Neural Volumetric Object Selection

Zhongzheng Ren, Aseem Agarwala†, Bryan Russell†, Alexander G. Schwing*, Oliver Wang†*

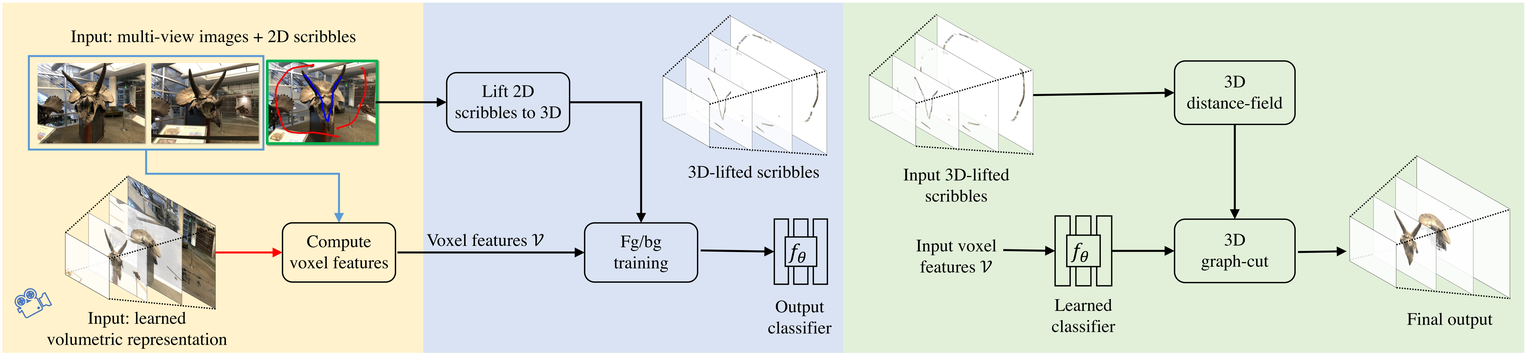
*UIUC †Adobe Research

Overview and contributions

Contributions: (1) approach for 3D object selection in neural volumes (e.g., MPI, NeRF)
(2) dataset for evaluation

2D images & scribbles 3D segmentation Novel views

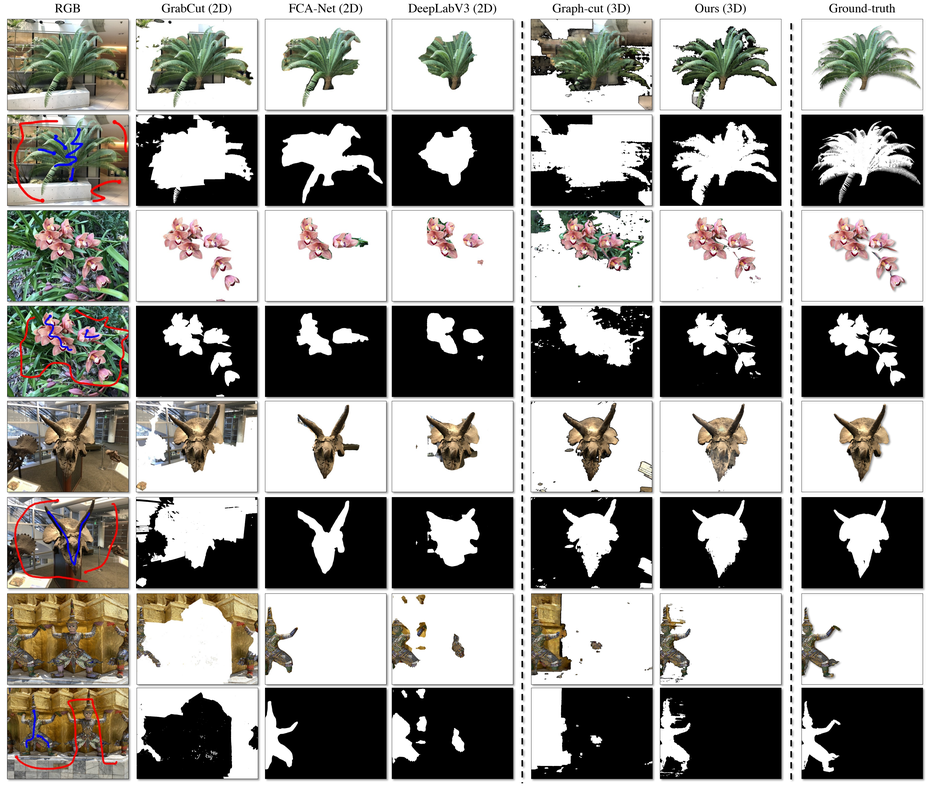
Approach



- **Input:** learned 3D volume $\mathscr{V} = \{\mathbf{v}_p\}$ where $\mathbf{v}_p$ is the **voxel feature embedding**
- **Task:** 3D foreground/background segmentation $f_\theta(\mathbf{v}_p) : \mathbb{R}^C \rightarrow [0,1]$
  - f_θ is implemented as MLP
- **Voxel feature embedding:** $\mathbf{v}_p = [\mathbf{v}_p^{MVS}; \mathbf{v}_p^{IBR}; \mathbf{v}_p^{XYZ}]$
  - $\mathbf{v}_p^{MVS}$: multi-view image feature embedding, following MVSNet
  - $\mathbf{v}_p^{IBR}$: discretized IBR feature (NeX for MPI / PlenOctrees for NeRF)
  - $\mathbf{v}_p^{XYZ}$: 3D positional encoding feature
- **Post-processing**: 3D graph-cut

https://jason718.github.io/nvos

Results



2D mask evaluation

Datasets	LLFF		Shiny	
Metrics	Acc.	IoU	Acc.	IoU
Graph-cut	88.6	59.0	86.1	48.0
GrabCut	78.1	49.0	66.2	31.1
DeepLabV3	91.5	56.6	88.6	50.4
DEXTR	89.7	34.5	59.6	40.4
FCA-Net	88.3	62.7	87.9	58.5
Graph-cut (3D)	73.6	39.4	78.3	32.4
Ours	92.0	70.1	90.7	69.3

Novel-view object rendering

Datasets	LLFF			Shiny		
Metrics	SSIM	PSNR	LPIPS	SSIM	PSNR	LPIPS
Graph-cut (3D)	0.600	15.03	0.415	0.454	12.83	0.477
Ours	0.767	18.40	0.213	0.612	15.73	0.319

Visual ablation