SCORE: Scene Context Matters in Open-Vocabulary Remote Sensing Instance Segmentation

Supplementary Material

In the supplementary materials, we provide more information on the datasets used for open-vocabulary remote sensing instance segmentation benchmark and include more qualitative results along with comparisons. Moreover, we show that **SCORE** can also enhance the performance for open-vocabulary remote sensing semantic segmentation task, which further unleashes the potential of our model.

A. Implementation Details

A.1. Remote Sensing Instance Segmentation

Training Dataset. Following the open-vocabulary benchmarks for natural images [55], we train the model on one dataset and evaluate its cross-dataset performance on other datasets. We select two datasets for training, *i.e.*, iSAID [64] and SIOR [48]. iSAID is a large scale instance segmentation dataset for remote sensing images. It contains 18732 images for training across 15 categories. SIOR is developed from aerial object detection dataset DIOR [25], with segmentation annotations generated in SAMRS [48], which contains 11725 images with 20 categories.

Evaluation Dataset. To evaluate the effectiveness of our method, we conduct cross-dataset evaluation on 4 aerial instance segmentation datasets, *i.e.*, NWPU-VHR-10 [7, 44], SOTA [48], FAST [48], and SIOR [48]. NWPU-VHR-10 is an aerial object detection dataset with instance masks further annotated by [44]. The test set contains 731 images from 10 aerial classes. SOTA, FAST and SIOR are segmentation datasets provided in SAMRS [48], which are developed from aerial object detection datasets DOTA-V2.0 [11], FAIR1M-2.0 [47], and DIOR [25], respectively. SOTA covers 874 images with 18 object categories for testing, FAST contains 3207 images across 37 fine-grained aerial object categories for testing, and SIOR is with 11738 testing samples. We provide the categories in each dataset in Table B.

A.2. Remote Sensing Semantic Segmentation

Training Dataset. Following the open-vocabulary benchmarks for remote sensing semantic segmentation [60], we train the model on their proposed LandDiscover50K dataset. It includes 51846 high-resolution remote sensing images annotated across 40 object categories.

Evaluation Dataset. To evaluate the effectiveness of our method, we follow the evaluation settings in [60] to conduct cross-dataset evaluation on 4 remote sensing semantic datasets, *i.e.*, FLAIR [13], FAST [48], Potsdam [20], and FloodNet [41]. Each dataset has its own bias towards different remote sensing categories. To illustrate, Potsdam [20] emphasizes the in-vocabulary performance with high category similarity to the training LandDiscover50K dataset, which contains 5472 images with 6 semantic categories. FloodNet [41] focuses more on the post-flood analysis, which contains 898 images with 9 semantic categories. FLAIR [13] is with 15700 images focusing on 12 large-scale landcover types. FAST [48] contains 3207 images, specializing in 37 fine-grained semantic classes for remote sensing. The combination of the four datasets enables a comprehensive evaluation of the open-vocabulary semantic segmentation tasks in remote sensing. We provide the categories in each dataset in Table C.

B. Additional Experiment Results

B.1. Additional Results on Semantic Segmentation

The proposed **SCORE** can also be applied to diverse segmentation related tasks, *e.g.* semantic segmentation. We provide the semantic segmentation results of our method in Table A. Our approach consistently outperforms existing across three of four benchmarks, demonstrating its effectiveness in open-vocabulary remote sensing semantic segmentation. Specifically, we achieve an average improvement of 1.13% over the current SOTA model [61]. Our method surpasses previous methods by a large margin, especially on FLAIR and FAST datasets, with gains up to 9.62%.

	LandDiscover50K				
Method	FLAIR	FAST	Potsdam	FloodNet	Average
CAT-SEG [10] [CVPR24]	19.71	15.55	39.57	35.91	27.69
GSNet [61] [AAAI25]	18.35	15.21	**43.29**	37.68	28.63
SCORE (Ours)	**29.33**	**21.51**	26.51	**41.70**	**29.76**

Table A. **Comparison with SOTA methods on open-vocabulary remote sensing semantic segmentation.** The model is trained on LandDiscover50K dataset and then tested on the four evaluation benchmarks to measure its cross-dataset generalization capabilities.

B.2. Additional Qualitative Results

We provide additional qualitative results of our proposed method on remote sensing instance segmentation task as shown in Figure A.

Dataset	#Category	Category Name
iSAID [64]	15	ship, storage tank, baseball diamond, tennis court, basketball court, ground track field, bridge, large vehicle, small vehicle, helicopter, swimming pool, roundabout, soccer ball field, plane, harbor
SIOR [48]	20	airplane, airport, baseball field, basketball court, bridge, chimney, expressway service area, expressway toll station, dam, golffield, ground track field, harbor, overpass, ship, stadium, storage tank, tennis court, train station, vehicle, windmill
NWPU [7, 44]	10	airplane, ship, storage tank, baseball diamond, tennis court, basketball court, ground track field, harbor, bridge, vehicle
FAST [48]	37	A220, A321, A330, A350, ARJ21, baseball field, basketball court, Boeing737, Boeing747, Boeing777, Boeing787, bridge, bus, C919, cargo truck, dry cargo ship, dump truck, engineering ship, excavator, fishing boat, football field, intersection, liquid cargo ship, motorboat, other-airplane, other-ship, other-vehicle, passenger ship, roundabout, small car, tennis court, tractor, trailer, truck tractor, tugboat, van, warship
SOTA [48]	18	large vehicle, swimming pool, helicopter, bridge, plane, ship, soccer ball field, basketball court, ground track field, small vehicle, baseball diamond, tennis court, roundabout, storage tank, harbor, container crane, airport, helipad

Table B. **Category Names for datasets used in our instance segmentation benchmarks.**

Dataset	#Category	Category Name
LandDiscover50K [61]	40	background, bare land, grass, pavement, road, tree, water, agriculture land, buildings, forest land, barren land, urban land, large vehicle, swimming pool, helicopter, bridge, plane, ship, soccer ball field, basketball court, ground track field, small vehicle, baseball diamond, tennis court, roundabout, storage tank, harbor, container crane, airport, helipad, chimney, expressway service area, expresswalltoll station, dam, golf field, overpass, stadium, train station, vehicle, windmill
FLAIR [13]	12	building, pervious surface, impervious surface, bare soil, water, coniferous, deciduous, brushwood, vineyard, herbaceous vegetation, agricultural land, plowed land
FAST [48]	37	A220, A321, A330, A350, ARJ21, baseball field, basketball court, Boeing737, Boeing747, Boeing777, Boeing787, bridge, bus, C919, cargo truck, dry cargo ship, dump truck, engineering ship, excavator, fishing boat, football field, intersection, liquid cargo ship, motorboat, other-airplane, other-ship, other-vehicle, passenger ship, roundabout, small car, tennis court, tractor, trailer, truck tractor, tugboat, van, warship
Potsdam [20]	6	impervious surface, building, low vegetation, tree, car, clutter
FloodNet [41]	9	building-flooded, building-non-flooded, road-flooded, road-non-flooded, water, tree, vehicle, pool, grass

Table C. **Category Names for datasets used in our semantic segmentation benchmarks.**

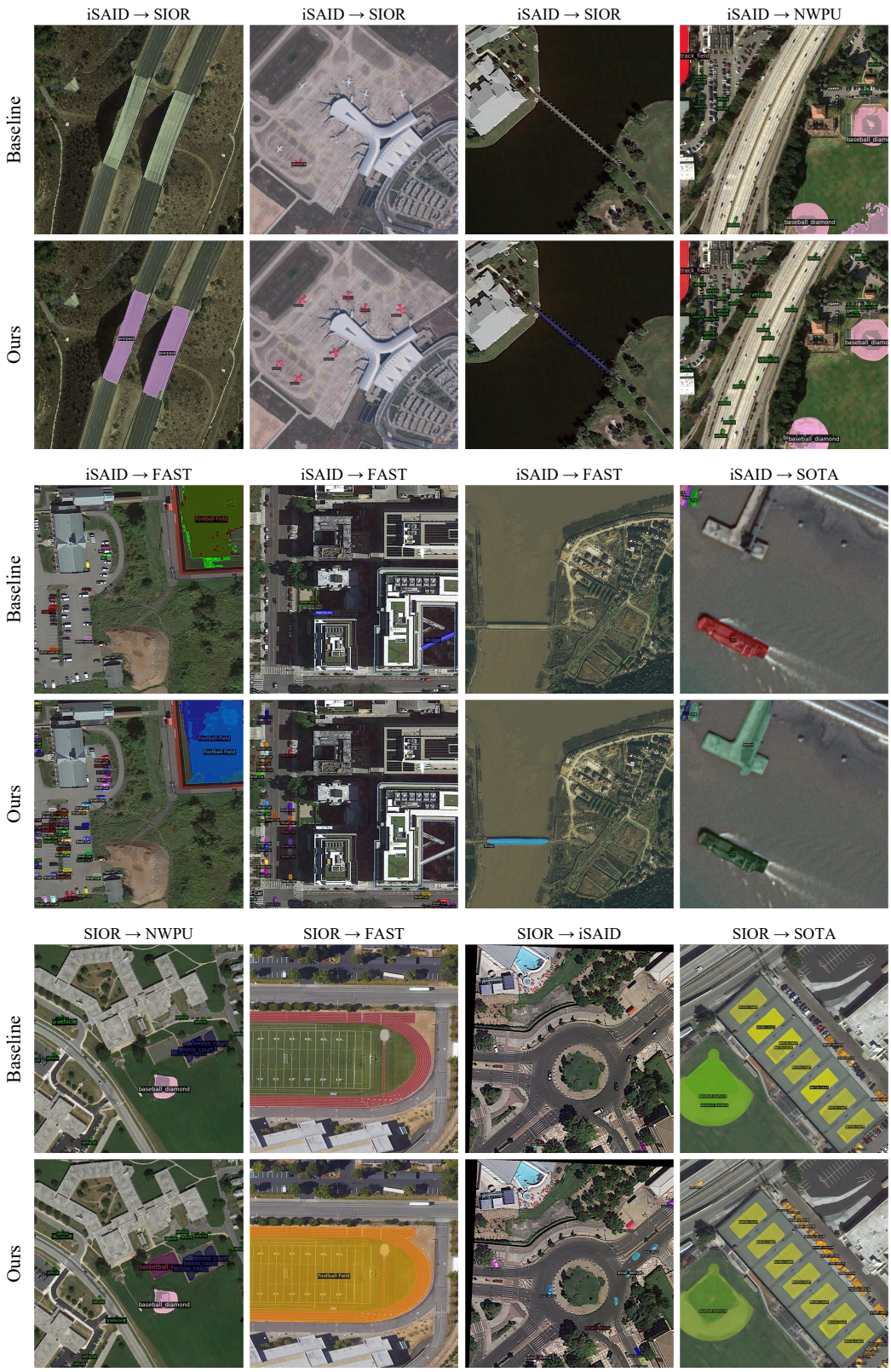

Figure A. **Additional qualitative results between the baseline and our model.**