# WALT: Watch And Learn 2D Amodal Representation from Time-lapse Imagery


N. Dinesh Reddy, Robert Tamburo, and Srinivasa G. Narasimhan

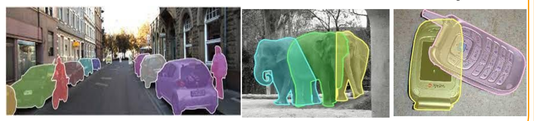
Carnegie Mellon University


Project Page: walt.cs.cmu.edu

## Human Amodal Supervision is Hard

**Goal: Learning Automatic 2D Amodal Representation**

**Issues: Current datasets are insufficient and imprecise**



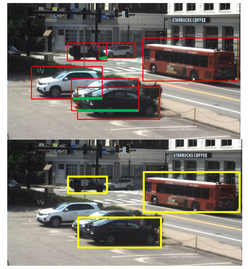
*Human amodal annotations are sparse and imprecise [1][2]*

## CG for Amodal leaves Domain Gap

*Synthetic datasets have domain gap with real data[4]*

## Unoccluded Object Detection is Much Easier

*Classify unoccluded and occluded objects using planar based IOU*

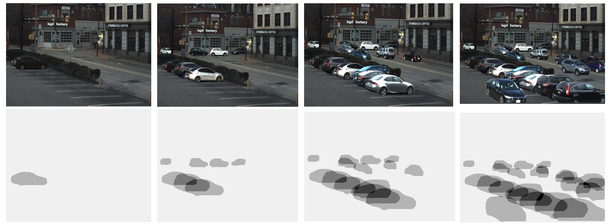


## Supervision From Time-Lapse

*Input Time-Lapse Video*

*Unoccluded Object Detected*

*Mined Unoccluded Objects*



*Generate Realistic GroundTruth using Compositing*

## Clip-Art WALT(CWALT)

**Unoccluded Objects composited in layers onto same scene**

*BackGround Image for each camera of WALT* | *Clip-Art Image mined Unoccluded objects* | *Amodal Ground-Truth Segmentation*

## WALT Dataset

**Consists of 12(4K and 1080p) and expanding cameras capturing data over a year in short bursts.**

| Type | #Cameras | #FPS | #Duration | #Images | #UnOcc Obj |
|---|---|---|---|---|---|
| Ours | 6 | 60 | 3s | 10 Mil | 21 Mil |
| Youtube | 6 | 30 | 3s | 5 Mil | 15 Mil |
| Total | **12** | - | - | **15 Mil** | **36 Mil** |

*Our Camera*

*Youtube Live Stream*

## Comparisons

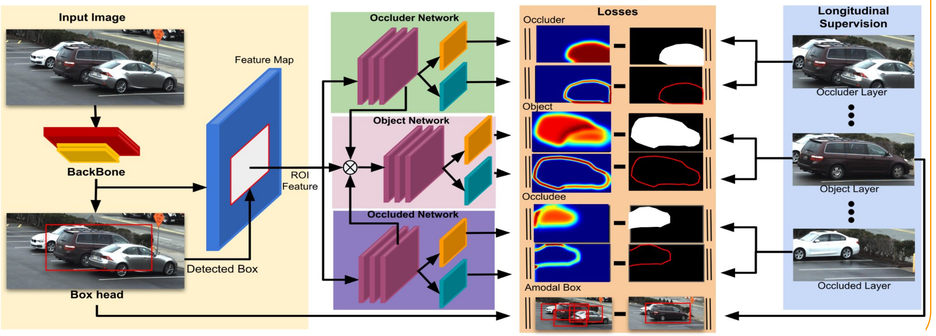
| Method | KINS [1] | COCOA [3] | SWALT γ=0.01 | SWALT γ=0.5 |
|---|---|---|---|---|
| ASN[1] | 24.9 | 29.6 | 79.4 | 76.91 |
| BCN[2] | 27.3 | 32.7 | 82.79 | 77.44 |
| Ours | **27.9** | **33.1** | **83.6** | **78.2** |

| Method | CWALT | SWALT γ=0.01 | SWALT γ=0.5 |
|---|---|---|---|
| ASN[1] | 66.1 | 83.1 | 81.9 |
| BCN[2] | 73.2 | 89.9 | 88.3 |
| Ours | **75.3** | **92.1** | **91.7** |

*CWALT = Clip-Art WALT Dataset*
*SWALT = Stationary Object Based WALT Dataset*

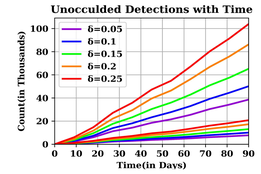
## WALT Amodal Network

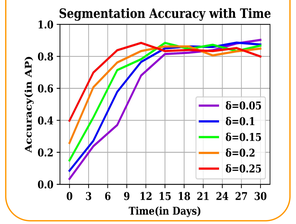
**The Network learns Amodal object- Occluder and occluded objects interactions**



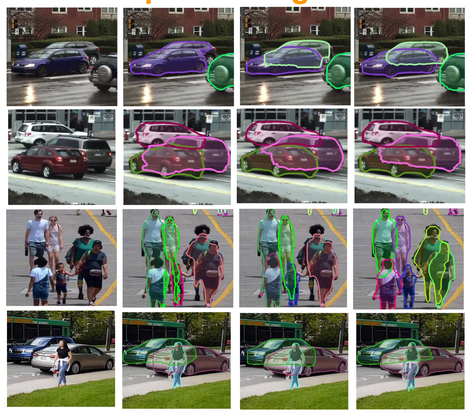

## Speed-Up Training

**Unoccluded object tracking improves training speed**



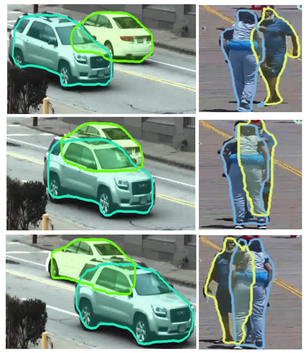




## WALT Improves Segmentation



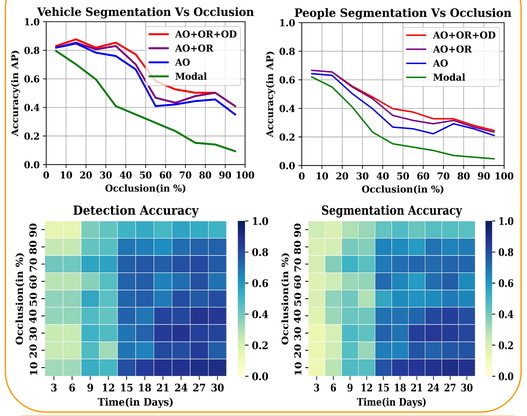
*Input Image* | *ASN[1] (CVPR 19)* | *BCN[2] (CVPR 21)* | *WALT-NET (Ours)*

## WALT Improves Tracking



*Accurate amodal segmentation enhances tracking accuracy*

## Robust to Occlusions





## Acknowledgements

**This work was sponsored in part by an ARL Grant W911QX20F016, NSF CNS-2038612, and DOT RITA Mobility-21 Grant 69A3551747111, and a Qualcomm Innovation Fellowship.**



*[1] Lu et al. Amodal instance segmentation with kins dataset. CVPR 19*
*[2] Lei et al. Deep occlusion segmentation overlapping bilayers. CVPR 21*
*[3] yan et al. Semantic amodal segmentation. CVPR 17*
*[4] Hu et al. SAIL-VOS: Semantic Amodal Instance Level VOS. CVPR 19*