BiViT: Extremely Compressed Binary Vision Transformers Supplementary Material

Yefei He[1] Zhenyu Lou[1] Luoming Zhang[1] Jing Liu[2] Weijia Wu[1]
Hong Zhou[1†] Bohan Zhuang[2†]

[1]Zhejiang University, China
[2]ZIP Lab, Monash University, Australia

A. More experimental results on COCO

In this section, we present additional experimental results of BiViT on the COCO 2017 validation set. As shown in Table A, our method outperforms the competitive BiBERT by a large margin in all metrics.

Table A. Comparisons of different methods and backbones on COCO 2017 validation set.

Framework	Method	Task	AP	AP_{50}	AP_{75}	AP_S	AP_M	AP_L
Mask R-CNN	FP	Object Detection	43.7	66.6	47.7	28.5	47.0	57.3
	BiBERT		32.0	53.9	33.7	19.6	34.1	41.3
	Ours		**33.4**	**55.0**	**35.2**	**20.8**	**35.6**	**42.0**
	FP	Instance Segmentation	39.8	63.3	42.7	24.2	43.1	54.6
	BiBERT		30.4	51.0	31.9	16.7	32.5	41.4
	Ours		**31.6**	**51.7**	**33.4**	**18.3**	**34.0**	**42.1**
Cascade Mask R-CNN	FP	Object Detection	48.1	67.1	52.2	30.4	51.5	63.1
	BiBERT		39.2	57.3	42.5	23.7	41.3	51.1
	Ours		**40.8**	**59.2**	**44.1**	**25.2**	**43.3**	**52.8**
	FP	Instance Segmentation	41.7	64.4	45.0	41.7	64.4	45.0
	BiBERT		34.5	54.5	36.8	19.1	36.5	46.6
	Ours		**35.7**	**56.5**	**38.2**	**20.1**	**38.2**	**48.5**

†H. Zhou and B. Zhuang are corresponding authors.