KI DATA TOOLING
The Data Kit for Automotive AI

Towards an End-to-end Amodal Video Instance Segmentation Challenge

Jasmin Breitenstein,
Tim Fingscheidt | TU BS

ML Function Development

Motivation

Amodal perception allows to hallucinate the full shape of occluded objects. In automated driving, knowing at all times the precise location of all instances is safety-relevant. Prior art [1,2] considers amodal segmentation only on images, however, temporal information is an important cue for solving partial and full occlusions. Our new baseline for end-to-end amodal video instance segmentation (VIS) is to our knowledge the first end-to-end trainable method that provides not only image-level, but also video-level results for amodal VIS.

Experimental Setup

We train our amodal VIS network VATrack [3] on an adapted version of the SAIL-VOS [1] dataset coined SAIL-VOS-cut. This adaptation gets rid of jump cuts in the video but does not affect the video content.

Results on Image Level and on Video Level

Method	DNN	V	A	AP	AP_{50}	AP_{50}^{P}	AP_{50}^{L}	AP_{50}^{M}	AP_{50}^{S}
MaskAmodal [1]	?		✓	13.0	23.0	24.3	16.7	36.6	21.5
MaskJoint [1]	?	✓	✓	14.1	24.8	24.3	18.9	37.8	21.5
MaskAmodal*	RX101		✓	16.3	25.6	27.4	17.1	35.2	24.2
MaskJoint*	RX101	✓	✓	16.7	25.6	26.9	17.3	33.0	22.3
AmodalTrack (MT)	RX101		✓	15.9	25.7	24.9	17.8	36.8	22.8
VATrack (MT)	RX101	✓	✓	16.4	26.0	24.9	18.0	**38.6**	22.5
AmodalTrack (QD)	RX101		✓	17.8	27.4	29.2	18.6	34.7	26.8
VATrack (QD)	RX101	✓	✓	**18.3**	**28.6**	**29.7**	**20.1**	38.1	**26.9**

*Table 1: Amodal image-level results on SAIL-VOS-cut for image-based and video-based methods with ✓ indicating whether visible (V) or amodal (A) masks are predicted. Best results in **bold**. (© TUBS)*

Table 1 shows that temporal context can improve results and that joint prediction mostly improves the image-level results.

Method	DNN	V	A	AP	AP_{50}	AP_{50}^{P}	AP_{50}^{L}	AP_{50}^{M}	AP_{50}^{S}
AmodalTrack (MT)	RX101		✓	2.4	3.1	3.8	1.7	3.8	1.4
VATrack (MT)	RX101	✓	✓	2.3	3.1	3.8	1.7	3.7	1.5
AmodalTrack (QD)	RX101		✓	13.1	20.5	21.0	10.7	29.4	14.7
VATrack (QD)	RX101	✓	✓	**14.1**	**22.3**	**22.0**	**12.8**	**32.8**	**15.6**

*Table 2: Amodal video-level results on SAIL-VOS-cut for video-based methods with ✓ indicating whether visible (V) or amodal (A) masks are predicted. Best results in **bold**. (© TUBS)*

Table 2 shows that also on video-level the joint prediction improves amodal VIS results.

Qualitative Results on SAIL-VOS

Figure 1: Qualitative results of VATrack (bottom) compared to MaskJoint [4]. VATrack detects and tracks all instances consistently across frames, while MaskJoint cannot exploit temporal context and thus, fails to detect severe occlusions. (© TUBS)

Towards a Challenge for Amodal VIS on Automotive Data

We aim to generate amodal data using the CARLA simulator [3]. We use the custom depth stencil to visualize occlusions. Additionally, we extend the technique of Bogdoll et al. [4] to generate deterministic trajectories.

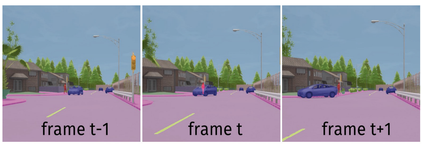

Figure 2: First amodal example data from the CARLA simulator with an occluded pedestrian in frame t visualized by the amodal segmentation (© TUBS)

Conclusions

VATrack is the first end-to-end amodal VIS method. We show that exploiting temporal context can improve amodal segmentation quality on both image- and video-level. We show works towards a challenge for amodal VIS on automotive data using the CARLA simulator.

References:

[1]: Y.-T. Hu, H.-S. Chen, K. Hui, J.-B. Huang, A. G. Schwing, "SAIL-VOS: Semantic Amodal Instance Level Video Object Segmentation - A Synthetic Dataset and Baselines", in Proc. of CVPR 2019
[2]: L. Ke, Y.-W. Tai, C.-K. Tang, „Deep Occlusion-Aware Instance Segmentation with Overlapping BiLayers", in Proc. of CVPR 2021
[3]: J. Breitenstein, K. Jin, A. Hakiri, M. Klingner, T. Fingscheidt, „End-to-end Amodal Video Instance Segmentation", in Proc. of BMVC-Workshops 2023
[4]: A. Dosovitskiy, G. Ros, F. Codevilla, A. Lopez, V. Koltun, „CARLA: An Open Urban Driving Simulator", in Proc. of CoRL 2017
[5]: D. Bogdoll, S. Guneshka, J. M. Zöllner, „One Ontology to Rule Them All: Corner Case Scenarios for Autonomous Driving", in Proc. of ECCV-Workshops 2022

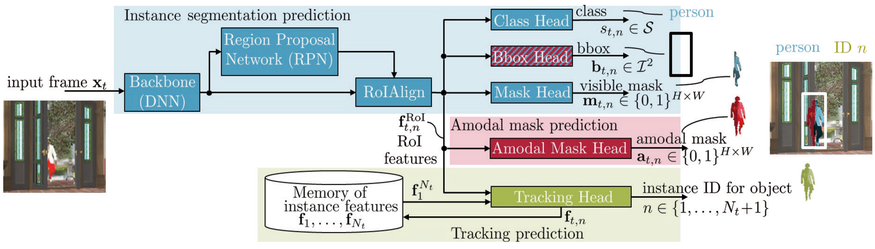

Figure 3: Our proposed VATrack method for end-to-end amodal video instance segmentation can simultaneously predict amodal and visible instance masks, while also tracking the instances throughout the sequence. The instance segmentation is based on Mask R-CNN, the tracking is based on QDTrack and MaskTrack R-CNN. (© TUBS)

Partners

External partners

For more information contact:
j.breitenstein@tu-bs.de
t.fingscheidt@tu-bs.de

KI Data Tooling is a project of the KI Familie. It was initiated and developed by the VDA Leitinitiative autonomous and connected driving and is funded by the Federal Ministry for Economic Affairs and Climate Action.

Supported by:
Federal Ministry for Economic Affairs and Climate Action
on the basis of a decision by the German Bundestag