JUNE 18-22, 2023
CVPR
VANCOUVER, CANADA

Pixels, Regions, and Objects: Multiple Enhancement for Salient Object Detection

Yi Wang[1], Ruili Wang[2*], Xin Fan[1], Tianzhu Wang[2], Xiangjian He[3*]

[1]Dalian University of Technology (China)

[2]Massey University (New Zealand)

[3]University of Nottingham Ningbo (China)

Introduction

Salient Object Detection (SOD)

- Aiming for precise *localization* and *segmentation* of the most eye catching regions that conform to Human Visual System (HVS)
- Criteria: accurate boundary, high resolution, and computational efficiency

GT   Image

Figure 1. Salient object detection examples.

Challenging Issues

Accurate boundaries for

- partially obstructed objects
- multiple connected objects
- camouflage objects
- small objects
- geometrically complex objects

....

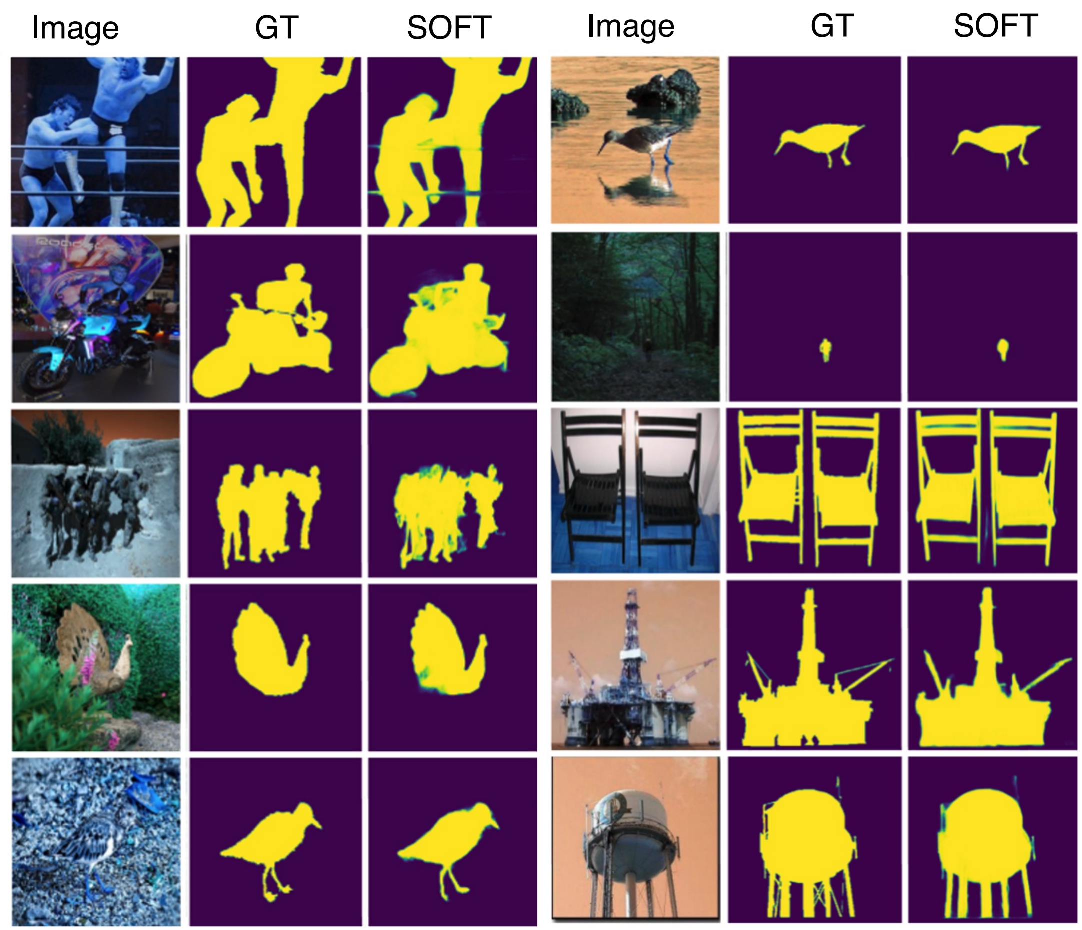


Figure 2. Challenging examples in SOD.

Motivations

To effectively integrate human visual system (HVS) mechanisms [1-4] to improve detection accuracy.

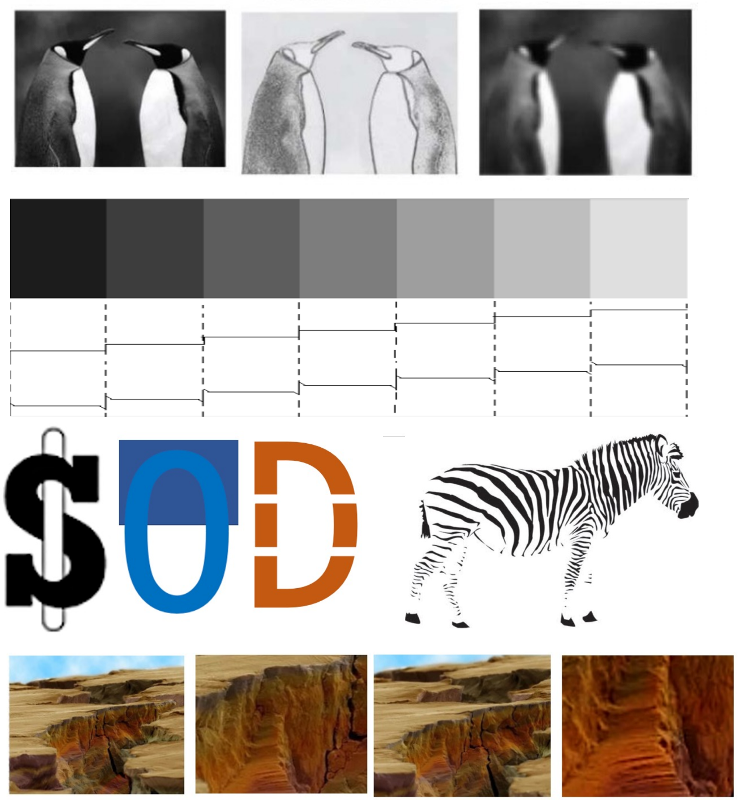
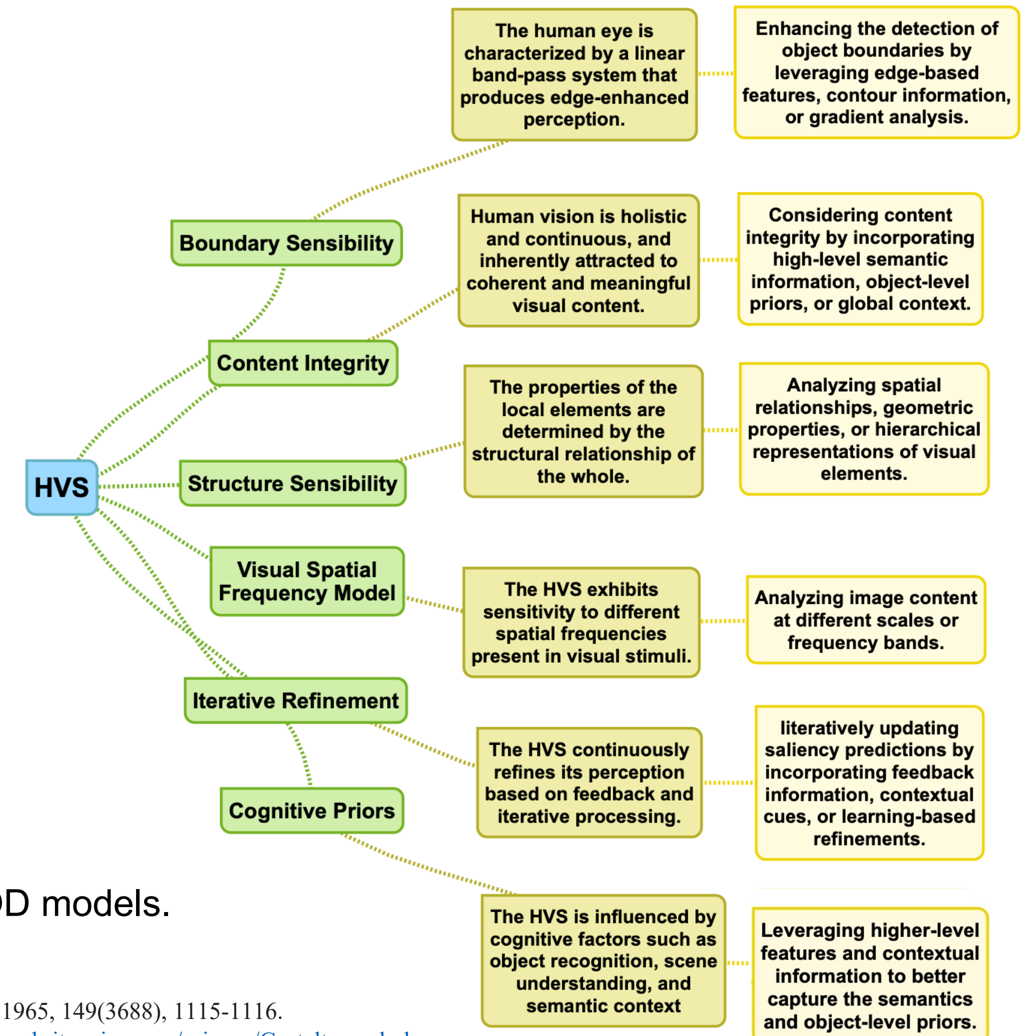


Figure 3. Some HVS mechanisms used in SOD models.



[1] Understanding vision: theory, models, and data. OUP Oxford, 2014.
[2] Spatial vision, 1988, no. 14, iSBN 13:9781423734710.
[3] Color adaptation of edge-detectors in the human visual system. Science, 1965, 149(3688), 1115-1116.
[4] "Gestalt psychology". Encyclopedia Britannica, 27 Apr. 2023, https://www.britannica.com/science/Gestalt-psychology

Methodology

Two-stream Network Structure with Boundary Enhancement

- Supervised boundary feature (high frequencies) learning by the frequency decomposed GT maps.
- Adaptative Inner body feature (low-frequencies) learning.

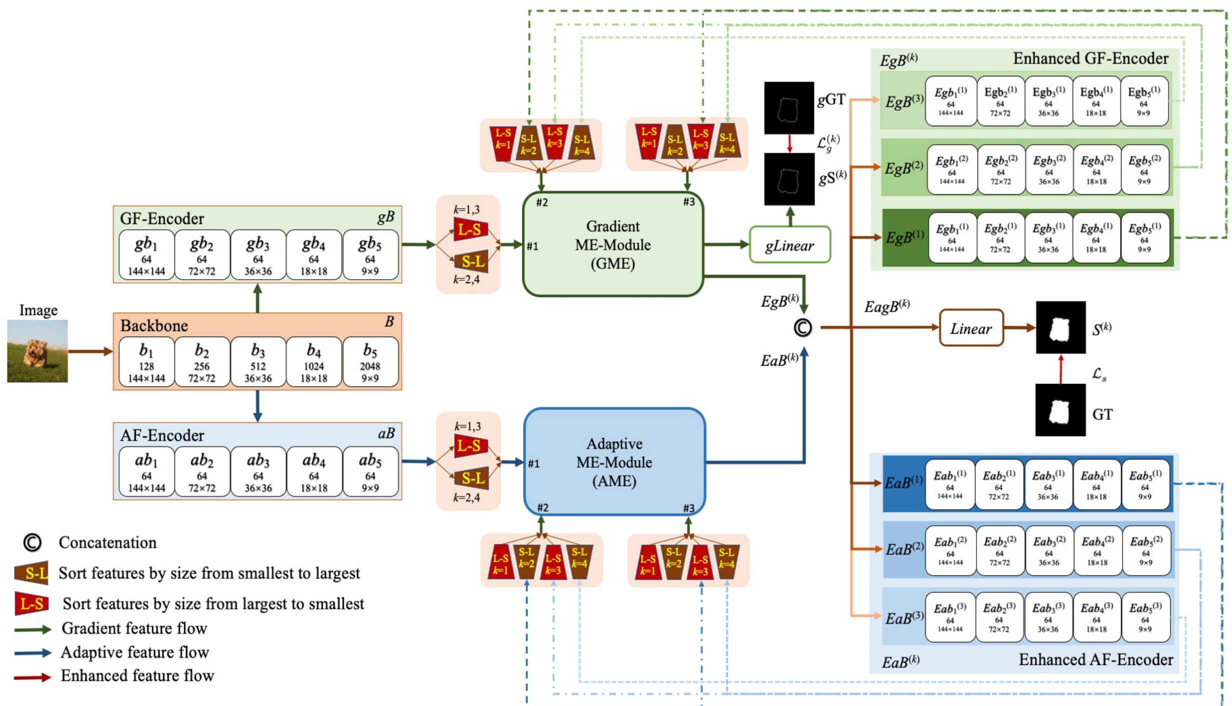


Figure 4. Illustration of the overall architecture and the pipeline of the MENet.

Multiscale Feature Enhancement

- ASPP diversifies visual fields
- ME-Attention module emphasizes location
- ME-module can output high- and low-level features merely by changing the input order of multiscale feature maps in terms of spatial size

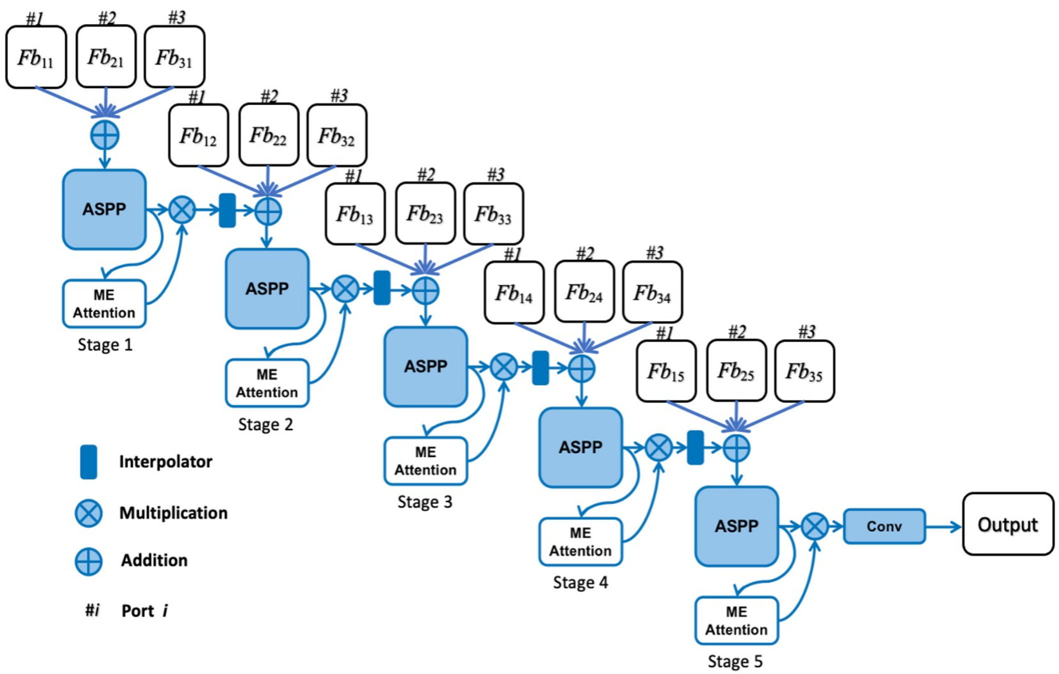


Figure 5. Illustration of the ME-Module.

Methodology

Iterative Enhancement

- Global- and detail features complement each other and promote each other.
- Odd number iterations (k=1,3) extract global features
- Even number iterations (k=2,4) refine detail features

$$EgB^{(k)} = GME(V(gB), V(EgB^{(k-1)}), V(EgB^{(k-2)})), \quad (1)$$

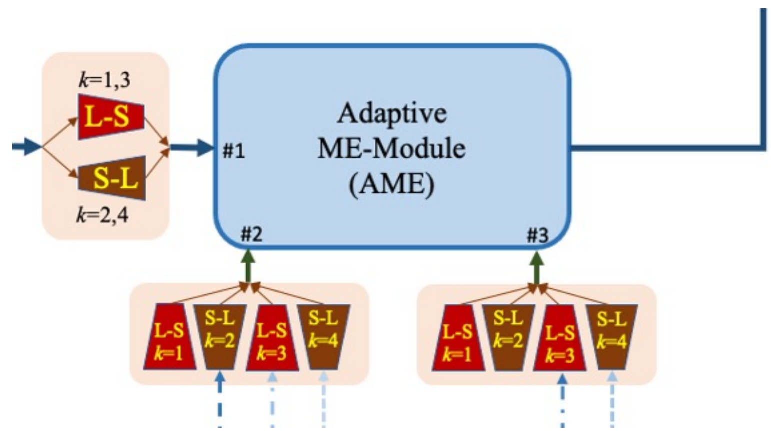
where the $GME(p_1, p_2, p_3)$ represents the Graduate ME-Module and $V(\;)$ is used to get the reverse of the feature list. If $k \leq 0$, we let $p_i = Null$.

$$EaB^{(k)} = AME(V(aB), V(EaB^{(k-1)}), V(EaB^{(k-2)})), \quad (2)$$

where $AME(q_1, q_2, q_3)$ represents the Adaptive ME-Module. If $k \leq 0$, we let $q_i = Null$.

$$S^{(k)} = Linear(Concat(EgB^{(k)}, EaB^{(k)})), \quad (3)$$

where $Linear(\;)$ is the linear layer and $Concat(\;,\;)$ is the concatenation operation.



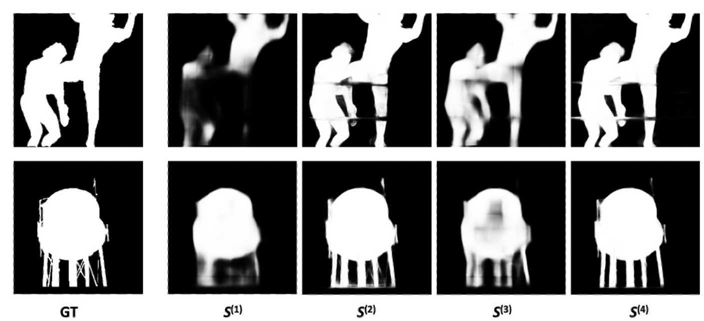


Figure 6. Visualized examples of the saliency map ($S^{(k)}$).

Methodology

New hybrid Loss

$$\mathcal{L}_s = \beta_1 \mathcal{L}_{s_{bce}} + \beta_2 \mathcal{L}_{s_{reg}} + \beta_3 \mathcal{L}_{s_{obj}}$$

- Pixel-level loss $\mathcal{L}_{s_{bce}} = -\sum (G log S + (1-G) log(1-S))$
- Region-level loss $\mathcal{L}_{s_{reg}} = 1 - \sum_{i=1}^{4} \omega_i (\theta_1 SSIM_i + \theta_2 IoU_i)$

S and G are evenly divided into four equal sub-regions.

- Object-level loss $\mathcal{L}_{s_{obj}} = 1 - \frac{1}{\left(\frac{\mu_{S_o}^2 + \mu_{G_o}^2}{2\mu_{S_o}\mu_{G_o}} + \lambda \frac{\sigma_{S_o}}{\mu_{S_o}}\right)} = 1 - \frac{2\mu_{S_o}}{\mu_{S_o}^2 + 1 + 2\lambda\sigma_{S_o}}$

  - Inspired by SSIM[5] and S-measure[6]
  - GT maps usually have sharp foreground-background contrast and a uniform distribution
  - the luminance component of SSIM
  - the coefficient of variation (i.e., the ratio of mean to deviation)
  - only consider foregrounds S_o and G_o

[5] Image quality assessment: from error visibility to structural similarity. IEEE TIP, 13(4):600–612, 2004.
[6] Structure measure: A new way to evaluate foreground maps. IJCV, 129(9):2622–2638, 2021.

Experiments

Ablation Study

Number of iterative enhancements

- Four-round enhancement setting achieves the best results on all databases.

Table 1. Comparison of iterative enhancement times.

Dataset	ITimes	$MAE\downarrow$	$F_\beta^{max}\uparrow$	$mF_\beta\uparrow$	$mE_m\uparrow$	$S_m\uparrow$
OMRON	1	0.0617	0.7867	0.7657	0.8701	0.7827
	2	0.0483	0.8304	0.8067	0.8854	0.8409
	3	0.0504	0.8066	0.7956	0.8755	0.8325
	4	**0.0450**	**0.8337**	**0.8178**	**0.8911**	**0.8496**
DUTS-TE	1	0.0496	0.8560	0.8386	0.9025	0.8363
	2	0.0310	0.9071	0.8762	0.9313	0.8969
	3	0.0389	0.8816	0.8690	0.9144	0.8783
	4	**0.0281**	**0.9123**	**0.8930**	**0.9368**	**0.9049**
HKU-IS	1	0.0504	0.9028	0.8864	0.9344	0.8649
	2	0.0266	0.9453	0.9184	0.9621	0.9205
	3	0.0383	0.9231	0.9102	0.9497	0.9024
	4	**0.0234**	**0.9483**	**0.9319**	**0.9657**	**0.9274**
PASCAL-S	1	0.0868	0.8525	0.8373	0.8611	0.8166
	2	0.0576	0.8856	0.8567	0.9059	0.8652
	3	0.0685	0.8726	0.8605	0.8728	0.8587
	4	**0.0535**	**0.8896**	**0.8701**	**0.9131**	**0.8721**
ECSSD	1	0.0673	0.9163	0.8988	0.9065	0.8617
	2	0.0344	0.9511	0.9298	0.9484	0.9213
	3	0.0503	0.9336	0.9208	0.9205	0.9016
	4	**0.0307**	**0.9549**	**0.9422**	**0.9544**	**0.9279**
SOD	1	0.1449	0.8388	0.7809	0.7497	0.7028
	2	0.0895	0.8700	0.8640	0.8352	0.8063
	3	0.1147	0.8548	0.8277	0.7795	0.7666
	4	**0.0874**	**0.8780**	**0.8684**	**0.8381**	**0.8089**

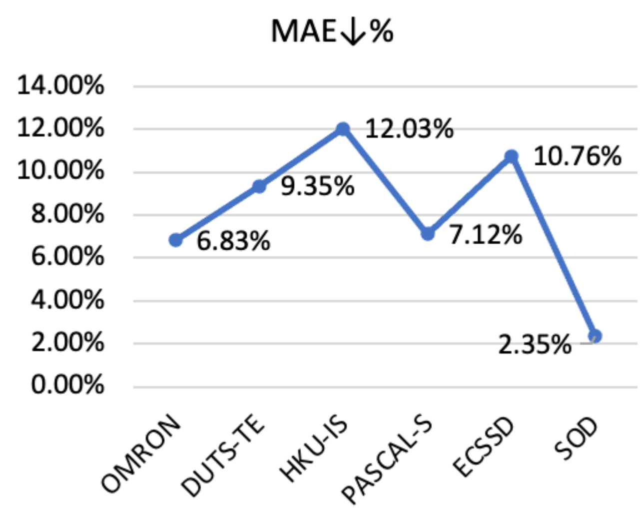


Figure 7. MAE reduction of iterative times from 2 to 4.

Experiments

Loss combinations

Table 2. Ablation test for the MENet loss setting.

Dataset	No.	$\mathcal{L}_g$	$\mathcal{L}_{s_{bce}}$	$\mathcal{L}_{s_{reg}}$	$\mathcal{L}_{s_{obj}}$	$MAE\downarrow$	$MaxF\uparrow$	$mF\uparrow$	$mE_m\uparrow$	$S_m\uparrow$
OMRON	1		✓			0.0518	0.8141	0.7933	0.8684	0.8407
	2		✓	✓4		0.0498	0.8075	0.7892	0.8632	0.8388
	3		✓	✓4	✓	0.0492	0.8281	0.8093	0.8840	0.8437
	4	✓	✓			0.0486	0.8221	0.8046	0.8762	0.8362
	5	✓	✓	✓4		0.0470	0.8190	0.8017	0.8762	0.8451
	6	✓	✓	✓4	✓	**0.0450**	**0.8337**	**0.8178**	**0.8911**	**0.8496**
	7	✓	✓	✓1	✓	0.0472	0.8156	0.8027	0.8713	0.8339
DUTS-TE	1		✓			0.0308	0.9030	0.8816	0.9288	0.8991
	2		✓	✓4		0.0305	0.8978	0.8755	0.9247	0.8945
	3		✓	✓4	✓	0.0295	0.9097	0.8871	0.9339	0.9007
	4	✓	✓			0.0301	0.9049	0.8797	0.9285	0.8935
	5	✓	✓	✓4		0.0295	0.9053	0.8856	0.9319	0.8993
	6	✓	✓	✓4	✓	**0.0281**	**0.9123**	**0.8930**	**0.9368**	**0.9049**
	7	✓	✓	✓1	✓	0.0295	0.9060	0.8887	0.9302	0.8980
HKU-IS	1		✓			0.0237	0.9453	0.9269	0.9639	0.9266
	2		✓	✓4		0.0259	0.9394	0.9209	0.9584	0.9199
	3		✓	✓4	✓	0.0252	0.9450	0.9269	0.9621	0.9220
	4	✓	✓			0.0283	0.9411	0.9206	0.9555	0.9140
	5	✓	✓	✓4		0.0250	0.9438	0.9271	0.9605	0.9226
	6	✓	✓	✓4	✓	**0.0234**	**0.9483**	**0.9319**	**0.9657**	**0.9274**
	7	✓	✓	✓1	✓	0.0255	0.9434	0.9247	0.9622	0.9223
PASCAL-S	1		✓			0.0572	0.8794	0.8608	0.9026	0.8663
	2		✓	✓4		0.0552	0.8836	0.8639	0.9054	0.8681
	3		✓	✓4	✓	0.0565	0.8839	0.8653	0.9100	0.8670
	4	✓	✓			0.0606	0.8831	0.8619	0.8985	0.8587
	5	✓	✓	✓4		0.0557	0.8865	0.8674	0.9055	0.8652
	6	✓	✓	✓4	✓	**0.0535**	**0.8896**	**0.8701**	**0.9132**	**0.8721**
	7	✓	✓	✓1	✓	0.0576	0.8845	0.8652	0.9024	0.8678
ECSSD	1		✓			0.0308	0.9524	0.9374	0.9542	0.9252
	2		✓	✓4		0.0315	0.9494	0.9343	0.9526	0.9243
	3		✓	✓4	✓	**0.0297**	0.9536	0.9384	0.9554	**0.9290**
	4	✓	✓			0.0344	0.9477	0.9310	0.9484	0.9193
	5	✓	✓	✓4		0.0305	0.9537	0.9393	0.9544	0.9267
	6	✓	✓	✓4	✓	0.0307	**0.9549**	**0.9422**	0.9545	0.9279
	7	✓	✓	✓1	✓	0.0326	0.9514	0.9247	**0.9622**	0.9223
SOD	1		✓			0.0910	0.8772	**0.8707**	0.8307	0.8024
	2		✓	✓4		0.0910	0.8595	0.8543	0.8137	0.7987
	3		✓	✓4	✓	**0.0841**	0.8648	0.8595	0.8250	**0.8089**
	4	✓	✓			0.0947	0.8725	0.8650	0.8150	0.7949
	5	✓	✓	✓4		0.0886	0.8667	0.8608	0.8133	0.8019
	6	✓	✓	✓4	✓	0.0874	**0.8780**	0.8684	**0.8381**	**0.8089**
	7	✓	✓	✓1	✓	0.0944	0.8729	0.8675	0.8171	0.7994

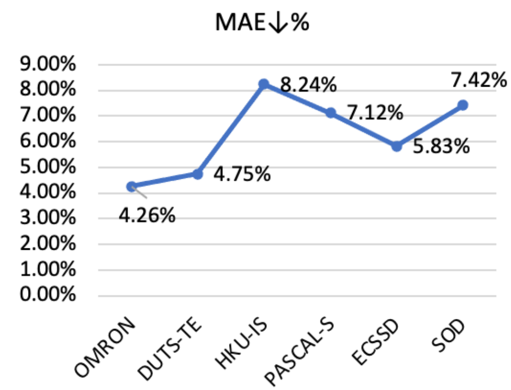


Figure 8. MAE reduction by adding region-level loss.

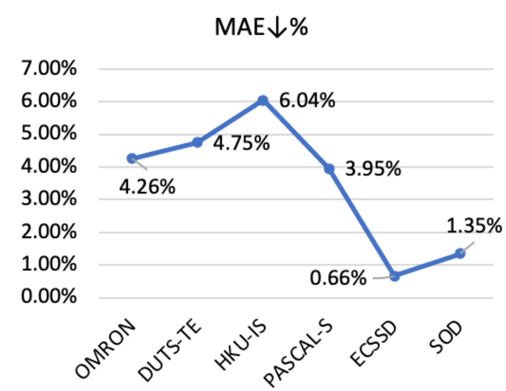


Figure 9. MAE reduction by adding object-level loss.

Experiments

Quantitative Comparisons

Table 3. Quantitative comparisons

			OMRON	(5,168 images)				DUTS-TE	(5,019 images)				SOD	(300 images)			
Year	Method	Backbone	$MAE\downarrow$	$MaxF\uparrow$	$mF\uparrow$	$mE_m\uparrow$	$S_m\uparrow$	$MAE\downarrow$	$MaxF\uparrow$	$mF\uparrow$	$mE_m\uparrow$	$S_m\uparrow$	$MAE\downarrow$	$MaxF\uparrow$	$mF\uparrow$	$mE_m\uparrow$	$S_m\uparrow$
2019	MLMSNet	VGG-16	0.0635	0.7740	0.7455	0.8387	0.8093	0.0484	0.8511	0.8137	0.8631	0.8618	0.1060	0.8517	0.8291	0.8019	0.7898
2019	AFNet	VGG-16	0.0573	0.7972	0.7766	0.8595	0.8263	0.0453	0.8623	0.8340	0.8929	0.8672	-	-	-	-	-
2019	EGNet	VGG-16	0.0564	0.8087	0.7855	0.8642	0.8357	0.0431	0.8764	0.8472	0.8927	0.8786	0.1100	0.8589	0.8426	0.8209	0.7882
2019	EGNet	ResNet-50	0.0528	0.8155	0.7942	0.8738	0.8412	0.0386	0.8880	0.8597	0.9040	0.8873	0.0969	0.8778	0.8610	0.8422	0.8067
2019	CPD	VGG-16	0.0567	0.7935	0.7800	0.8685	0.8178	0.0425	0.8638	0.8458	0.9038	0.8669	0.1125	0.8480	0.8365	0.8124	0.7715
2019	CPD	ResNet-50	0.0560	0.7966	0.7807	0.8726	0.8248	0.0429	0.8649	0.8431	0.9009	0.8691	0.1097	0.8568	0.8376	0.8174	0.7711
2019	BASNet	ResNet-34	0.0565	0.8053	0.7906	0.8691	0.8362	0.0472	0.8589	0.8416	0.8790	0.8660	0.1124	0.8487	0.8368	0.7793	0.7721
2020	AADFNet	ResNet-50	0.0488	0.8143	0.8050	0.8744	0.8389	0.0314	0.8993	0.8911	0.9225	0.8914	0.0903	0.8677	0.8579	0.8051	0.7929
2020	GateNet	VGG-16	0.0613	0.7940	0.7691	0.8534	0.8209	0.0448	0.8695	0.8388	0.8856	0.8705	-	-	-	-	-
2020	GateNet	ResNet-50	0.0552	0.8180	0.7914	0.8682	0.8382	0.0399	0.8870	0.8552	0.9004	0.8852	-	-	-	-	-
2020	GateNet	ResNet-101	0.0547	0.8210	0.7944	0.8736	0.8449	0.0380	0.8919	0.8615	0.9075	0.8910	-	-	-	-	-
2020	U2Net	RSU	0.0544	0.8226	0.8023	0.8716	0.8467	0.0443	0.8719	0.8479	0.8840	0.8738	0.1061	0.8588	0.8428	0.7993	0.7891
2020	MINet	VGG-16	0.0572	0.7936	0.7755	0.8644	0.8218	0.0395	0.8761	0.8550	0.9070	0.8753	-	-	-	-	-
2020	MINet	ResNet-50	0.0559	0.8098	0.7911	0.8734	0.8329	0.0373	0.8833	0.8597	0.9132	0.8842	-	-	-	-	-
2020	LDF	ResNet-50	0.0517	0.8199	0.8015	0.8814	0.8392	0.0336	0.8968	0.8779	0.9232	0.8924	-	-	-	-	-
2021	SAC	ResNet-101	0.0523	0.8287	0.8092	0.8833	0.8487	0.0339	0.8944	0.8732	0.9208	0.8957	0.0934	**0.8804**	0.8695	0.8482	0.8087
2021	CANet	CNN	0.0581	0.8101	0.7796	0.8593	0.8356	0.0437	0.8755	0.8382	0.8896	0.8781	0.0992	0.8650	0.8406	0.8331	0.8007
2021	SGL-KRN	ResNet-50	0.0492	0.7961	0.7830	0.8783	0.8464	0.0337	0.8833	0.8649	0.9311	0.8929	-	-	-	-	-
2021	PA-KRN	ResNet-50	0.0496	0.8101	0.7956	0.8880	0.8533	0.0328	0.8945	0.8761	0.9353	0.9005	-	-	-	-	-
2022	ICON	ResNet-50	0.0569	0.8254	0.8013	0.8791	0.8445	0.0370	0.8917	0.8665	0.9142	0.8889	**0.0841**	0.8790	**0.8711**	**0.8516**	**0.8238**
2022	EDN	VGG-16	0.0565	0.7818	0.7686	0.8628	0.8376	0.0410	0.8636	0.8457	0.9118	0.8829	-	-	-	-	-
2022	EDN	ResNet-50	0.0494	0.7992	0.7880	0.8774	0.8495	0.0351	0.8784	0.8634	0.9250	0.8924	-	-	-	-	-
2023	**MENet(Ours)**	ResNet-50	**0.0450**	**0.8337**	**0.8178**	**0.8911**	**0.8496**	**0.0281**	**0.9123**	**0.8930**	**0.9368**	**0.9049**	0.0874	0.8780	0.8684	0.8381	0.8089

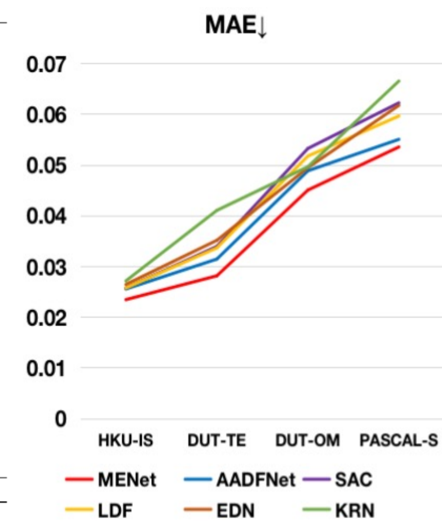
Project page：https://github.com/yiwangtz/MENet

Experiments

- MENet achieves new state-of-the-art results in models using ResNet-50,-101 or VGG-16 as the backbone.
- Inference for a testing image scaled to [352,352] takes just 0.022s (45 fps)

Table 4. Quantitative comparisons

			HKU-IS (4,447 images)					PASCAL-S (850 images)					ECSSD (1,000 images)				
Year	Method	Backbone	$MAE\downarrow$	$MaxF\uparrow$	$mF\uparrow$	$mE_m\uparrow$	$S_m\uparrow$	$MAE\downarrow$	$MaxF\uparrow$	$mF\uparrow$	$mE_m\uparrow$	$S_m\uparrow$	$MAE\downarrow$	$MaxF\uparrow$	$mF\uparrow$	$mE_m\uparrow$	$S_m\uparrow$
2019	MLMSNet	VGG-16	0.0387	0.9207	0.8891	0.9379	0.9066	0.0736	0.8552	0.8254	0.8447	0.8443	0.0446	0.9284	0.9007	0.9161	0.9112
2019	AFNet	VGG-16	0.0358	0.9226	0.8998	0.9475	0.9055	0.0700	0.8629	0.8409	0.8851	0.8494	0.0418	0.9350	0.9157	0.9414	0.9135
2019	EGNet	VGG-16	0.0345	0.9273	0.9050	0.9503	0.9100	0.0776	0.8585	0.8371	0.8714	0.8475	0.0405	0.9434	0.9232	0.9408	0.9193
2019	EGNet	Resnet-50	0.0309	0.9352	0.9122	0.9564	0.9180	0.0740	0.8653	0.8437	0.8772	0.8521	0.0374	0.9474	0.9288	0.9469	0.9246
2019	CPD	VGG-16	0.0333	0.9239	0.9075	0.9501	0.9042	0.0721	0.8612	0.8441	0.8837	0.8446	0.0402	0.9360	0.9233	0.9433	0.9103
2019	CPD	ResNet-50	0.0342	0.9250	0.9047	0.9503	0.9056	0.0706	0.8595	0.8414	0.8873	0.8484	0.0371	0.9393	0.9244	0.9494	0.9182
2019	BASNet	ResNet-34	0.0322	0.9284	0.9113	0.9458	0.9090	0.0758	0.8539	0.8344	0.8527	0.8380	0.0370	0.9425	0.9274	0.9210	0.9163
2020	AADFNet	ResNet-50	0.0255	0.9415	**0.9339**	0.9592	0.9190	0.0550	0.8797	0.8677	0.9051	0.8658	**0.0280**	0.9543	**0.9478**	0.9529	0.9299
2020	GateNet	VGG-16	0.0361	0.9287	0.9036	0.9470	0.9100	0.0684	0.8696	0.8439	0.8692	0.8574	0.0418	0.9413	0.9191	0.9314	0.9169
2020	GateNet	ResNet-50	0.0337	0.9335	0.9097	0.9534	0.9154	0.0680	0.8690	0.8459	0.8842	0.8580	0.0408	0.9454	0.9250	0.9431	0.9198
2020	GateNet	ResNet-101	0.0320	0.9375	0.9136	0.9567	0.9195	0.0668	0.8702	0.8468	0.8924	0.8622	0.0357	0.9508	0.9301	0.9501	0.9302
2020	U2Net	RSU	0.0312	0.9352	0.9133	0.9484	0.9161	0.0740	0.8592	0.8386	0.8500	0.8444	0.0330	0.9510	0.9325	0.9251	0.9276
2020	MINet	VGG-16	0.0316	0.9302	0.9133	0.9540	0.9119	0.0645	0.8650	0.8450	0.8961	0.8544	0.0370	0.9435	0.9296	0.9475	0.9192
2020	MINet	ResNet-50	0.0292	0.9349	0.9166	0.9600	0.9189	0.0643	0.8665	0.8461	0.8981	0.8563	0.0342	0.9475	0.9309	0.9532	0.9250
2020	LDF	ResNet-50	0.0275	0.9394	0.9224	0.9597	0.9196	0.0596	0.8741	0.8577	0.9048	0.8630	0.0335	0.9501	0.9379	0.9509	0.9245
2021	SAC	ResNet-101	0.0257	0.9416	0.9260	0.9636	0.9253	0.0622	0.8772	0.8585	0.9022	0.8656	0.0309	0.9512	0.9376	**0.9586**	**0.9312**
2021	CANet	CNN	0.0371	0.9297	0.8977	0.9455	0.9100	0.0728	0.8662	0.8392	0.8790	0.8552	0.0441	0.9378	0.9103	0.9362	0.9154
2021	SGL-KRN	ResNet-50	0.0280	0.9301	0.9154	0.9539	0.9206	0.0678	0.8502	0.8373	0.8941	0.8556	0.0360	0.9368	0.9241	0.9462	0.9231
2021	PA-KRN	ResNet-50	0.0271	0.9349	0.9198	0.9561	0.9235	0.0665	0.8530	0.8388	0.8964	0.8578	0.0323	0.9425	0.9301	0.9503	0.9278
2022	ICON	ResNet-50	0.0289	0.9395	0.9196	0.9585	0.9202	0.0644	0.8757	0.8514	0.8931	0.8611	0.0318	0.9503	0.9336	0.9543	0.9290
2022	EDN	VGG-16	0.0286	0.9286	0.9141	0.9504	0.9208	0.0650	0.8555	0.8406	0.8955	0.8605	0.0336	0.9408	0.9285	0.9508	0.9283
2022	EDN	ResNet-50	0.0264	0.9325	0.9196	0.9548	0.9241	0.0617	0.8600	0.8489	0.9015	0.8646	0.0320	0.9410	0.9304	0.9508	0.9267
2023	**MENet(Ours)**	ResNet-50	**0.0234**	**0.9483**	0.9319	**0.9657**	**0.9274**	**0.0535**	**0.8896**	**0.8701**	**0.9132**	**0.8721**	0.0307	**0.9549**	0.9422	0.9544	0.9279



Experiments

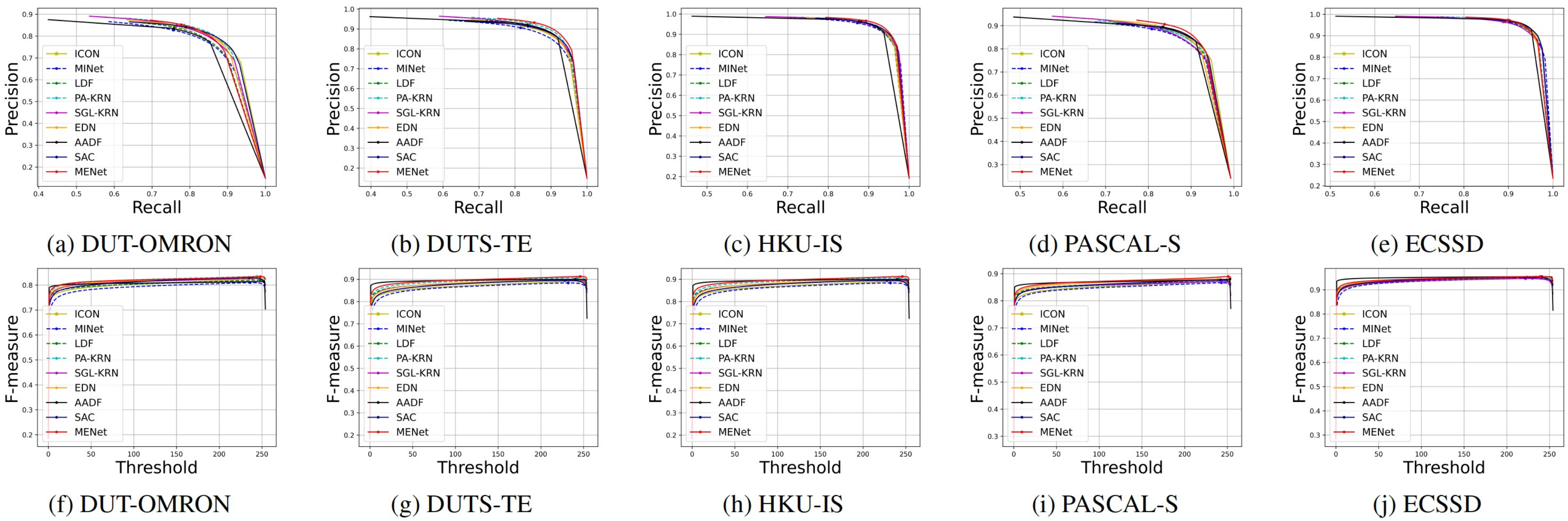


Figure 10. PR-curves (Row 1) and Fm-curves (Row 2) for SOD methods.

Experiments

Qualitative Comparisons

The boundaries of the targets are more precise and complete.

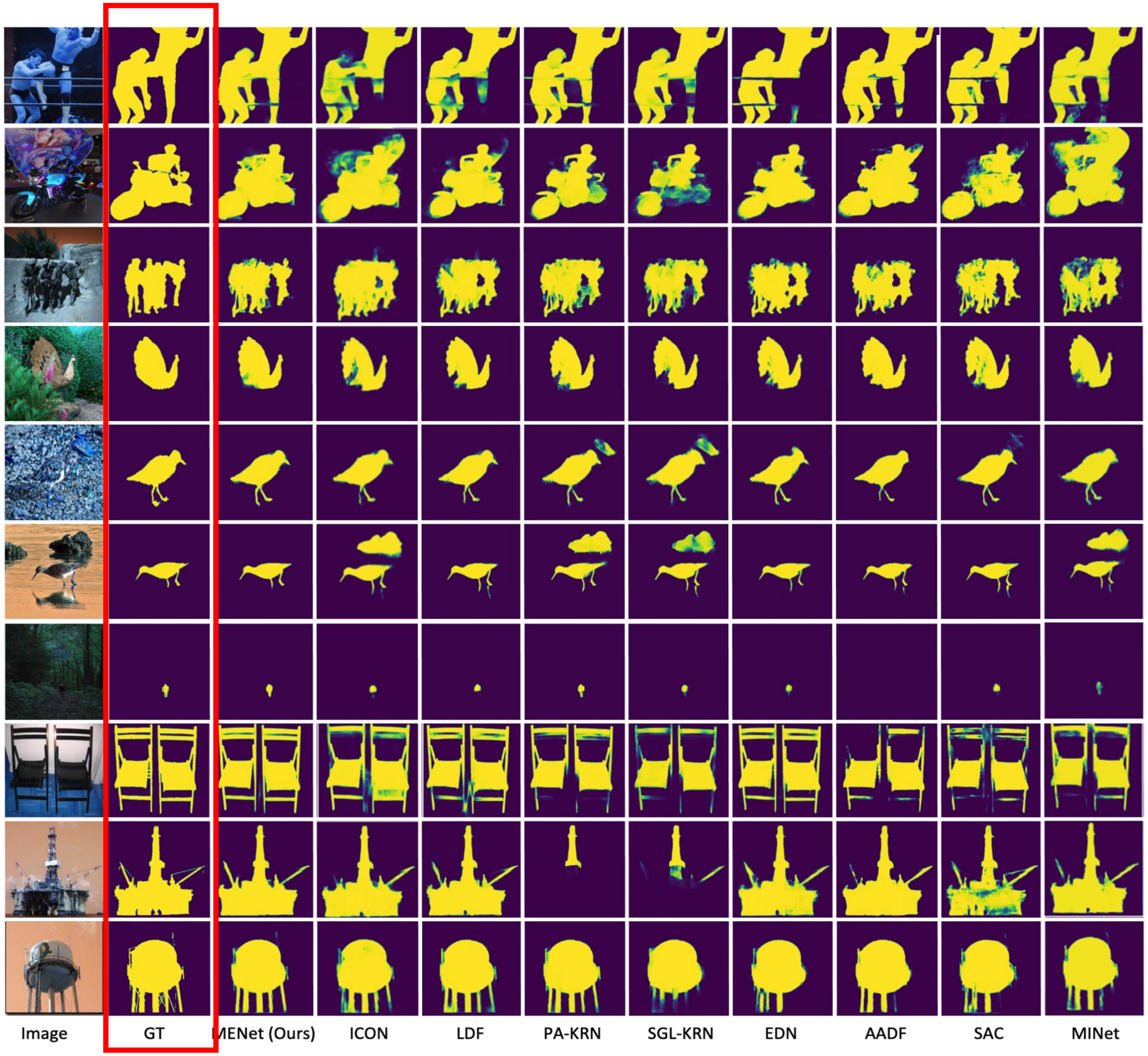


Figure 11. Qualitative comparisons with recent state-of-the-art methods.

Conclusion

- Fully leverage the human visual system (HVS) and cognition mechanisms
- A new hybrid loss that measures pixel-, region- and object-level similarities
- A multi-enhancement network with multiscale feature aggregation
- State-of-the-art results in models using ResNet-50,-101 or VGG-16 as the backbone.

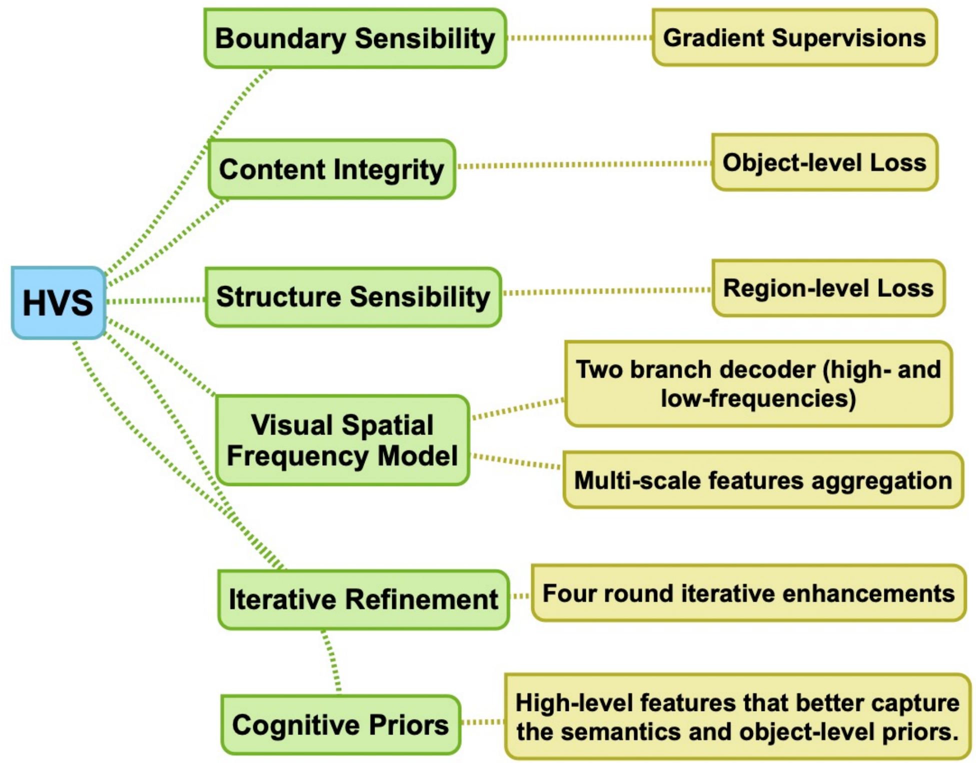


Figure 12. HVS mechanisms used in the proposed MENet.

Thank you!