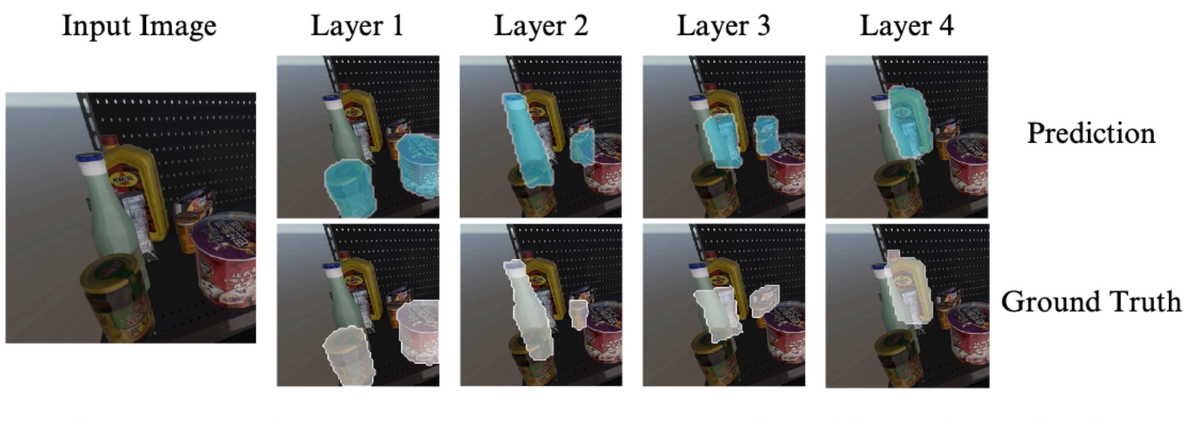
Sequential Amodal Segmentation via Cumulative Occlusion Learning

Jiayang Ao[#], Qiuhong Ke*, and Krista A. Ehinger[#]
The University of Melbourne[#], Monash University*
jaao@unimelb.edu.au

Introduction



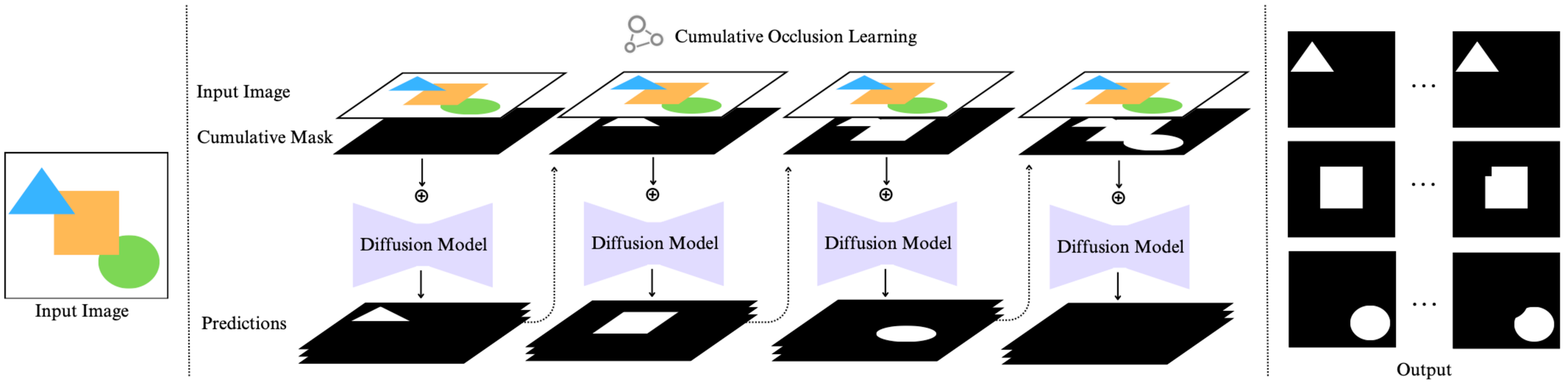
Our approach generates **amodal masks** for objects layer by layer, including both visible and **hidden parts**.

Contributions:
- Handles unlimited layers of occlusion
- Simultaneously predicts amodal masks & occlusion order
- Accounts for uncertainty in occluded regions

Check Our Website

BMVC 2024

Model Architecture



- Human perception uses objects in the foreground to infer occluded objects behind them.
- Similarly, our approach employs a **cumulative mask** that aggregates the masks of previously identified objects.
- This keeps a clear record of areas already segmented and directs **focus to hidden regions**.

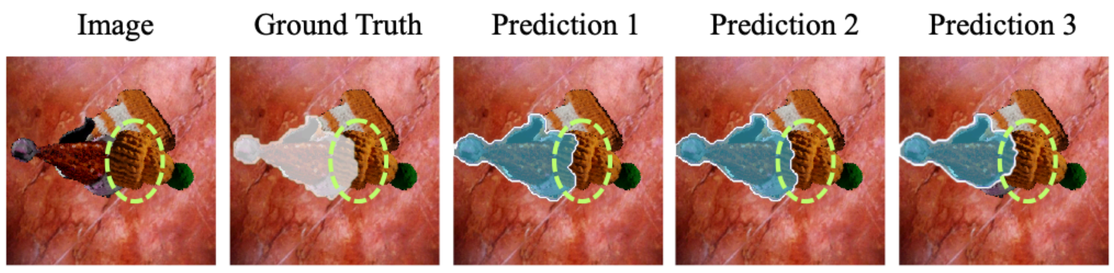
Experimental Results

Dataset	Layer / Method	1 IoU / AP	2 IoU / AP	3 IoU / AP	4 IoU / AP	5 IoU / AP
Intra-AFruit	DIS	89.5 / 90.7	81.6 / 82.6	52.4 / 52.6	9.8 / 12.4	0.5 / 2.0
	Ours	**94.3 / 94.7**	**87.4 / 88.2**	**76.2 / 77.3**	**26.7 / 27.6**	**7.2 / 7.4**
ACOM	DIS	31.6 / 34.8	26.6 / 28.7	1.6 / 10.2	0.2 / 6.0	0.1 / 2.5
	Ours	**57.1 / 57.8**	**44.8 / 45.4**	**28.8 / 30.0**	**12.2 / 14.2**	**1.9 / 3.6**
MUVA	DIS	68.2 / 71.5	19.3 / 27.3	0.1 / 8.6	0.2 / 3.4	0 / 0.5
	Ours	**77.0 / 79.3**	**48.7 / 51.2**	**25.4 / 27.8**	**8.5 / 9.9**	**1.0 / 1.1**

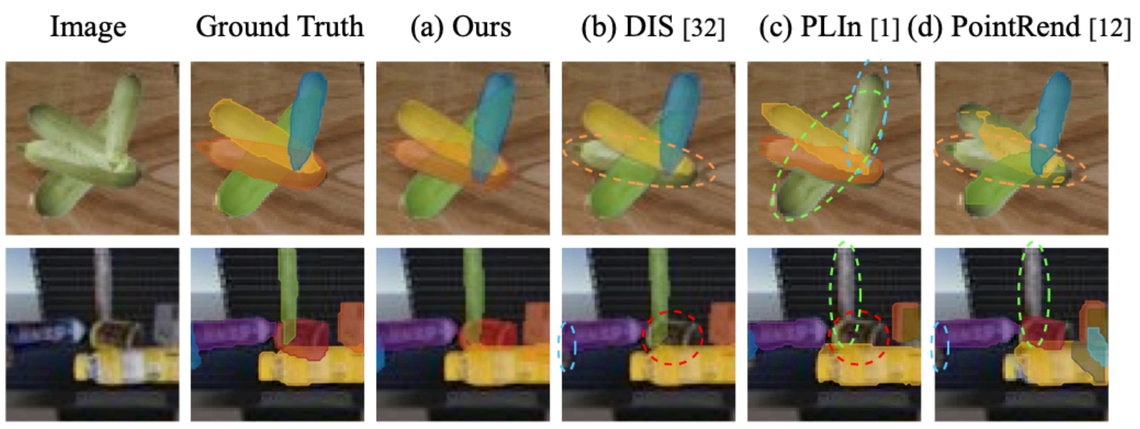
Comparison with diffusion-based segmentation model [32] without cumulative occlusion learning. Ours exhibits great improvement in **deeper-layer scenes**.

Method	Intra-AFruit AP w/ Layer	Intra-AFruit AP w/o Layer	ACOM AP w/ Layer	ACOM AP w/o Layer	MUVA AP w/ Layer	MUVA AP w/o Layer
PointRend [12]	N/A	70.9	N/A	22.0	N/A	38.9
AISFormer [30]	N/A	70.4	N/A	34.9	N/A	49.7
PLIn [1]	42.2	78.9	3.9	17.0	16.3	47.3
Ours	**84.6**	**92.6**	**45.4**	**65.5**	**53.1**	**55.7**

Comparison with category-specific segmentation models. Despite being trained on class-agnostic data, ours surpasses those trained on category-labelled data.



Our approach considers the **diversity** of possible amodal masks, especially for occluded regions (indicated by dashed circles).



-- Dashed circles indicate objects that missed being predicted.

Our approach performs better at segmenting objects/providing more plausible amodal masks than others.