Context-Aware Relative Object Queries to Unify Video Instance and Panoptic Segmentation

Anwesa Choudhuri, Girish Chowdhary, Alexander Schwing, University of Illinois at Urbana-Champaign

https://anwesachoudhuri.github.io/ContextAwareRelativeObjectQueries/

JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

Goal

General and seamless approach for

Video Instance Segmentation

Multi-Object Tracking and Segmentation

Video Panoptic Segmentation

Prior Work

- Absolute positional encodings [1, 2, 8]
- Process full video at once [2] or post-processing to merge frames [1]
- Two types of object queries with heuristic post processing [8]

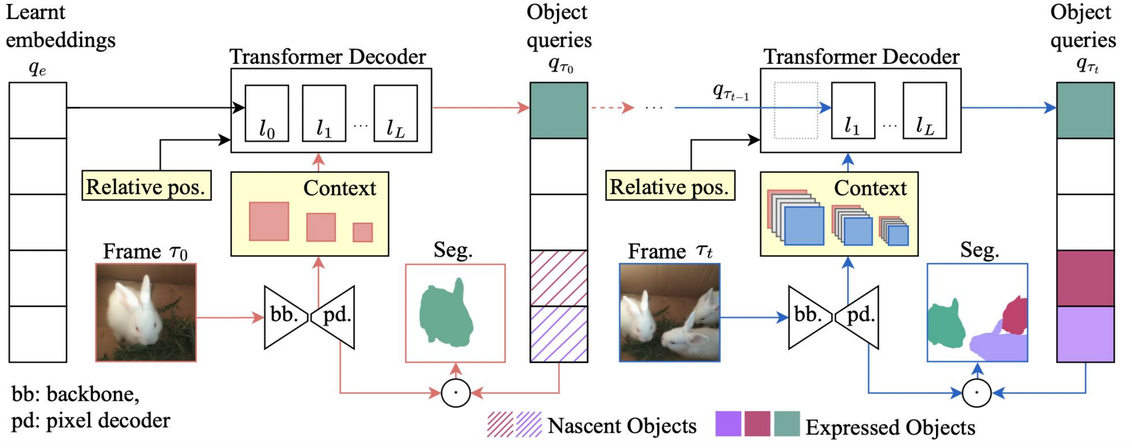
Our Work

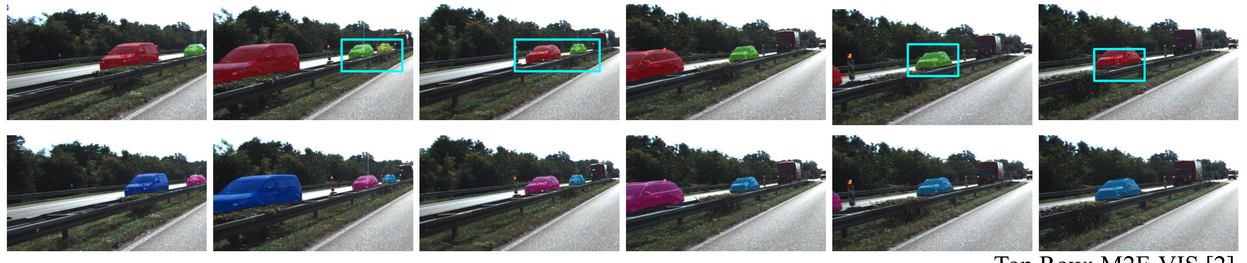
- Relative positional encodings
- Frame-by-frame, but no heuristics or post processing
- Only one type of object query

References

[1] Huang et al., NeurIPS 2022
[2] Cheng et al., arXiv 2022
[3] Qiao et al., CVPR 2021
[4] Kim et al., CVPR 2020
[5] Choudhuri et al., ICCV 2021
[6] Xu et al., ECCV 2020
[7] Meinhardt et al., CVPR 2022

Context-Aware Relative Object Queries





Top Row: M2F-VIS [2]
Bottom Row: Our Approach

Ablation

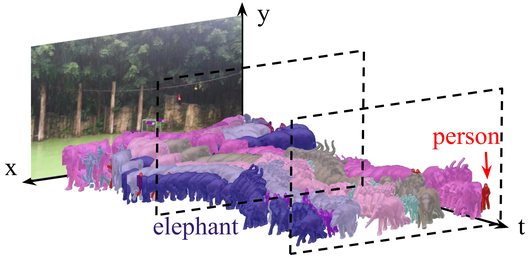
KITTI-MOTS	Car			Pedestrian		
Method	HOTA	DetA	AssA	HOTA	DetA	AssA
Ours (Rel. pos.)	**83.2**	**84.5**	**85.0**	**64.1**	**64.4**	**63.7**
Ours (Abs. pos.)	70.0	78.6	62.7	52.0	58.0	46.4

OVIS	
Method	AP
Ours	**25.8**
[7]-like prop.	16.9
w/o. prop.	9.8

	MOTS 2020
Method	sMOTSA
Pt.Track [6]	58.1
Tr.Former [7]	58.7
Ours	**61.2**

	KITTI-MOTS	
Method	HOTA (car, ped.)	AssA (car, ped.)
ASMOTS [5]	68.7, 58.8	62.2, 57.6
Pt.Track [6]	61.6, 54.4	48.8, 48.0
Ours	**74.5, 62.2**	**64.0, 58.4**

Results



Method	OVIS (AP)	YTVIS 2021 (AP)	YTVIS 2019 (AP)
MinVIS [1]	25.0	**44.2**	**47.4**
M2F-VIS [2]	17.3	40.6	46.4
Ours	**25.8**	43.3	46.7

Cityscapes-VPS

Method	Depth	VPQ	VPQ_{th}	VPQ_{st}
Vip-DeepLab [3]	✓	63.1	49.5	73.0
VPS-Net [4]		57.5	44.8	66.7
Ours		**63.0**	**48.0**	**72.8**