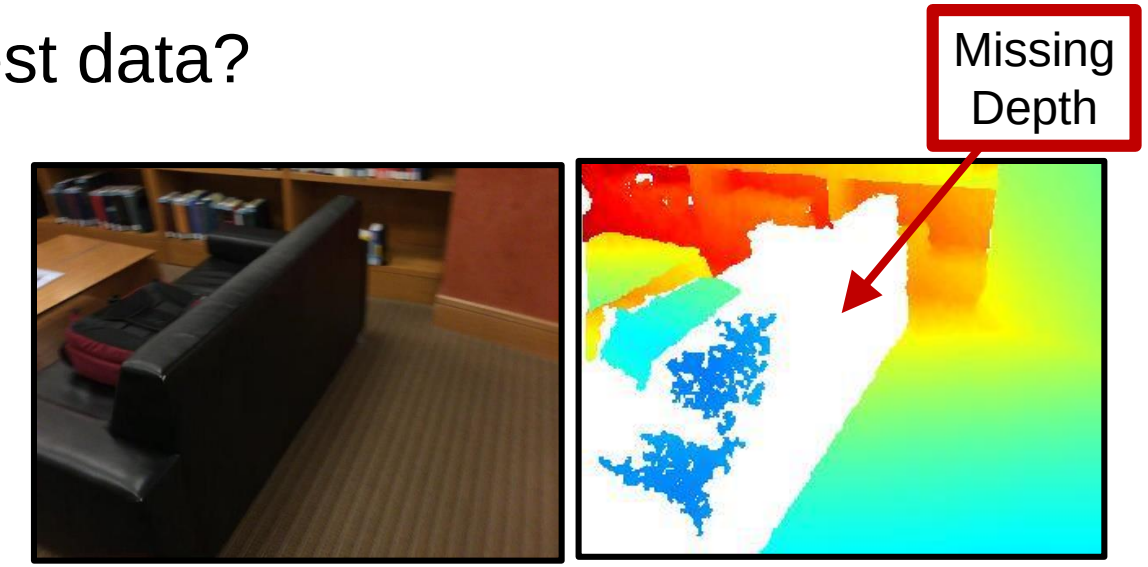
Rationale 2: depths can be estimated robustly from normals

- Solving the linearized equations guarantees a globally optimal solution



Deep Depth Completion: Data

Where get real training/test data?



Color

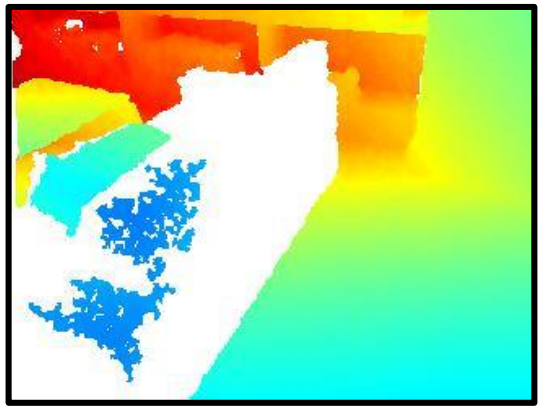
Raw Depth

Deep Depth Completion: Data

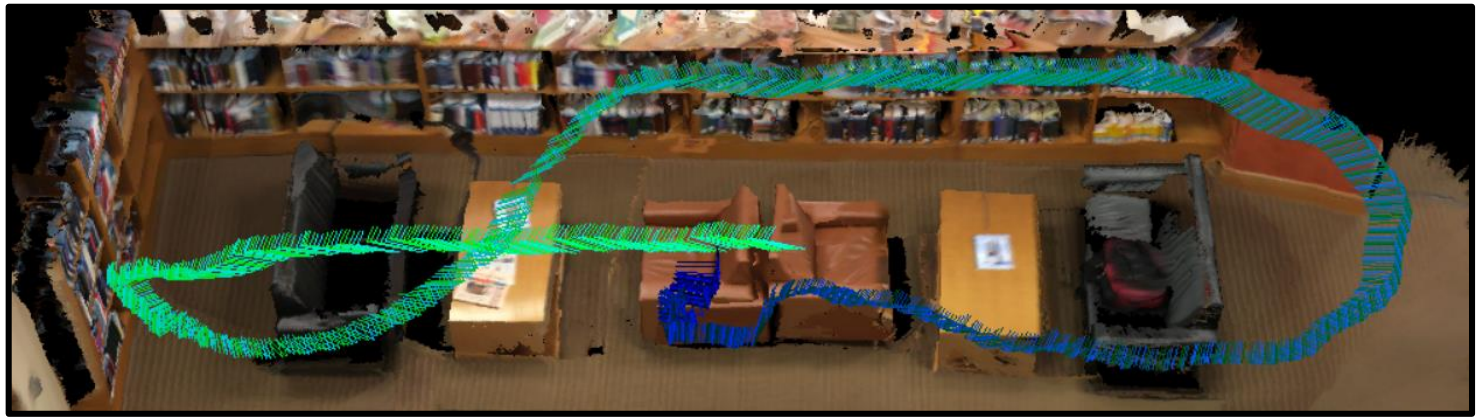
Where get real training/test data?

- Complete depths by rendering RGB-D SLAM surface reconstructions (ScanNet, Matteport3D)

Color



Raw Depth



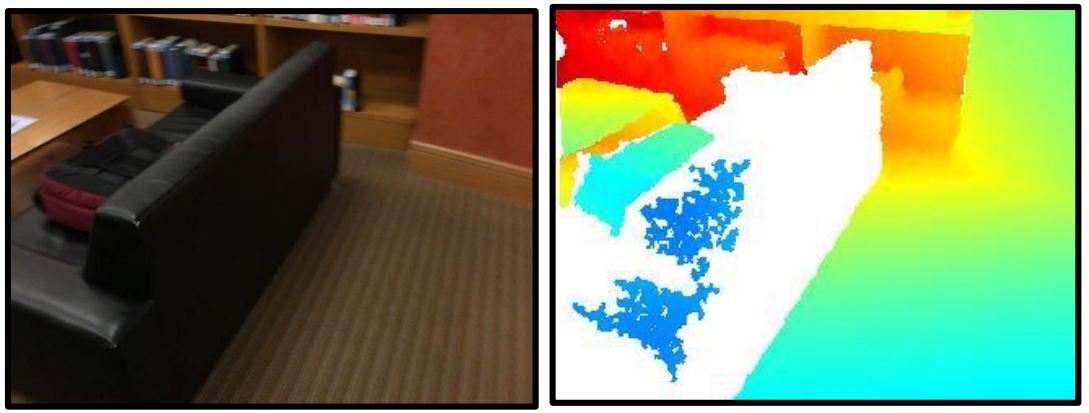
ScanNet Surface Reconstruction

A. Dai, A.X. Chang, M. Savva, M. Halber, T. Funkhouser, M. Niessner., "ScanNet: Richly-annotated 3D Reconstructions of Indoor Scenes," CVPR 2017

Deep Depth Completion: Data

Where get real training/test data?

- Complete depths by rendering RGB-D SLAM surface reconstructions (ScanNet, Matteport3D)



Color Raw Depth

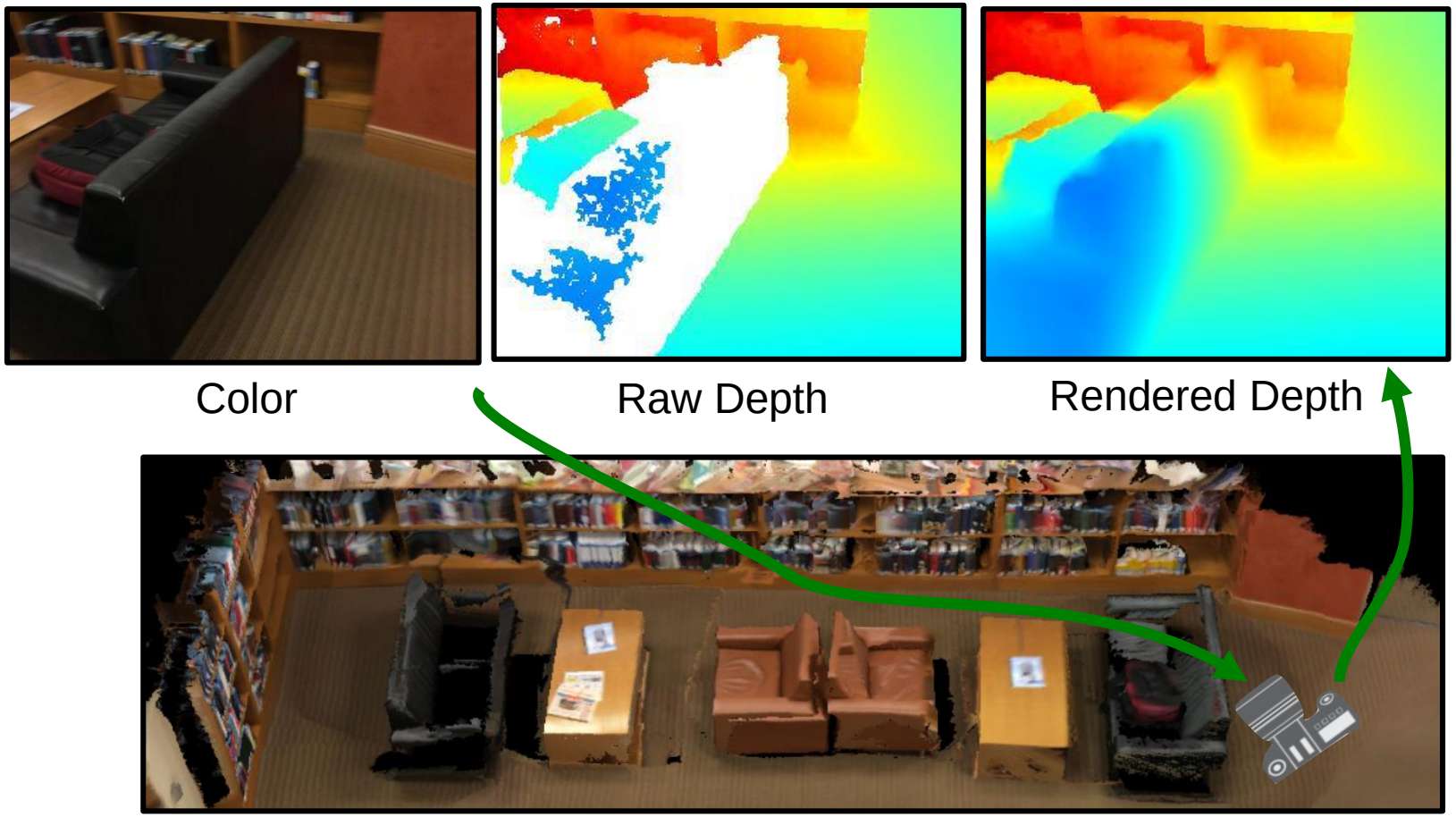
ScanNet Surface Reconstruction

A. Dai, A.X. Chang, M. Savva, M. Halber, T. Funkhouser, M. Niessner., "ScanNet: Richly-annotated 3D Reconstructions of Indoor Scenes," CVPR 2017

Deep Depth Completion: Data

Where get real training/test data?

- Complete depths by rendering RGB-D SLAM surface reconstructions (ScanNet, Matteport3D)



A. Dai, A.X. Chang, M. Savva, M. Halber, T. Funkhouser, M. Niessner., "ScanNet: Richly-annotated 3D Reconstructions of Indoor Scenes," CVPR 2017

Deep Depth Completion: Results

Comparisons to other depth completion methods:

Method	Rel↓	RMSE↓	1.05↑	1.10↑	1.25↑	1.25^2↑	1.25^3↑
Smooth	0.151	0.187	32.80	42.71	57.61	72.29	80.15
Bilateral [64]	0.118	0.152	34.39	46.50	61.92	75.26	81.84
Fast [5]	0.127	0.154	33.65	45.08	60.36	74.52	81.79
TGV [20]	0.103	0.146	37.40	48.75	62.97	75.00	81.71
Garcia et.al [6]	0.115	0.144	36.78	47.13	61.48	74.89	81.67
FCN [13]	0.167	0.241	16.43	31.13	57.62	75.63	84.01
Ours	**0.089**	**0.116**	**40.63**	**51.21**	**65.35**	**76.74**	**82.98**

[5] J. T. Barron and B. Poole. The fast bilateral solver. ECCV 2016.
[6] D. Garcia. Robust smoothing of gridded data in one and higher dimensions with missing values. Comp. stat. & data anal., 2010.
[13] Y. Zhang et al. Physically-based rendering for indoor scene understanding using convolutional neural networks. CVPR 2017.
[20] D. Ferstl et al. Image guided depth upsampling using anisotropic total generalized variation. ICCV 2013.
[64] N. Silberman, D. Hoiem, P. Kohli, and R. Fergus. Indoor segmentation and support inference from rgbd images. ECCV 2012.

Deep Depth Estimation: Results

Comparison to other depth estimation methods:

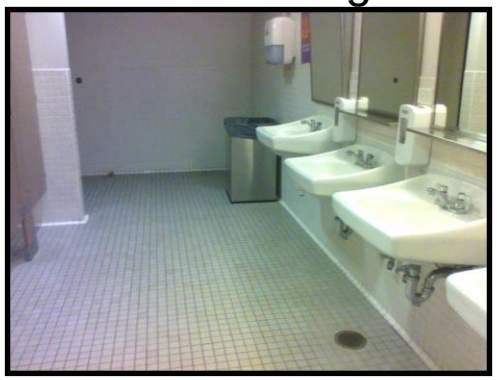
Obs	Meth	Rel↓	RMSE↓	1.05↑	1.10↑	1.25↑	1.25^2↑	1.25^3↑
Y	Laina [37]	0.190	0.374	17.90	31.03	54.80	75.97	85.69
	Chakr. [7]	0.161	0.320	21.52	35.5	58.75	77.48	85.65
	Ours	**0.130**	**0.274**	**30.60**	**43.65**	**61.14**	**75.69**	**82.65**
N	Laina [37]	0.384	0.572	8.86	16.67	34.64	55.60	69.21
	Chakr. [7]	0.352	0.610	11.16	20.50	37.73	57.77	70.10
	Ours	**0.283**	**0.537**	**17.27**	**27.42**	**44.19**	**61.80**	**70.90**

[7] Chakrabarti, A. et al., Depth from a single image by harmonizing overcomplete local network predictions. NIPS 2016.
[37] Laina, C. et al., Deeper depth prediction with fully convolutional residual networks. 3DV 2016.

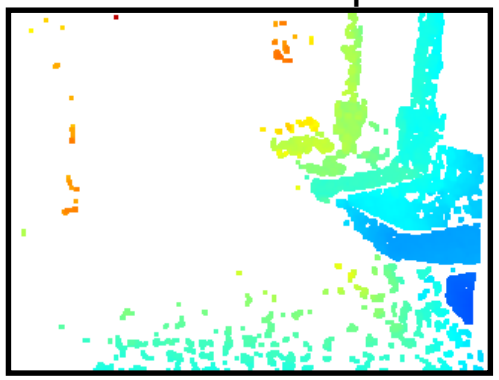
Deep Depth Completion: Results

Intel RealSense R200 examples:

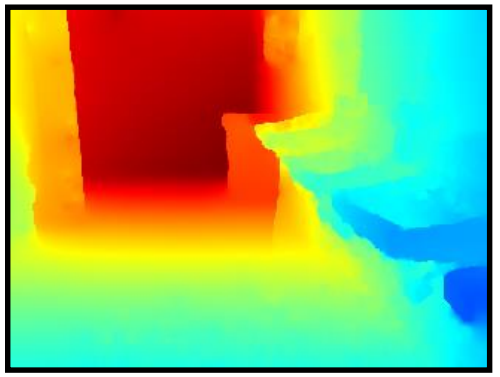
Color Image



Sensor Depth



Completed Depth



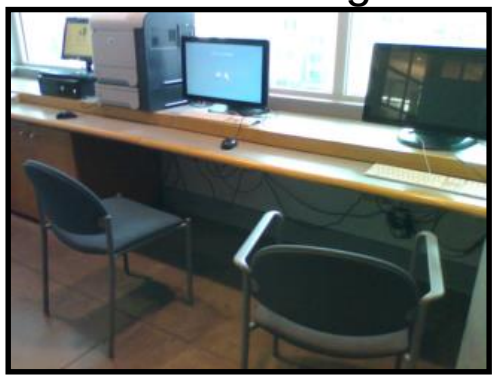
Sensor Point Cloud

Completed Point Cloud

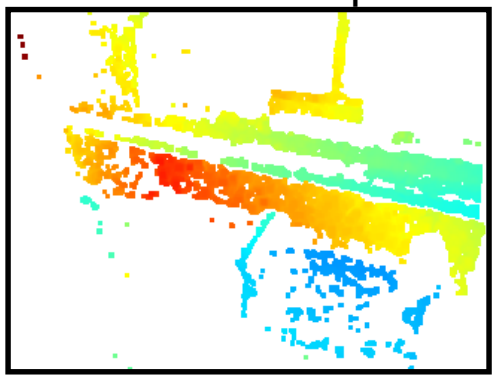
Deep Depth Completion: Results

Intel RealSense R200 examples:

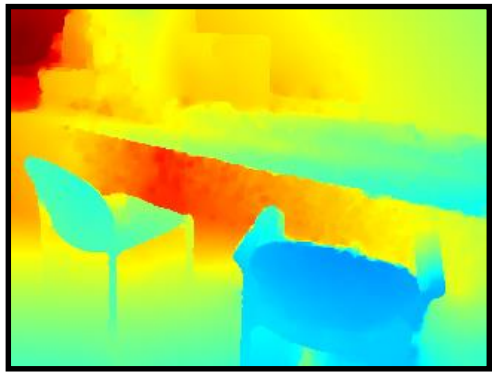
Color Image



Sensor Depth



Completed Depth



Sensor Point Cloud

Completed Point Cloud

Talk Outline (Part 2)

Introduction

Three recent projects

- Deep depth completion [CVPR 2018]
- Semantic scene completion [CVPR 2017]
- Semantic view extrapolation [CVPR 2018]

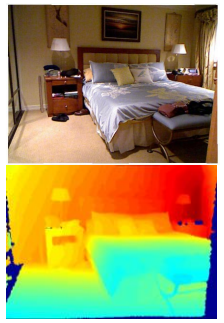
Common themes

Future work

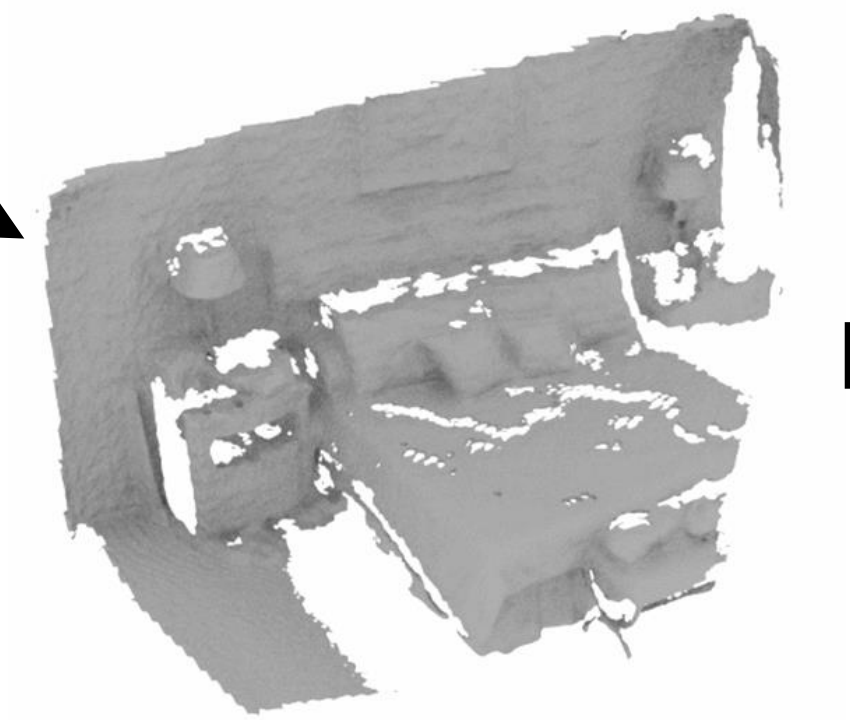
Shuran Song, Fisher Yu, Andy Zeng,
Angel Chang, Manolis Savva, and Thomas Funkhouser,
"Semantic Scene Completion from a Single Depth Image,"
CVPR 2017 (oral)

Semantic Scene Completion

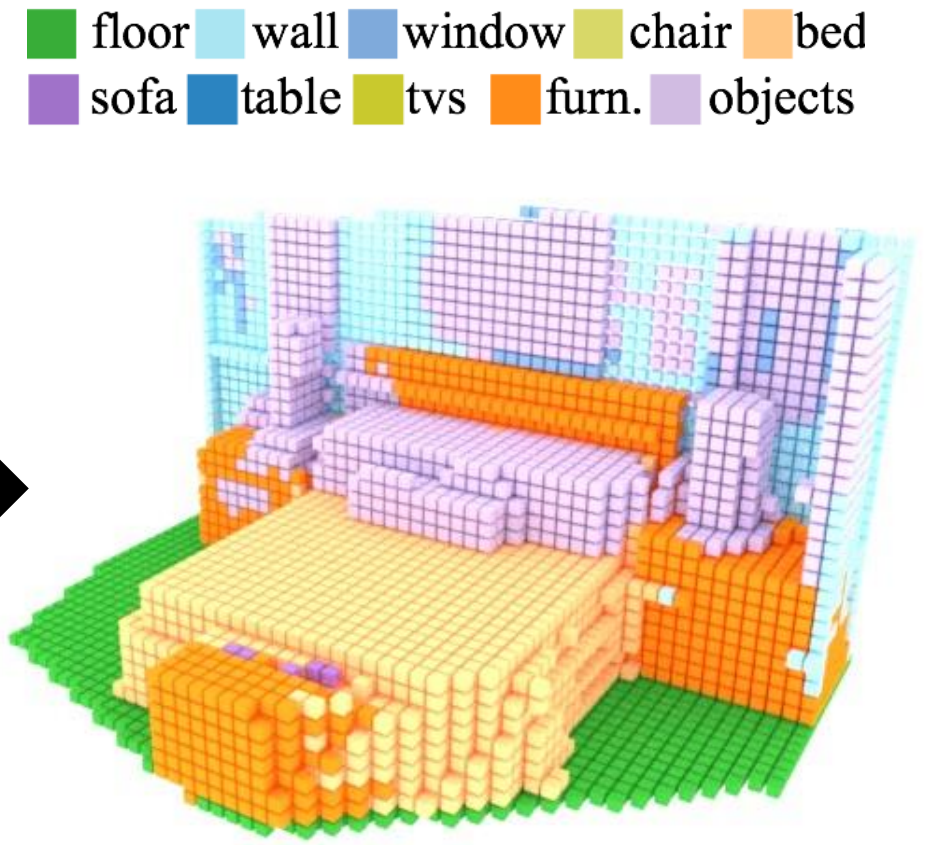
Goal: estimate the semantics and geometry occluded from a depth camera



RGB-D Image



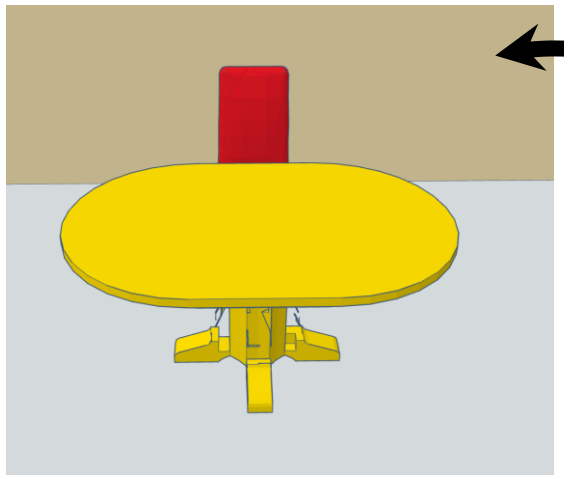
Input: Single view depth map



Output: Semantic scene completion

Semantic Scene Completion

Formulation: given a depth image, label all voxels by semantic class



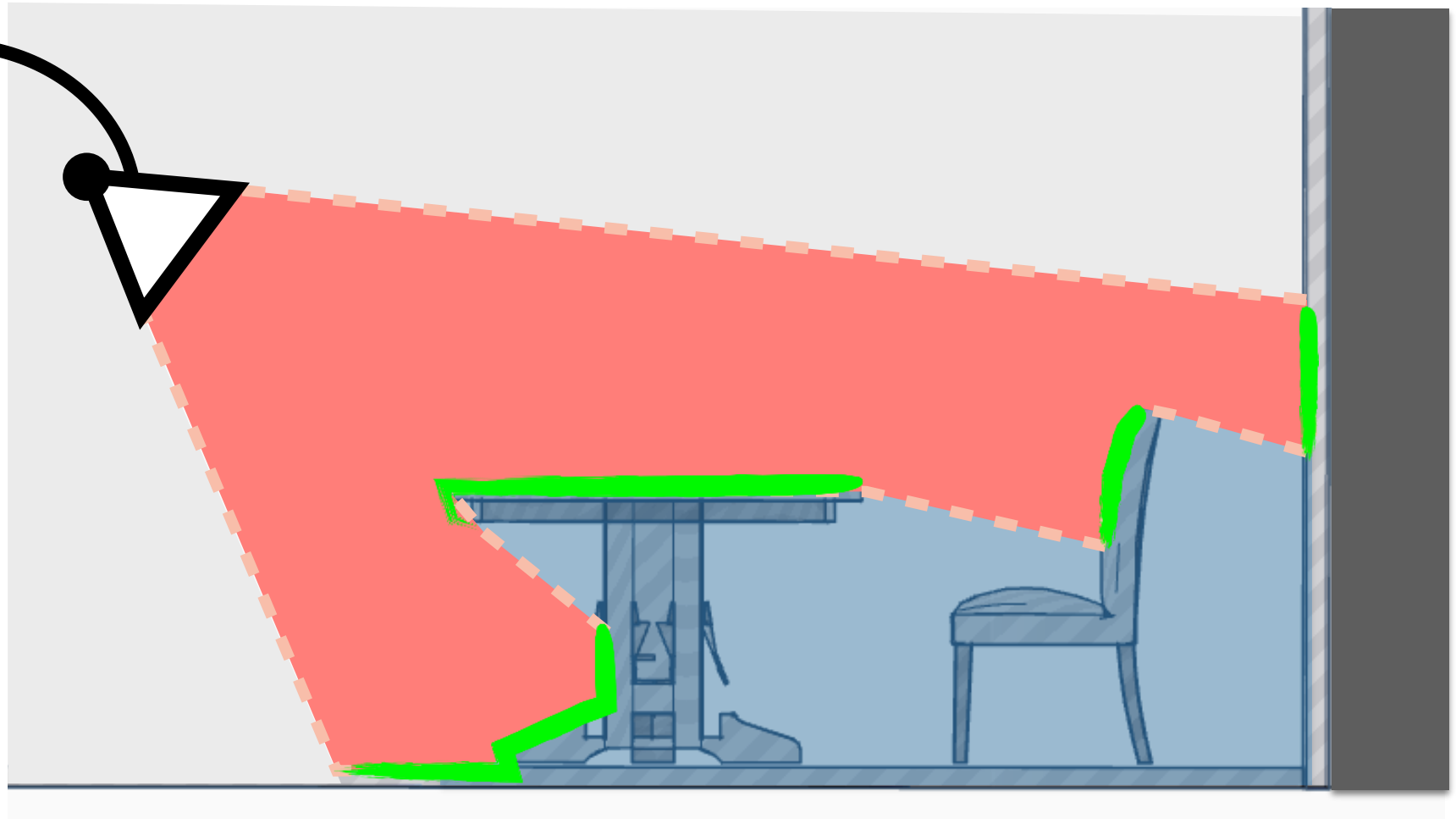
visible surface

free space

occluded space

outside view

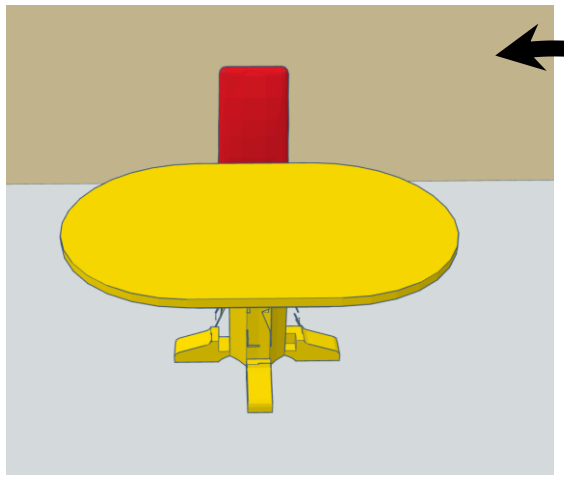
outside room



3D Scene

Semantic Scene Completion

Formulation: given a depth image, label all voxels by semantic class



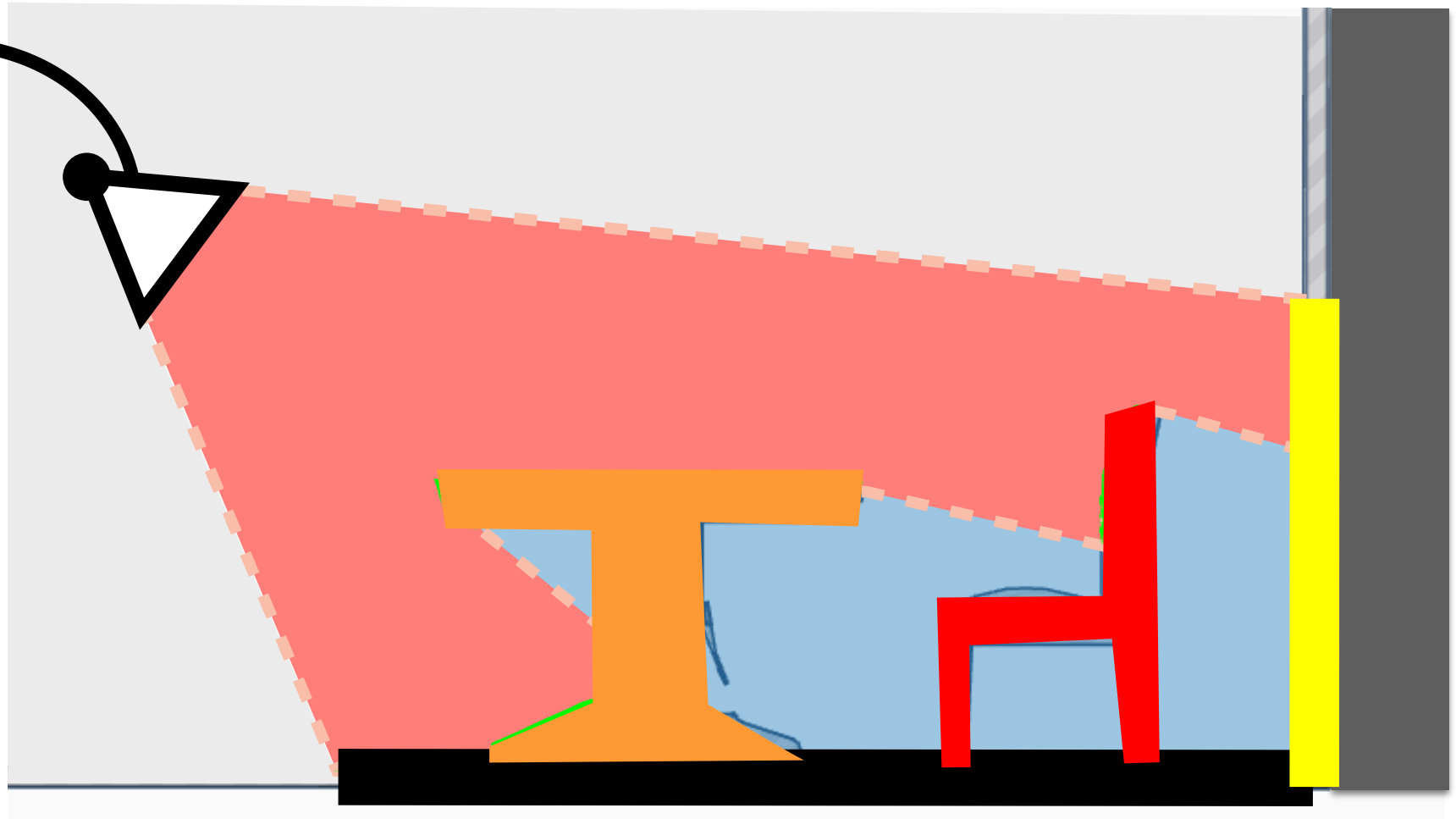
visible surface

free space

occluded space

outside view

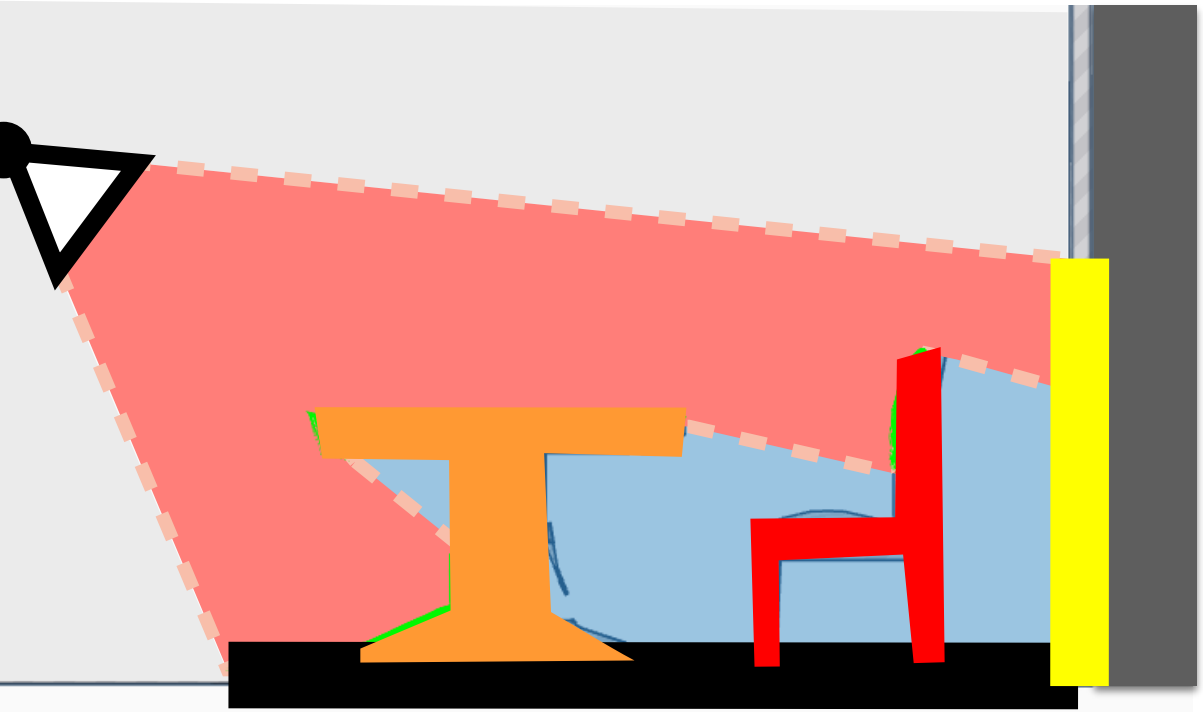
outside room



3D Scene

Semantic Scene Completion

Prior work: segmentation **OR** completion

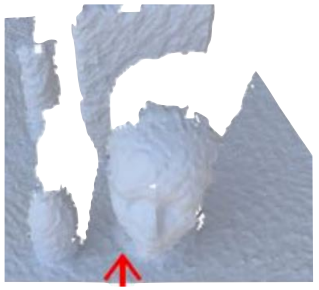


3D Scene

surface segmentation

Silberman *et al.*

scene completion



Firman *et al.*

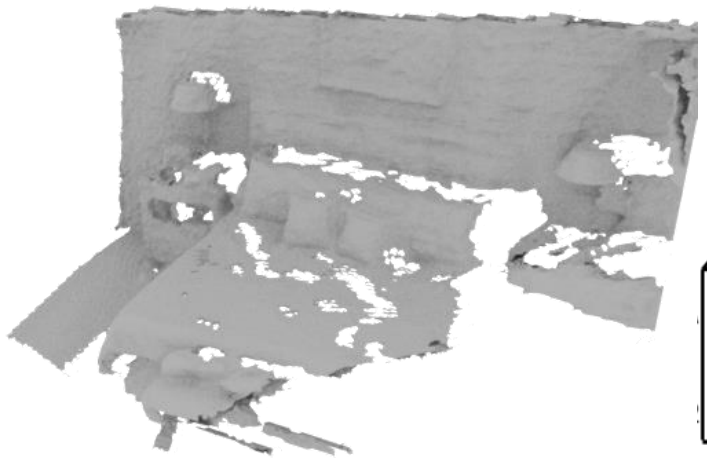
The occupancy and the object identity are tightly intertwined !

semantic scene completion

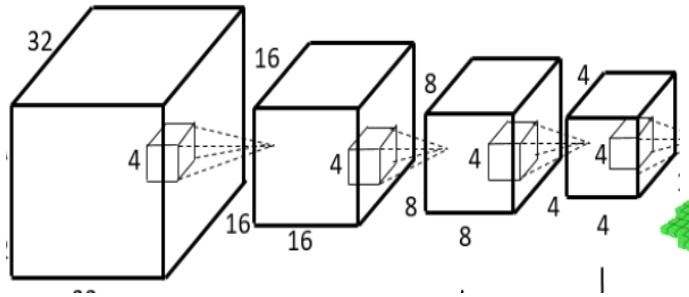
This paper

Semantic Scene Completion

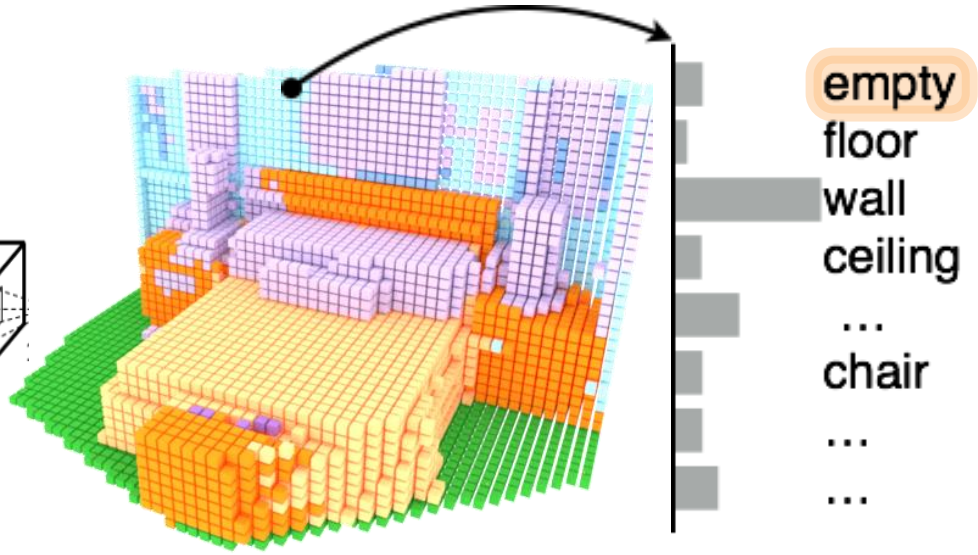
Approach: end-to-end 3D deep network



SSCNet



Prediction: N+1 classes

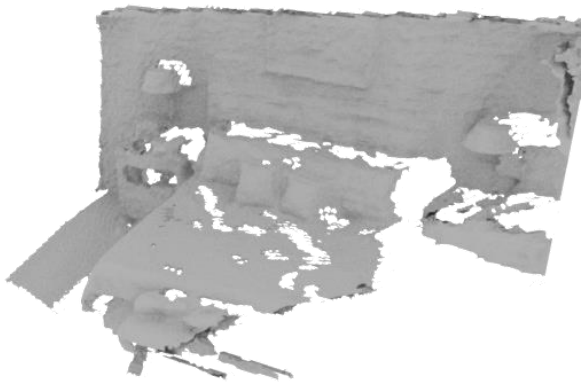


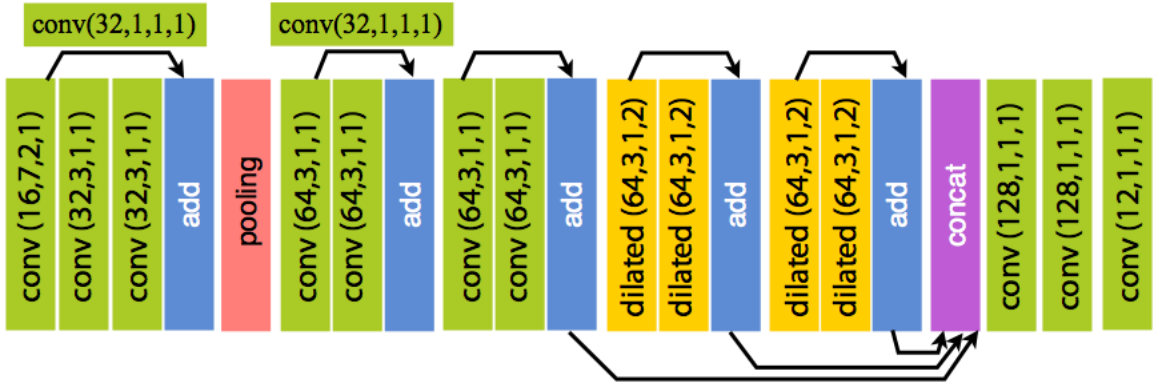
Input:
Single view depth map

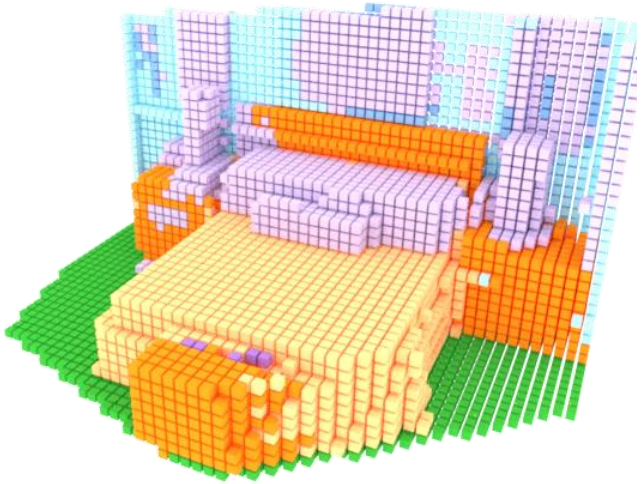
Output:
Volumetric occupancy + semantics

Simultaneously predict voxel occupancy and semantics classes by a single forward pass.

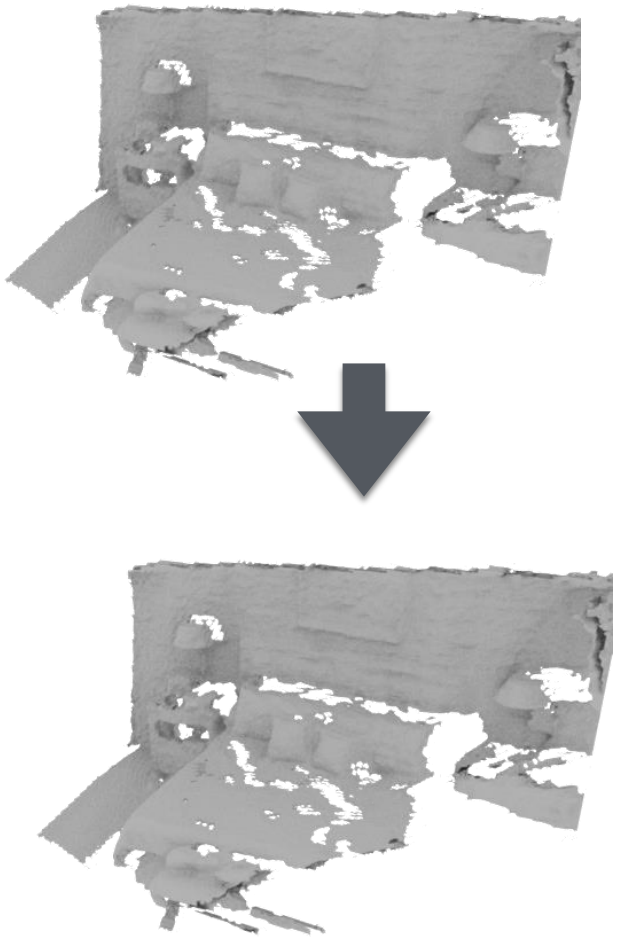
Semantic Scene Completion: Network Architecture

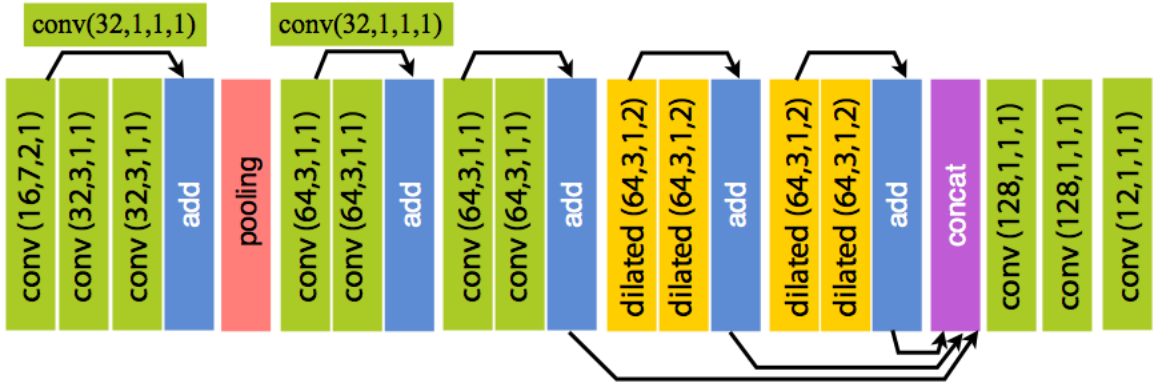


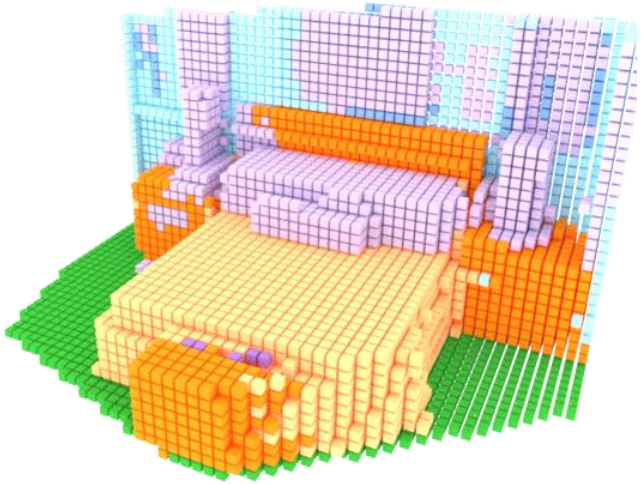




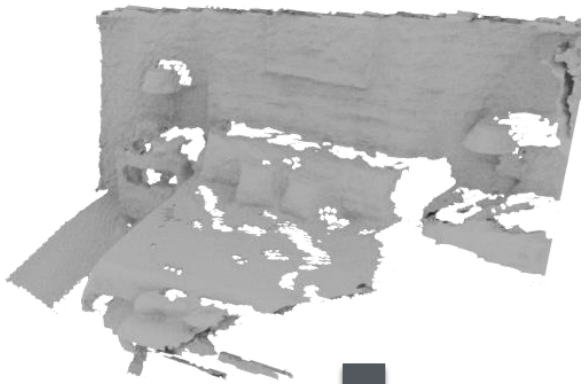
Semantic Scene Completion: Network Architecture

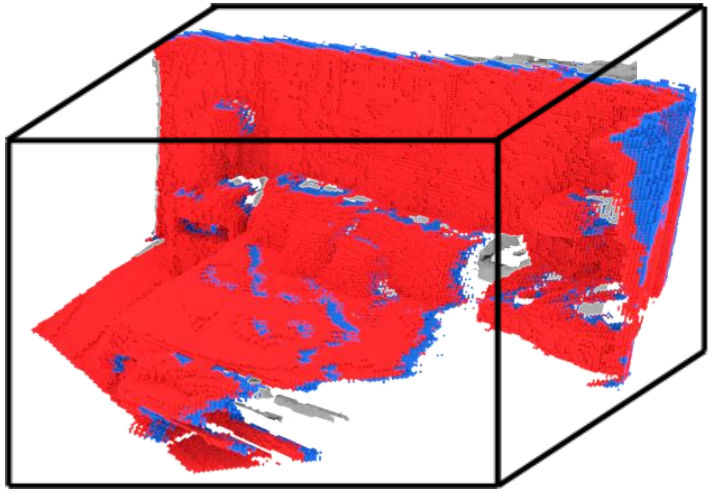




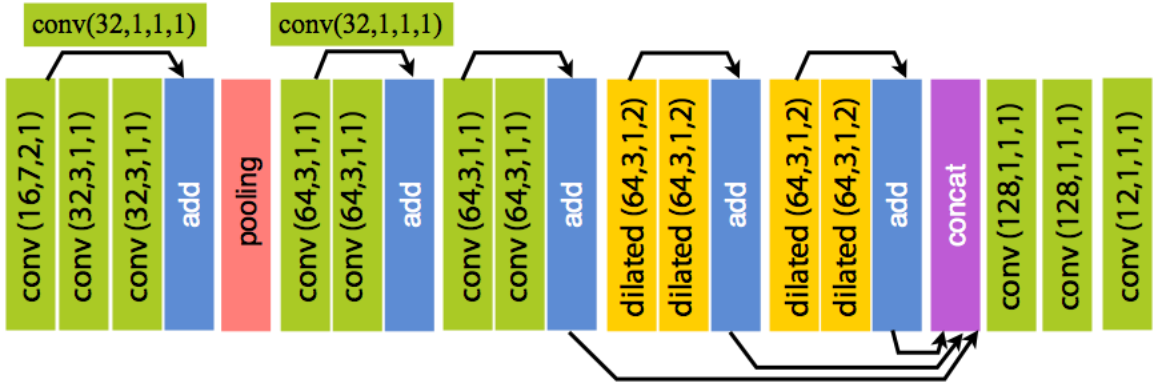


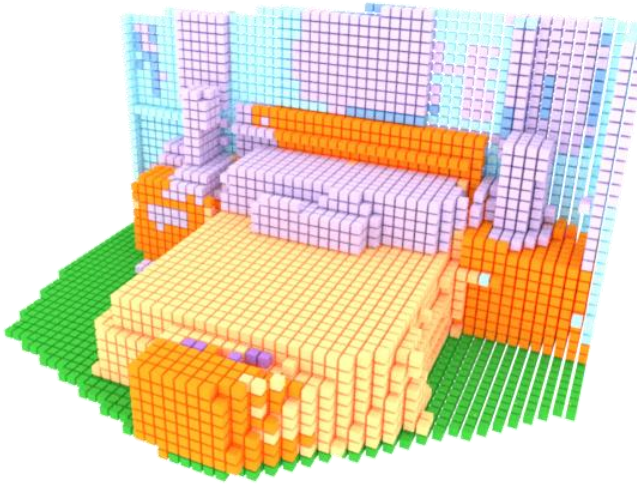
Semantic Scene Completion: Network Architecture



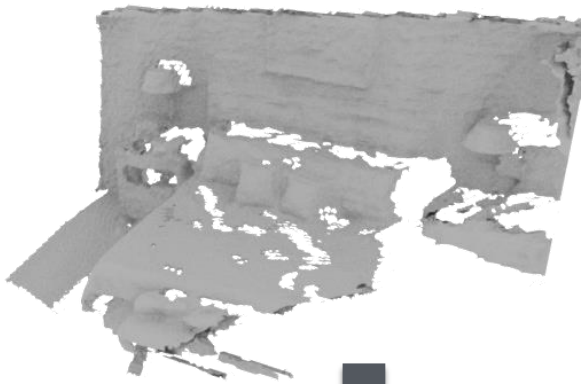


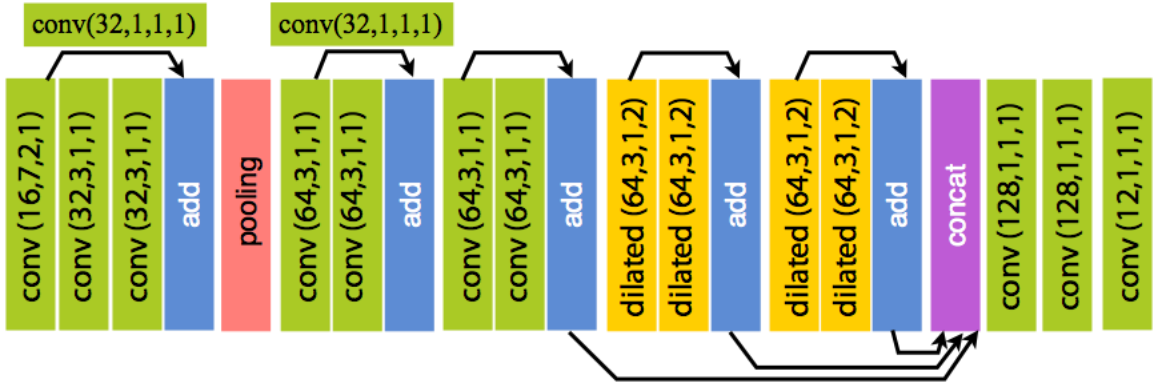
Voxel size: 0.02 m

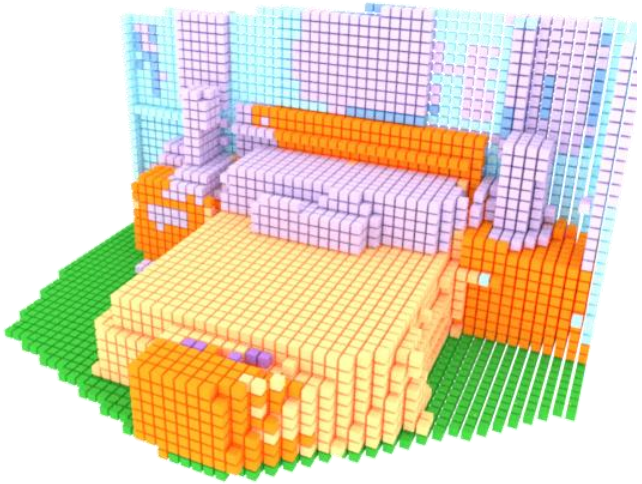


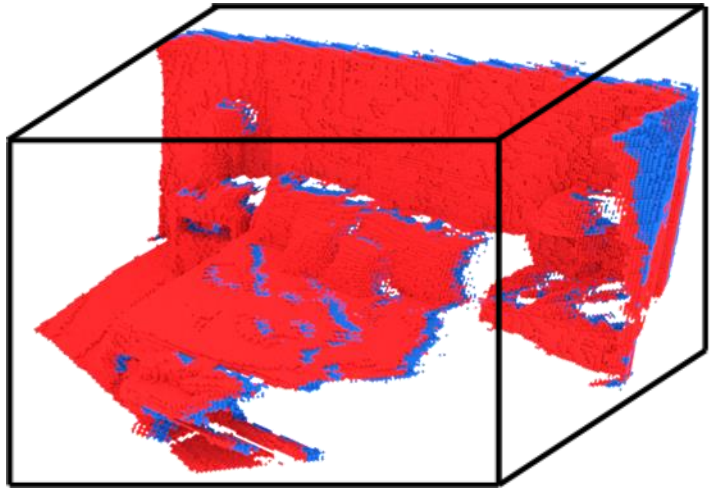


Semantic Scene Completion: Network Architecture

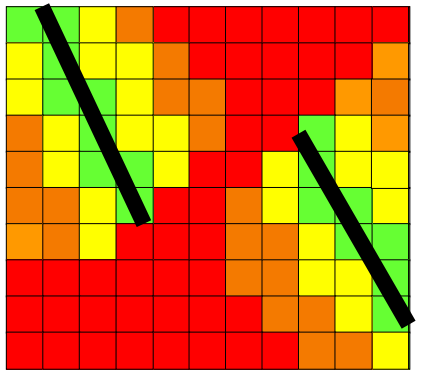








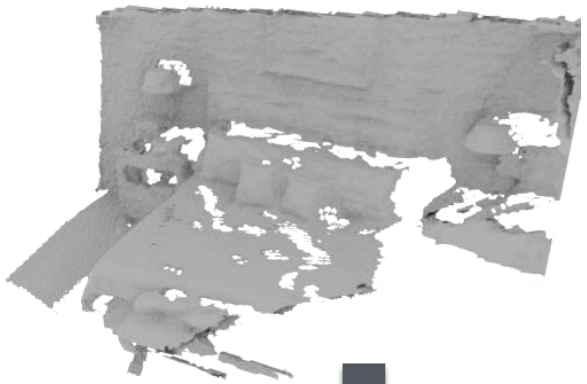
Voxel size: 0.02 m

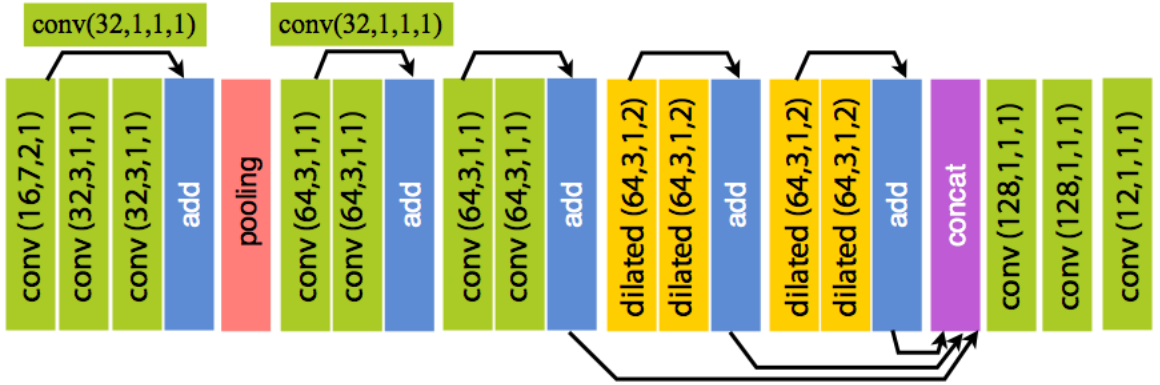


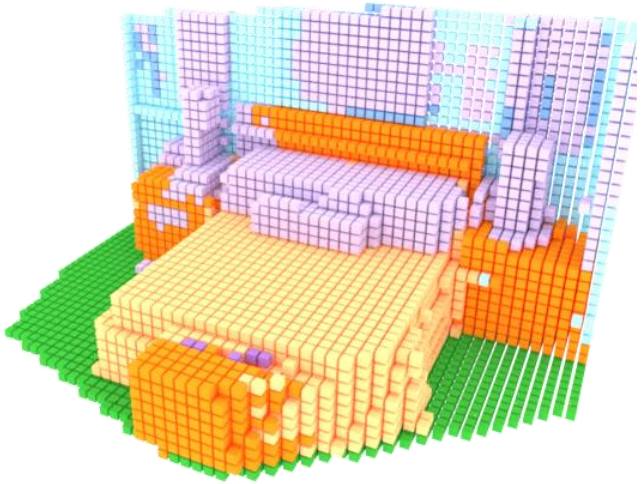
View

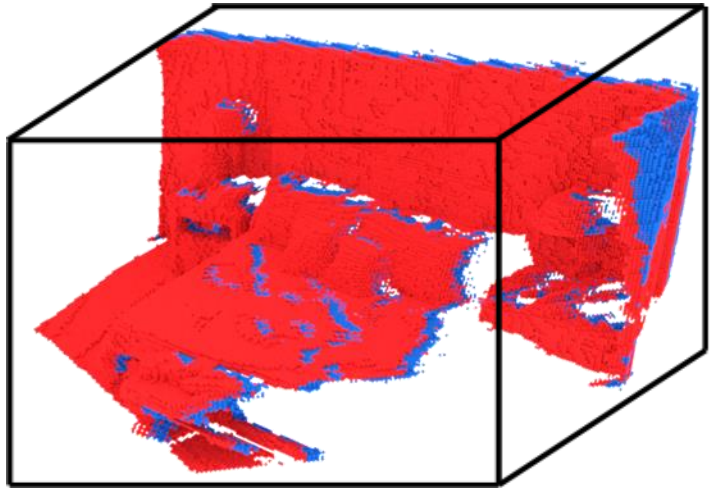
Standard TSDF

Semantic Scene Completion: Network Architecture

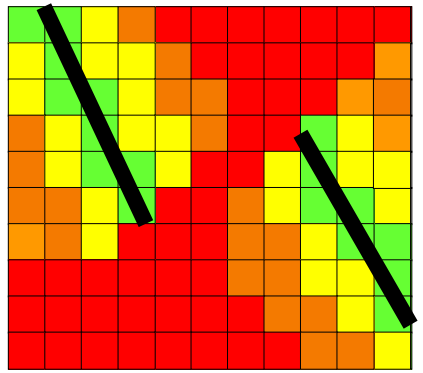






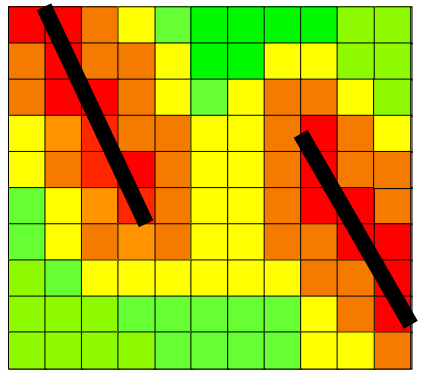


Voxel size: 0.02 m



View

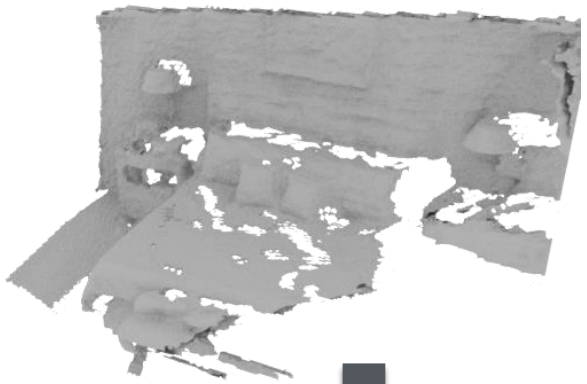
Standard TSDF

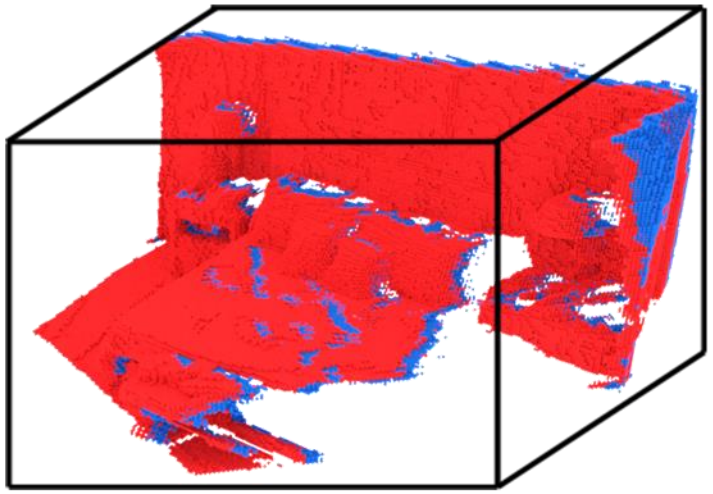


Flipped TSDF

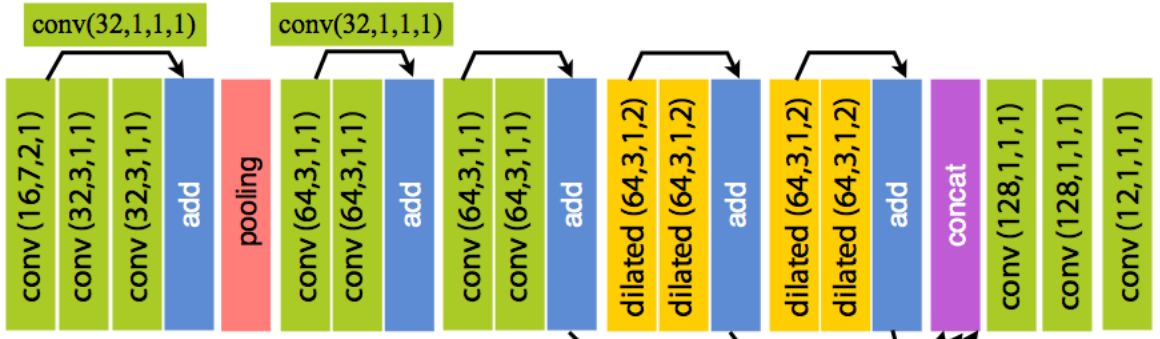
Encode 3D space using flipped TSDF

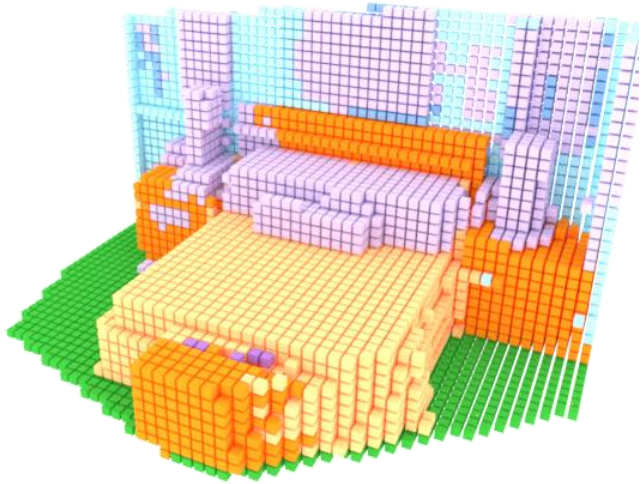
Semantic Scene Completion: Network Architecture

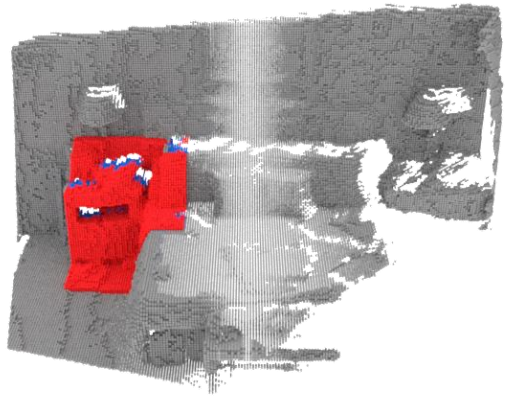




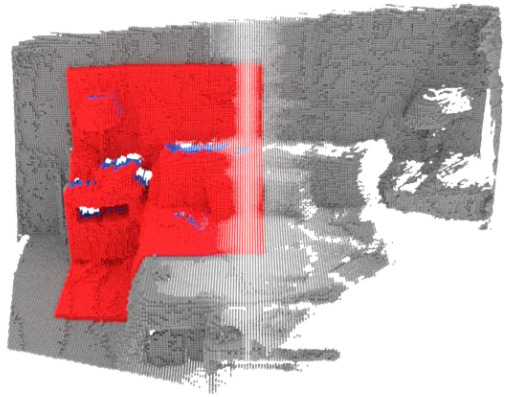
Voxel size: 0.02 m



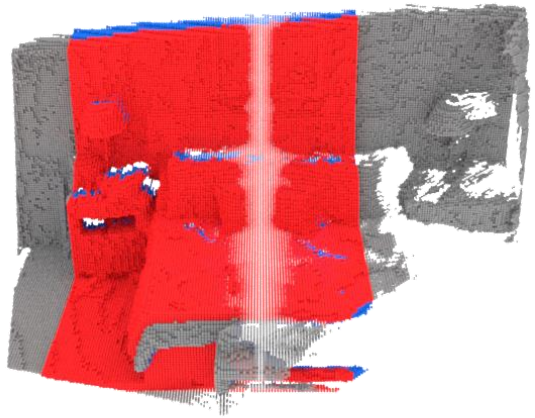




Receptive field: 0.98 m



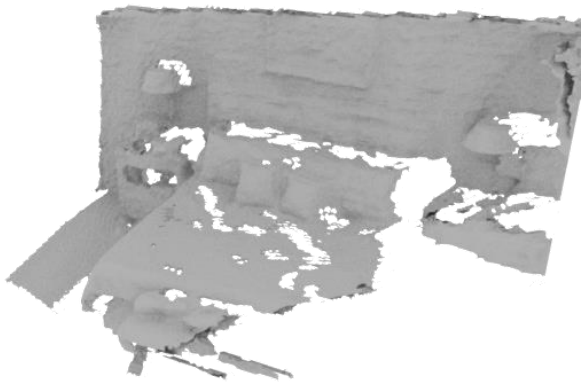
Receptive field:1.62 m

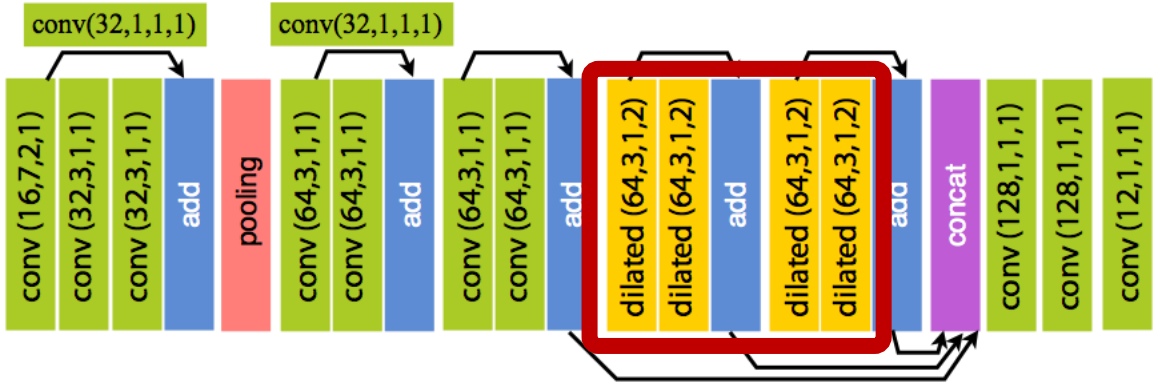


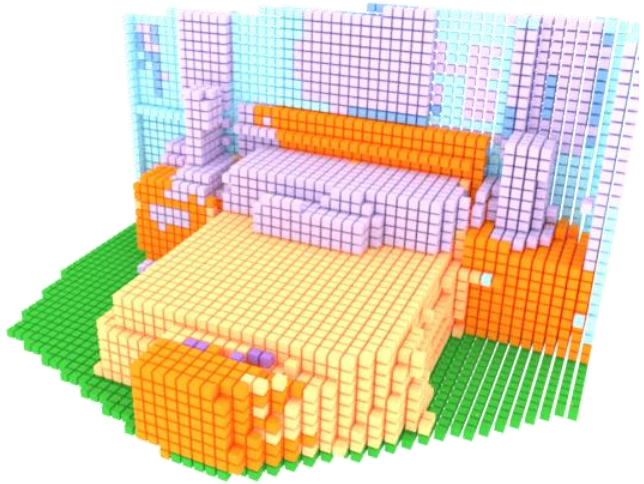
Receptive field: 2.26 m

Extract features for different physical scales

Semantic Scene Completion: Network Architecture





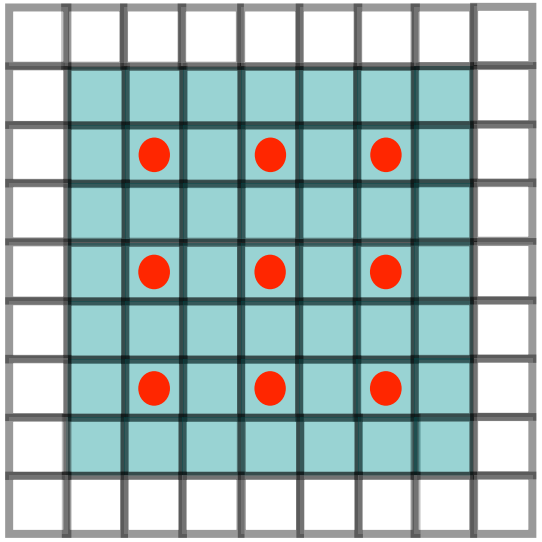


receptive field

learnable parameter

Receptive Field = 7x7x7

Parameters = 27



Dilated Convolutions

Larger receptive field with **same** number of parameters and **same** output resolution!

F. Yu et al., Multi-Scale Context Aggregation by Dilated Convolutions, ICLR 2016

Semantic Scene Completion: Data

Where get training data?

NYUv2

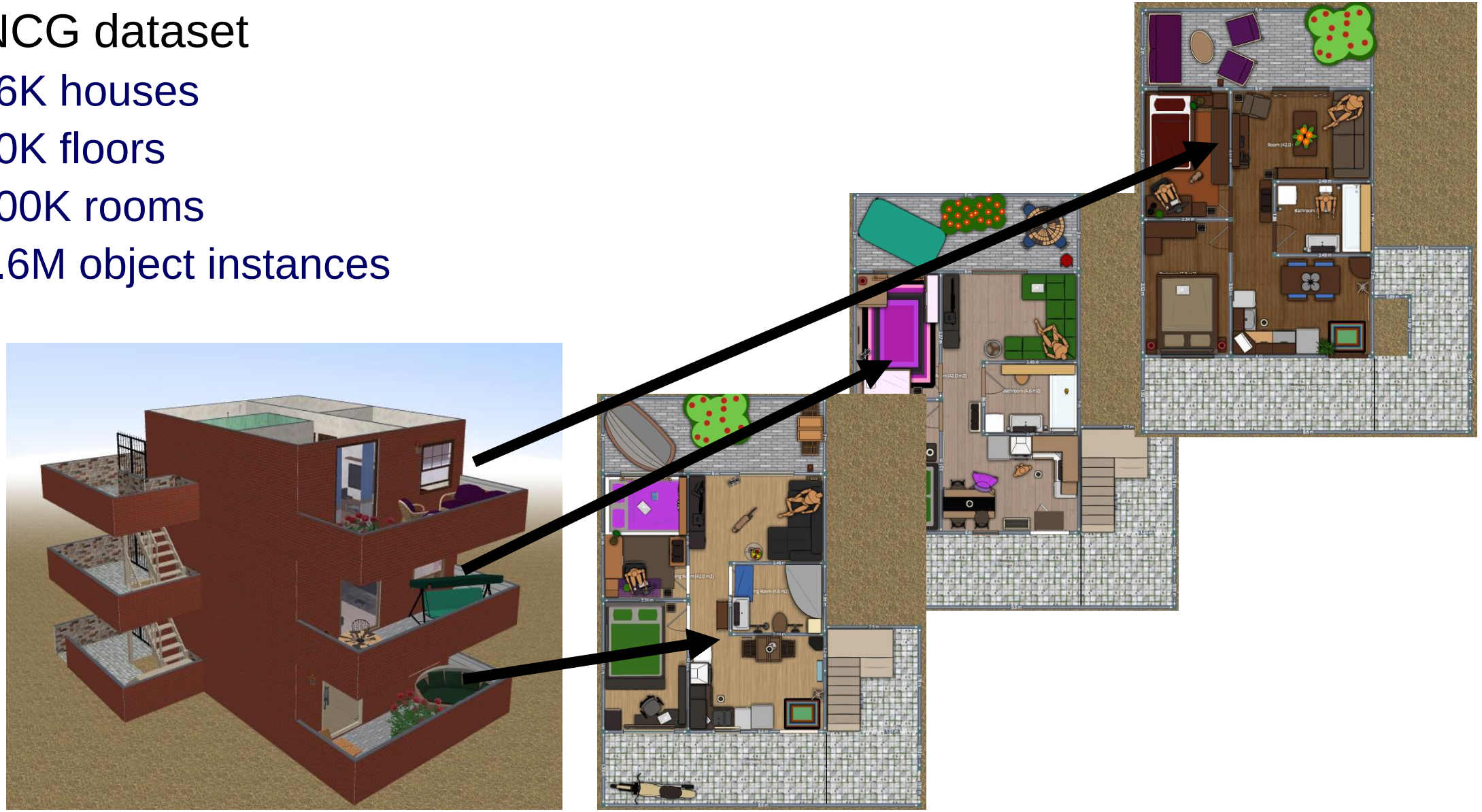
Small number of objects labeled with CAD models
(suitable for testing, not training)

N. Silberman, P. Kohli, D. Hoiem, R. Fergus, Indoor Segmentation and Support Inference from RGBD Images, ECCV 2012
R. Guo, C. Zou, D. Hoiem, Predicting Complete 3D models of Indoor Scenes, arXiv 2015

Semantic Scene Completion: Data

SUNCG dataset

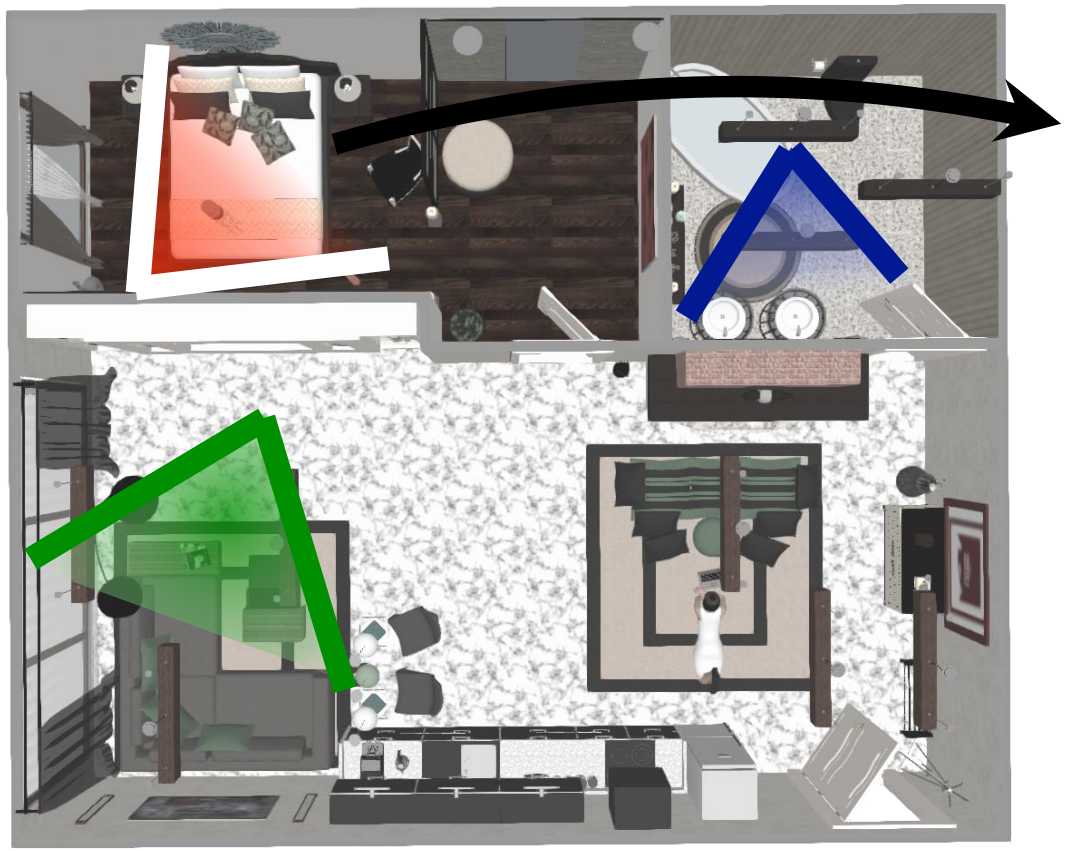
- 46K houses
- 50K floors
- 400K rooms
- 5.6M object instances



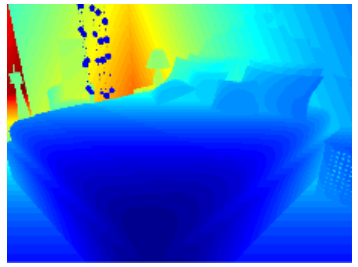
Semantic Scene Completion: Data

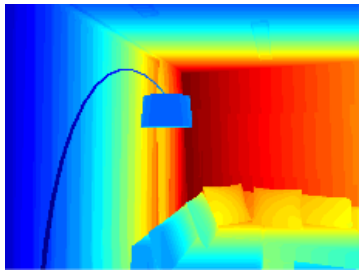
SUNCG dataset

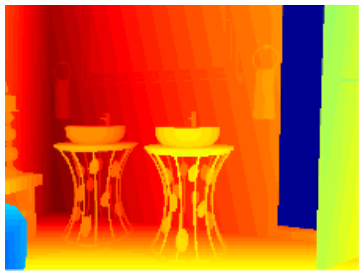
synthetic camera views



depth







ground truth semantic scene completion

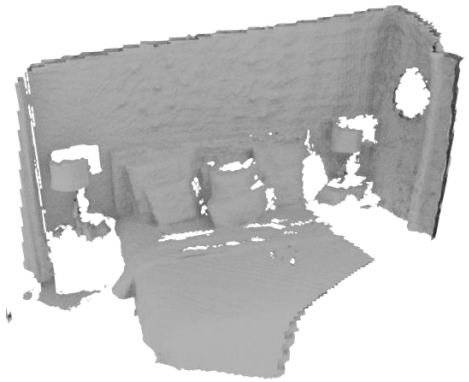
Semantic Scene Completion: Experiments

Pre-train on SUNCG

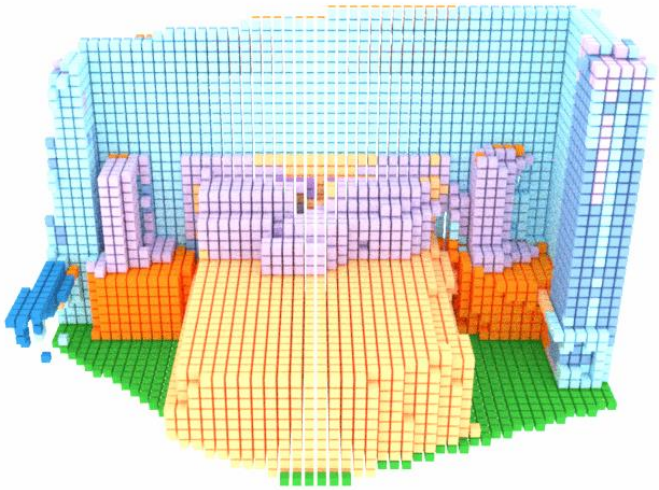
Fine-tune and test on NYUv2

Semantic Scene Completion: Results

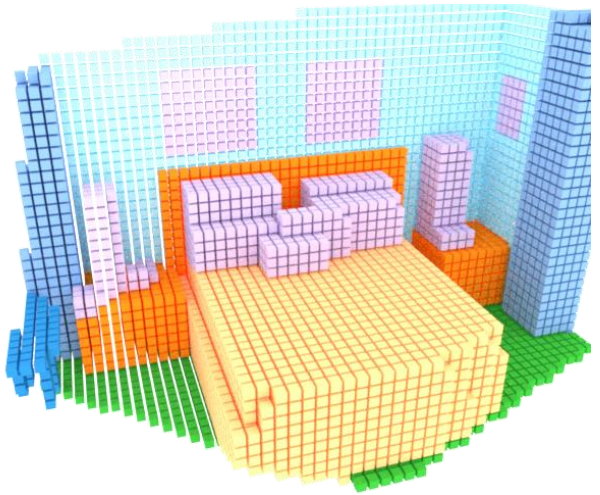
Input Color



Input Depth



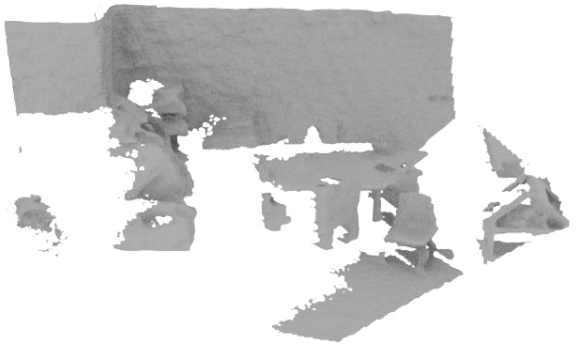
Our Result



Ground Truth

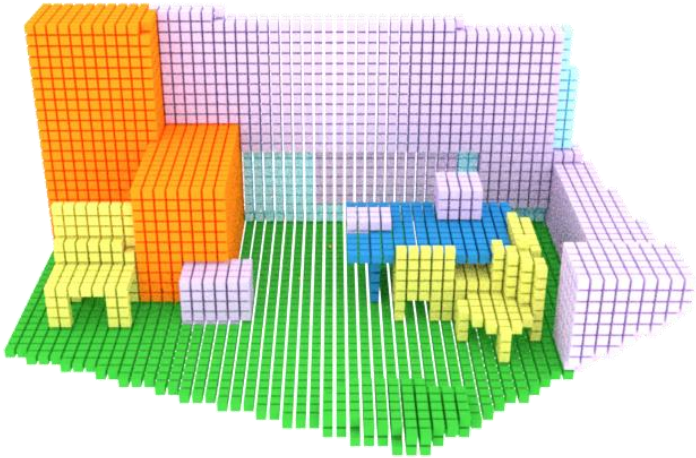
Semantic Scene Completion: Results

Input Color



Input Depth

Our Result



Ground Truth

Semantic Scene Completion: Results

Result 1: better than previous volumetric completion algorithms

method	training	prec.	recall	IoU
Zheng *et al.* [36]	NYU	60.1	46.7	34.6
Firman *et al.* [3]	NYU	66.5	69.7	50.8
SSCNet completion	NYU	66.3	96.9	64.8
SSCNet joint	NYU	75.0	92.3	70.3
SSCNet joint	SUNCG+NYU	**75.0**	**96.0**	**73.0**

Comparison to previous algorithms for volumetric completion

Semantic Scene Completion: Results

Result 2: better than previous semantic labeling algorithms

	scene completion			semantic scene completion											
method (train)	prec.	recall	IoU	ceil.	floor	wall	win.	chair	bed	sofa	table	tvs	furn.	objs.	avg.
Lin *et al.* (NYU) [17]	58.5	49.9	36.4	0	11.7	13.3	**14.1**	9.4	29	24	6.0	7.0	16.2	1.1	12.0
Geiger and Wang (NYU) [4]	65.7	58	44.4	10.2	62.5	19.1	5.8	8.5	40.6	27.7	7.0	6.0	22.6	5.9	19.6
SSCNet (NYU)	57.0	**94.5**	55.1	15.1	**94.7**	24.4	0	12.6	32.1	35	13	7.8	27.1	10.1	24.7
SSCNet (SUNCG)	55.6	91.9	53.2	5.8	81.8	19.6	5.4	12.9	34.4	26	13.6	6.1	9.4	7.4	20.2
SSCNet (SUNCG+NYU)	**59.3**	92.9	**56.6**	**15.1**	94.6	**24.7**	10.8	**17.3**	**53.2**	**45.9**	**15.9**	**13.9**	**31.1**	**12.6**	**30.5**

Comparison to previous algorithms for semantic labeling with 3D model fitting

Talk Outline (Part 3)

Introduction

Three recent projects

- Deep depth completion [CVPR 2018]
- Semantic scene completion [CVPR 2017]
- Semantic view extrapolation [CVPR 2018]

Common themes

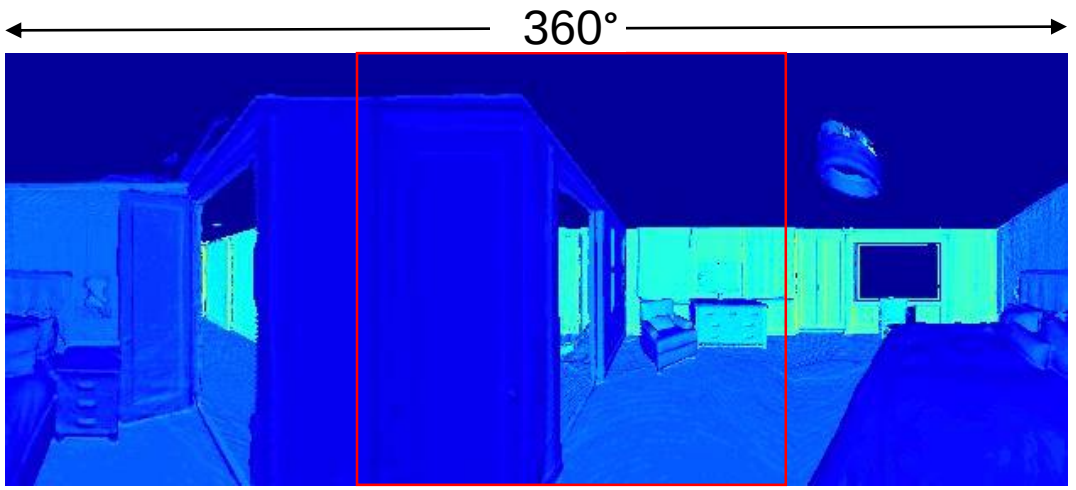
Future work

Shuran Song, Andy Zeng, Angel X. Chang,
Manolis Savva, Silvio Savarese, and Thomas Funkhouser,
"Im2Pano3D: Extrapolating 360 Structure and Semantics
Beyond the Field of View,"
CVPR 2018 (oral)

Semantic View Extrapolation

Goal: given an RGB-D image, predict 3D structure and semantics outside view

Input: RGB-D Image

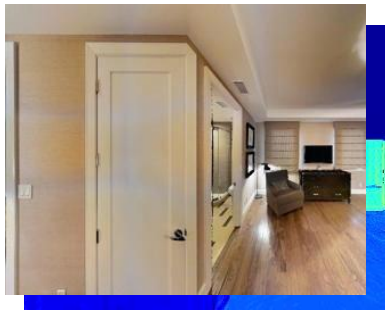


Output 1: 3D structure

Output 2: semantic segmentation

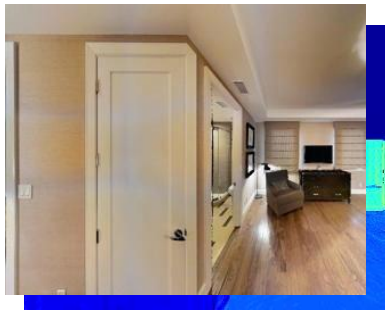
Semantic View Extrapolation

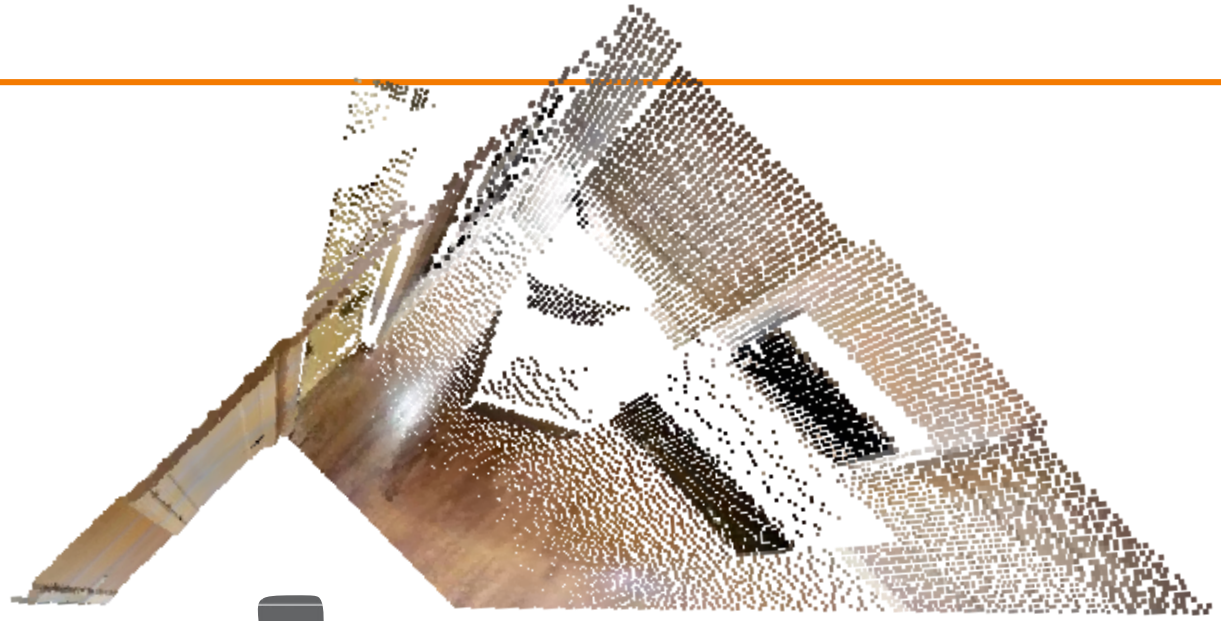
Input:
RGB-D Image



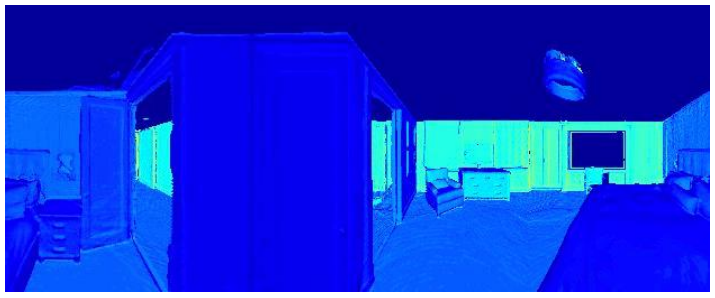
Semantic View Extrapolation

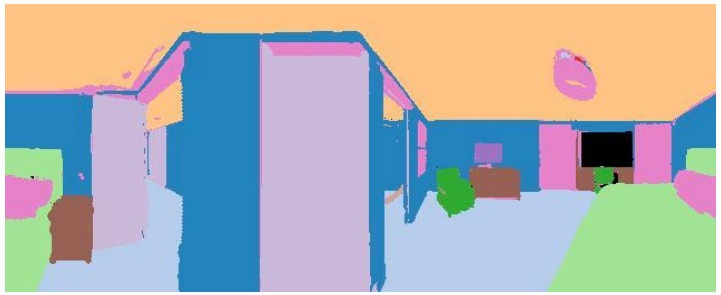
Input:
RGB-D Image



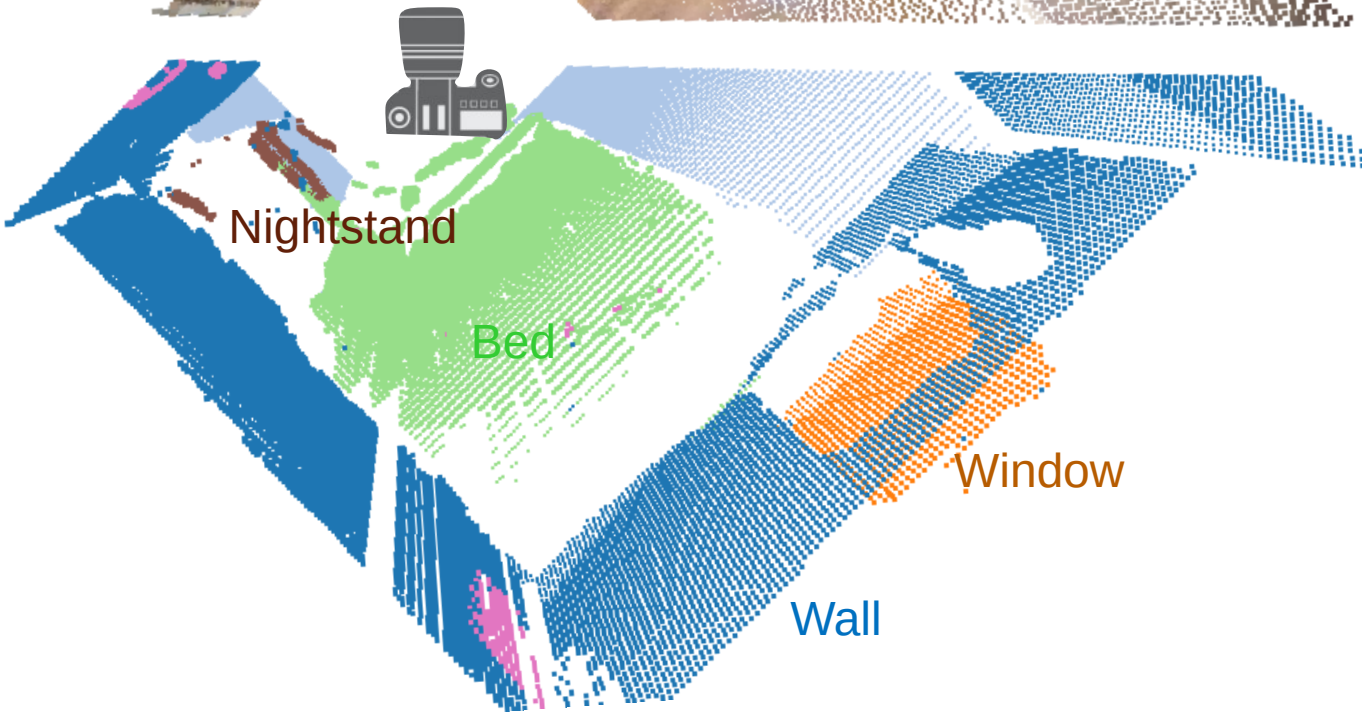


Output:
360° panorama
with 3D structure
& semantics





360°



Semantic View Extrapolation

Prior work: extrapolating appearance (color) outside field of view

Pathak et al. CVPR 2017

Semantic View Extrapolation

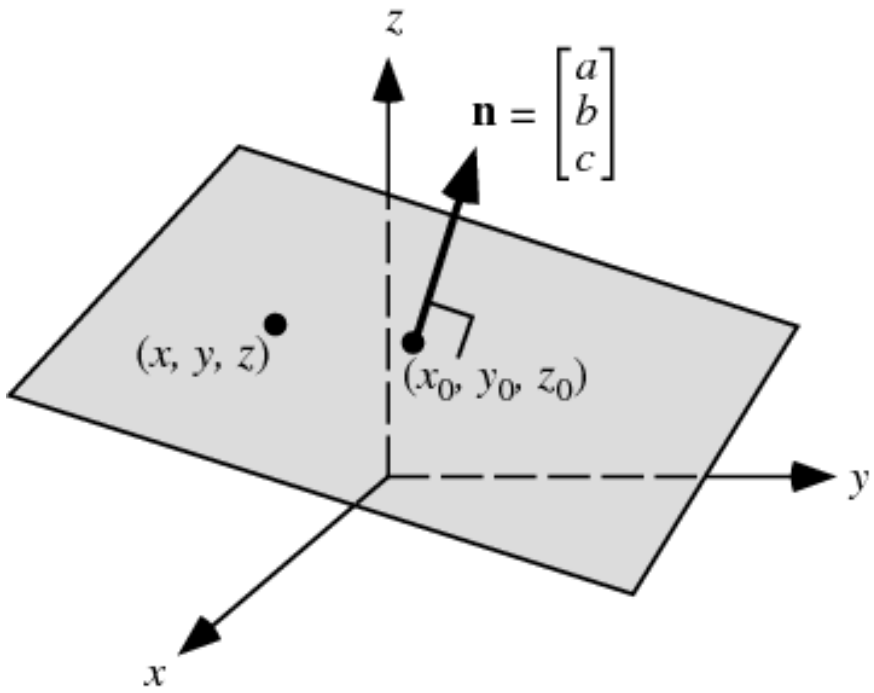
Our work: **predicting 3D structure and semantics** for full 360° panorama

3D structure

Semantic segmentation

Semantic View Extrapolation

3D structure representation: plane equation per pixel (normal and offset)

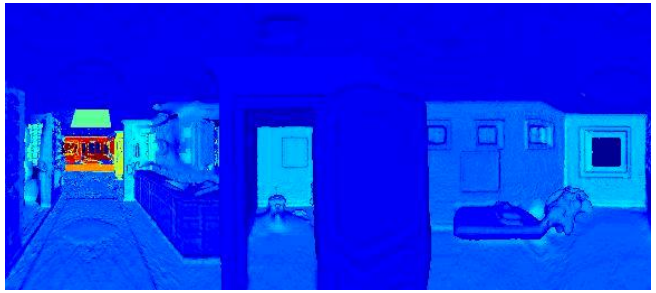


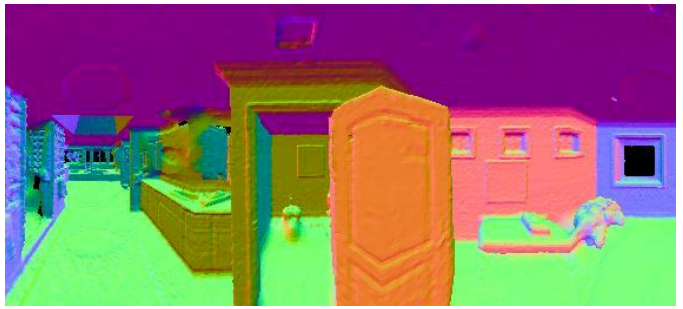
Plane Equation

$ax + by + cz - d = 0$

(a,b,c) = normal

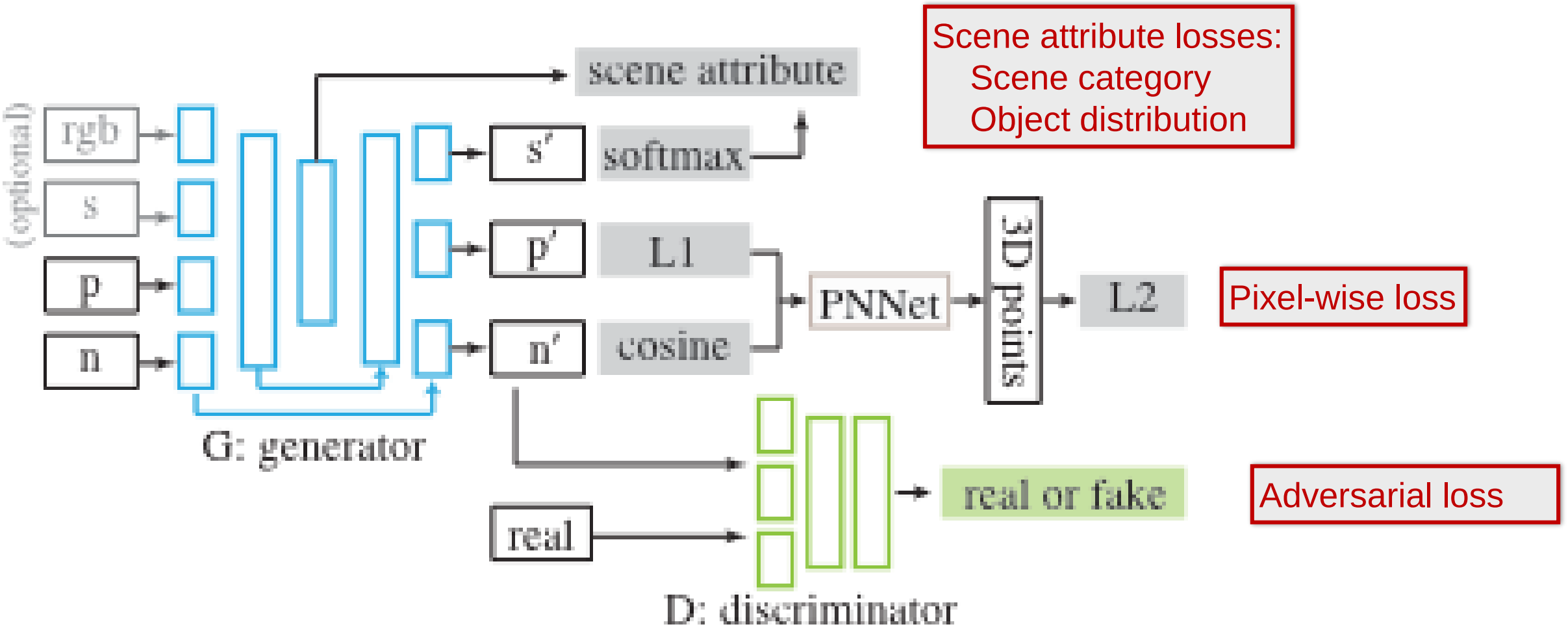
d = plane offset from origin





Similar to first project

Semantic View Extrapolation: Network Architecture

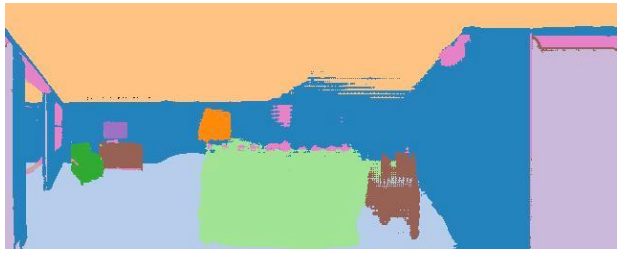


Semantic View Extrapolation: Training Objectives

Semantic View Extrapolation: Training Objectives

Every pixel is correct

$$L_{recon} = \sum_{i=0}^{n} -log(\frac{e^{f_{yi}}}{\sum_j e^{f_j}})$$



Prediction

Ground truth

- Hard for even humans to do.
- Lose the ability to generalize.

Semantic View Extrapolation: Training Objectives

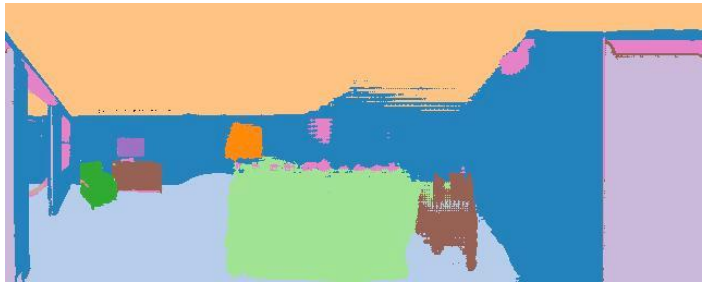
Every pixel is correct

Prediction is plausible

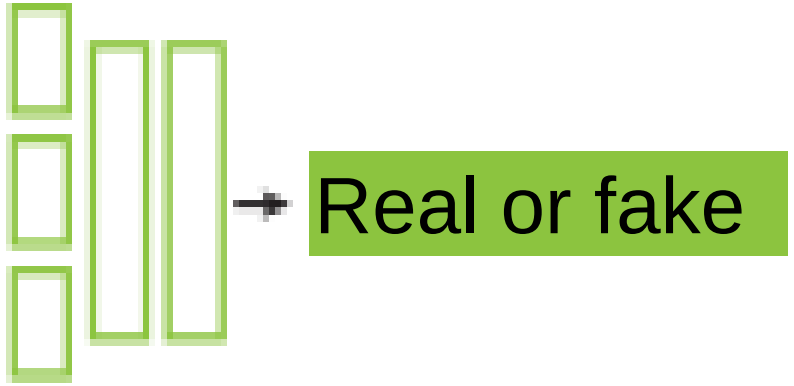
L_{recon}

$$\min_G \max_D \quad \mathbb{E}_{x \in \mathcal{X}}[\log(D(x))] + \mathbb{E}_{z \in \mathcal{Z}}[\log(1 - D(G(z)))] = L_{adv}$$

G:generator D: discriminator



Prediction



Adversarial loss

Goodfellow et al. 2014

Semantic View Extrapolation: Training Objectives

Every pixel is correct

L_{recon}

Similar scene attributes

$L_{attribute}$

Prediction is plausible

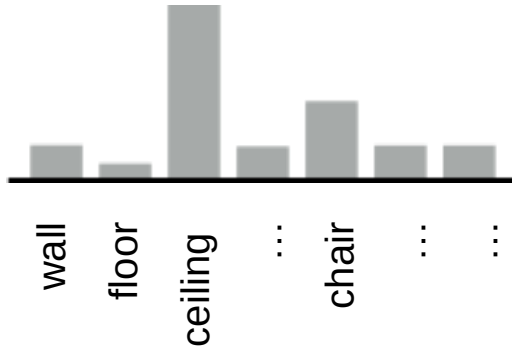
L_{adv}

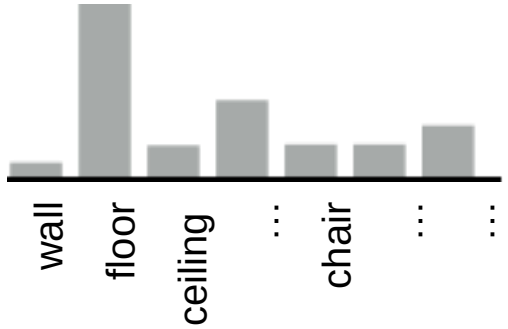
Scene Category

Object Distribution

Prediction

$h(x_c)$





Ground truth

y_c

$$L_{dis} = \sum_c |y_c - h(x_c)|$$

Semantic View Extrapolation: Training Objectives

Every pixel is correct

L_{recon}

Similar scene attribute

$L_{attribute}$

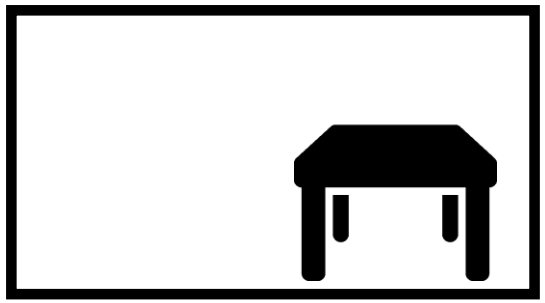
Scene Category

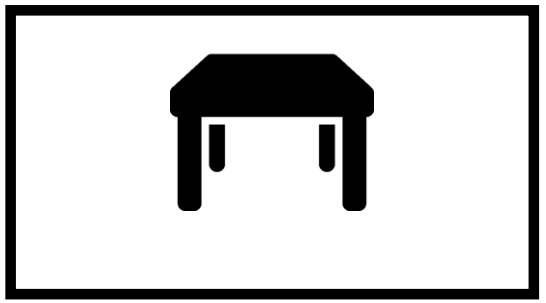
Object Distribution

Prediction is plausible

L_{adv}

Prediction





Ground truth

Semantic View Extrapolation: Training Objectives

Every pixel is correct

L_{recon}

Similar scene attribute

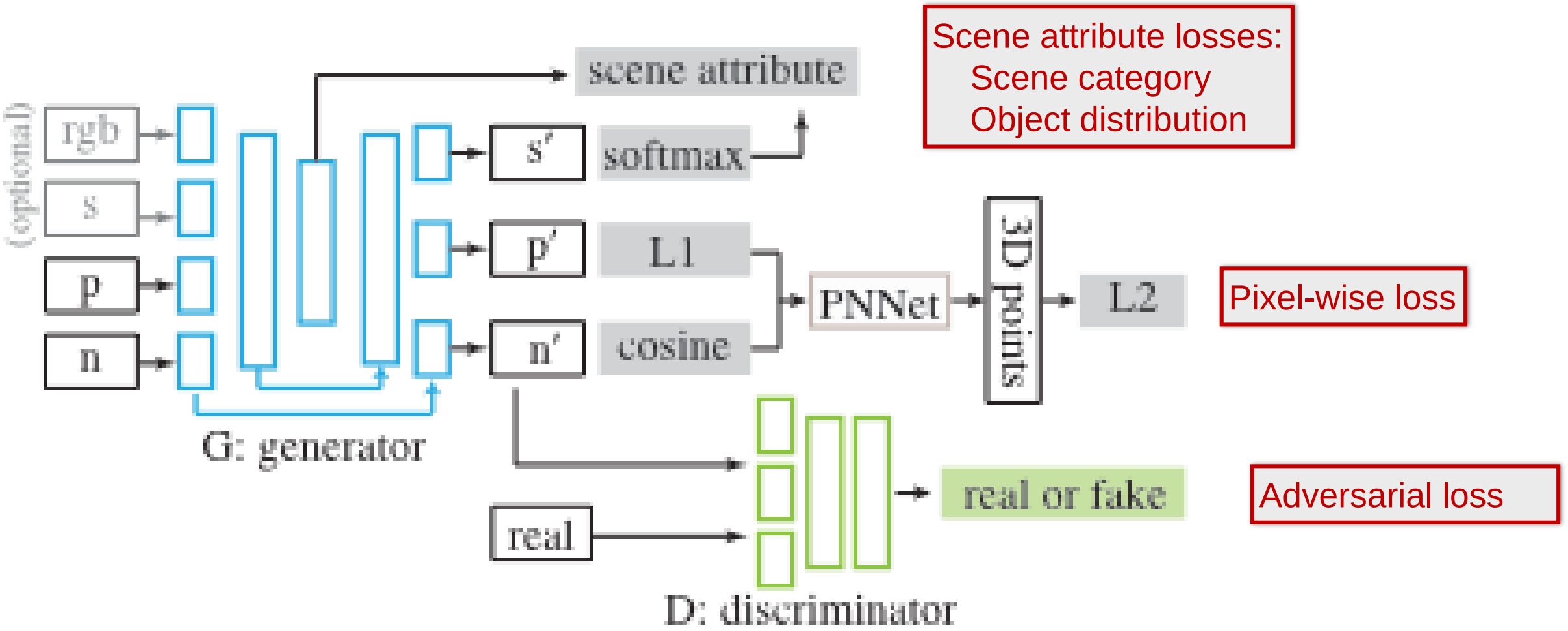
$L_{attribute}$

Prediction is plausible

L_{adv}

$$L = \lambda_1 L_{recon} + \lambda_2 L_{attribute} + \lambda_3 L_{adv}$$

Semantic View Extrapolation: Network Architecture



Semantic View Extrapolation: Data

Where get training/test data?

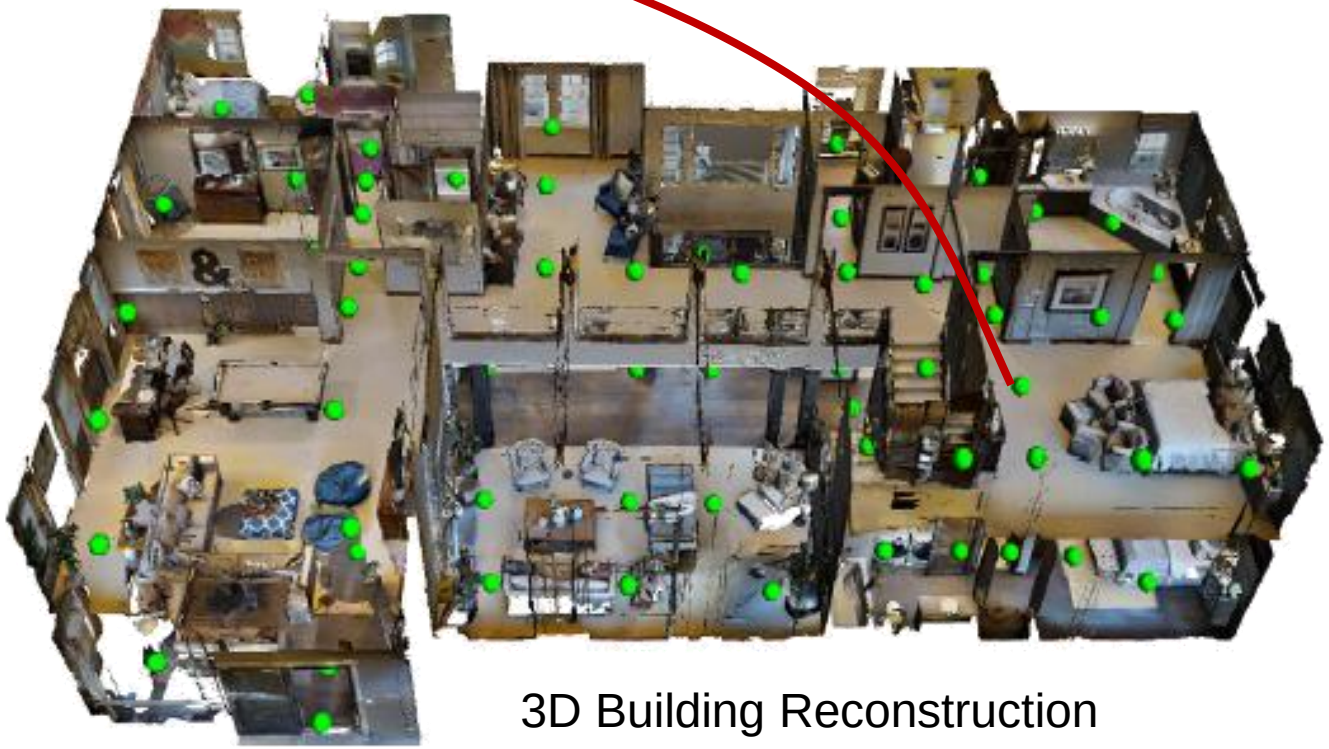
3D structure

Semantic segmentation

Semantic View Extrapolation: Data

Matterport3D dataset

Matterport Camera



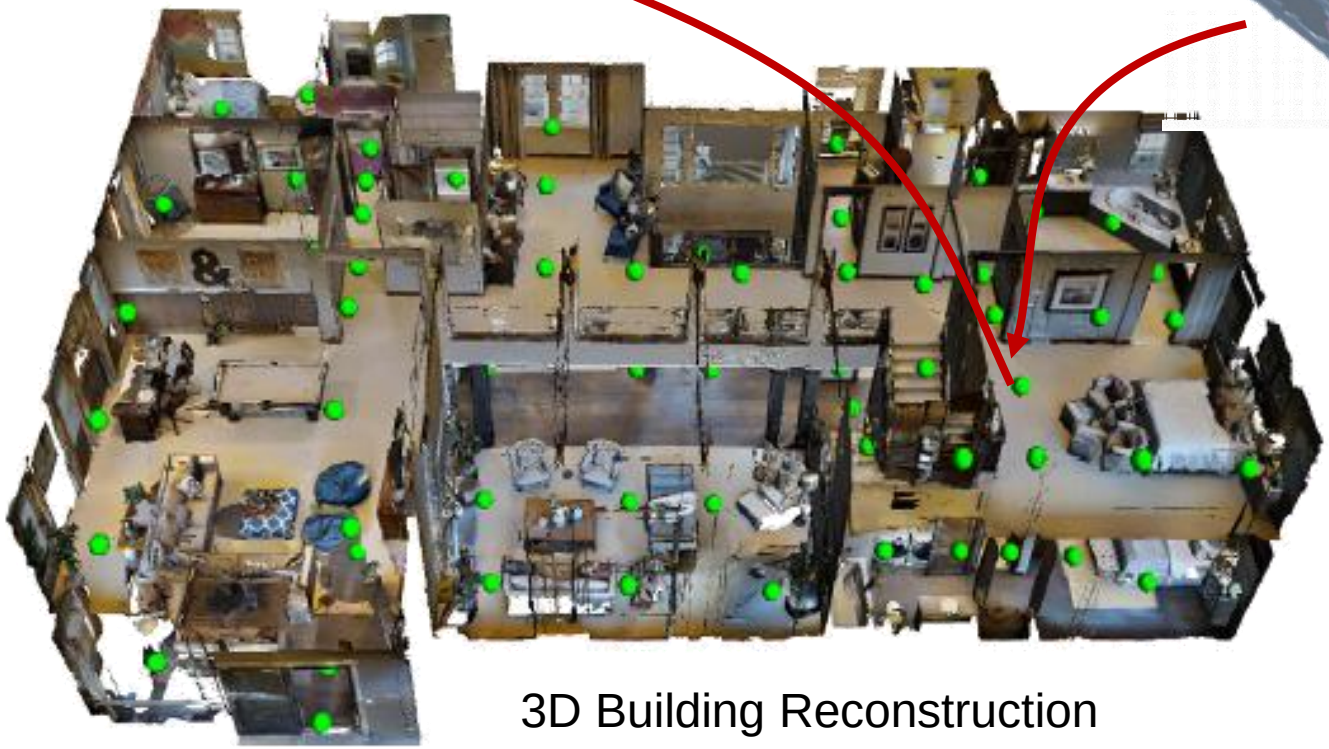
3D Building Reconstruction

A. Chang, A. Dai, T. Funkhouser, M. Halber, M. Niessner, M. Savva, S. Song, A. Zeng, Y Zhang, “Matterport3D: Learning from RGB-D Data in Indoor Environments,” 3DV 2017

Semantic View Extrapolation: Data

Matterport3D dataset

Matterport Camera



3D Building Reconstruction

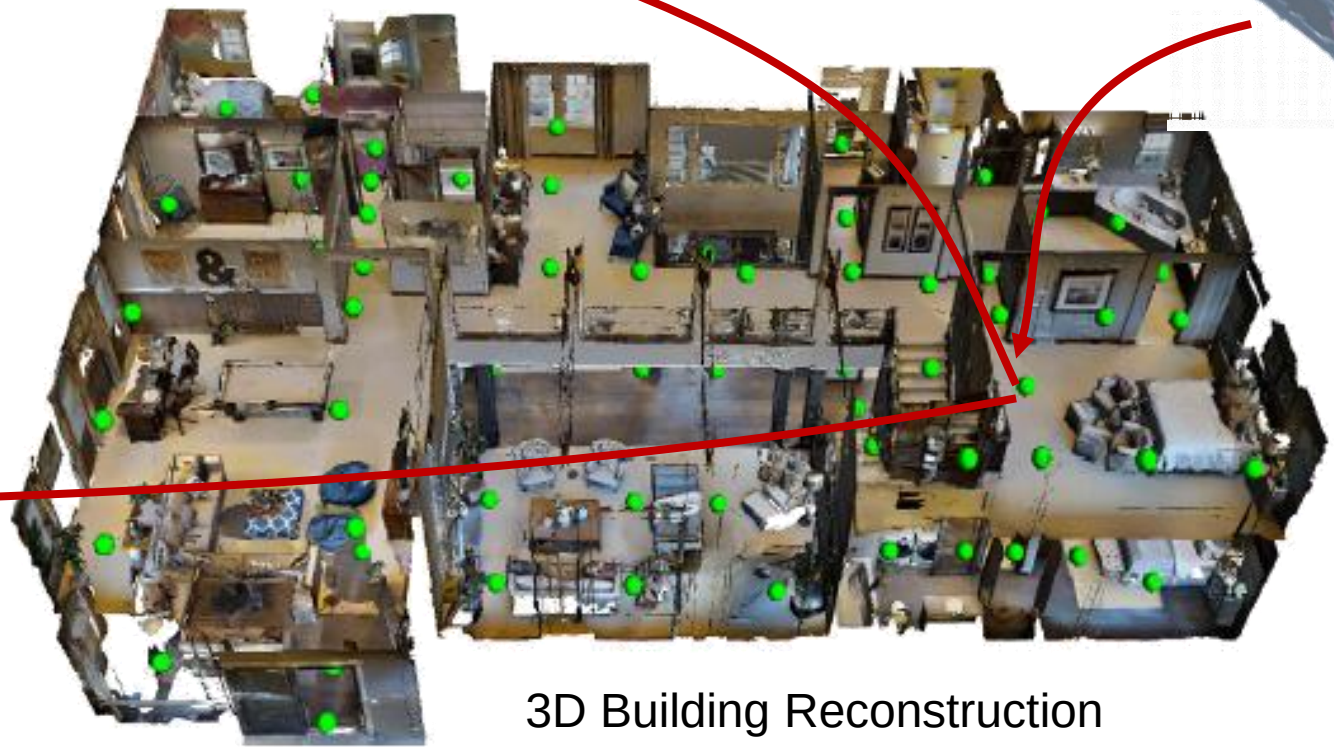
A. Chang, A. Dai, T. Funkhouser, M. Halber, M. Niessner, M. Savva, S. Song, A. Zeng, Y Zhang, “Matterport3D: Learning from RGB-D Data in Indoor Environments,” 3DV 2017

Semantic View Extrapolation: Data

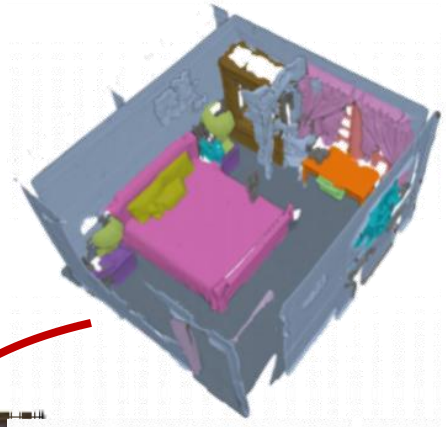
Matterport3D dataset

Matterport Camera

RGB-D Panorama
with Semantics



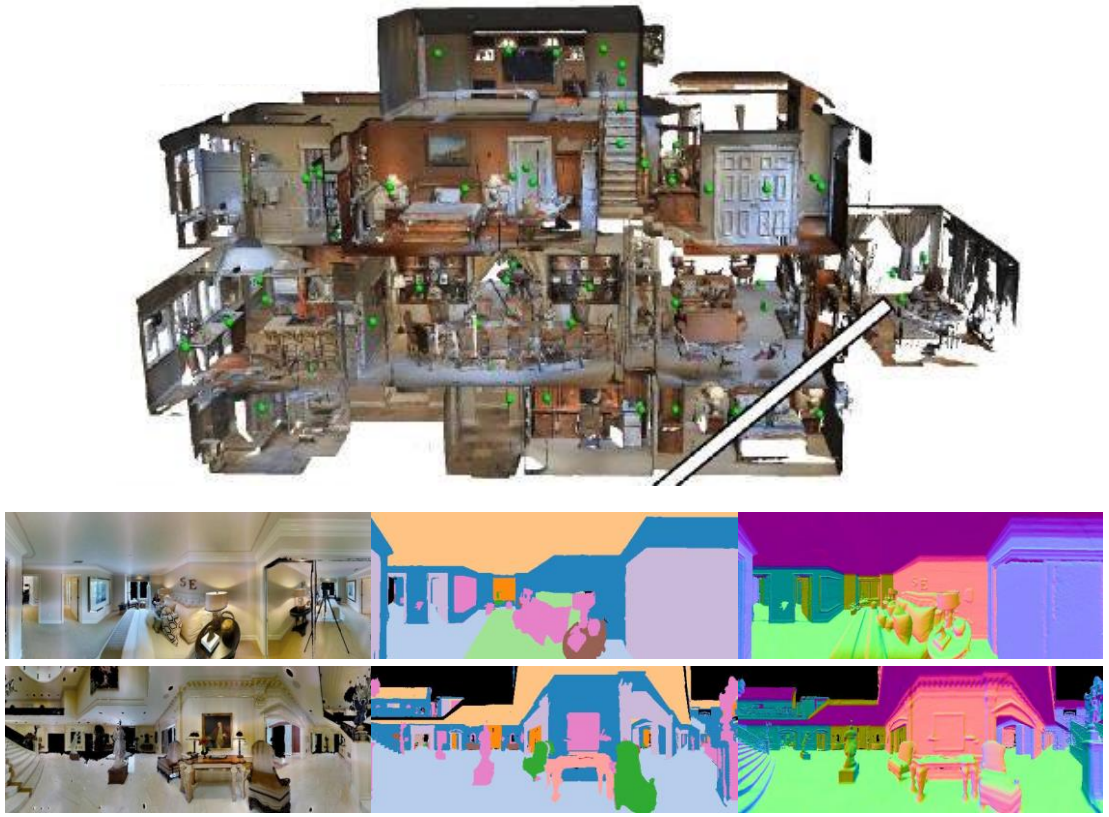
3D Building Reconstruction



A. Chang, A. Dai, T. Funkhouser, M. Halber, M. Niessner, M. Savva, S. Song, A. Zeng, Y Zhang, "Matterport3D: Learning from RGB-D Data in Indoor Environments," 3DV 2017

Semantic View Extrapolation: Experiments

Pre-train on SUNCG
58,866 synthetic panoramas



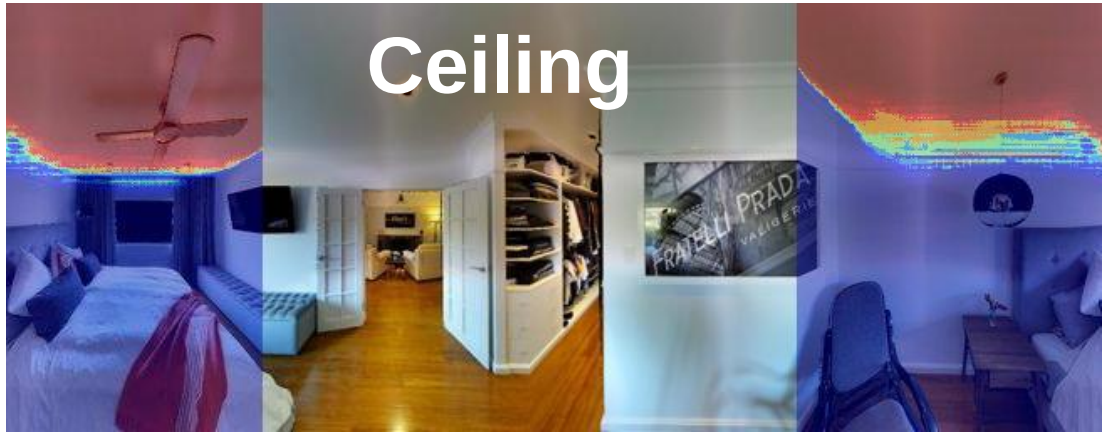
Fine-tune and test on Matterport3D
5,315 real panoramas

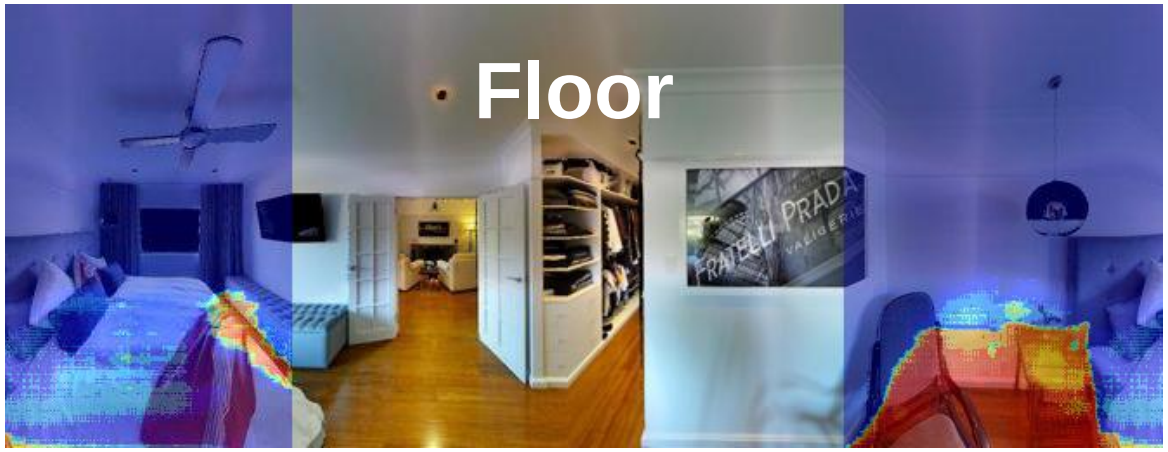
Semantic View Extrapolation: Results

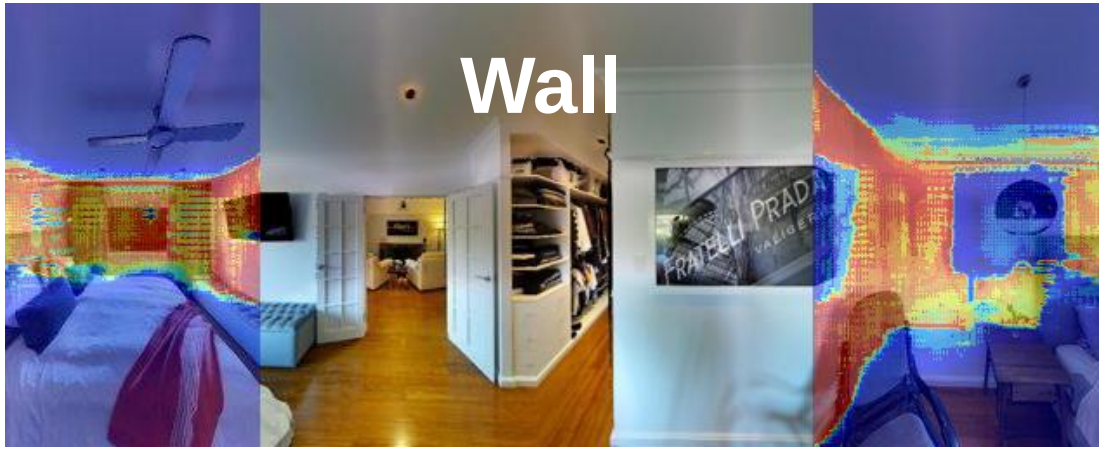
Input Observation

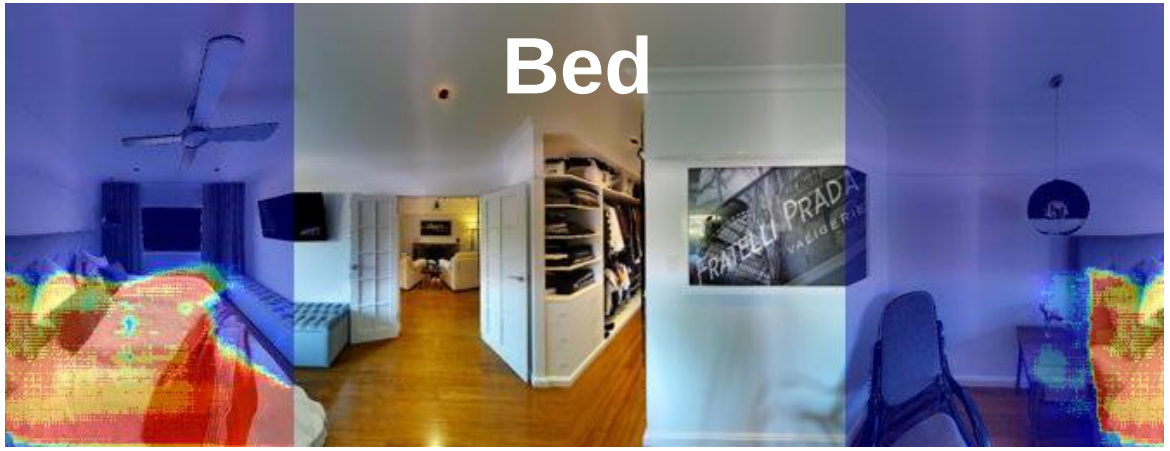
Semantic View Extrapolation: Results

Prediction



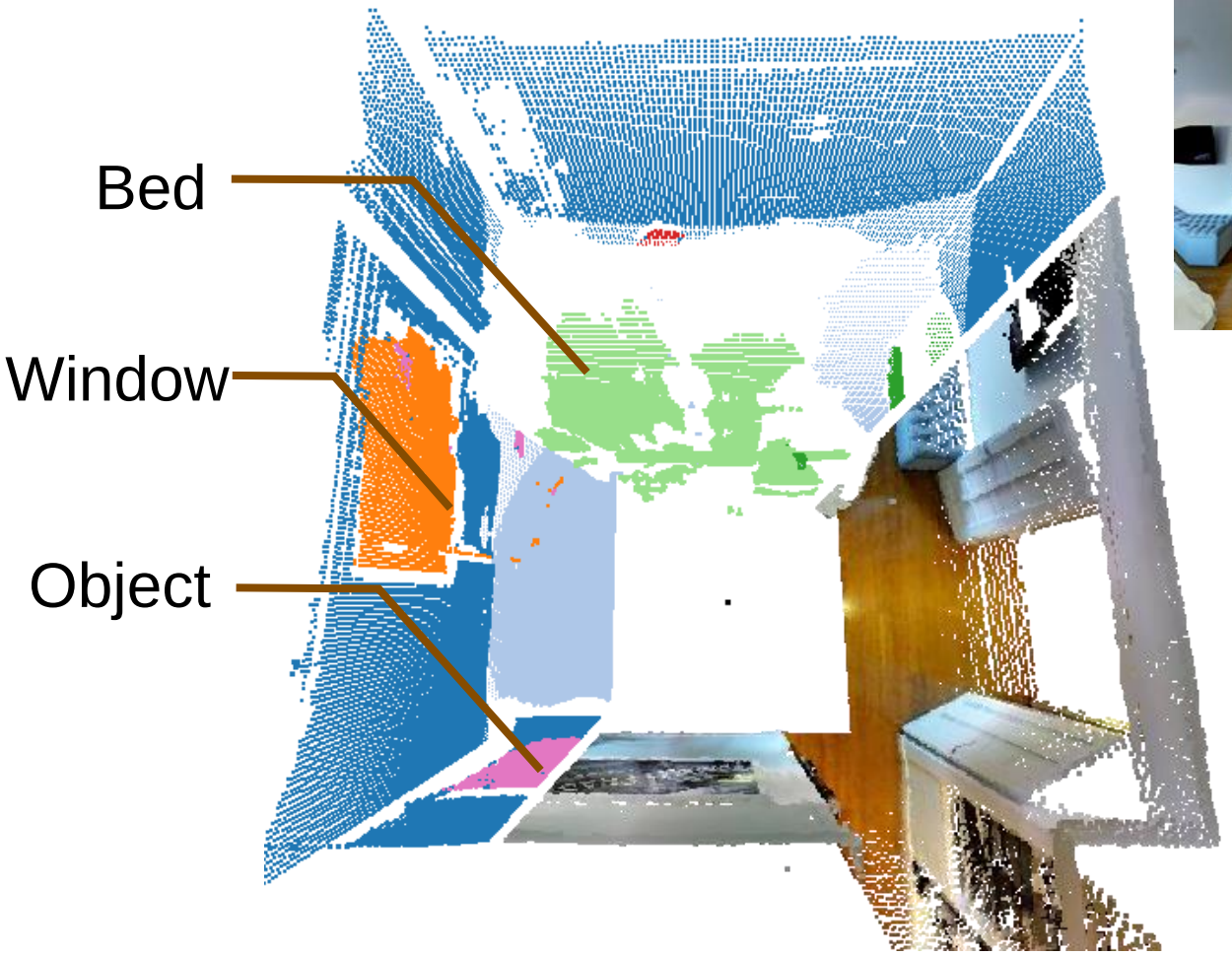






Semantic View Extrapolation: Results

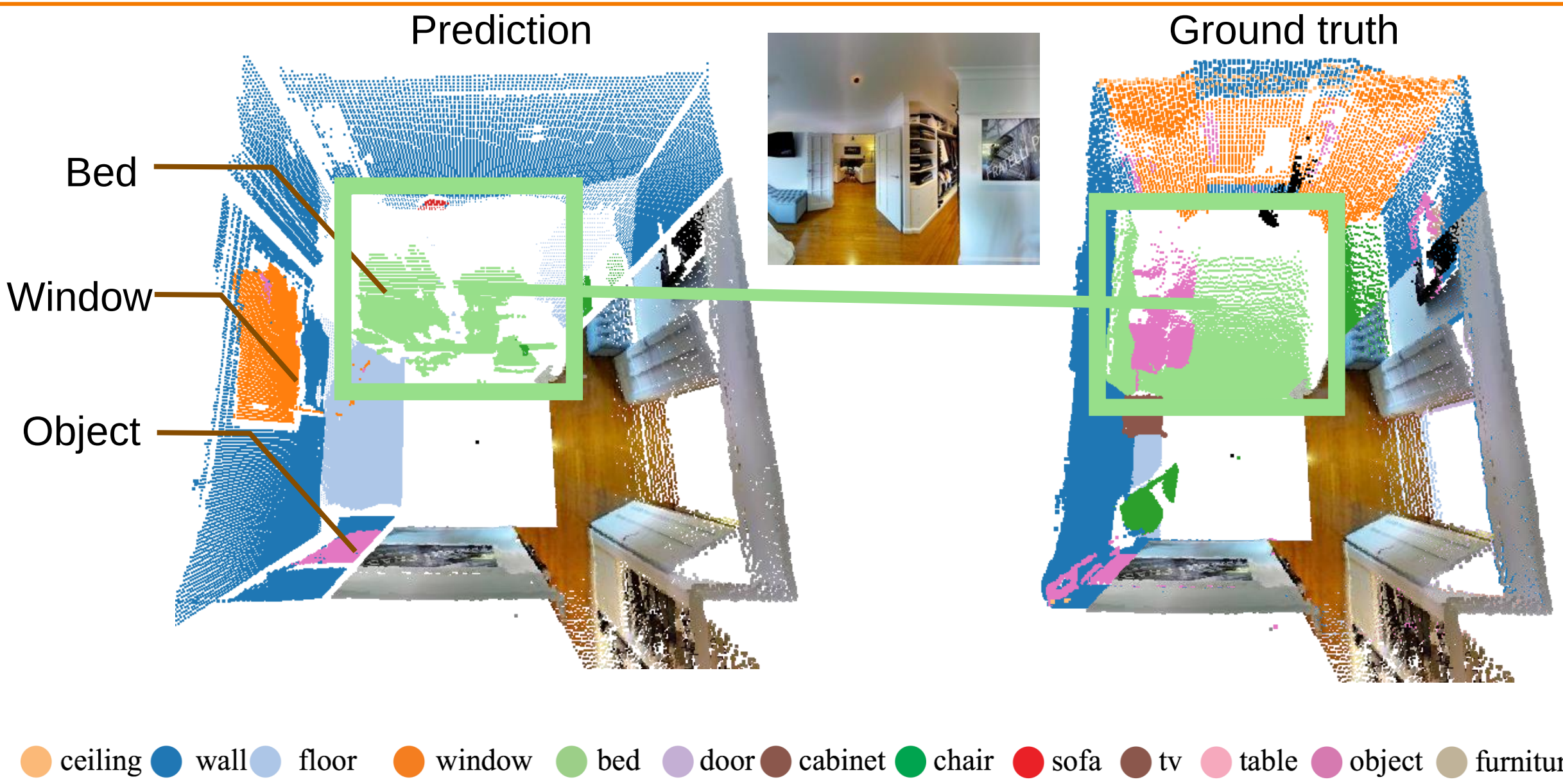
Prediction



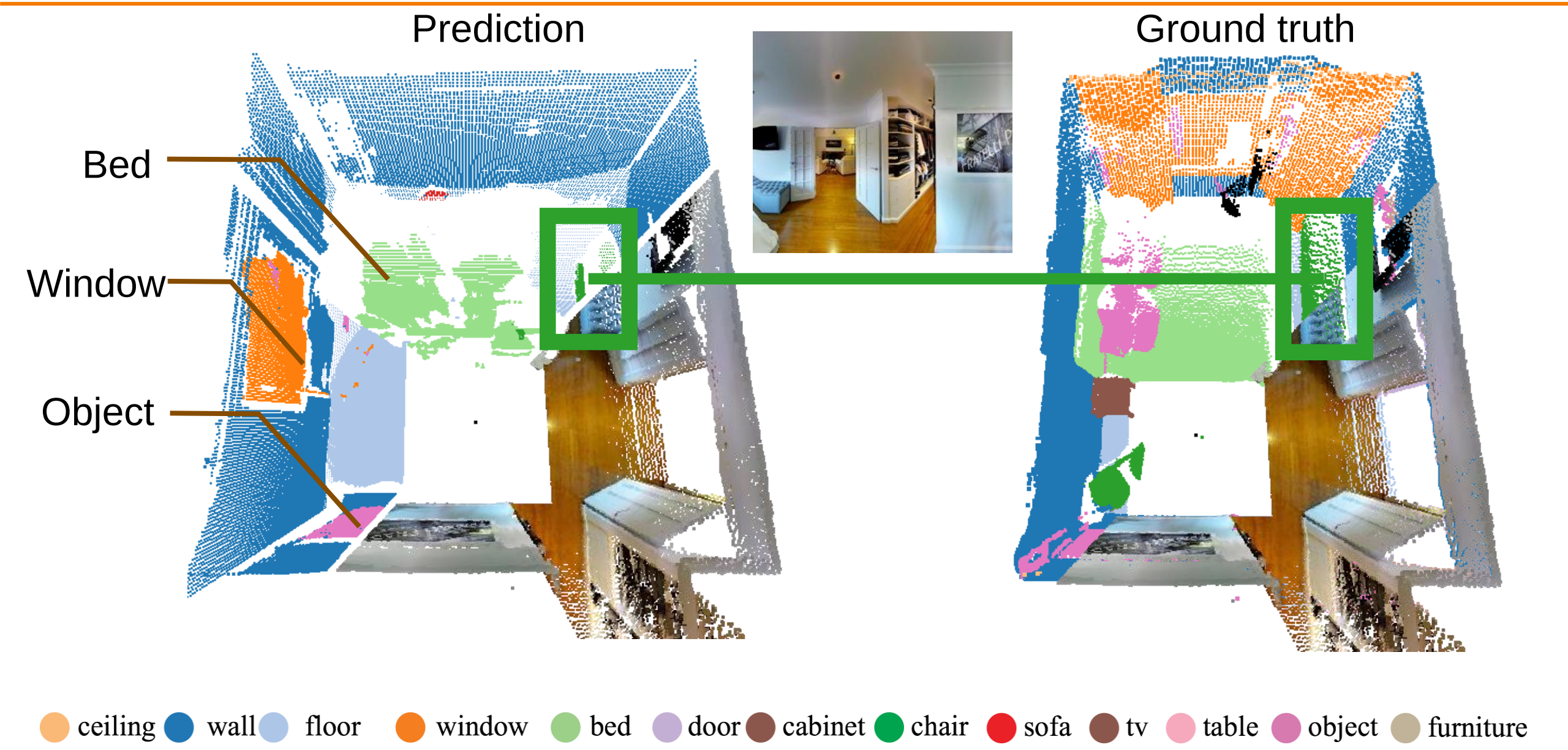
Ground truth

ceiling wall floor window bed door cabinet chair sofa tv table object furniture

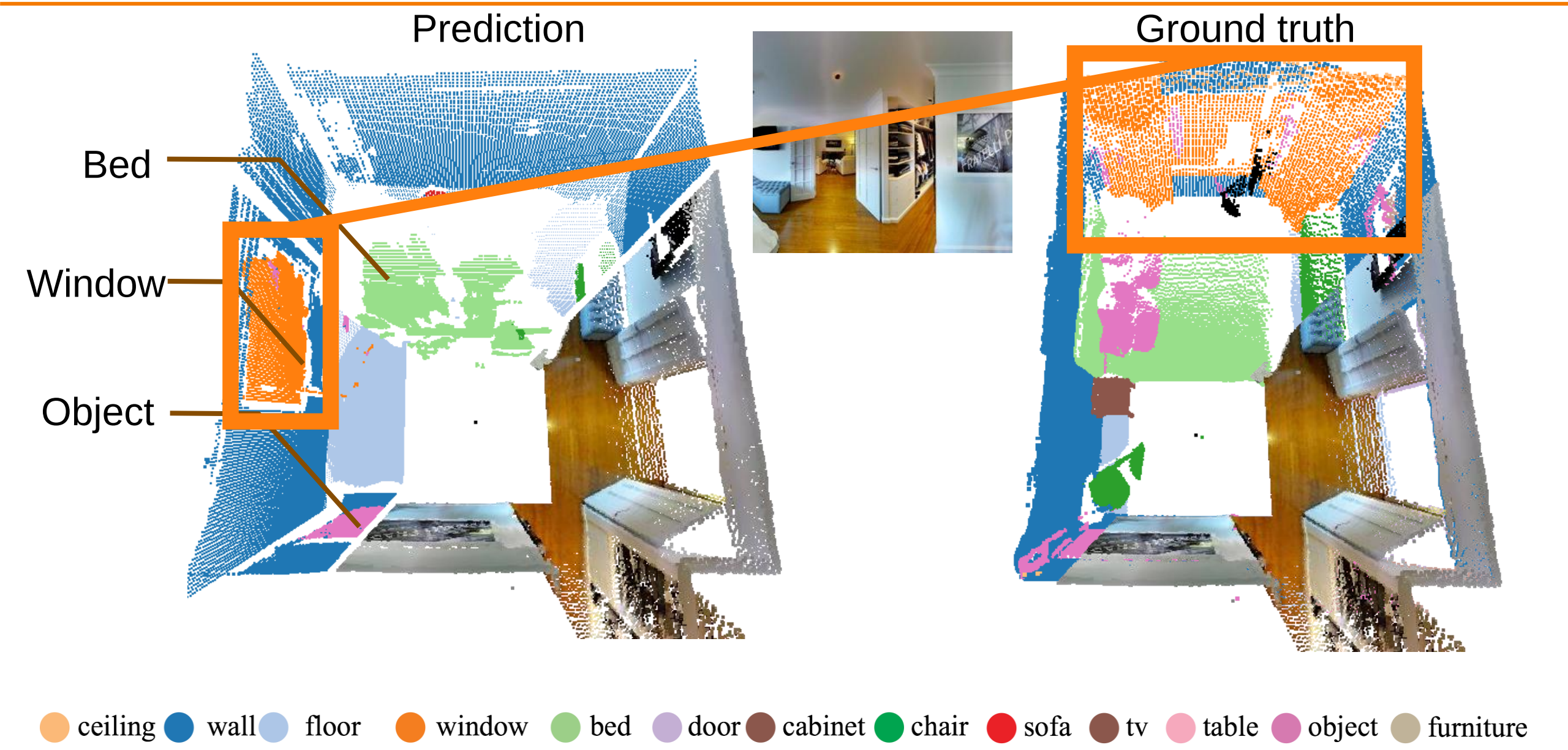
Semantic View Extrapolation: Results



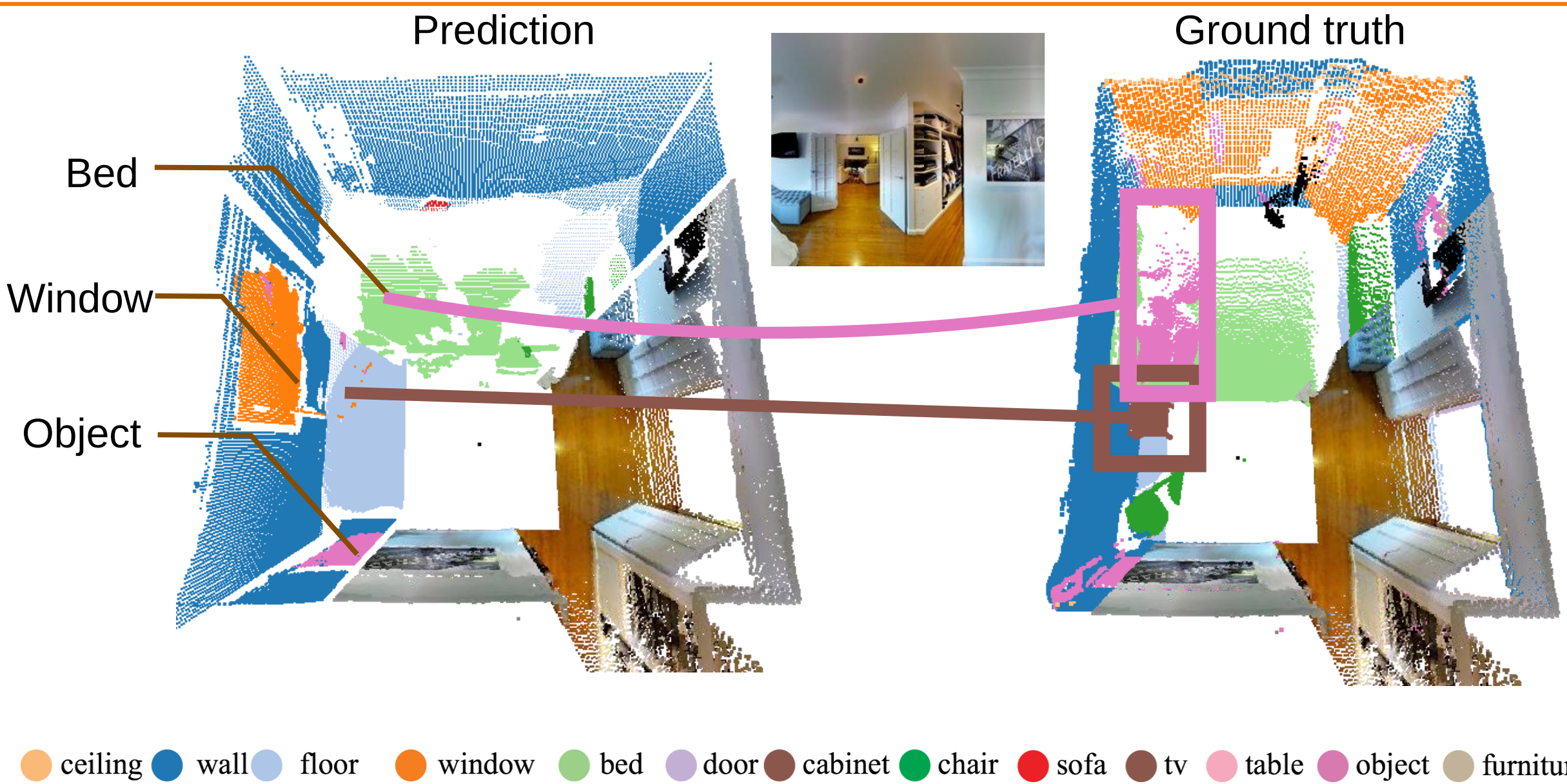
Semantic View Extrapolation: Results



Semantic View Extrapolation: Results

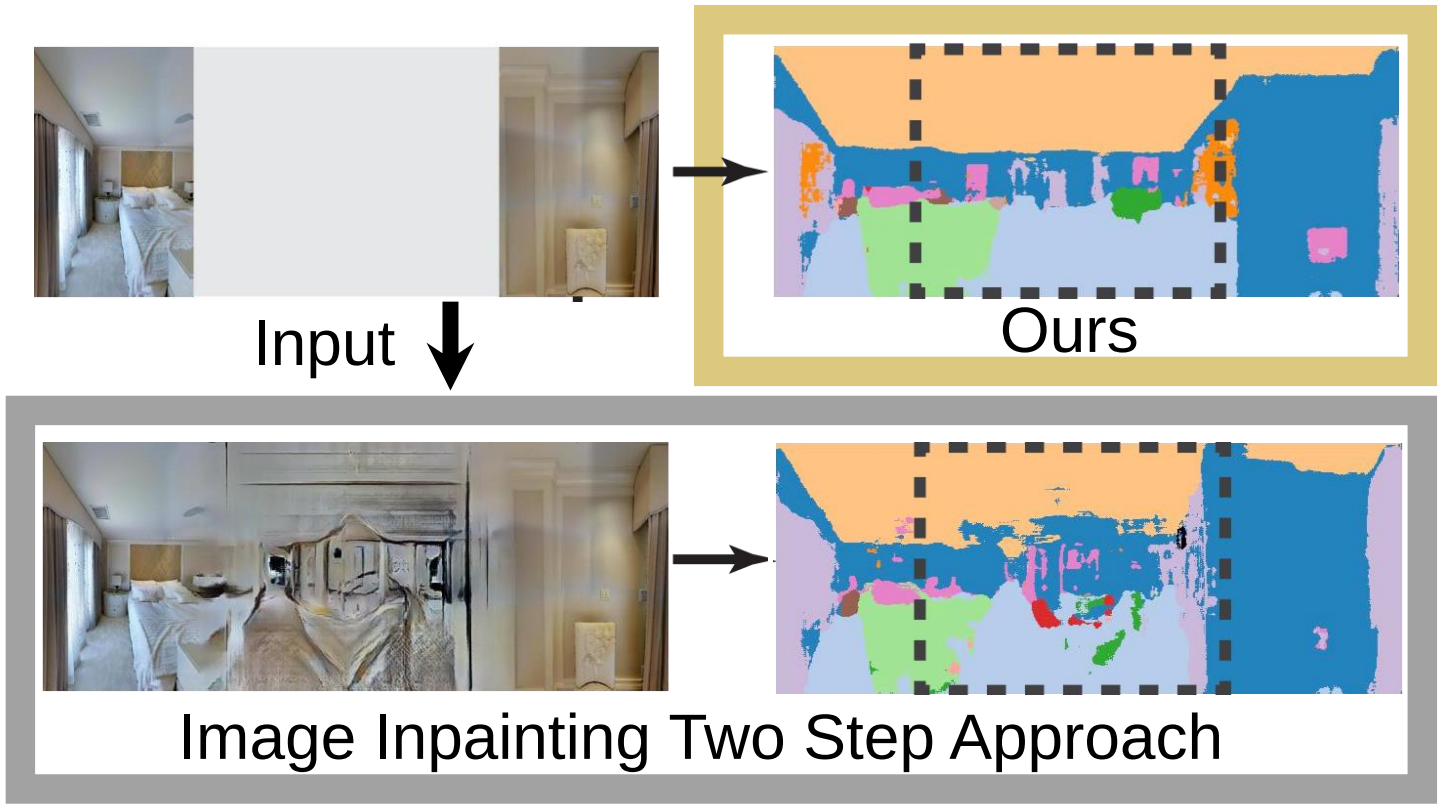


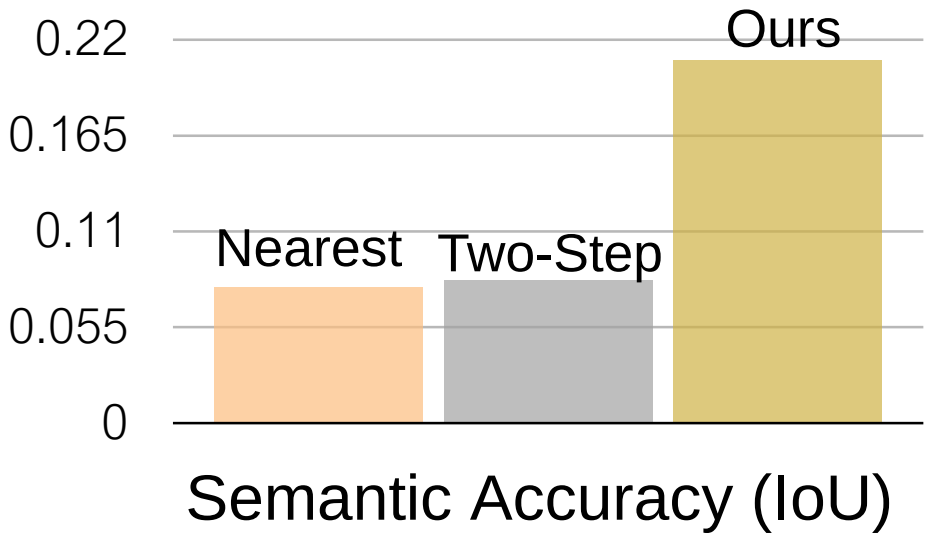
Semantic View Extrapolation: Results

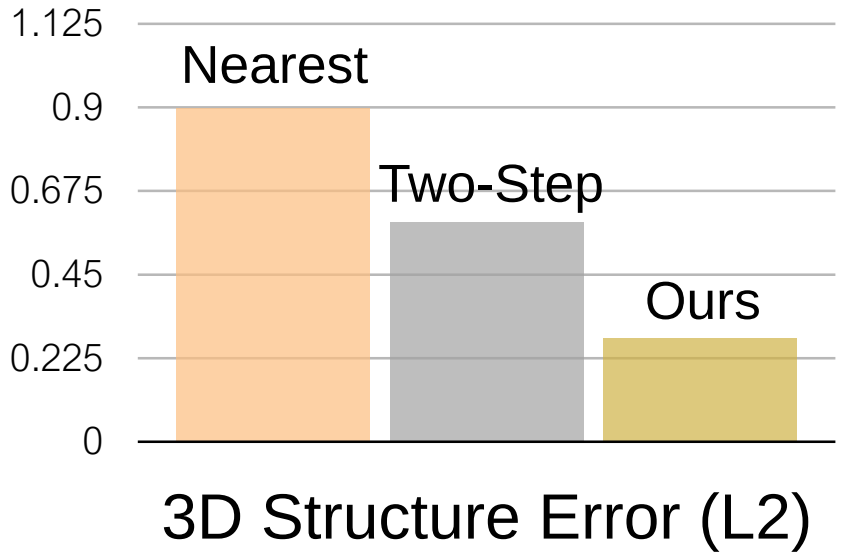


Semantic View Extrapolation: Results

Comparison to alternative completion methods

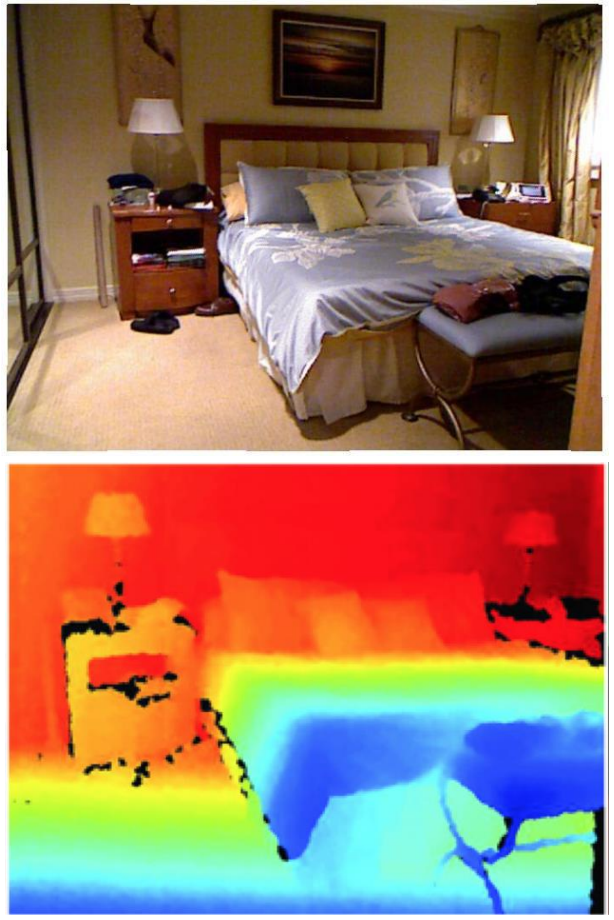




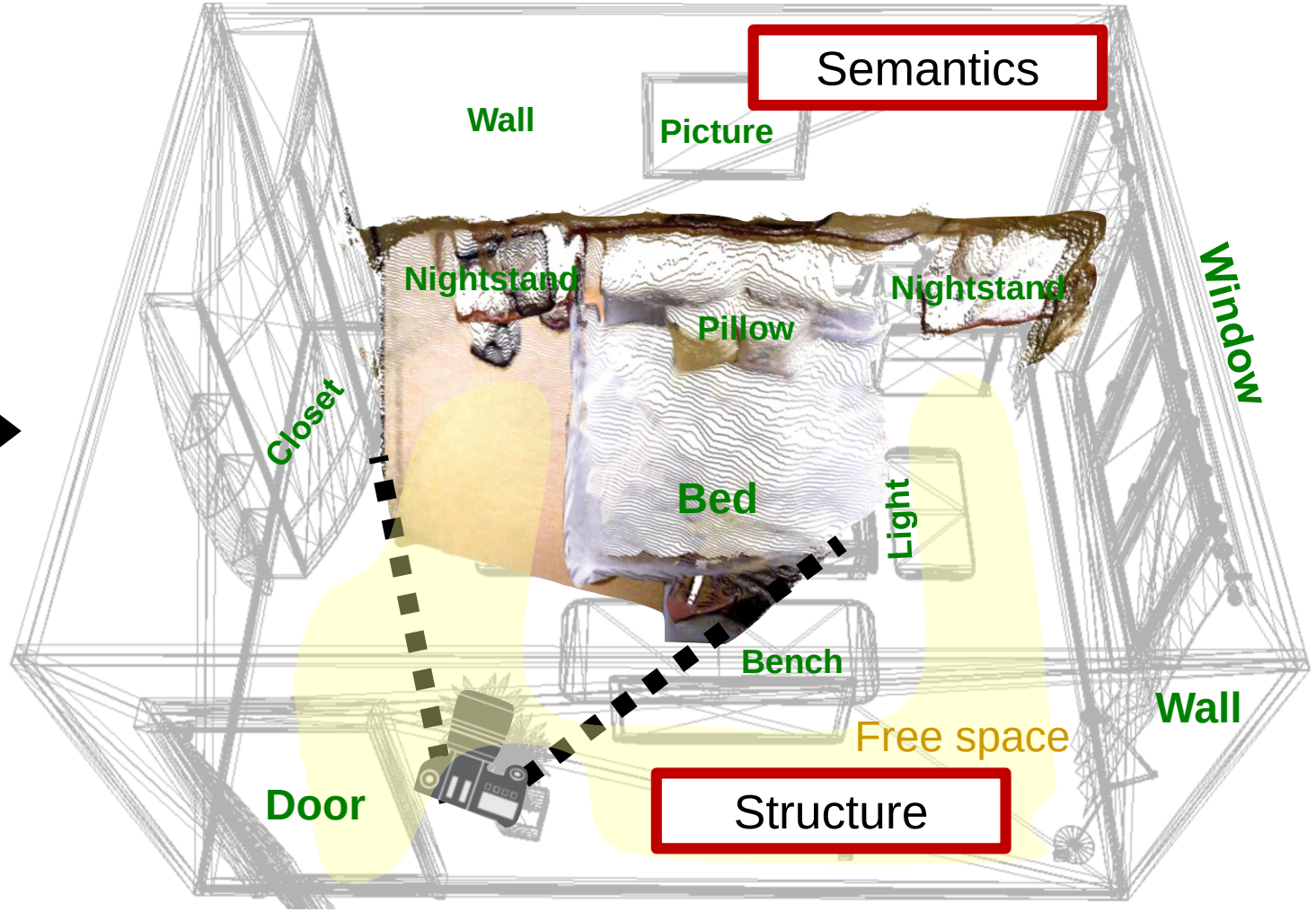


Summary

Scene understanding from partial observation …



Input: RGB-D Image



Output: complete, annotated 3D representation

Talk Outline

Introduction

Three recent projects

- Deep depth completion [CVPR 2018]
- Semantic scene completion [CVPR 2017]
- Semantic view extrapolation [CVPR 2018]

Common themes

Future work

Common Themes

Geometric representation

- Choice of 3D representation is critical
- Choosing the most obvious representation is usually not best

Large-scale context

- Global context is very important … even for simply estimating depth
- Can leverage larger contexts with global minimization, dilated convolutions, etc.

3D Dataset curation

- Synthetic 3D datasets very useful for training
- Real 3D datasets are important for testing. More needed

Common Themes

Geometric representation

- Choice of 3D representation is critical
- Choosing the most obvious representation is usually not best

Large-scale context

- Global context is very important … even for simply estimating depth
- Can leverage larger contexts with global minimization, dilated convolutions, etc.

3D Dataset curation

- Synthetic 3D datasets very useful for training
- Real 3D datasets are important for testing. More needed

Common Themes

Geometric representation

- Choice of 3D representation is critical
- Choosing the most obvious representation is usually not best

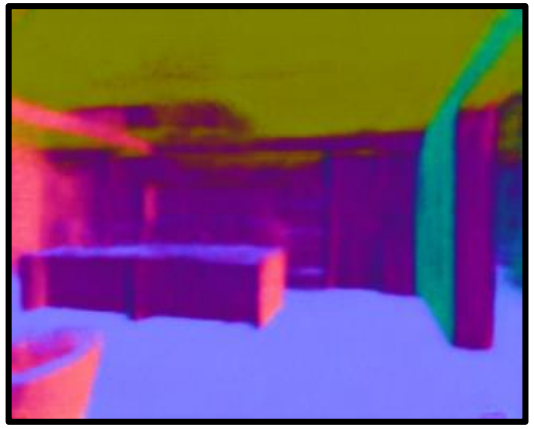
Large

- Glo
- Ca

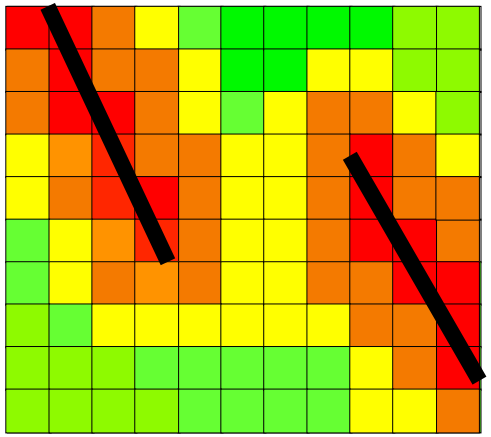
3D Da

- Sy
- Re

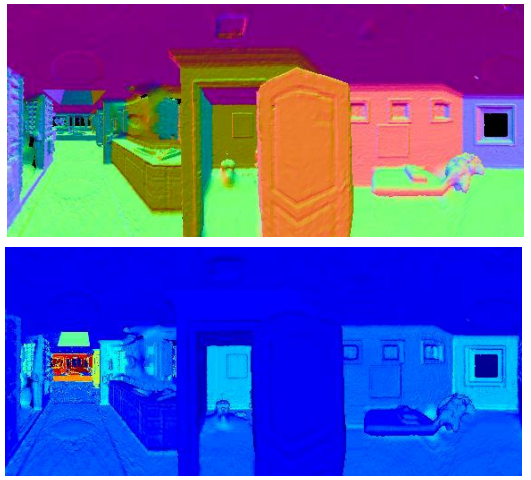
etc.



Surface Normals



Flipped TSDF



Plane Equations

Common Themes

Geometric representation

- Choice of 3D representation is critical
- Choosing the most obvious representation is usually not best

Large-scale context

- Global context is very important … even for simply estimating depth
- Can leverage larger contexts with global minimization, dilated convolutions, etc.

3D Dataset curation

- Synthetic 3D datasets very useful for training
- Real 3D datasets are important for testing. More needed

Common Themes

Geometric representation

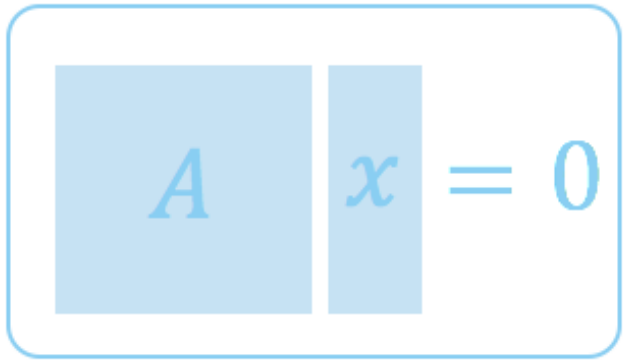
- Choice of 3D representation is critical
- Choosing the most obvious representation is usually not best

Large-scale context

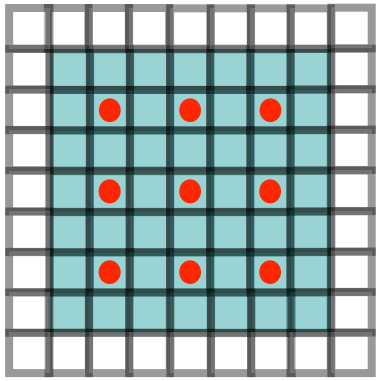
- Global context is very important … even for simply estimating depth
- Can leverage larger contexts with global minimization, dilated convolutions, etc.

3D

-
-



Global Solution to
Linear System of Equations



Dilated
Convolutions

Panoramic
Representations

Common Themes

Geometric representation

- Choice of 3D representation is critical
- Choosing the most obvious representation is usually not best

Large-scale context

- Global context is very important … even for simply estimating depth
- Can leverage larger contexts with global minimization, dilated convolutions, etc.

3D Dataset curation

- Synthetic 3D datasets very useful for training
- Real 3D datasets are important for testing. More needed

Common Themes

Geometric represe

- Choice of 3D repr
- Choosing the mos

Large-scale contex

- Global context is
- Can leverage larg

3D Dataset curatio

- Synthetic 3D datasets very useful for training
- Real 3D datasets are important for testing. More needed

	Synthetic	RGB-D Image	RGB-D Video
Object	ShapeNet	Intel RealSense	Redwood
Room	SUNCG	SUN RGB-D	ScanNet
Multiroom	SUNCG	Matterport3D	SUN3D

Largest 3D datasets available today for indoor environments

Talk Outline

Introduction

Three recent projects

- Deep depth completion [CVPR 2018]
- Semantic scene completion [CVPR 2017]
- Semantic view extrapolation [CVPR 2018]

Common themes

Future work

Future work

Large-scale scenes

Self-supervision

Active sensing

Acknowledgments

Princeton students and postdocs:

- Angel X. Chang, Kyle Genova, Maciej Halber, Manolis Savva, Elena Sizikova, Shuran Song, Fisher Yu, Yinda Zhang, Andy Zeng

Google collaborators:

- Martin Bokeloh, Alireza Fathi, Sean Fanello, Aleksey Golovinskiy, Shahram Izadi, Sameh Khamis, Adarsh Kowdle, Johnny Lee, Christoph Rhemann, Jurgen Sturm, Vladimir Tankovich, Julien Valentin, Stefan Welker

Other collaborators:

- Angela Dai, Vladlen Koltun, Matthias Niessner, Alberto Rodriquez, Silvio Savarese, Yifei Shi, Jianxiong Xiao, Kai Xu

Data:

- SUN3D, NYU, Trimble, Planner5D, Matterport

Funding:

- NSF, Google, Intel, Facebook, Amazon, Adobe, Pixar

Thank You!