Towards High-Resolution Salient Object Detection(Supplementary Material)

Yi Zeng[1], Pingping Zhang[1], Jianming Zhang[2], Zhe Lin[2], Huchuan Lu[1*]
[1] Dalian University of Technology, China
[2] Adobe Research, USA
{dllgzy, jssxzhpp}@mail.dlut.edu.cn, {jianmzha, zlin}@adobe.com, lhchuan@dlut.edu.cn

1. Network Architecture for GSN and LRN

We present the detailed network architecture of Global Semantic Network (GSN) and Local Refinement Network (LRN) as follows. The unique layers for LRN are shown in red.

Layer	Layer type	Input	Filter size	Stride	Zero-Padding	Dilation rate	Output	Output channel
1	Conv+Relu	Image	3×3	1	1	-	Conv1-1	64
2	Conv+Relu	Conv1-1	3×3	1	1	-	Conv1-2	64
3	Max-Pool	Conv1-2	2×2	2	0	-	Pool1	64
4	Conv+Relu	Pool1	3×3	1	1	-	Conv2-1	128
5	Conv+Relu	Conv2-1	3×3	1	1	-	Conv2-2	128
6	Max-Pool	Conv2-2	2×2	2	0	-	Pool2	128
7	Conv+Relu	Pool2	3×3	1	1	-	Conv3-1	256
8	Conv+Relu	Conv3-1	3×3	1	1	-	Conv3-2	256
9	Conv+Relu	Conv3-2	3×3	1	1	-	Conv3-3	256
10	Max-Pool	Conv3-3	2×2	2	0	-	Pool3	256
11	Conv+Relu	Pool3	3×3	1	1	-	Conv4-1	512
12	Conv+Relu	Conv4-1	3×3	1	1	-	Conv4-2	512
13	Conv+Relu	Conv4-2	3×3	1	1	-	Conv4-3	512
14	Max-Pool	Conv4-3	2×2	2	0	-	Pool4	512
15	Conv+Relu	Pool4	3×3	1	1	-	Conv5-1	512
16	Conv+Relu	Conv5-1	3×3	1	1	-	Conv5-2	512
17	Conv+Relu	Conv5-2	3×3	1	1	-	Conv5-3	512
18	Conv+Relu	Conv5-3	3×3	1	1	1	ASPP-5-1	32
19	Conv+Relu	Conv5-3	3×3	1	3	3	ASPP-5-3	32
20	Conv+Relu	Conv5-3	3×3	1	5	5	ASPP-5-5	32
21	Conv+Relu	Conv5-3	3×3	1	7	7	ASPP-5-7	32
22	Concat	ASPP-5-1 ASPP-5-3 ASPP-5-5 ASPP-5-7	-	-	-	-	Concat1	128
23	Conv+Relu	Concat1	3×3	1	1	-	Conv5-S	2
24	Deconv+Relu	Conv5-S	4×4	2	1	-	UpMap1	2
25	Conv+Relu	Conv4-3	3×3	1	1	1	ASPP-4-1	32

*Corresponding author.

Layer	Layer type	Input	Filter size	Stride	Zero-Padding	Dilation rate	Output	Output channel
26	Conv+Relu	Conv4-3	3×3	1	3	3	ASPP-4-3	32
27	Conv+Relu	Conv4-3	3×3	1	5	5	ASPP-4-5	32
28	Conv+Relu	Conv4-3	3×3	1	7	7	ASPP-4-7	32
29	Concat	ASPP-4-1 ASPP-4-3 ASPP-4-5 ASPP-4-7	-	-	-	-	Concat2	128
30	Conv+Relu	Concat2	3×3	1	1	-	Conv4-S	2
31	Elt-Sum	Conv4-S UpMap1	-	-	-	-	FuseMap1	2
32	Concat	FuseMap1 Guidance1	-	-	-	-	Concat3	2
33	Conv+Relu	Concat3	3×3	1	1	-	Map1	2
34	Deconv+Relu	Map1	4×4	2	1	-	UpMap2	2
35	Conv+Relu	Conv3-3	3×3	1	1	1	ASPP-3-1	32
36	Conv+Relu	Conv3-3	3×3	1	3	3	ASPP-3-3	32
37	Conv+Relu	Conv3-3	3×3	1	5	5	ASPP-3-5	32
38	Conv+Relu	Conv3-3	3×3	1	7	7	ASPP-3-7	32
39	Concat	ASPP-3-1 ASPP-3-3 ASPP-3-5 ASPP-3-7	-	-	-	-	Concat4	128
40	Conv+Relu	Concat4	3×3	1	1	-	Conv3-S	2
41	Elt-Sum	Conv3-S UpMap2	-	-	-	-	FuseMap2	2
42	Concat	FuseMap2 Guidance2	-	-	-	-	Concat5	2
43	Conv+Relu	Concat5	3×3	1	1	-	Map2	2
44	Deconv	Map2	8×8	4	2	-	UpMap3	2
45	SoftMax	UpMap3	–	-	-	-	SalMap	2

2. Shape complexity for different datasets

To evaluated shape complexity, we adopt a kind of area-perimeter approach. Intuitively, the ratio of area to perimeter can reflect the shape complexity to some extend. A region with longer perimeter and smaller area has more complex boundaries. This type of region is more challenging in dense prediction tasks. Although straightforward, this simplistic ratio suffers from the fact that the measure varies when shape size changes. C_{IPQ} [7] can largely remedy this problem and has become one of the most widely accepted metrics of this class. This metric can be formulated as:

$$C_{IPQ} = \frac{4\pi A}{P^2} \tag{1}$$

Table 1 shows quantitative comparisons with other datasets in term of C_{IPQ} .

Dataset	HRSOD-Test	HRSOD-TR	DUTS-Test	DUTS-TR	ECSSD	HKU-IS	PASCAL-S	THUR	THUS
C_{IPQ}	0.028	0.034	0.028	0.023	0.062	0.042	0.035	0.040	0.072

Table 1. The C_{IPQ} scores of different datasets (smaller is better).

3. Detailed algorithm for APS

More details about APS are shown in Algorithm 1.

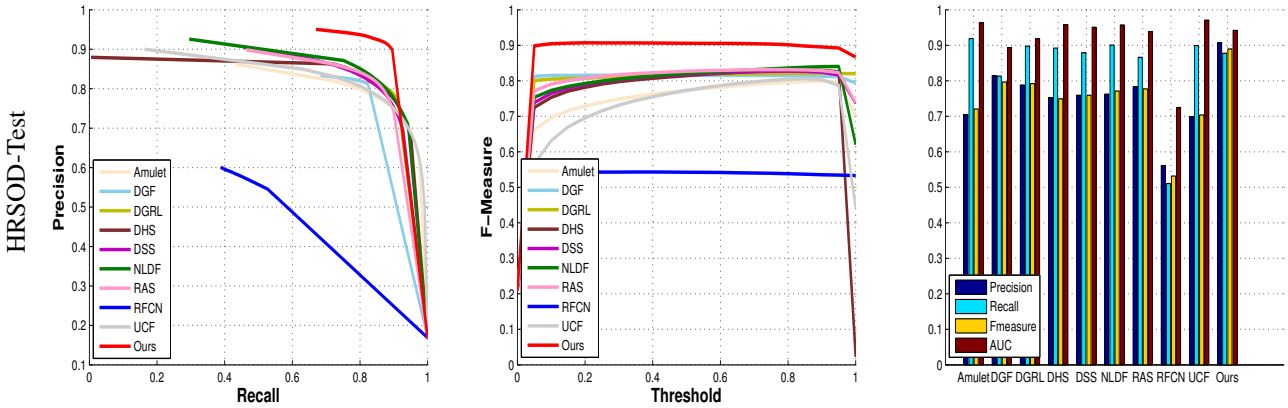
Algorithm 1 Attended Patch Sampling.

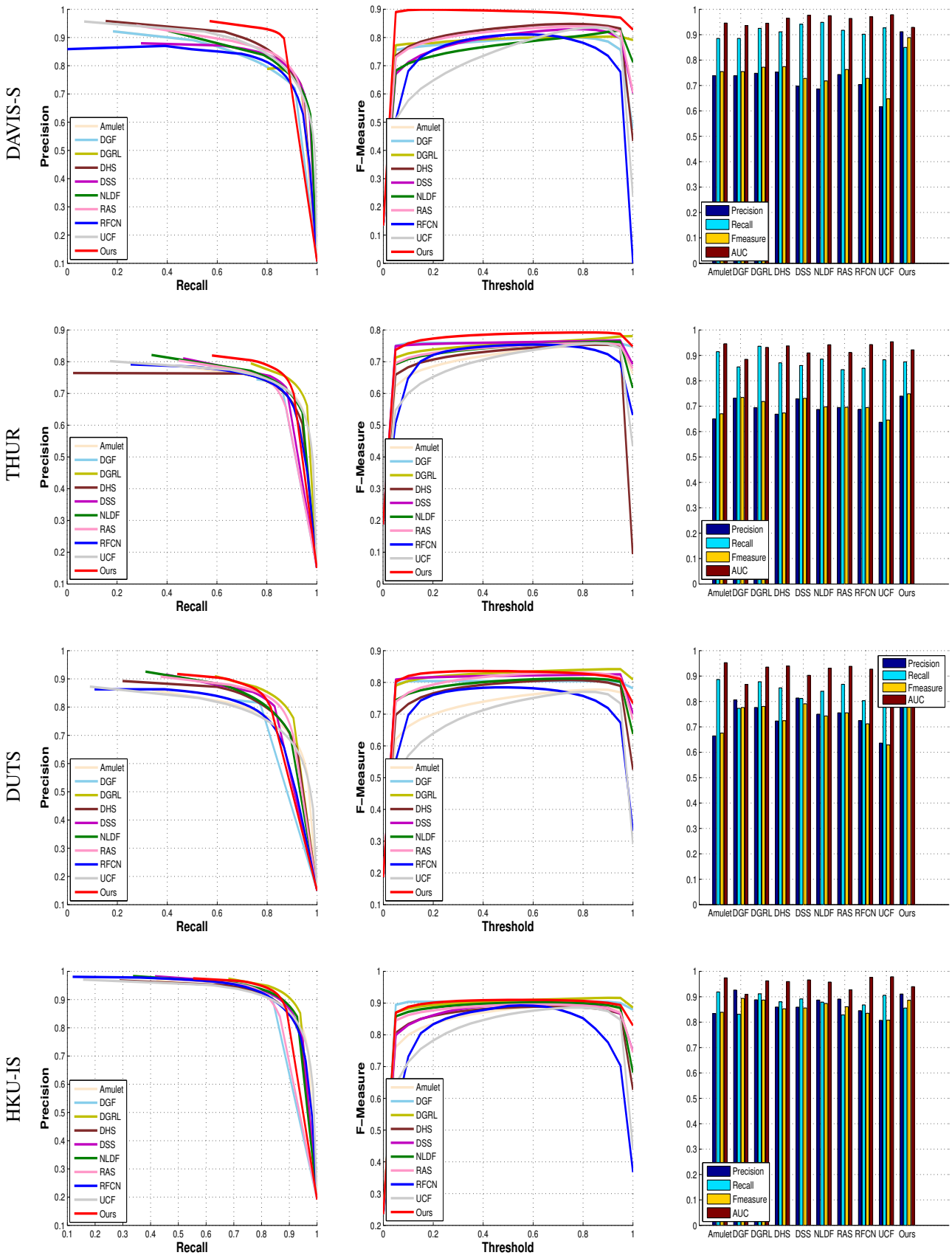
Input: RGB image I_i, ground truth label L_i, base cropping size D.
Output: RGB patch set $\{P_m^{I_i}\}_{m=1}^{M}$, ground truth patch set $\{P_m^{L_i}\}_{m=1}^{M}$.

Generate attention map A_i from F_i.
$N_x = \lceil w/D \rceil + n$
for $t = 1, \ldots, N_x + 1$ **do**
  $C = D + r$
  $X_t = min\{X_L + (t-1) \times \lceil w/N_x \rceil, X_R\}$
  $Y = \{y \mid A_i(X_t, y) = 1\}$
  $D_y = max\{Y\}$ - $min\{Y\}$
  if $D_y < 0.5 \times$C **then**
    Randomly generate a $y \in Y$ and $a_x \leftarrow X_t, a_y \leftarrow y$
  else if $0.5 \times C < D_y < C$ **then**
    $a_x(1) \leftarrow X_t, a_y(1) \leftarrow min\{Y\}$
    $a_x(2) \leftarrow X_t, a_y(2) \leftarrow max\{Y\}$
  else if $D_y > C$ **then**
    $N_y = \lceil D_y/C \rceil$
    $d = \lceil D_y/N_y \rceil$
    for $t_y = 1, \ldots, N_y + 1$ **do**
      $a_x(t_y) \leftarrow X_t$
      $a_y(t_y) \leftarrow min\{min\{Y\} + d \times (t_y - 1), max\{Y\}\}$
    end for
  end if
  Taking C as cropping size, $(a_x(j), a_y(j))_{j=1}^{J}$ as center pixels, crop $\{P_j^{I_i}\}_{j=1}^{J}$ and $\{P_j^{L_i}\}_{j=1}^{J}$ from I_i and L_i, respectively (J is the actual number of sampled pixels according to the size of D_y.).
end for

4. Quantitative Comparison

We choose two high-resolution datasets (*i.e.*, HRSOD-Test and DAVIS-S) and three widely used benchmark datasets (*i.e.*, THUR [2], HKU-IS [4], and DUTS [8]) to evaluate all methods. We compare PR curves, F-measures curves and F-measure scores with 9 state-of-the-art deep learning-based methods, including RFCN [9], DHS [5], UCF [13], Amulet [12], NLDF [6], DSS [3], RAS [1] DGF [11] and DGRL [10]:




DAVIS-S
THUR
DUTS
HKU-IS
Precision
Recall
F−Measure
Threshold
Amulet
DGF
DGRL
DHS
DSS
NLDF
RAS
RFCN
UCF
Ours
Fmeasure
AUC

5. Visual Comparison

We present qualitative results of our method and other 9 state-of-the-art methods on 2 high-resolution datasets: HRSOD-Test and DAVIS-S.These results are better to be viewed by zooming in.

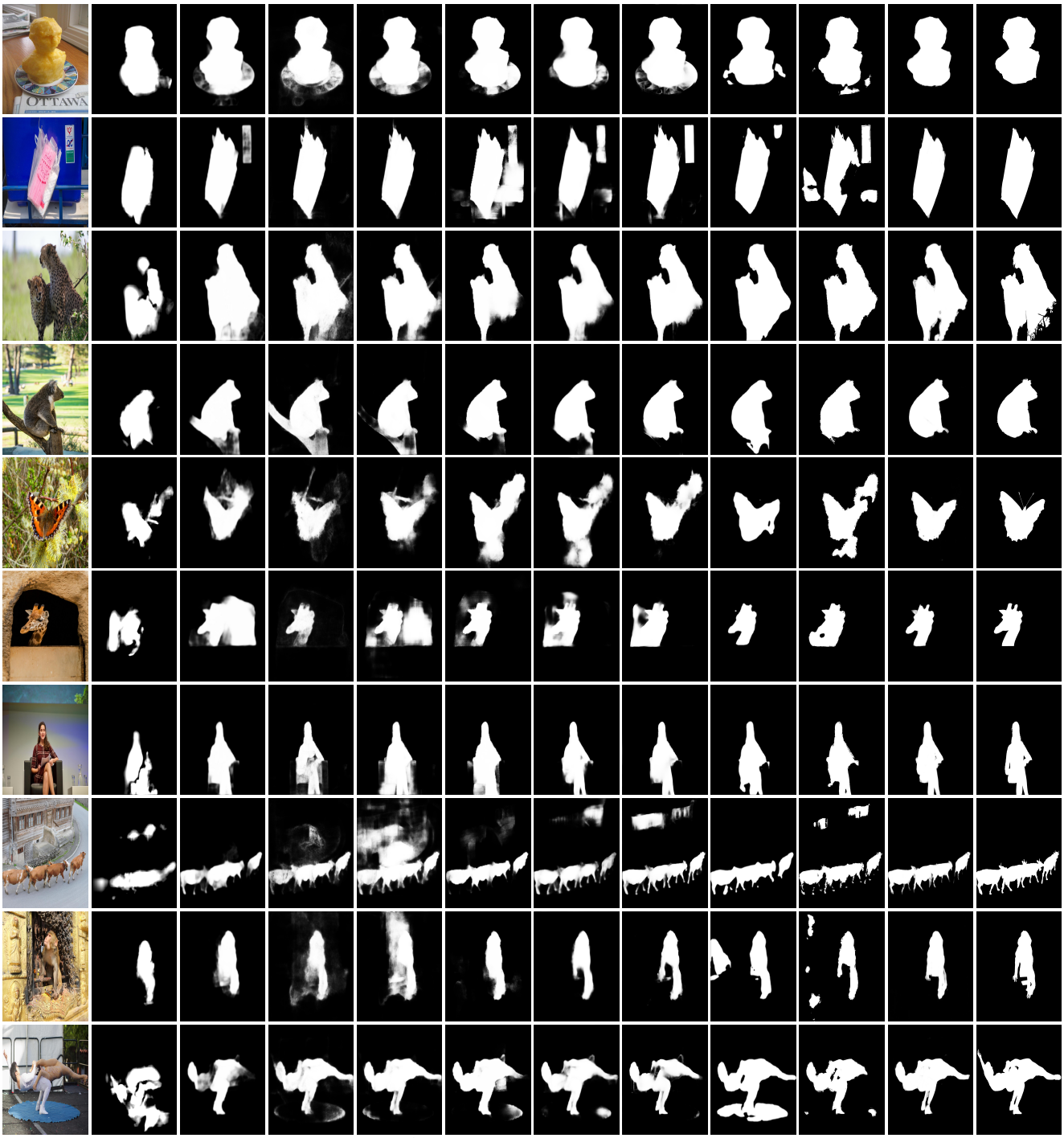


Figure 1. HRSOD-Test dataset

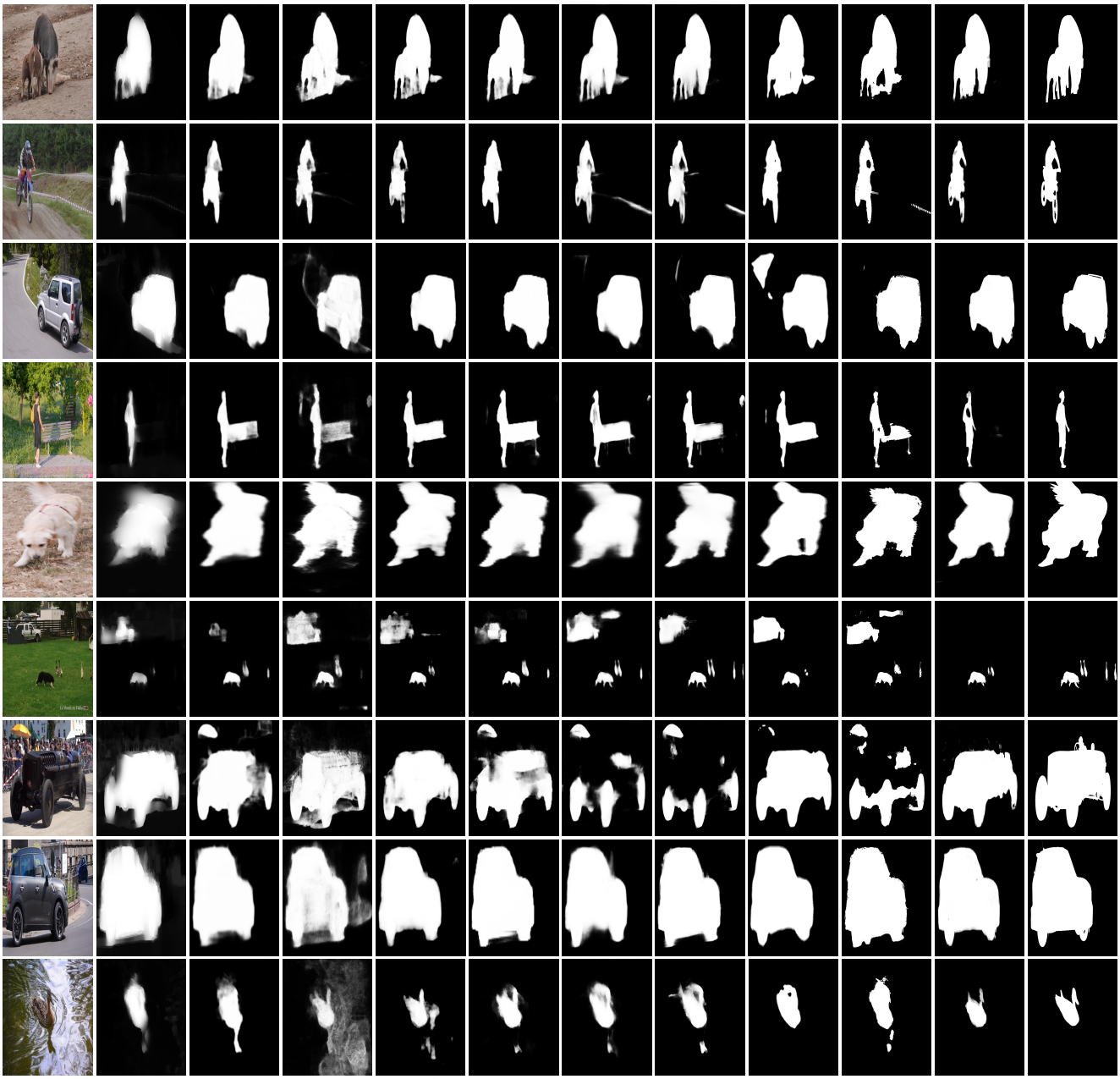


Figure 2. DAVIS-S dataset

References

[1] Shuhan Chen, Xiuli Tan, Ben Wang, and Xuelong Hu. Reverse attention for salient object detection. In *European Conference on Computer Vision*, 2018.

[2] Ming-Ming Cheng, Niloy J Mitra, Xiaolei Huang, and Shi-Min Hu. Salientshape: Group saliency in image collections. *The Visual Computer*, 30(4):443–453, 2014.

[3] Qibin Hou, Ming-Ming Cheng, Xiaowei Hu, Ali Borji, Zhuowen Tu, and Philip Torr. Deeply supervised salient object detection with short connections. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5300–5309. IEEE, 2017.

[4] Guanbin Li and Yizhou Yu. Visual saliency based on multiscale deep features. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 5455–5463, 2015.

[5] Nian Liu and Junwei Han. Dhsnet: Deep hierarchical saliency network for salient object detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 678–686, 2016.

[6] Zhiming Luo, Akshaya Kumar Mishra, Andrew Achkar, Justin A Eichel, Shaozi Li, and Pierre-Marc Jodoin. Non-local deep features for salient object detection. In *CVPR*, volume 2, page 7, 2017.

[7] Robert Osserman et al. The isoperimetric inequality. *Bulletin of the American Mathematical Society*, 84(6):1182–1238, 1978.

[8] Lijun Wang, Huchuan Lu, Yifan Wang, Mengyang Feng, Dong Wang, Baocai Yin, and Xiang Ruan. Learning to detect salient objects with image-level supervision. In *CVPR*, 2017.

[9] Linzhao Wang, Lijun Wang, Huchuan Lu, Pingping Zhang, and Xiang Ruan. Saliency detection with recurrent fully convolutional networks. In *European Conference on Computer Vision*, pages 825–841. Springer, 2016.

[10] Tiantian Wang, Lihe Zhang, Shuo Wang, Huchuan Lu, Gang Yang, Xiang Ruan, and Ali Borji. Detect globally, refine locally: A novel approach to saliency detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3127–3135, 2018.

[11] Huikai Wu, Shuai Zheng, Junge Zhang, and Kaiqi Huang. Fast end-to-end trainable guided filter. In *CVPR*, 2018.

[12] Pingping Zhang, Dong Wang, Huchuan Lu, Hongyu Wang, and Xiang Ruan. Amulet: Aggregating multi-level convolutional features for salient object detection. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 202–211, 2017.

[13] Pingping Zhang, Dong Wang, Huchuan Lu, Hongyu Wang, and Baocai Yin. Learning uncertain convolutional features for accurate saliency detection. In *Computer Vision (ICCV), 2017 IEEE International Conference on*, pages 212–221. IEEE, 2017.