2D Amodal Instance Segmentation Guided by 3D Shape Prior

Zhixuan Li[1,2], Weining Ye[2], Tingting Jiang✉[1,2], and Tiejun Huang[2]

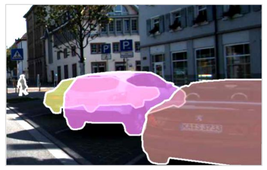
[1] *Advanced Institute of Information Technology, Peking University, Hangzhou, China*

[2] *National Engineering Research Center of Visual Technology, School of Computer Science, Peking University, Beijing, China*

What is amodal instance segmentation?

Predict the complete mask of the occluded instance, including both **visible** and **invisible** regions.

Visible Instance Segmentation

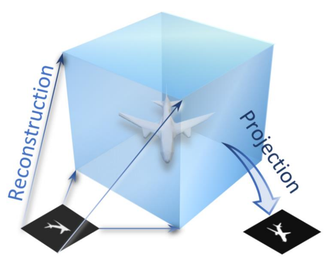


Amodal Instance Segmentation

Observations:

2D amodal mask is the **projection** of 3D model

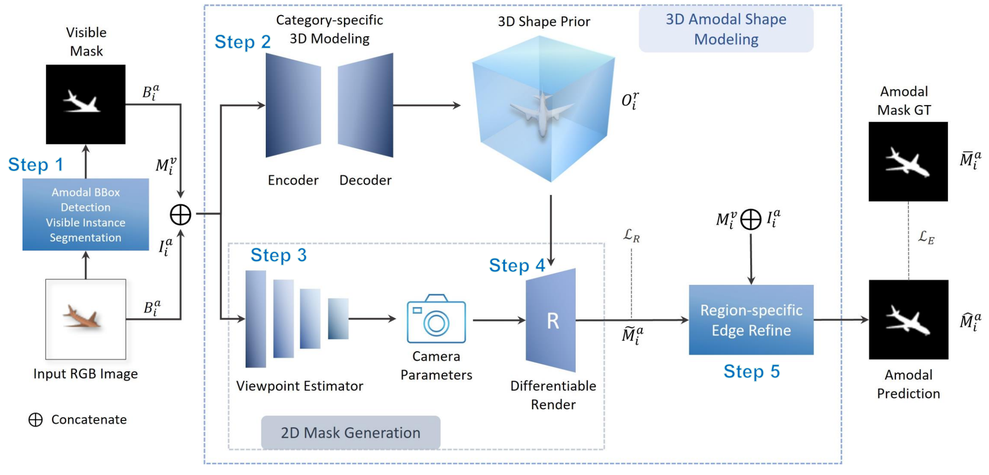
3D model can be **reconstructed** from 2D instances



Our Purpose:

1. Build a bridge between 2D and 3D

2. Use 3D model as shape prior



Our approach:

Step 1: Visible Instance Segmentation

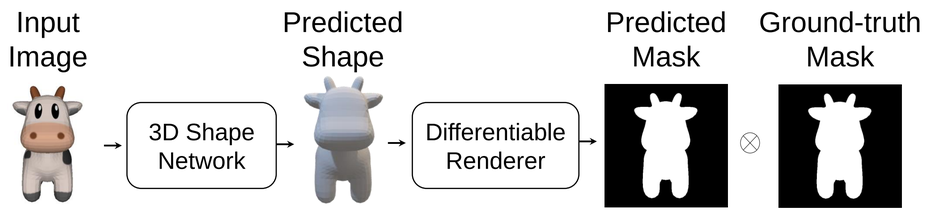
Step 2: Single-view **Unsupervised** 3D Reconstruction

Step 3: Unsupervised Learning for Viewpoint Estimation

Step 4: Differentiable Render for Projecting 3D to 2D

Step 5: Region-specific Edge Refine

What is unsupervised 3D reconstruction



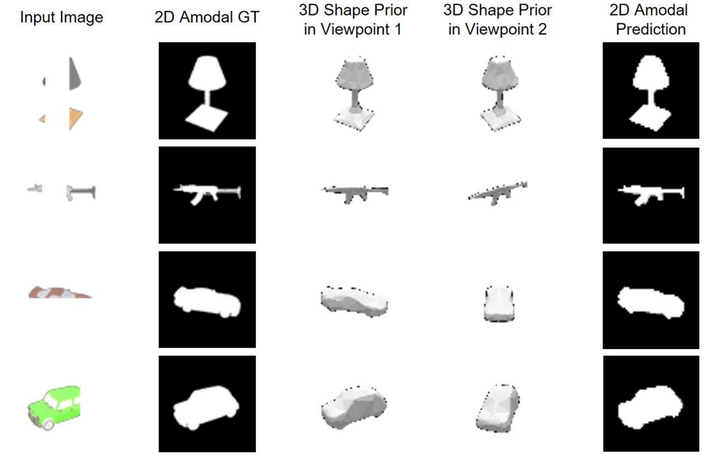
Experiments:

Comparison with SOTA

Methods	Airplane	Bench	Dresser	Car	Chair	Display	Lamp	Speaker	Rifle	Sofa	Table	Phone	Vessel	mIoU
Deocclusion [36]$_{CVPR'20}$	24.9	67.4	45.3	58.8	83.7	78.4	77.9	15.2	48.7	48.1	39.5	23.8	71.9	52.2
Mask-RCNN [9]$_{ICCV'17}$	73.4	66.0	92.4	93.5	**89.3**	**90.0**	77.4	88.5	30.0	86.0	73.1	89.8	80.5	79.2
ORCNN [7]$_{WACV'19}$	71.5	61.1	92.0	92.7	88.8	88.8	79.5	88.7	32.8	85.6	72.5	89.0	80.0	78.7
BCNet [17]$_{CVPR'21}$	73.0	75.1	93.8	89.4	86.6	88.7	**81.6**	90.2	32.8	83.4	77.5	88.7	74.8	78.2
ShapeDict [30]$_{AAAI'21}$	75.2	68.5	93.7	**93.6**	88.4	89.3	78.1	88.6	34.4	87.3	74.8	90.7	80.9	80.3
Ours (no pretrain)	**77.9**	**80.8**	**94.2**	92.8	79.7	87.5	67.8	**90.5**	**69.9**	**90.3**	**86.2**	**92.1**	**81.3**	**83.9**

On the ShapeNet dataset

Qualitative Results



On the ShapeNet dataset

References:

[7] Follmann, P., Konig, R., Hartinger, P., Klostermann, M., Bottger, T.: Learning to see the invisible: End-to-end trainable amodal instance segmentation. *WACV*. 2019

[9] He, K., Gkioxari, G., Dollar, P., Girshick, R.: Mask R-CNN. *ICCV*. 2017

[17] Ke, L., Tai, Y.W., Tang, C.K.: Deep occlusion-aware instance segmentation with overlapping bilayers. *CVPR*. 2021

[30] Xiao, Y., Xu, Y., Zhong, Z., Luo, W., Li, J., Gao, S.: Amodal segmentation based on visible region segmentation and shape prior. *AAAI*. 2020

[36] Zhan, X., Pan, X., Dai, B., Liu, Z., Lin, D., Loy, C.C.: Self-supervised scene deocclusion. *CVPR*. 2020