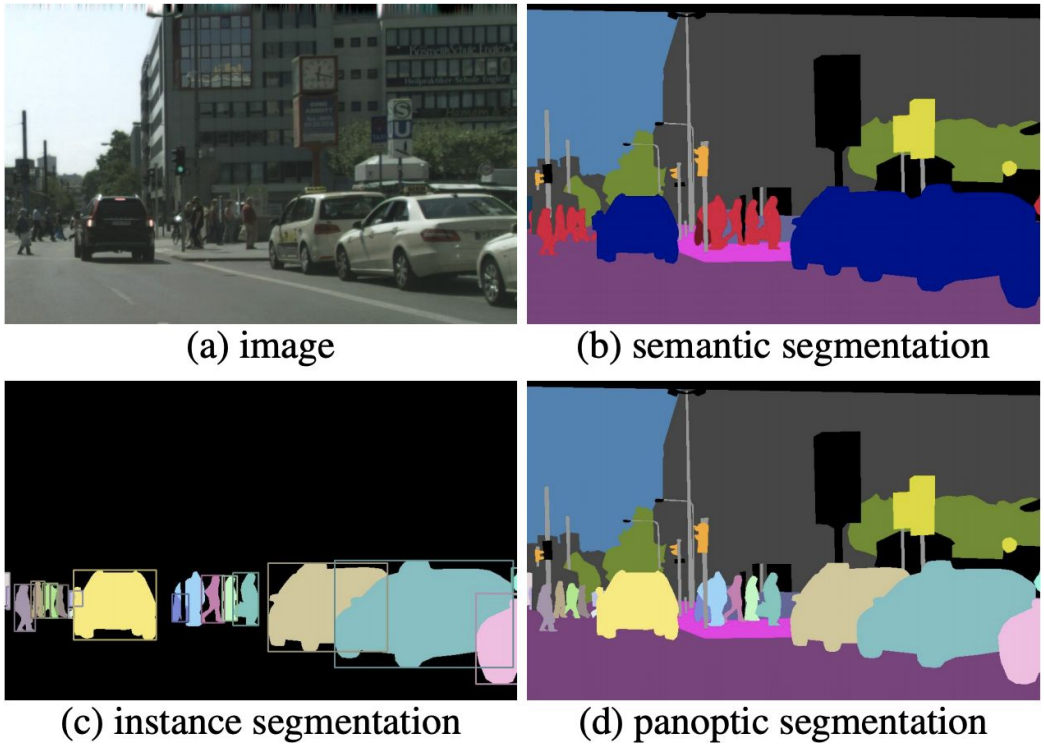
MaX-DeepLab: End-to-End Panoptic Segmentation with Mask Transformers

CVPR 2021

Huiyu Wang, Yukun Zhu, Hartwig Adam, Alan Yuille, Liang-Chieh Chen

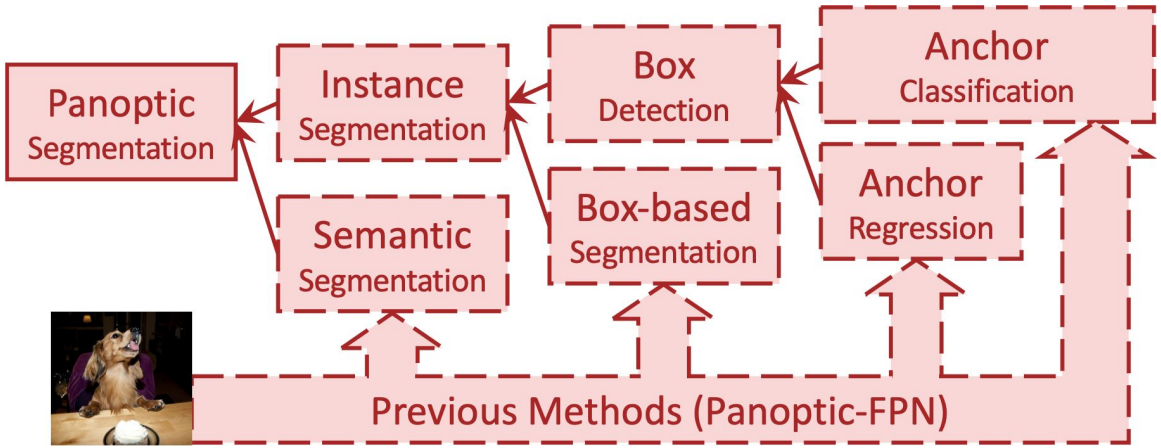
Johns Hopkins University, Google Research

Panoptic Segmentation



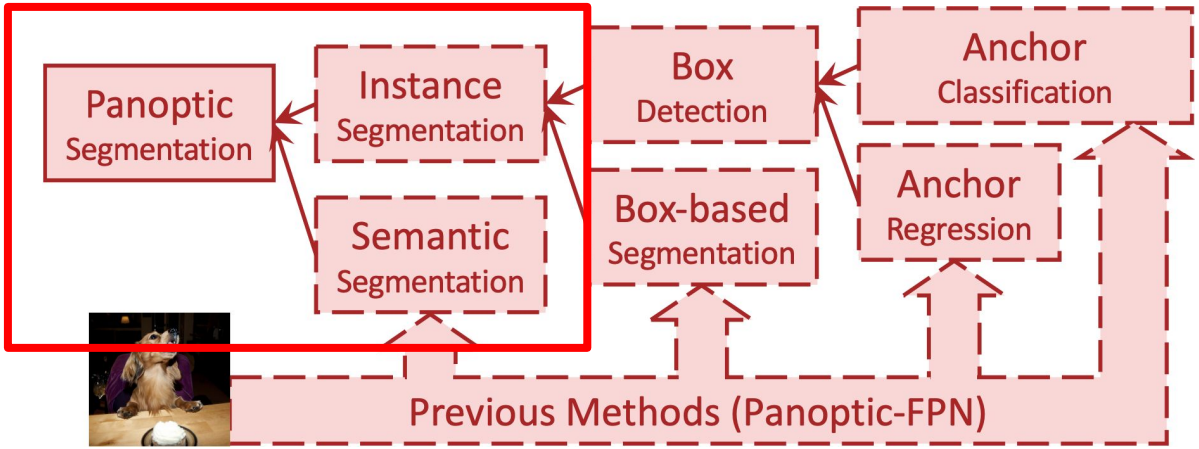
Kirillov, A., et al. Panoptic Segmentation. CVPR 2019.
Cordts, M., et al. The cityscapes dataset for semantic urban scene understanding. CVPR 2016.

Surrogate Sub-Tasks



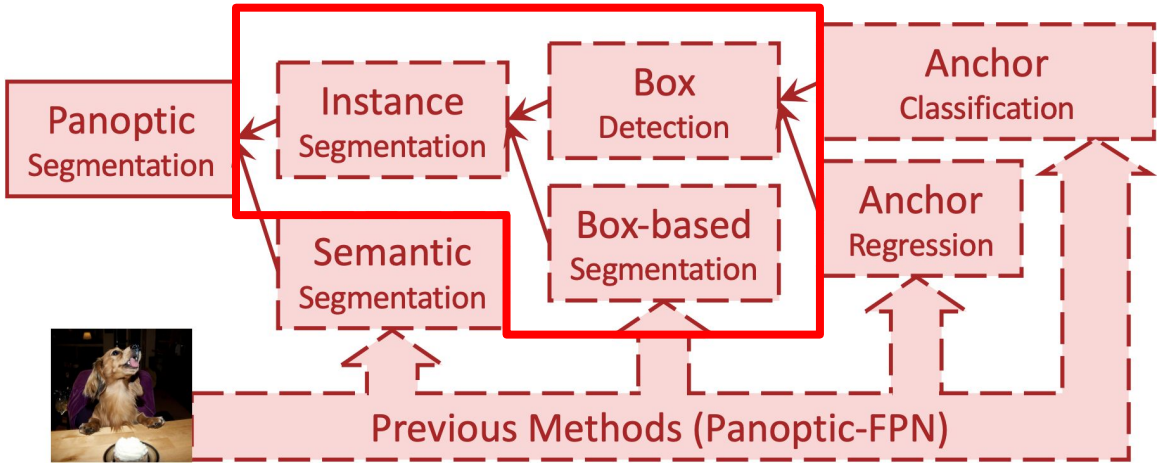
Kirillov, A., et al. Panoptic Feature Pyramid Networks. CVPR 2019.

Surrogate Sub-Tasks



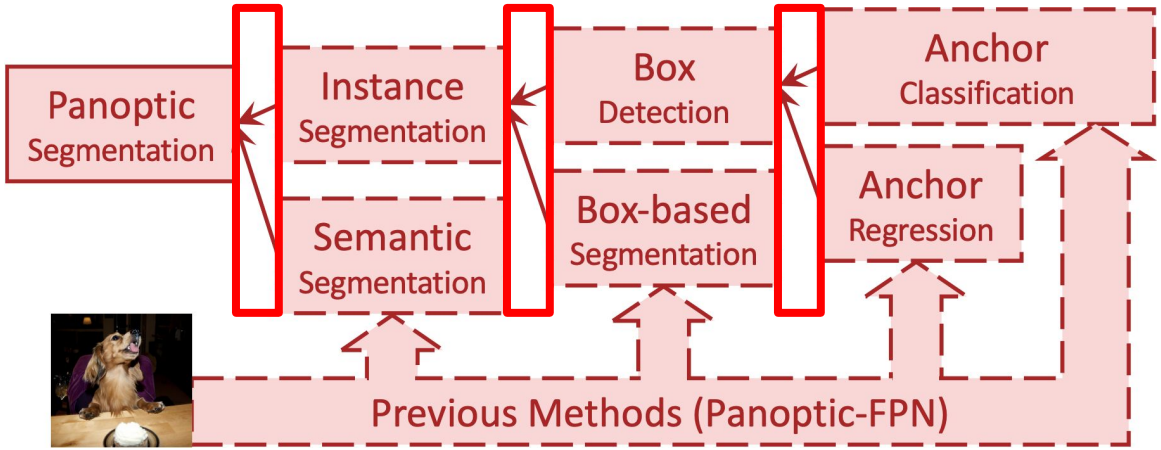
Kirillov, A., et al. Panoptic Feature Pyramid Networks. CVPR 2019.

Surrogate Sub-Tasks



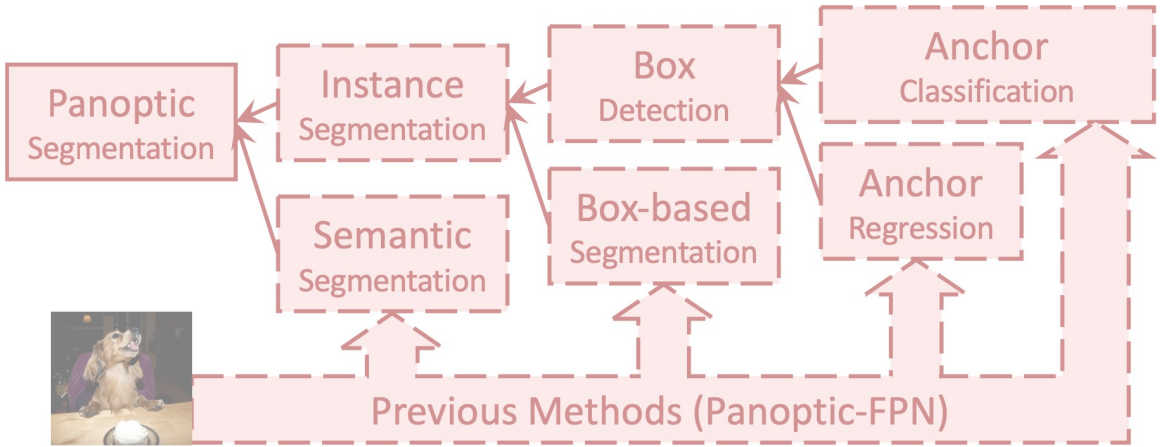
Kirillov, A., et al. Panoptic Feature Pyramid Networks. CVPR 2019.

Surrogate Sub-Tasks

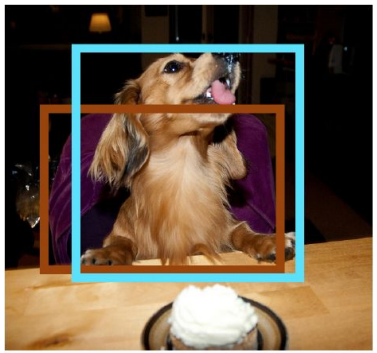


Kirillov, A., et al. Panoptic Feature Pyramid Networks. CVPR 2019.

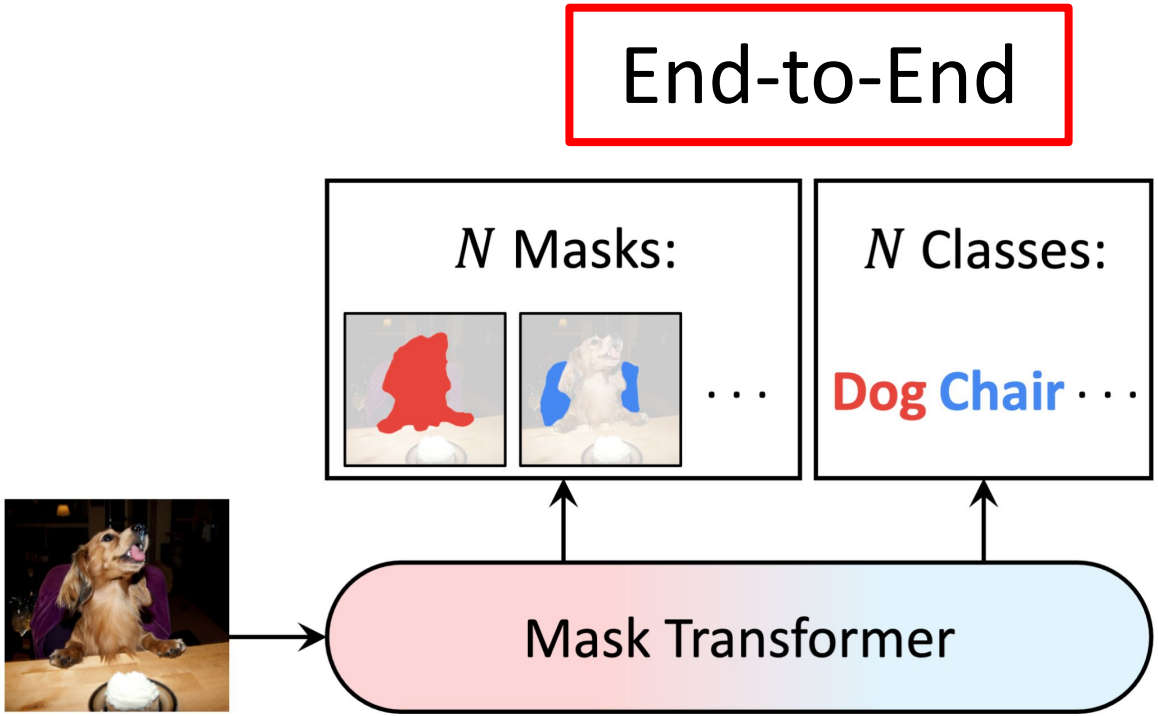
Surrogate Sub-Tasks



Fails!



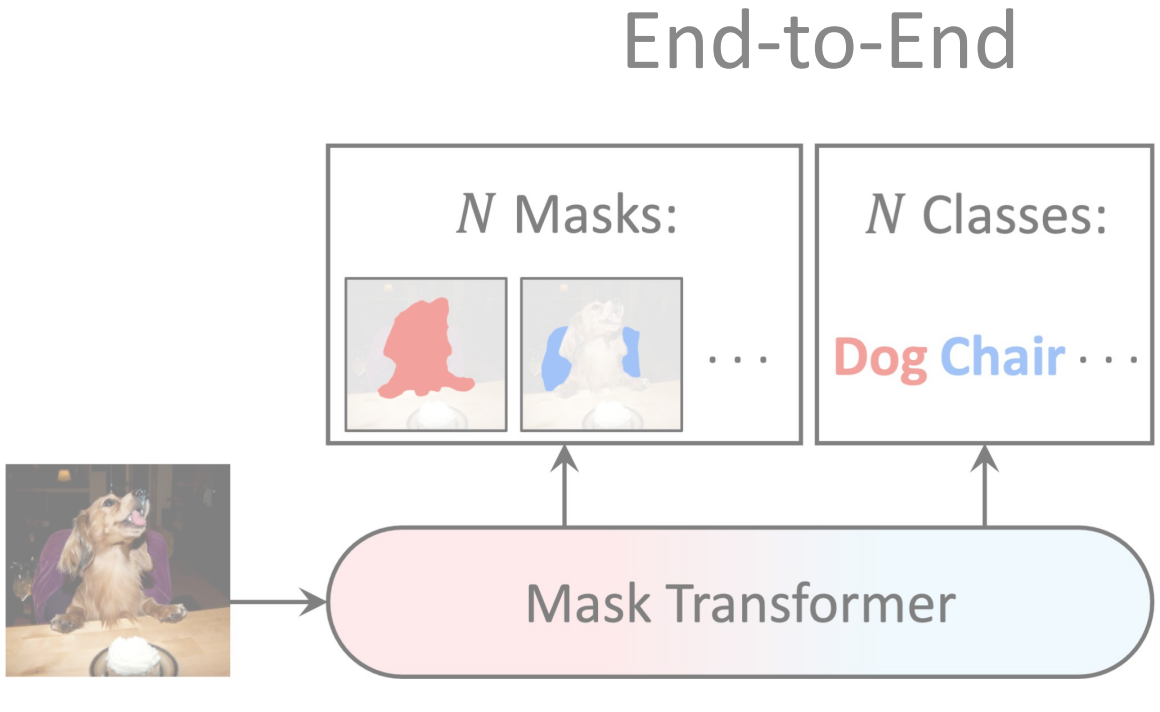
MaX-DeepLab

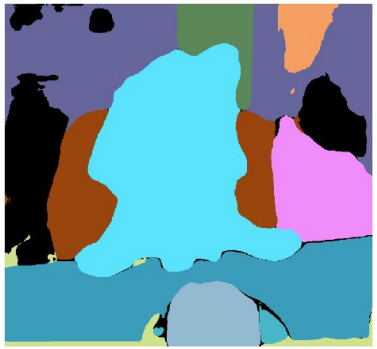


MaX-DeepLab

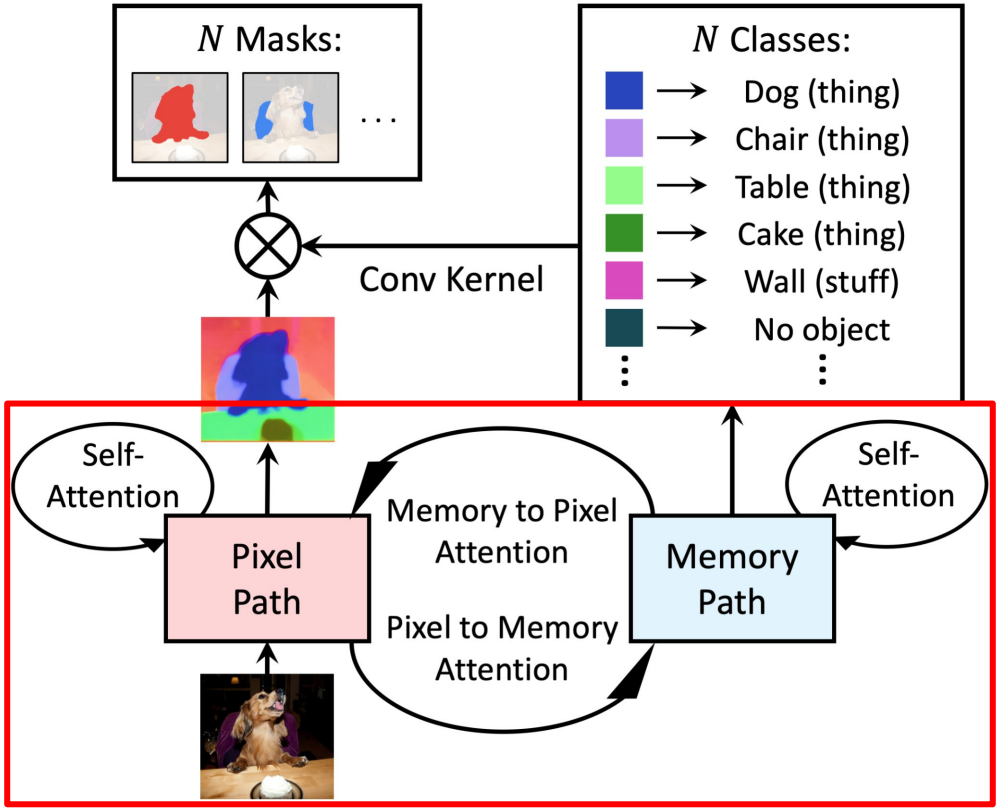
Solves the Case!

End-to-End

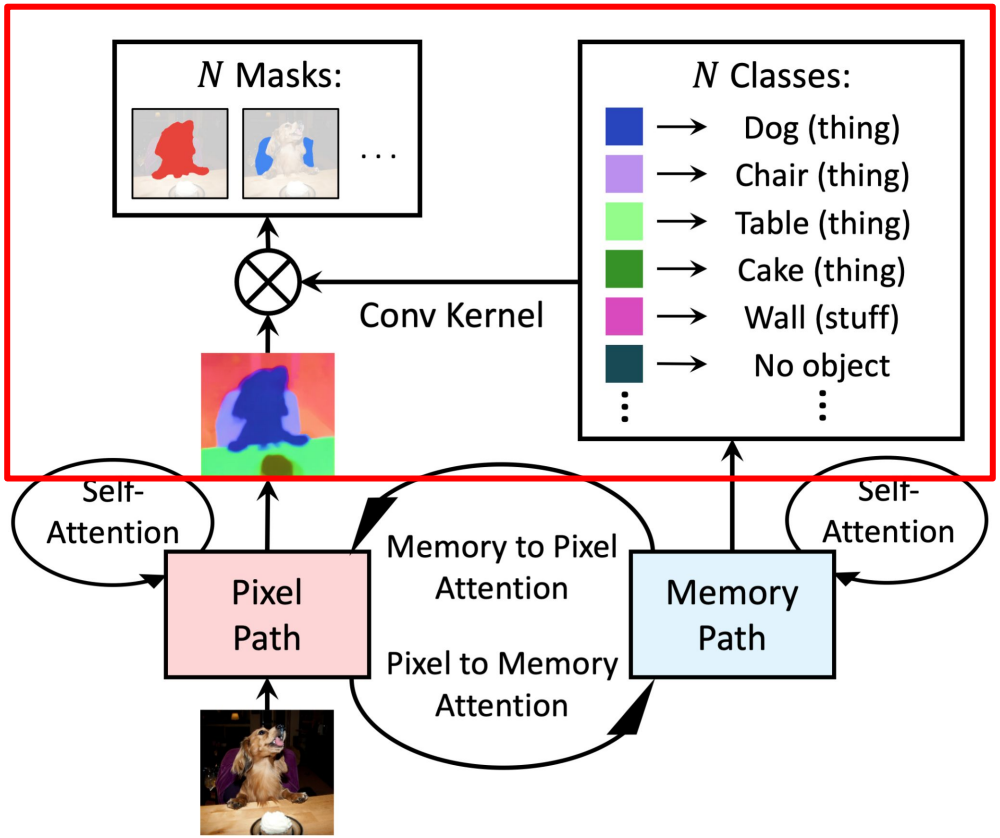




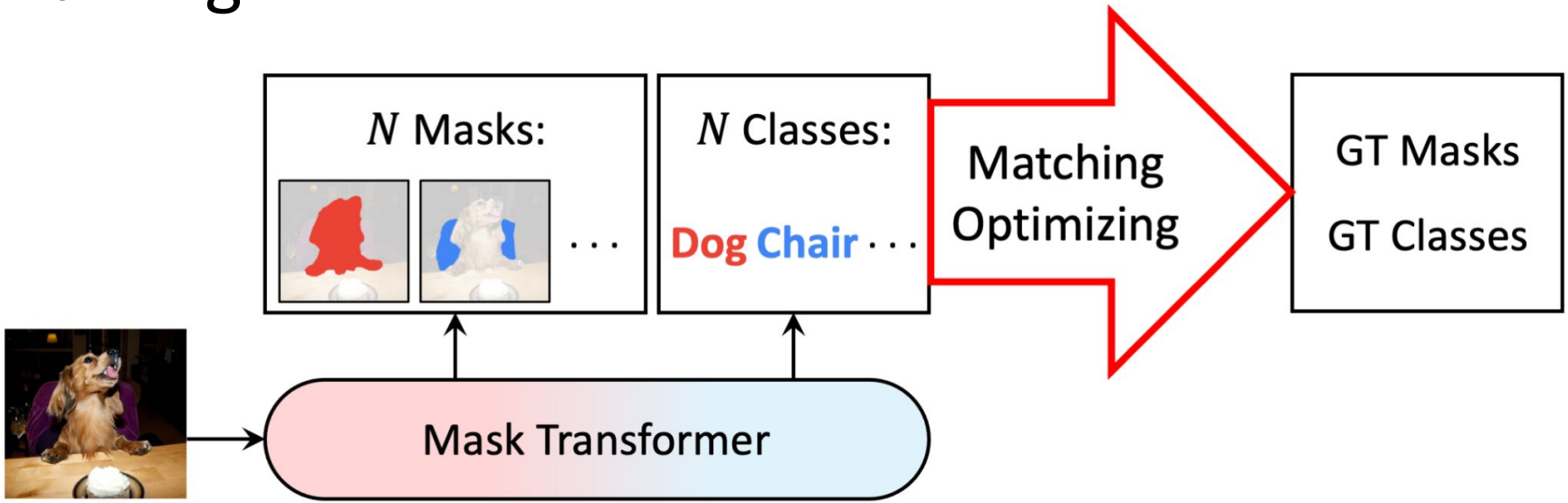
Mask Transformer



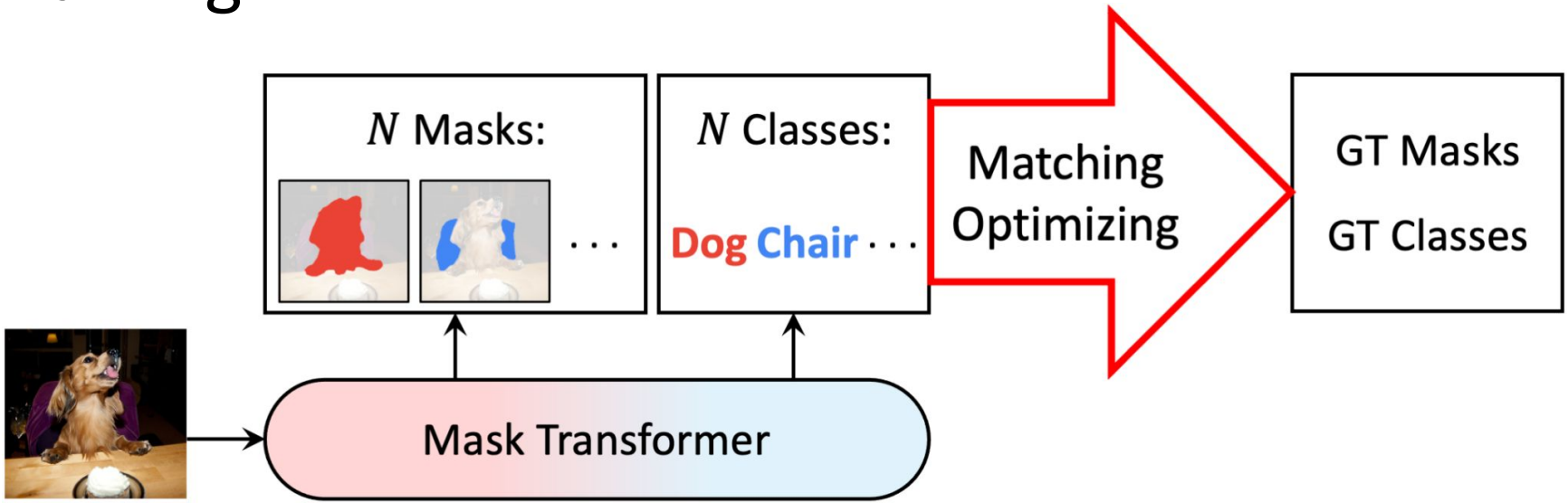
Mask Transformer



Training

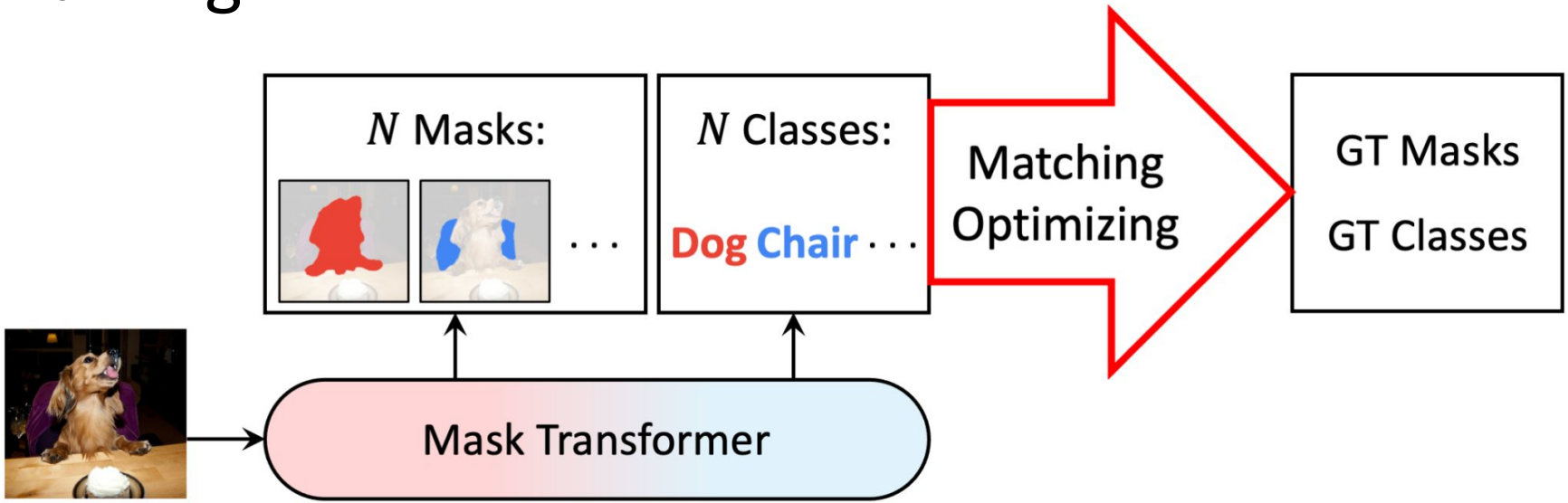

N Masks:
N Classes:
Dog Chair . . .
Matching
Optimizing
GT Masks
GT Classes
Mask Transformer

Training



- Panoptic Quality = Recognition Quality x Segmentation Quality

Training



- Panoptic Quality = Recognition Quality x Segmentation Quality
- PQ-style Similarity = Class Similarity x Mask Dice Score

Results on COCO (test-dev)

Method	Backbone	TTA	PQ	PQ^{Th}	PQ^{St}
Box-based panoptic segmentation methods					
Panoptic-FPN	RN-101		40.9	48.3	29.7
DETR	RN-101		46.0	-	-
UPSNet	DCN-101	✓	46.6	53.2	36.7
DetectoRS	RX-101	✓	49.6	57.8	37.1
Box-free panoptic segmentation methods					
Panoptic-DeepLab	X-71	✓	41.4	45.1	35.9
Axial-DeepLab-L	AX-L		43.6	48.9	35.6
Axial-DeepLab-L	AX-L	✓	44.2	49.2	36.8
MaX-DeepLab-S	MaX-S		49.0	54.0	41.6
MaX-DeepLab-L	MaX-L		**51.3**	57.2	42.4

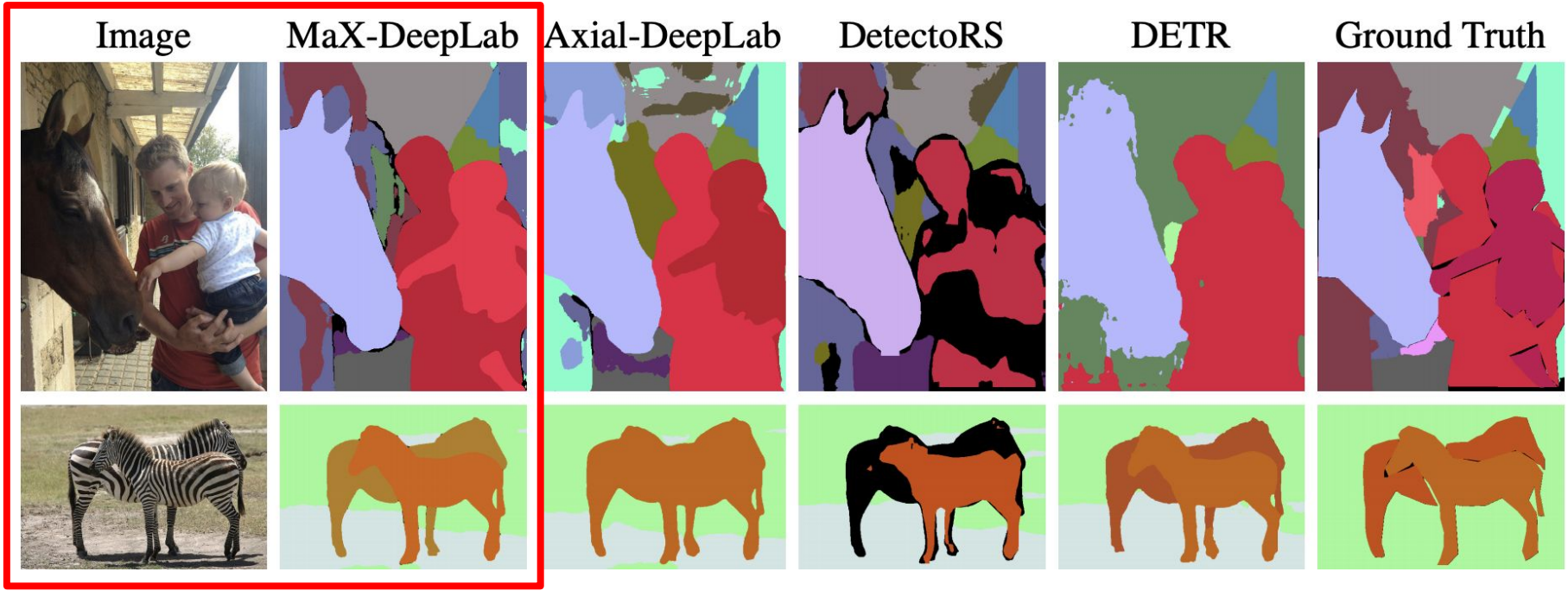
Lin, T.Y., et al. Microsoft coco: Common objects in context. ECCV 2014.
Xiong, Y. et al. UPSNet: A Unified Panoptic Segmentation Network. CVPR 2019.
Qiao, S. et al. DetectoRS: Detecting Objects with Recursive Feature Pyramid and Switchable Atrous Convolution. CVPR 2021.

Results on COCO (test-dev)

Method	Backbone	TTA	PQ	PQ^{Th}	PQ^{St}
Box-based panoptic segmentation methods					
Panoptic-FPN	RN-101		40.9	48.3	29.7
DETR	RN-101		46.0	-	-
UPSNet	DCN-101	✓	46.6	53.2	36.7
DetectoRS	RX-101	✓	49.6	57.8	37.1
Box-free panoptic segmentation methods					
Panoptic-DeepLab	X-71	✓	41.4	45.1	35.9
Axial-DeepLab-L	AX-L		43.6	48.9	35.6
Axial-DeepLab-L	AX-L	✓	44.2	49.2	36.8
MaX-DeepLab-S	MaX-S		49.0	54.0	41.6
MaX-DeepLab-L	MaX-L		**51.3**	57.2	42.4

Carion, N. et al. End-to-End Object Detection with Transformers. ECCV 2020.
Cheng, B. et al. Panoptic-DeepLab: A Simple, Strong, and Fast Baseline for Bottom-Up Panoptic Segmentation. CVPR 2020.
Wang, H. et al. Axial-DeepLab: Stand-Alone Axial-Attention for Panoptic Segmentation. ECCV 2020.

Visualizations



http://farm8.staticflickr.com/7295/9090127695_6dc690f776_z.jpg (http://creativecommons.org/licenses/by-nc/2.0/)
http://farm1.staticflickr.com/6/10184440_354f384aac_z.jpg (http://creativecommons.org/licenses/by-nc-nd/2.0/)

Attention Maps

Conclusion

- **End-to-end** panoptic segmentation
- Mask transformer
- PQ-style objective
- Code: https://github.com/google-research/deeplab2