SkySense-O: Towards Open-World Remote Sensing Interpretation with Vision-Centric Visual-Language Modeling

Supplementary Material

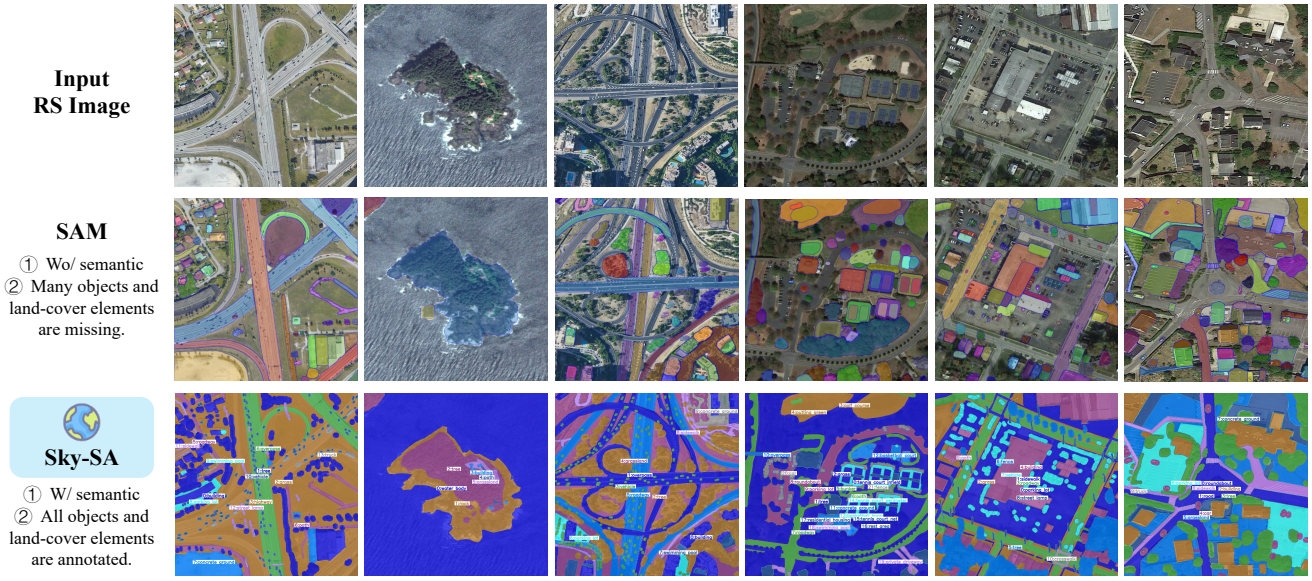

Figure 1. Comparison between the Sky-SA dataset and results produced directly using SAM. Comprehensive segmentations of all pixels in remote sensing with semantic features are outstandingly shown in our dataset.

Table of Contents:

In this paper, we present SkySense-O, a open-world remote sensing interpretation model, which is mainly driven by the proposed semantically-rich and fine-grained Sky-SA dataset. Due to space limitations, some details are reduced in the draft version. We provide the materials to supplement our paper and divide them into 8 sections as follows.

- §Sec. 1: Key Advantages over SAM.
- §Sec. 2: Sky-SA Dataset Details.
- §Sec. 3: Vision-Centric Knowledge Graph Details.
- §Sec. 4: Generalization Ability to Unseen Categories.
- §Sec. 5: Experiment Implementation Details.
- §Sec. 6: Limitations.
- §Sec. 7: More Image-Text Alignment Visualizations.
- §Sec. 8: More Predictions across Different Datasets.

1. Key Advantages over SAM

In this work, we aim to achieve pixel-level open-world remote sensing interpretation. In the domain of natural images, a common approach leverages the combination of two foundational models: the Segment Anything Model (SAM) and Contrastive Language–Image Pre-training (CLIP). This combination method decomposes the open-world interpretation task into two stages, namely segmentation and recognition. Although these approaches have achieved significant success on natural images, we find that both SAM and CLIP exhibit suboptimal performance when applied to remote sensing images. This observation motivates us to develop a RS-specific foundational model. In this section, we primarily discuss the segmentation limitations of SAM in remote sensing images, which is the main reason that drives us to annotate the Sky-SA dataset.

Why Not Directly Use SAM? Despite numerous existing works [10, 12, 19] that directly employ SAM to generate pre-segmented results, we observe that SAM often strug-

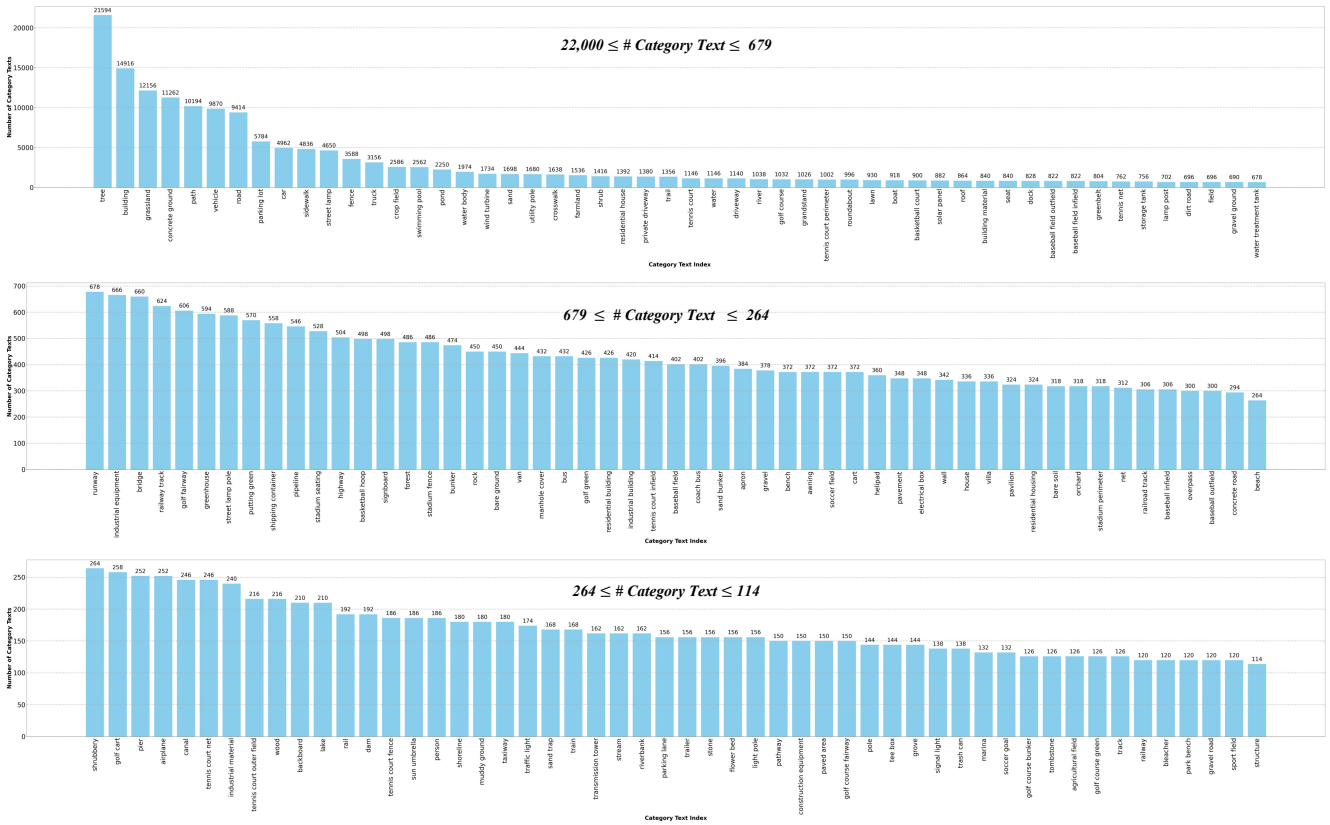

Figure 2. Distribution statistics of the number of category texts in the Sky-SA dataset.

gles to successfully segment all relevant regions in remote sensing images (see Fig. 1). This limitation is particularly pronounced in the following two conditions:

1) Land-cover Elements (*e.g.*, overpasses, roads, soil): As primary categories in remote sensing imagery, land-cover elements frequently span extensive spatial areas and often intersect or overlap with one another. Such intricate spatial distribution present significant challenges for SAM in achieving accurate segmentation.

2) Targets with Large Scale Variations (*e.g.*, vehicles and buildings): The targets in remote sensing images usually exhibit substantial variations in spatial scales. However, when performing zero-shot segmentation, SAM typically relies on spatially uniform sampling of visual prompts to guide the segmentation process. This strategy makes it difficult for SAM to effectively handle situations with large-scale variations. For example, if the sampling interval of point prompts is large, small targets may be missed. Conversely, if the sampling interval is small, large targets may be unnecessarily subdivided into multiple parts, increasing the complexity of assigning semantic labels to the segments.

2. Sky-SA Dataset Details

In this section, we present numerous examples from Sky-SA dataset to illustrate the high quality and diversity of the annotated masks, which we have extensively verified. Beyond its use in training *SkySense-O* to be robust and general, we hope the Sky-SA dataset becomes a valuable resource for research aiming to build new foundation models.

The Necessity of Manual Annotation. Despite the popularity of current self-supervised paradigms, achieving high fine-grained local segmentation remains challenging through self-supervision manner alone. This is probably due to insufficient propagation of local image-text pair data within the network. Therefore, we adopt the same manual annotation strategy as SAM, namely a closed loop of model prediction and manual correction. Specifically, the data annotation can be divided into two stages: *assisted-manual*, *semi-automatic*. In the first stage, we utilize GPT-4V to assign anticipated word labels to the images. We then apply Grounding DINO and SAM to sparsely pre-annotate target categories within the images. Building upon these preliminary annotations, annotators employ adding and erasing tools to adjust the category labels and fill in any missing annotations, similar to a classic segmentation setup. In the second stage, *SkySense-O* already possess certain segmenta-

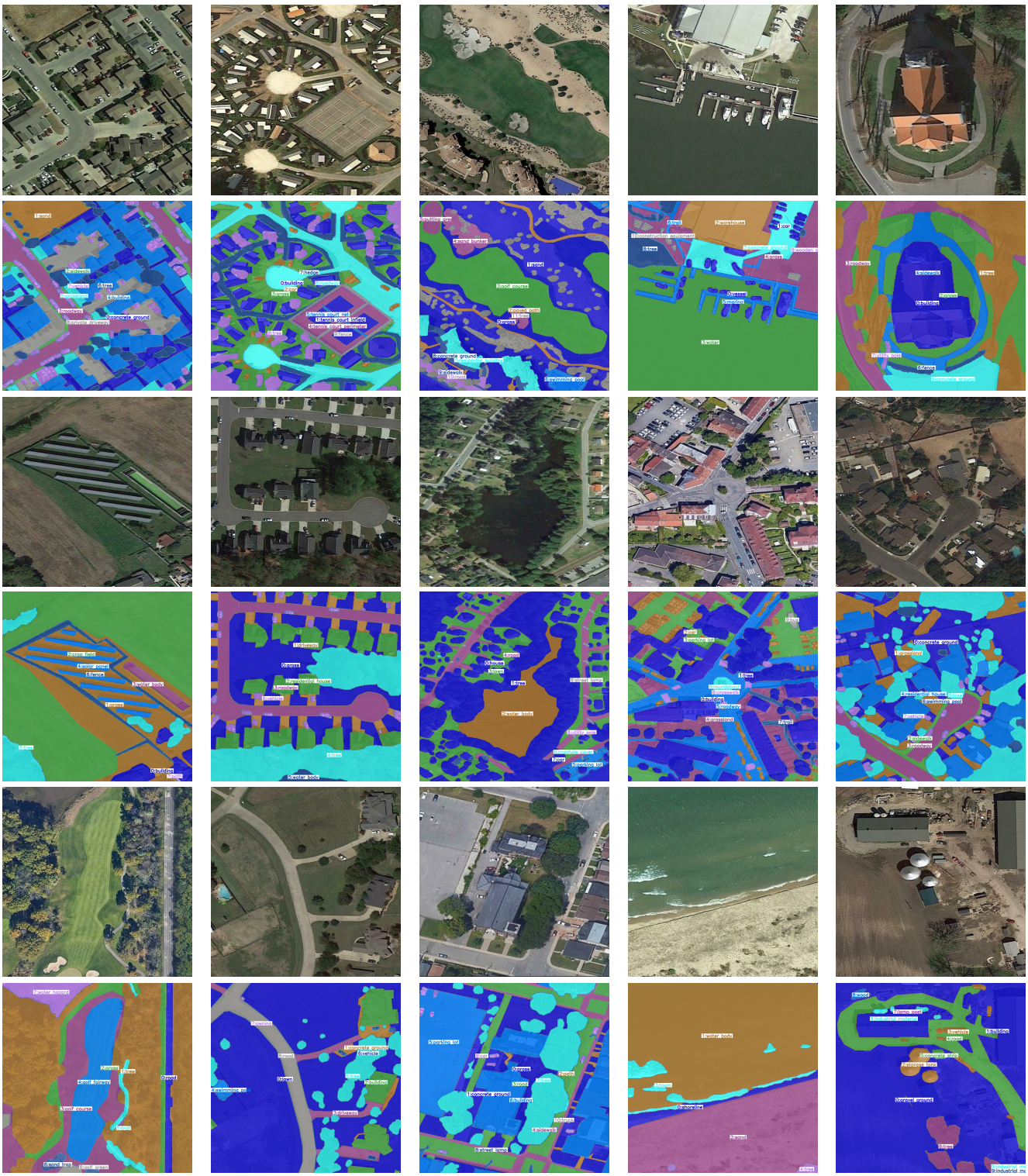

Figure 3. Example images and annotations with open texts and masks from our newly proposed dataset, **Sky-SA**, which is characterized by its textual openness and high density of masks.

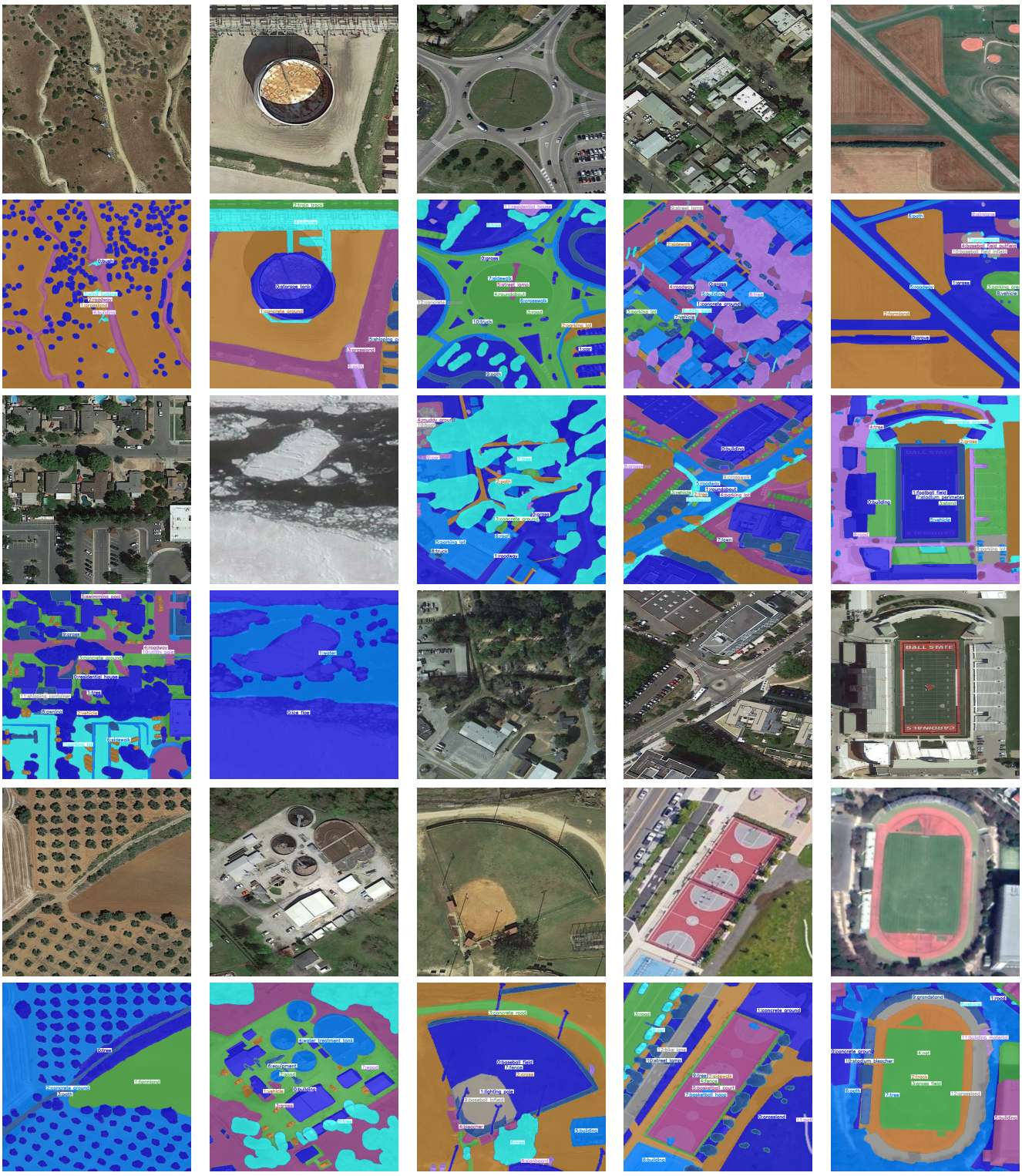

Figure 4. Another images and annotations examples with open texts and masks from our newly proposed dataset, **Sky-SA**, which is characterized by its textual openness and high density of masks.

Figure 5. Another images and annotations examples with open texts and masks from our newly proposed dataset, **Sky-SA**, which is characterized by its textual openness and high density of masks.

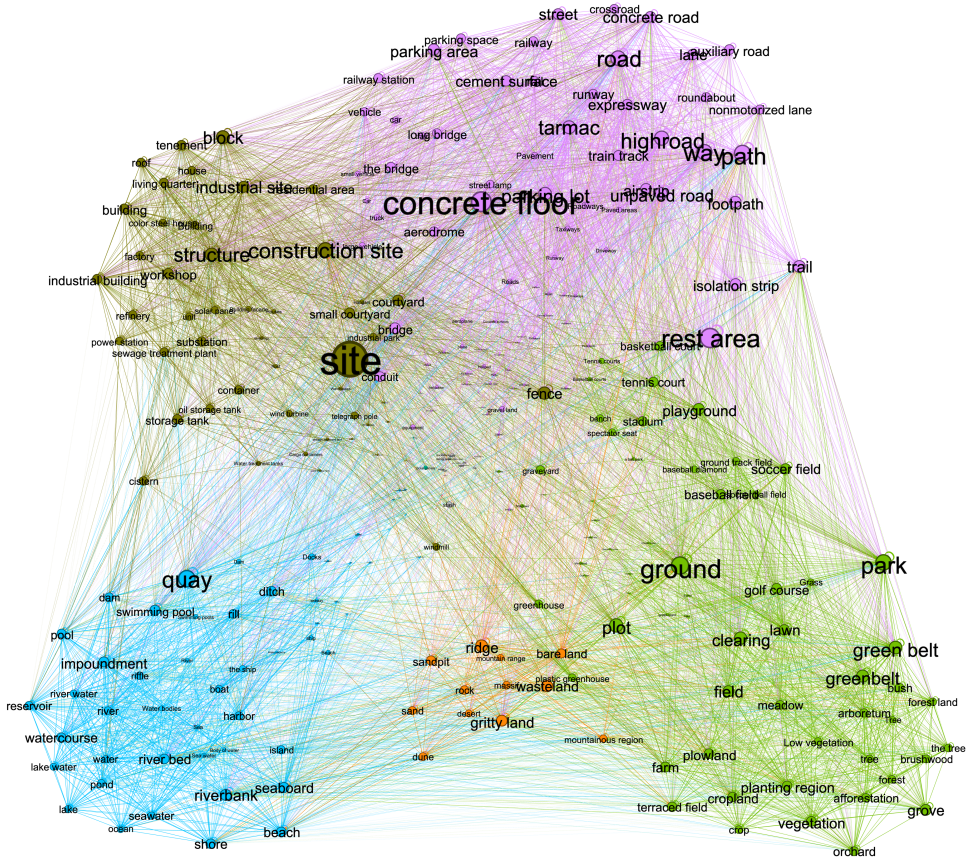

Figure 6. Zoomed version of the proposed vision-centric knowledge graph for more node details.

Model	Publication	plane	soccer ball field	small vehicle	harbor	others	Average
SAN [17]	CVPR 2023	23.54	5.03	24.38	2.37	26.75	16.41
CAT-SEG [3]	CVPR 2024	28.08	6.72	17.07	4.43	71.93	25.64
SkySense-O	-	**33.36**	**43.58**	**31.74**	**17.55**	**90.73**	**43.39**
	-	(+5.28)	(+36.86)	(+7.36)	(+13.12)	(+18.80)	(+17.75)

Table 1. Zero-shot remote sensing semantic segmentation performances for unseen classes on the SOTA dataset. Bold indicates the highest performances in mIoU (%).

Model	Publication	A220	dry cargo ship	intersection	liquid cargo ship	motorboat	excavator	others	Average
SAN [17]	CVPR 2023	0.00	4.50	0.26	2.81	0.18	0.27	13.03	3.01
CAT-SEG [3]	CVPR 2024	0.01	0.21	1.49	3.12	0.25	0.75	63.33	9.88
SkySense-O	-	**2.26**	**12.00**	**3.28**	**10.04**	**2.06**	**2.65**	**95.21**	**18.21**
	-	(+2.25)	(+7.50)	(+1.79)	(+6.92)	(+1.81)	(+1.90)	(+31.88)	(+8.33)

Table 2. Zero-shot remote sensing semantic segmentation performances for unseen classes on the FAST dataset. Bold indicates the highest performances in mIoU (%).

tion abilities and we use it to automatically generate masks for new RS images. Annotators can then perform corrections and completions on the predictions from *SkySense-O*, mirroring the process in the first stage. This closed-loop

-Goal-

Given a source entity and a target entity list observed from remote sensing images, where each target entity is enclosed in quotation marks and separated by comma, you are an intelligent assistant that helps a human analyst to measure the appearance similarity between the source entity and each target entity from the target entity list.

-Steps-

1. Describe the common shape, color, material of the source entity and all target entities in remote sensing images.
- entity_name: Name of the source entity
- entity_type: Source entity / Target entity
- entity_description: Comprehensive description of the source entity's appearance in remote sensing images

Format each entity as ("entity"{tuple_delimiter}<entity_name>{tuple_delimiter}<entity_type>{tuple_delimiter}<entity_description>{tuple_delimiter})

2. From the entity attibutes in step 1, score the imaging appearance similarity between the source entity and each target entity in the target entity list.
For each pair of source entity and target entity, extract the following information:
- source_entity: name of the source entity, as identified in step 1
- target_entity: name of the target entity, as identified in step 1
- relationship_strength_score: a numeric score indicating strength of the relationship between the source entity and target entity
- relationship_description: explanation as to why you give the relationship strength score between the source entity and the target entity.

Format each relationship as
("relationship"{tuple_delimiter}<source_entity>{tuple_delimiter}<target_entity>{tuple_delimiter}<relationship_strength_score>{tuple_delimiter}<relationship_description>)

3. Return output in English as a single list of all the entities and relationships identified in steps 1 and 2. Use **{record_delimiter}** as the list delimiter.

4. When finished, output {completion_delimiter}. Do not output full information; the output format MUST be following example.

-Audience-
The output audience is the semantic segmentation algorithm in remote sensing, where the labeled entity in the ground truth will be replaced with the target entity you give high score, but the mask label will remain unchanged. After the entity is changed, the source mask can still match the replaced entity.

######################
-Examples-
######################
Example 1:

Source entity: "Apple tree"
Target entity list:
["Apple tree", "Apple", "Orange tree", "The apple tree"]
##############################
Output:
("relationship"{tuple_delimiter}"Apple tree"{tuple_delimiter}"Apple tree"{tuple_delimiter}"10"{tuple_delimiter}'Apple tree' and 'Apple tree' represent the same object on remote sensing images.) {record_delimiter}
("relationship"{tuple_delimiter}"Apple tree"{tuple_delimiter}"tree"{tuple_delimiter}"10"{tuple_delimiter}'Apple tree' mask can be relaced by 'tree' mask because they have similar visual characteristics in remote sensing semantic segmentation.) {record_delimiter}
("relationship"{tuple_delimiter}"tree"{tuple_delimiter}"Apple tree"{tuple_delimiter}"7"{tuple_delimiter}'tree' mask often not be relaced by 'apple tree' mask because 'apple tree' is just a subset of 'tree' in remote sensing semantic segmentation.) {record_delimiter}
("relationship"{tuple_delimiter}"Apple tree"{tuple_delimiter}"Apple"{tuple_delimiter}"0"{tuple_delimiter}'Apple tree' mask can not be relaced by 'Apple' mask because they have completely different visual shapes and colors in remote sensing semantic segmentation.){record_delimiter}
("relationship"{tuple_delimiter}"Apple tree"{tuple_delimiter}"Orange tree"{tuple_delimiter}"7"{tuple_delimiter}'Apple tree' mask may be relaced by 'Orange tree' mask because they are both trees and have similar looks on remote sensing images.){record_delimiter}
("relationship"{tuple_delimiter}"Apple tree"{tuple_delimiter}"The apple tree"{tuple_delimiter}"10"{tuple_delimiter}'Apple tree' and 'The apple tree' represent the same object on remote sensing images.){record_delimiter}

Figure 7. Main prompts used in design of vision-centric agent.

Model	Publication	airport	train station	windmill	harbor	others	Average
SAN [17]	CVPR 2023	1.77	4.59	1.19	5.50	20.26	6.66
CAT-SEG [3]	CVPR 2024	2.86	8.43	5.99	8.17	61.48	17.39
SkySense-O	-	**3.36**	**27.87**	**16.99**	**22.05**	**89.25**	**31.90**
	-	(+0.50)	(+19.44)	(+11.00)	(+13.88)	(+27.77)	(+14.51)

Table 3. Zero-shot remote sensing semantic segmentation performances for unseen classes on the SIOR dataset. Bold indicates the highest performances in mIoU (%).

Source Dataset	Samples Num
NWPU-RESISC-45 [2]	9000
UCM-Landuse [9]	2100
RSITMD [18]	948
EarthVQA [13]	145368
RSVQA-LR [7]	47173
DOTA-v2.0 [4]	20000
FAIR1M [11]	40000
FIT-RS (subset) [8]	327955

Table 4. Data details of the samples in the instruction fine-tuning dataset for VQA experiments.

approach substantially enhances annotation efficiency.

3. Vision-Centric Knowledge Graph Details

In this work, we observe that when using text embeddings as prompts, it is challenging to decouple categories with strong contextual relationships, such as *car* and *parking lot*. In contrast, we note that these categories exhibit significant visual differences. This observation naturally inspire us to consider leveraging the visual features of these categories to help decouple them, which could be an effective solution. As illustrated in the main text, to realize this insight, we construct a visual-centric agent utilizing GPT-4V to obtain the visual relevance score for each text pair in the Sky-SA dataset. The scoring employs a meticulously designed chain-of-thought process that assesses a series of visual features, including appearance, color distribution, and overall visual shape. Examples of the prompts used and details of the constructed graphs are illustrated in Fig. 6 and Fig. 7, which demonstrate that our constructed knowledge graph performs relationship clustering centered on visual features, verifying that this is a direct yet effective approach.

Why Not Use Visual Representation from Foundation Model? We indeed experiment with directly using visual foundation models to score visual similarity based on the corresponding visual region features of these categories. While this manner appears more reasonable and elegant, our preliminary experiments indicate that the proper visual similarity matching by existing foundation models may be limited to larger-scale regions, for dense or small objects, such as car and parking lot, the unsuitable receptive fields and information compression in visual embeddings may hinder effective visual correspondence. We hope this issue can be decently addressed in future research.

What Insight Does the Approach Inspire? Recently, some open-world models [14, 16] employed text-to-image diffusion models as additional agent inputs. Compared with them, we find that using large language models (LLMs) can also achieve accurate visual similarity scoring. The success of this approach confirms that LLMs inherently contain sufficient visual knowledge, and effective visual matching could be achieved without the additional step of text-to-image generation. This insight also resonates with the current mainstream MLLMs alignment paradigm via aligning visual models to large language models. Moreover, this phenomenon suggests that visual tokens could be compressed into large language models, as in recent work [1]. We hope that this observation can inspire more alignment designs for MLLMs.

4. Generalization Ability to Unseen Categories

In this section, to validate the generalization ability of SkySense-O when extended to open-world models, we compare the performance of SkySense-O with other models on unseen categories in the Sky-SA dataset (see Table 1 2 3).

5. Experiment Implementation Details

Details of VQA Baselines. As mentioned in the main paper, we employ SkySense-O in MLLM as the vision encoder for experimentation. We set the batch size to 256 for both the pre-training and fine-tuning stages. For the instruction fine-tuning dataset, we select a 328k subset of FIT-RS and collect existing public datasets as in [8]. These public datasets include 3 scene classification datasets (NWPU [2], UCM [9], and RSITMD [18]), 2 VQA datasets (EarthVQA [13], and RSVQA-LR [7]), and 2 object detection datasets (DOTA-v2.0 [4] and FAIR1M [11]) to enrich the instruction dataset. Using the same instruction fine-tuning dataset, We compare the SkySense-O with CLIP-L-14 and CLIP-H-14, the results are shown in Tab. 5, SkySense-O demonstrates superior performance compared to both CLIP-L and CLIP-H.

Evaluation Details of Few-Shot Baselines. We follow the one-shot evaluation approach as in DINOv [6]. In this approach, for a test image containing multiple categories, we evaluate each category separately using class-specific prompts. For example, for a test image containing categories of 'water,' 'building,' and 'road,' we perform independent inference for each category using class-specific prompts. Specifically, to assess the 'water' in the test image, we randomly select another image containing 'water' from the test set and use it, along with its 'water' label, as a prompt. The model then predicts the 'water' regions in the test image based on this prompt. This procedure is repeated for each category present in the image. It is important to note that this one-shot evaluation focuses only on the categories present in the image, which may lead to higher performance metrics compared to zero-shot evaluation, as it

	SIRI-WHU [20]	AID-VQA [15]	RSVQA-HR [7]
Visual Encoder	Acc	Acc	Avg Acc
CLIP-L-14	70.12	91.80	74.50
CLIP-H-14	70.63	91.55	75.45
SkySense-O†(ours)	**74.79** (+4.16)	**94.10** (+2.30)	**78.09** (+2.64)

Table 5. Comparison of SkySense-O with CLIP series in zero-shot VQA task.

does not account for categories absent from the image.

Upsampling Visual Decoder. In our upsampling visual decoder, we start by taking the features in the last layer of the Swin model. Initially, the extracted feature maps have a resolution of 24×24 pixels, after processing them with the transposed convolution operation, we increase their resolution to 24×24 pixels to 384×384 pixels. Moreover, we employ the visual decoder architecture similar to CAT-SEG [3], the spatial aggregation and the class aggregation modules are introduced for processing the rough image-text alignment results. All hyperparameters are kept constant across the evaluation datasets.

Text Prompt Templates. To extract text embeddings from the text encoder, we construct sentences using class names, for example, `"A remote sensing image of {class}"`. Although we do not delve into the use of handcrafted prompts in this study, we acknowledge it as a potential area for future research.

6. Limitations

SkySense-O specializes in leveraging textual prompts for remote sensing interpretation tasks, an effective approach for common RS categories. However, for less frequent categories in Earth observation tasks, such as pig farms and wildfires, the incorporation of visual prompts may be a more appropriate choice. Consequently, constructing a more comprehensive and flexible prompt encoder is an intriguing topic for achieving open-world remote sensing interpretation.

7. More Image-Text Alignment Visualizations

In the main text, we follow MaskCLIP [5] and adopt the image-text correlation in the representation space to evaluate the effectiveness of vision-language foundation models. First, we provide here the list of categories used for the x-axis of the histogram in Fig.7(c) of the main text. This list includes a total of 131 categories, which are as follows: {*airplane, airport, baseball field, basketball court, bridge, expressway service area, dam, golf field, harbor, ship, football field, storage tank, tennis court, train station, vehicle, windmill, swimming pool, impervious surfaces, building, low vegetation, tree, way, clearing, parking lot, wasteland, highroad, lane, field, water, river, cropland, plowland, lake, unpaved road, pond, expressway, fence, solar panel, path, unit, ridge, brushwood, sandpit, runway, greenhouse, footpath, parking space, courtyard, mountain range, island, beach, orchard, roof, site, rock, graveyard, massif, block, planting region, stadium, sand, lake water, street lamp, seawater, train track, house, plot, conduit, impoundment, sewage treatment plant, shore, dune, ditch, park, river water, seaboard, forest land, crop, color steel house, bench, watercourse, reservoir, container, isolation strip, stash, parking area, train, small courtyard, afforestation, farm, refinery, street, gritty land, crossroad, trail, auxiliary road, spectator seat, plastic greenhouse, terraced field, airstrip, river bed, seat, bare land, rill, riffle, nonmotorized lane, telegraph pole, misc, construction site, separator, grandstand, hamlet, cultivated land, gravel land, gully, oil tank, mountain road, rockily, dumping ground, solar power plant, gazebo, a small island, equipment, industrial park, green plant, sewage treatment tank, pitcher, canal, cloud, the helipad, seaway*}. Then, in Fig. 8, we present more image-text correlation maps to further validate the effectiveness of SkySense-O in image-text alignment in open-world scenes.

8. More Predictions across Different Datasets

In this section, we provide numerous results visualizations on open-world RS interpretation across various datasets as shown in Fig. 9 10 11. These visualizations are utilized to provide a more comprehensive and intuitive assessment of SkySense-O's performance.

References

[1] Anonymous. LLaVA-mini: Efficient image and video large multimodal models with one vision token. In *Submitted to The Thirteenth International Conference on Learning Representations*, 2024. under review. 8

[2] Gong Cheng, Junwei Han, and Xiaoqiang Lu. Remote sensing image scene classification: Benchmark and state of the art. *Proceedings of the IEEE*, 105(10):1865–1883, 2017. 8

[3] Seokju Cho, Heeseong Shin, Sunghwan Hong, Anurag Arnab, Paul Hongsuck Seo, and Seungryong Kim. Cat-seg: Cost aggregation for open-vocabulary semantic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4113–4123, 2024. 6, 7, 9

[4] Jian Ding, Nan Xue, Gui-Song Xia, Xiang Bai, Wen Yang, Michael Ying Yang, Serge Belongie, Jiebo Luo, Mihai

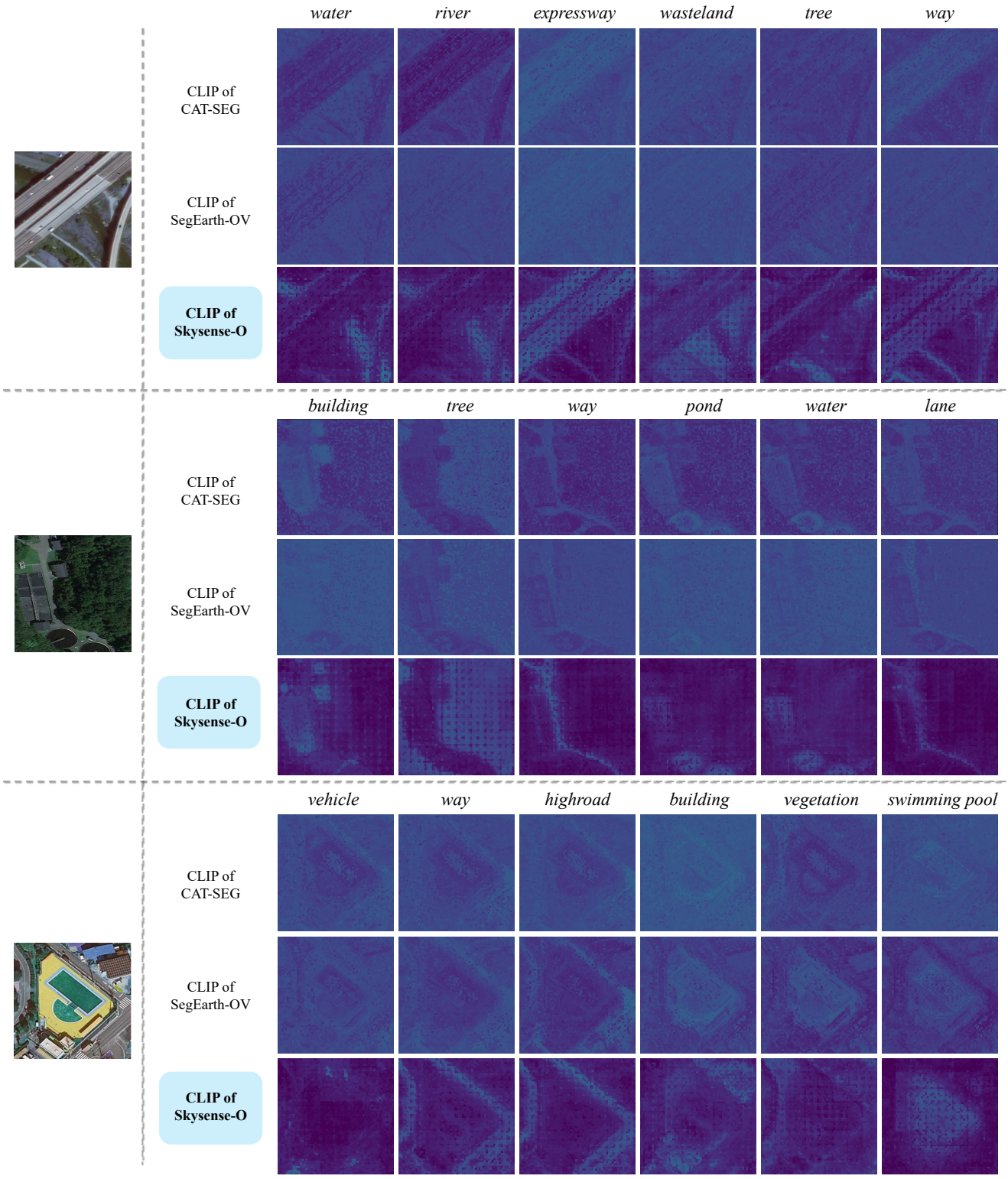

Figure 8. More visualizations of dense features in the image-text alignment for different categories.

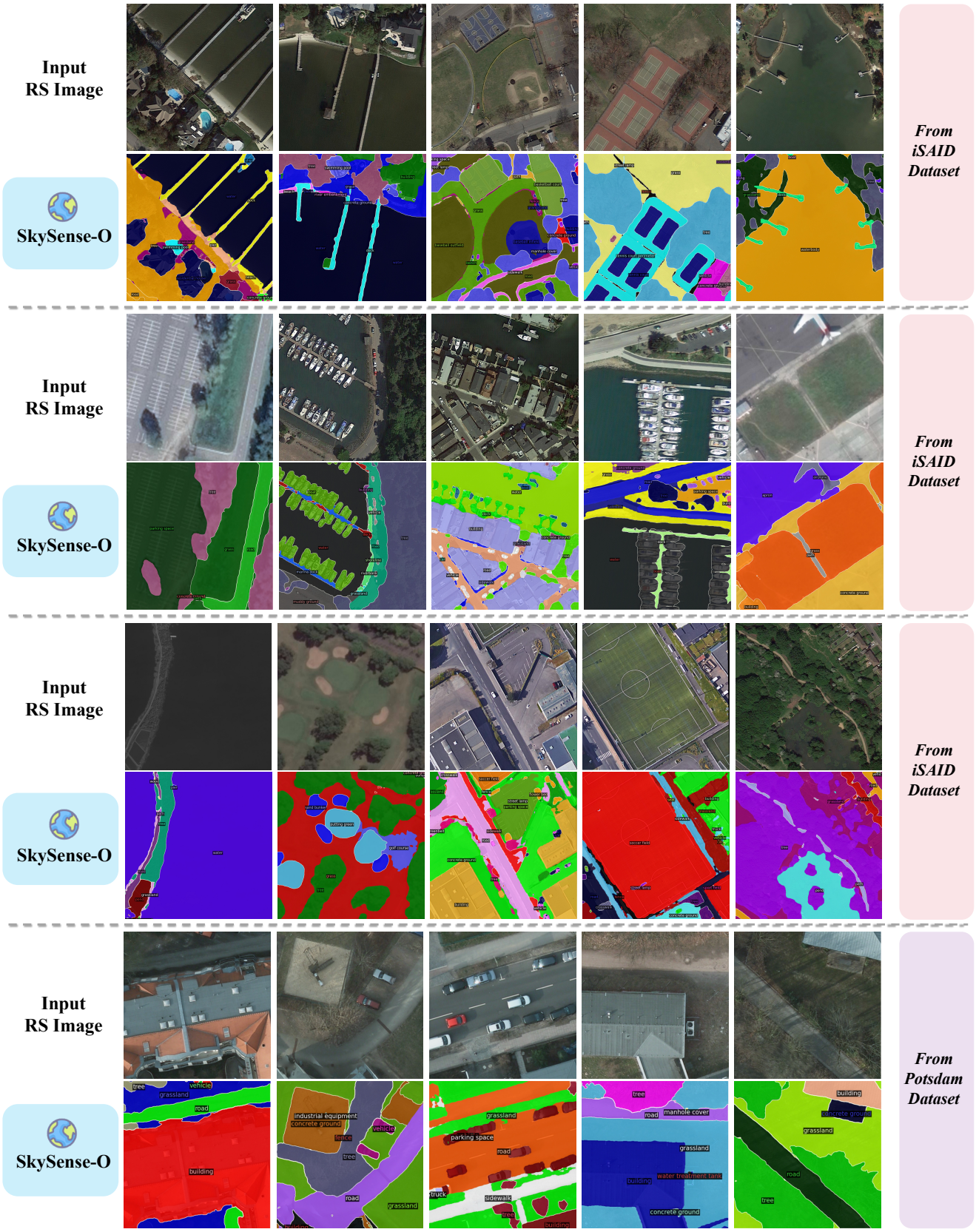

Figure 9. Visualizations of open-world interpretation results by SkySense-O on iSAID and Potsdam datasets.

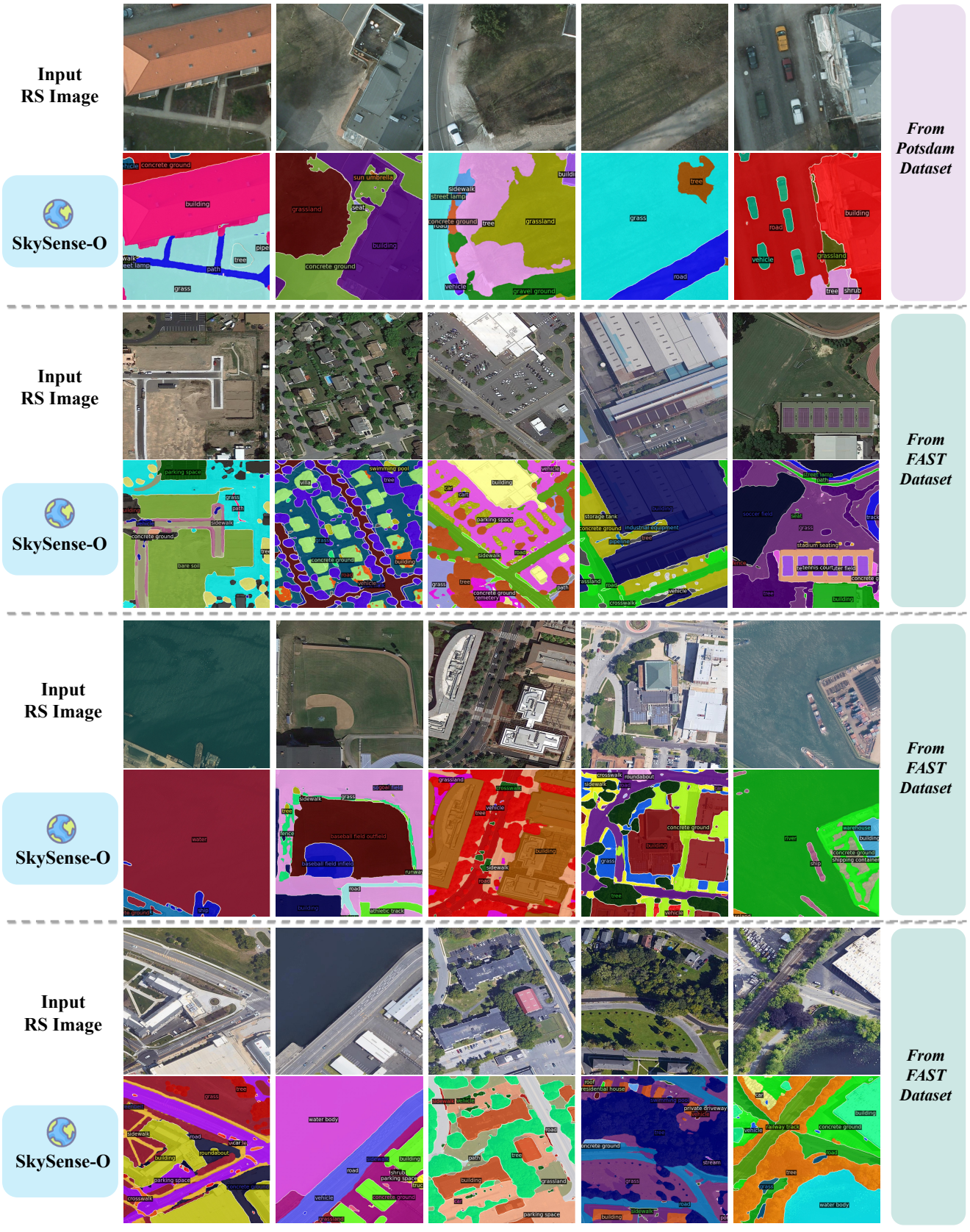

Figure 10. Visualizations of open-world interpretation results by SkySense-O on Potsdam and FAST datasets.

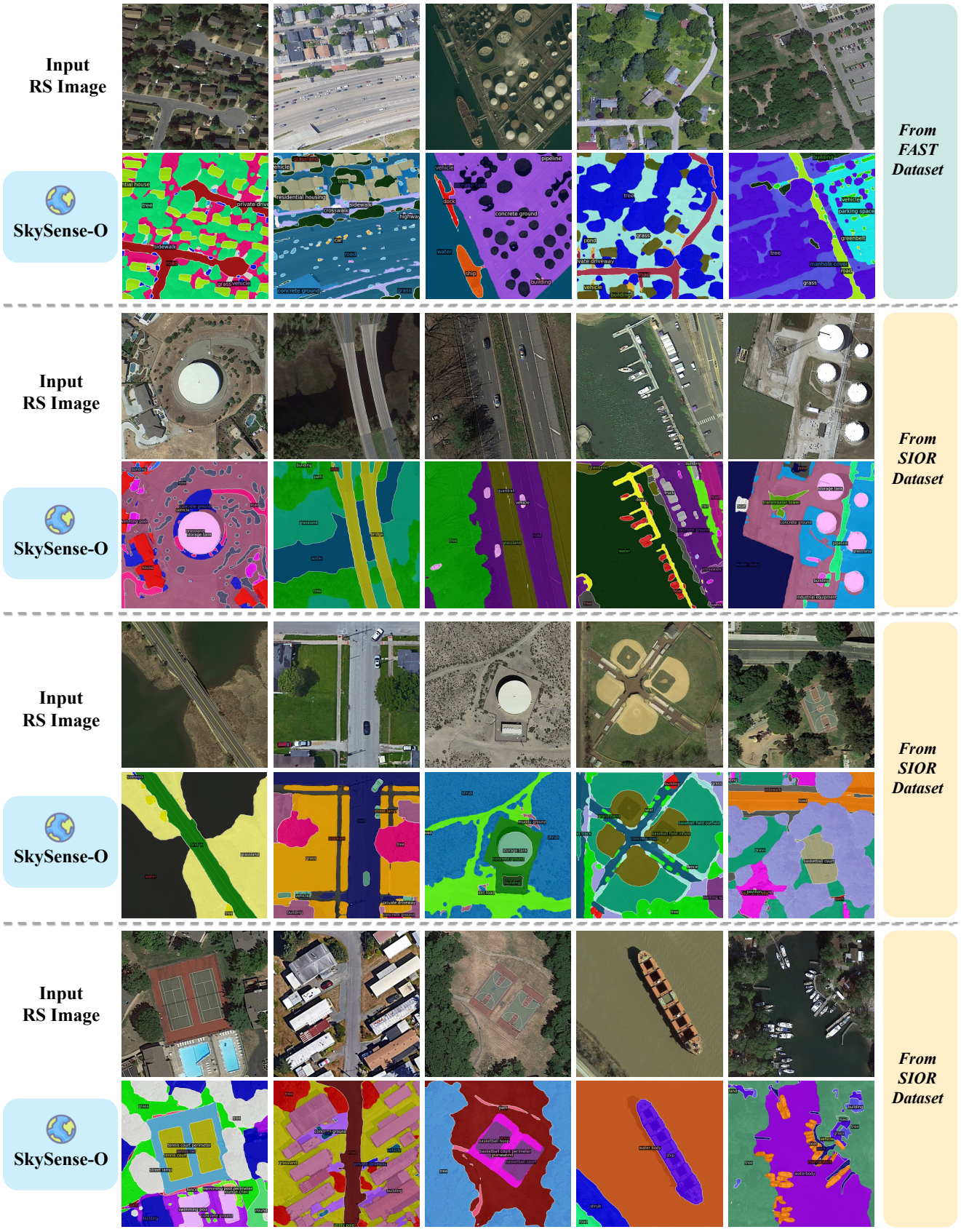

Figure 11. Visualizations of open-world interpretation results by SkySense-O on FAST and SIOR datasets.

Datcu, Marcello Pelillo, et al. Object detection in aerial images: A large-scale benchmark and challenges. *IEEE transactions on pattern analysis and machine intelligence*, 44(11):7778–7796, 2021. 8

[5] Xiaoyi Dong, Jianmin Bao, Yinglin Zheng, Ting Zhang, Dongdong Chen, Hao Yang, Ming Zeng, Weiming Zhang, Lu Yuan, Dong Chen, et al. Maskclip: Masked self-distillation advances contrastive language-image pretraining. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10995–11005, 2023. 9

[6] Feng Li, Qing Jiang, Hao Zhang, Tianhe Ren, Shilong Liu, Xueyan Zou, Huaizhe Xu, Hongyang Li, Chunyuan Li, Jianwei Yang, et al. Visual in-context prompting. *arXiv preprint arXiv:2311.13601*, 2023. 8

[7] Sylvain Lobry, Diego Marcos, Jesse Murray, and Devis Tuia. Rsvqa: Visual question answering for remote sensing data. *IEEE Transactions on Geoscience and Remote Sensing*, 58(12):8555–8566, 2020. 8, 9

[8] Junwei Luo, Zhen Pang, Yongjun Zhang, Tingzhu Wang, Linlin Wang, Bo Dang, Jiangwei Lao, Jian Wang, Jingdong Chen, Yihua Tan, et al. Skysensegpt: A fine-grained instruction tuning dataset and model for remote sensing vision-language understanding. *arXiv preprint arXiv:2406.10100*, 2024. 8

[9] Bo Qu, Xuelong Li, Dacheng Tao, and Xiaoqiang Lu. Deep semantic understanding of high resolution remote sensing image. In *2016 International conference on computer, information and telecommunication systems (Cits)*, pages 1–5. IEEE, 2016. 8

[10] Tianhe Ren, Shilong Liu, Ailing Zeng, Jing Lin, Kunchang Li, He Cao, Jiayu Chen, Xinyu Huang, Yukang Chen, Feng Yan, Zhaoyang Zeng, Hao Zhang, Feng Li, Jie Yang, Hongyang Li, Qing Jiang, and Lei Zhang. Grounded sam: Assembling open-world models for diverse visual tasks, 2024. 1

[11] Xian Sun, Peijin Wang, Zhiyuan Yan, Feng Xu, Ruiping Wang, Wenhui Diao, Jin Chen, Jihao Li, Yingchao Feng, Tao Xu, et al. Fair1m: A benchmark dataset for fine-grained object recognition in high-resolution remote sensing imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, 184:116–130, 2022. 8

[12] Di Wang, Jing Zhang, Bo Du, Minqiang Xu, Lin Liu, Dacheng Tao, and Liangpei Zhang. Samrs: Scaling-up remote sensing segmentation dataset with segment anything model, NeurIPS 2023. 1

[13] Junjue Wang, Zhuo Zheng, Zihang Chen, Ailong Ma, and Yanfei Zhong. Earthvqa: Towards queryable earth via relational reasoning-based remote sensing visual question answering. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 5481–5489, 2024. 8

[14] Yuan Wang, Rui Sun, Naisong Luo, Yuwen Pan, and Tianzhu Zhang. Image-to-image matching via foundation models: A new perspective for open-vocabulary semantic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3952–3963, 2024. 8

[15] Gui-Song Xia, Jingwen Hu, Fan Hu, Baoguang Shi, Xiang Bai, Yanfei Zhong, Liangpei Zhang, and Xiaoqiang Lu. Aid: A benchmark data set for performance evaluation of aerial scene classification. *IEEE Transactions on Geoscience and Remote Sensing*, 55(7):3965–3981, 2017. 9

[16] Jiarui Xu, Sifei Liu, Arash Vahdat, Wonmin Byeon, Xiaolong Wang, and Shalini De Mello. Open-vocabulary panoptic segmentation with text-to-image diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2955–2966, 2023. 8

[17] Mengde Xu, Zheng Zhang, Fangyun Wei, Han Hu, and Xiang Bai. Side adapter network for open-vocabulary semantic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2945–2954, 2023. 6, 7

[18] Zhiqiang Yuan, Wenkai Zhang, Kun Fu, Xuan Li, Chubo Deng, Hongqi Wang, and Xian Sun. Exploring a fine-grained multiscale method for cross-modal remote sensing image retrieval. *arXiv preprint arXiv:2204.09868*, 2022. 8

[19] Jian Liu Dongqi Tang Xinjie Luo Chi Qin Lei Zhang Yuqian Yuan, Wentong Li and Jianke Zhu. Osprey: Pixel understanding with visual instruction tuning, 2023. 1

[20] Qiqi Zhu, Yanfei Zhong, Bei Zhao, Gui-Song Xia, and Liangpei Zhang. Bag-of-visual-words scene classifier with local and global features for high spatial resolution remote sensing imagery. *IEEE Geoscience and Remote Sensing Letters*, 13(6):747–751, 2016. 9