PoolNet+: Exploring the Potential of Pooling for Salient Object Detection

Supplementary Material

Jiang-Jiang Liu, Qibin Hou, Zhi-Ang Liu, Ming-Ming Cheng

1 Salient Object Detection

1.1 PR and F-measure Curves

We show the comparisons of PR and F-measure curves among seven state-of-the-art salient object detection methods, PoolNet-V+ and PoolNet-R+ on five datasets in Fig. 1. As can be seen, the PR and F-measure curves by PoolNet-R+ (solid red ones) are more outstanding compared to all other previous approaches. Even the PoolNet-V+ (solid cyan ones), which utilizes a less powerful backbone, still performs comparably to the other methods. Specifically, in the top row of Fig. 1, as the recall scores approach 1, our precision scores are much higher than other methods. This phenomenon reveals that the false positives in our saliency map are low, and the detected salient objects are more integral, which is essential for salient object detection approaches. This conclusion can also be drawn from the bottom row of Fig. 1, where the F-measure curves of our approach are more convex.

2 Generalization to Other Tasks

This section investigates the generalization ability of the proposed approach by applying it to two popular related low-level vision tasks, including RGB-D salient object detection and camouflaged object detection.

2.1 RGB-D Salient Object Detection

2.1.1 Related Work

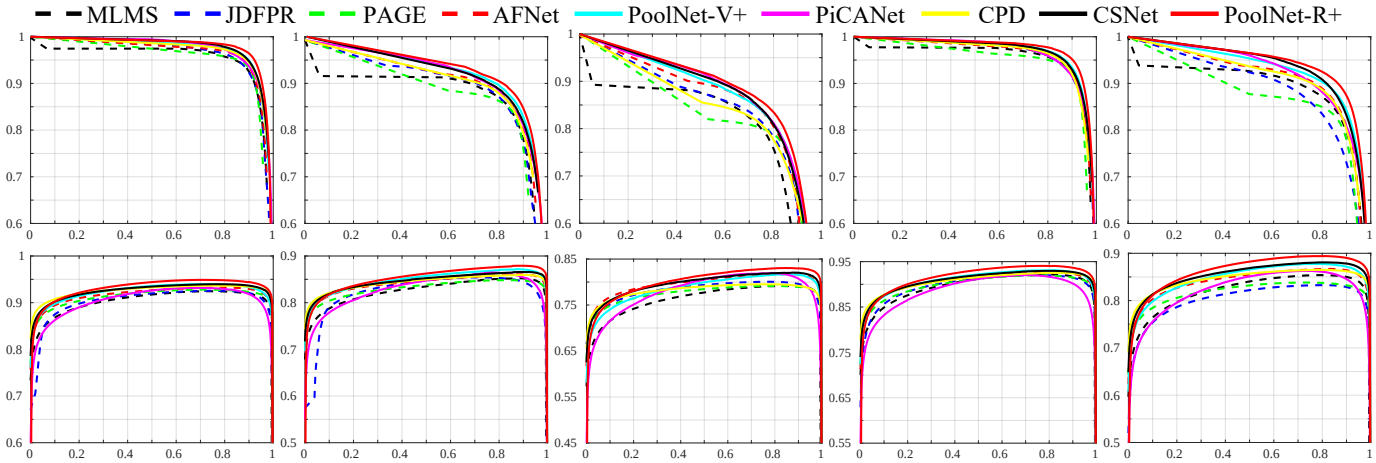
RGB-D salient object detection aims at predicting the most visually prominent objects or regions in an RGB-D image. Previous RGB-D salient object detection methods mainly relied on handcrafted saliency priors of the RGB and depth maps [1], [2]. Deep-learning-based methods rise in recent years with the flourish of CNNs. Early deep methods [3], [4] combined the manually designed contrast cues with the multi-scale features extracted by CNNs. Recent deep methods directly utilized the raw RGB-D images in an end-to-end scheme. [5] made predictions by combining side outputs of the backbone network. [6] progressively integrated RGB and depth features extracted by respective independent networks. To better utilize the depth information, [7] transferred the architecture designed for RGB images to fuse the depth representation. Based on the contrast prior, [8] introduced the idea of complementary information into the RGB feature extractor by fusing enhanced depth features. [9] leveraged cross-modal interactions to diversify the fusion ways to address the insufficient combination between RGB and depth features. [10] designed a three-stream attention-aware network to select complementary representations effectively. [11] used residual connections to extract and fuse multi-level paired complementary cues from RGB and depth streams. [12] integrated multiple attention strategies to propagate contexts accurately. [13] extended a simple general structure for cross-modal feature learning.

2.1.2 Implementation Details

To meet the input format of RGB-D salient object detection, which takes in an RGB image along with the corresponding depth image, we add a sequence of five 1×1 convolutional layers to extract the depth information based on PoolNet-R+ to build our RGB-D model. We follow [14], [15] and directly fuse features from the RGB and depth branches of the corresponding stages for simplicity. We use the Adam [16] optimizer with an initial learning rate of 1e-4 and a batch size of 10. Our network is trained for 200 epochs, and the learning rate is divided by 10 for every 60 epochs. We use the joint set, including 1485 samples from NJU [17] and 700 samples from NLPR [18], for model training as commonly done [6], [11], [15]. In addition to the rest images in NJU and NLPR, the STERE [19], DES [20], and SIP [21] sets are used for testing. In both training and testing phases, the input images are resized to 352×352. We employ the same four metrics as salient object detection for performance evaluation.

2.1.3 Comparisons to the State-of-the-Arts

The quantitative comparisons of our approach with ten previous state-of-the-art methods are shown in Table 1. We could see that our approach surpasses all the other methods on all five datasets in terms of all three metrics. Compared to the previous best method, D3Net [13], our approach outperforms it with average promotions of 2.1%, 1.5%, and 2.2% in F-measure, MAE, and S-measure, respectively, across all five datasets. It is noteworthy that the numbers of parameters and MAdds of our approach are the smallest in both terms. Additionally, we plot the PR and F-measure curves in Fig. 2. As can be seen, the curves obtained by our approach (red ones) are particularly more convex than the rest ones. Unlike most of the previous methods that usually utilize a heavy independent network playing the role of depth feature extraction, our approach requires only several convolutional layers, *i.e.*, 0.18M additional parameters and 0.12G additional MAdds in this case. We attribute this to the powerful feature extraction capability of the proposed PoolNet-R+, in which case the supplementary depth features

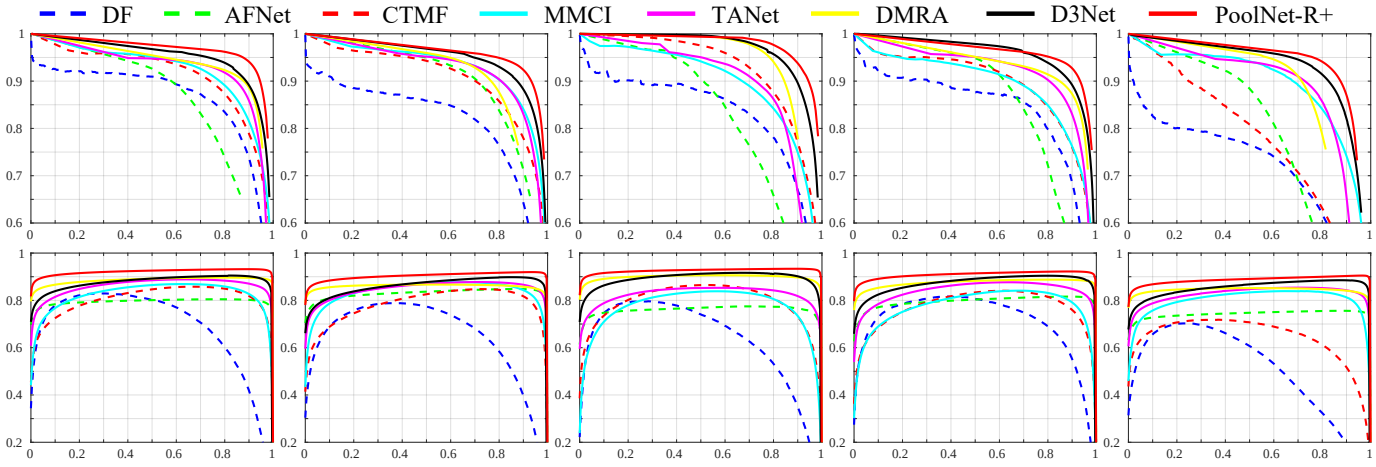


(a) ECSSD [22] (b) PASCAL-S [23] (c) DUT-OMRON [24] (d) HKU-IS [25] (e) DUTS-TE [26]

Fig. 1. PR and F-measure curves on five salient object detection datasets. Top: precision (vertical) recall (horizontal) curves. Bottom: F-measure (vertical) and threshold (horizontal) curves. We show results using two different backbone networks: VGG-16 and ResNet-50, which correspond to PoolNet-V+ and PoolNet-R+, respectively. Our curves are better than those of other state-of-the-art methods.

Method	Params (M)	MAdds (G)	NJU [17]			STERE [19]			DES [20]			NLPR [18]			SIP [21]		
			$F_\beta \uparrow$	MAE $\downarrow$	$S_m \uparrow$	$F_\beta \uparrow$	MAE $\downarrow$	$S_m \uparrow$	$F_\beta \uparrow$	MAE $\downarrow$	$S_m \uparrow$	$F_\beta \uparrow$	MAE $\downarrow$	$S_m \uparrow$	$F_\beta \uparrow$	MAE $\downarrow$	$S_m \uparrow$
DF_{17} [3]	-	-	0.829	0.141	0.763	0.788	0.141	0.757	0.796	0.093	0.752	0.817	0.085	0.802	0.703	0.185	0.653
$CTMF_{18}$ [7]	-	-	0.858	0.085	0.849	0.848	0.064	0.875	0.865	0.055	0.863	0.841	0.056	0.860	0.718	0.139	0.715
PCF_{18} [6]	133.40	-	0.888	0.059	0.877	-	-	-	-	-	-	0.888	0.044	0.874	-	-	-
$AFNet_{19}$ [27]	-	-	0.805	0.100	0.772	0.848	0.075	0.825	0.775	0.068	0.770	0.816	0.058	0.799	0.756	0.118	0.720
$MMCI_{19}$ [9]	-	-	0.869	0.079	0.858	0.877	0.068	0.873	0.838	0.065	0.848	0.841	0.059	0.856	0.839	0.082	0.813
$TANet_{19}$ [10]	232.45	-	0.889	0.060	0.878	0.878	0.060	0.871	0.853	0.046	0.858	0.877	0.041	0.886	0.854	0.075	0.835
$CPFP_{19}$ [8]	69.50	-	-	-	-	0.891	0.051	0.879	0.883	0.038	0.872	0.892	0.031	0.899	0.873	0.064	0.850
$DMRA_{19}$ [11]	59.66	-	0.896	0.051	0.886	0.867	0.066	0.835	0.906	0.035	0.883	0.888	0.031	0.899	0.852	0.085	0.806
$S2MA_{20}$ [12]	86.65	108.22	0.899	0.058	0.887	0.895	0.051	0.890	-	-	-	0.912	0.030	0.916	0.893	0.057	0.872
$D3Net_{20}$ [13]	45.23	55.17	0.905	0.051	0.893	0.898	0.054	0.889	0.917	0.033	0.898	0.904	0.033	0.905	0.885	0.063	0.864
PoolNet-R+	34.30	14.15	**0.932**	**0.033**	**0.921**	**0.920**	**0.035**	**0.912**	**0.934**	**0.019**	**0.924**	**0.922**	**0.023**	**0.921**	**0.905**	**0.049**	**0.881**

Table 1
Quantitative RGB-D salient object detection results on five widely used datasets. As can be seen, our approach achieves the best results on all five datasets in terms of F-measure, MAE, and S-measure.



(a) NJU [17] (b) STERE [23] (c) DES [20] (d) NLPR [18] (e) SIP [21]

Fig. 2. PR and F-measure curves on five RGB-D salient object detection datasets. Top: precision (vertical) recall (horizontal) curves. Bottom: F-measure (vertical) and threshold (horizontal) curves.

Method	Params (M)	MAdds (G)	CHAMELEON [28]					CAMO [29]					COD10K [30]				
			$F_\beta \uparrow$	$S_m \uparrow$	$E_\phi \uparrow$	$F_\beta^w \uparrow$	MAE $\downarrow$	$F_\beta \uparrow$	$S_m \uparrow$	$E_\phi \uparrow$	$F_\beta^w \uparrow$	MAE $\downarrow$	$F_\beta \uparrow$	$S_m \uparrow$	$E_\phi \uparrow$	$F_\beta^w \uparrow$	MAE $\downarrow$
FPN_{17} [31]	36.32	16.84	0.749	0.794	0.783	0.590	0.075	0.681	0.684	0.677	0.483	0.131	0.600	0.697	0.691	0.411	0.075
$MaskRCNN_{17}$ [32]	-	-	0.647	0.643	0.778	0.518	0.099	0.567	0.574	0.715	0.430	0.151	0.511	0.613	0.748	0.402	0.080
$PSPNet_{17}$ [33]	46.71	37.47	0.747	0.773	0.758	0.555	0.085	0.656	0.663	0.659	0.455	0.139	0.575	0.678	0.680	0.377	0.080
$UNet{+}{+}_{18}$ [34]	36.63	105.70	0.632	0.695	0.762	0.501	0.094	0.557	0.599	0.653	0.392	0.149	0.499	0.623	0.672	0.350	0.086
$PiCANet_{18}$ [35]	47.22	54.06	0.777	0.769	0.749	0.536	0.085	0.596	0.609	0.584	0.356	0.156	0.587	0.649	0.643	0.322	0.090
$MSRCNN_{19}$ [36]	-	-	0.671	0.637	0.686	0.443	0.091	0.653	0.617	0.669	0.454	0.133	0.605	0.641	0.706	0.419	0.073
$BASNet_{19}$ [37]	87.06	97.48	0.633	0.687	0.721	0.474	0.118	0.584	0.618	0.661	0.413	0.159	0.504	0.634	0.678	0.365	0.105
$PFANet_{19}$ [38]	16.29	27.82	0.602	0.679	0.648	0.378	0.144	0.631	0.659	0.622	0.391	0.172	0.549	0.636	0.618	0.286	0.128
HTC_{19} [39]	-	-	0.502	0.517	0.489	0.204	0.129	0.432	0.476	0.442	0.174	0.172	0.505	0.548	0.520	0.221	0.088
CPD_{19} [40]	47.85	-	0.824	**0.853**	0.866	0.706	0.052	0.724	0.726	0.729	0.550	0.115	0.669	0.747	0.770	0.508	0.059
$EGNet_{19}$ [41]	111.66	120.85	**0.830**	0.848	0.870	0.702	0.050	0.733	0.732	0.768	0.583	0.104	0.683	0.737	0.779	0.509	0.056
PoolNet-R+	34.12	14.03	0.828	0.838	**0.887**	**0.723**	**0.042**	**0.761**	**0.750**	**0.785**	**0.620**	**0.095**	**0.718**	**0.754**	**0.800**	**0.559**	**0.049**

Table 2
Quantitative camouflaged object detection results on three widely used datasets. As can be seen, our approach achieves the best results on almost all datasets in all five metrics.

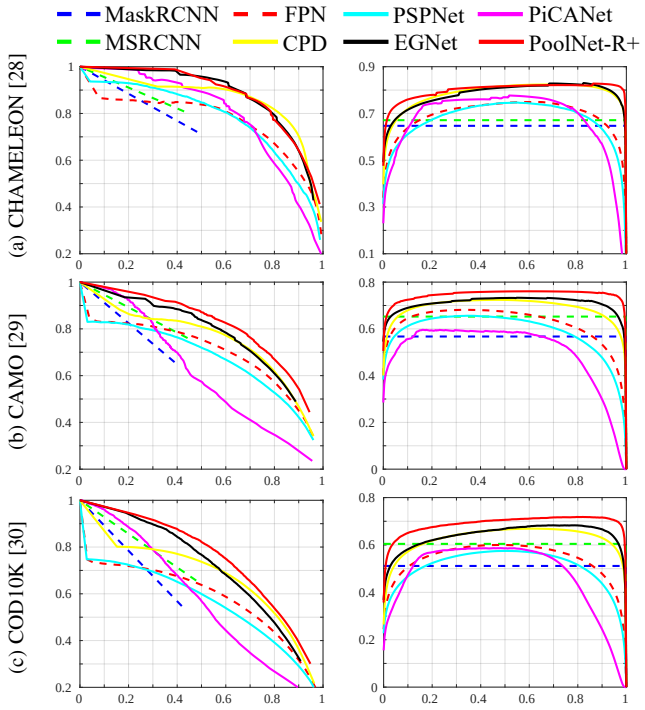


Fig. 3. PR and F-measure curves on three camouflaged object detection datasets. Left: precision (vertical) recall (horizontal) curves. Right: F-measure (vertical) and threshold (horizontal) curves.

extracted by a few convolutional layers are sufficient enough to make good predictions. We argue that designing more powerful depth feature extracting branches or more advanced cross-modal feature integration strategies between the RGB and depth branches can further promote the performances. The above phenomenon shows that our proposed approach can adapt and perform well even when transferred and applied to a task that takes input data from different domains. It verifies the generalization ability and robustness of the proposed approach.

2.2 Camouflaged Object Detection

2.2.1 Related Work

Unlike salient object detection that aims to detect the most attractive objects, camouflaged object detection intends to discover the camouflaged objects hidden in the surrounding. Early research in camouflaged object detection [42]–[45] can be traced decades ago. [46] pointed out that the colors of camouflaged objects are similar to their backgrounds and the boundaries between them and the background are ambiguous. [47] simulated the visual angle of predators to discover the hidden targets by combining the noteworthy features of camouflaged objects. [30] built a large-scale benchmark and proposed a deep network to solve this problem effectively. [48] introduced contradictory features and designed a similarity feature module for better detecting the salient and camouflaged targets.

2.2.2 Implementation Details

Like salient object detection, camouflaged object detection also takes RGB images as input and makes binary predictions. To this end, we directly apply the proposed PoolNet-R+ for camouflaged object detection without network modification. Following [30], we use the stochastic gradient descent (SGD) optimizer with a momentum of 0.9 and weight decay of 5e-5 for optimization. We train our network for 32 epochs with a batch size of 30. The initial learning rate is set to 5e-3, and the warming-up and cosine updating schedules are applied. Random horizontal flipping and cropping are used for data augmentation. The network is trained on the training subset of COD10K [30]. For testing, CHAMELEON [28], CAMO [29], and the testing subset of COD10K [30] are used. In both training and testing phases, the input images are resized to 352×352. We use six metrics for evaluation, including PR curves, F-measure (F_β), S-measure (S_m), E-measure (E_ϕ), weighted F-measure (F_β^w), and mean absolute error (MAE).

2.2.3 Comparisons to the State-of-the-Arts

We compare our approach with 11 state-of-the-art methods numerically in Table 2. One can see that our approach achieves the best results on three widely used datasets in terms of almost all metrics. Specifically, though the F-measure and S-measure scores of our approach on the CHAMELEON [28] dataset (76 test images) do not rank first, the performances of ours on the larger COD10K [30] dataset (2026 test images) outperform the previous state-of-the-art methods by large margins. This phenomenon could better demonstrate the effectiveness of our approach. As for the consumption of computational resources, only the parameters of PFANet [38] are lower than ours. However, PFANet [38] performs much worse than our approach (averagely more than 17%, 12%, and 19% in terms of F-measure, S-measure, and E-measure, respectively)

and requires 98% more MAdds. Compared to the previous best-performing method, our approach still averagely outperforms about 2% in F-measure and E-measure, while requiring only 30% and 12% parameters and MAdds, respectively. Similar phenomena can be observed in the PR and F-measure curves that our approach (red) is more outstanding than the others, as shown in Fig. 3. The above experimental results show that our approach performs well when applied to a task with a very different purpose (*i.e.*, salient *v.s.* camouflaged), verifying our approach's good generalization ability and robustness.

REFERENCES

[1] Y. Cheng, H. Fu, X. Wei, J. Xiao, and X. Cao, "Depth enhanced saliency detection method," in *Proceedings of international conference on internet multimedia computing and service*, 2014, pp. 23–27.

[2] C. Zhu, G. Li, W. Wang, and R. Wang, "An innovative salient object detection using center-dark channel prior," in *Int. Conf. Comput. Vis. Worksh.*, 2017, pp. 1509–1515.

[3] L. Qu, S. He, J. Zhang, J. Tian, Y. Tang, and Q. Yang, "Rgbd salient object detection via deep fusion," *IEEE Trans. Image Process.*, vol. 26, no. 5, pp. 2274–2285, 2017.

[4] R. Shigematsu, D. Feng, S. You, and N. Barnes, "Learning rgb-d salient object detection using background enclosure, depth contrast, and top-down features," in *Int. Conf. Comput. Vis.*, 2017, pp. 2749–2757.

[5] Z. Liu, S. Shi, Q. Duan, W. Zhang, and P. Zhao, "Salient object detection for rgb-d image by single stream recurrent convolution neural network," *Neurocomputing*, vol. 363, pp. 46–57, 2019.

[6] H. Chen and Y. Li, "Progressively complementarity-aware fusion network for rgb-d salient object detection," in *IEEE Conf. Comput. Vis. Pattern Recog.*, 2018, pp. 3051–3060.

[7] J. Han, H. Chen, N. Liu, C. Yan, and X. Li, "Cnns-based rgb-d saliency detection via cross-view transfer and multiview fusion," *IEEE transactions on cybernetics*, vol. 48, no. 11, pp. 3171–3183, 2017.

[8] J.-X. Zhao, Y. Cao, D.-P. Fan, M.-M. Cheng, X.-Y. Li, and L. Zhang, "Contrast prior and fluid pyramid integration for rgbd salient object detection," in *IEEE Conf. Comput. Vis. Pattern Recog.*, 2019, pp. 3927–3936.

[9] H. Chen, Y. Li, and D. Su, "Multi-modal fusion network with multi-scale multi-path and cross-modal interactions for rgb-d salient object detection," *Pattern Recognition*, vol. 86, pp. 376–385, 2019.

[10] H. Chen and Y. Li, "Three-stream attention-aware network for rgb-d salient object detection," *IEEE Trans. Image Process.*, vol. 28, no. 6, pp. 2825–2835, 2019.

[11] Y. Piao, W. Ji, J. Li, M. Zhang, and H. Lu, "Depth-induced multi-scale recurrent attention network for saliency detection," in *Int. Conf. Comput. Vis.*, 2019, pp. 7254–7263.

[12] N. Liu, N. Zhang, and J. Han, "Learning selective self-mutual attention for rgb-d saliency detection," in *IEEE Conf. Comput. Vis. Pattern Recog.*, 2020, pp. 13 756–13 765.

[13] D.-P. Fan, Z. Lin, Z. Zhang, M. Zhu, and M.-M. Cheng, "Rethinking rgb-d salient object detection: Models, data sets, and large-scale benchmarks," *IEEE Trans. Neural Netw. and Learn. Syst.*, 2020.

[14] K. Fu, D.-P. Fan, G.-P. Ji, and Q. Zhao, "Jl-dcf: Joint learning and densely-cooperative fusion framework for rgb-d salient object detection," in *IEEE Conf. Comput. Vis. Pattern Recog.*, 2020, pp. 3052–3062.

[15] D.-P. Fan, Y. Zhai, A. Borji, J. Yang, and L. Shao, "Bbs-net: Rgb-d salient object detection with a bifurcated backbone strategy network," in *Eur. Conf. Comput. Vis.* Springer, 2020, pp. 275–292.

[16] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Int. Conf. Learn. Represent.*, 2015.

[17] R. Ju, L. Ge, W. Geng, T. Ren, and G. Wu, "Depth saliency based on anisotropic center-surround difference," in *IEEE Int. Conf. Image Process.*, 2014, pp. 1115–1119.

[18] H. Peng, B. Li, W. Xiong, W. Hu, and R. Ji, "RGBD salient object detection: a benchmark and algorithms," in *Eur. Conf. Comput. Vis.*, 2014, pp. 92–109.

[19] Y. Niu, Y. Geng, X. Li, and F. Liu, "Leveraging stereopsis for saliency analysis," in *IEEE Conf. Comput. Vis. Pattern Recog.*, 2012, pp. 454–461.

[20] Y. Cheng, H. Fu, X. Wei, J. Xiao, and X. Cao, "Depth enhanced saliency detection method," in *ACM International Conference on Internet Multimedia Computing and Service*, 2014, p. 23.

[21] D.-P. Fan, Z. Lin, Z. Zhang, M. Zhu, and M.-M. Cheng, "Rethinking rgb-d salient object detection: Models, data sets, and large-scale benchmarks," *IEEE Trans. Neural Netw. and Learn. Syst.*, 2020.

[22] Q. Yan, L. Xu, J. Shi, and J. Jia, "Hierarchical saliency detection," in *IEEE Conf. Comput. Vis. Pattern Recog.*, 2013, pp. 1155–1162.

[23] Y. Li, X. Hou, C. Koch, J. M. Rehg, and A. L. Yuille, "The secrets of salient object segmentation," in *IEEE Conf. Comput. Vis. Pattern Recog.*, 2014, pp. 280–287.

[24] C. Yang, L. Zhang, H. Lu, X. Ruan, and M.-H. Yang, "Saliency detection via graph-based manifold ranking," in *IEEE Conf. Comput. Vis. Pattern Recog.*, 2013, pp. 3166–3173.

[25] G. Li and Y. Yu, "Visual saliency based on multiscale deep features," in *IEEE Conf. Comput. Vis. Pattern Recog.*, 2015, pp. 5455–5463.

[26] L. Wang, H. Lu, Y. Wang, M. Feng, D. Wang, B. Yin, and X. Ruan, "Learning to detect salient objects with image-level supervision," in *IEEE Conf. Comput. Vis. Pattern Recog.*, 2017, pp. 136–145.

[27] N. Wang and X. Gong, "Adaptive fusion for rgb-d salient object detection," *IEEE Access*, vol. 7, pp. 55 277–55 284, 2019.

[28] P. Skurowski, H. Abdulameer, J. Błaszczyk, T. Depta, A. Kornacki, and P. Kozieł, "Animal camouflage analysis: Chameleon database," *Unpublished Manuscript*, 2018.

[29] T.-N. Le, T. V. Nguyen, Z. Nie, M.-T. Tran, and A. Sugimoto, "Anabranch network for camouflaged object segmentation," *Computer Vision and Image Understanding*, vol. 184, pp. 45–56, 2019.

[30] D.-P. Fan, G.-P. Ji, G. Sun, M.-M. Cheng, J. Shen, and L. Shao, "Camouflaged object detection," in *IEEE Conf. Comput. Vis. Pattern Recog.*, 2020, pp. 2777–2787.

[31] T.-Y. Lin, P. Dollár, R. B. Girshick, K. He, B. Hariharan, and S. J. Belongie, "Feature pyramid networks for object detection," in *IEEE Conf. Comput. Vis. Pattern Recog.*, 2017.

[32] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask r-cnn," in *Int. Conf. Comput. Vis.*, 2017, pp. 2961–2969.

[33] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, "Pyramid scene parsing network," in *IEEE Conf. Comput. Vis. Pattern Recog.*, 2017.

[34] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, "Unet++: A nested u-net architecture for medical image segmentation," in *Deep learning in medical image analysis and multimodal learning for clinical decision support*. Springer, 2018, pp. 3–11.

[35] N. Liu, J. Han, and M.-H. Yang, "Picanet: Learning pixel-wise contextual attention for saliency detection," in *IEEE Conf. Comput. Vis. Pattern Recog.*, 2018, pp. 3089–3098.

[36] Z. Huang, L. Huang, Y. Gong, C. Huang, and X. Wang, "Mask scoring r-cnn," in *IEEE Conf. Comput. Vis. Pattern Recog.*, 2019, pp. 6409–6418.

[37] X. Qin, Z. Zhang, C. Huang, C. Gao, M. Dehghan, and M. Jagersand, "Basnet: Boundary-aware salient object detection," in *IEEE Conf. Comput. Vis. Pattern Recog.*, 2019.

[38] T. Zhao and X. Wu, "Pyramid feature attention network for saliency detection," in *IEEE Conf. Comput. Vis. Pattern Recog.*, 2019.

[39] K. Chen, J. Pang, J. Wang, Y. Xiong, X. Li, S. Sun, W. Feng, Z. Liu, J. Shi, W. Ouyang *et al.*, "Hybrid task cascade for instance segmentation," in *IEEE Conf. Comput. Vis. Pattern Recog.*, 2019, pp. 4974–4983.

[40] Z. Wu, L. Su, and Q. Huang, "Cascaded partial decoder for fast and accurate salient object detection," in *IEEE Conf. Comput. Vis. Pattern Recog.*, 2019.

[41] J.-X. Zhao, J.-J. Liu, D.-P. Fan, Y. Cao, J. Yang, and M.-M. Cheng, "Egnet:edge guidance network for salient object detection," in *Int. Conf. Comput. Vis.*, 2019.

[42] R. R. Behrens, "The theories of abbott h. thayer: Father of camouflage," *Leonardo*, vol. 21, no. 3, pp. 291–296, 1988.

[43] ——, "Seeing through camouflage: Abbott thayer, background-picturing and the use of cutout silhouettes," *Leonardo*, vol. 51, no. 1, pp. 40–46, 2018.

[44] I. Cuthill, "Camouflage," *Journal of Zoology*, vol. 308, no. 2, pp. 75–92, 2019.

[45] M. Stevens and S. Merilaita, "Animal camouflage: current issues and new perspectives," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 364, no. 1516, pp. 423–427, 2009.

[46] I. C. Cuthill, M. Stevens, J. Sheppard, T. Maddocks, C. A. Párraga, and T. S. Troscianko, "Disruptive coloration and background pattern matching," *Nature*, vol. 434, no. 7029, pp. 72–74, 2005.

[47] T. W. Pike, "Quantifying camouflage and conspicuousness using visual salience," *Methods in Ecology and Evolution*, vol. 9, no. 8, pp. 1883–1895, 2018.

[48] A. Li, J. Zhang, Y. Lv, B. Liu, T. Zhang, and Y. Dai, "Uncertainty-aware joint salient object and camouflaged object detection," *arXiv preprint arXiv:2104.02628*, 2021.