Computer Vision

3D understanding

Tae-Hyun Oh (오태현)
전자전기공학과
POSTECH

Slide by Jongha Kim (김종하)

TAs: {Dongmin Choi , Jongha Kim, Juyong Lee, Sungbin Kim} (in alphabetic order)

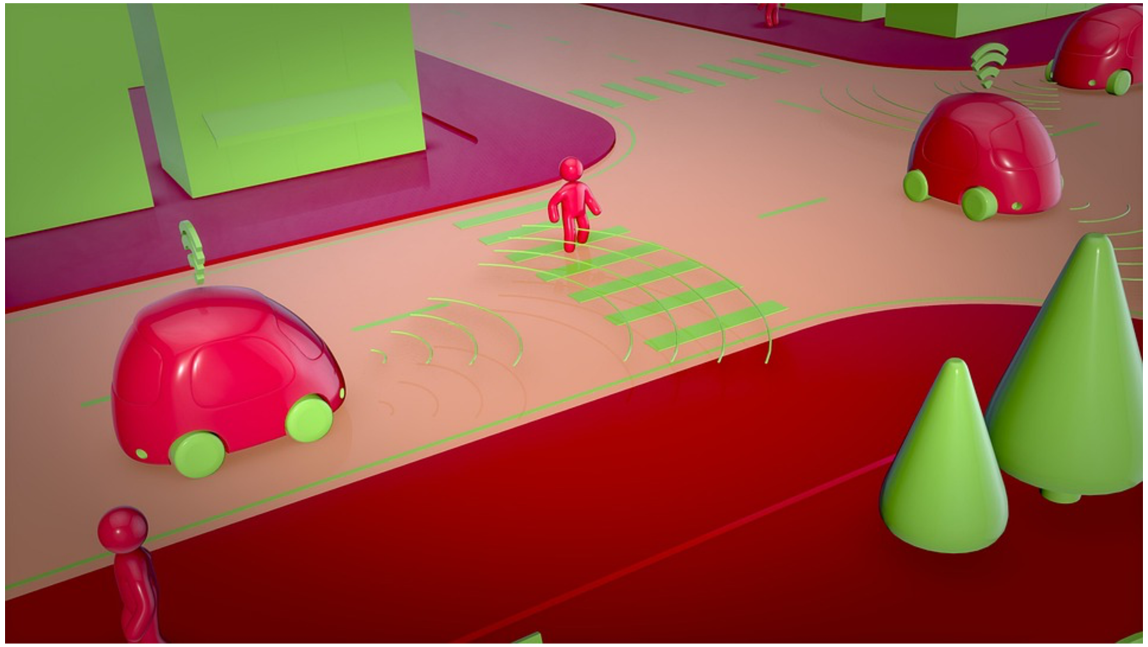
1.

Seeing the world in 3D perspective

1.1 Why is 3D important?

We live in a 3D space

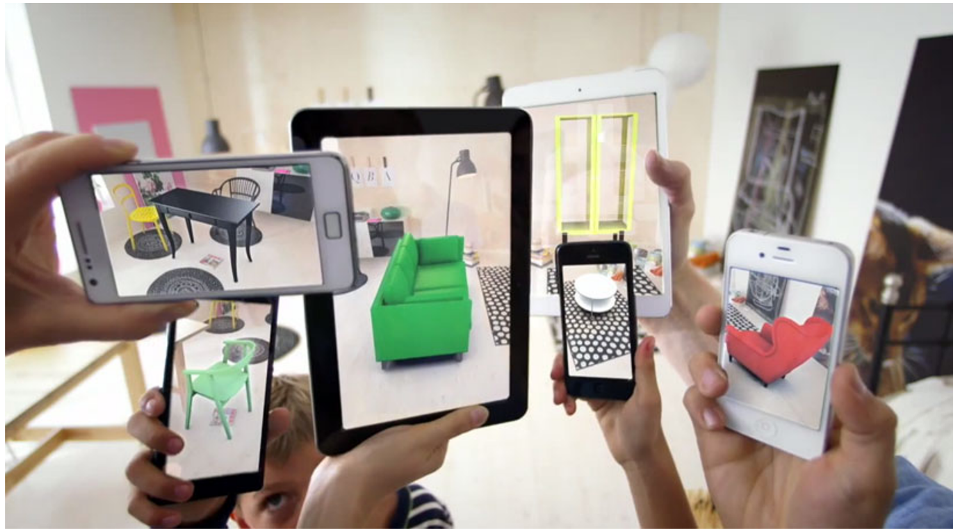
- AI agents operate in the real world, which is a 3D space



For AI driving agents, understanding the 3D space is crucial!

1.1 Why is 3D important?

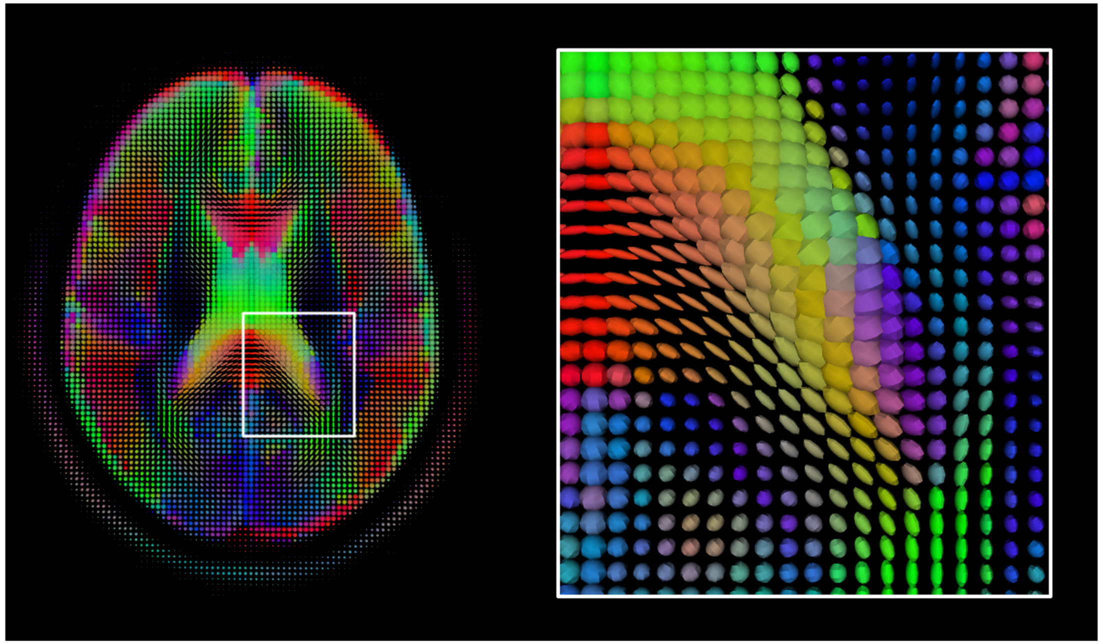
3D applications – AR/VR



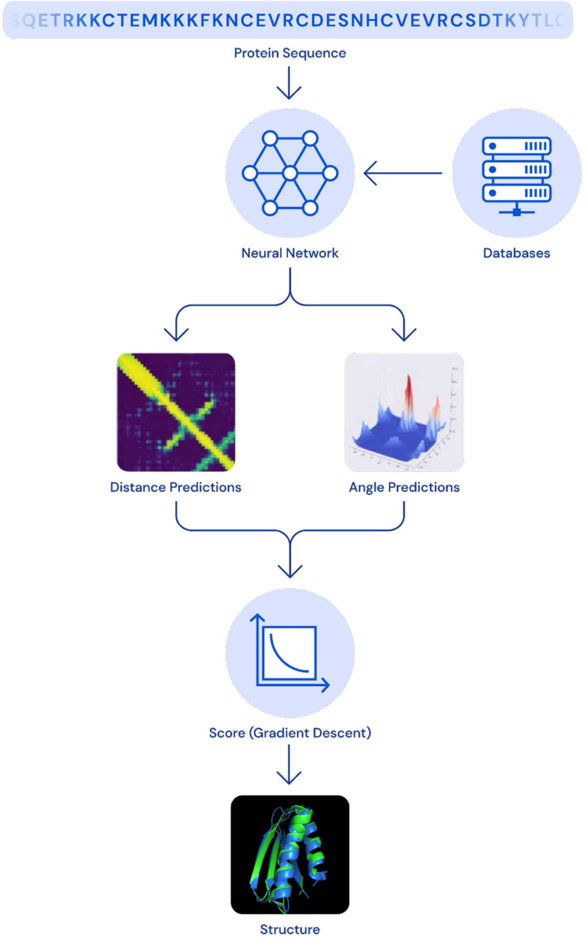
1.1 Why is 3D important?

3D applications – 3D printing

3D applications – Medical applications



Neuroimaging

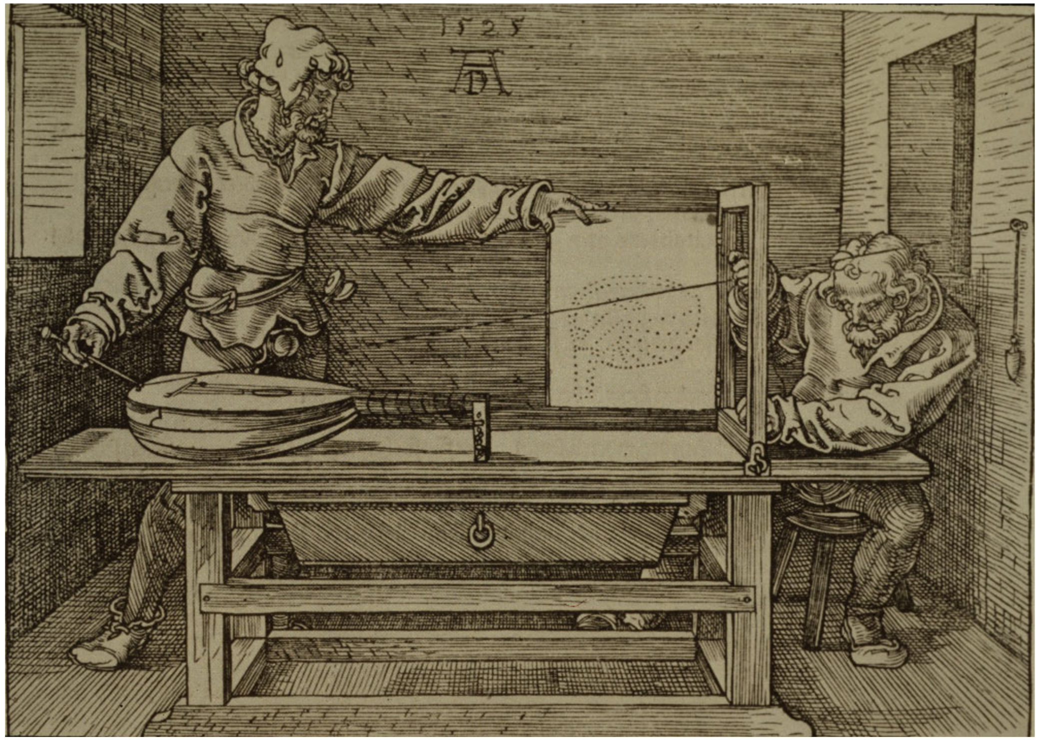


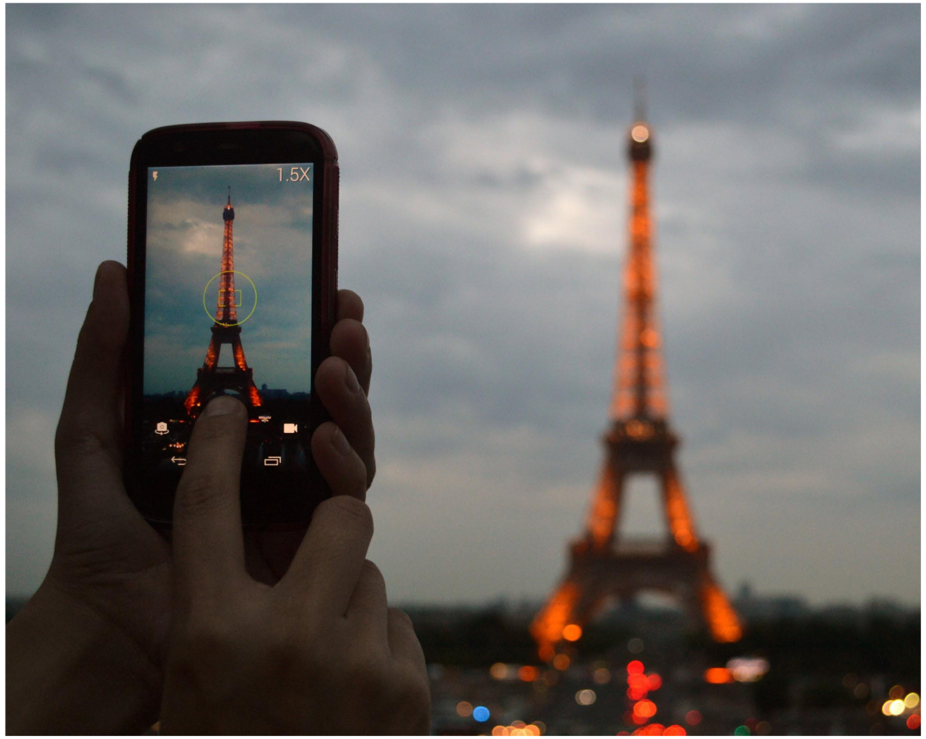
Protein sequence analysis & synthesis

[https://deepmind.com/blog/article/AlphaFold-Using-AI-for-scientific-discovery]

1.2 The way we observe 3D

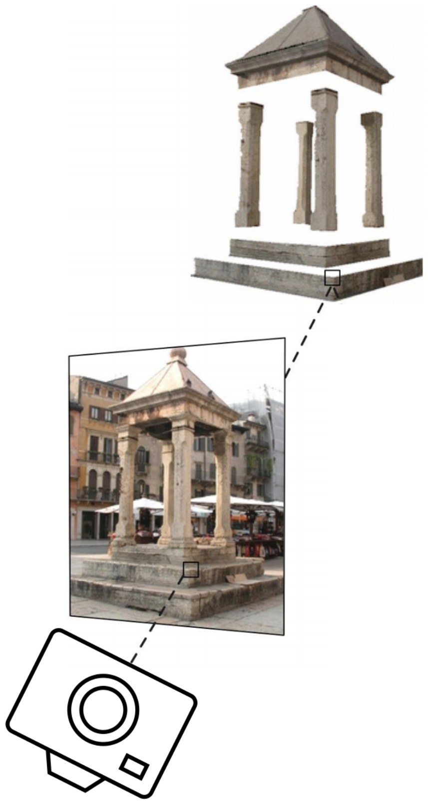
An image is a projection of the 3D world onto a 2D space





1.2 The way we observe 3D

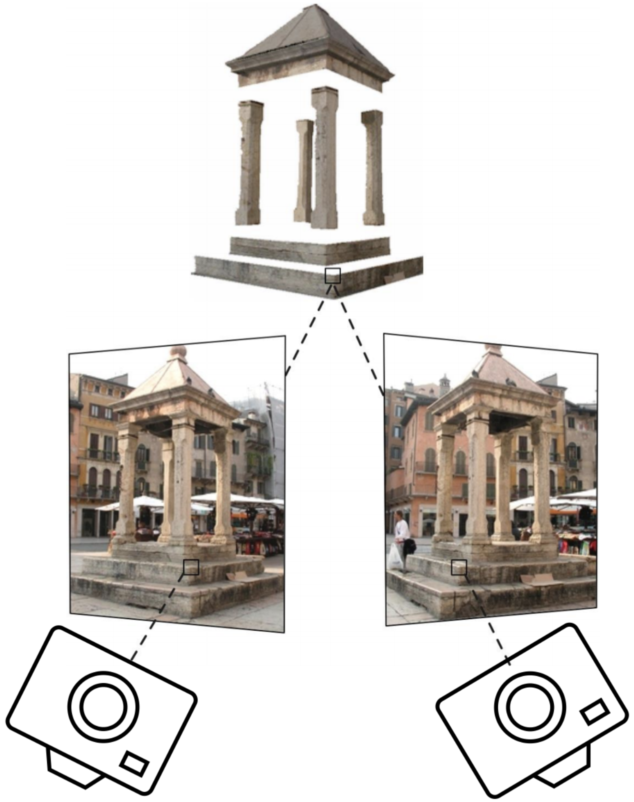
The camera is a projection device of the 3D scene onto a 2D image plane



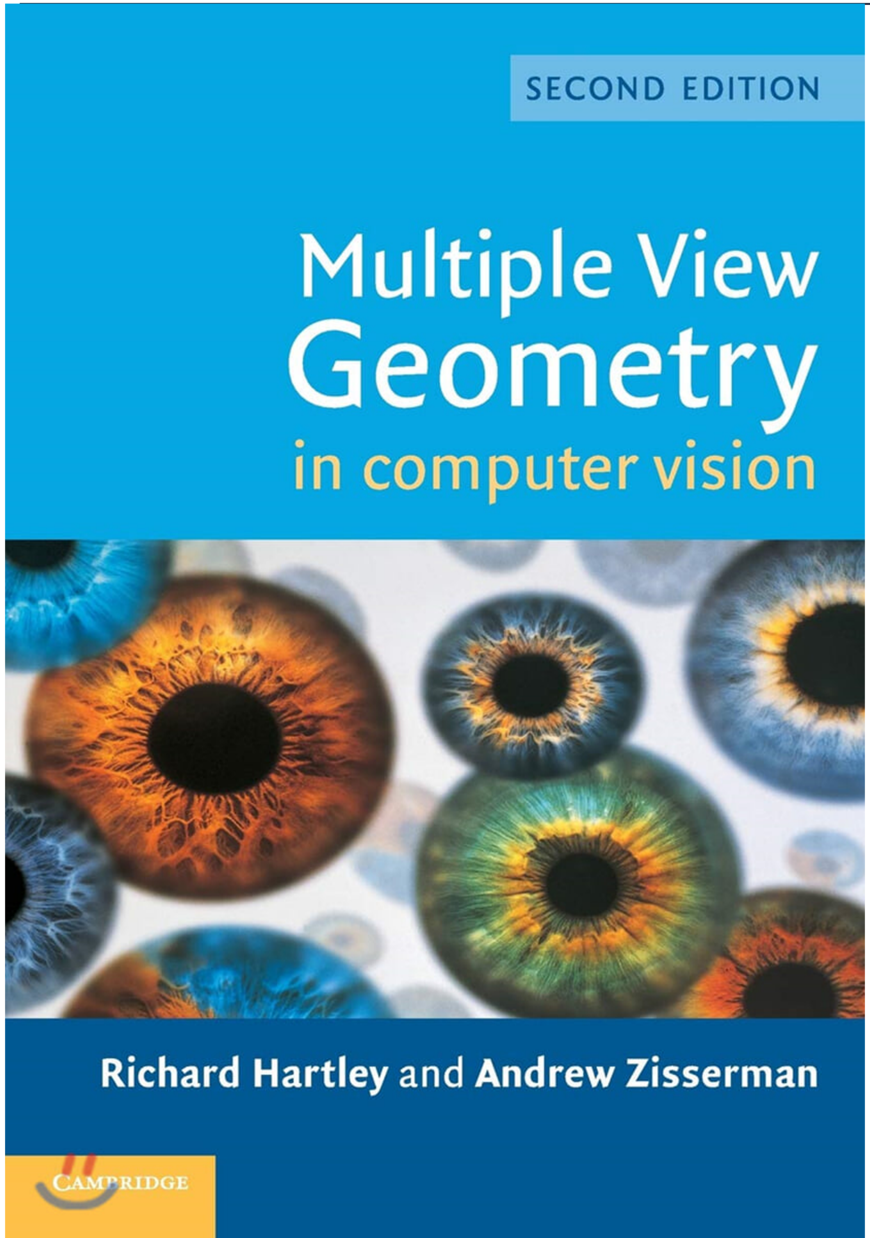
1.2 The way we observe 3D

Triangulation – The way to obtain a 3D point from 2D images

Two-view geometry



1.2 The way we observe 3D

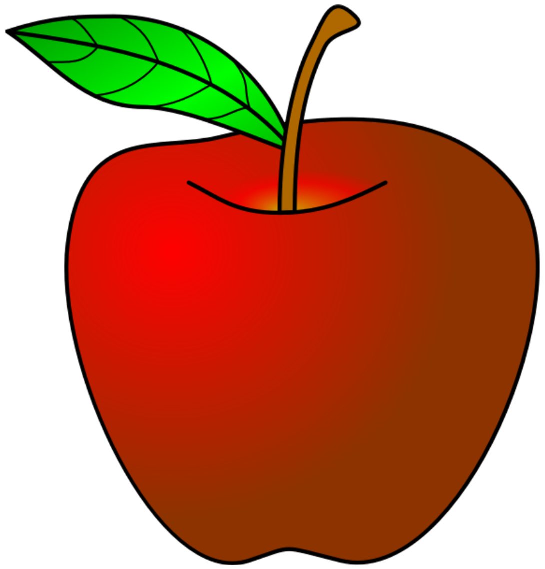


If you are interested in more details about 3D,
check this Bible.
(PDF available online)

1.3 3D data representation

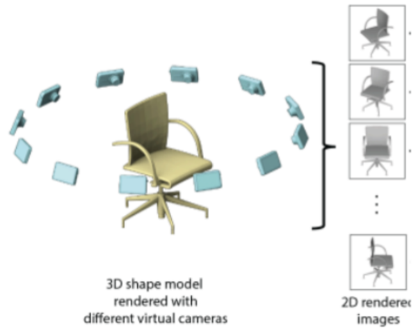
How is 3D data represented in computer?

- A 2D image is represented by RGB values of each pixel in 2D array structure

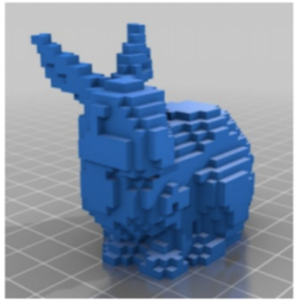


255	3	143
33	123	52
111	0	83

3D data representation is not unique

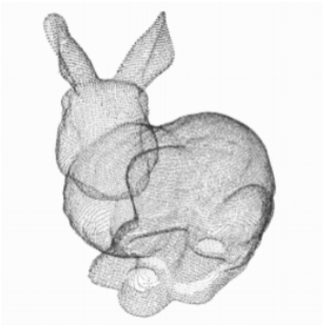


Multi-view images

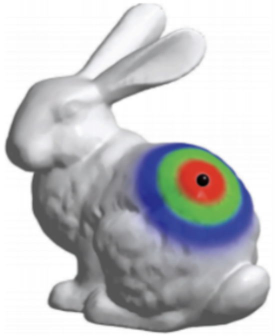


Volumetric (voxel)

Part assembly



Point cloud



Mesh (Graph CNN)

$$F(x) = 0$$

Implicit shape

1.4 3D datasets

ShapeNet

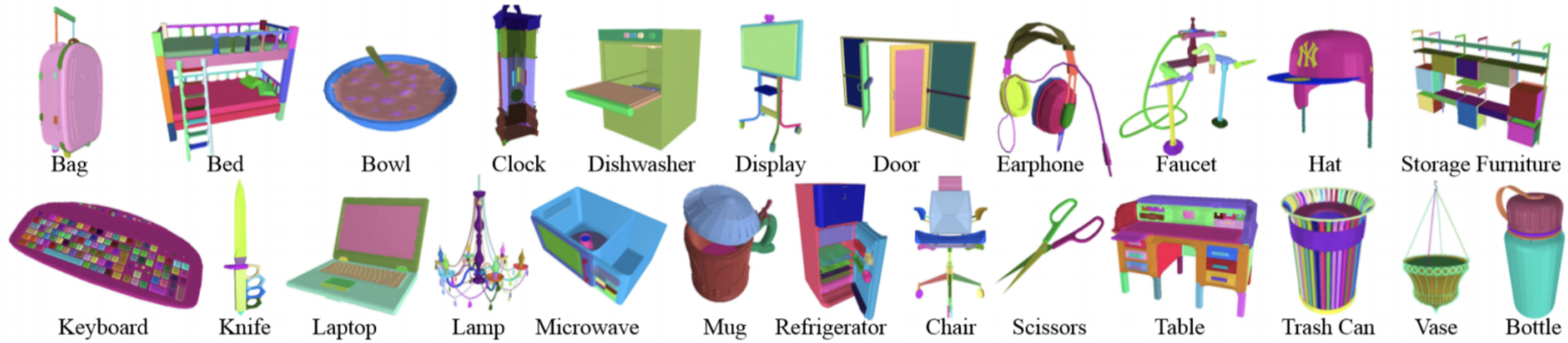
[Chang et al., ArXiv 2015]

- Large scale synthetic objects (51,300 3D models with 55 categories)

PartNet (ShapeNetPart2019)

[Mo et al., CVPR 2019]

- Fine-grained dataset, useful for segmentation (573,585 part instances in 26,671 3D models)



1.4 3D datasets

SceneNet

[McCormac et al., ICCV 2017]

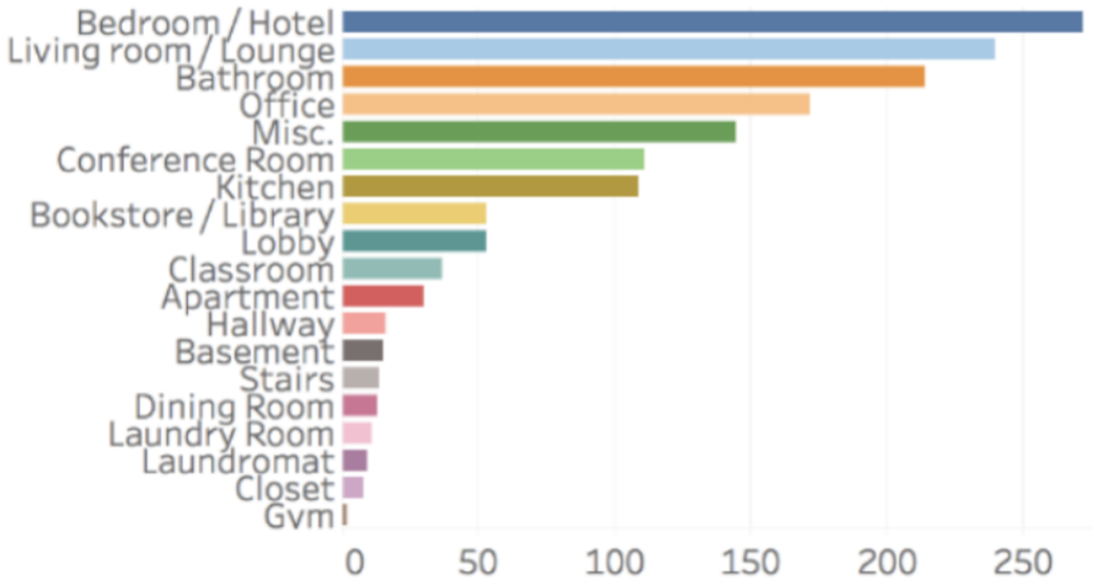
- 5 million RGB-Depth synthetic indoor images

1.4 3D datasets

ScanNet

[Dai et al., CVPR 2017]

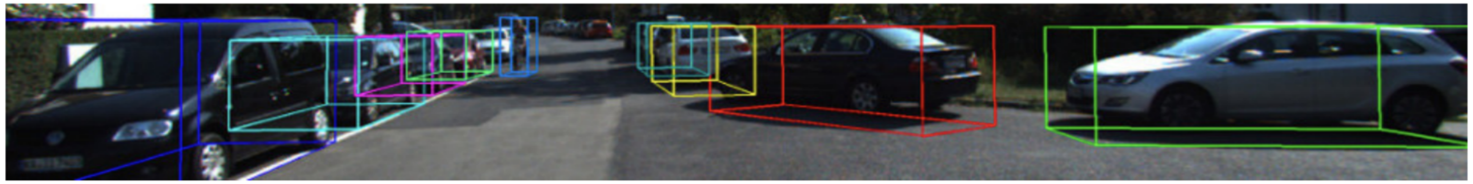
- RGB-Depth dataset with 2.5 million views obtained from more than 1500 scans



1.4 3D datasets

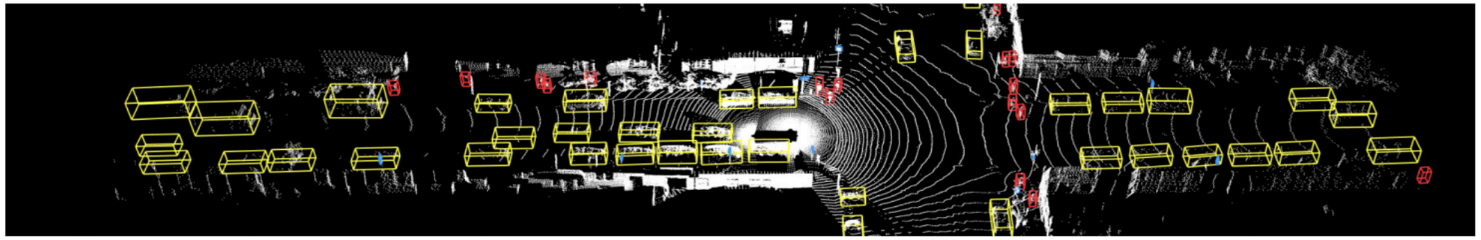
Outdoor 3D scene datasets (typically for autonomous vehicle applications)

KITTI: LiDAR data, labeled by 3D b.boxes



Semantic KITTI: LiDAR data, labeled per point

Waymo Open Dataset: LiDAR data, labeled by 3D b.boxes

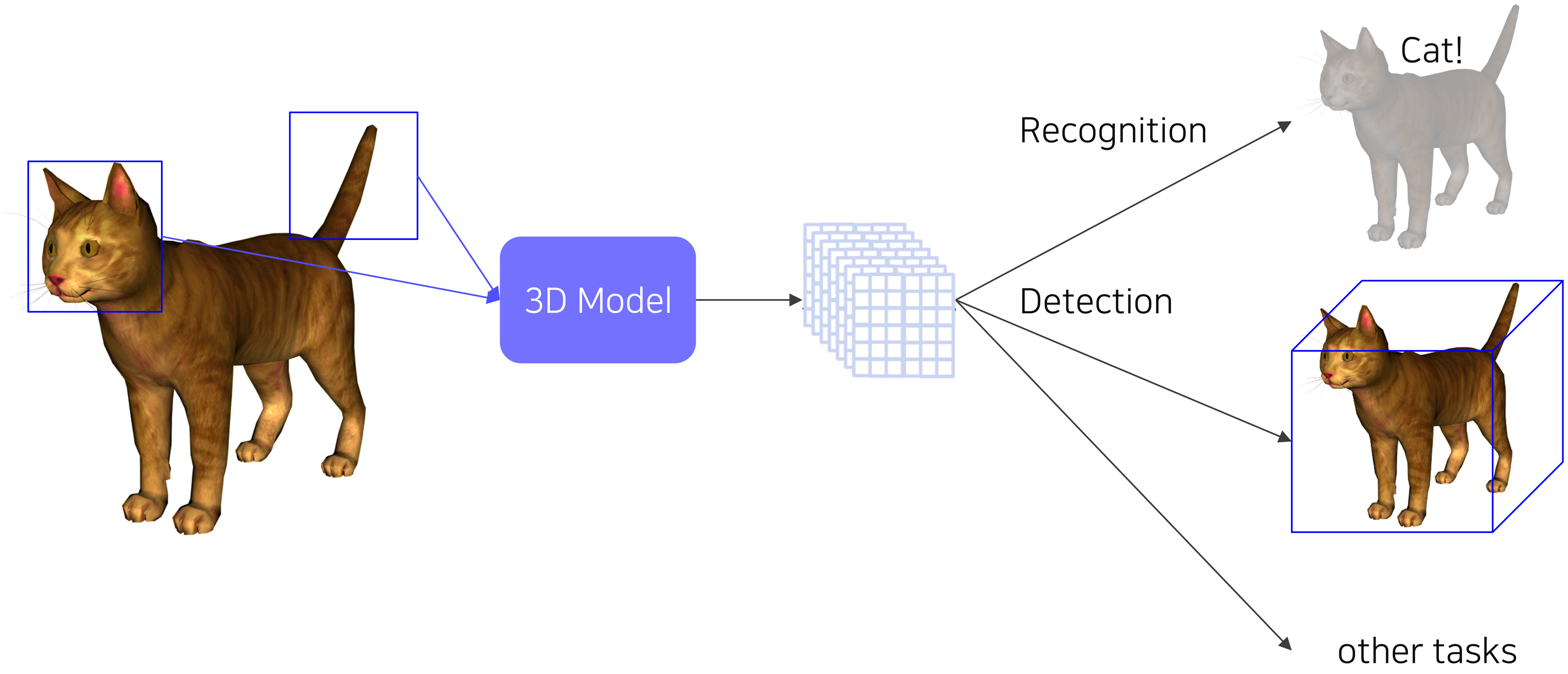


2.

3D tasks

2.1 3D recognition

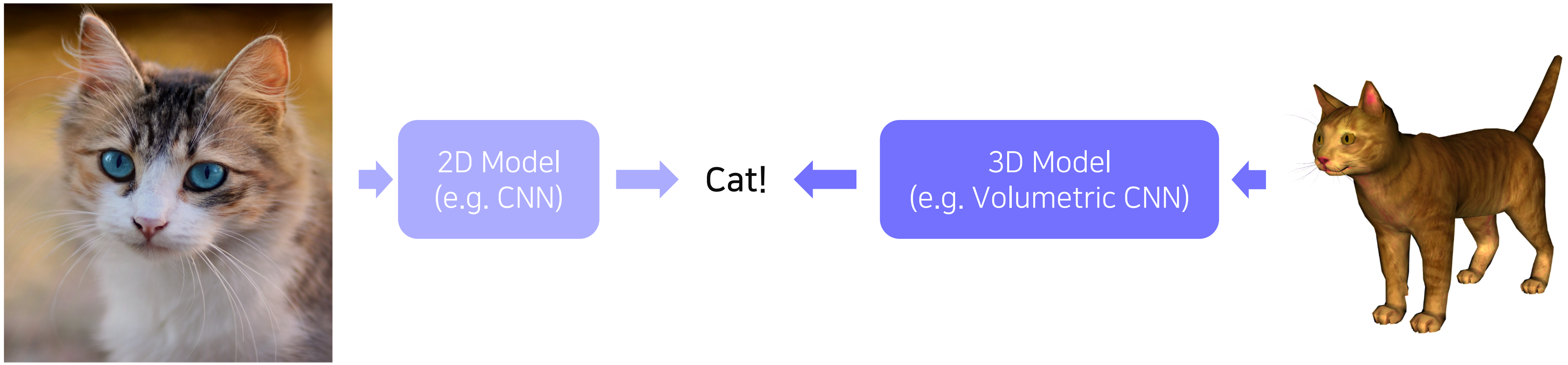
Various tasks for 3D data



2.1 3D recognition

3D object recognition

- Recognizing a 3D object like the object recognition in 2D image



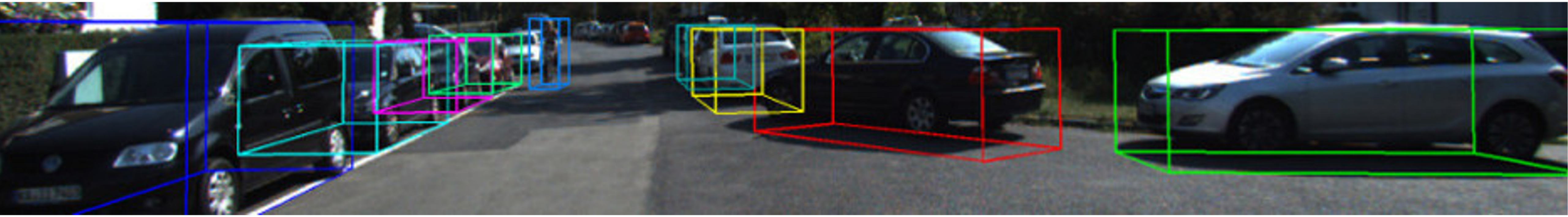
Cat in a 2D image space

Cat in a 3D space

2.2 3D object detection

3D object detection

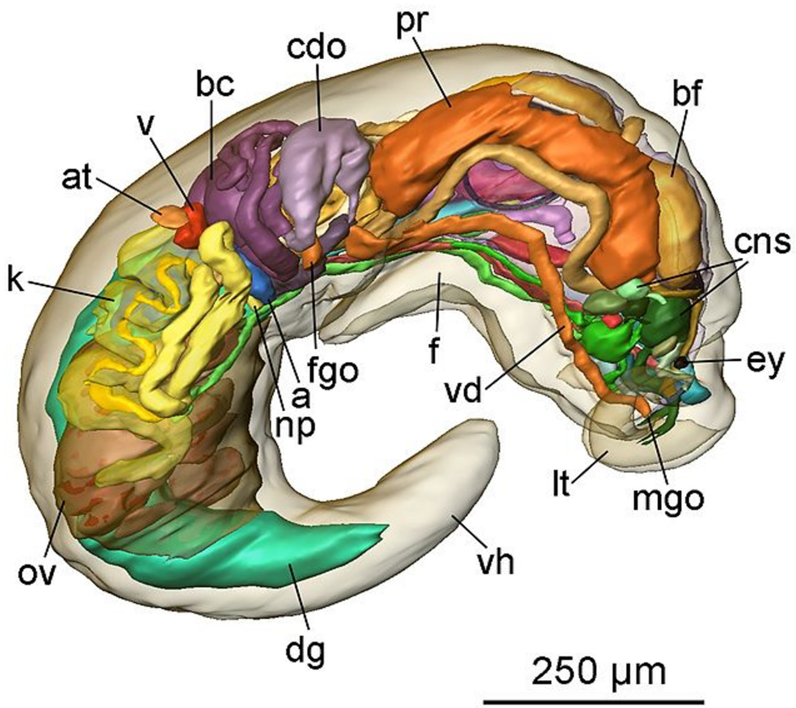
- Detecting 3D object locations in image or 3D spaces
- Useful for autonomous driving applications

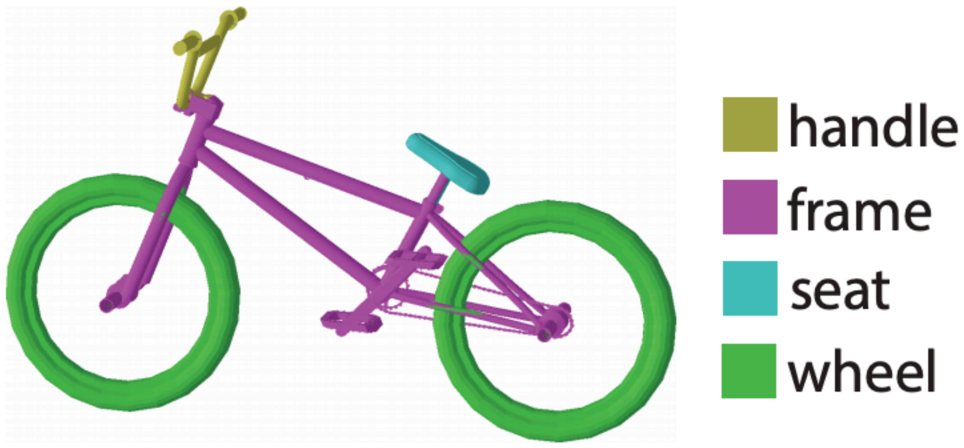


2.3 3D semantic segmentation

3D semantic segmentation

- Semantic segmentation of 3D data, such as neuroimaging



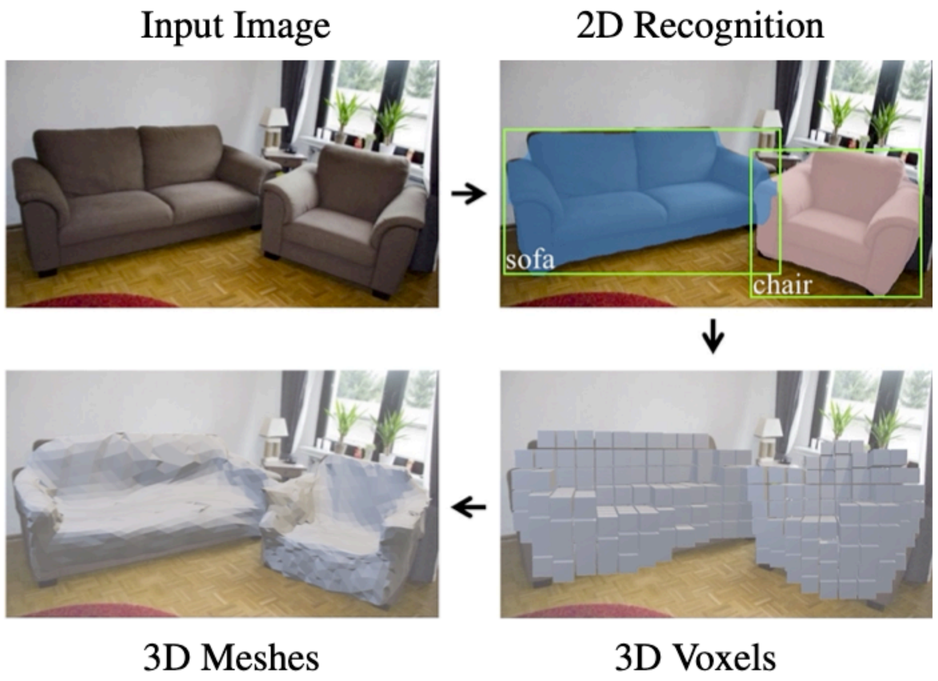


2.4 Conditional 3D generation

Mesh R-CNN

[Gkioxari et al., ICCV 2019]

- Input: a 2D image, output: 3D meshes of detected objects
- Can be implemented by modification from Mask R-CNN

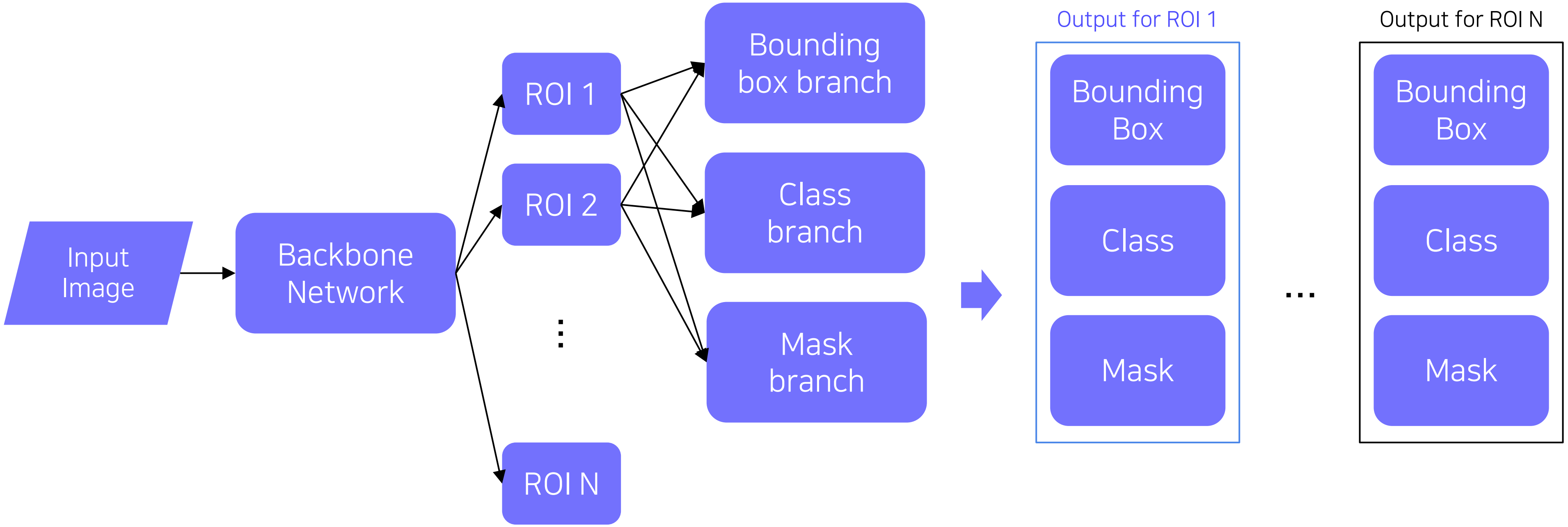


2.4 Conditional 3D generation

Recap: Branches in Mask R-CNN

[Gkioxari et al., ICCV 2019]

- Mask R-CNN segments objects by predicting ”box”, “classes”, and “mask”
- Branches infer each output from a shared feature corresponding to each RoI

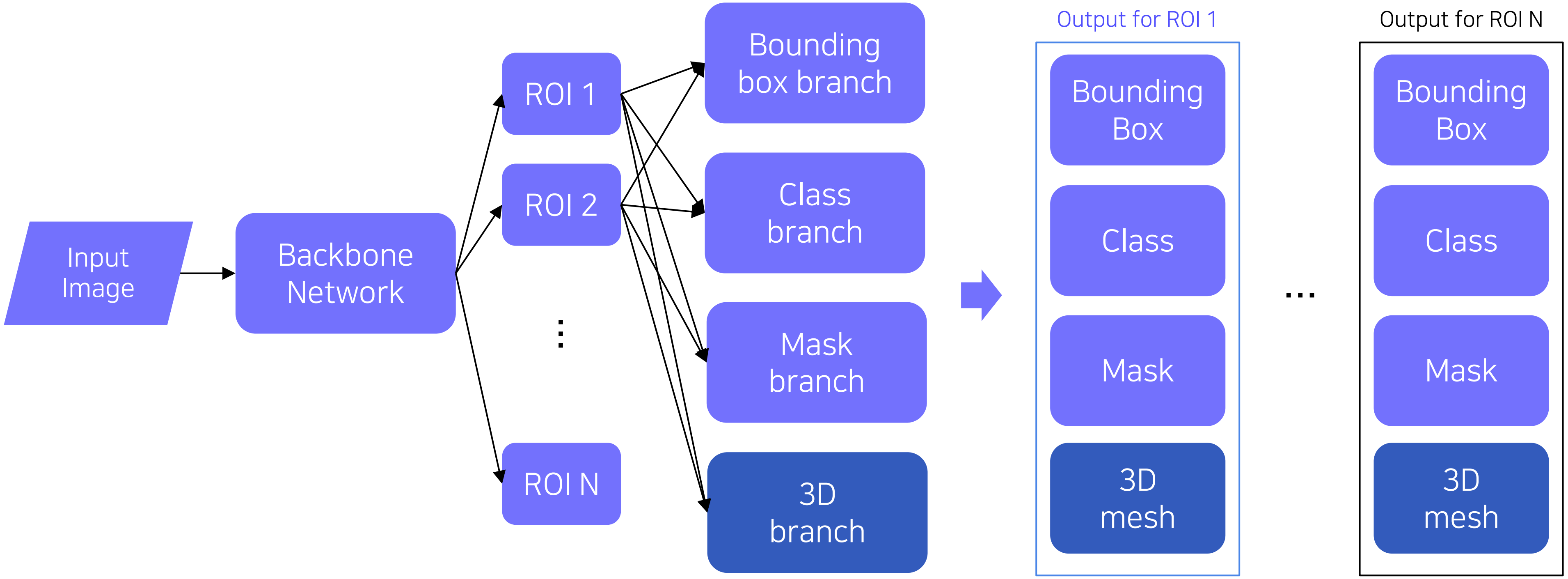


2.4 Conditional 3D generation

Mask R-CNN vs. Mesh R-CNN

[Gkioxari et al., ICCV 2019]

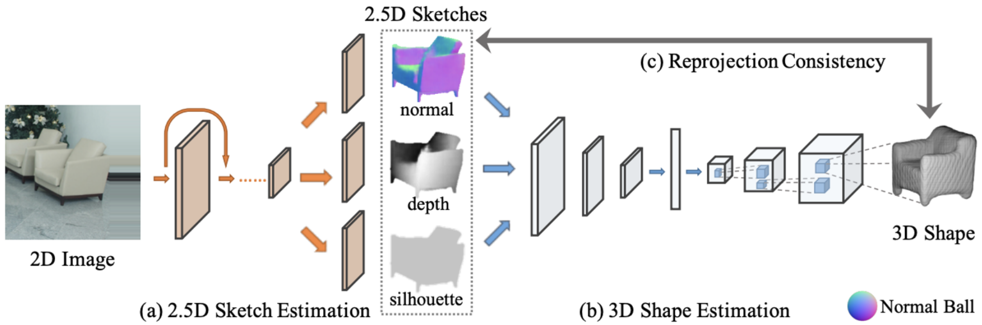
- Mesh R-CNN: "3D branch" is added to Mask R-CNN
- The 3D branch outputs a 3D mesh of an object



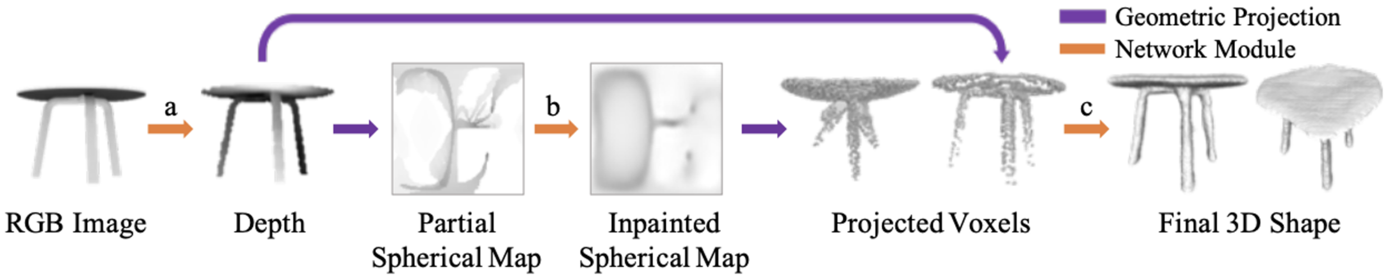
2.4 Conditional 3D generation

More complex 3D reconstruction models

- Decomposing 3D object reconstruction into multiple sub-problems
- Sub-problems: physically meaningful disentanglement (Surface normal, depth, silhouette, ⋯)



[Wu et al., NeurIPS 2017]



[Zhang et al., NeurIPS 2018]

Conclusion

[https://clearpathrobotics.com/blog/2017/01/3d-lidar-true-3d-sensing-spinning-2d-alternatives/]

3.

3D application example

3.1 Photo refocusing

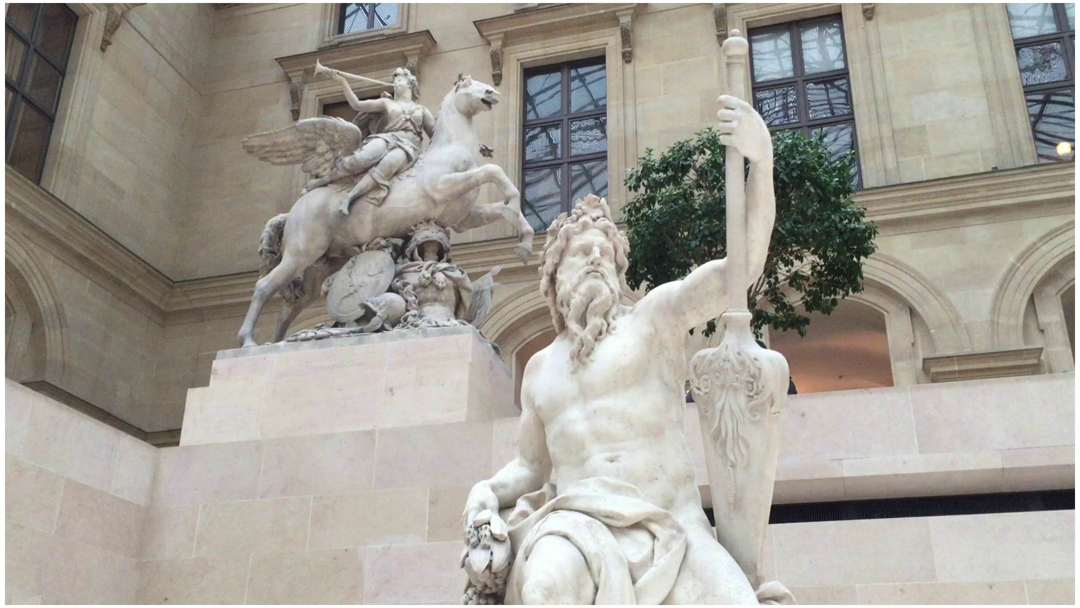
Defocusing a photo using depth map

Expected result of the post-refocused image

3.1 Photo refocusing

Implement the post-refocusing feature in your phone

- Implementing an after-refocusing image feature given depth map
- Almost same with "portrait mode" in your smartphone camera app
- Depth map can be estimated by depth sensors or neural networks



Original image

Inferred depth map of the image

Image samples from [Ha et al., High-quality Depth from Uncalibrated Small Motion Clip, CVPR 2016]

3.1 Photo refocusing

Defocusing a photo using depth map

1. Set a depth threshold range $[\mathrm{D}_{min}, \mathrm{D}_{max}]$ you want to focus
   (in this example, we will consider only D_{max} for simplicity)

2. Compute a mask of "focusing area" and "defocusing area" by depth map thresholding

3. Generate a blurred version of the input image

4. Compute "Masked focused image" and "Masked defocused image"

5. Blend masked images to generate a refocused image

3.1 Photo refocusing

Defocusing a photo using depth map

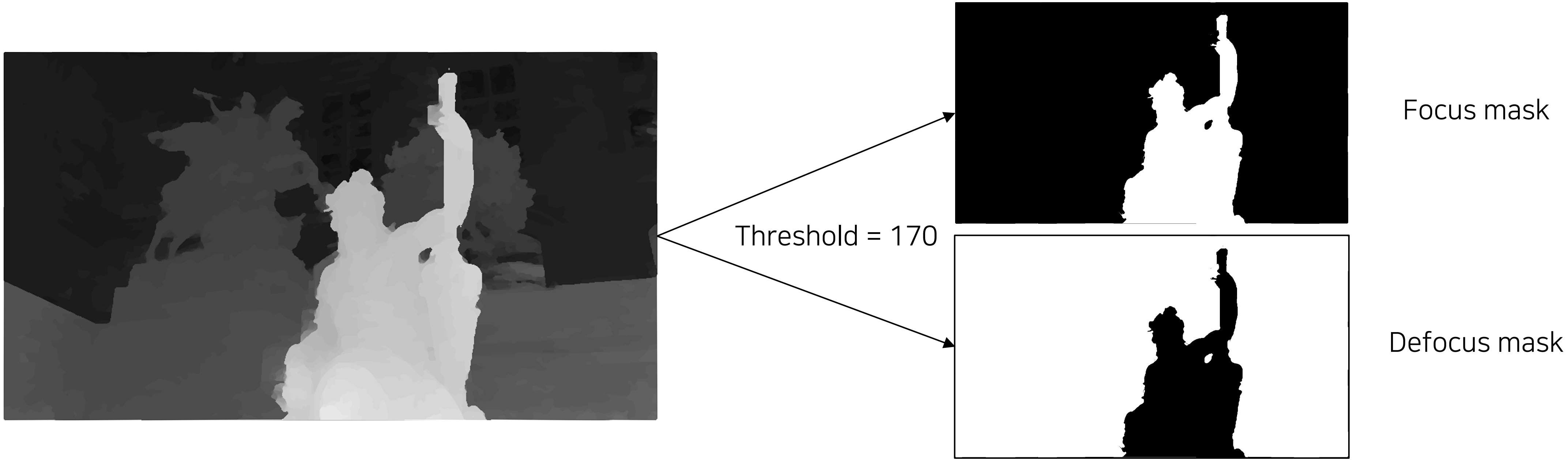
1. Set a depth threshold range $[\mathrm{D}_{min}, \mathrm{D}_{max}]$ you want to focus
(in this example, we will consider only D_{max} for simplicity)

The depth values are normalized to integer values $\{0, \cdots, 255\}$
You can arbitrarily set a threshold depth you want to focus

3.1 Photo refocusing

Defocusing a photo using depth map

2. Compute a mask of "focusing area" and "defocusing area" by depth map thresholding

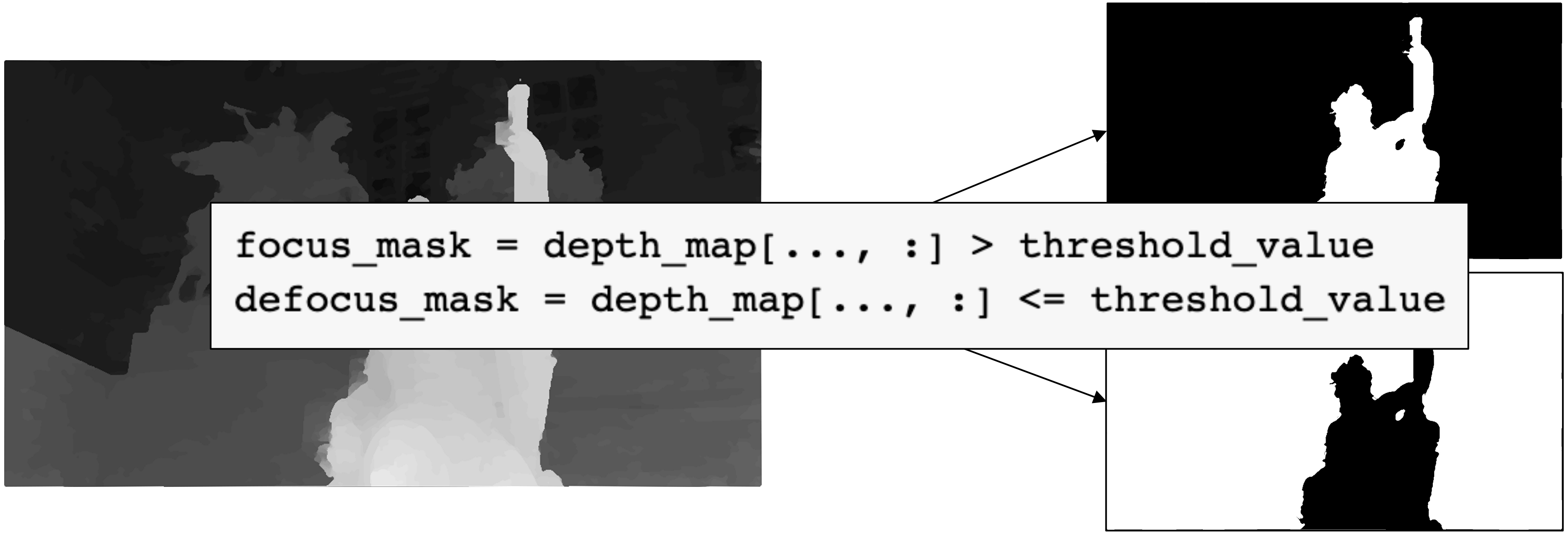


The depth values are normalized to integer values {0,⋯,255}
You can arbitrarily set a threshold depth you want to focus

3.1 Photo refocusing

Defocusing a photo using depth map

2. Compute a mask of “focusing area” and “defocusing area” by depth map thresholding



```
focus_mask = depth_map[..., :] > threshold_value
defocus_mask = depth_map[..., :] <= threshold_value
```

Focus mask

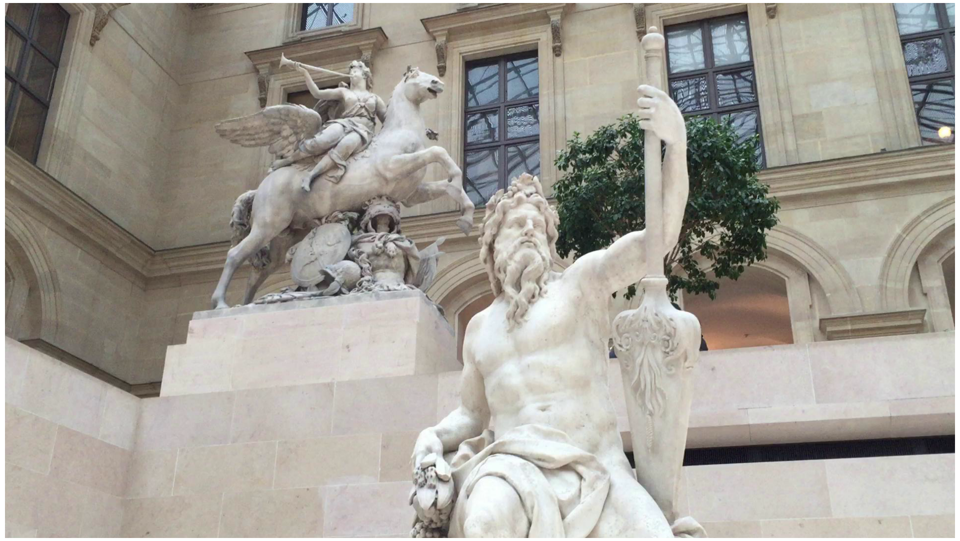
Defocus mask

The depth values are normalized to integer values {0,⋯,255}
You can arbitrarily set a threshold depth you want to focus

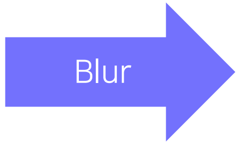
3.1 Photo refocusing

Defocusing a photo using depth map

3. Generate a blurred version of the input image



Original image



Blurred image with a kernel of size 20×20

3.1 Photo refocusing

Defocusing a photo using depth map

3. Generate a blurred version of the input image
(you can set blur parameters or method you preferred)

```
blurred_image = cv2.blur(original_image, (20, 20))
```

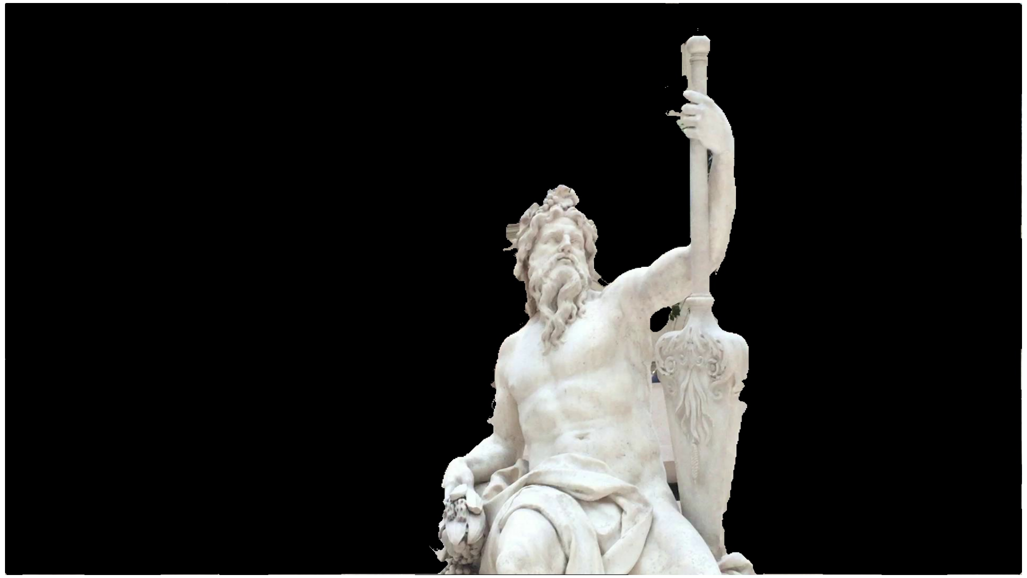
Original image

Blurred image with a kernel of size 20×20

3.1 Photo refocusing

Defocusing a photo using depth map

4. Compute ”Masked focus image” and “Masked defocused image”
   by using the precomputed mask



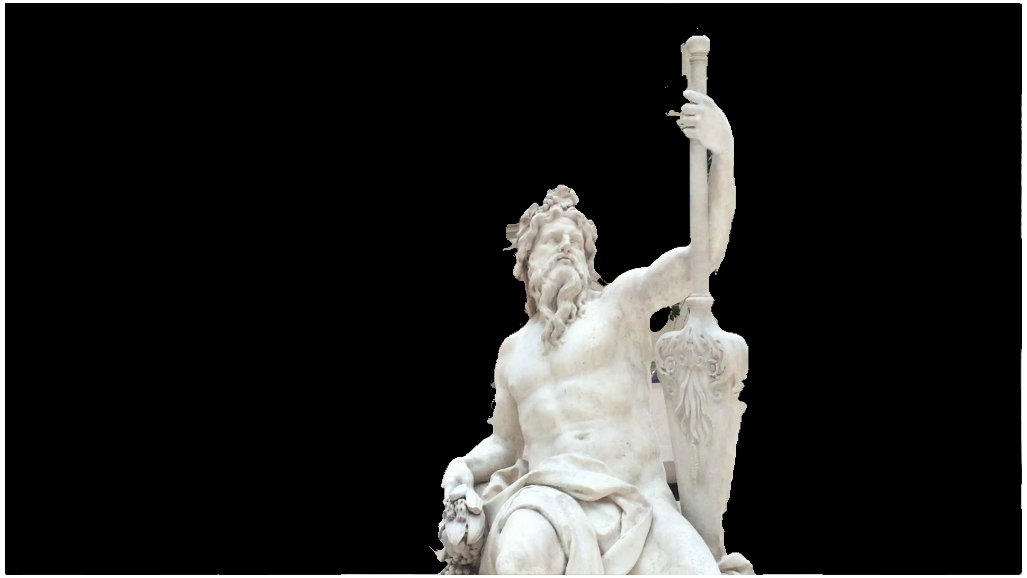
Masked focused image

Masked defocused image

3.1 Photo refocusing

Defocusing a photo using depth map

4. Compute "Masked focus image" and "Masked defocused image"
   by using the precomputed mask



`focused_with_mask = focus_mask * original_image`

`defocused_with_mask = defocus_mask * blurred_image`

3.1 Photo refocusing

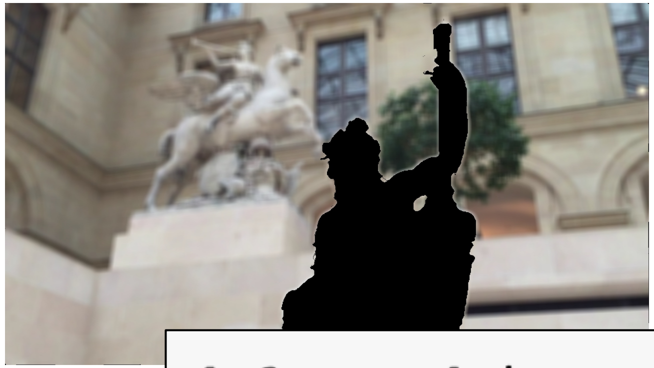
Defocusing a photo using depth map

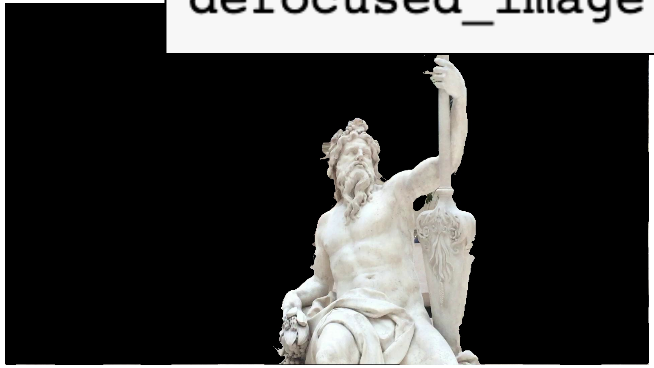
5. Blend masked images to generate a refocused image

3.1 Photo refocusing

Defocusing a photo using depth map

5. Blend masked images to generate a refocused image





```
defocused_image = focused_with_mask + defocused_with_mask
```

3.1 Photo refocusing

Defocusing a photo using depth map

Example results

Output with threshold = 50

Output with threshold = 170

Do it yourself to only focus on the middle object
(Hint: use a range of thresholds $[D_{min}, D_{max}]$ to generate masks)

Reference

1. 3D understanding

- Chang et al., ShapeNet: An Information-Rich 3D Model Repository, ArXiv 2015
- Mo et al., PartNet: A Large-Scale Benchmark for Fine-Grained and Hierarchical Part-Level 3D Object Understanding, CVPR 2019
- McCormac et al., SceneNet RGB-D: Can 5M Synthetic Images Beat Generic ImageNet Pre-Training on Indoor Segmentation?, ICCV 2017
- Dai et al., ScanNet: Richly-Annotated 3D Reconstructions of Indoor Scenes, CVPR 2017

2. 3D tasks

- Gkioxari et al., Mesh R-CNN, ICCV 2019
- Wu et al., MarrNet: 3D Shape Reconstruction via 2.5D Sketches, NeurIPS 2017
- Zhang et al., Learning to Reconstruct Shapes from Unseen Classes, NeurIPS 2018

End of Document
Thank You.