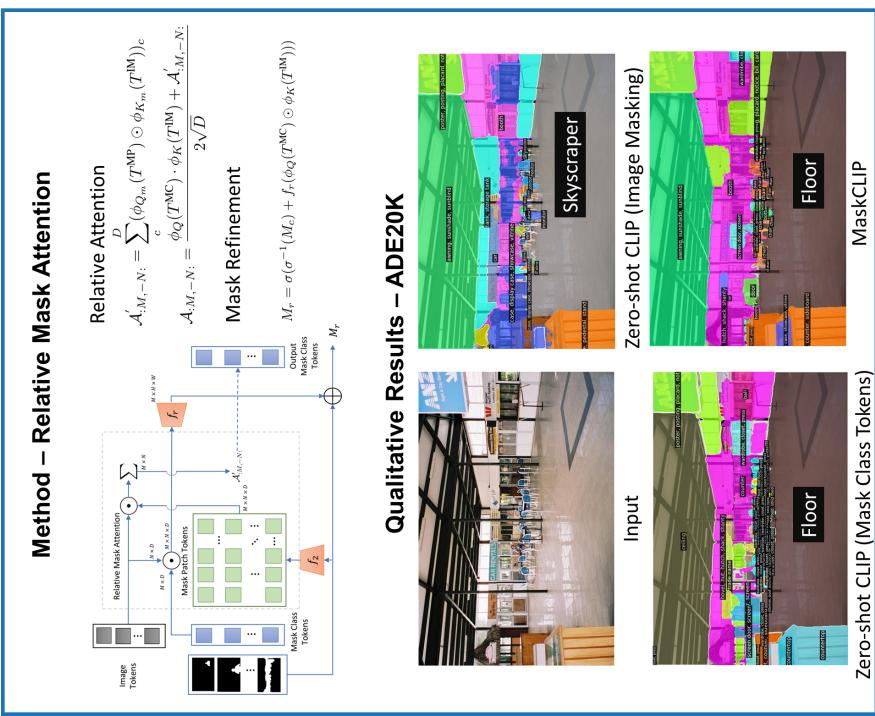
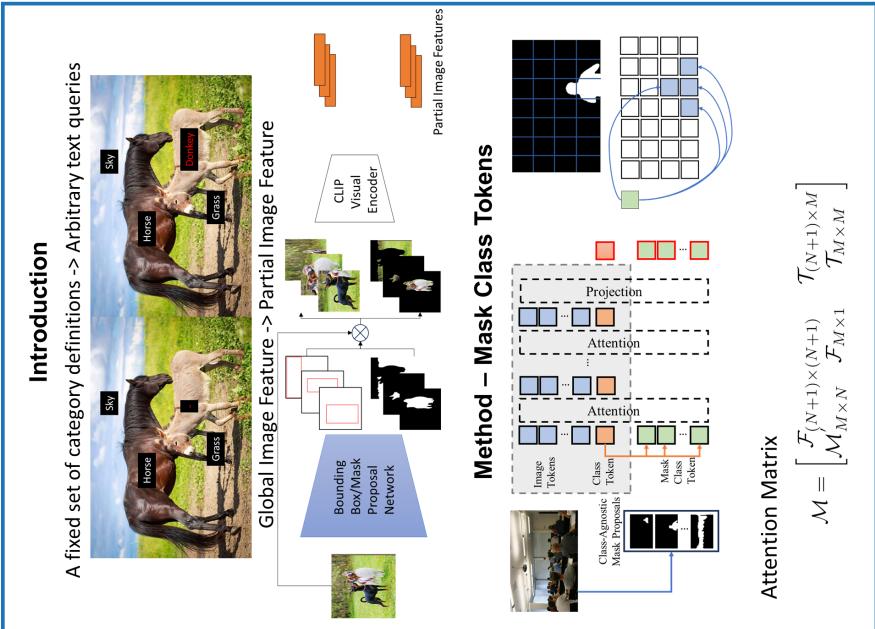


Open-Vocabulary Universal Image Segmentation with MaskCLIP

Zheng Ding, Jieke Wang, Zhuowen Tu
UC San Diego



Qualitative Results - Arbitrary Text Inputs

Quantitative Results

Method	Training Data	A-150↑	A-847↑	P-459↑	P-59↑
ALIGN	None	10.7	4.1	3.7	15.7
ALIGN w/ proposals	Masks	12.9	5.8	4.8	22.4
LSeg+	Masks + Labels	18.0	3.8	7.8	46.5
OpenSeg	Masks + Captions	21.0	6.3	9.0	42.1
SimSeg	Masks + Labels	20.5	7.0	-	47.7
ZS CLIP (Image Masking)	Masks	13.8	5.2	5.2	25.3
ZS CLIP (Mask Class Tokens)	Masks	14.9	5.6	5.3	26.1
MaskCLIP	Masks + Labels	23.7	8.2	10.0	45.9

FLOPs Comparison

Method	TFLOPS
RegionCLIP	9.5
ZegFormer	10.3
SimSeg	9.6
CLIP Baseline	10.5
MaskCLIP	0.3

Please refer to our project website for more details:
<https://maskclip.github.io>