Surfaces, Objects, Procedures: Integrating Learning and Graphics for 3D Scene Understanding

Jiajun Wu


Stanford University

Xiuming Zhang

Zhoutong Zhang

Jiayuan Mao

Yikai Li

Bill Freeman

Josh Tenenbaum

Noah Snavely

Jun-Yan Zhu

3D Scene Understanding

How many cars are there in the scene?

What's the color of the closest car?

How would the car look like from the front?

What's going to happen?

What if it is rainy?

Also applies to **novel** objects

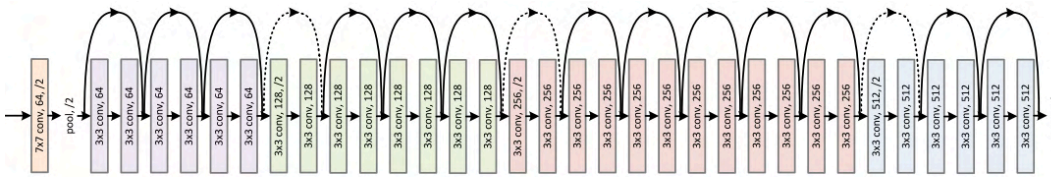
2D Image

Vision

Graphics

3D Shape

Deep Learning



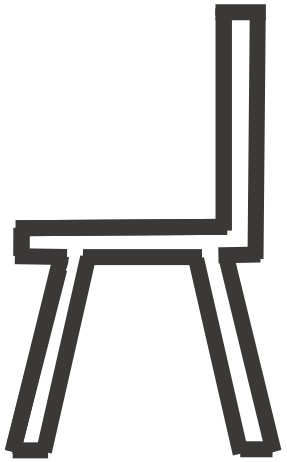
2D Image

Graphics

3D Shape

Forward: Image Formation

Image



Visible Surface

3D Shape

Inverse: Shape Estimation

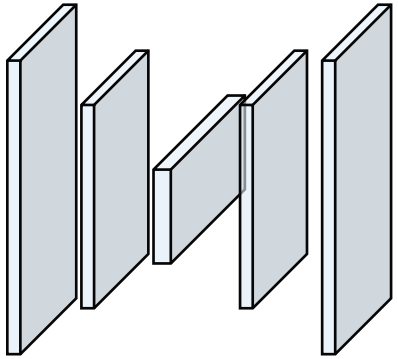
Depth Estimation

Depth Map

Shape Completion

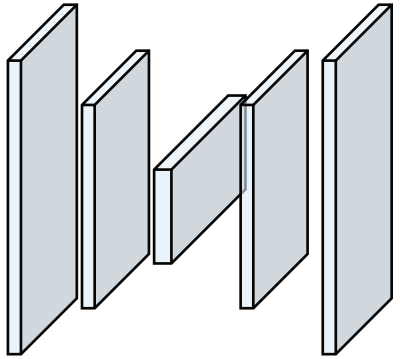
Inverting the Graphics Engine

Image



Depth Estimation

Depth

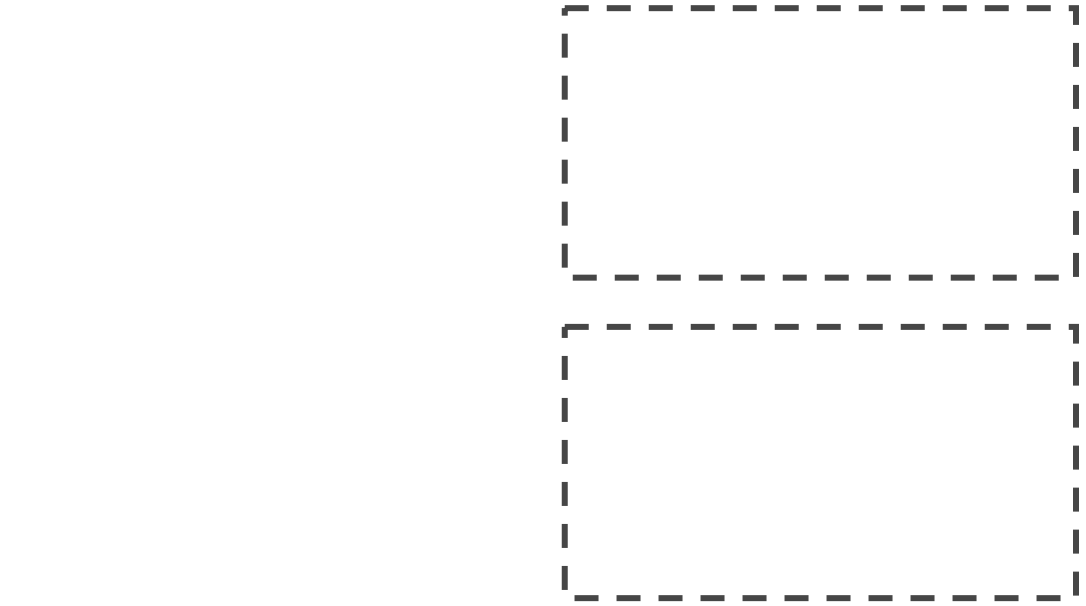


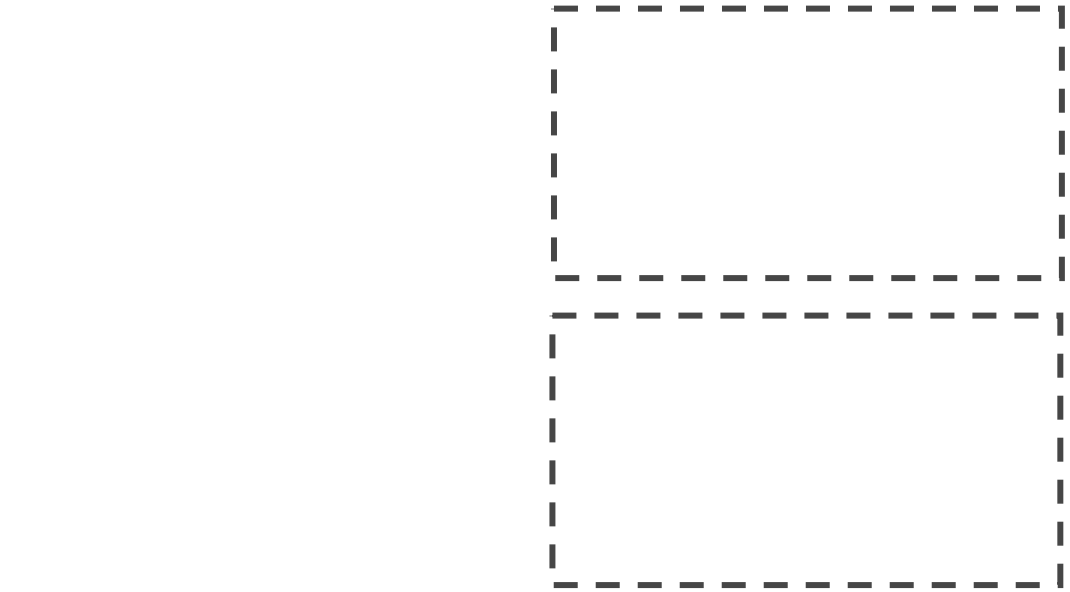
Shape Completion

Shape

Trained on:
ShapeNet [Chang et al., 2016]

with Yifan Wang

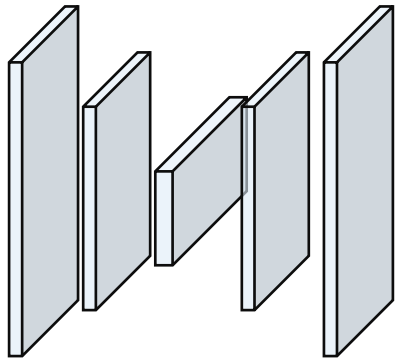




Wu*, Wang*, Xue, Sun, Freeman, Tenenbaum. NeurIPS'17

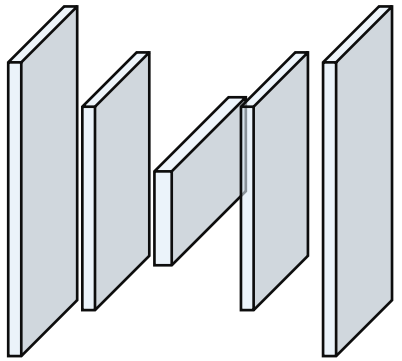
Generalization to Unseen Classes

Image



Depth Estimation

Depth



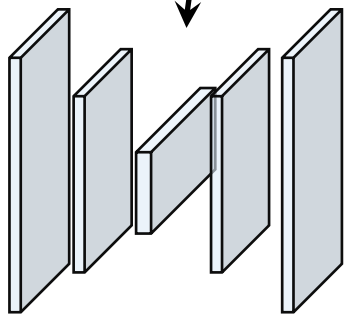
Shape Completion

3D-R2N2 [ECCV'16]
DRC [CVPR'17]
PSGN [CVPR'17]
OGN [ICCV'17]
MarrNet [NeurIPS'17]
AtlasNet [CVPR'18]
ShapeHD [ECCV'18]
Multi-View [ECCV'18]

If we take models trained on chairs, planes, cars

Training Images

Test Image (Table)



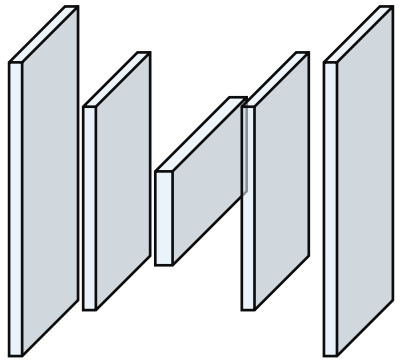
Network

Direct Prediction

What do single-view 3D reconstruction networks learn? Tatarchenko*, Richter*, Ranftl, Li, Koltun, Brox. CVPR'19

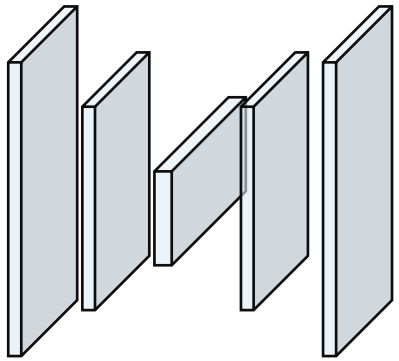
Inverting the Graphics Engine

Image



Depth Estimation

Depth



Shape Completion

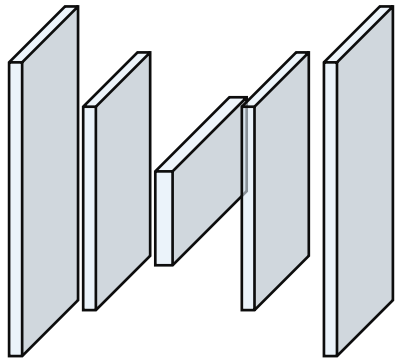
Shape

Projecting depth into 3D is a deterministic, fully differentiable process.

Shape completion network is over-parameterized: learning a deterministic mapping.

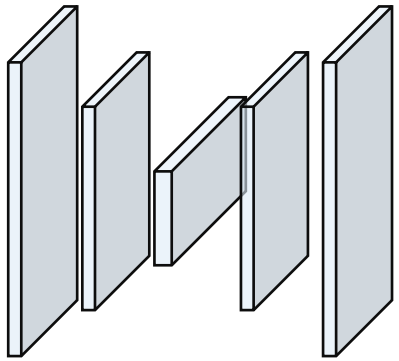
Inverting the Graphics Engine

Image



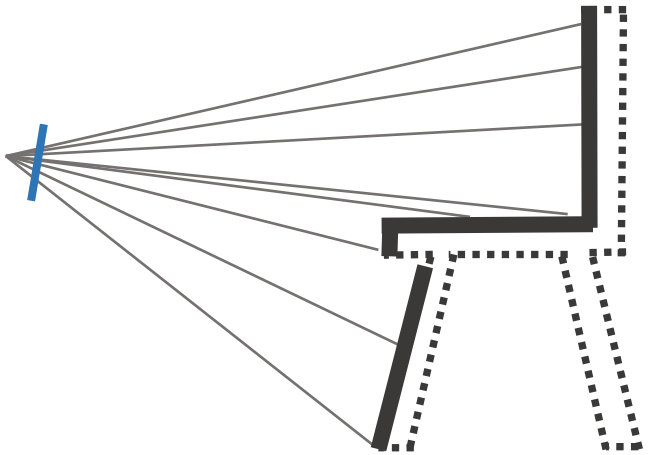
Depth Estimation

Depth

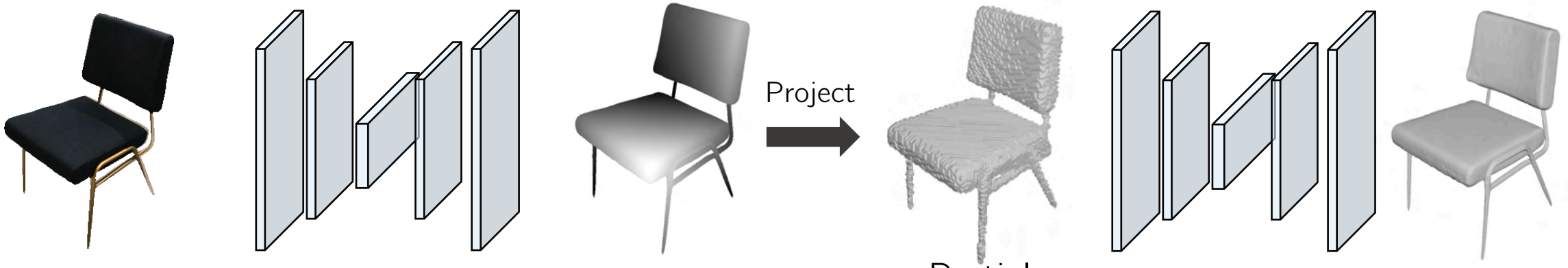


Shape Completion

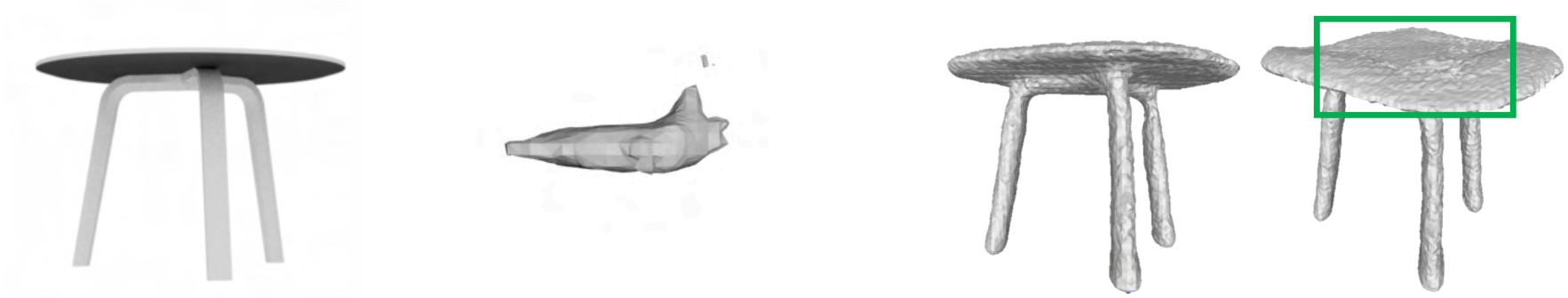
Shape



Inverting the Graphics Engine

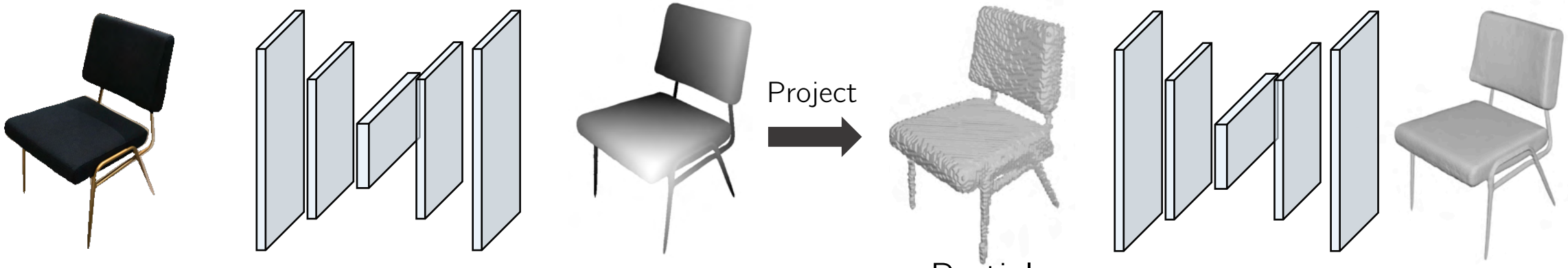


If we take models trained on chairs, planes, cars



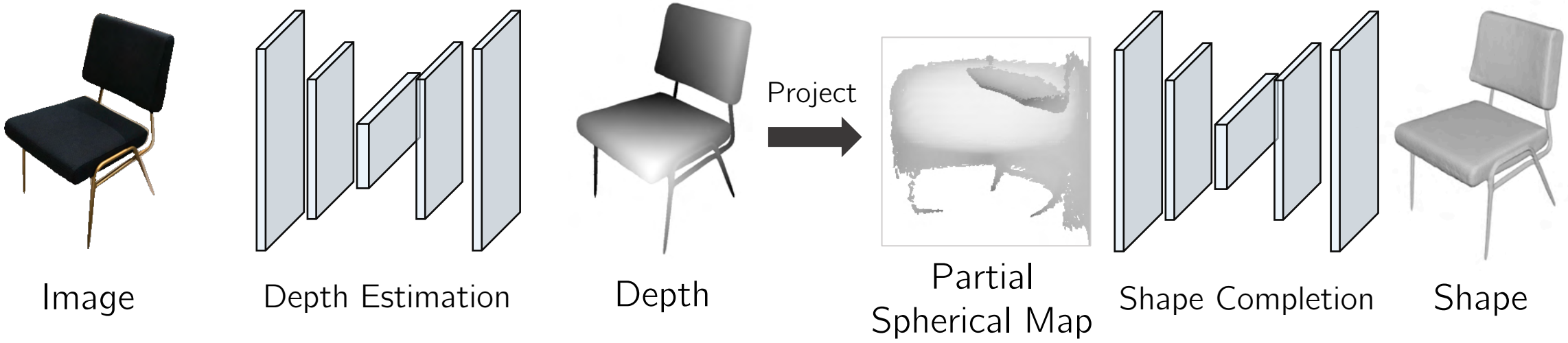
Input Image (Table) Direct Prediction Modeling Depth and Projection

Inverting the Graphics Engine

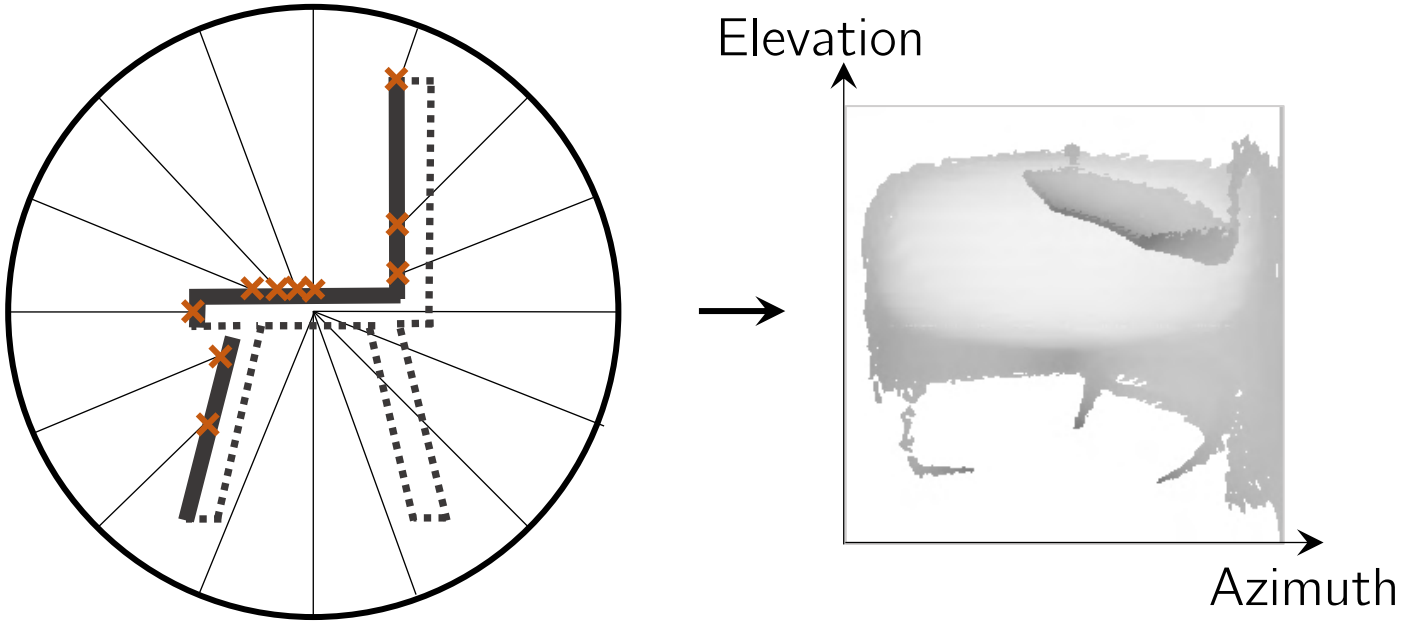


Partial surface in 3D is very sparse
-> hard for the completion net to capture surface features.

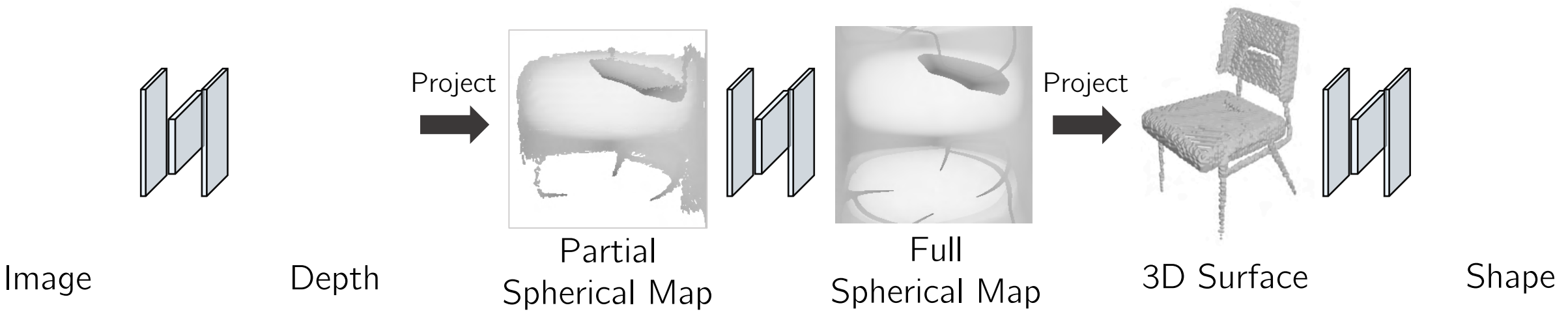
Inverting the Graphics Engine

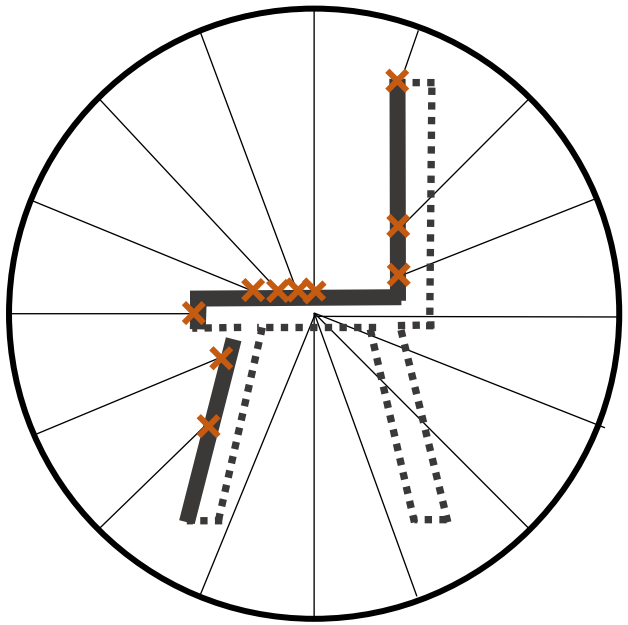


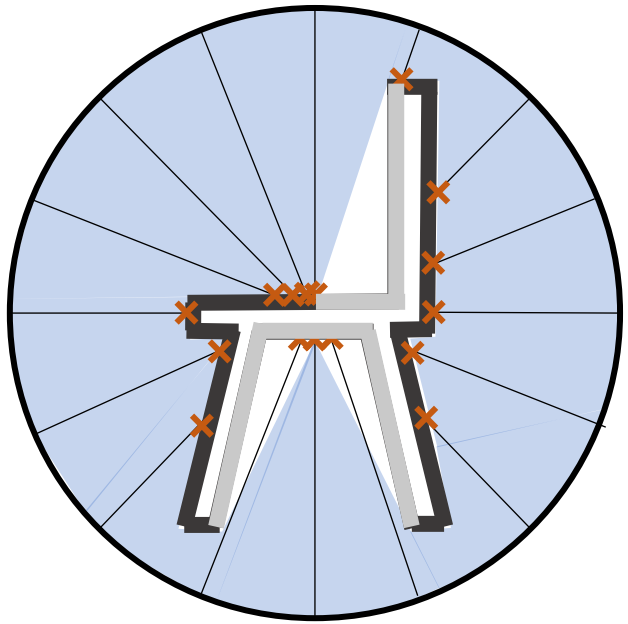
Spherical map as a surrogate representation for surfaces in 3D



Inverting the Graphics Engine







Learning to Reconstruct an Unseen 3D Object (GenRe)


with Xiuming Zhang Zhoutong Zhang

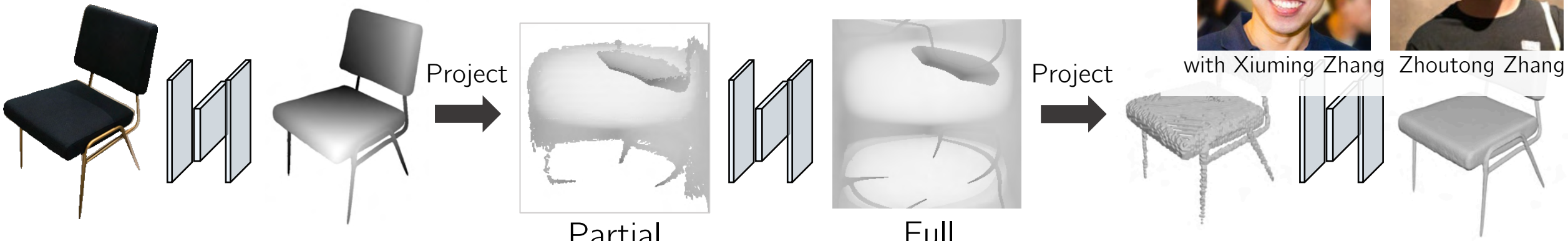
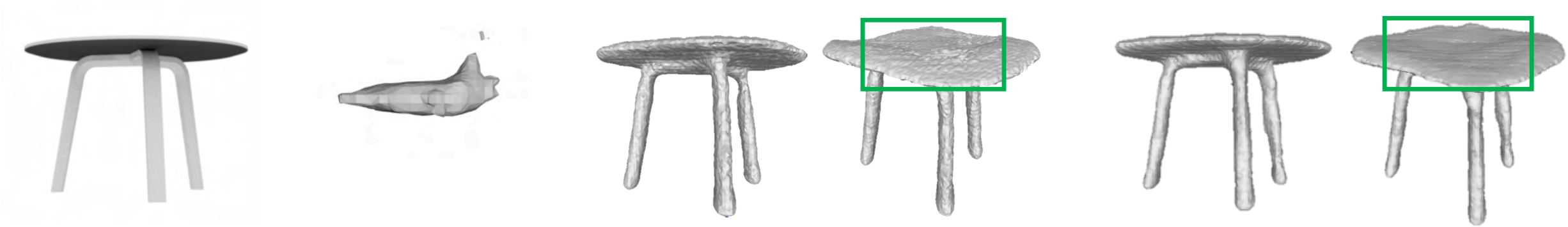


Image | Depth | Partial Spherical Map | Full Spherical Map | 3D Surface | Shape

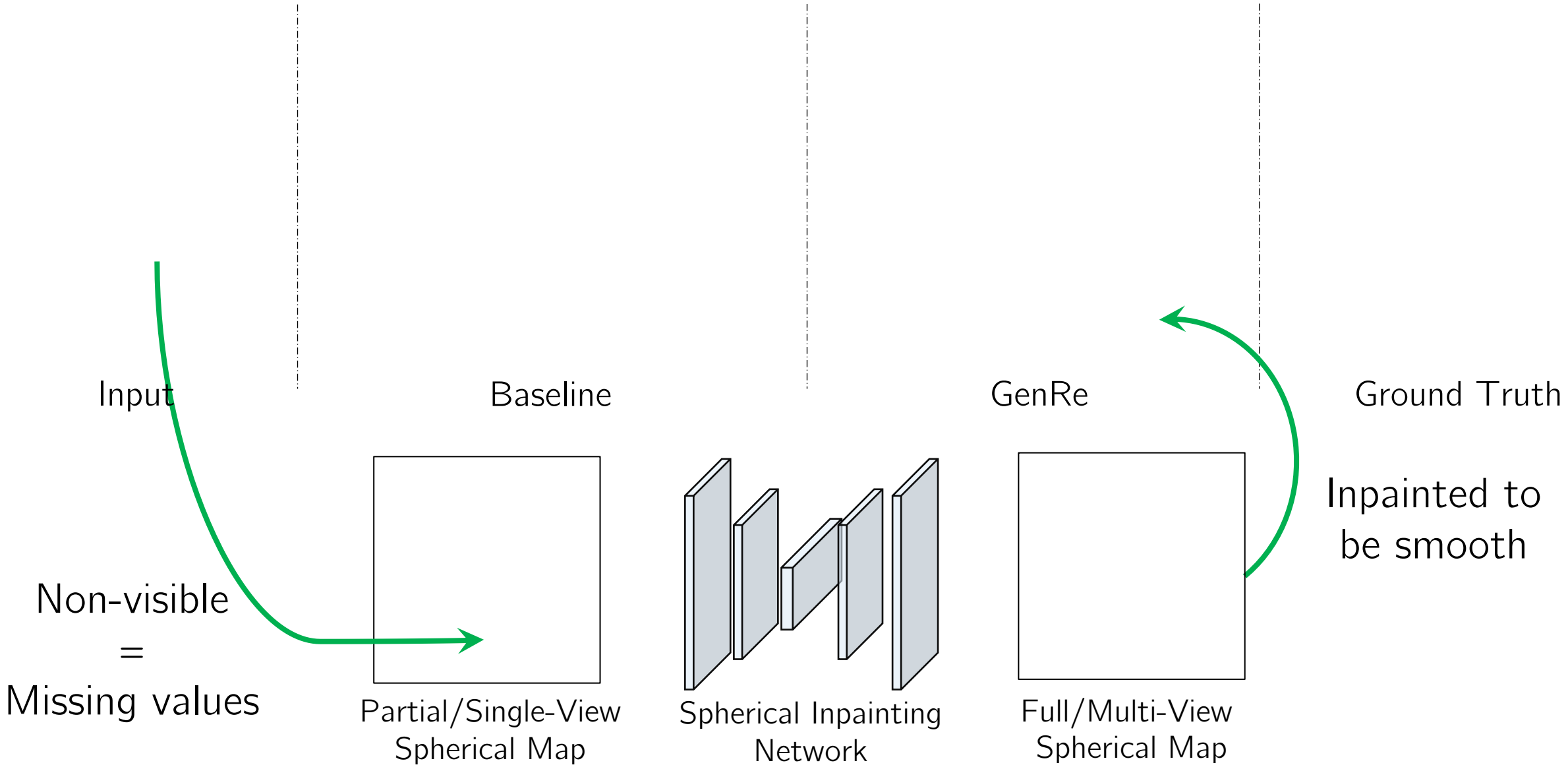
If we take models trained on chairs, planes, cars



Input Image (Table) | Direct Prediction | Without Spherical Maps | With Spherical Maps

Zhang*, Zhang*, Zhang, Tenenbaum, Freeman, Wu. NeurIPS'18

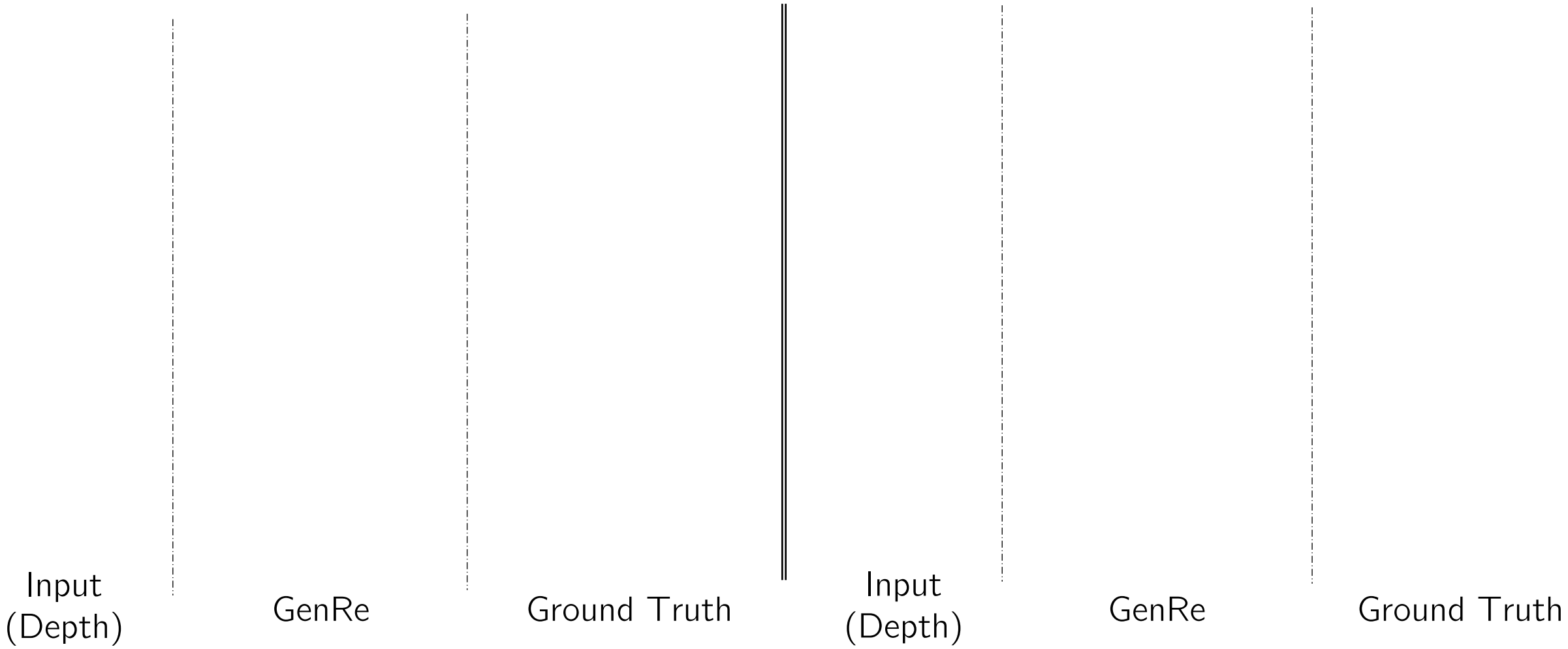
Results: Generalizing to Unseen Classes



Results: Generalizing to Unseen Classes

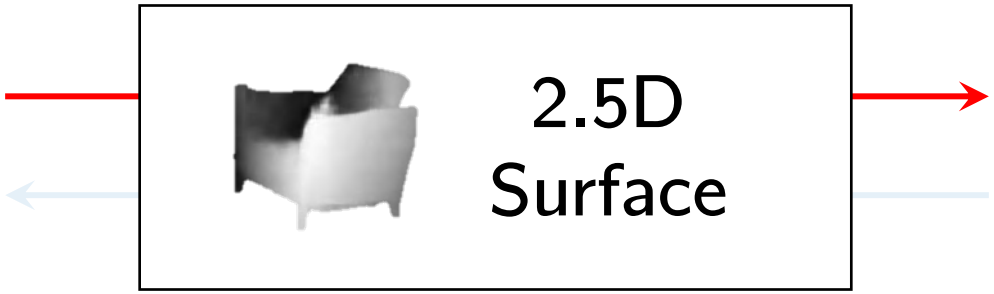
Training: cars, chairs, airplanes
Testing: beds, tables, benches

Results: Generalizing to Non-Rigid Classes



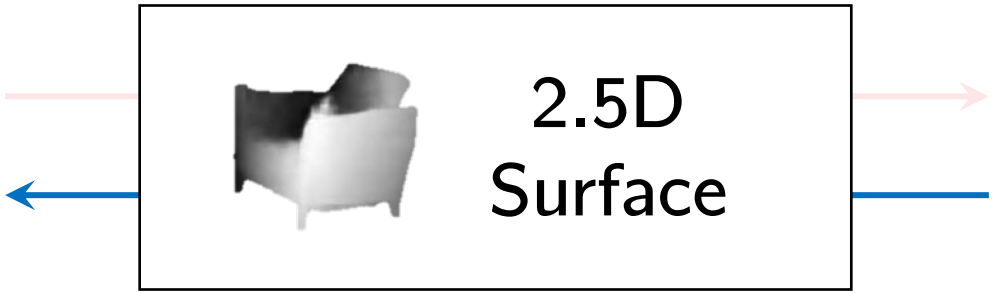
Training: cars, chairs, airplanes
Testing: humans, horses

2D Image



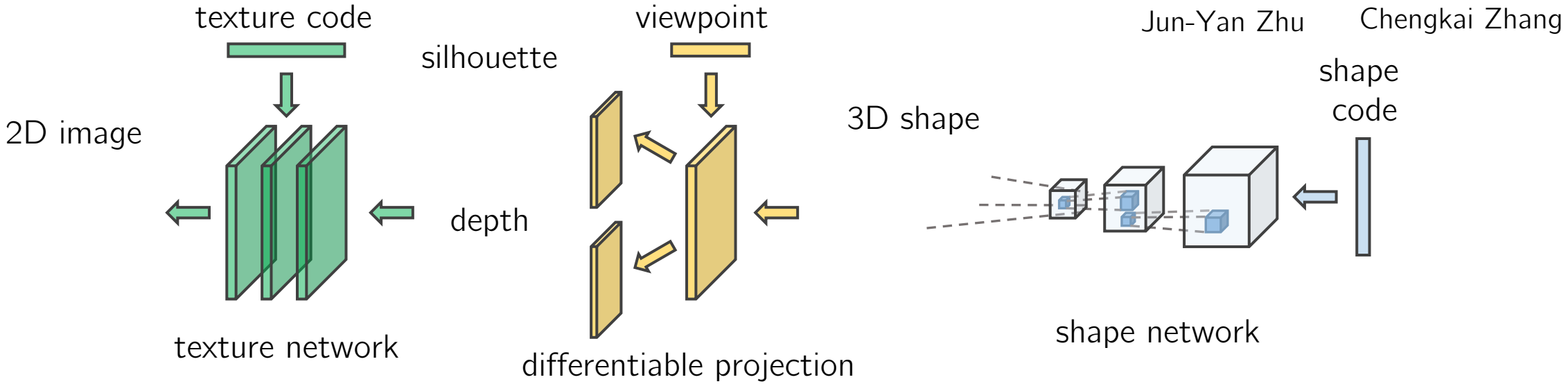
3D Shape

2D Image



3D Shape

Shape and Texture Synthesis



Zhu, Zhang, Zhang, Wu, Torralba, Tenenbaum, Freeman. NeurIPS'18

Wu*, Zhang*, Xue, Freeman, Tenenbaum. NeurIPS'16

Editing viewpoint, shape, and texture

viewpoint

shape

texture

viewpoint

shape

texture

Interpolation in the latent space

shape

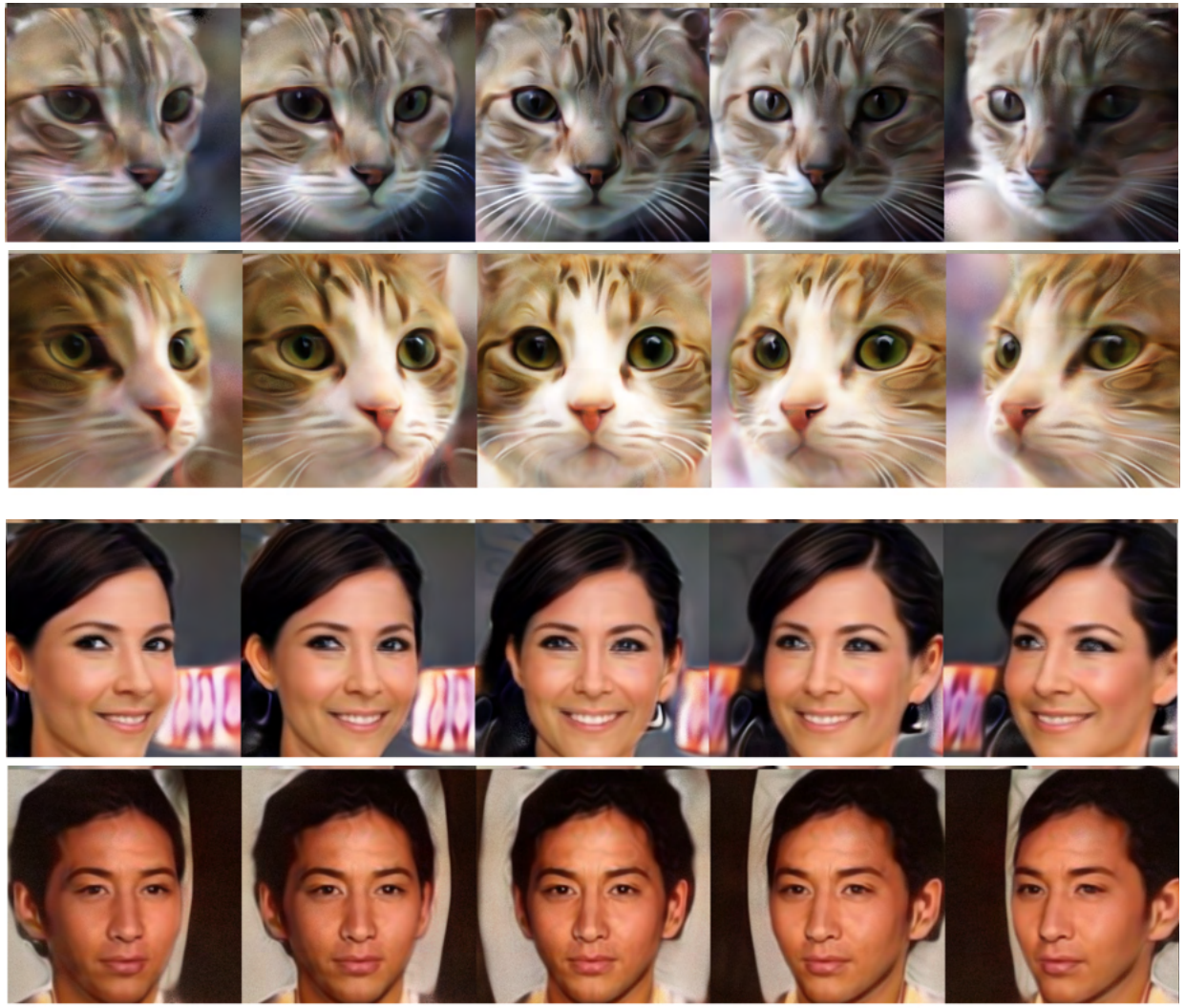
texture

both

Transferring shape and texture

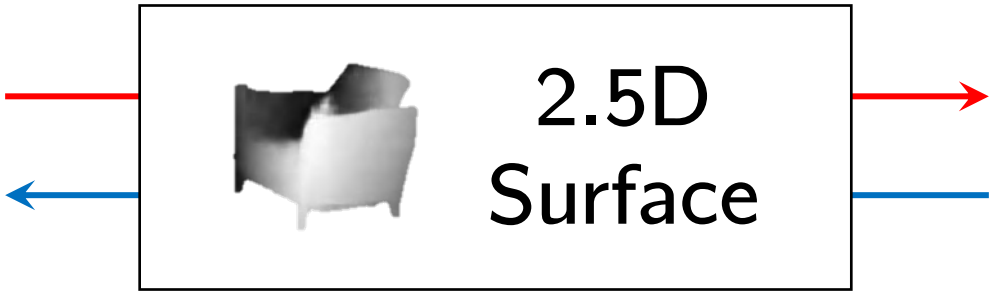
Zhu, Zhang, Zhang, Wu, Torralba, Tenenbaum, Freeman. NeurIPS'18

π-GAN: Implicit Representations



Chan, Monteiro, Kellnhofer, Wu, Wetzstein. CVPR 2021

2D Image

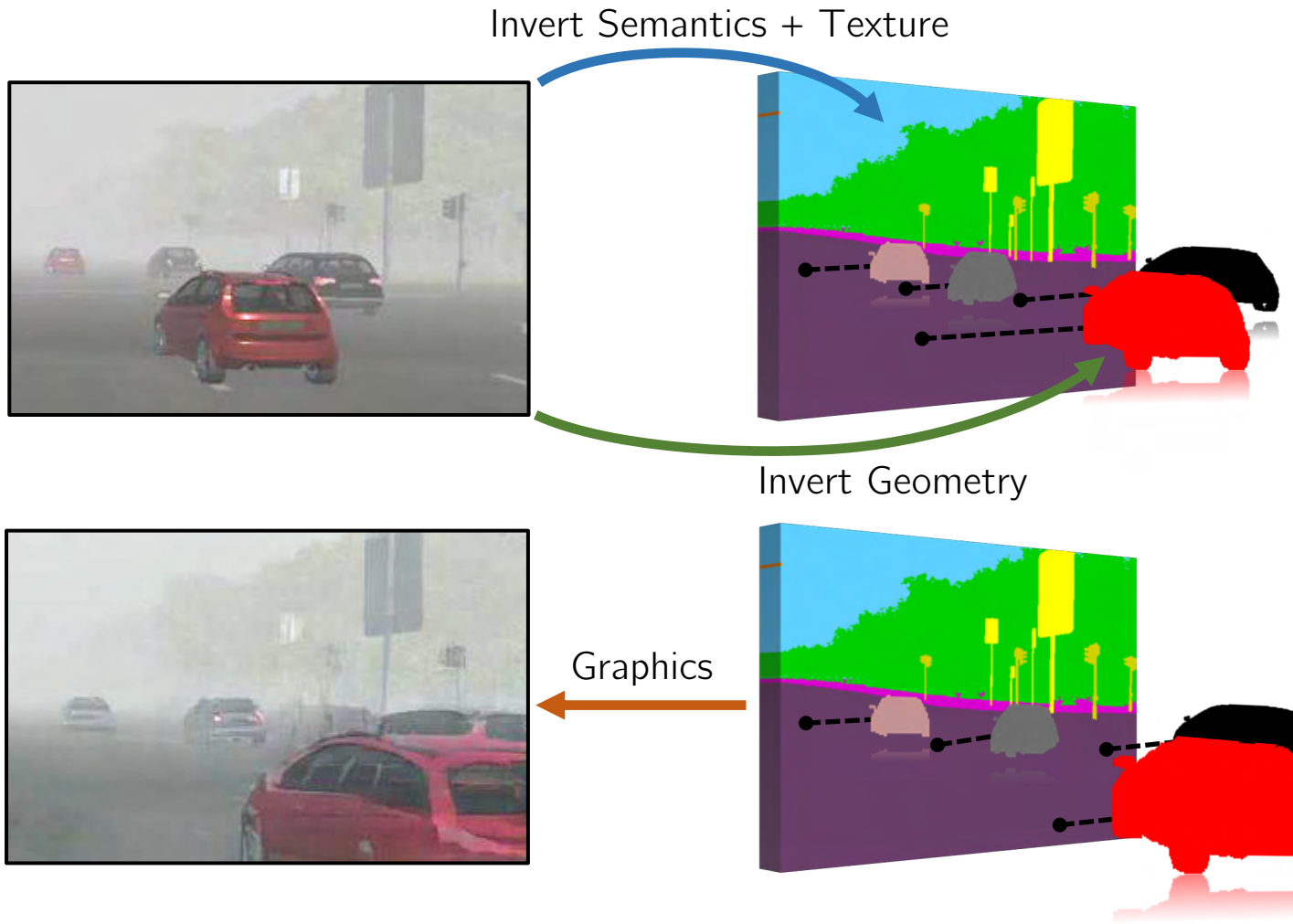


3D Shape

Extension to Scenes

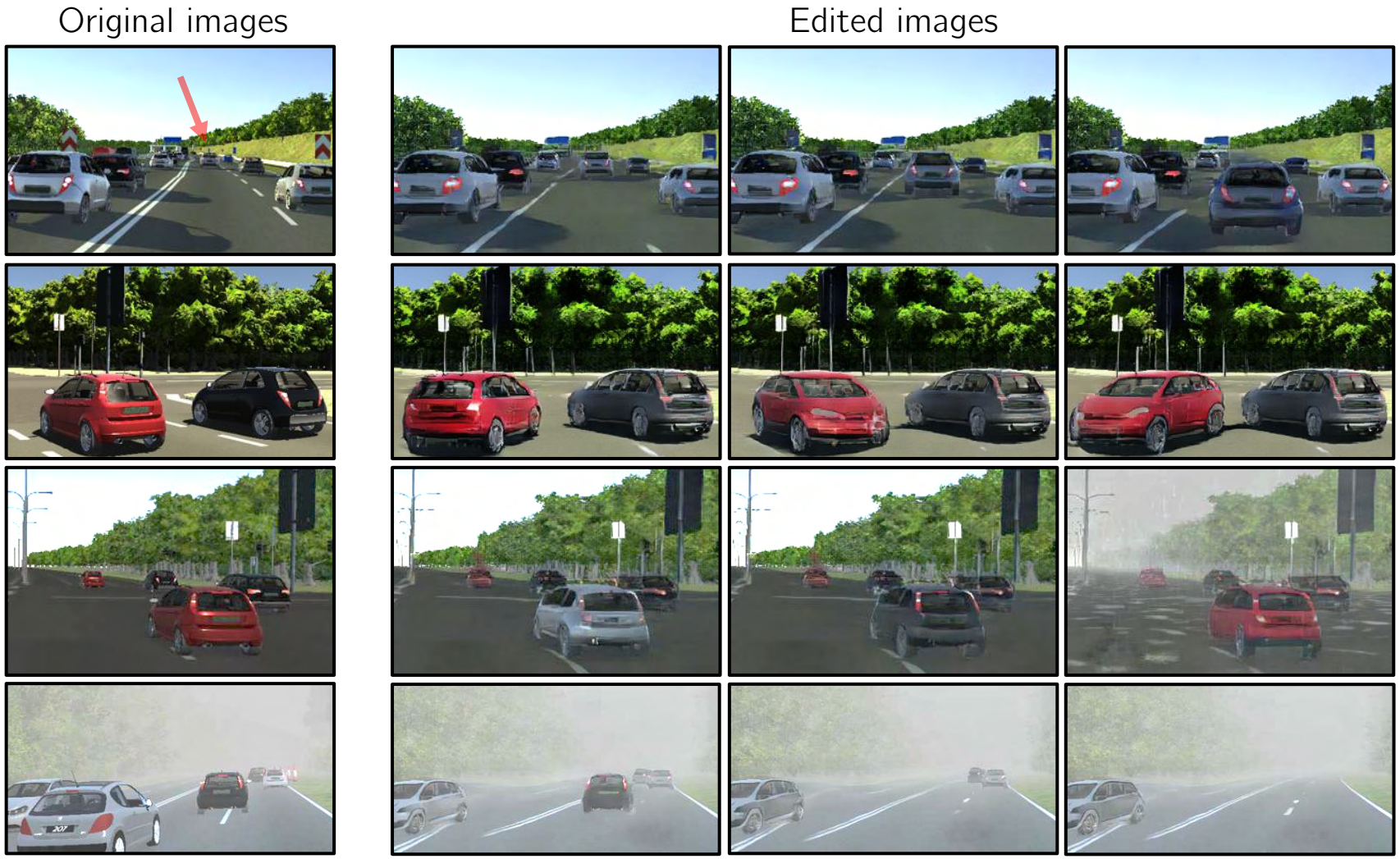

Shunyu Yao


Harry Hsu



Yao*, Hsu*, Zhu, Wu, Torralba, Freeman, Tenenbaum. NeurIPS'18

Image Editing on Virtual KITTI



Yao*, Hsu*, Zhu, Wu, Torralba, Freeman, Tenenbaum. NeurIPS'18

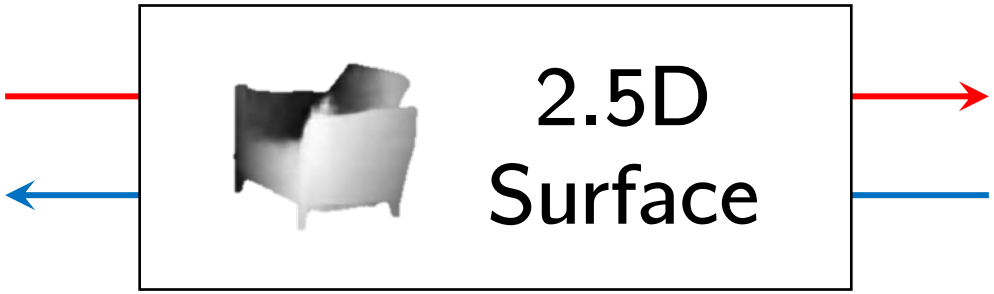
Image Editing on CityScapes (Real Images)

Original images

Edited images

Yao*, Hsu*, Zhu, Wu, Torralba, Freeman, Tenenbaum. NeurIPS'18

2D Image

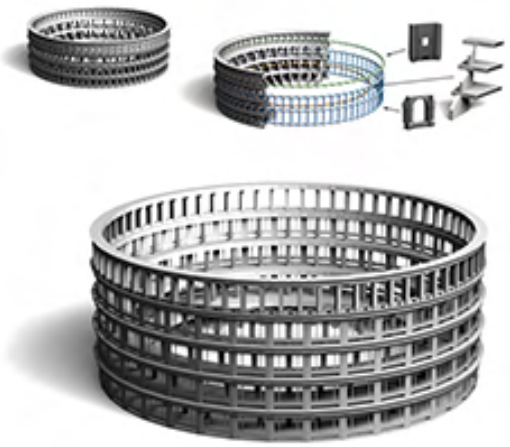


3D Shape

L-Systems

start: X
rules: (X → F+[[X]-X]-F[-FX]+X)
(F → FF)

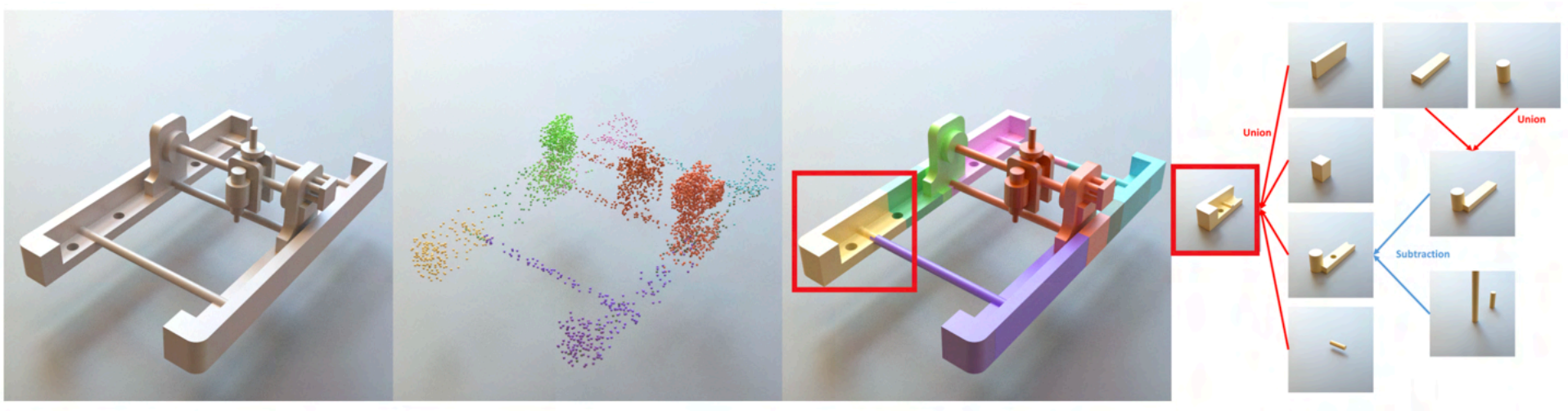
PCFG, Symmetry



Mitra and Pauly. 2008

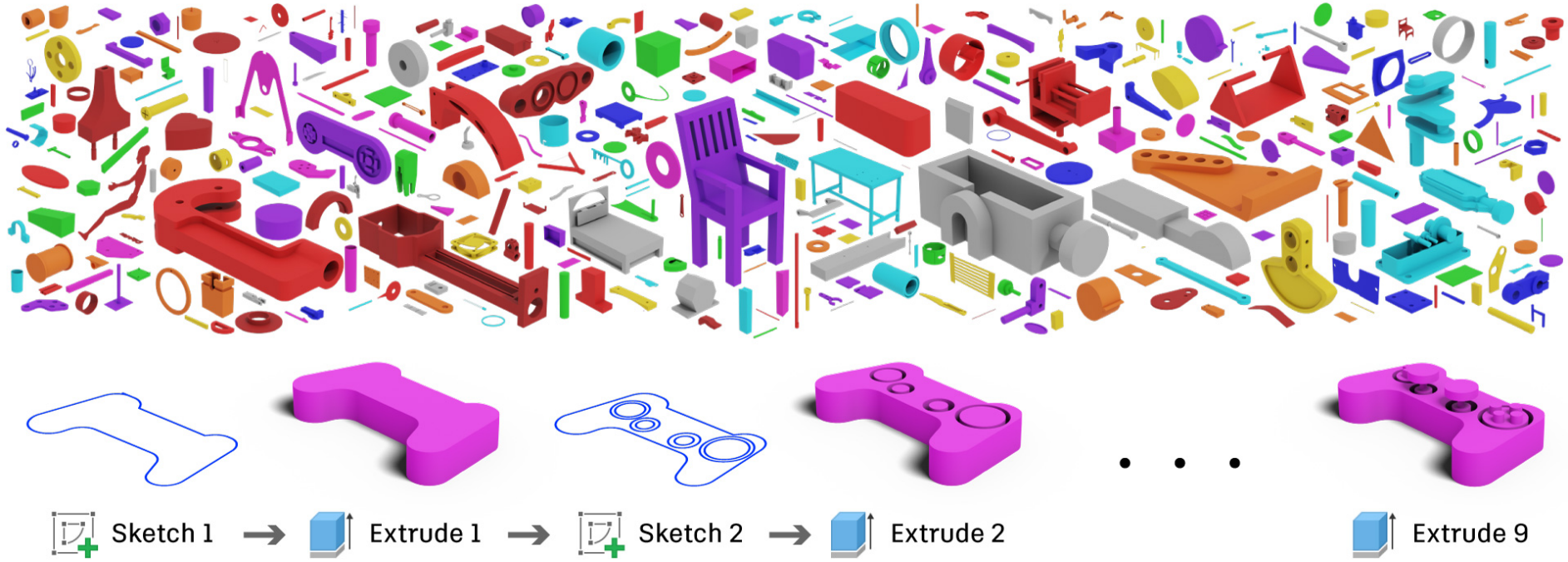
InverseCSG

Du et al. SIGAsia 2018

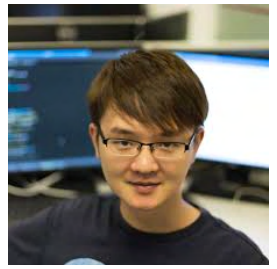


Fusion 360 Gallery

Willis et al. arXiv 2020



From 3D Reconstruction to Abstraction



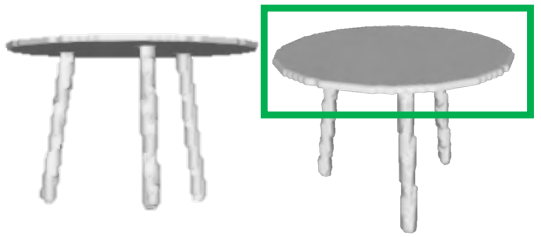
Yonglong Tian

Input Image (Table)

Direct Prediction

GenRe Without Spherical Maps

GenRe



GenRe + Shape Programs

Shapes often have abstract, program-like structure.

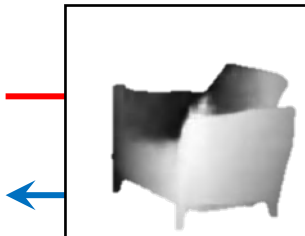
```
Draw("Top", "Circle")

For(i < 2, "trans")
  For(j < 2, "trans")
    Draw("Cuboid")

For(i < 2, "trans")
  Draw("Rect")
```

Shape domain-specific language (Shape DSL)

Program	→	Statement; Program
Statement	→	Draw(Semantics, Shape, Position_Params, Geometry_Params)
Statement	→	For(For_Params); Program; EndFor
Semantics	→	semantics 1 | semantics 2 | semantics 3 | ...
Shape	→	Cuboid | Cylinder | Rectangle | Circle | Line | ...
Position_Params	→	(x, y, z)
Geometry_Params	→	$(g_1, g_2, g_3, g_4, ...)$
For_Params	→	Translation_Params | Rotation_Params
Translation_Params	→	(times i, orientation u)
Rotation_Params	→	(times i, angle θ, axis a)

Tian, Luo, Sun, Ellis, Freeman, Tenenbaum, Wu. ICLR'19

2D Image



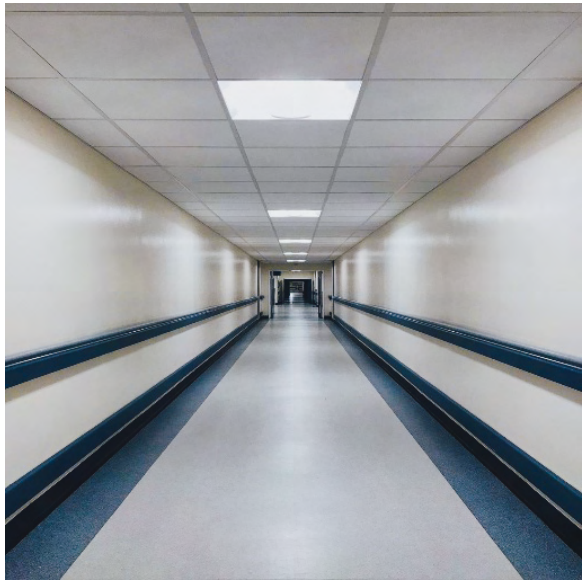
2.5D Surface

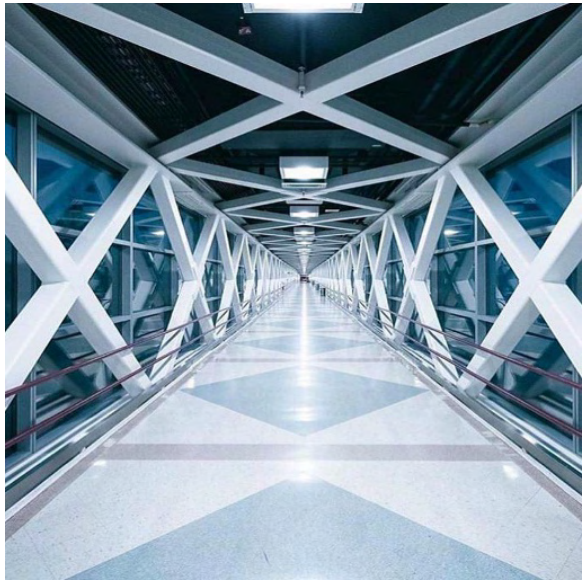
3D Shape

```
draw(Back, Cuboid)

for i in range(0, 1):
    draw(Side, Cuboid, i)

draw(Bottom, Cuboid)
```

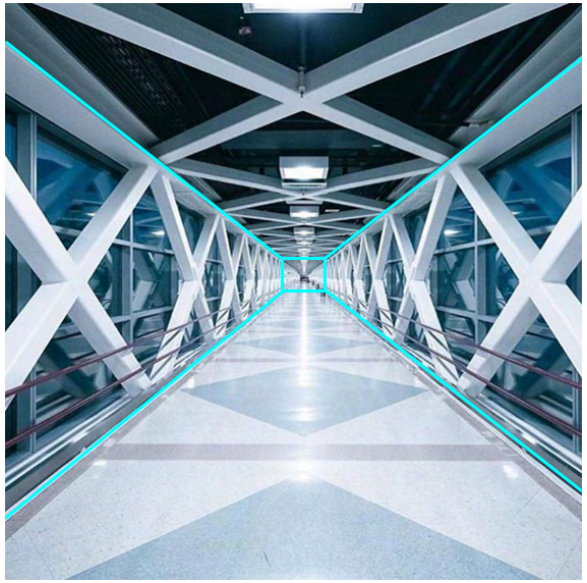
Programs

Rich Structure in 3D Scenes

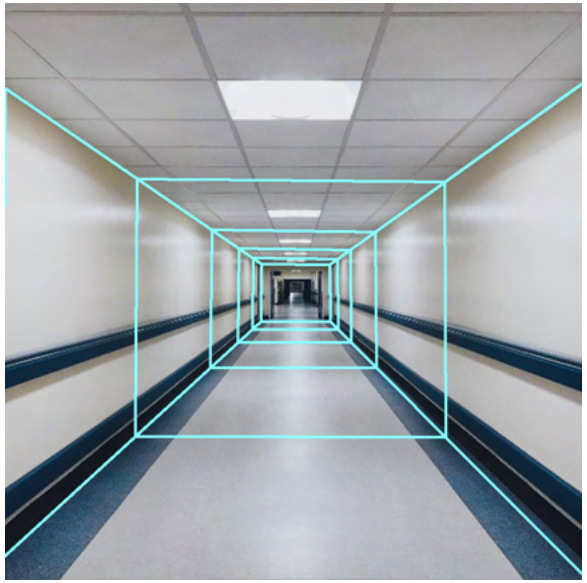


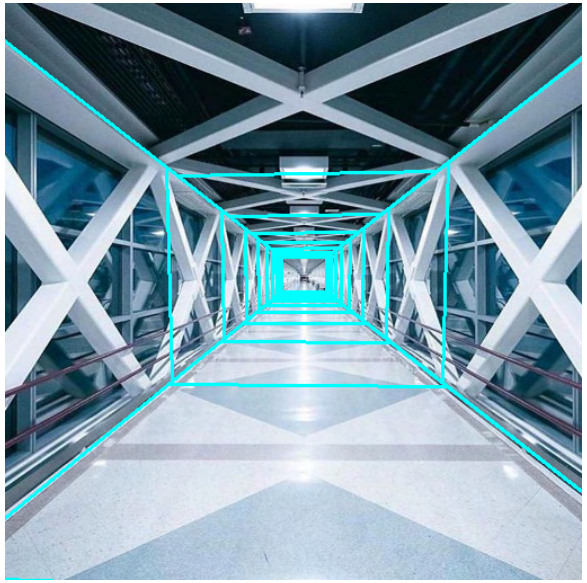


Planes/Surfaces, Symmetry



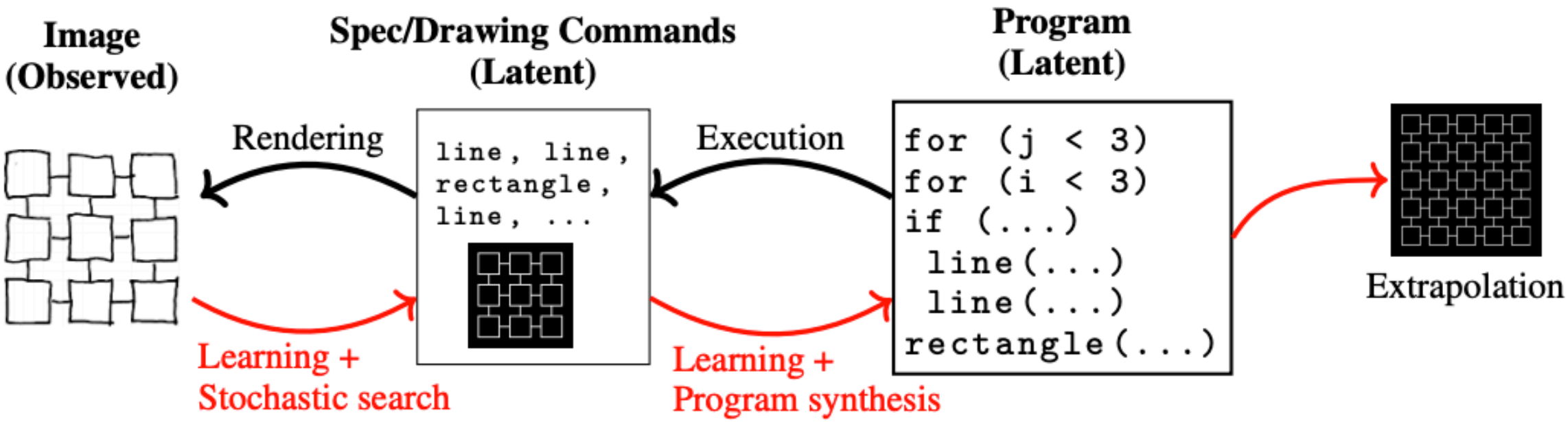
Repetition



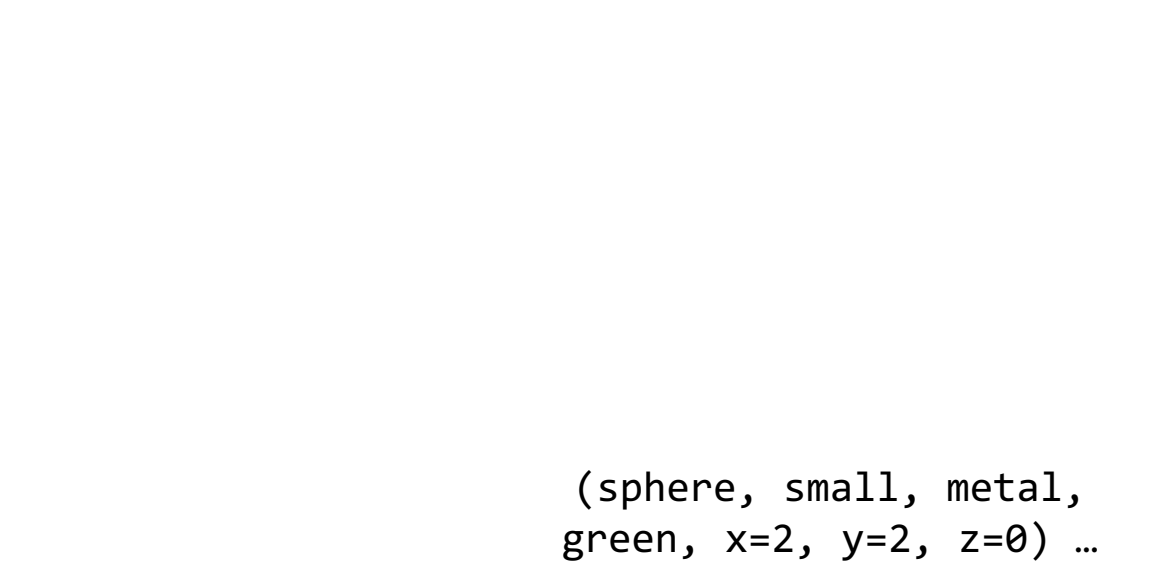


How can we leverage such structural information for 3D scene understanding and editing?

Program Synthesis for Visual Data

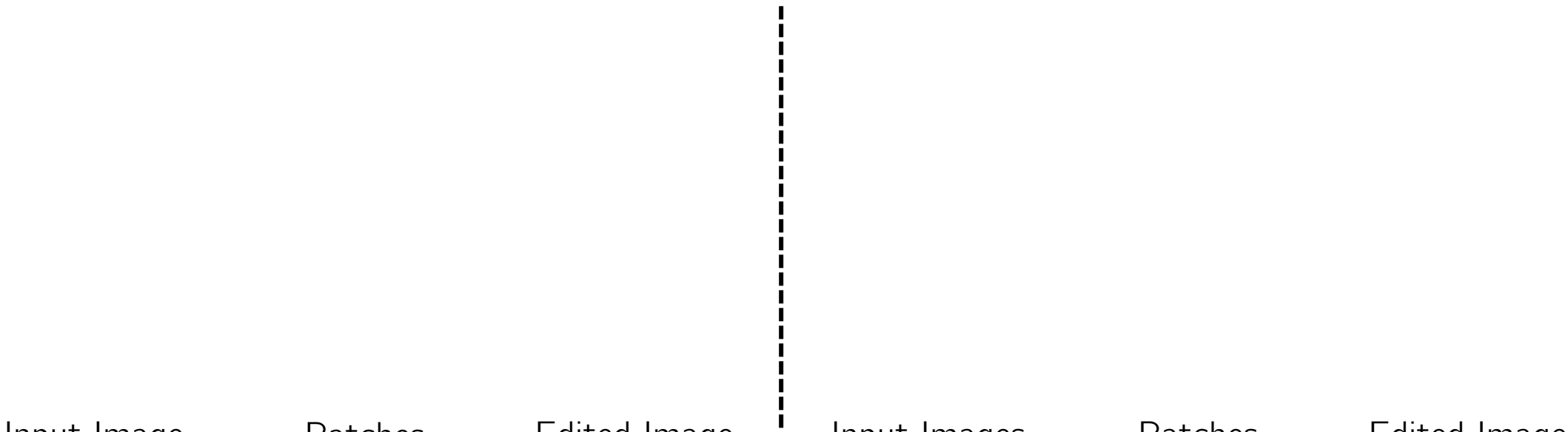


[On 2D Line Drawings] Ellis, Ritchie, Solar-Lezama, Tenenbaum. NeurIPS 2018




```
for i in range(0, 5):
  for j in range(0, 3 - i):
    sphere(pos = (1 + i, 1 + j, 0),
           color = 6 - j)
for i in range(0, 4):
  cylinder(pos = (3, 3, i),
           color = 7)
cylinder(pos = (4, 2, 0),
         color = 3)
```

(a) Input Image (b) Object & Group Detection (c) Program



[On 3D Primitives] Liu, Wu, Ritchie, Freeman, Tenenbaum, Wu. ICLR 2019

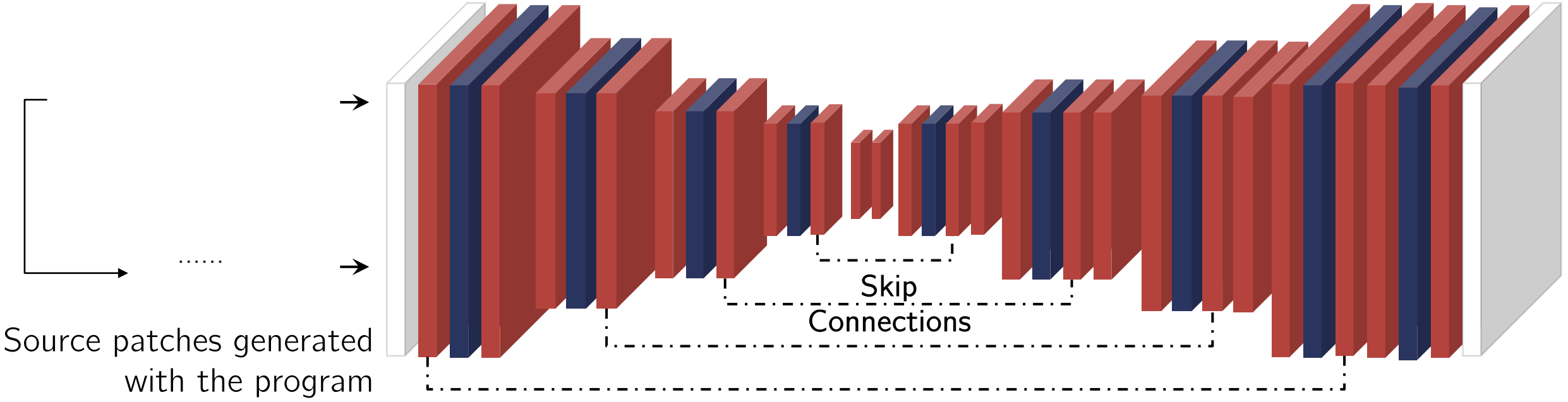
Generalizing to Natural Images

```
for i in range(0, 5):
  for j in range(0, 9):
    if i + j >= 8:
      draw(
        33 + 45 * i + 24 * j,
        6 + 45 * j,

      )
```

Image | Repeated Objects (Centroids) | Program

Key Idea: Internal learning (single-image learning)

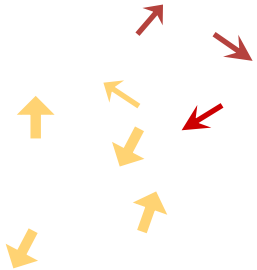
[Internal Statistics] Zontak and Irani. CVPR 2011
[Repeated Pattern Detection] Lettry et al. WACV 2017



Input (Corrupted) Output (Inpainted)

Input (Partial) Output (Extrapolated)

Input (Regular) Output (Irregular)



A. Inpainting

B. Extrapolation

C. Regularity Editing

Mao*, Zhang*, Li, Freeman, Tenenbaum, Wu. ICCV 2019

```
!"# i $% range(0, 5):
  !"# j $% range(0, 9):
    $! i + j >= 8:
      draw(
        33 + 45 * i + 24 * j,
        6 + 45 * j,

      )
```

Image

Repeated Objects (Centroids)

Program

Corrupted

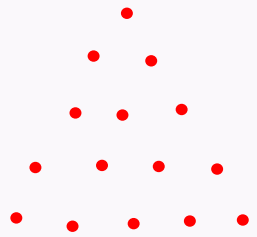
Inpainted

Partial

Extrapolated

Original

Irregular



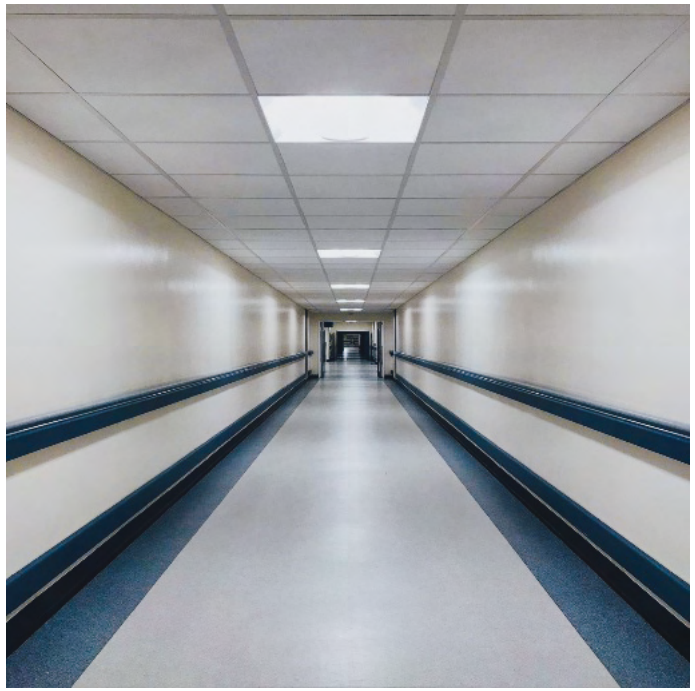
A. Inpainting

B. Extrapolation

C. Regularity Editing

Mao*, Zhang*, Li, Freeman, Tenenbaum, Wu. ICCV 2019

Generalizing to Multiple Planes



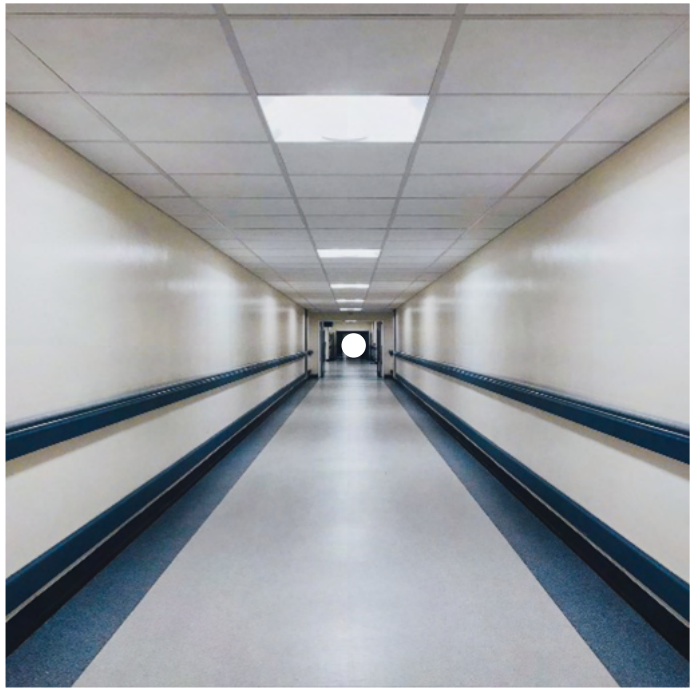
Input Image

→

```
:(;C&?(#&)*+/DE-
6/78&%(/9
:(;78&%()<"=>9+/%"#?&8>9-
!"# $ $% #&%'()*+ ,-./
    0#&1)/////////+/
          2 3 4, 5/$-
6/78&%(/@
:(;78&%()<"=>@+/%"#?&8>@-
!"# $ $% #&%'()*+ ,-./
    0#&1)/////////+/
          2 3 4, 5/$-
6/78&%(/A/BBB
```

Program

Visual Cue Extraction



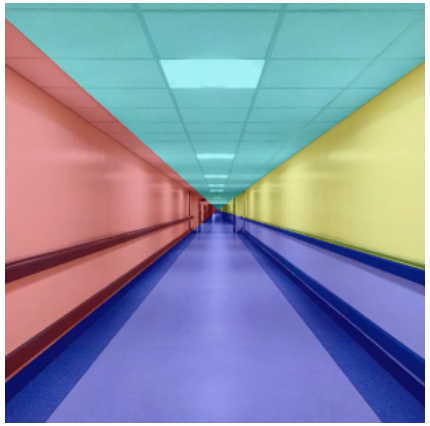
Vanishing Point
(NeurVPS [Zhou et al. 2019])

Wireframes
(L-CNN [Zhou et al. 2019])

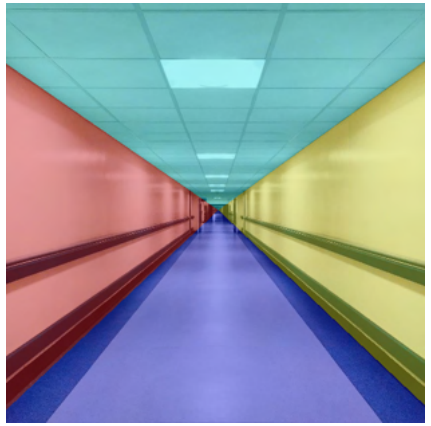
Bottom-up visual cues for program synthesis:

Lines and shapes	for	line drawings
Repeated objects	for	single-plane images
Vanishing point and wireframes	for	multi-plane images

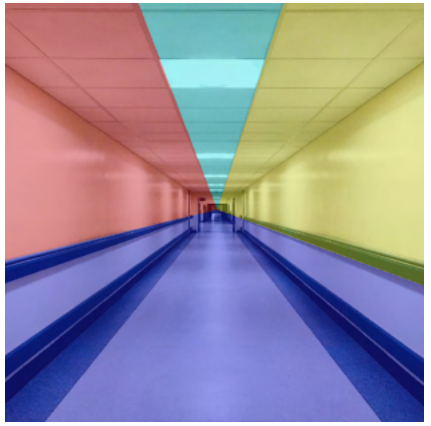
Candidate Plane Partition Generation



Candidate 1



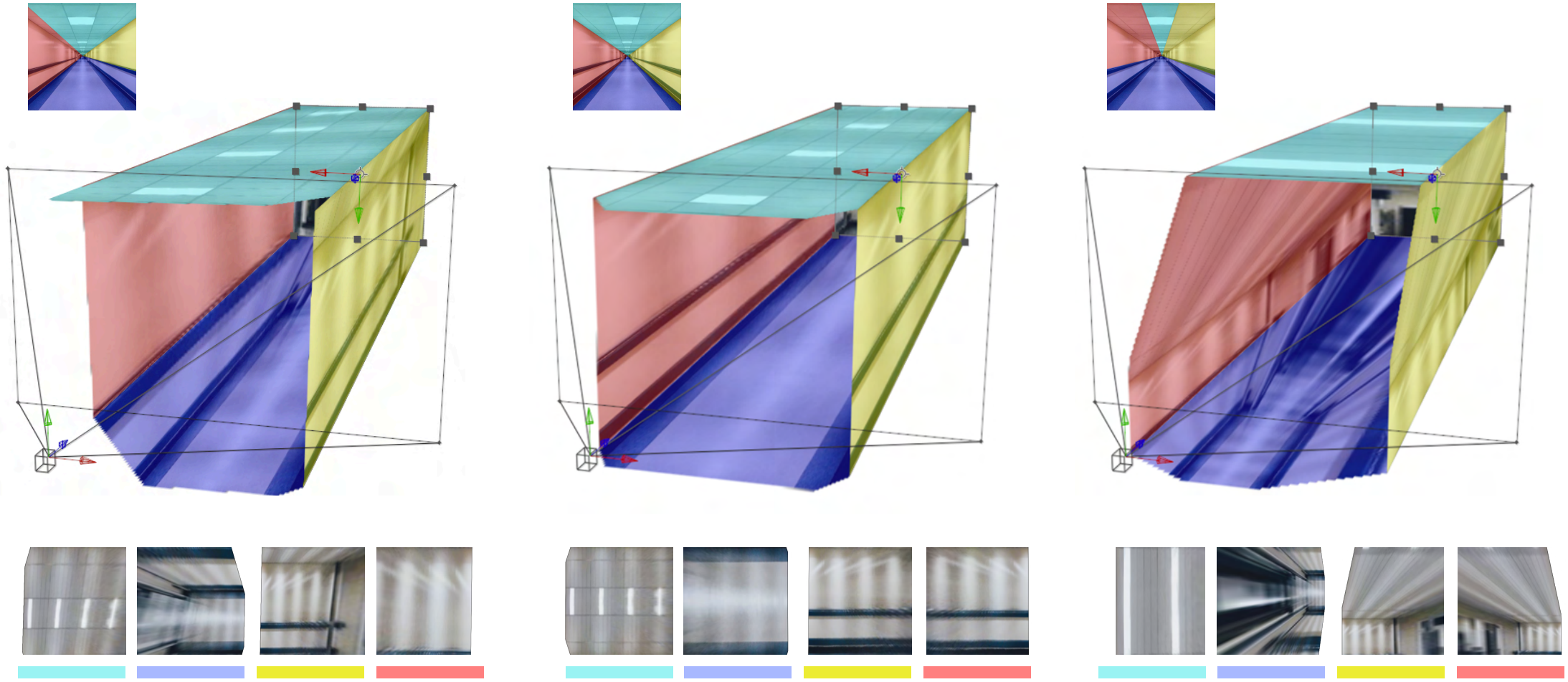
Candidate 2



Candidate 3

Li*, Mao*, Zhang, Freeman, Tenenbaum, Snavely, Wu. NeurIPS 2020

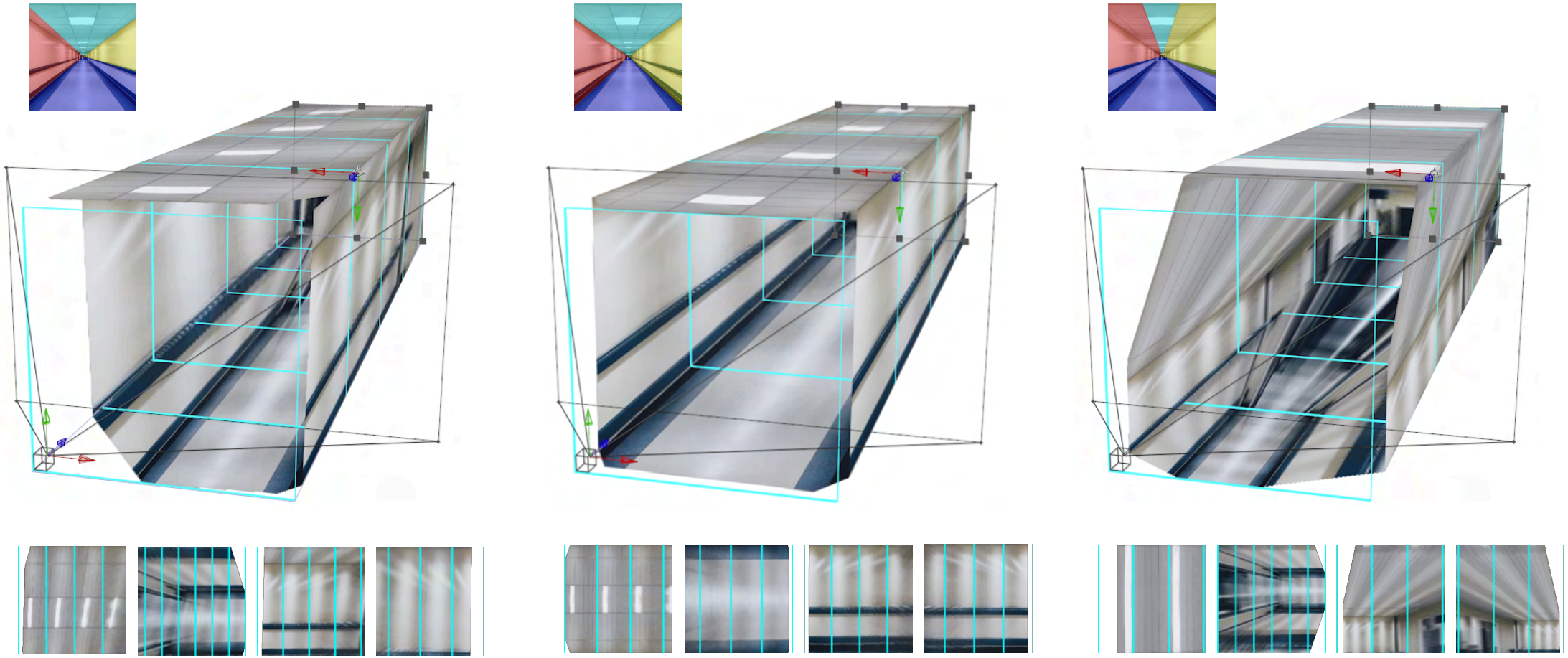
Plane Rectification



Li*, Mao*, Zhang, Freeman, Tenenbaum, Snavely, Wu. NeurIPS 2020

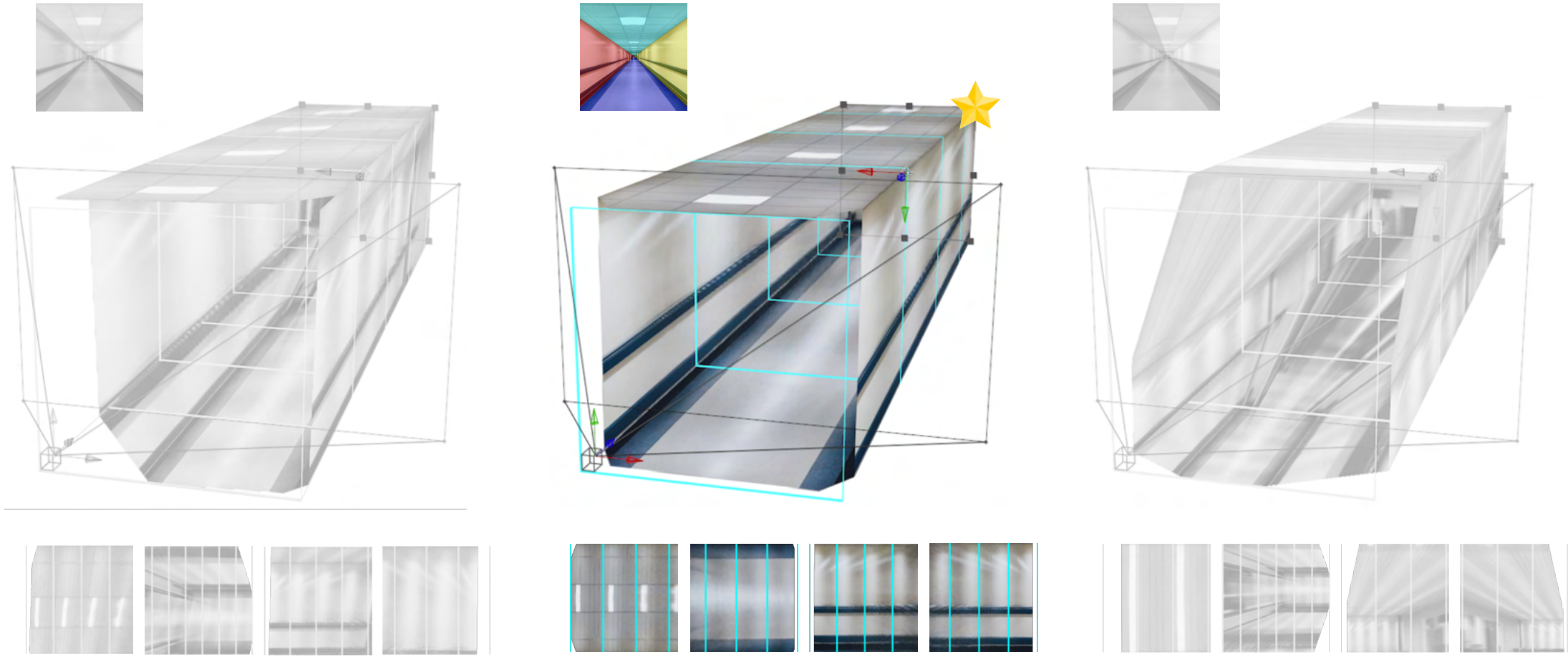
Visual Program Synthesis

Plane partition and program synthesis **help each other**.



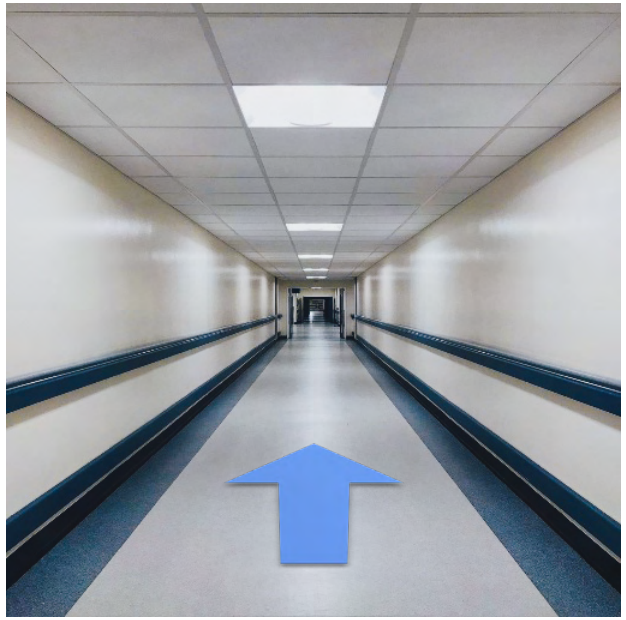
Li*, Mao*, Zhang, Freeman, Tenenbaum, Snavely, Wu. NeurIPS 2020

Visual Program Synthesis

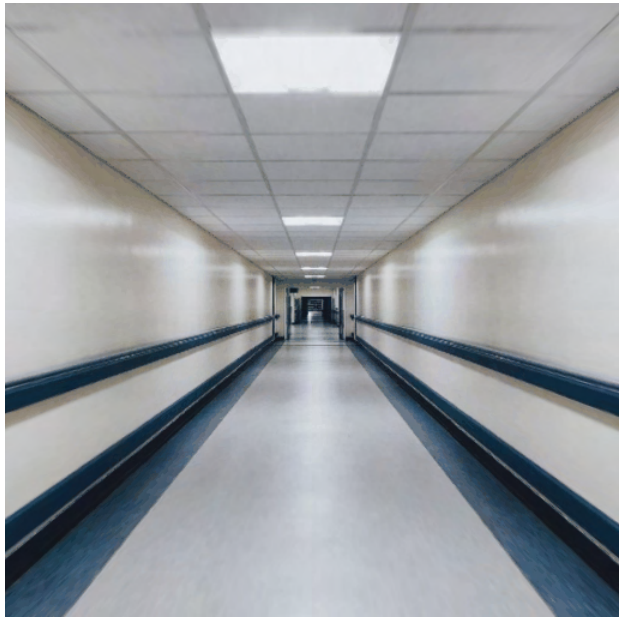


Li*, Mao*, Zhang, Freeman, Tenenbaum, Snavely, Wu. NeurIPS 2020

View Synthesis



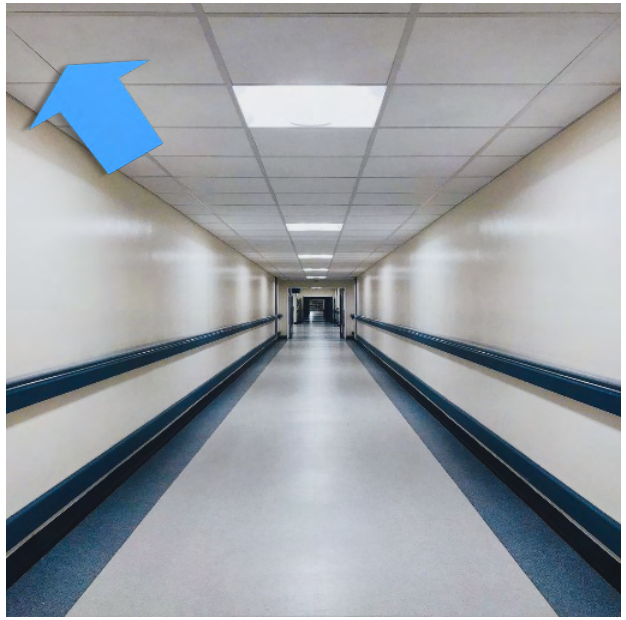
Input Image



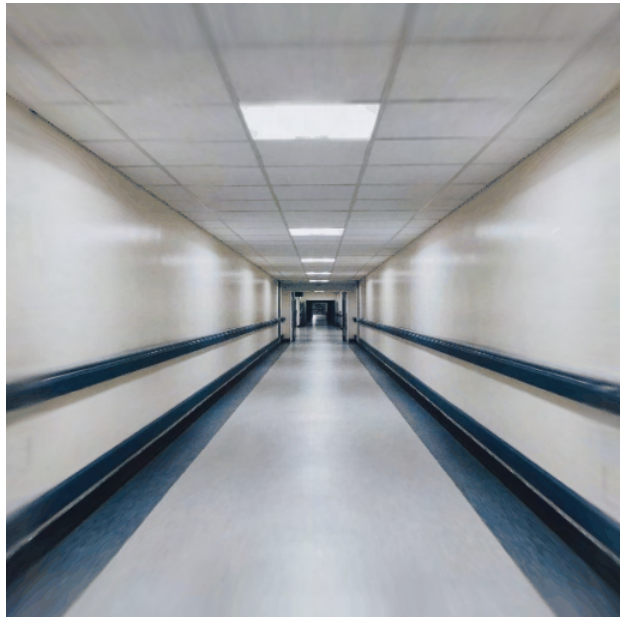
Ours

SynSin [Wiles et al. 2020]

View Synthesis



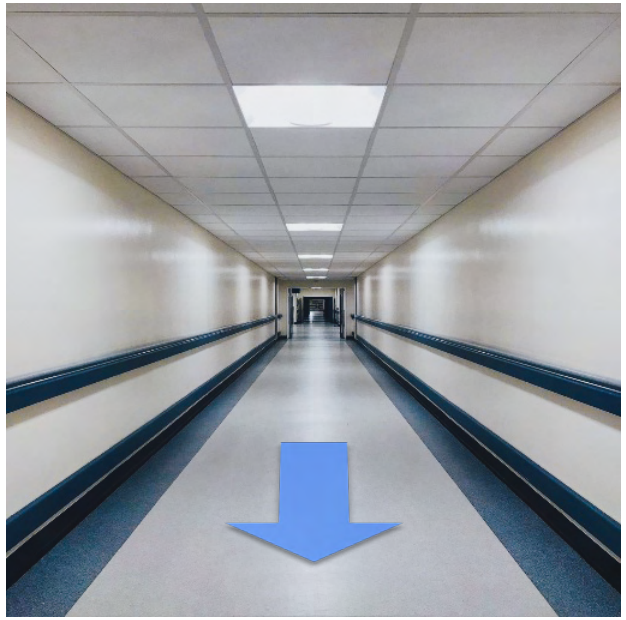
Input Image



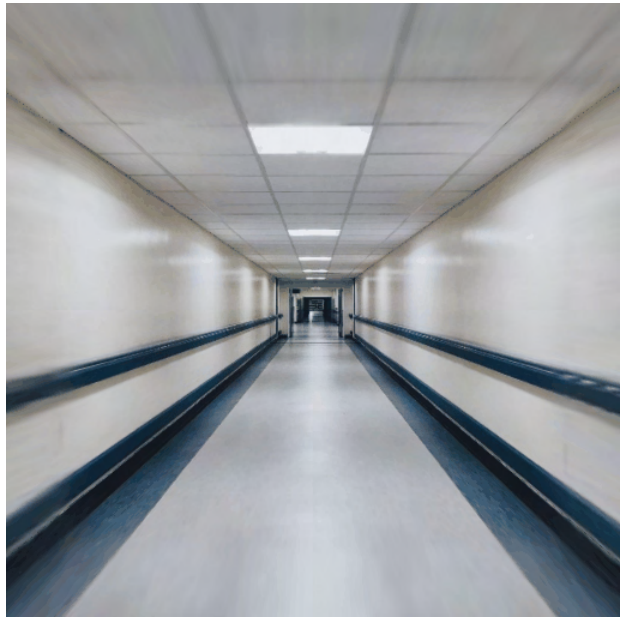
Ours

SynSin [Wiles et al. 2020]

View Synthesis



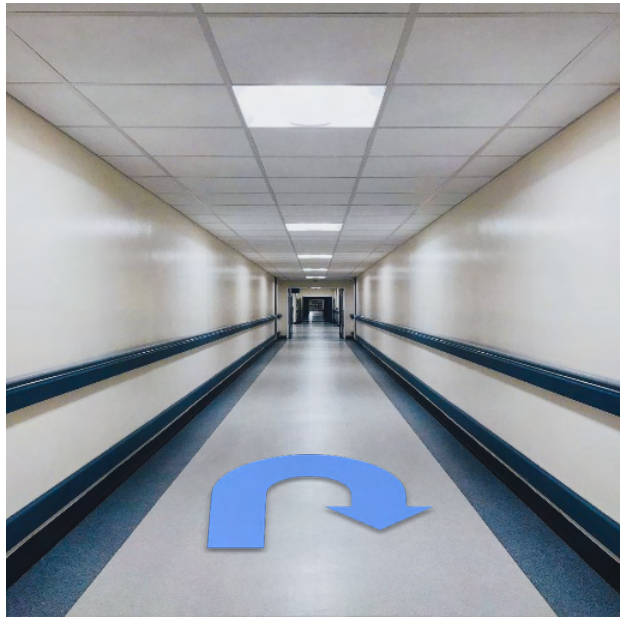
Input Image



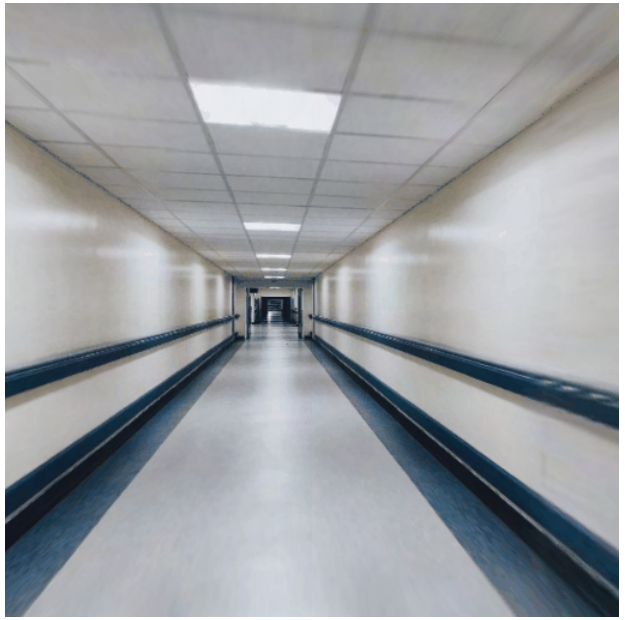
Ours

SynSin [Wiles et al. 2020]

View Synthesis



Input Image



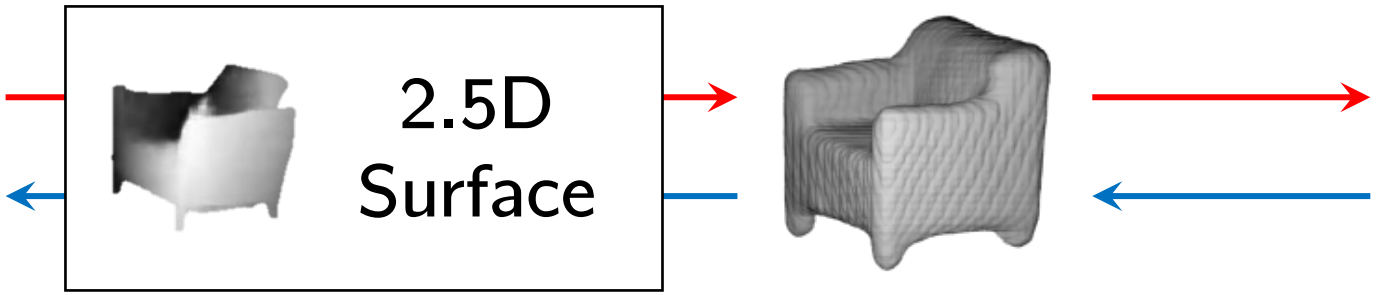
Ours

SynSin [Wiles et al. 2020]

Image Extrapolation

Jiayuan Mao Yikai Li

Li*, Mao*, Zhang, Freeman, Tenenbaum, Snavely, Wu. NeurIPS 2020




```
draw(Back, Cuboid)

for i in range(0, 1):
    draw(Side, Cuboid, i)

draw(Bottom, Cuboid)
```

2D Image

3D Shape

Programs

Key innovations:

- Neural nets for recognition +
- Domain knowledge for generalization: surfaces, objects, procedures

Takeaways:

- Think about the causal structure behind visual data.
- Model what we can, learn what we can't.