

Learning 3D Scene Priors with 2D Supervision



Yinyu Nie¹, Angela Dai¹, Xiaoguang Han², Matthias Nießner¹

<https://yinyunie.github.io/sceneprior-page/>

¹ Technical University of Munich

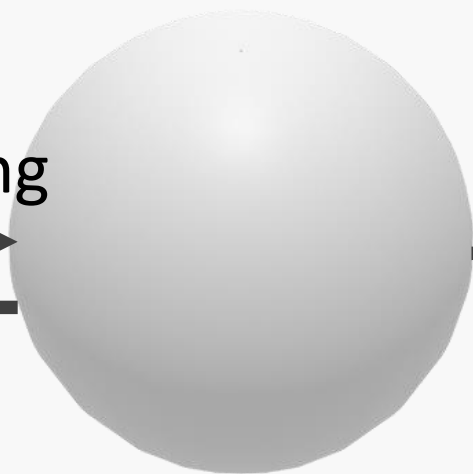
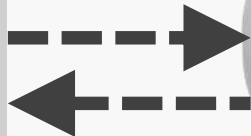
² The Chinese University of Hong Kong (Shenzhen)

Task: To Learn 3D Scene Priors from 2D



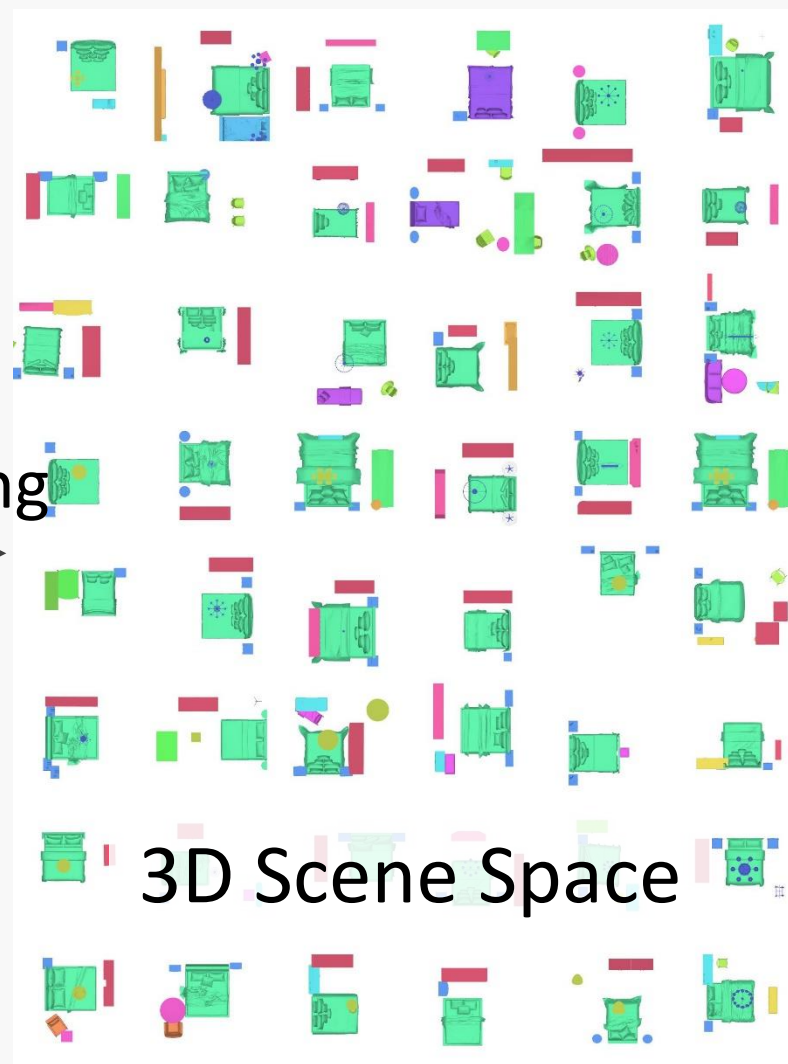
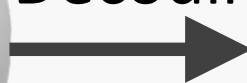
2D multi-view data

Learning



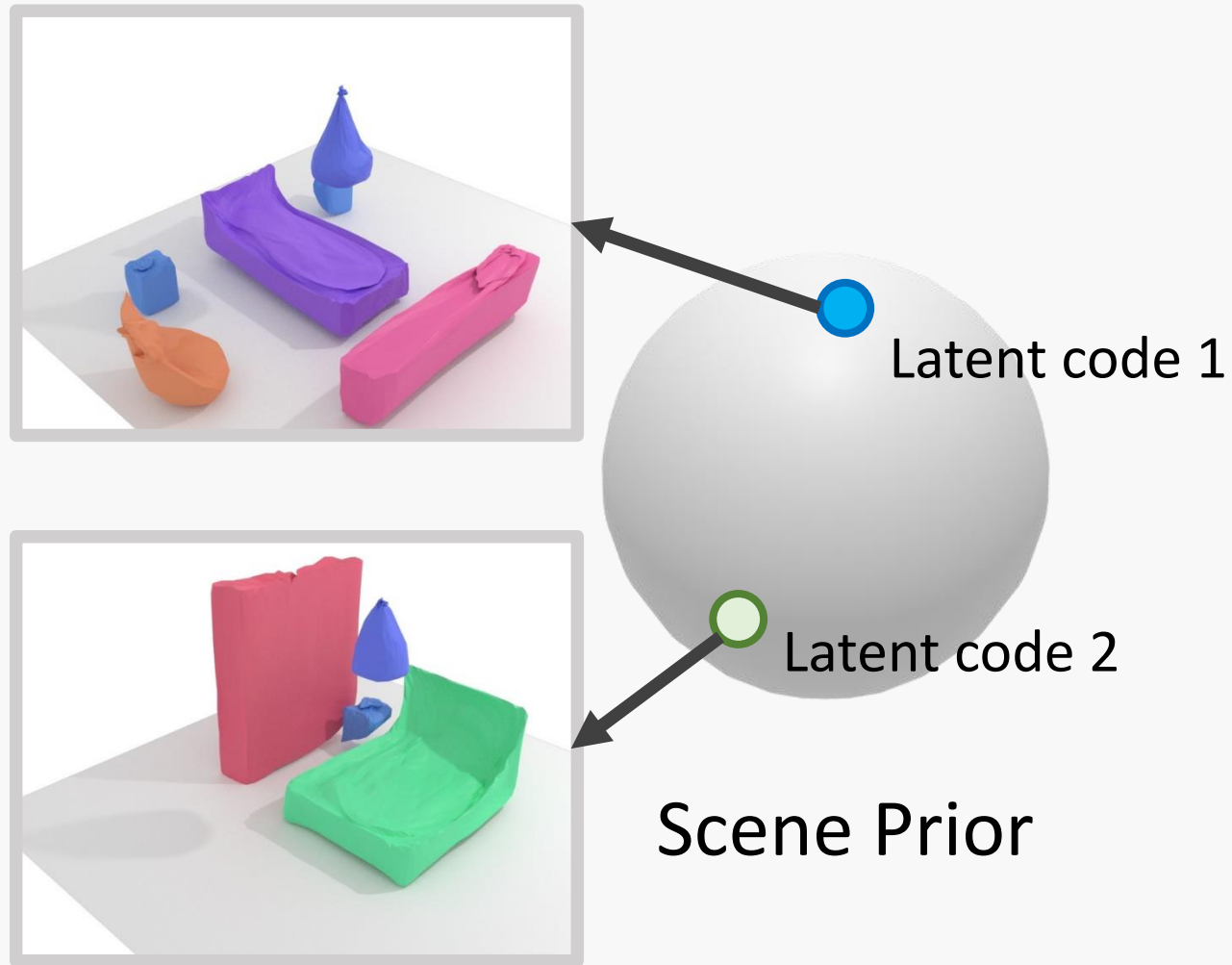
Latent Space
(Scene Prior)

Decoding

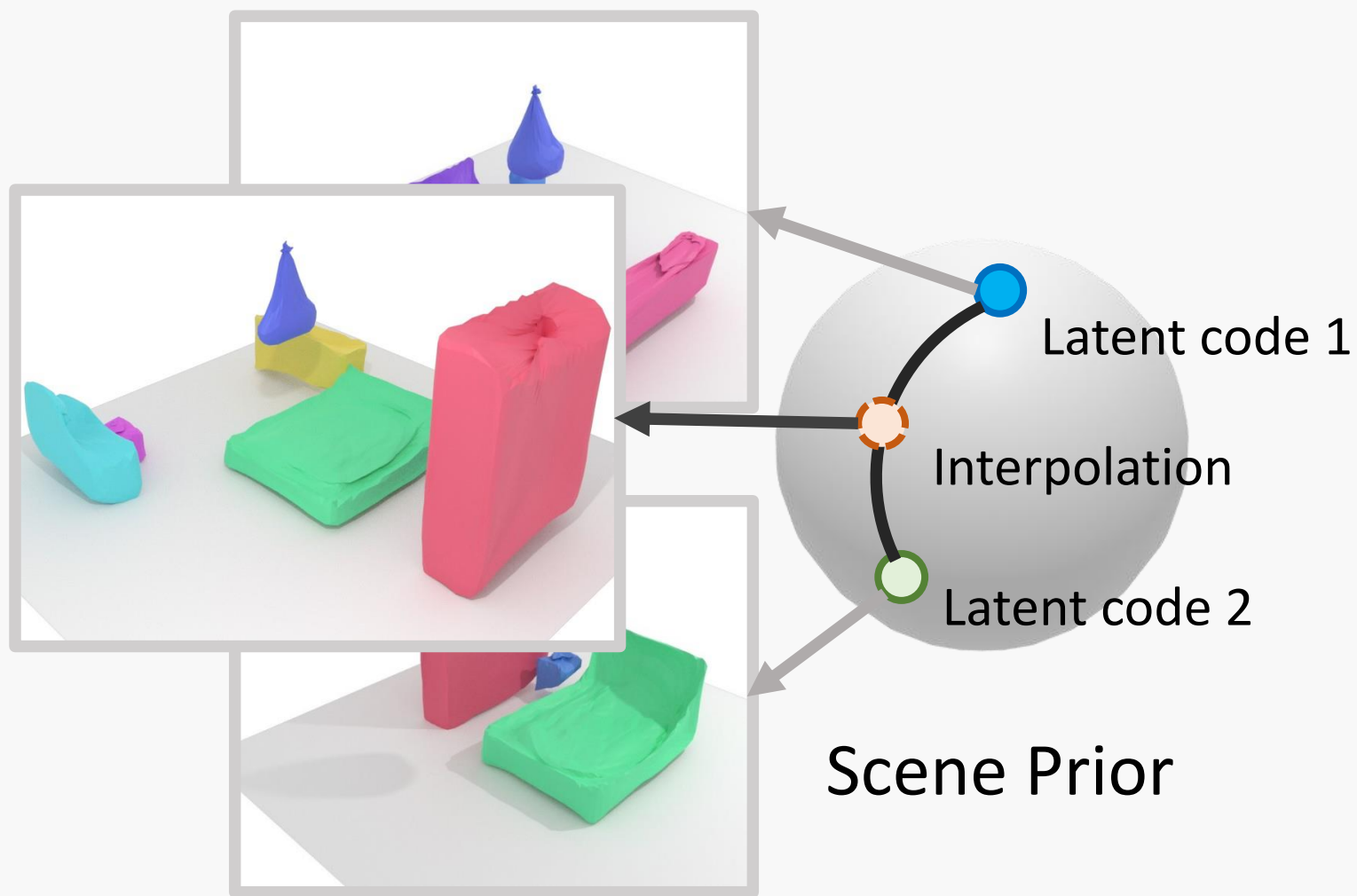


3D Scene Space

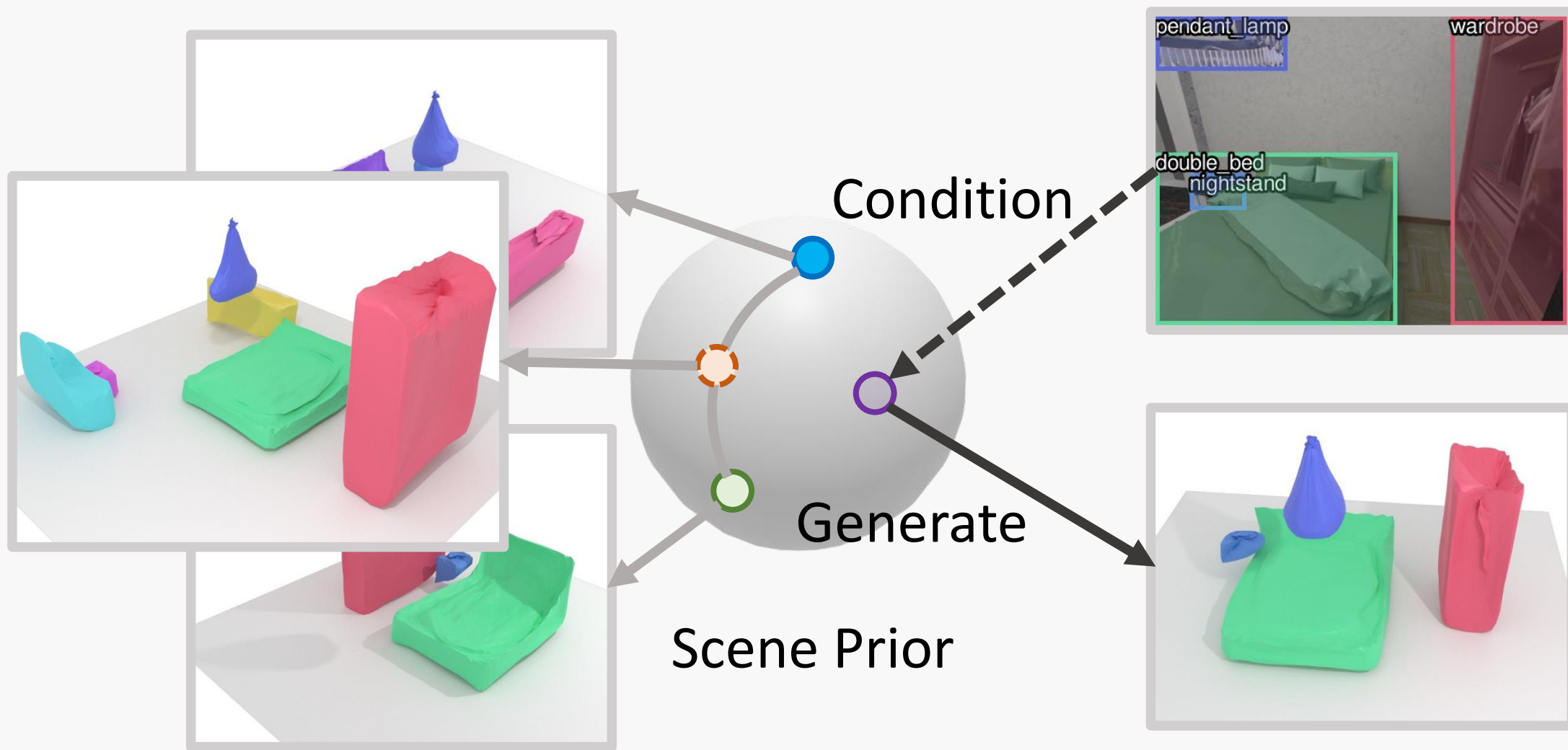
Downstream Application: Scene Synthesis



Downstream Application: Scene Interpolation



Downstream Application: Single-view Reconstruction



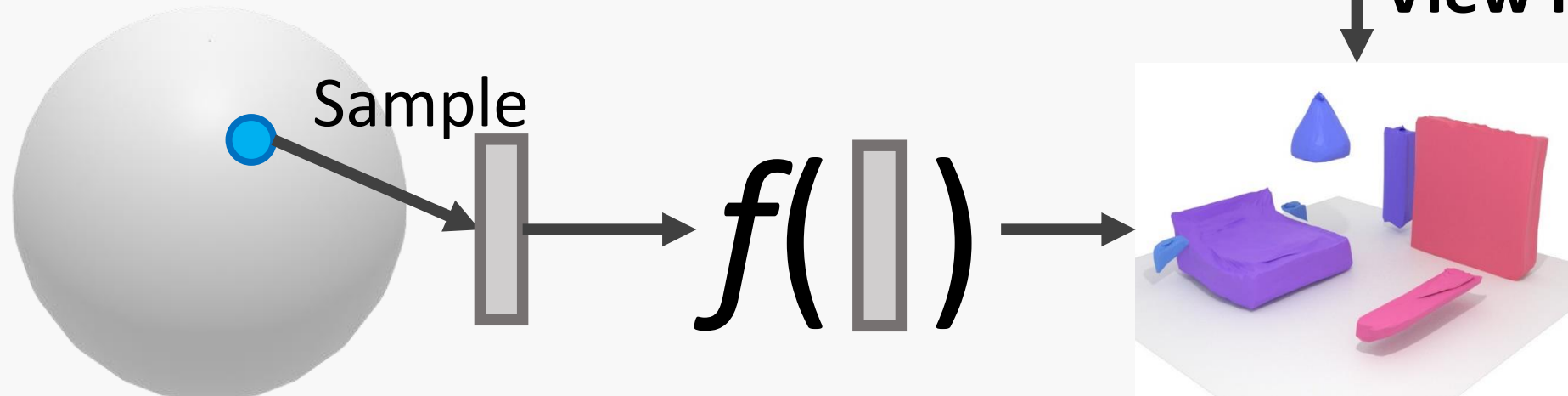
A 3D rendered scene of a living room. In the center, there is a purple sofa and a yellow armchair. To the left of the sofa is a green ottoman and a green chair. To the right of the sofa is a small pink side table. In the background, there is a blue circular rug and a blue chandelier hanging from the ceiling. The word "Method" is overlaid in the center of the scene.

Method

Method Overview



↕ **View loss**

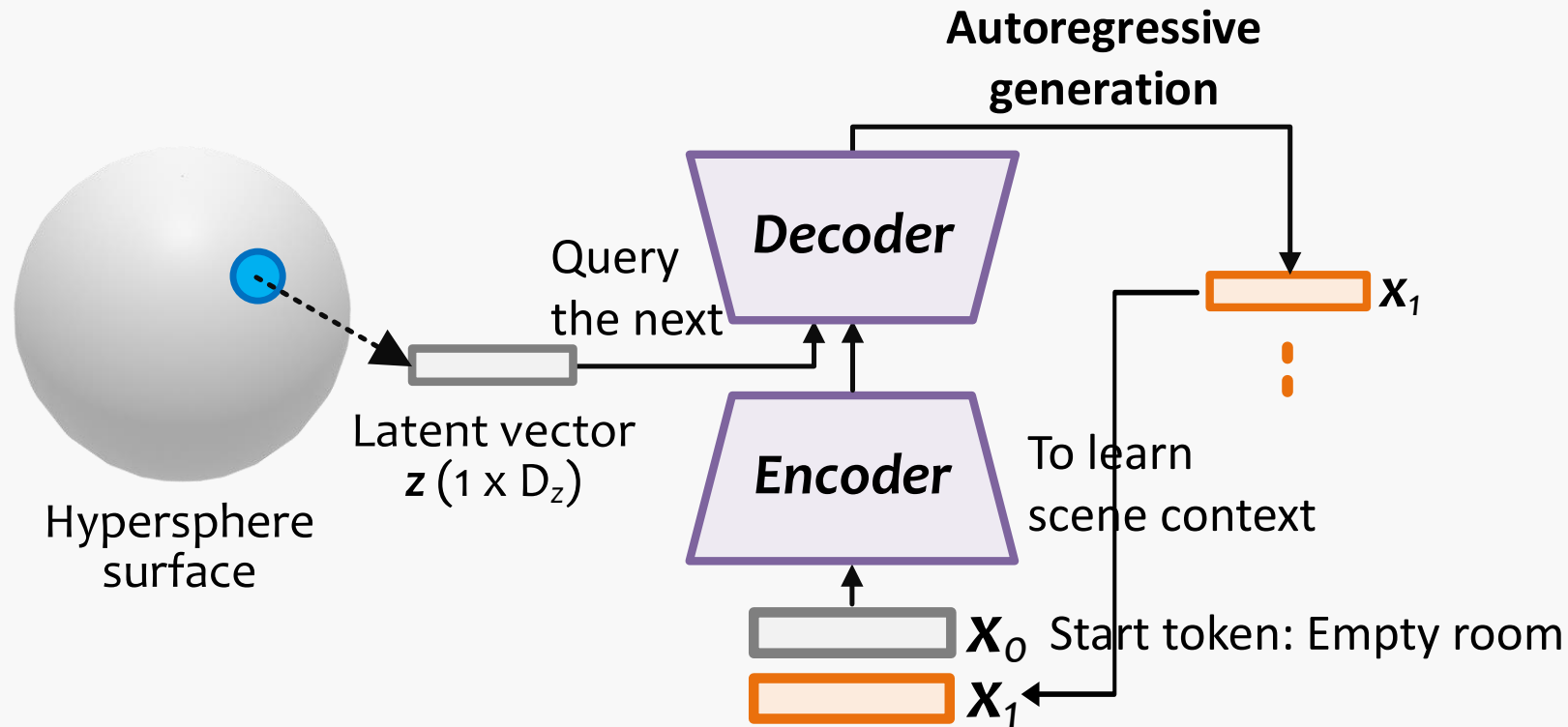


Latent Space

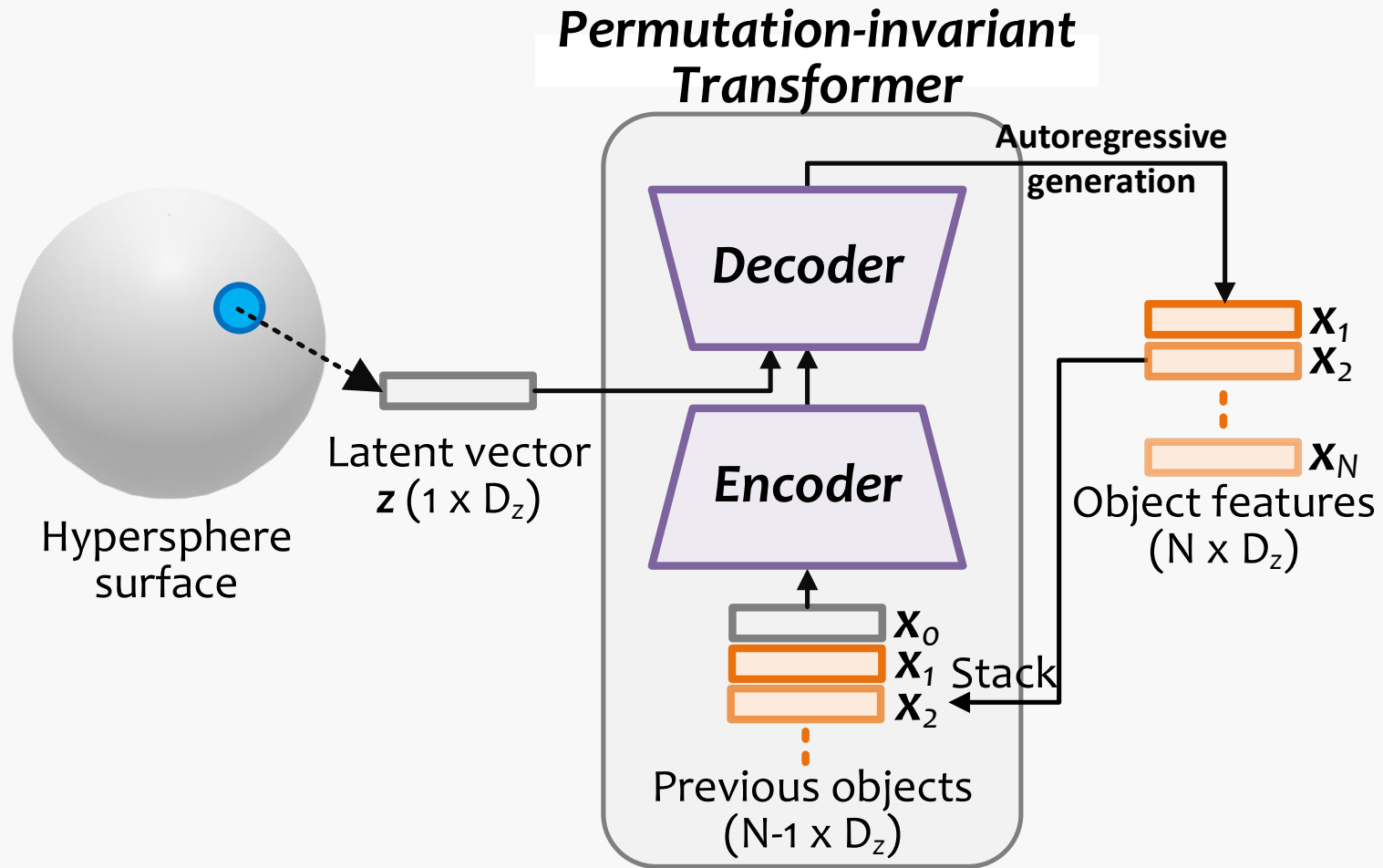
Decoder

Scene Space

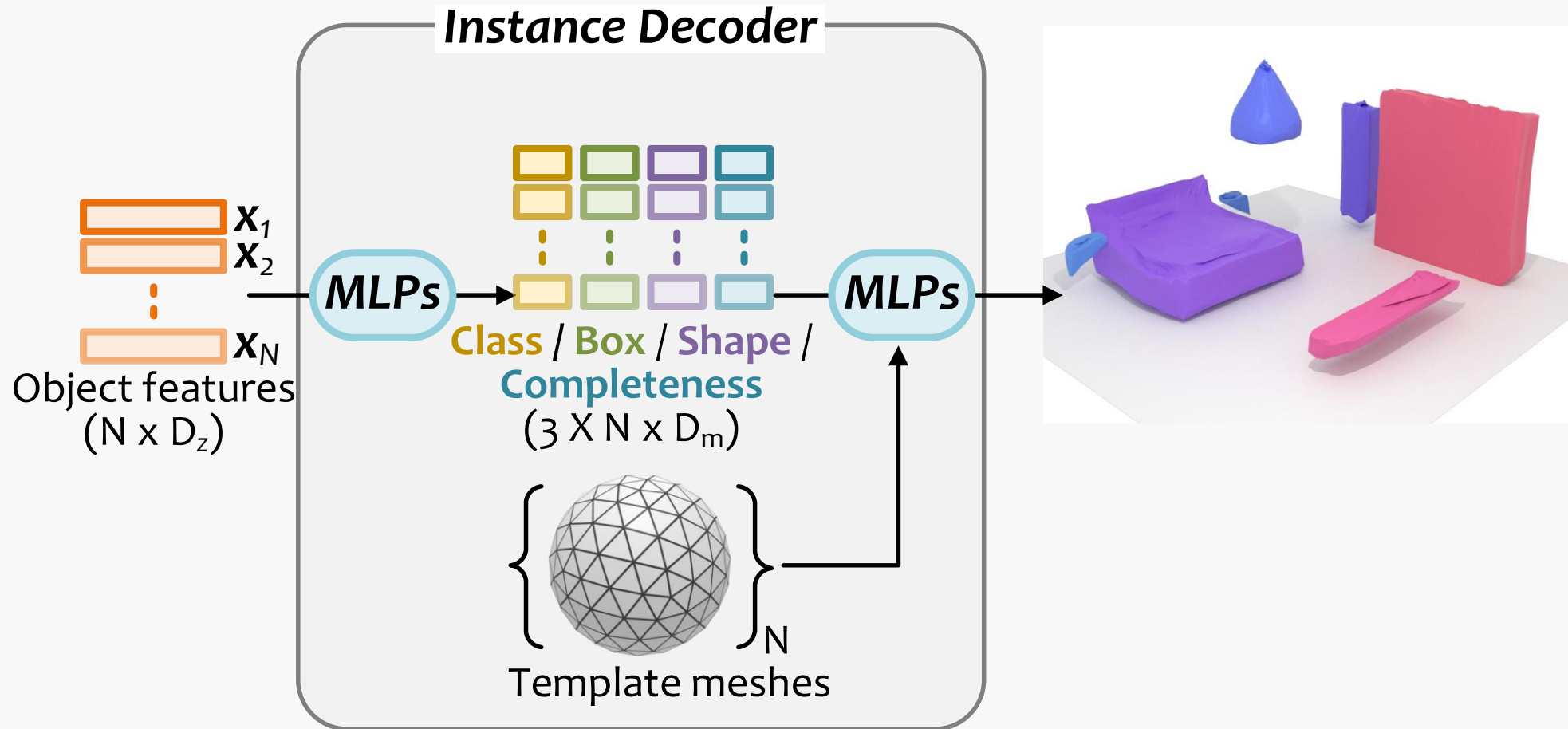
Network: Permutation-invariant Transformer



Network: Permutation-invariant Transformer

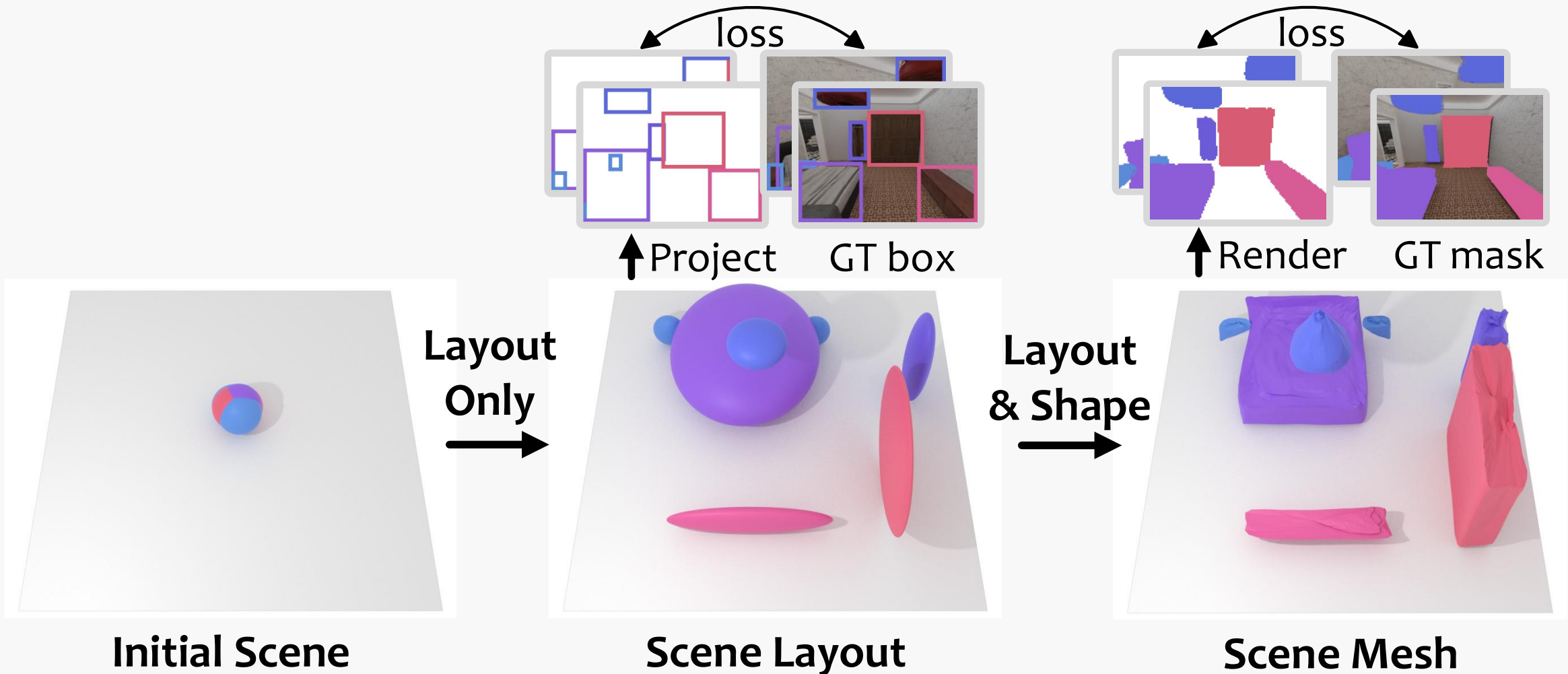


Network: Layout and Shape Decoder



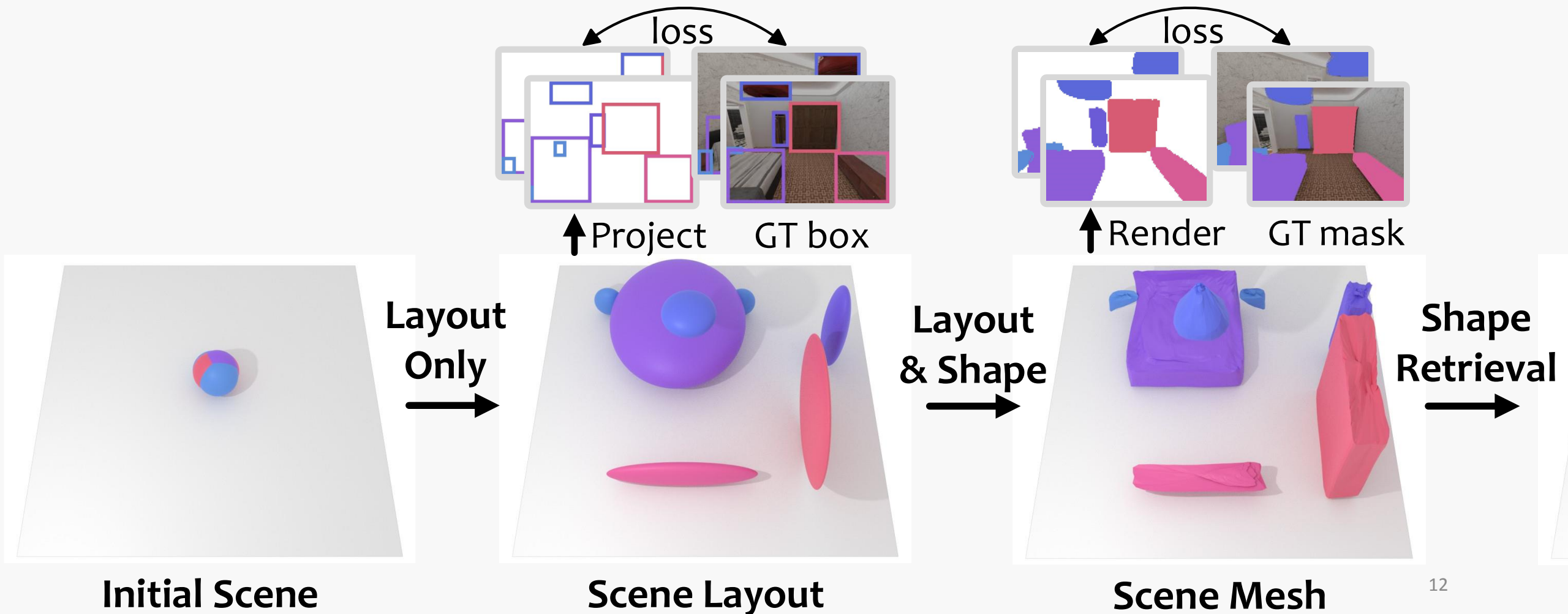
2D View Loss:

Latent codes and decoder are trained together



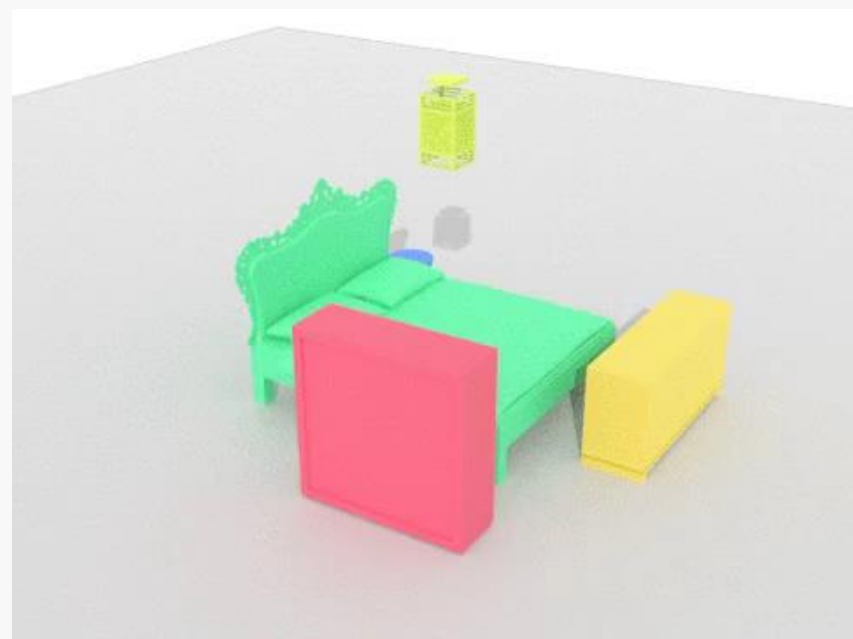
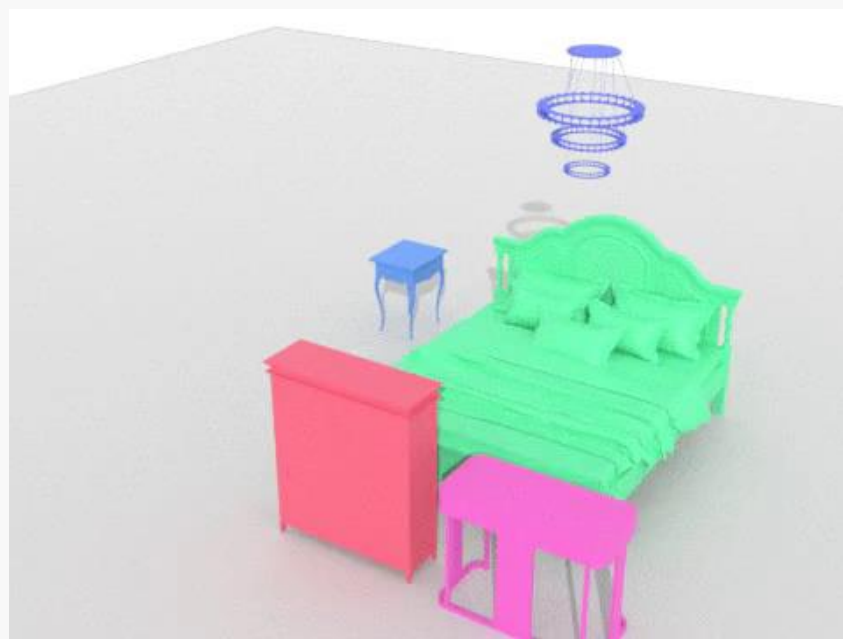
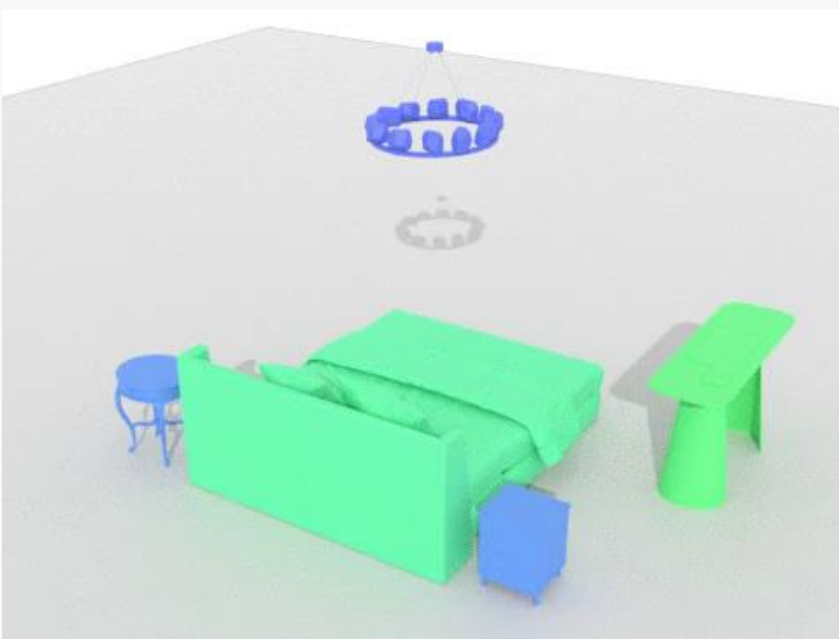
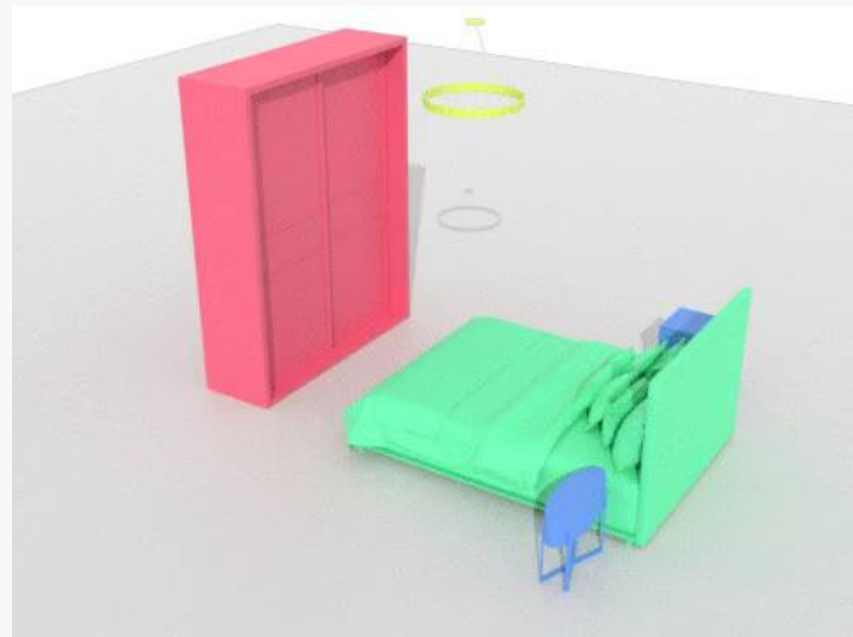
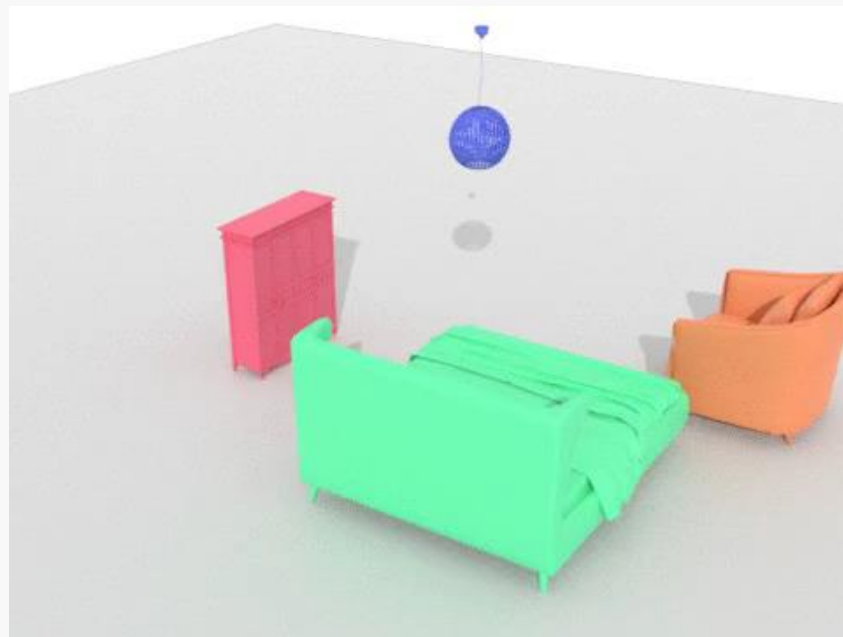
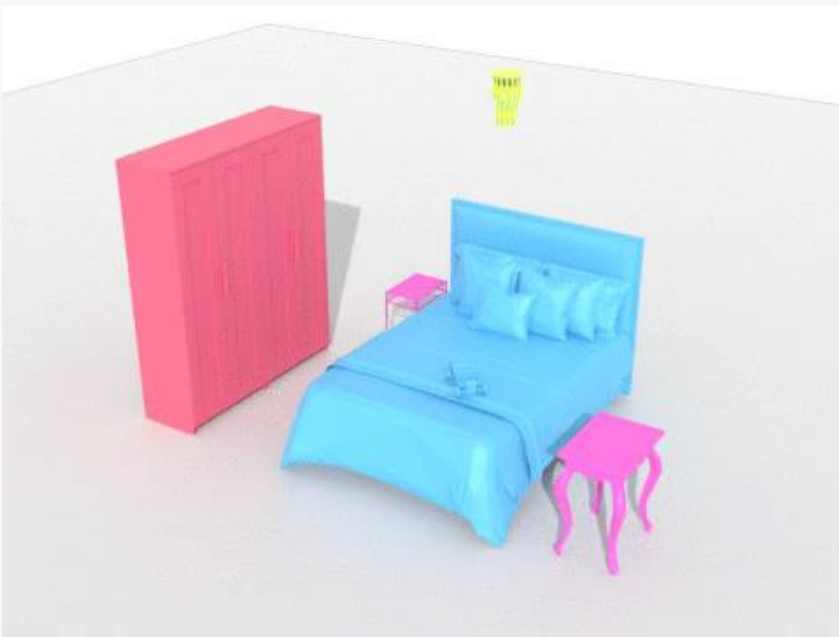
2D View Loss:

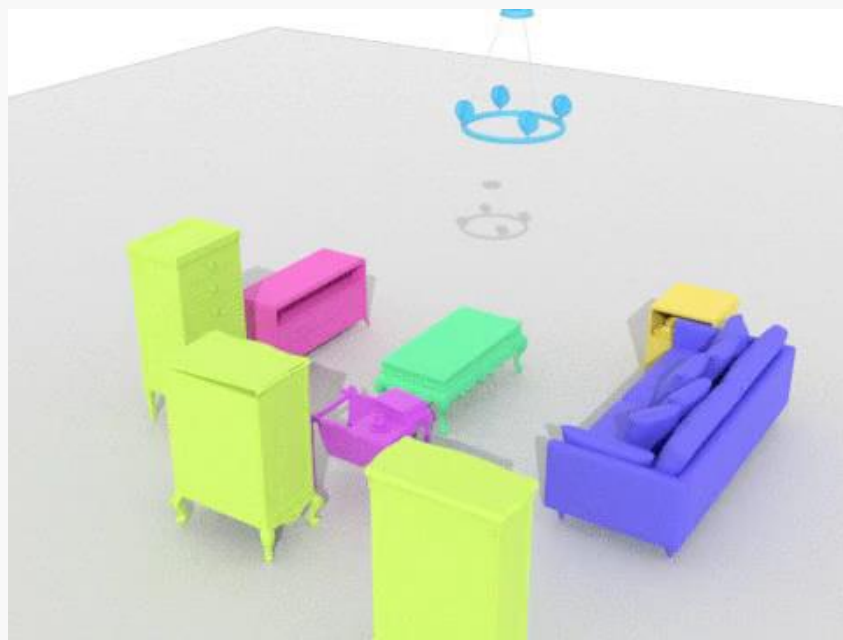
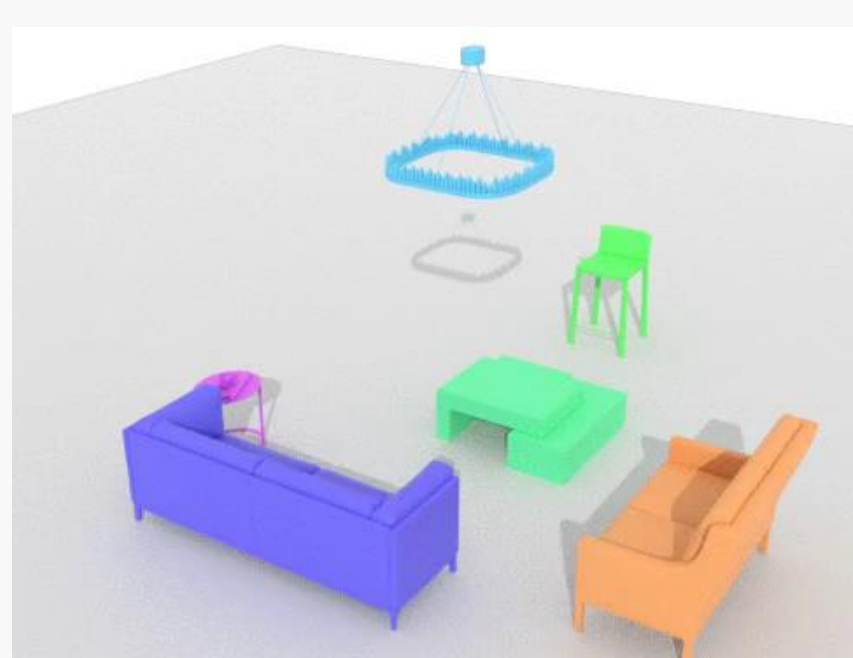
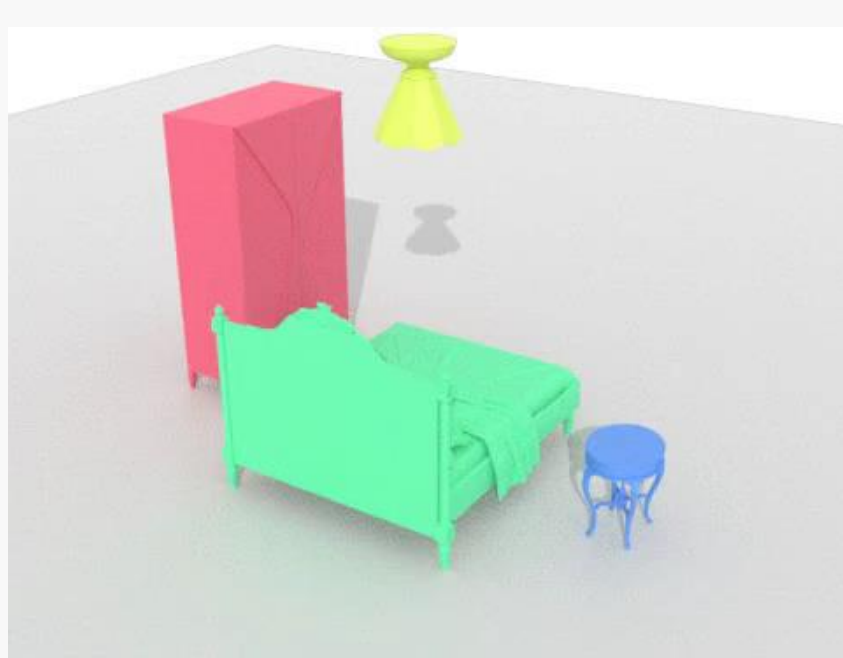
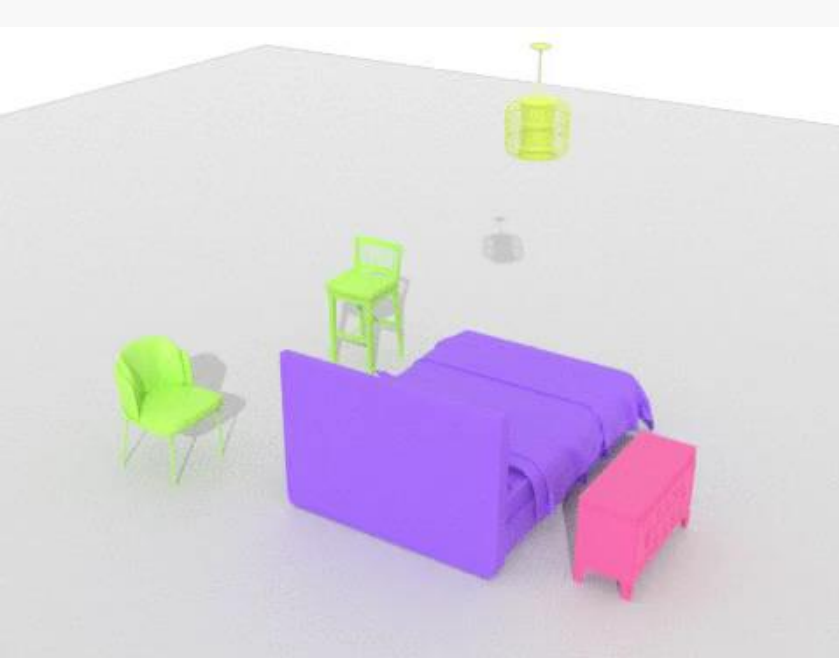
Latent codes and decoder are trained together

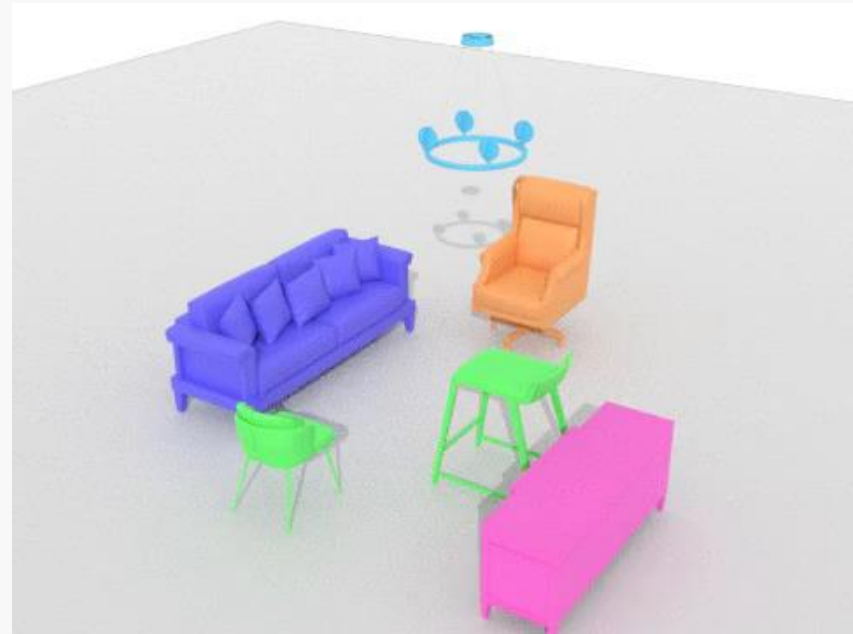
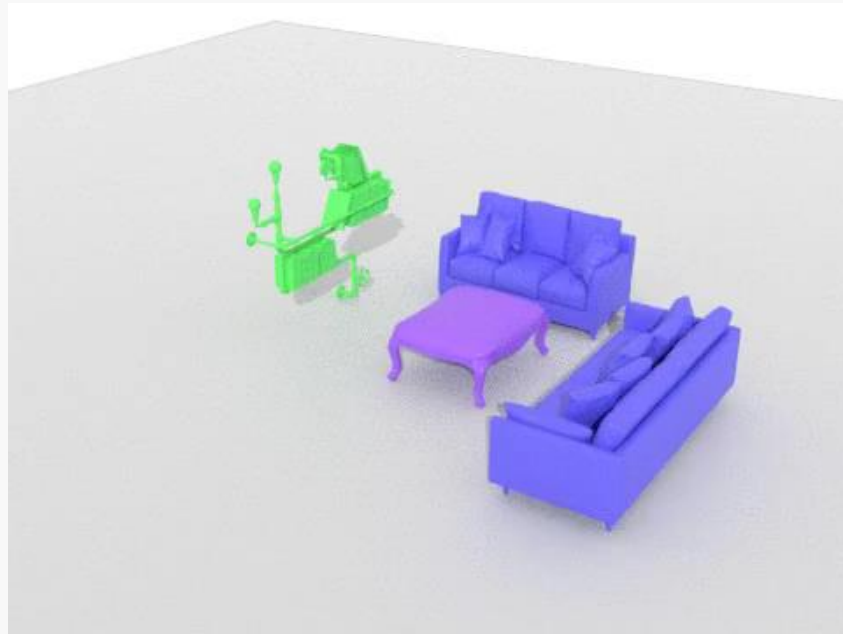


3D Scene Synthesis



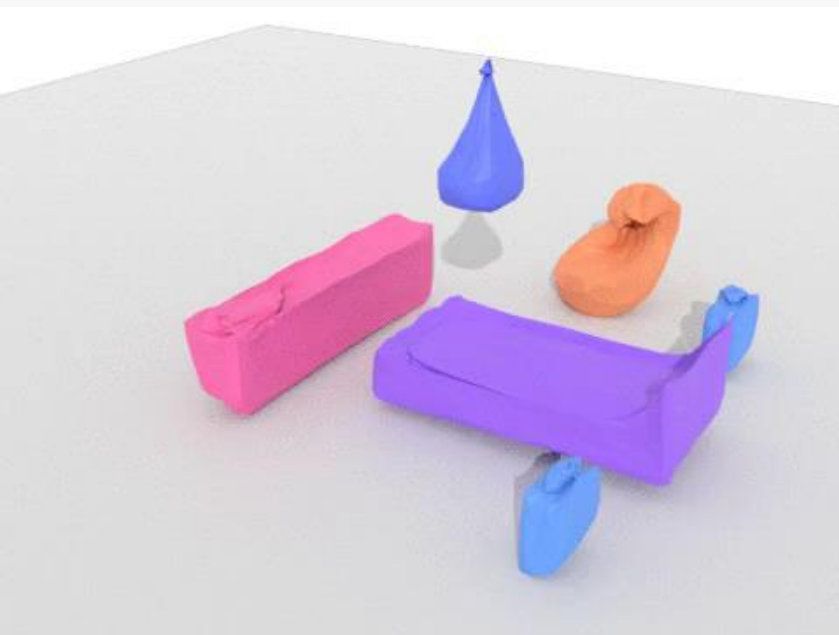




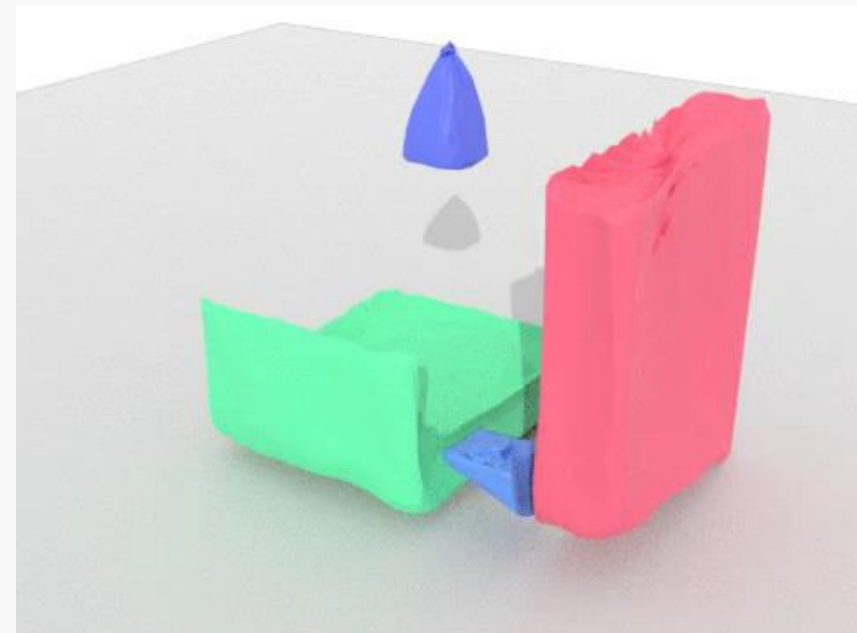
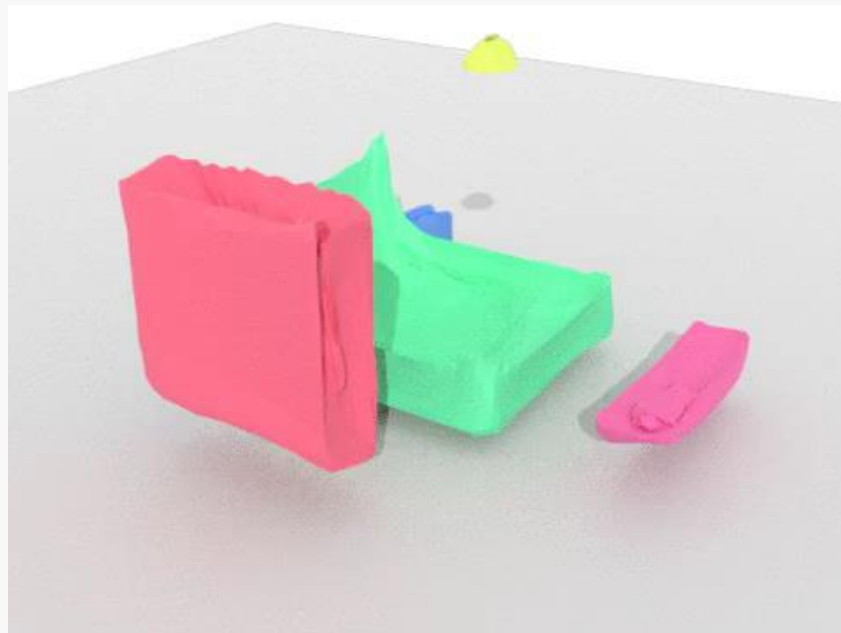
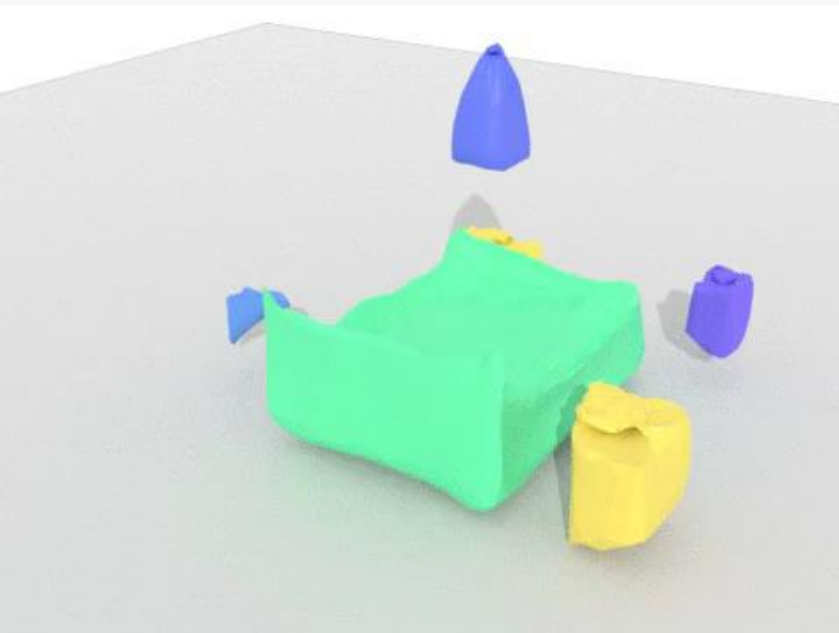
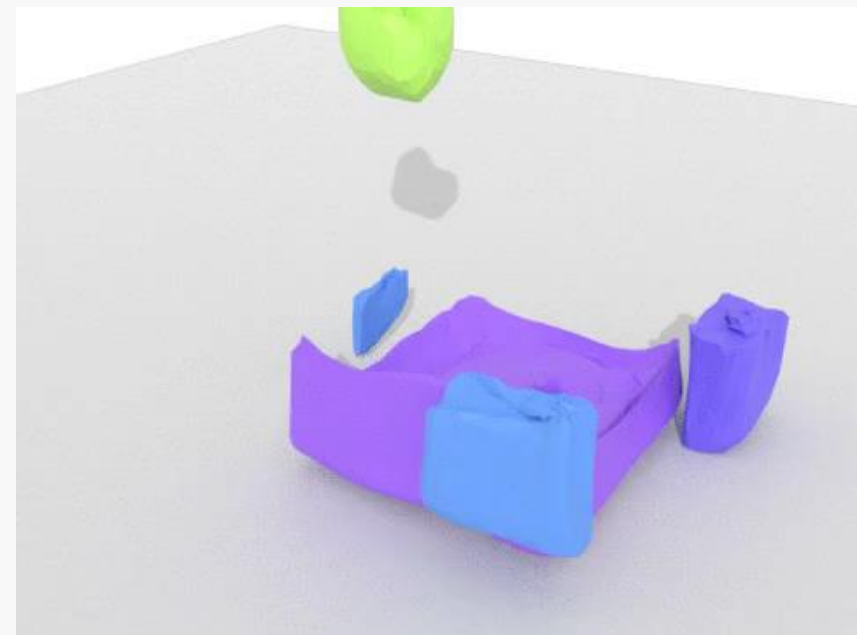
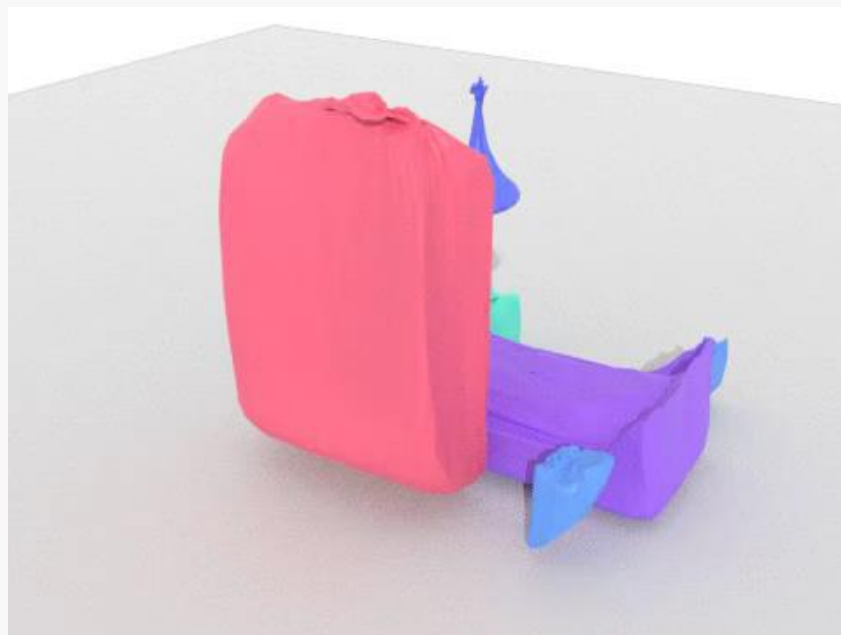


A 3D rendered scene featuring a minimalist interior. On the left is a light brown armchair. In the center is a light green sofa. To the right is a light pink rectangular table. Above the sofa hangs a blue spherical pendant light. The scene is set against a plain white background.

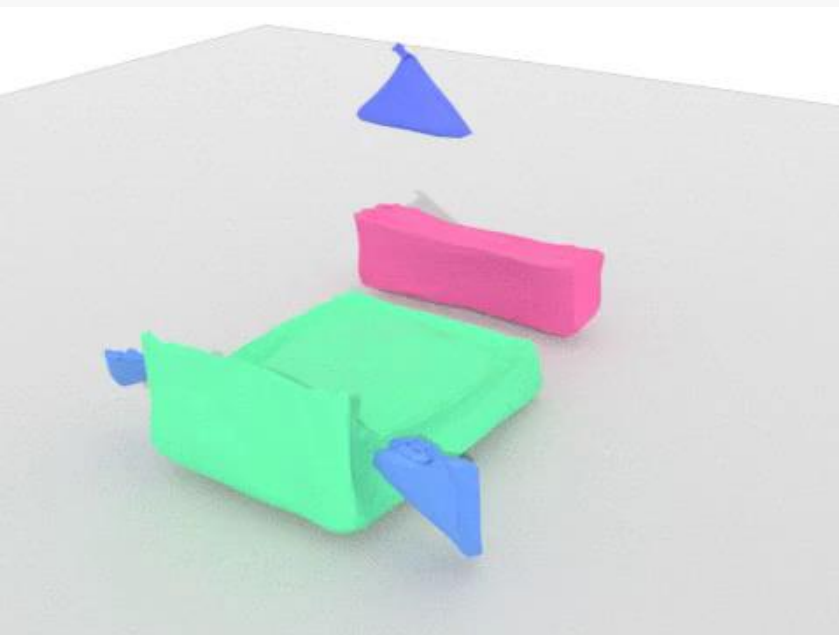
3D Scene Interpolation



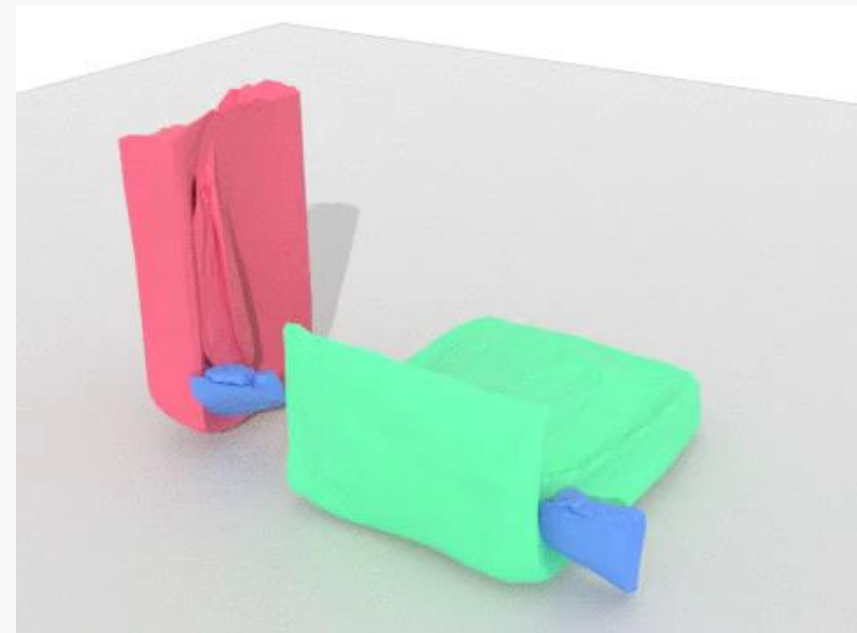
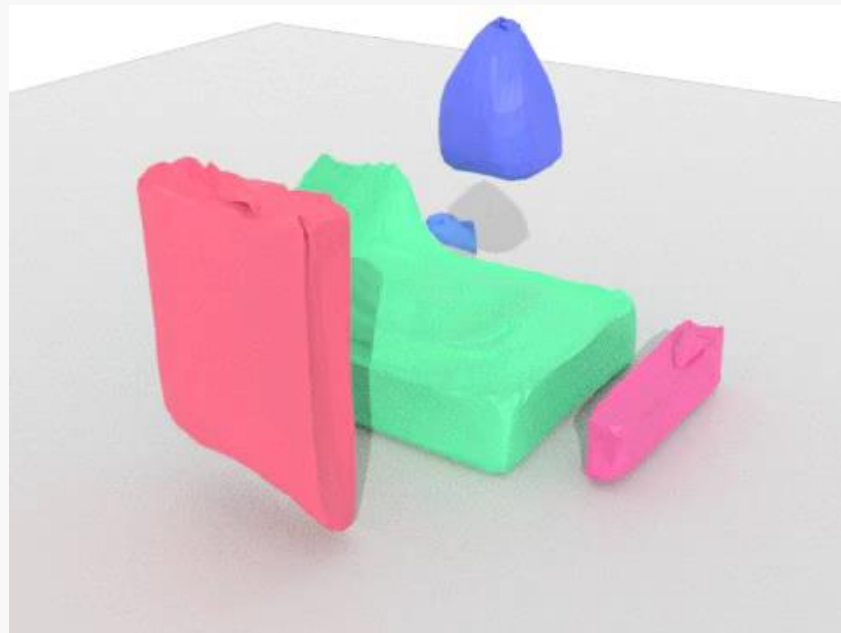
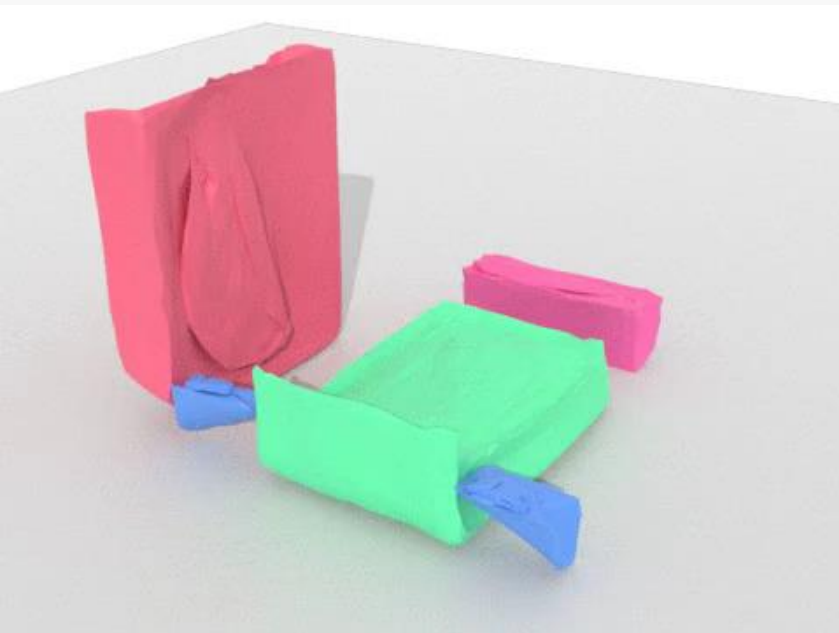
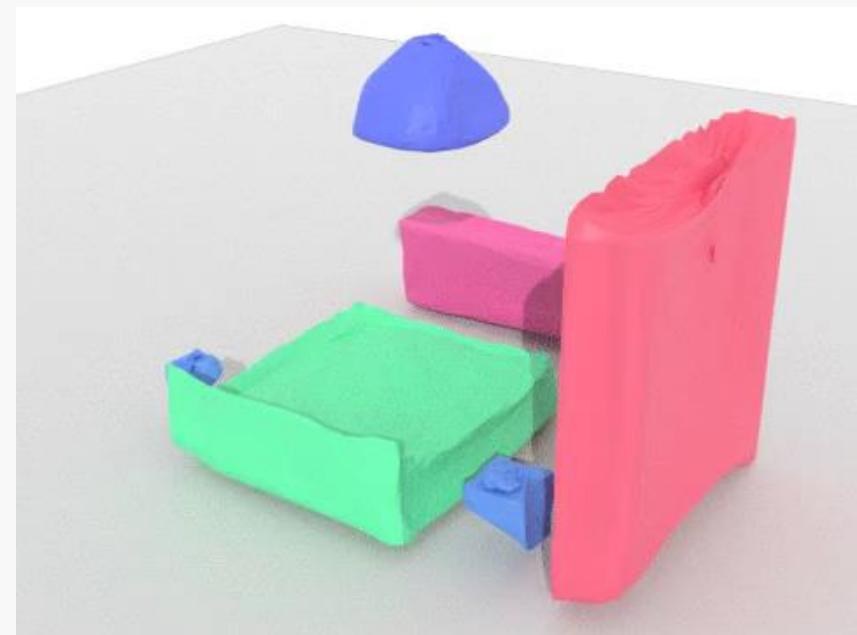
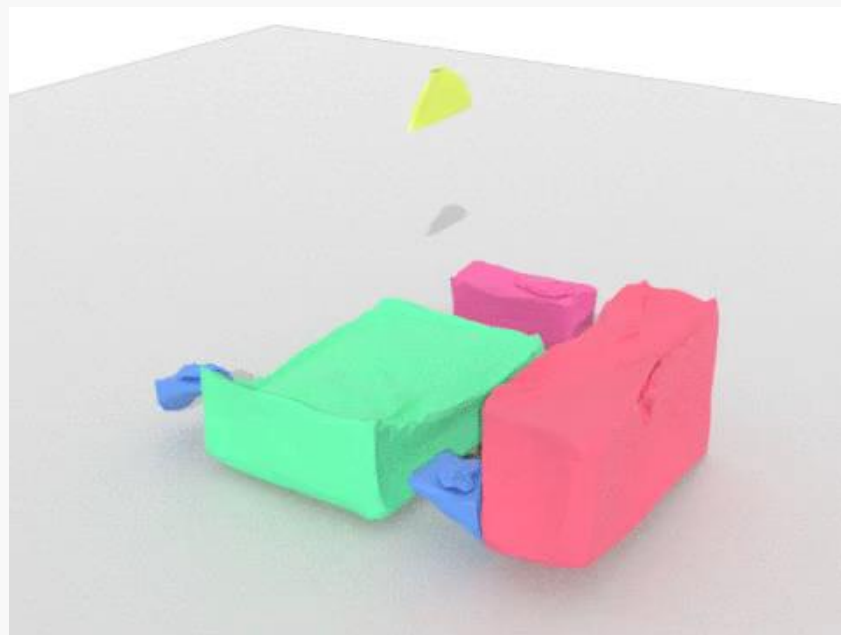
Sample 1



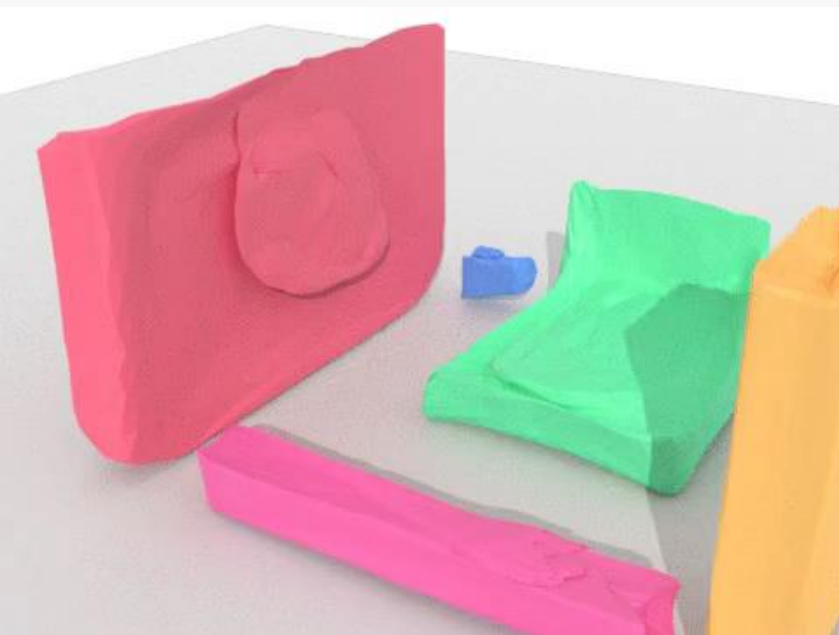
Sample 2



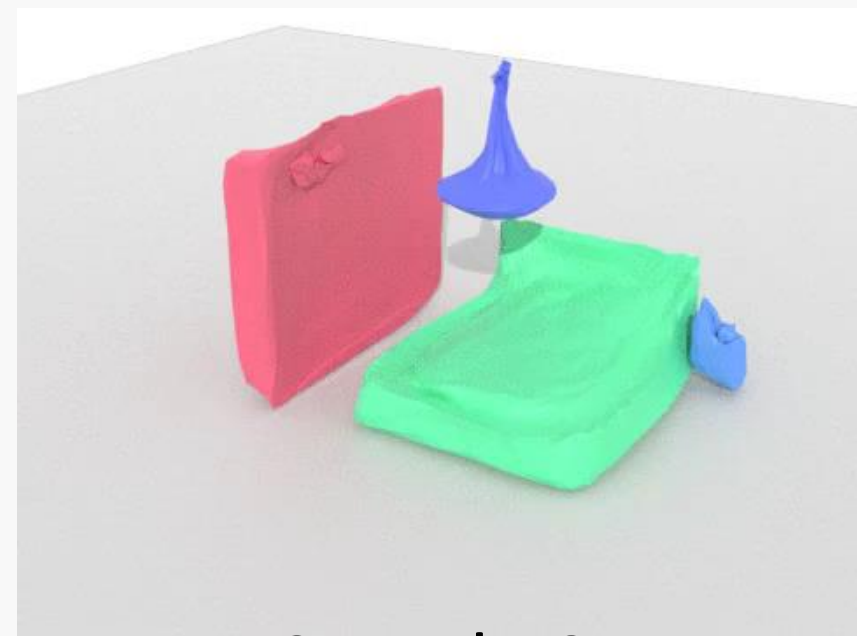
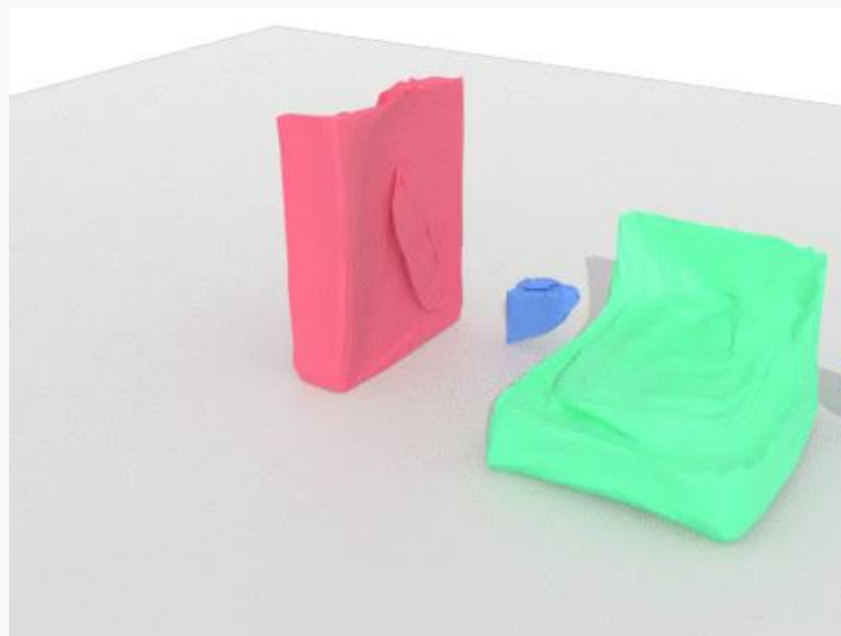
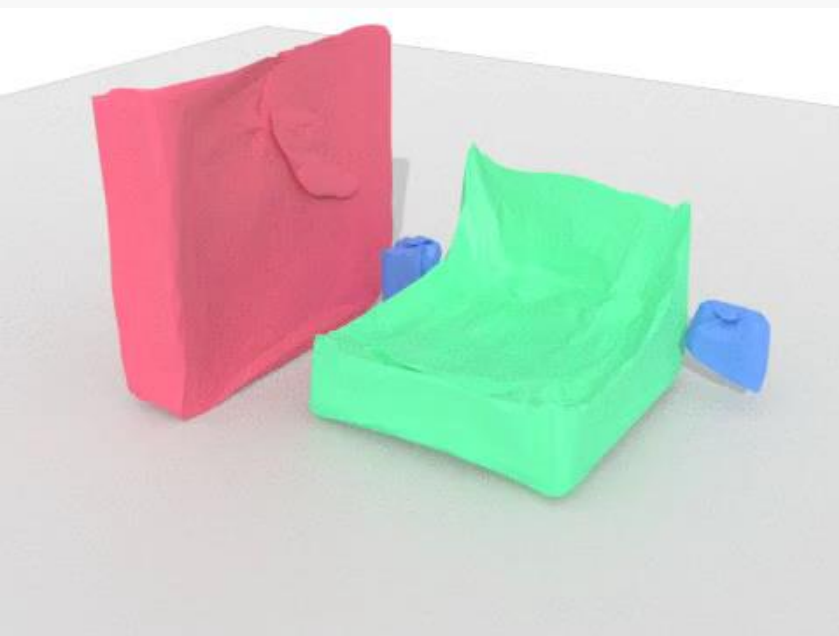
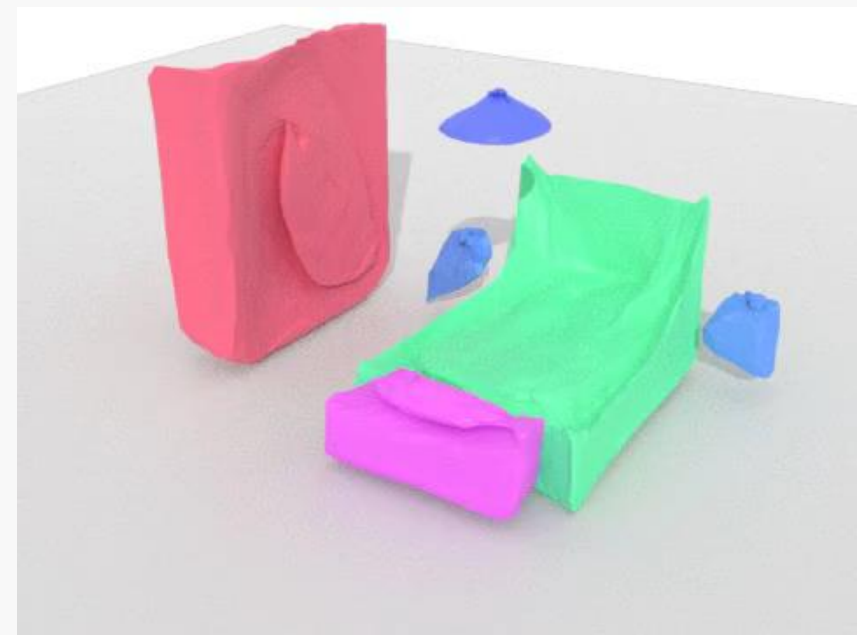
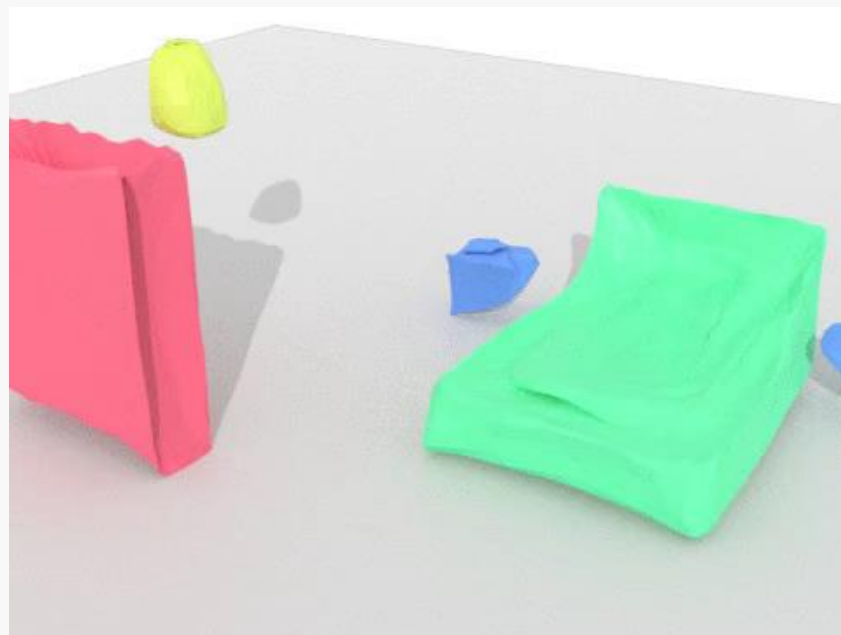
Sample 1



Sample 2



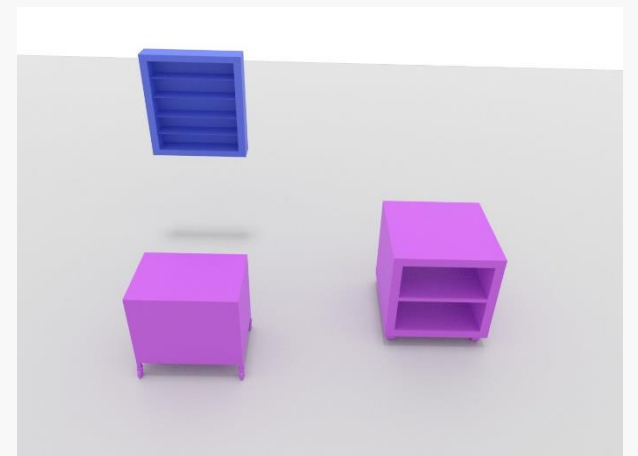
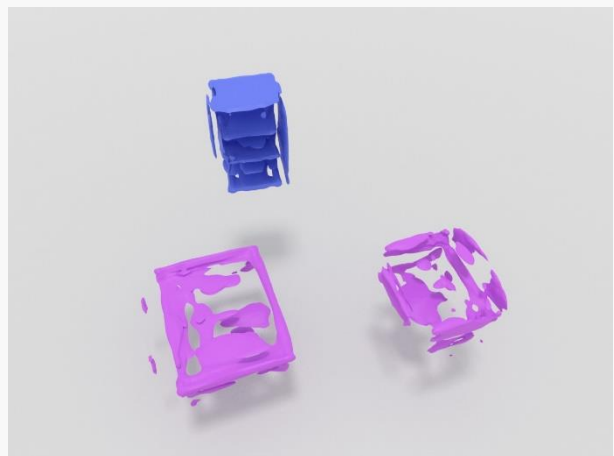
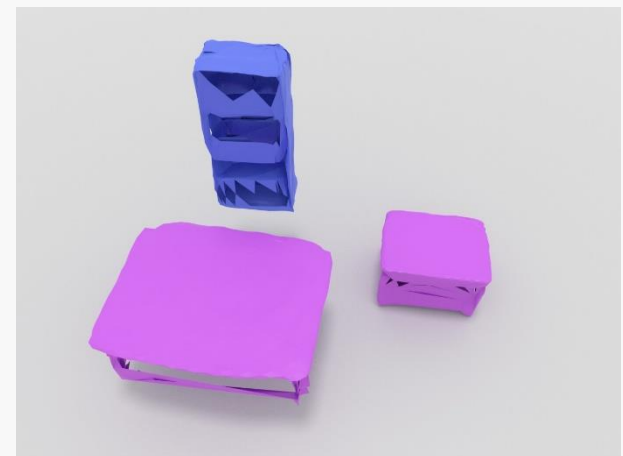
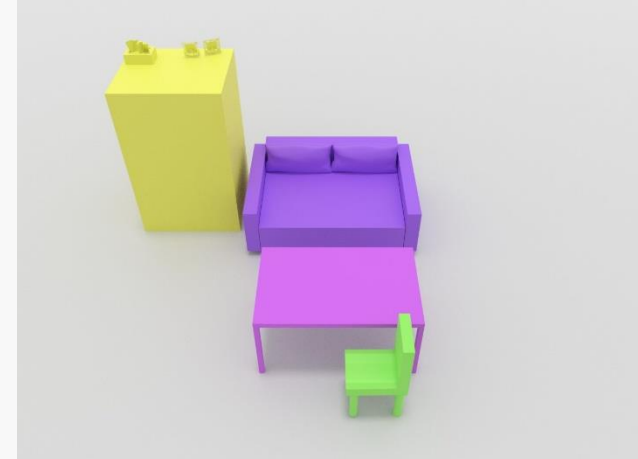
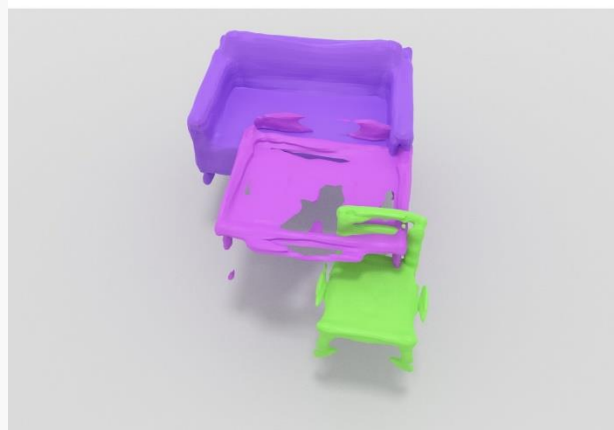
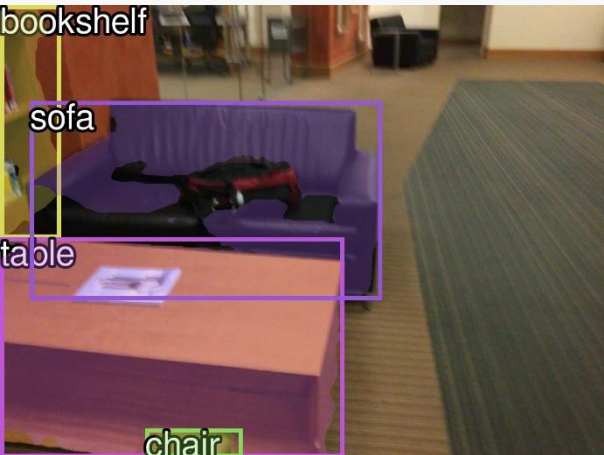
Sample 1



Sample 2



Single-view Reconstruction



Input

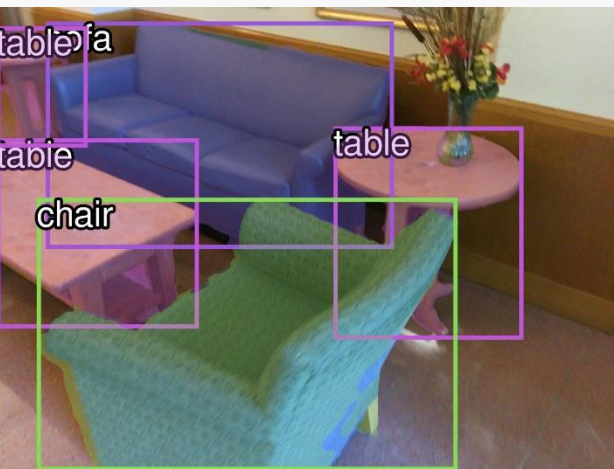
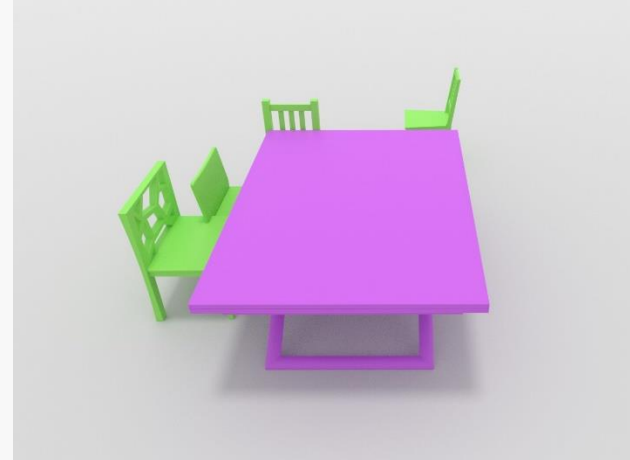
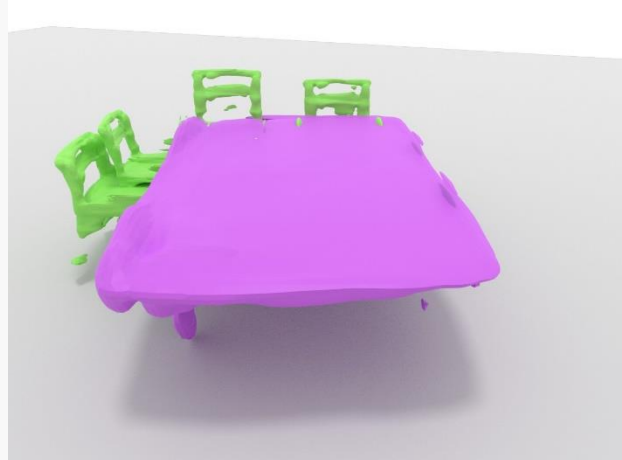
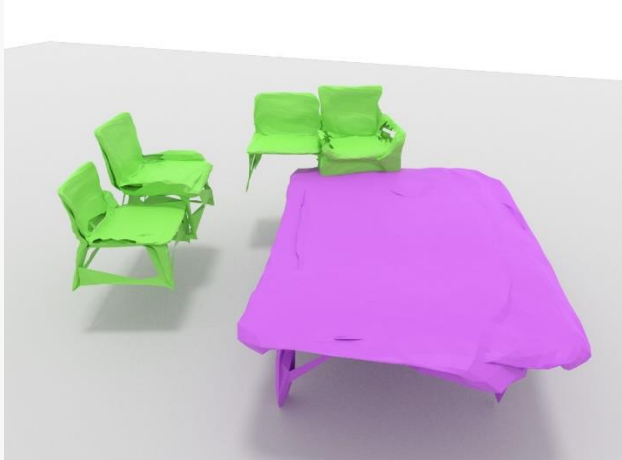
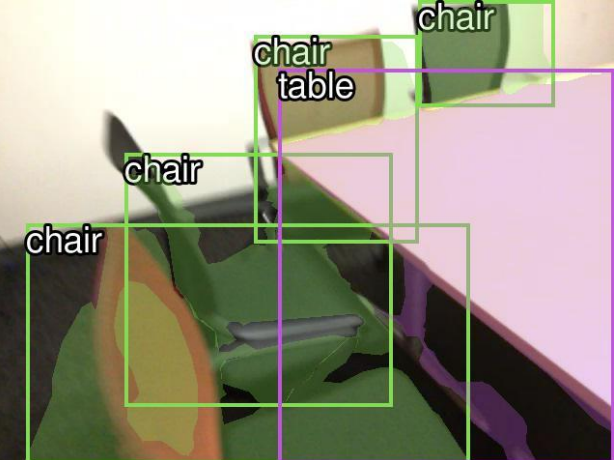
Total3D [1]

Im3D [2]

Ours

[1] Total3DUnderstanding, CVPR'20

[2] Im3D, CVPR'21



Input

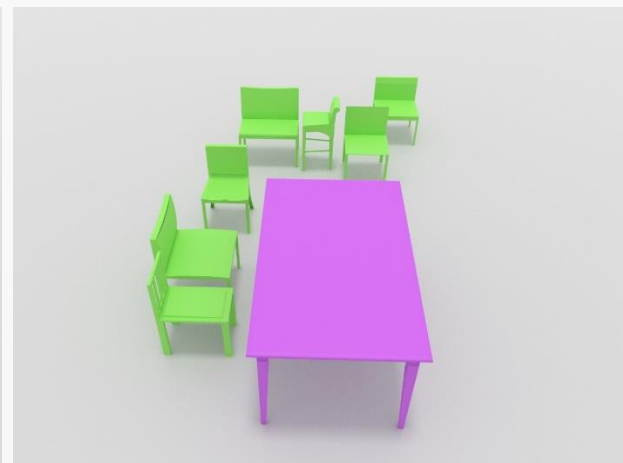
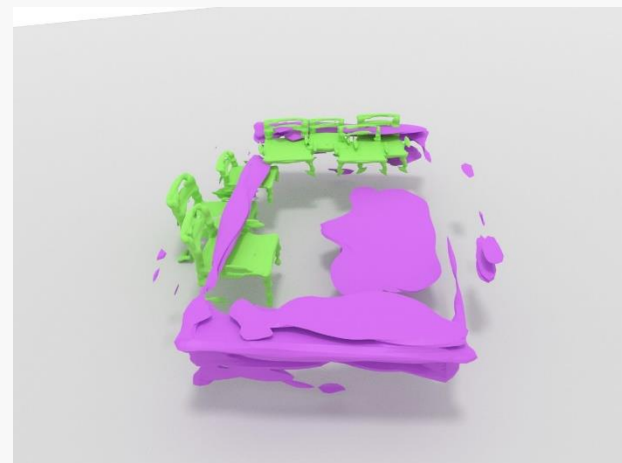
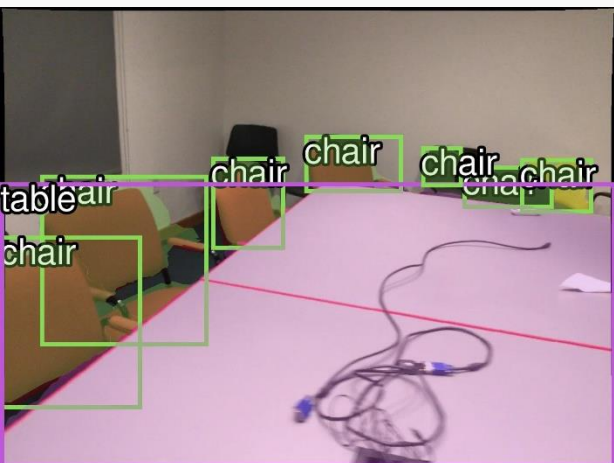
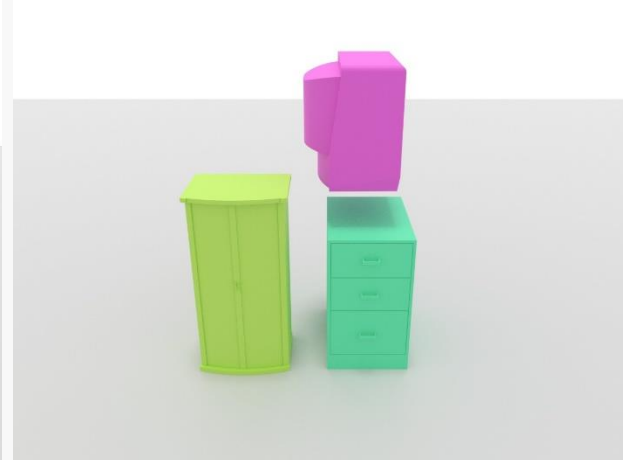
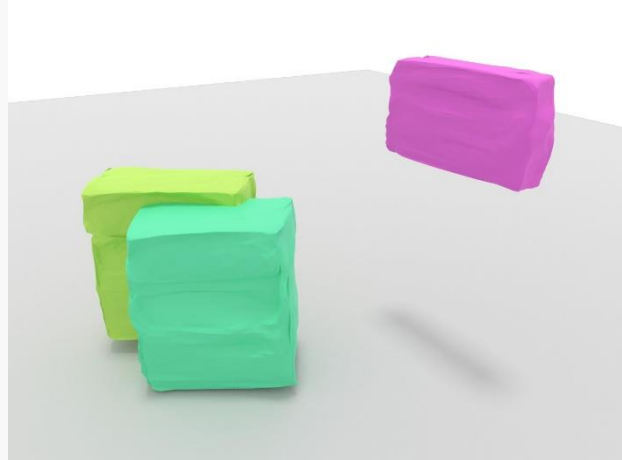
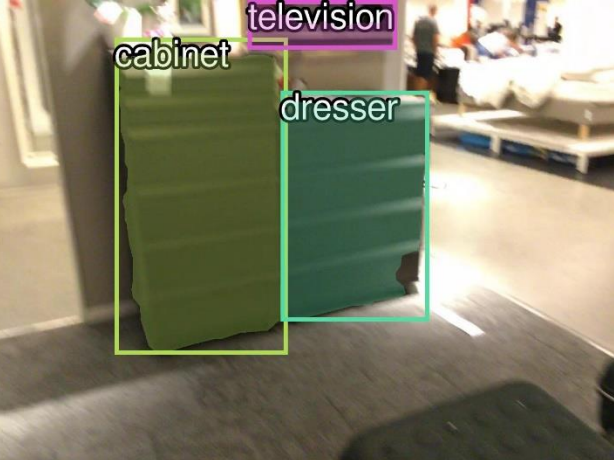
Total3D [1]

Im3D [2]

Ours

[1] Total3DUnderstanding, CVPR'20

[2] Im3D, CVPR'21



Input

Total3D [1]

Im3D [2]

Ours

[1] Total3DUnderstanding, CVPR'20

[2] Im3D, CVPR'21

Learning 3D Scene Priors with 2D Supervision



Yinyu Nie¹, Angela Dai¹, Xiaoguang Han², Matthias Nießner¹

¹ Technical University of Munich

² The Chinese University of Hong Kong (Shenzhen)