Beyond the Vision: Amodal Object Segmentation for Lab-Driven Applications

Ricardo Ornelas, Riley Menter, Shreyangshu Bera, Varsha Vinod

UC MERCED | UC RIVERSIDE

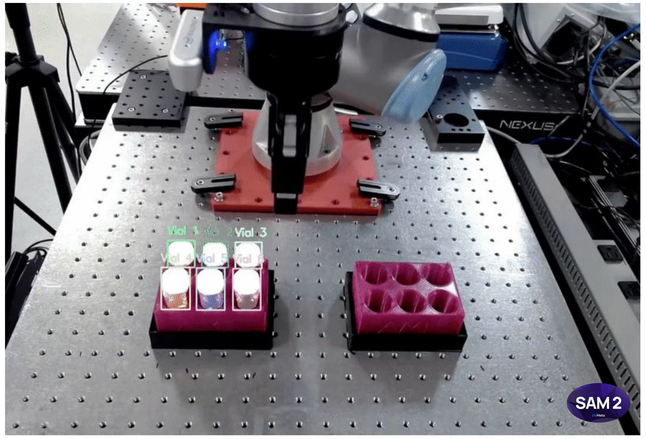
Lawrence Livermore National Laboratory

Problem Statement

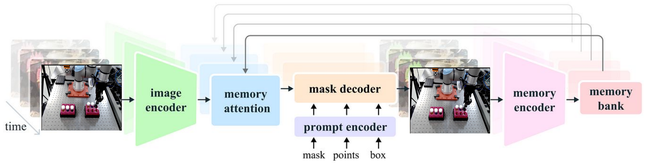
Given an RGB scene and the modal mask of an object, the objective is to reconstruct the object's complete amodal RGB appearance of the occluded regions. We address this task in both static images and dynamic video sequences.

SAM2 Model

Meta's SAM2 model segments objects in videos using prompts, applies consistent visuals (boxes, titles, modal masks)



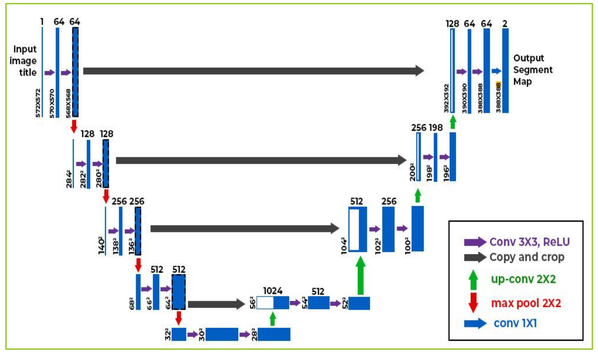
How does it work?



SAM2 segments objects across frames using memory and user prompts

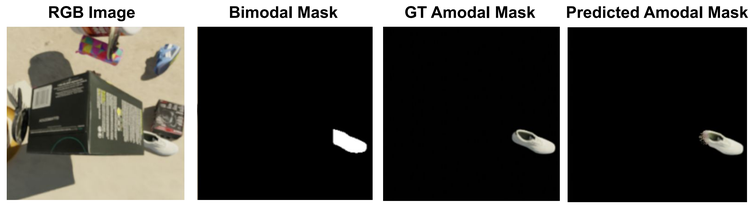
Model Architecture

UNet: symmetric encoder-decoder with skip connections

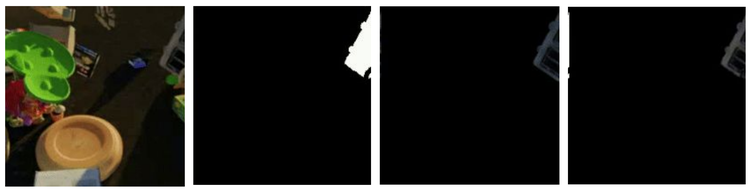


Methods in Training and Evaluation

- Pixel imbalance during training and evaluation
- Imbalance between black and white/rgb pixels lead to model biasing
- Black images to achieve low loss and high accuracy

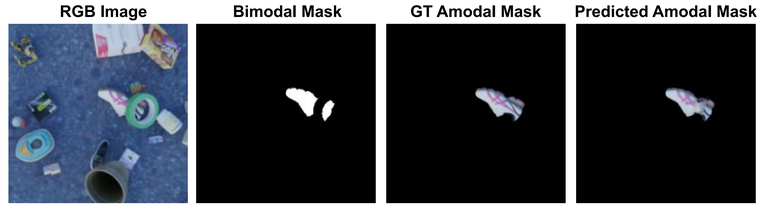


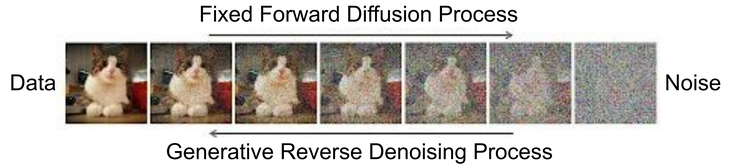
IoU: 0.962, Dice: 0.981, Pixel Accuracy: 0.999



IoU: 0.650, Dice: 0.980, Pixel Accuracy: 0.790

Diffusion Model (Pix2gestalt)





Future Work

- Finetune decoder of SAM2 for amodal mask
- Object re-identification (building off SAM2)
- Use diffusion model for Videos
- Look into efficient data acquisition methods for training

Conclusion

- Explore different training architectures
- Fine tuning for overcoming pixel imbalances
- With more time and effort, we can take a step towards self-operating labs

G6 Members