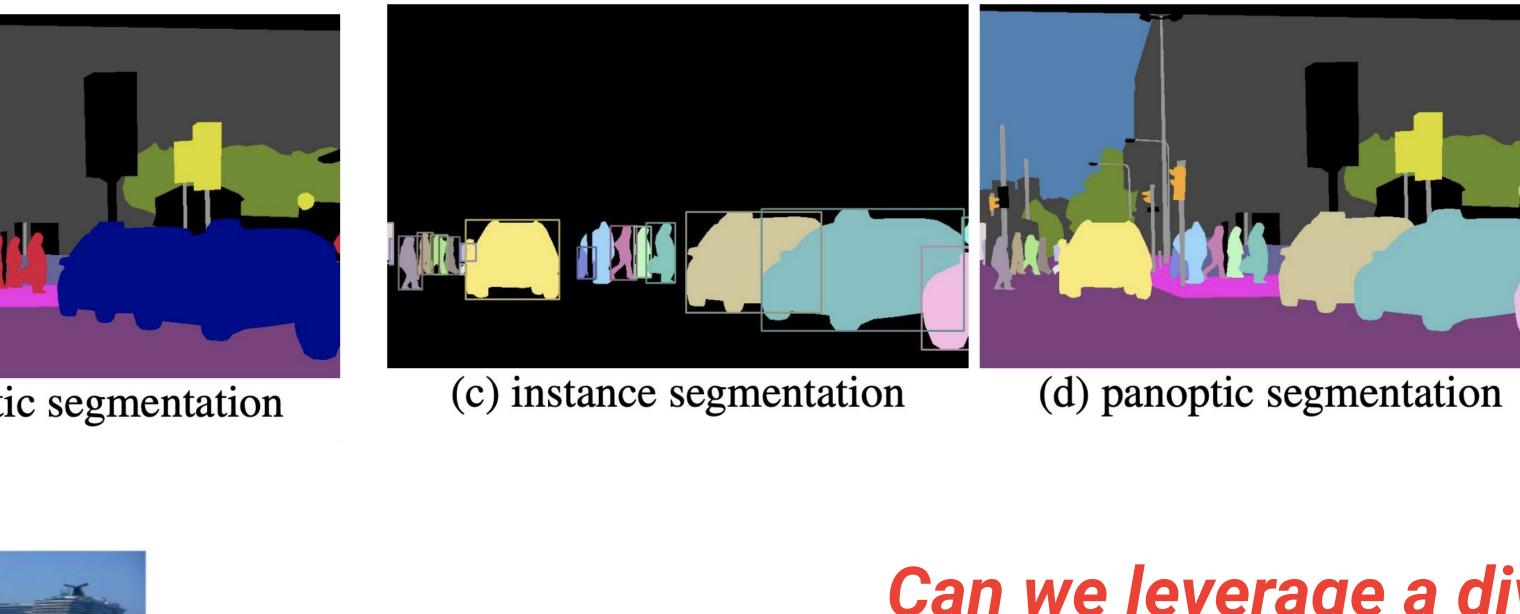# DaTaSeg: Taming a Universal Multi-Dataset Multi-Task Segmentation Model

Xiuye Gu, Yin Cui*, Jonathan Huang*, Abdullah Rashwan*, Xuan Yang*, Xingyi Zhou*, Golnaz Ghiasi, Weicheng Kuo, Huizhong Chen, Liang-Chieh Chen, David A Ross
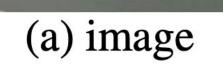
Introduction

(a) image   (b) semantic segmentation   (c) instance segmentation   (d) panoptic segmentation

Motivation
- Close relationship among different segmentation tasks
- Current situation: Task-specific model architectures
- Segmentation masks are expensive to annotate
  → Segmentation datasets are small

Can we leverage a diverse set of segmentation data to co-train a single model for all these segmentation tasks?

DaTaSeg: a universal multi-dataset multi-task segmentation model

- Single model to handle multiple segmentation tasks:
  - MaskFormer, Mask2Former, K-Net, OneFormer, etc.
  - Train a **separate** set of model weights for each dataset
- Single model on multiple segmentation datasets
  - UniSeg, MSeg, LMSeg
  - Train on a single segmentation task
- Ours: Train a **single** segmentation model on multiple datasets for multiple tasks

Method

- Universal representation for all tasks:
  - (mask proposal, class prediction) pairs
- Adopt different MERGE operations
- Adopt different postprocessing

Overview of DaTaSeg Architecture
- Use text embeddings from the same semantic embedding space as classifiers
- Share all network parameters among datasets

Release a DaTaSeg Objects365 Instance Segmentation Dataset

Quantitative Results

Main Results

Backbone	Model	Fully-Supervised		Weakly-Supervised Transfer	
		ADE semantic mIoU	COCO panoptic PQ	ADE semantic → panoptic PQ	O365 box → instance mask AP
ResNet50	Separate	42.0	48.2	26.9	12.3
	DaTaSeg	48.1 (+6.1)	49.0 (+0.8)	29.8 (+2.9)	14.3 (+2.0)
ViTDet-B	Separate	46.3	51.9	27.5	14.7
	DaTaSeg	51.4 (+5.1)	52.8 (+0.9)	32.9 (+5.4)	16.1 (+1.4)
ViTDet-L	DaTaSeg	54.0	53.5	33.4	16.4

- Train on ADE20k semantic, COCO panoptic, and Objects365 detection
- DaTaSeg outperforms separately trained models on ALL datasets
- DaTaSeg enables weakly-supervised knowledge transfer
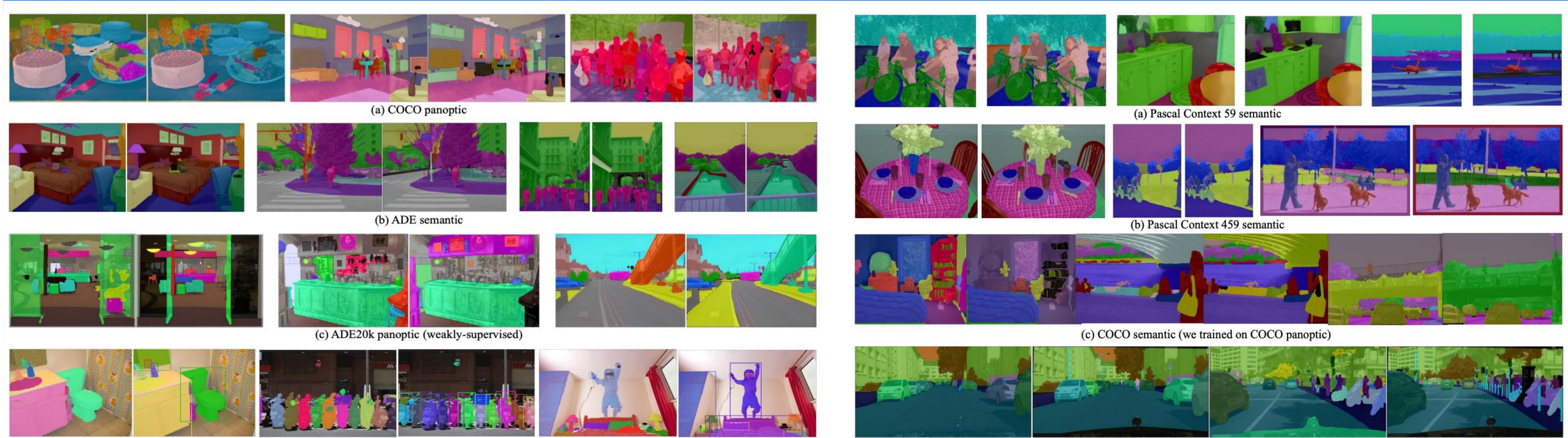- DaTaSeg scales with backbones

DaTaSeg enables open-vocabulary segmentation

Method	Backbone	Training data	PC-59 mIoU	PC-459 mIoU	COCO mIoU	Cityscapes PQ
ODISE [63]	UNet+M2F	LAION+CLIP+COCO	55.3	13.8	52.4	23.9
MaskCLIP [12]	R-50	COCO pan+CLIP	45.9	10.0	–	–
OVSeg [35]	R-101c	COCO stuff+cap	55.3	11.0	–	–
	Swin-B		55.7	12.4	–	–
OpenSeg [15]	R-101	COCO pan+cap	42.1	9.0	36.1	–
	Eff-b7	COCO+Loc. Narr.	44.8	11.5	38.0	–
DaTaSeg	R-50	COCO panoptic	50.9	11.1	57.7	30.0
	ViTDet-B	+ADE semantic	51.1	11.6	62.7	28.0
	ViTDet-L	+O365 bbox	51.4	11.1	62.9	29.8

DaTaSeg scales with number of training datasets

Training sets			Fully-Supervised		Weakly-Supervised Transfer	
ADE semantic	COCO panoptic	O365 bbox	ADE semantic mIoU	COCO panoptic PQ	ADE semantic → panoptic PQ	O365 box → instance mask AP
✓			42.0			
	✓			8.4	26.7	1.0
		✓	15.3		11.6	5.8
			11.0	48.2	5.8	12.3
✓	✓		46.3 (+4.3)	12.4		
		✓	18.3	14.0	26.4 (-0.3)	15.0 (+2.7)
✓	✓	✓	47.3 (+5.3)	48.9 (+0.7)	12.3	15.2 (+2.9)
			48.1 (+6.1)	49.0 (+0.8)	30.5 (+3.8)	5.6
✓	✓	✓			29.8 (+2.9)	14.3 (+2.0)

Qualitative Results

(a) COCO panoptic

(b) ADE semantic

(c) ADE20k panoptic (weakly-supervised)

(d) Objects365 instance (weakly-supervised)

(a) Pascal Context 59 semantic

(b) Pascal Context 459 semantic

(c) COCO semantic (we trained on COCO panoptic)

(d) Cityscapes panoptic