Panoptic Perception: A Novel Task and Fine-grained Dataset for Universal Remote Sensing Image Interpretation

Danpei Zhao, *Member, IEEE,* Bo Yuan, Ziqiang Chen, Tian Li, Zhuoran Liu, Wentao Li, and Yue Gao
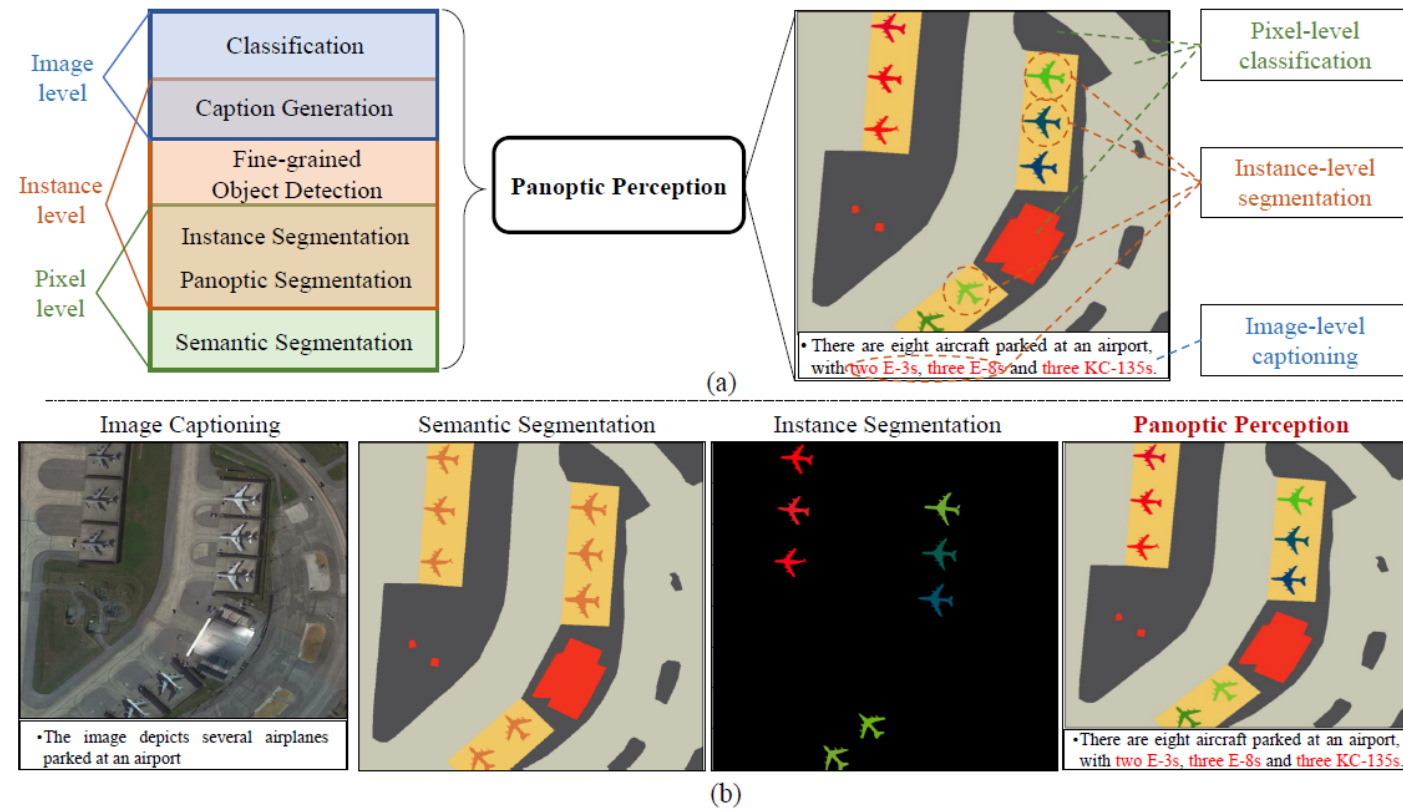
AIRVIC lab

Fig. 1. The proposed fine-grained Panoptic Perception task: (a) The definition and scope of panoptic perception Task. (b) Qualitative comparison of panoptic perception and other interpretation tasks.

TABLE I
THE PROPOSED FINEGRIP DATASET.

Dataset	Class num.	Sample num.	Bbox	Masks	Caption	Fine grained	Cross modal
MAR20 [40]	20	3842	✓	✗	✗	✓	✗
BSB [7]	15	3400	✓	✓	✗	✗	✗
FineGrip	25	2649	✓	✓	✓	✓	✓

Fig. 2. Number of per-category masks in the FineGrip dataset across training and validation sets.

TABLE II
COMPARISONS BETWEEN OUR PROPOSED PANOPTIC PERCEPTION METHOD AND OTHER INTERPRETATION APPROACHES ON FINEGRIP. THE BEST RESULTS ARE MARKED IN BOLD.

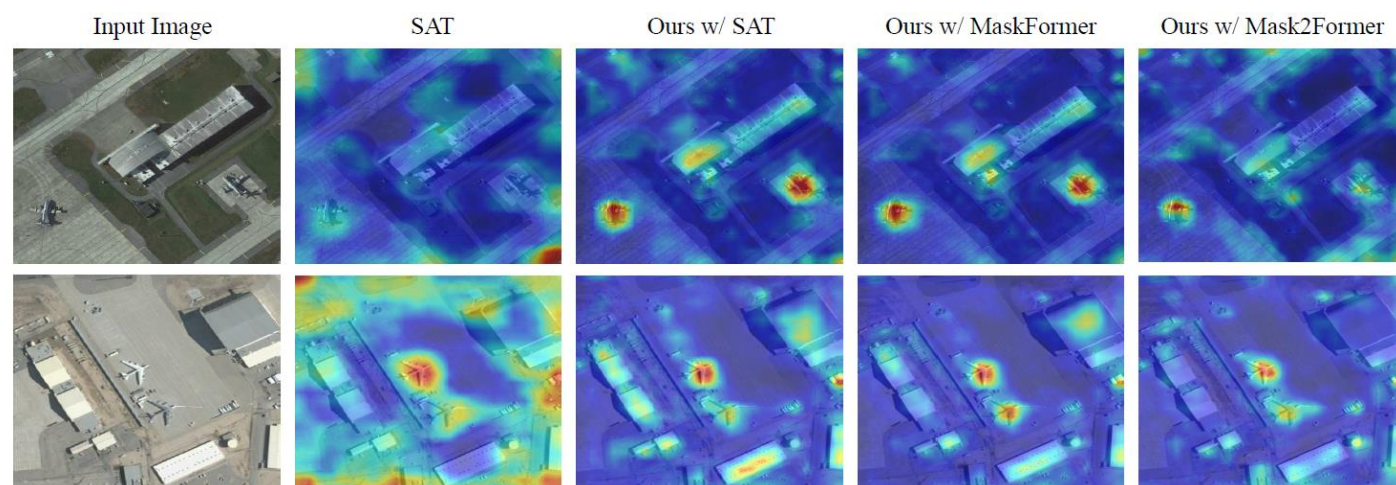| Task	Method	All			Things			Stuff			BLEU score	
		PQ	SQ	RQ	PQ^{Th}	SQ^{Th}	RQ^{Th}	PQ^{st}	SQ^{St}	RQ^{St}	beam size=3	beam size=5
Panoptic Segmentation	Panoptic FPN [48]	54.1	78.6	68.3	55.8	79.0	**70.7**	47.1	77.1	58.8	-	-
	MaskFormer [43]	50.2	79.4	63.1	49.9	78.8	63.3	51.6	81.8	62.5	-	-
	Mask2Former [49]	55.2	**81.0**	67.9	54.8	80.5	67.9	57.0	83.0	68.0	-	-
Caption Generation	SAT [50]	-	-	-	-	-	-	-	-	-	30.4	29.4
	SAT w/o attention	-	-	-	-	-	-	-	-	-	36.2	35.9
	MLAT [51] w/o LSTM	-	-	-	-	-	-	-	-	-	35.9	35.8
	MLAT	-	-	-	-	-	-	-	-	-	41.2	40.3
Panoptic Perception	Ours w/ SAT	49.6	79.5	62.1	48.5	78.7	61.4	53.8	**82.6**	64.7	**45.2**	**44.0**
	Ours w/ MaskFormer	50.9	79.3	63.8	49.8	78.6	63.2	55.1	82.1	66.1	42.4	41.7
	Ours w/ Mask2Former	**56.5**	80.9	**69.6**	**56.3**	80.6	69.7	**57.3**	82.2	**68.9**	42.3	41.5

Fig. 9. Saliency maps of features from the last layer of the ResNet-50 backbone for different models. The regions with colors closer to red indicate areas of higher feature response intensity.
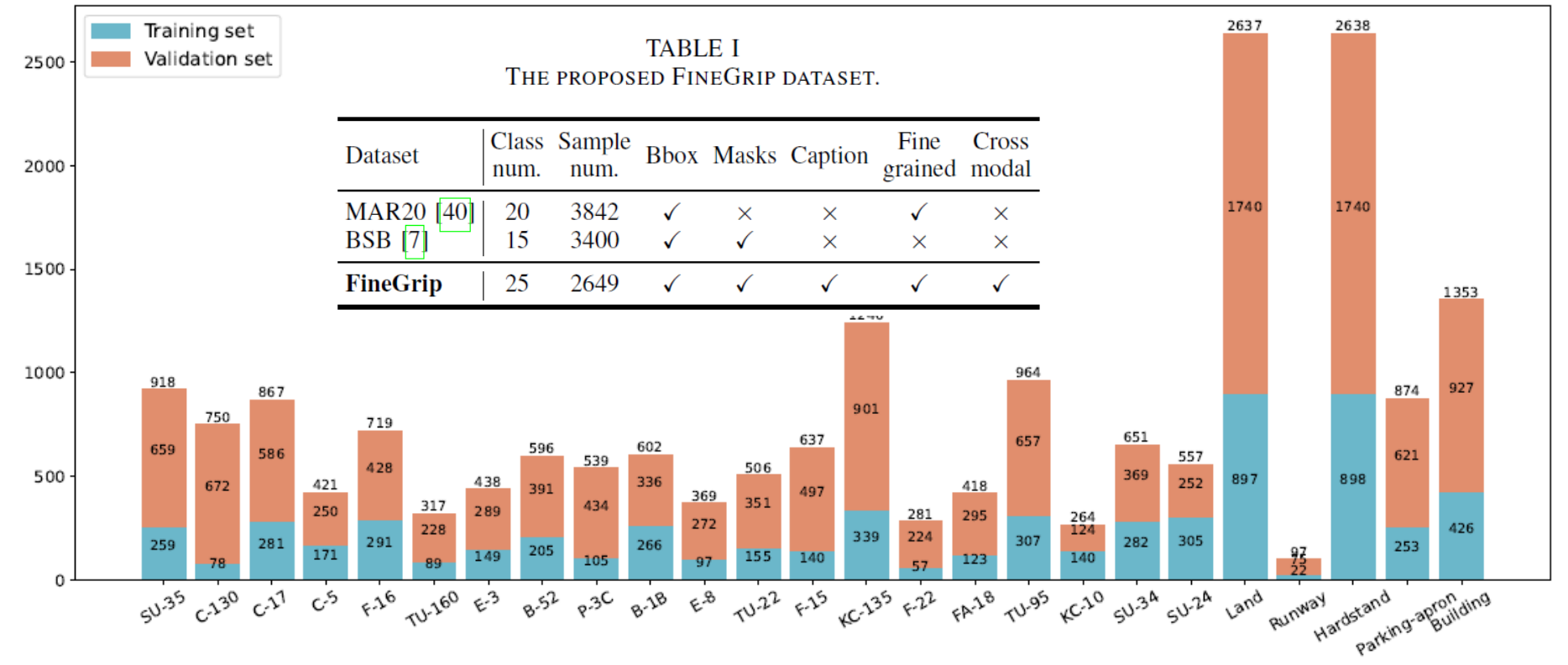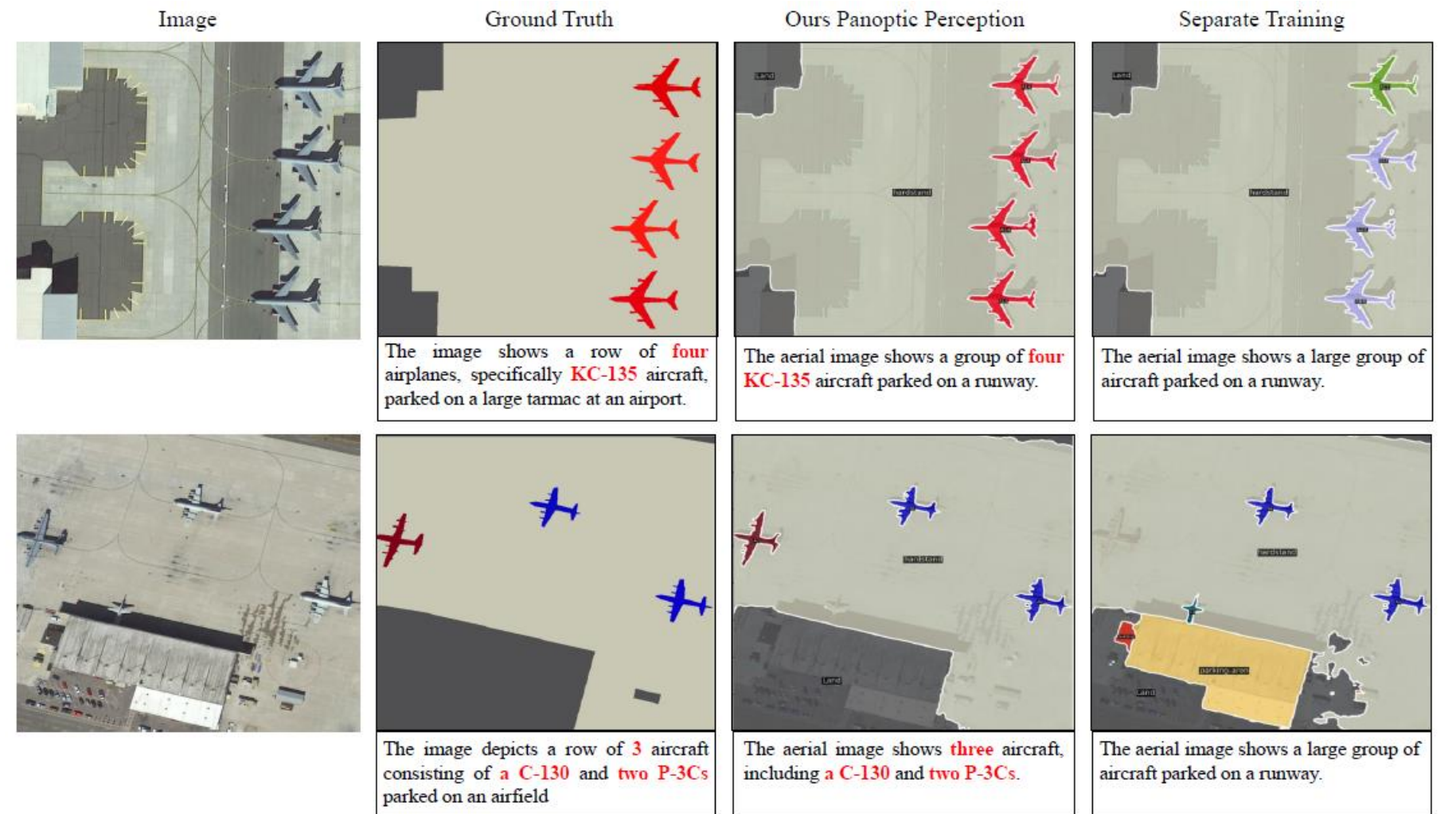

Fig. 8. Panoptic perception results from the jointly trained and separately trained models. The first column displays the ground-truth mask and one of the five caption sentences for each image.