Paper ID: 443, Presentation ID: Fr4P1T4.3

Title: DAANet: Dual Attention Aggregating Network for Salient Object Detection

The 2023 IEEE International Conference on Robotics and Biomimetics

IEEE ROBIO 2023 Koh Samui Thailand

Yijie Li[1], Hewei Wang[1], Zhenqi Li[1], Shaofan Wang[1], Soumyabrata Dev[2], Guoyu Zuo[1]

[1]Beijing University of Technology
[2]University College Dublin

Motivation

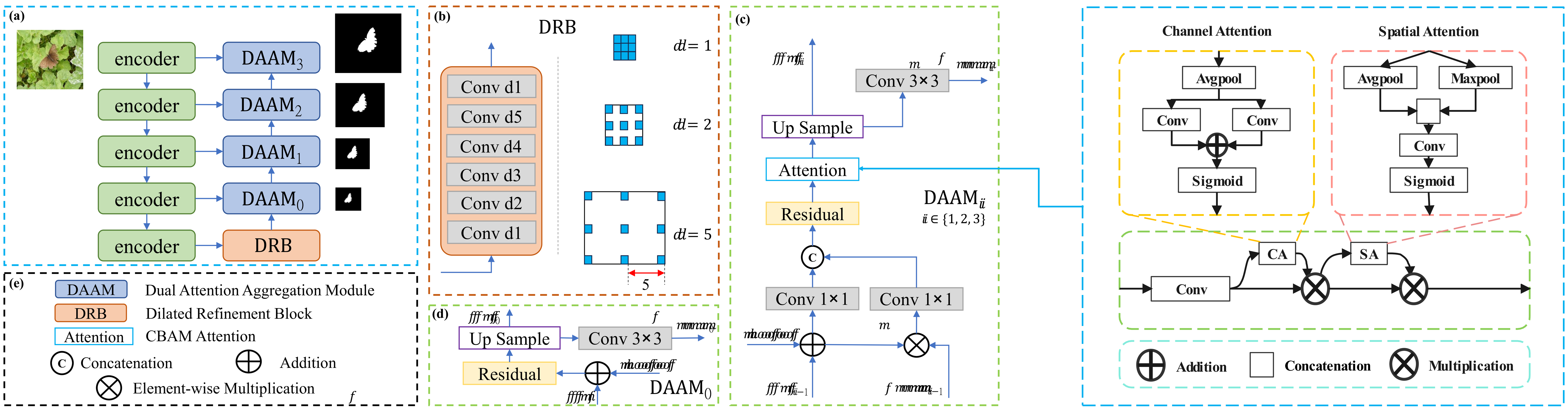
To address the two current issues:

- Several methods are estimated based on an encoder-decoder structure with a straightforward design in decoders, leading to a sub-optimal capacity of capturing the details in a specific salient region.
- Multiple existing salient object detection methods directly use the feature maps from the backbone encoder without further processing which results in the limitation of the receptive field.

Contributions

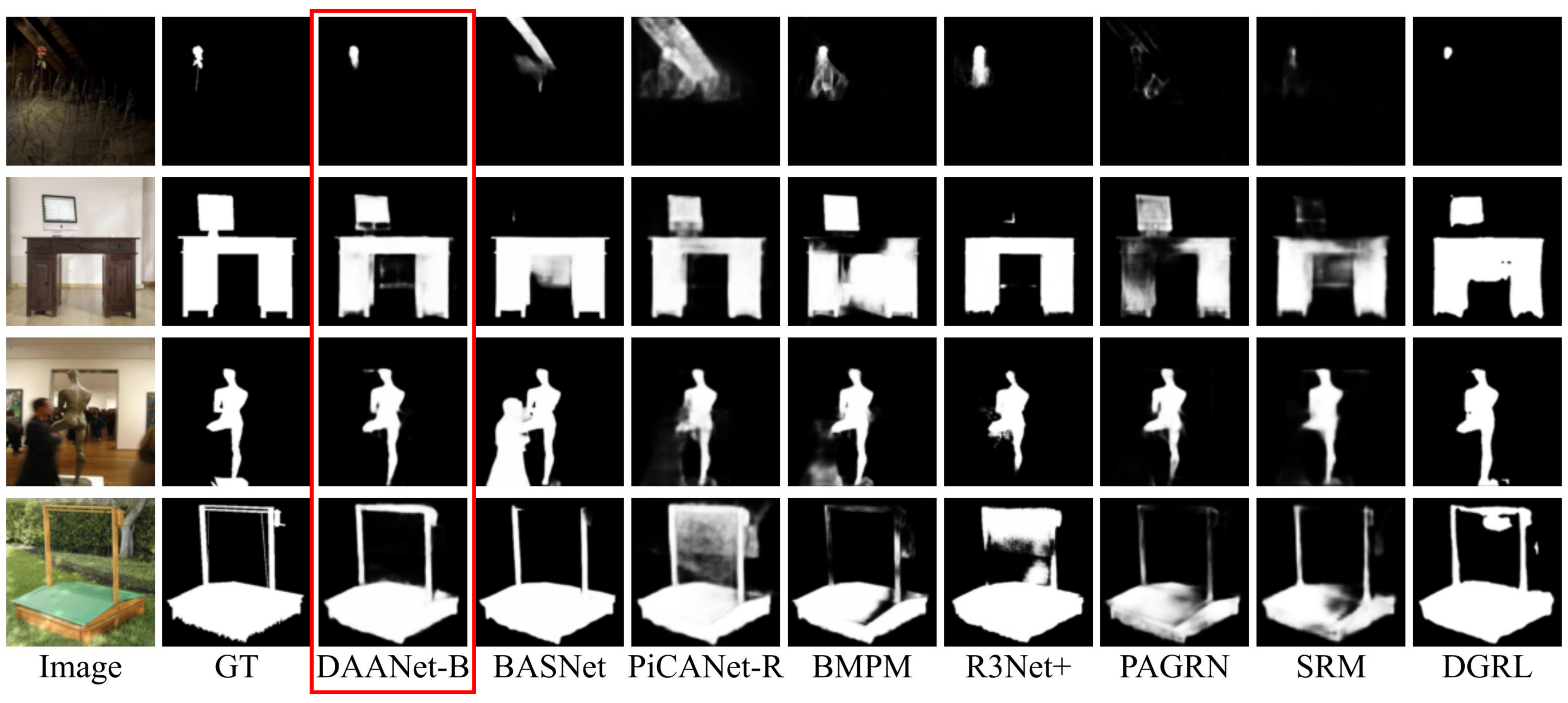
- We proposed a novel decoder module for salient object detection: a dual attention aggregation module (DAAM), which can achieve better performance than baseline models.
- We introduced the dilated refinement block (DRB) for salient object detection to expand the receptive field and refine the feature maps output by the backbone encoder.
- We designed and conducted a thorough evaluation and comparison with twelve methods on six benchmark datasets, and DAANet can achieve advanced performance with a light-weighted configuration.

Methodology



Results

Qualitative comparison with previous approaches



We use combined BCE and IOU as objective functions,

$$\mathcal{L}_{\text{bce}} = -\frac{1}{NM}\sum_{n=0}^{N}\sum_{i=0}^{M} t_i \log(p_i) + (1-t_i)\log(1-p_i)$$

$$\mathcal{L}_{\text{iou}} = 1 - \frac{1}{N}\sum_{n=0}^{N}\frac{\sum_{i=0}^{M} t_i p_i}{\sum_{i=0}^{M}(t_i + p_i - t_i p_i)}$$

when applying these objective functions to all output, the total loss can be formulated as,

$$\mathcal{L}_{\text{total}} = \sum_{i=1}^{4}\beta_i(\alpha_1\mathcal{L}_{\text{bce}} + \alpha_2\mathcal{L}_{\text{iou}})$$

Quantitative ablation study on DUTS-TE dataset

No.	Backbone	FPN configs			DUTS-TE	
		IOU	DAAM	DRB	mF_β	MAE
1	VGG16				0.749	0.053
2	VGG16	✓			0.773	0.049
3	VGG16	✓	✓		0.792	0.044
4	VGG16	✓	✓	✓	0.795	0.043
5	ResNet50	✓	✓		0.801	0.042
6	ResNet50	✓	✓	✓	**0.801**	**0.042**
7	MobileNetV2	✓	✓		0.767	0.051
8	MobileNetV2	✓	✓	✓	0.762	0.052

Quantitative comparison with previous methods on the DUTS-TE dataset

Methods	Size(MB)	Training Data		DUTS-TE [21]		
		Datasets	#Images	$maxF_\beta \uparrow$	$mF_\beta \uparrow$	$MAE \downarrow$
UCF [27]	**117.9**	MSRA10K	10,000	0.771	0.629	0.117
Amulet [28]	132.6	MSRA10K	10,000	0.778	0.676	0.085
DSS [29]	447.3	MSRA-B	2,500	0.826	0.791	0.057
PAGRN [30]	-	DUTS-TR	10,553	0.855	0.788	0.056
BMPM [15]	-	DUTS-TR	10,553	0.851	0.751	0.049
AFNet [31]	143.9	DUTS-TR	10,553	**0.862**	**0.797**	**0.046**
RAS [32]	**81.0**	MSRA-B	2,500	0.831	0.755	0.060
PiCANet [33]	153.3	DUTS-TR	10,553	0.851	0.755	0.054
DAANet-A	89.8	DUTS-TR	10,553	0.867	0.795	**0.043**
DGRL [34]	648.0	DUTS-TR	10,553	0.829	0.798	0.050
SRM [35]	213.1	DUTS-TR	10,553	0.827	0.757	0.059
PiCANet-R [33]	197.2	DUTS-TR	10,553	**0.860**	0.764	0.051
BASNet [36]	348.5	DUTS-TR	10,553	0.859	**0.796**	0.048
DAANet-B	229.0	DUTS-TR	10,553	**0.870**	**0.801**	**0.042**
DAANet-C	**15.8**	DUTS-TR	10,553	0.836	0.767	0.051

Conclusion

- The paper presents DAANet, an advanced salient object detection model, employing a dual attention aggregation module (DAAM) and dilated refinement block (DRB). This approach, leveraging an ImageNet pre-trained backbone and FPN architecture, significantly enhances feature capture and the accuracy of salient masks
- Future research will concentrate on enhancing DAANet's performance, exploring additional improvements in salient object detection, and adapting the model for efficient deployment in domain-specific applications, particularly on resource-constrained platforms such as mobile devices and embedded systems.