JUNE 18-22, 2023
CVPR
VANCOUVER, CANADA

Poster: June 20, 2023 @ TUE-PM-280

Open-Vocabulary Semantic Segmentation with Mask-adapted CLIP

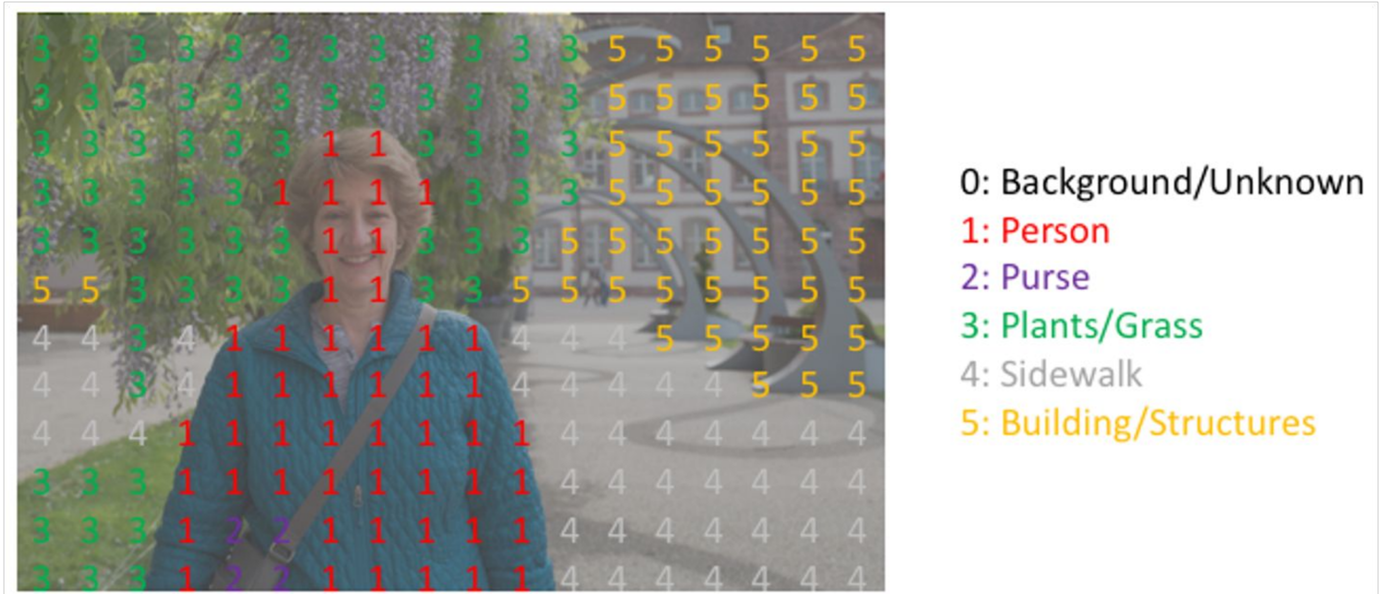
https://jeff-liangf.github.io/projects/ovseg/

Feng (Jeff) Liang, Bichen Wu, Xiaoliang Dai, Kunpeng Li, Yinan Zhao, Hang Zhang, Peizhao Zhang, Peter Vajda, Diana Marculescu

TEXAS
The University of Texas at Austin

Meta

Traditional segmentation



[Source: Jeremy Jordan]

Traditional segmentation model can only segment the classes in the training dataset

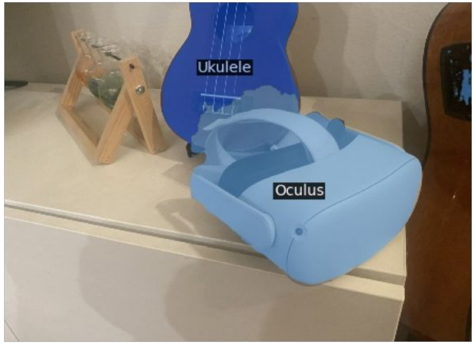
Open-vocabulary segmentation

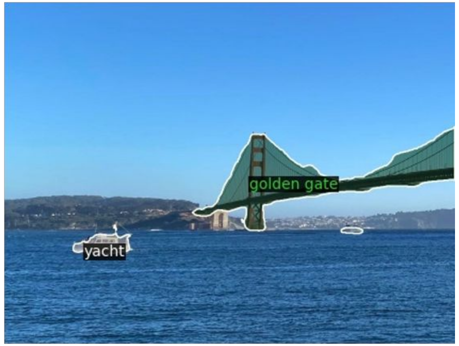


Where is Saturn V, blossom?

Where is Oculus, Ukulele?

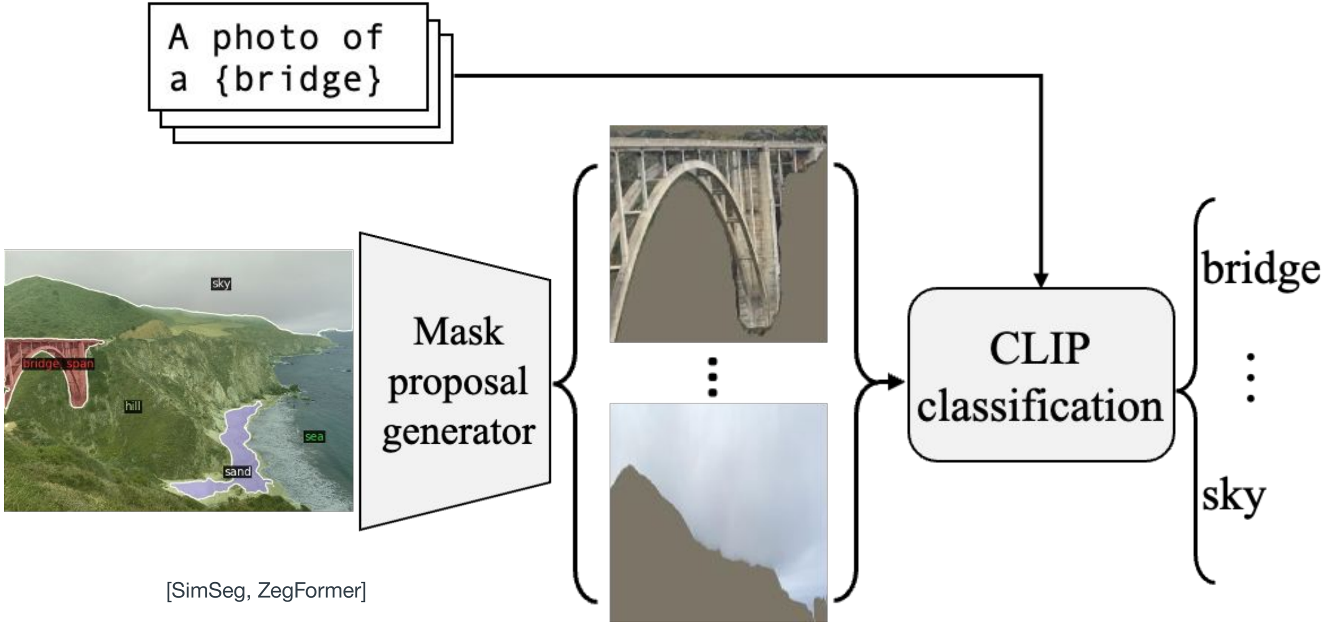
Where is Golden gate, yacht?





Open-vocab segmentation model can segment any arbitrary class defined by user

Two-stage open-vocabulary baseline

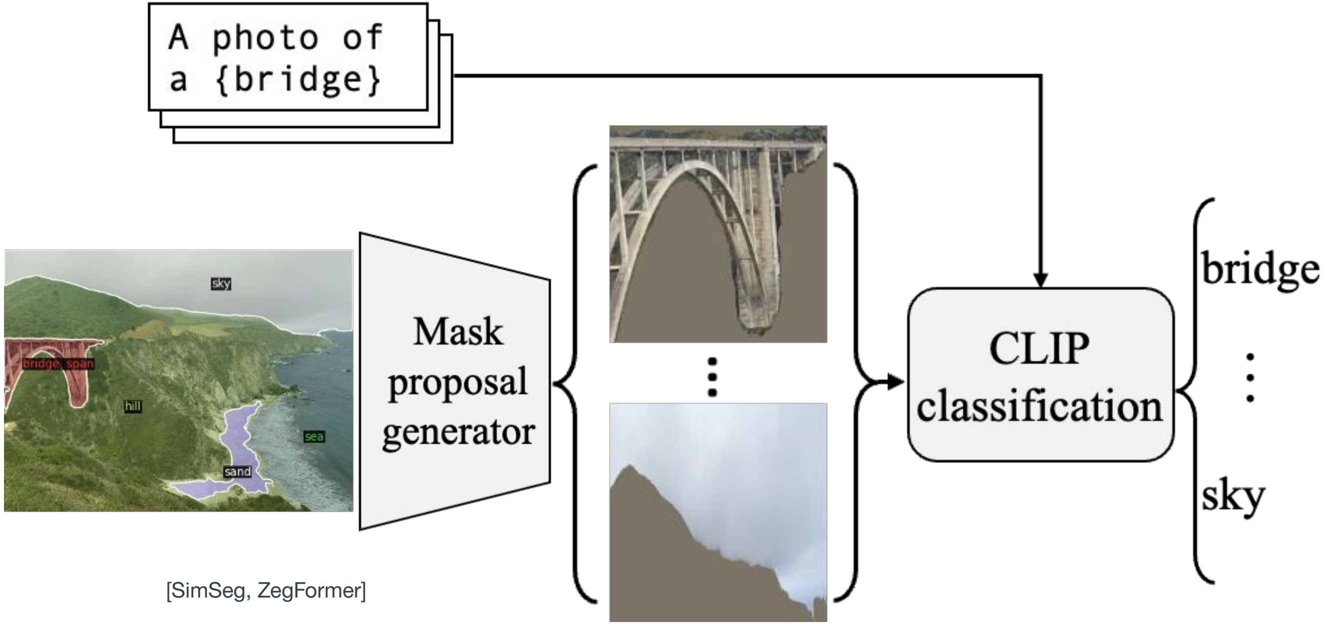


[SimSeg, ZegFormer]

First generate 'class-agnostic' mask proposals

Then use the pre-trained CLIP to do open-vocabulary classification

Two-stage open-vocabulary baseline

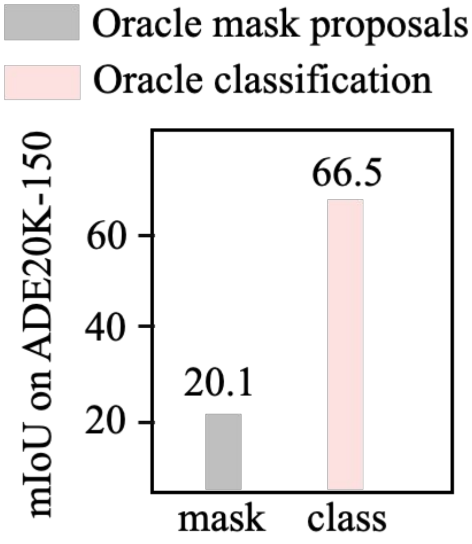


[SimSeg, ZegFormer]

The success of two-stage approaches lies on two assumptions:
(1) Mask proposals: Proposal generator can generalize to unseen categories
(2) Classification: Pre-trained CLIP can perform good classification on mask images

Bottleneck analysis of two-stage baseline

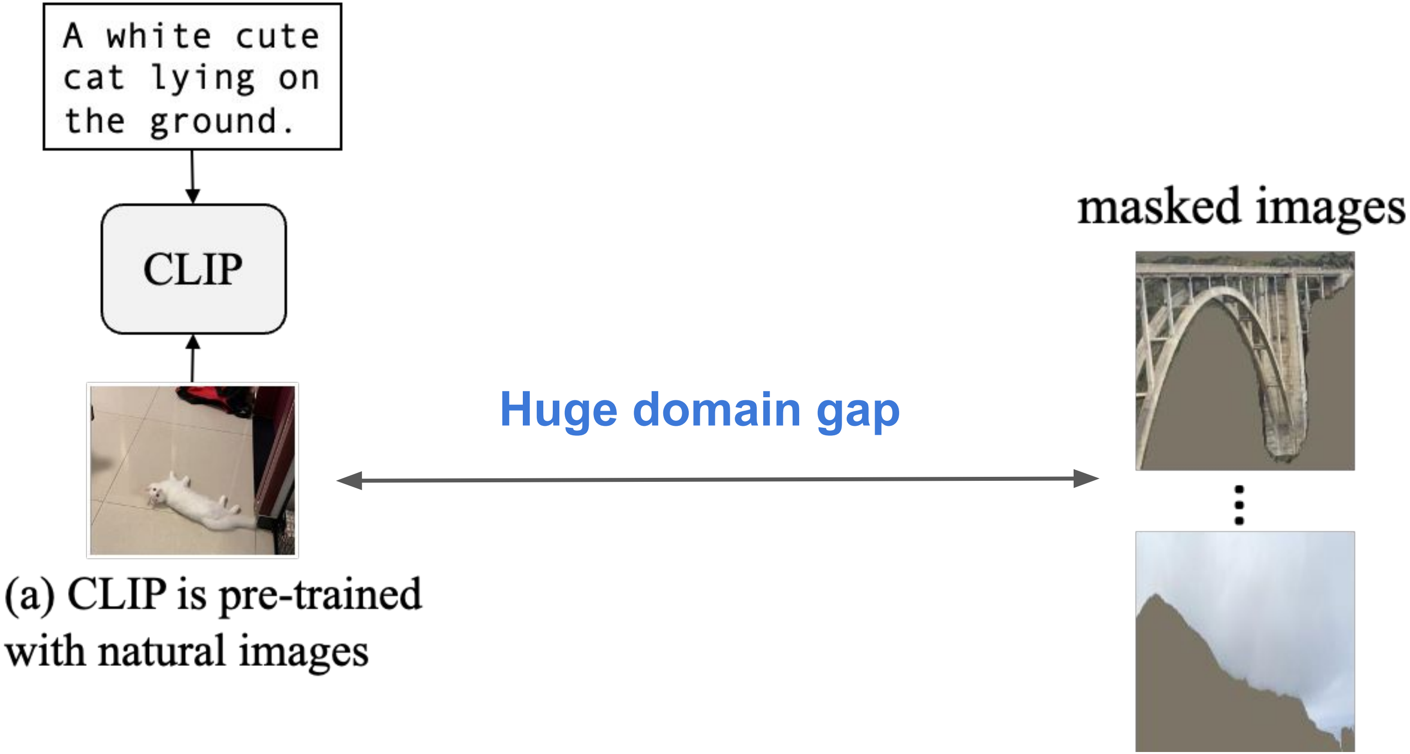
Bottleneck analysis
(1) We use oracle (ground-truth) mask proposals and perform CLIP classification over them
(2) We assume an "oracle" classifier but an ordinary mask proposal generator – a MaskFormer pre-trained on the COCO dataset



(c) Bottleneck analysis

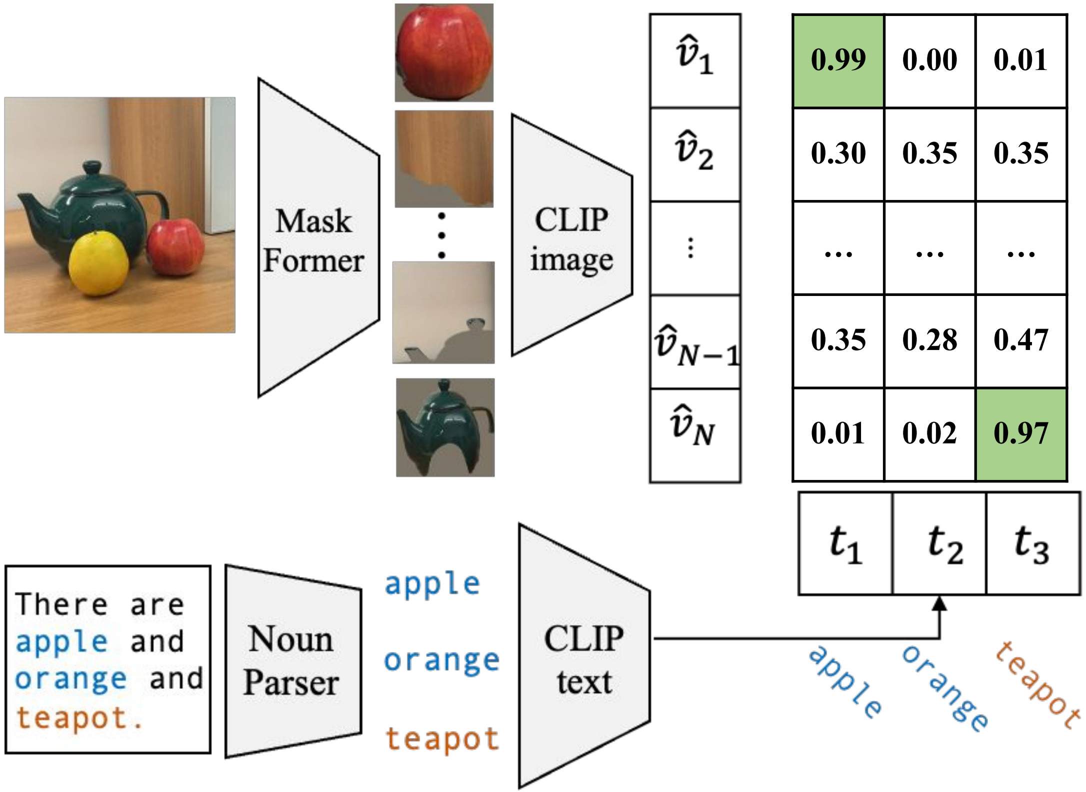
CLIP can not perform well on mask proposals, making CLIP the major bottleneck.

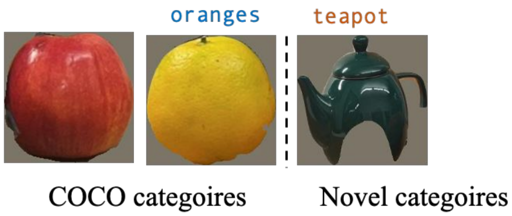
CLIP can not perform well on mask proposals



Collecting mask-text pairs to finetune CLIP

We propose adapt CLIP to masked images with collected diverse mask-category pairs from captions

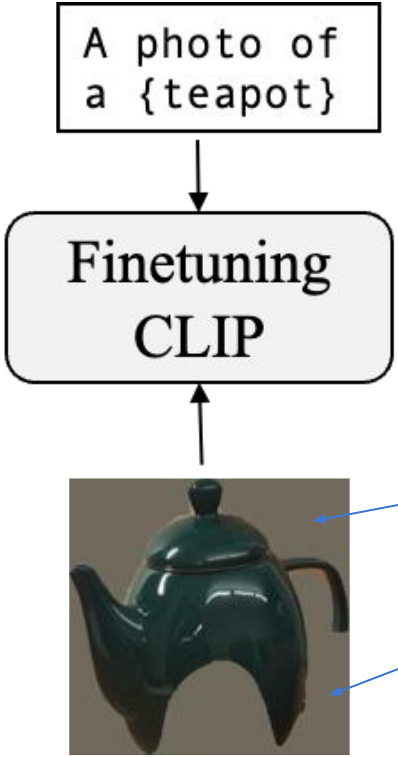




We can collect 440K pairs with 12K nouns

Mask prompt tuning for CLIP

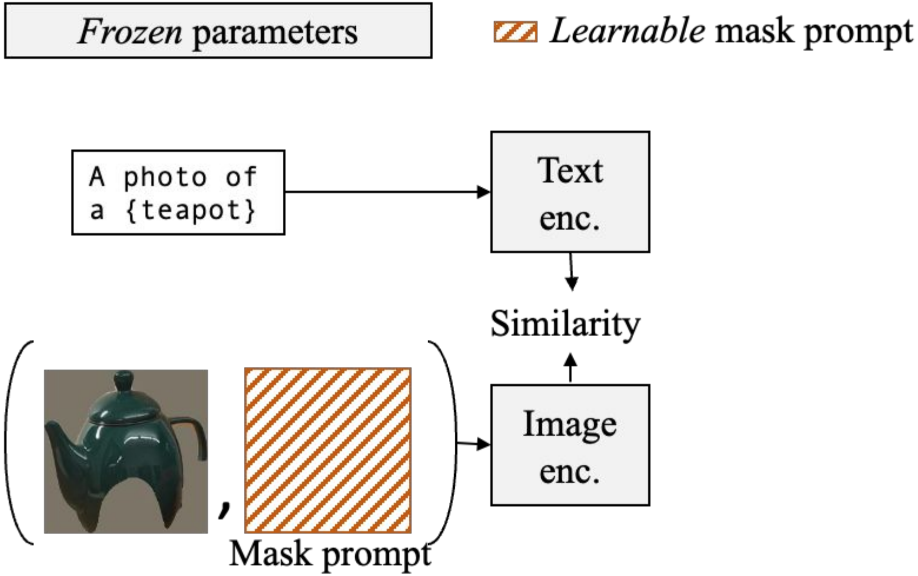
After collecting the data, how can we finetune the CLIP?



The most notable difference between a masked image and a natural image is that background pixels in a masked image are **masked out**.

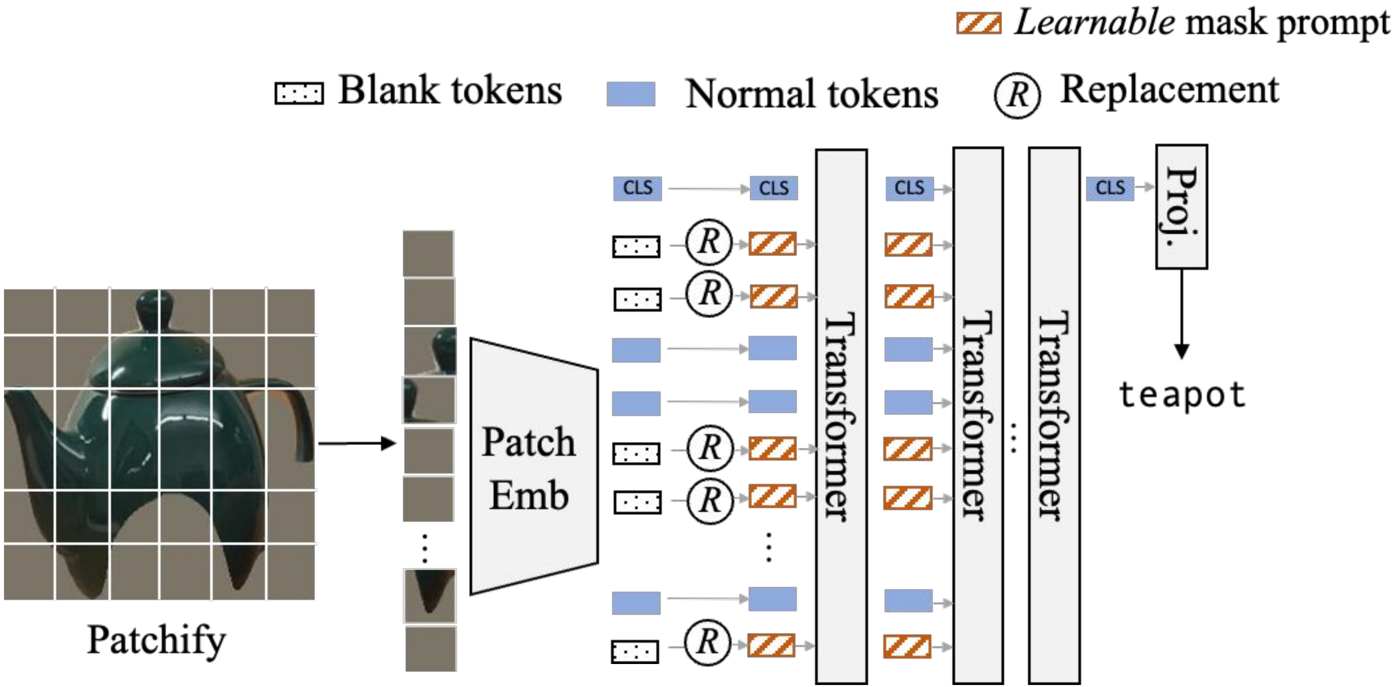
Mask prompt tuning for CLIP

We only need to finetune the ‘blank areas’ whiling keep the entire CLIP model frozen.



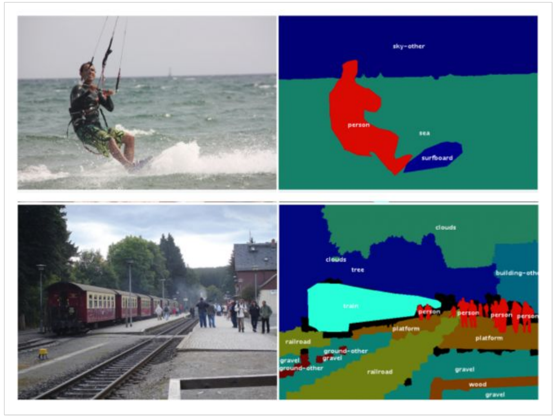
Mask prompt tuning for CLIP

We replace the blank tokens with leanable tokens



Evaluation setting

Training

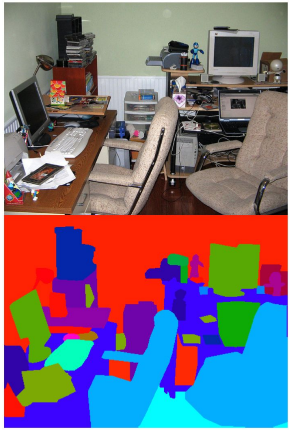


COCO-stuff
# of classes: 171

Zero-shot transfer

Evaluating

ADE-20k
# of cls: 150 or 857



Pascal Context
# of cls: 59 or 459

……

Performance

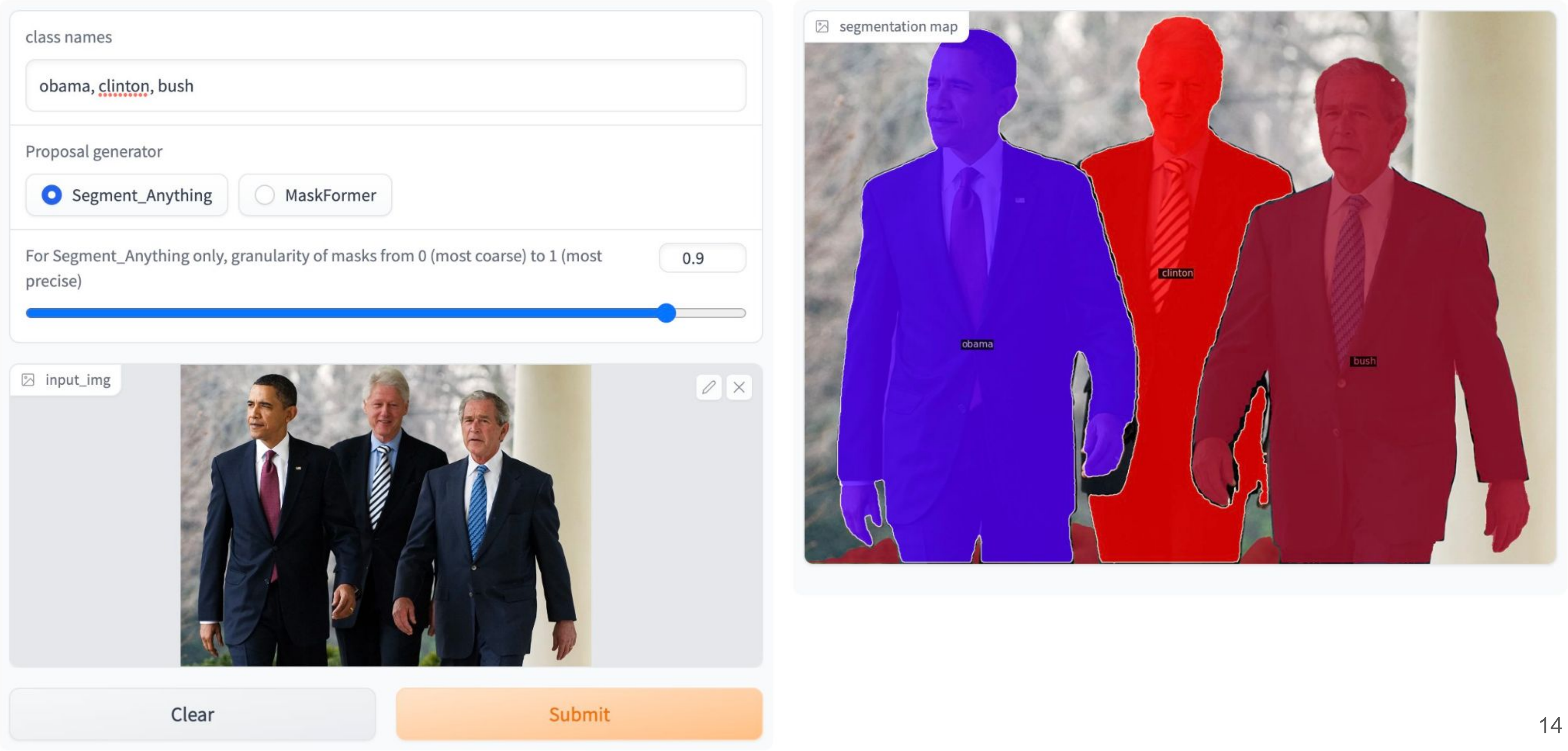
• A-847/150: ADE with 847/150 classes • PC-459/59: Pascal Context with 459/59 classes • PAS-20: VOC 2012 with 20 classes

method	backbone	training dataset	A-847	PC-459	A-150	PC-59	PAS-20
		Open-vocabulary generalist models.					
SPNet (Xian et al., 2019)	R-101	PASCAL-15	-	-	-	24.3	18.3
ZS3Net (Bucher et al., 2019)	R-101	PASCAL-15	-	-	-	19.4	38.3
LSeg (Li et al., 2022)	R-101	PASCAL-15	-	-	-	-	47.4
LSeg+ (Ghiasi et al., 2021)	R-101	COCO Panoptic	2.5	5.2	13.0	36.0	59.0
SimBaseline (Xu et al., 2021)	R-101c	COCO-Stuff-156	-	-	15.3	-	74.5
ZegFormer (Ding et al., 2022)	R-50	COCO-Stuff-156	-	-	16.4	-	80.7
OpenSeg (Ghiasi et al., 2021)	R-101	COCO Panoptic	4.0	6.5	15.3	36.9	60.0
OVSeg (Ours)	R-101c	COCO-Stuff-171	**7.1**	**11.0**	**24.8**	**53.3**	**92.6**
LSeg+ (Ghiasi et al., 2021)	Eff-B7	COCO Panoptic	3.8	7.8	18.0	46.5	-
OpenSeg (Ghiasi et al., 2021)	Eff-B7	COCO Panoptic	6.3	9.0	21.1	42.1	-
OVSeg (Ours)	Swin-B	COCO-Stuff-171	**9.0**	**12.4**	**29.6**	**55.7**	**94.5**
		Supervised specialist models.					
FCN (Long et al., 2015)	FCN-8s	Same as test	-	-	29.4	37.8	-
Deeplab (Chen et al., 2017)	R-101	Same as test	-	-	-	45.7	77.7
SelfTrain (Zoph et al., 2020)	Eff-L2	Same as test	-	-	-	-	90.0

Our model outperforms the state-of-the-art OpenSeg by a +8.5% margin.

For the first-time, we show open-vocabulary generalist models can match the performance of supervised specialist model.

OVSeg + Segment_Anything



Feel free to try our model /codes /demo !

FLOWCODE
PRIVACY.FLOWCODE.COM

OVSeg project

EnyAC
enyac.org

Our EnyAC group