SED: A Simple Encoder-Decoder for Open-Vocabulary Semantic Segmentation

Supplementary Material

Num.	SED (ours)			CAT-Seg			A-150		PC-459	
	HECG	GFD	CER	Enc.	Dec.	Feat. Backbone	mIoU	Time (ms)	mIoU	Time (ms)
(a)	✓	✓	✓				31.6	**82**	**18.6**	**120**
(b)	✓	✓					**31.8**	178	**18.6**	468
(c)		✓		✓			27.2	287	16.7	806
(d)				✓	✓		25.9	323	16.0	916
(e)				✓	✓	✓	27.2	362	16.6	1004

Table 1. **Impact of transforming SED to CAT-Seg.**

1. Impact of transforming SED to CAT-Seg

To better show the efficacy of our SED, we perform the experiment by transforming SED to CAT-Seg. On A-150, when replacing our HECG by CAT-Seg encoder (b→c), mIoU drops from 31.8 to 27.2, and inference time increases from 178ms to 287ms. When further replacing our GFD by CAT-Seg decoder (c→d), mIoU drops from 27.2 to 25.9, and inference time increases from 287ms to 323ms. The similar tendency is observed on PC-459. Moreover, our SED (a) *significantly* outperforms CAT-Seg (e) at faster speed.