

Learning Synergistic Attention for Light Field Salient Object Detection (Supplementary Material)

Yi Zhang¹

yi.zhang1@insa-rennes.fr

Geng Chen^{*2}

geng.chen.cs@gmail.com

Qian Chen³

poly@mail.ustc.edu.cn

YuJia Sun⁴

thograce@163.com

Yong Xia²

yxia@nwpu.edu.cn

Olivier Deforges¹

odeforge@insa-rennes.fr

Wassim Hamidouche¹

Wassim.Hamidouche@insa-rennes.fr

Lu Zhang¹

lu.ge@insa-rennes.fr

¹ Univ Rennes, INSA Rennes, CNRS, IETR (UMR 6164), France

² National Engineering Laboratory for Integrated Aero-Space-Ground-Ocean Big Data Application Technology, School of Computer Science and Engineering, Northwestern Polytechnical University, China

³ School of Information Science and Technology, University of Science and Technology of China, China

⁴ School of Computer Science, Inner Mongolia University, China

Abstract

This document provides additional information regarding the evaluation metrics and experiments. The four evaluation metrics adopted in the experiments are detailed in this document. Besides, we show more qualitative results to further demonstrate the effectiveness of the proposed *SA-Net*.

1 Metrics

In this work, we evaluate all 28 benchmark models and our *SA-Net* with four widely used SOD metrics with respect to the ground-truth binary mask and predicted saliency map. The F-Measure (F_β) [1] and mean absolute error (MAE) [2] focus on the local (per-pixel) match between ground truth and prediction, while S-Measure (S_α) [3] pays attention to the object structure similarities. Besides, E-Measure (E_ϕ) [4] considers both the local and global information.

* Corresponding Author: Geng Chen (geng.chen.cs@gmail.com)

© 2021. The copyright of this document resides with its authors.

It may be distributed unchanged freely in print or electronic forms.

- **MAE** computes the mean absolute error between the ground truth $G \in \{0, 1\}$ and a normalized predicted saliency map $P \in [0, 1]$, i.e.,

$$MAE = \frac{1}{W \times H} \sum_{i=1}^W \sum_{j=1}^H |G(i, j) - P(i, j)|, \quad (1)$$

where H and W denotes height and width, respectively.

- **F-Measure** gives a single value (F_β) which considers both the *Precision* and *Recall*, thus being defined as:

$$F_\beta = \frac{(1 + \beta^2) Precision \times Recall}{\beta^2 Precision + Recall}, \quad (2)$$

with

$$Precision = \frac{|M \cap G|}{|M|}, Recall = \frac{|M \cap G|}{|G|}, \quad (3)$$

where M denotes a binary mask converted from a predicted saliency map and G is the ground truth. Multiple M are computed by taking different thresholds of $[0, 255]$ on the saliency map. Note that the β^2 is set to 0.3 according to [10]. Notably, the adaptive F-Measure-based results reported in our manuscript are calculated by applying an adaptive threshold algorithm [10].

- **S-Measure** evaluates the structure similarities between salient objects in ground-truth foreground maps and predicted saliency maps:

$$S = \alpha \times S_o + (1 - \alpha) \times S_r. \quad (4)$$

where S_o and S_r denote the object-/region-based structure similarities, respectively. $\alpha \in [0, 1]$ is set as 0.5 so that equal weights are assigned to both the object-level and region-level assessments [9].

- **E-Measure** is a cognitive vision-inspired metric to evaluate both the local and global similarities between two binary maps. Specifically, it is defined as:

$$E_\phi = \frac{1}{W \times H} \sum_{x=1}^W \sum_{y=1}^H \phi(P(x, y), G(x, y)), \quad (5)$$

where ϕ represents the enhanced alignment matrix [9]. Similar to F_β , adaptive E-Measure is adopted for the evaluation in our manuscript.

2 Qualitative Results

Comparison of Ablation Models. Due to the page limit, we only show partial visual results of ablation studies in our manuscript. To further illustrate the benefit of each key component in our *SA-Net*, we show complete qualitative results for all six ablation models in Figure 1. As can be observed, each component improves the quality of predicted saliency maps and contributes to the superior performance of *SA-Net*.

Comparison with State-of-the-Arts. To further demonstrate the effectiveness of our *SA-Net*, we show extensive visual results of our method as well as the competing models upon the three benchmark datasets (Figure 2 to 5). Overall, our proposed *SA-Net* depicts fine object structures and possesses less false positive and false negative, thus giving predictions closest to ground truths.

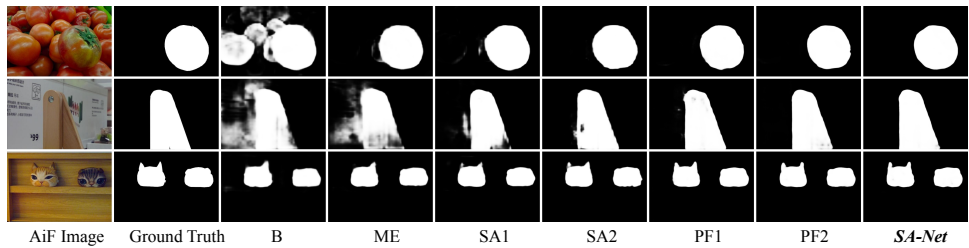


Figure 1: Visual results of ablation studies.

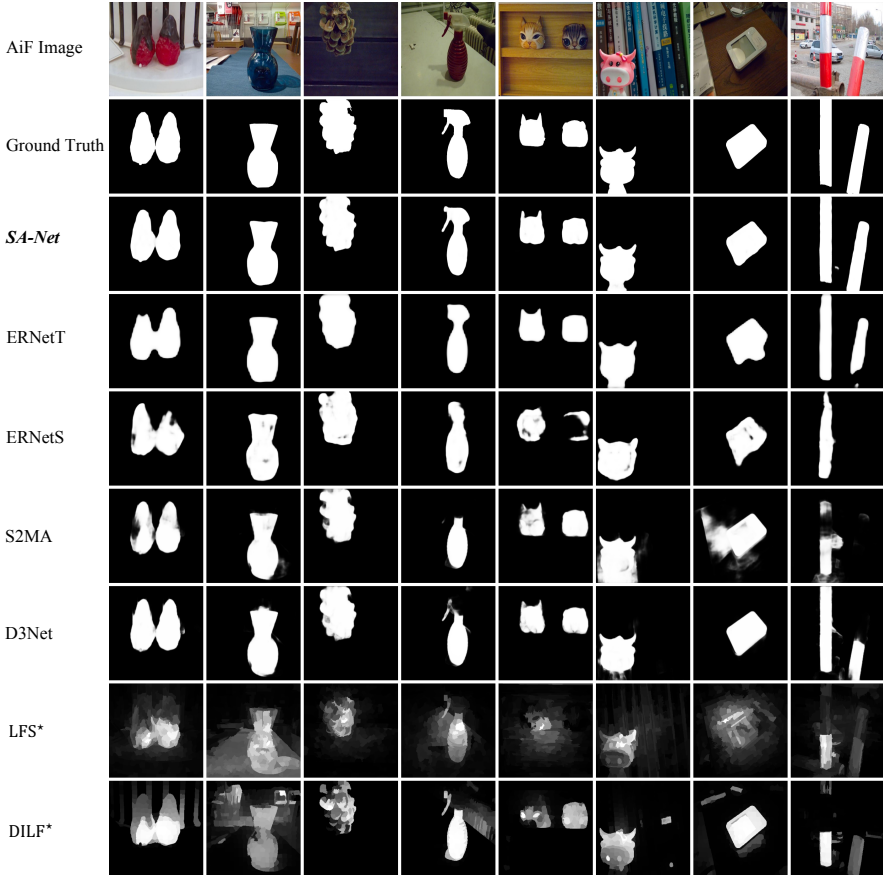


Figure 2: Visual comparison of our *SA-Net* and state-of-the-art SOD models upon DUT-LF [17]. ★ indicates tradition methods.

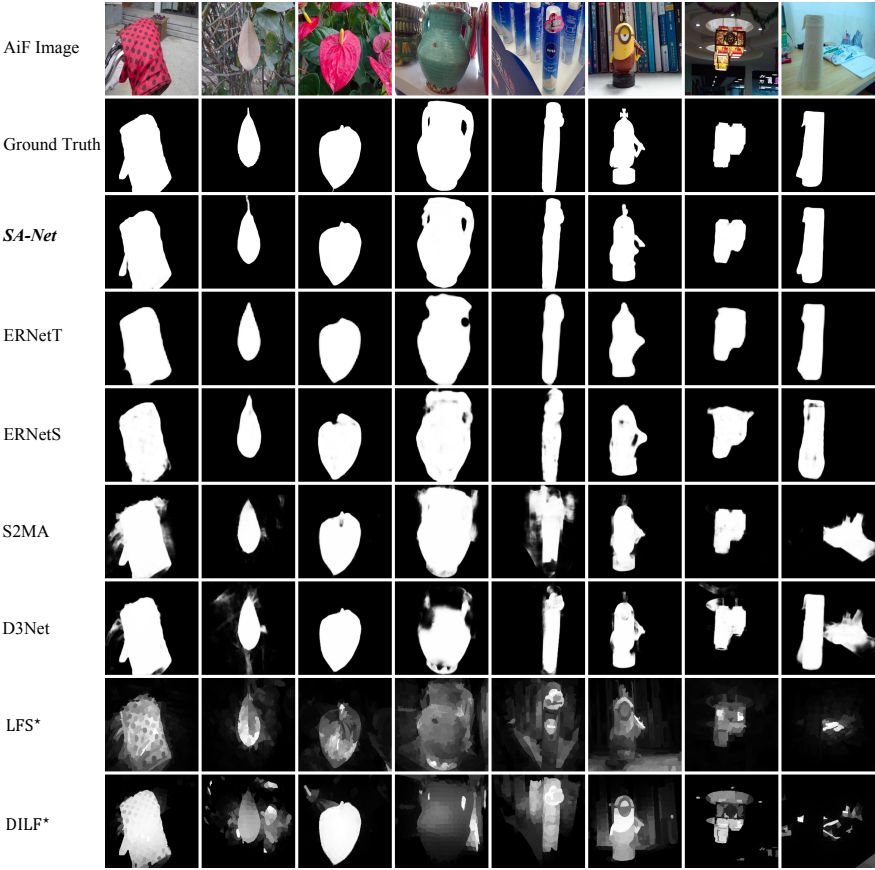


Figure 3: Visual comparison of our SA-Net and state-of-the-art SOD models upon DUT-LF [17]. ★ indicates tradition methods.

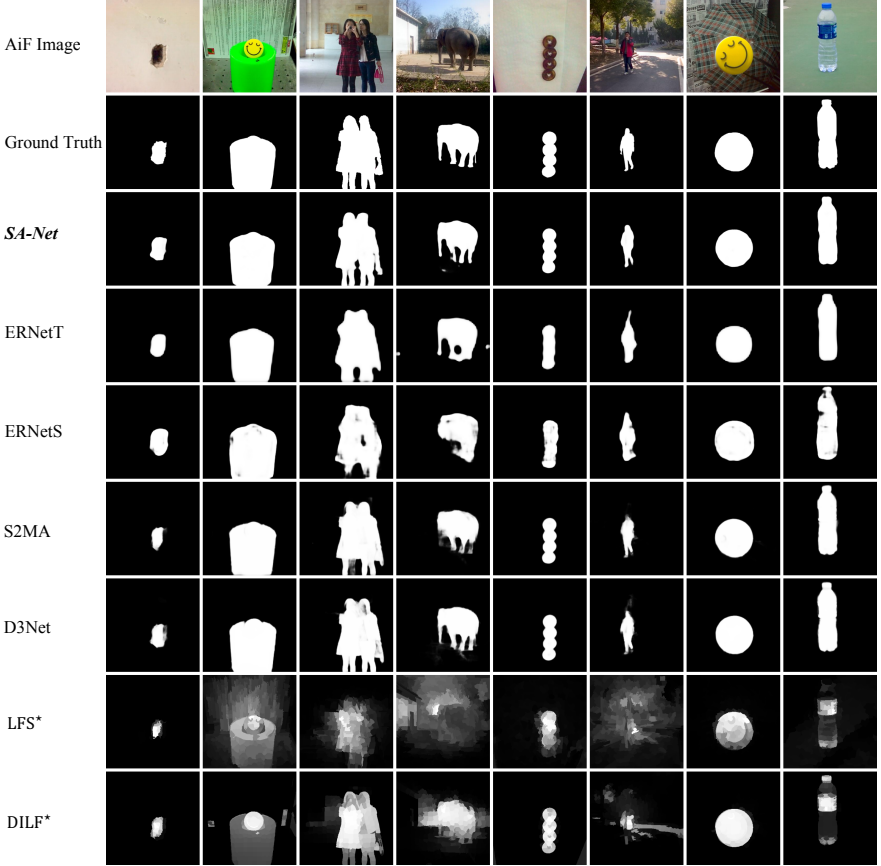


Figure 4: Visual comparison of our *SA-Net* and state-of-the-art SOD models upon HFUT [8]. ★ indicates tradition methods.

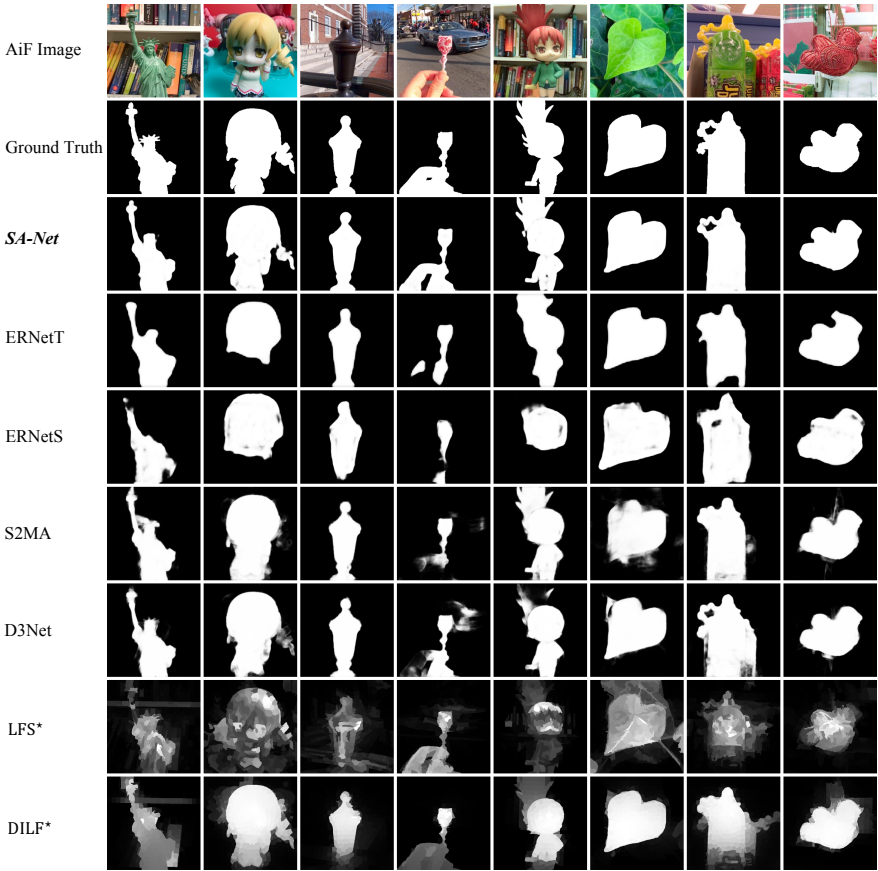


Figure 5: Visual comparison of our *SA-Net* and state-of-the-art SOD models upon LFSD [1].
★ indicates tradition methods.

References

- [1] Radhakrishna Achanta, Sheila Hemami, Francisco Estrada, and Sabine Susstrunk. Frequency-tuned salient region detection. In *CVPR*, 2009.
- [2] Ali Borji, Ming-Ming Cheng, Huaizu Jiang, and Jia Li. Salient object detection: A benchmark. *TIP*, 2015.
- [3] Deng-Ping Fan, Ming-Ming Cheng, Yun Liu, Tao Li, and Ali Borji. Structure-measure: A new way to evaluate foreground maps. In *ICCV*, 2017.
- [4] Deng-Ping Fan, Cheng Gong, Yang Cao, Bo Ren, Ming-Ming Cheng, and Ali Borji. Enhanced-alignment measure for binary foreground map evaluation. In *IJCAI*, 2018.
- [5] Nianyi Li, Jinwei Ye, Yu Ji, Haibin Ling, and Jingyi Yu. Saliency detection on light field. In *CVPR*, 2014.
- [6] Federico Perazzi, Philipp Krähenbühl, Yael Pritch, and Alexander Hornung. Saliency filters: Contrast based filtering for salient region detection. In *CVPR*, 2012.
- [7] Tiantian Wang, Yongri Piao, Huchuan Lu, Xue chun Li, and Li Zhang. Deep learning for light field saliency detection. In *ICCV*, 2019.
- [8] Jun Zhang, Meng Wang, Liang Lin, Xun Yang, Jun Gao, and Yong Rui. Saliency detection on light field: A multi-cue approach. *TOMM*, 2017.