RSSOD-BENCH: A LARGE-SCALE BENCHMARK DATASET FOR SALIENT OBJECT DETECTION IN OPTICAL REMOTE SENSING IMAGERY

Zhitong Xiong[1], *Yanfeng Liu*[2,3], *Qi Wang*[3], *Xiao Xiang Zhu*[1]

[1] Data Science in Earth Observation, Technical University of Munich (TUM), Ottobrunn, Germany
[2] School of Computer Science, and [3] School of Artificial Intelligence, Optics and Electronics (iOPEN), Northwestern Polytechnical University, Xi'an 710072, Shaanxi, P. R. China

ABSTRACT

We present the RSSOD-Bench dataset for salient object detection (SOD) in optical remote sensing imagery. While SOD has achieved success in natural scene images with deep learning, research in SOD for remote sensing imagery (RSSOD) is still in its early stages. Existing RSSOD datasets have limitations in terms of scale, and scene categories, which make them misaligned with real-world applications. To address these shortcomings, we construct the RSSOD-Bench dataset, which contains images from four different cities in the USA [1]. The dataset provides annotations for various salient object categories, such as buildings, lakes, rivers, highways, bridges, aircraft, ships, athletic fields, and more. The salient objects in RSSOD-Bench exhibit large-scale variations, cluttered backgrounds, and different seasons. Unlike existing datasets, RSSOD-Bench offers uniform distribution across scene categories. We benchmark **23** different state-of-the-art approaches from both the computer vision and remote sensing communities. Experimental results demonstrate that more research efforts are required for the RSSOD task.

Index Terms— benchmark, dataset, remote sensing, salient object detection

1. INTRODUCTION

Automatically extracting salient objects from images can serve as an important pre-processing step for numerous computer vision and remote sensing tasks [1–4]. To name a few, salient object detection (SOD) has been applied in self-supervised learning [5], image quality assessment [6], image retrieval [7], etc. SOD aims to extract visually distinctive objects from diverse complicated backgrounds at a pixel level. Namely, given an input image, two steps are required for SOD models to successfully detect the salient objects: 1) determine correct salient areas from cluttered backgrounds; 2) accurately segment the pixels of salient objects.

For natural scene images, SOD has achieved remarkable success with the advent of deep learning. However, research on SOD for remote sensing and Earth observation data (RSSOD) is still in its infancy. Natural scene images are usually object-centric. In contrast, remote sensing imagery, captured through high-altitude shooting, usually covers a larger range of scenes with diverse ground objects and complicated backgrounds. Considering these differences, several datasets dedicated to remote sensing data have been released to foster the research of novel methods. The ORSSD dataset [8] contains 800 images (600 for training and 200 for testing) collected from Google Earth. EORSSD [9] is an extended version of ORSSD with 2,000 images (1400 for training, 600 for testing) in total. ORSI-4199 [10] is a large dataset, which contains 4,199 images with pixel-level annotations.

Table 1. Statistics comparison with existing RSSOD datasets.

Datasets	#Total Images	#Cities	#Training	#Validation	#Test
ORSSD	800	—	600	0	200
EORSSD	2,000	—	1,400	0	600
ORSI-4199	4,199	—	2,000	0	2,199
Ours	**6,000**	**4**	**3,000**	**600**	**2,400**

Several shortcomings of existing RSSOD datasets hinder further research and progress of the SOD. The first limitation is that the scale of existing datasets is relatively small. As presented in Table 1, we list the statistical information of different existing datasets. The number of images in ORSSD and EORSSD datasets is less than 2,000, which is not enough to align the performance of models well with real-world scenarios. In contrast, our dataset contains 6,000 images collected in four different cities, which is larger than the existing ones.

The second limitation is that the images are limited to some specific scene categories. The remote sensing images of existing datasets are usually collected in several scene categories and not uniformly sampled from the Earth's surface. This makes the data distribution does not align well with real-world applications. Considering this problem, we introduce RSSOD-Bench to facilitate the research in the community.

2. DATASET CONSTRUCTION

For the RSSOD-Bench dataset, we annotate the salient objects that are naturally distinct from the background and are

[1]The RSSOD-Bench dataset can be accessed via https://github.com/EarthNets/Dataset4EO

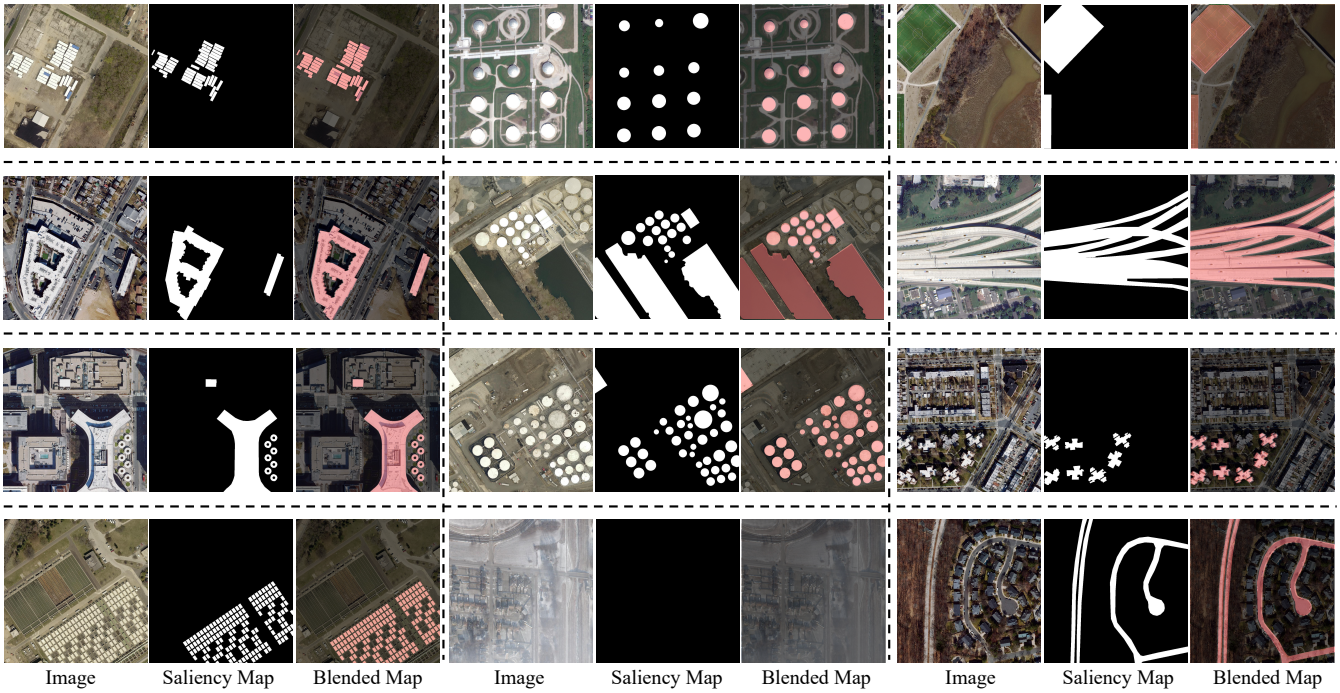

Fig. 1. Illustration of some visualization examples in the RSSOD-Bench dataset. Several challenging examples are presented, including tiny, large, dense, incomplete, and irregular salient objects.

associated with certain object categories useful for specific tasks. Specifically, the following objects are annotated: buildings, lakes, rivers, highways, bridges, aircraft, ships, athletic fields, badminton courts, volleyball courts, baseball fields, basketball courts, gymnasiums, storage tanks, etc. As presented in Fig. 1, the salient objects in RSSOD-Bench have large scale variations with tiny and large regions. Also, there are severely cluttered backgrounds with different seasons. In some scenes, the salient objects can be very dense. Note that, the remote sensing images in the RSSOD-Bench dataset are uniformly distributed regarding scene categories. This is different from existing ones that are usually collected from several scene categories.

3. METHODS AND RESULTS

We report four commonly-used metrics in the field of SOD, including the **MAE** [30], **F-Measure** [12], **S-Measure** [31], and **E-Measure** [32]. **MAE** quantifies the pixel-level disparity between the predicted saliency map (SM) and the ground truth (GT). **F-Measure** (F_β) is a composite metric that combines precision and recall to assess the similarity between SM and GT, with different weights. **S-Measure** (S_m) utilizes balanced structural information from object-aware and region-aware levels to evaluate the structural likeness between SM and GT. **E-Measure** (E_m) is a metric that signifies the degree of pixel-level correspondence and image-level statistics.

To comprehensively validate existing state-of-the-art methods on our proposed dataset, we conduct extensive experiments to benchmark and compare their performance. Specifically, LC [11] and FT [12] are classical saliency detection methods with no use of deep learning models. As there are considerable SOD methods proposed in the computer vision (CV) community, we choose ten typical deep learning-based methods, including DSS [13], NLDF [14], RAS [15], PoolNet [16], and so forth. Compared with classical methods, deep learning methods from the CV community can obtain clearly better results. For example, GateNet [21] and PFSNet [22] can achieve a S_m of over 0.82. Furthermore, 11 RSSOD methods are compared on our dataset, including the SARNet [23], FSMINet [24], MCCNet [26], and so on. The methods from the remote sensing community achieve state-of-the-art performance.

As presented in Table 2, we use different colors (i.e., red, green, and blue) to highlight the best, second-best, and third-best quantitative results. Typically, several latest algorithms, MJRBM-R [10], EMFI-R [28], HFANet-R [1], ACCo-V [29], and ACCo-R [29] works well on the proposed RSSOD-Bench dataset. This is in line with the expectations of these methods, and also reflects the feasibility of the proposed dataset.

However, as visualized in Fig. 2, the segmentation results of hard examples are still not satisfactory on the proposed dataset. This indicates that more research efforts are required to enhance the SOD performance for real-world applications.

Table 2. Comparison study with state-of-the-art methods on the proposed RSSOD-Bench dataset.

Methods	Publications	MAE ↓	F_β ↑	S_m ↑	E_m ↑
LC [11]	MM'06	0.1834	0.2467	0.5258	0.4826
FT [12]	CVPR'09	0.1588	0.2586	0.5417	0.4859
DSS [13]	CVPR'17	0.0622	0.6761	0.7627	0.7022
NLDF [14]	CVPR'17	0.0535	0.6765	0.7798	0.7632
RAS [15]	ECCV'18	0.0476	0.7158	0.7923	0.7524
PoolNet [16]	CVPR'19	0.0609	0.6915	0.7684	0.7092
PFAN [17]	CVPR'19	0.0430	0.7207	0.8046	0.7684
SCRN [18]	ICCV'19	0.0415	0.7435	0.8117	0.7582
F3Net [19]	AAAI'20	0.0456	0.7089	0.8043	0.7679
MINet [20]	CVPR'20	0.0396	0.7121	0.8008	0.8049
GateNet [21]	ECCV'20	0.0385	0.7369	0.8224	0.7877
PFSNet [22]	AAAI'21	0.0409	0.7457	0.8271	0.8027
SARNet [23]	RS'21	0.0402	0.7298	0.8242	0.8165
DAFNet [9]	TIP'21	0.0541	0.7006	0.7939	0.7389
FSMINet [24]	GRSL'22	0.0406	0.7388	0.8186	0.8113
MSCNet [25]	ICPR'22	0.0444	0.7486	0.8102	0.8016
MCCNet [26]	TGRS'22	0.0417	0.7525	0.8287	0.8122
CorrNet [27]	TGRS'22	0.0449	0.7415	0.7914	0.7503
MJRBM-R [10]	TGRS'22	0.0378	0.7531	0.8313	0.7899
EMFI-R [28]	TGRS'22	0.0377	0.7765	0.8400	0.8244
HFANet-R [1]	TGRS'22	0.0393	0.7635	0.8277	0.8020
ACCo-V [29]	TCYB'23	0.0367	0.7630	0.8406	0.8225
ACCo-R [29]	TCYB'23	0.0385	0.7583	0.8347	0.8226

4. CONCLUSION

We introduce the RSSOD-Bench dataset for salient object detection (SOD) in remote sensing imagery. Addressing the limitations of existing RSSOD datasets, RSSOD-Bench comprises carefully chosen images from four US cities, exhibiting diverse salient objects, varied backgrounds, and seasonal variations. Unlike previous datasets, RSSOD-Bench ensures uniform scene category distribution. We evaluate **23** state-of-the-art approaches from computer vision and remote sensing communities. While these methods perform well on RSSOD-Bench, there is still room for improving SOD accuracy compared to existing datasets. Therefore, further research efforts and advanced models are needed to enhance performance.

5. REFERENCES

[1] Q. Wang, Y. Liu, Z. Xiong, and Y. Yuan, "Hybrid feature aligned network for salient object detection in optical remote sensing imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–15, 2022.

[2] Y. Liu, Z. Xiong, Y. Yuan, and Q. Wang, "Distilling knowledge from super-resolution for efficient remote sensing salient object detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, pp. 1–16, 2023.

[3] Z. Xiong, F. Zhang, Y. Wang, Y. Shi, and X. X. Zhu, "Earthnets: Empowering AI in Earth Observation," *arXiv preprint arXiv:2210.04936*, 2022.

[4] Y. Liu, Q. Li, Y. Yuan, Q. Du, and Q. Wang, "Abnet: Adaptive balanced network for multiscale object detection in remote sensing imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–14, 2022.

[5] W. Van Gansbeke, S. Vandenhende, S. Georgoulis, and L. Van Gool, "Unsupervised semantic segmentation by contrasting object mask proposals," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, 2021, pp. 10 052–10 062.

[6] K. Gu, S. Wang, H. Yang, W. Lin, G. Zhai, X. Yang, and W. Zhang, "Saliency-guided quality assessment of screen content images," *IEEE Trans. Multimedia*, vol. 18, no. 6, pp. 1098–1110, 2016.

[7] A. Babenko and V. Lempitsky, "Aggregating local deep features for image retrieval," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, 2015, pp. 1269–1277.

[8] C. Li, R. Cong, J. Hou, S. Zhang, Y. Qian, and S. Kwong, "Nested Network With Two-Stream Pyramid for Salient Object Detection in Optical Remote Sensing Images," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 11, pp. 9156–9166, 2019.

[9] Q. Zhang, R. Cong, C. Li, M.-M. Cheng, Y. Fang, X. Cao, Y. Zhao, and S. Kwong, "Dense Attention Fluid Network for Salient Object Detection in Optical Remote Sensing Images," *IEEE Trans. Image Process.*, vol. 30, pp. 1305–1317, 2021.

[10] Z. Tu, C. Wang, C. Li, M. Fan, H. Zhao, and B. Luo, "ORSI salient object detection via multiscale joint region and boundary model," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–13, 2022.

[11] Y. Zhai and M. Shah, "Visual attention detection in video sequences using spatiotemporal cues," in *Proc. 14th ACM Int. Conf. Multimedia*, Oct. 2006, pp. 815–824.

[12] R. Achanta, S. Hemami, F. Estrada, and S. Susstrunk, "Frequency-tuned salient region detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2009, pp. 1597–1604.

[13] Q. Hou, M.-M. Cheng, X. Hu, A. Borji, Z. Tu, and P. Torr, "Deeply Supervised Salient Object Detection with Short Connections," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 4, pp. 815–828, 2019.

[14] Z. Luo, A. Mishra, A. Achkar, J. Eichel, S. Li, and P.-M. Jodoin, "Non-local Deep Features for Salient Object Detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2017, pp. 6593–6601.

[15] S. Chen, X. Tan, B. Wang, and X. Hu, "Reverse Attention for Salient Object Detection," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Jul. 2018, pp. 234–250.

[16] J.-J. Liu, Q. Hou, M.-M. Cheng, J. Feng, and J. Jiang, "A Simple Pooling-Based Design for Real-Time Salient Object Detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2019, pp. 3912–3921.

[17] T. Zhao and X. Wu, "Pyramid Feature Attention Network for Saliency Detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2019, pp. 3080–3089.

[18] Z. Wu, L. Su, and Q. Huang, "Stacked Cross Refinement Network for Edge-Aware Salient Object Detection," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, 2019, pp. 7263–7272.

[19] J. Wei, S. Wang, and Q. Huang, "F³Net: Fusion, Feedback and Focus for Salient Object Detection," *Proc. AAAI Conf. Artif. Intell. (AAAI)*, vol. 34, no. 07, pp. 12 321–12 328, 2020.

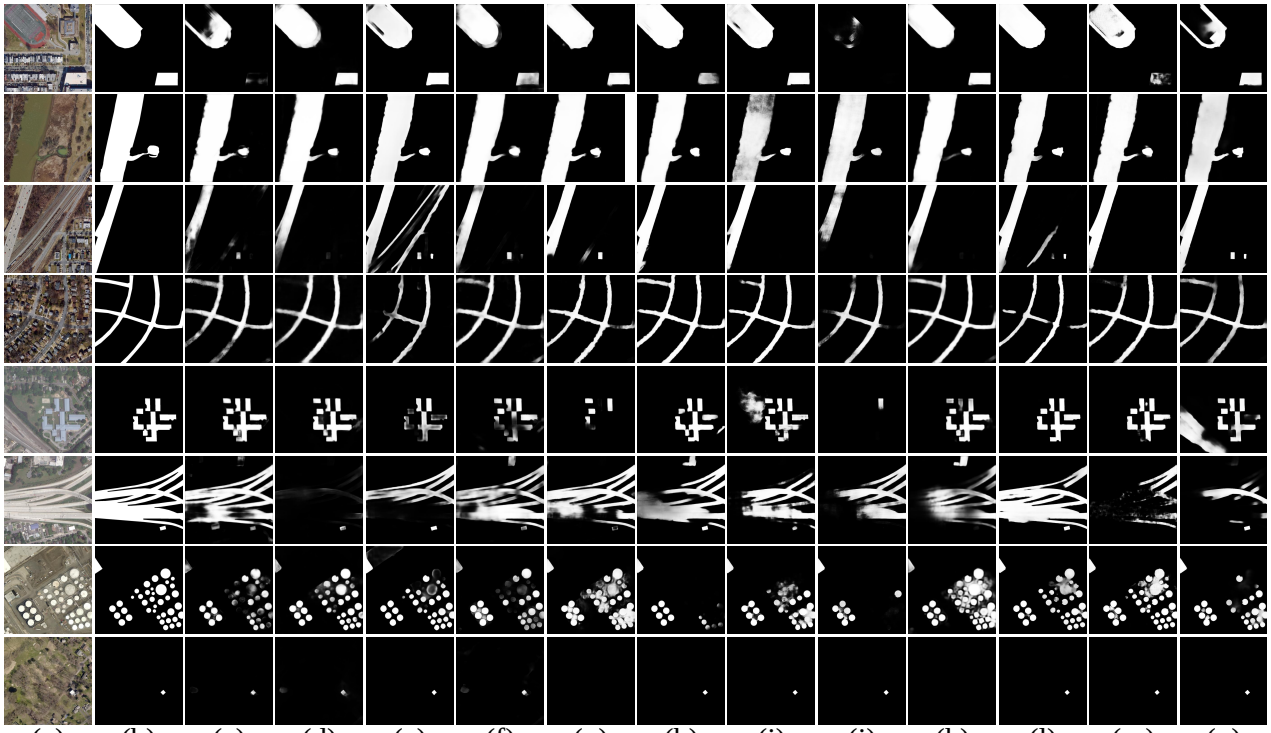

Fig. 2. Visualization saliency maps with 12 state-of-the-art methods on the proposed dataset, including five NSI-SOD approaches, and seven RSI-SOD algorithms, on different patterns. (a) Optical RSIs. (b) GT. (c) PFAN. (d) SCRN. (e) F3Net. (f) GateNet. (g) PFSNet. (h) FSMINet. (i) MCCNet. (j) CorrNet. (k) MJRBM-R. (l) EMFINet-R. (m) HFANet-R. (n) ACCo-V.

[20] Y. Pang, X. Zhao, L. Zhang, and H. Lu, "Multi-Scale Interactive Network for Salient Object Detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2020, pp. 9410–9419.

[21] X. Zhao, Y. Pang, L. Zhang, H. Lu, and L. Zhang, "Suppress and balance: A simple gated network for salient object detection," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2020, pp. 35–51.

[22] M. Ma, C. Xia, and J. Li, "Pyramidal Feature Shrinking for Salient Object Detection," *Proc. AAAI Conf. Artif. Intell. (AAAI)*, vol. 35, no. 03, pp. 2311–2318, 2021.

[23] Z. Huang, H. Chen, B. Liu, and Z. Wang, "Semantic-Guided Attention Refinement Network for Salient Object Detection in Optical Remote Sensing Images," *Remote Sens.*, vol. 13, no. 11, 2021.

[24] K. Shen, X. Zhou, B. Wan, R. Shi, and J. Zhang, "Fully squeezed multiscale inference network for fast and accurate saliency detection in optical remote-sensing images," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2022.

[25] Y. Lin, H. Sun, N. Liu, Y. Bian, J. Cen, and H. Zhou, "A lightweight multi-scale context network for salient object detection in optical remote sensing images," in *Proc. Int. Conf. Pattern Recognit. (ICPR)*, 2022.

[26] G. Li, Z. Liu, W. Lin, and H. Ling, "Multi-Content Complementation Network for Salient Object Detection in Optical Remote Sensing Images," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–13, 2022.

[27] G. Li, Z. Liu, Z. Bai, W. Lin, and H. Ling, "Lightweight salient object detection in optical remote sensing images via feature correlation," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–12, 2022.

[28] X. Zhou, K. Shen, Z. Liu, C. Gong, J. Zhang, and C. Yan, "Edge-Aware Multiscale Feature Integration Network for Salient Object Detection in Optical Remote Sensing Images," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–15, 2022.

[29] G. Li, Z. Liu, D. Zeng, W. Lin, and H. Ling, "Adjacent context coordination network for salient object detection in optical remote sensing images," *IEEE Trans. on Cybern.*, vol. 53, no. 1, pp. 526–538, 2023.

[30] F. Perazzi, P. Krähenbühl, Y. Pritch, and A. Hornung, "Saliency filters: Contrast based filtering for salient region detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2012, pp. 733–740.

[31] D.-P. Fan, M.-M. Cheng, Y. Liu, T. Li, and A. Borji, "Structure-measure: A new way to evaluate foreground maps," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, 2017, pp. 4558–4567.

[32] D.-P. Fan, C. Gong, Y. Cao, B. Ren, M.-M. Cheng, and A. Borji, "Enhanced-Alignment Measure for Binary Foreground Map Evaluation," in *Proc. Int. Joint Conf. Artif. Intell. (IJCAI)*, Jul. 2018, pp. 698–704.