Frequency-Spatial Entanglement Learning for Camouflaged Object Detection Supplementary Materials

Yanguang Sun[1], Chunyan Xu[1], Jian Yang[1], Hanyu Xuan[2] ✉, and Lei Luo[1] ✉

[1] PCA Lab, Nanjing University of Science and Technology, Nanjing, China
[2] School of Big Data and Statistics, Anhui University, Heifei, China
Sunyg@njust.edu.cn

We will provide more details and experiments that cannot be expanded in the main text to further complement the proposed FSEL method.

In this section, we first give more comparisons of subclasses in the COD10K [7] dataset. Then, we provide an extended application on salient object detection (SOD) and polyp segmentation (PS) to demonstrate the generalization ability of the proposed FSEL model.

1 More Comparisons

1.1 Evaluation of subclasses

Four subclasses. COD10K [7] dataset can be divided into four subclasses based on the species' habits, namely Aquatic, Terrestrial, Flying, and Amphibian. In Table 1, we provide an overall quantitative comparison of the proposed FSEL method and some existing COD models. From Table 1, it can be observed that regardless of categories, the proposed FSEL method demonstrates strong competitiveness. In particular, our FSEL model with the transformer backbone has improved 27.59%, 17.39%, 33.33%, and 35.29% under the four subclasses of $\mathcal{M}$ metrics compared to the recent FSPNet [13] method. Similarly, the proposed FSEL model in other frameworks has achieved excellent performance.

Sixty-six subclasses. According to the category of camouflaged objects, COD10K [7] dataset can also be divided into sixty-six subclasses, including Ant, Bat, BatFish, Bee, Bird, Bittern, Bug, Butterfly, Cat, Caterpillar, Centipede, Chameleon, Cheetah, Cicada, ClownFish, Crab, Crocodile, CrocodileFish, Deer, Dog, Dragonfly, Duck, Fish, Flounder, Frog, FrogFish, Frogmouth, Gecko, GhostPipefish, Giraffe, Grasshopper, Grouse, Heron, Human, Kangaroo, Katydid, LeafySeaDragon, Leopard, Lion, Lizard, Mantis, Mockingbird, Monkey, Moth, Octopus, Owl, Pagurian, Pipefish, Rabbit, Reccoon, Sciuridae, ScorpionFish, SeaHorse, Sheep, Shrimp, Slug, Snake, Spider, StarFish, StickInsect, Stingaree, Tiger, Toad, Turtle, Wolf, Worm. As shown in Table 2, we give the quantitative results of sixty-six subclasses under the S_m metric, and it can be seen that our FSEL exceeds existing COD methods (*e.g.*, FEDER [11], EVP [24], and FSPNet [13]) in the overwhelming majority of categories.

✉ Corresponding Author

Methods	Aquatic (474 images)						Terrestrial (699 images)						Flying (714 images)						Amphibian(124 images)					
	$\mathcal{M}\downarrow$	$F_\varphi^m\uparrow$	$F_\varphi^a\uparrow$	$F_\varphi^w\uparrow$	$S_m\uparrow$	$E_m\uparrow$	$\mathcal{M}\downarrow$	$F_\varphi^m\uparrow$	$F_\varphi^a\uparrow$	$F_\varphi^w\uparrow$	$S_m\uparrow$	$E_m\uparrow$	$\mathcal{M}\downarrow$	$F_\varphi^m\uparrow$	$F_\varphi^a\uparrow$	$F_\varphi^w\uparrow$	$S_m\uparrow$	$E_m\uparrow$	$\mathcal{M}\downarrow$	$F_\varphi^m\uparrow$	$F_\varphi^a\uparrow$	$F_\varphi^w\uparrow$	$S_m\uparrow$	$E_m\uparrow$
SINet [7]	0.065	0.755	0.670	0.644	0.778	0.836	0.046	0.665	0.577	0.565	0.754	0.810	0.040	0.737	0.614	0.580	0.797	0.817	0.042	0.783	0.684	0.654	0.827	0.847
UGTR [45]	0.050	0.789	0.707	0.687	0.808	0.860	0.036	0.721	0.609	0.606	0.786	0.820	0.026	0.805	0.697	0.700	0.841	0.870	0.029	0.813	0.749	0.738	0.857	0.883
JSOCOD [17]	0.049	0.786	0.734	0.705	0.803	0.877	0.037	0.708	0.647	0.624	0.776	0.863	0.026	0.794	0.733	0.719	0.834	0.903	0.031	0.792	0.765	0.742	0.841	0.903
MGL-S [47]	0.053	0.784	0.698	0.674	0.801	0.858	0.037	0.711	0.606	0.593	0.778	0.823	0.027	0.797	0.693	0.690	0.835	0.871	0.029	0.819	0.754	0.734	0.855	0.886
PFNet [28]	0.055	0.768	0.706	0.675	0.791	0.865	0.041	0.697	0.622	0.606	0.770	0.850	0.030	0.774	0.698	0.691	0.822	0.887	0.031	0.813	0.753	0.740	0.848	0.896
LSR [26]	0.053	0.781	0.727	0.694	0.801	0.875	0.038	0.704	0.639	0.611	0.770	0.858	0.027	0.783	0.726	0.707	0.828	0.907	0.030	0.805	0.774	0.751	0.845	0.922
PreyNet [48]	0.049	0.801	0.757	0.718	0.810	**0.892**	0.036	0.715	0.664	0.628	0.774	0.867	0.024	0.811	0.770	0.740	0.840	0.918	0.028	0.830	0.797	0.768	0.856	0.927
SegMaR [15]	0.048	0.787	0.724	0.704	0.812	0.878	0.037	0.713	0.642	0.623	0.780	0.861	0.027	0.792	0.725	0.713	0.834	0.899	0.031	0.823	0.777	0.755	0.854	0.913
FEDER [11]	0.046	0.808	**0.767**	0.733	0.817	**0.892**	**0.033**	0.729	0.679	0.651	0.785	**0.880**	0.022	0.827	0.775	0.759	0.850	**0.925**	**0.027**	0.824	0.794	0.774	0.857	**0.931**
Ours-R50	**0.045**	**0.812**	0.762	**0.735**	**0.819**	**0.892**	**0.033**	**0.751**	**0.686**	**0.671**	**0.801**	0.876	**0.021**	**0.838**	**0.778**	**0.771**	**0.859**	0.924	**0.027**	**0.844**	**0.805**	**0.778**	**0.867**	0.923
ZoomNet [30]	0.043	**0.820**	0.761	0.736	**0.827**	0.887	**0.029**	0.763	0.687	0.675	0.806	0.874	**0.020**	0.845	0.773	0.767	0.862	0.913	0.025	0.845	0.797	0.785	0.874	0.918
Ours-R50†	**0.042**	**0.820**	**0.772**	**0.743**	0.824	**0.898**	0.030	**0.769**	**0.701**	**0.692**	**0.814**	**0.882**	**0.020**	**0.848**	**0.786**	**0.783**	**0.867**	**0.925**	**0.024**	**0.853**	**0.816**	**0.803**	**0.879**	**0.932**
C2FNet [36]	0.052	0.784	0.727	0.701	0.805	0.879	0.037	0.706	0.644	0.627	0.781	0.864	0.026	0.802	0.734	0.724	0.838	0.908	0.030	0.809	0.771	0.752	0.849	0.919
FAPNet [53]	0.049	0.798	0.740	0.717	0.817	0.879	0.037	0.732	0.652	0.639	0.792	0.852	0.025	0.814	0.733	0.725	0.843	0.896	0.032	0.821	0.768	0.752	0.854	0.896
BSANet [55]	0.050	0.786	0.738	0.706	0.807	0.886	0.035	0.722	0.668	0.643	0.787	**0.875**	0.024	0.813	0.759	0.738	0.841	0.913	0.027	0.825	0.786	0.769	0.858	**0.931**
SINetv2 [6]	0.051	0.786	0.712	0.696	0.809	0.864	0.039	0.717	0.622	0.623	0.784	0.835	0.027	0.802	0.710	0.713	0.838	0.888	0.030	0.831	0.761	0.757	0.858	0.893
Ours-R2N	**0.042**	**0.817**	**0.763**	**0.745**	**0.829**	**0.901**	**0.032**	**0.748**	**0.677**	**0.671**	**0.807**	0.869	**0.020**	**0.840**	**0.775**	**0.773**	**0.864**	**0.925**	**0.024**	**0.852**	**0.808**	**0.797**	**0.878**	**0.931**
VST [23]	0.050	0.793	0.742	0.714	0.817	0.880	0.041	0.725	0.656	0.629	0.780	0.855	0.025	0.813	0.757	0.739	0.845	0.907	0.032	0.845	0.807	0.789	0.870	0.922
FSPNet [13]	0.037	0.842	0.764	0.763	0.847	0.905	0.027	0.768	0.688	0.679	0.819	0.880	0.020	0.839	0.759	0.761	0.868	0.913	0.023	0.846	0.786	0.782	0.878	0.927
EVP [24]	0.047	0.813	0.735	0.724	0.828	0.875	0.033	0.749	0.645	0.654	0.804	0.849	0.023	0.836	0.738	0.751	0.860	0.903	0.026	0.865	0.785	0.790	0.882	0.917
FPNet [5]	0.042	0.833	0.788	0.767	0.840	0.898	0.029	0.779	0.720	0.711	0.825	0.926	0.020	0.844	0.791	0.786	0.870	0.926	0.022	0.823	0.800	0.783	0.871	0.942
Ours-Pvt	**0.029**	**0.866**	**0.821**	**0.814**	**0.871**	**0.928**	**0.023**	**0.815**	**0.748**	**0.755**	**0.850**	**0.908**	**0.015**	**0.875**	**0.817**	**0.825**	**0.892**	**0.943**	**0.017**	**0.900**	**0.859**	**0.854**	**0.907**	**0.961**

Table 1: Quantitative results on four subclasses from COD10K [7] dataset. The best result is shown in **blod**. Note that "Ours-R50", "Ours-R2N", and "Ours-Pvt" present ResNet50 [12]/Res2Net [10]/PVTv2 [39] as backbone. "Ours-R50†" denotes using the same input strategy as ZoomNet [30].

These above results again demonstrate that the proposed FSEL model, through the joint optimization of components entanglement transformer block (ETB), joint domain perception module (JDPM), and dual-domain reverse parser (DRP) in the frequency and spatial domains, is conducive to generating discriminative features for better inferring camouflaged objects.

2 Extended applications on SOD and PS

Salient object detection (SOD) is opposite to camouflage object detection (COD) in image understanding, it aims to extract the most attractive objects or regions in a visual scene, which serves as a pre-processing and has been widely applied. Polyp segmentation (PS) is similar to COD because intestinal polyps have similar features to the surrounding environment. The application of polyp segmentation helps medical professionals to better analyze the condition of patients. To show great generalization and further demonstrate the rationality of our FSEL model, we evaluate the proposed FSEL method on SOD and PS tasks.

2.1 Experimental details

For model training, we only change the dataset without adjusting the model's hyper-parameters to ensure the effectiveness of evaluating the generalization ability of the proposed FSEL model in SOD and PS tasks.

SOD Datasets. We evaluate the proposed FSEL method on four widely-used SOD datasets, including ECSSD [44] (1,000 images), PASCAL-S [20] (850 images), HKU-IS [18] (4,447 images) and DUTS-TE [38] (5,019 images). We

Sub-classes	SINet	PFNet	LSR	JSOCOD	TINet	SegMaR	FEDER$_{23}$	Ours-R50	C2FNet	SINetv2	FAPNet	BSANet	Ours-R2	VST	FPNet	EVP	FSPNet	Ours-Pvt
Ant	0.633	0.655	0.695	0.636	0.677	**0.694**	0.661	0.685	0.653	0.663	0.670	**0.679**	0.654	0.665	0.670	0.675	0.726	**0.732**
Bat	0.769	0.838	0.782	0.795	0.819	0.813	**0.845**	**0.860**	0.816	0.847	0.826	0.847	**0.875**	0.855	0.889	0.893	0.875	**0.930**
BatFish	0.836	0.867	0.809	0.817	0.856	**0.878**	0.829	0.850	0.834	0.873	0.815	**0.883**	0.852	0.867	0.889	0.902	**0.907**	**0.907**
Bee	0.766	0.765	**0.811**	0.776	0.782	**0.811**	0.779	0.803	0.798	0.804	0.778	0.775	**0.822**	0.800	**0.856**	0.805	0.718	0.848
Bird	0.806	0.825	0.828	0.834	0.811	0.831	0.847	**0.861**	0.843	0.835	0.839	0.844	**0.871**	0.833	0.862	0.863	0.869	**0.894**
Bittern	0.844	0.855	0.860	0.865	0.863	0.874	0.876	**0.889**	0.840	0.849	0.870	0.847	**0.871**	0.873	0.888	0.869	0.863	**0.890**
Bug	0.844	0.839	0.828	0.842	0.825	0.846	0.881	**0.886**	0.847	0.856	0.868	0.860	**0.886**	0.825	0.890	0.871	0.873	**0.909**
Butterfly	0.828	0.861	0.870	0.867	0.857	0.849	0.898	**0.910**	0.885	0.883	0.880	**0.897**	0.893	0.892	0.897	0.892	0.884	**0.912**
Cat	0.711	0.751	0.745	0.742	0.740	0.750	0.775	**0.788**	0.767	0.771	0.770	0.774	**0.800**	0.760	0.803	0.804	0.825	**0.853**
Caterpillar	0.702	0.742	0.754	**0.783**	0.743	0.773	0.710	0.752	0.741	0.772	**0.778**	0.745	0.757	0.732	0.830	0.776	0.807	**0.849**
Centipede	0.677	0.682	0.744	0.645	0.678	0.599	**0.727**	0.708	0.731	0.762	**0.778**	0.690	0.674	0.739	0.804	0.760	0.791	**0.837**
Chameleon	0.759	0.775	0.786	0.813	0.805	**0.832**	0.822	0.828	0.809	0.803	0.827	0.795	**0.845**	0.828	0.840	0.827	0.842	**0.879**
Cheetah	0.785	0.793	0.812	0.830	0.814	0.821	**0.825**	0.815	0.830	0.825	0.801	0.832	**0.843**	0.811	0.857	0.831	0.849	**0.861**
Cicada	0.849	0.883	0.871	0.883	0.866	0.872	**0.888**	**0.893**	0.871	0.881	0.878	0.892	**0.905**	0.876	0.901	0.904	0.906	**0.930**
ClownFish	0.693	0.805	0.784	0.709	0.694	0.751	0.846	**0.850**	0.737	0.787	**0.804**	0.744	0.769	0.752	0.873	0.844	0.850	**0.859**
Crab	0.802	0.802	0.816	0.808	0.806	0.819	0.832	**0.840**	0.826	0.813	**0.839**	0.822	0.837	0.834	0.859	0.842	0.859	**0.863**
Crocodile	0.761	0.753	0.783	0.808	0.801	0.846	**0.861**	0.855	0.817	**0.825**	0.806	0.804	**0.825**	0.833	0.781	0.840	0.857	**0.863**
CrocodileFish	0.734	0.780	0.791	0.751	0.604	0.713	**0.797**	0.790	0.764	0.747	0.754	**0.804**	0.783	0.861	0.860	0.867	0.846	**0.889**
Deer	0.700	0.730	0.735	0.754	0.742	0.744	0.737	**0.759**	0.753	0.755	0.759	0.751	**0.777**	0.736	0.795	0.780	0.789	**0.815**
Dog	0.660	0.664	0.663	0.663	0.663	0.667	0.678	**0.694**	0.668	0.683	0.687	0.690	**0.705**	0.677	0.705	0.720	**0.745**	0.766
Dragonfly	0.772	0.818	0.820	0.824	0.800	0.833	0.810	**0.839**	0.822	0.835	0.809	0.824	**0.839**	0.830	0.837	0.866	0.877	**0.880**
Duck	0.690	**0.793**	0.701	0.704	0.689	0.739	0.765	0.780	0.742	0.727	0.704	**0.729**	0.727	0.758	0.758	0.740	0.772	**0.799**
Fish	0.766	0.798	0.815	0.819	0.768	**0.838**	0.820	0.831	0.816	0.833	0.816	0.823	**0.852**	0.825	0.860	0.853	0.853	**0.882**
Flounder	0.786	0.850	0.857	0.873	0.847	**0.888**	0.877	0.879	0.857	0.890	0.880	0.879	**0.900**	0.868	0.856	0.879	**0.922**	0.918
Frog	0.785	0.815	0.814	0.794	0.789	0.801	0.828	**0.831**	0.820	0.837	0.824	0.833	**0.850**	0.832	0.846	0.869	0.852	**0.897**
FrogFish	0.748	0.840	0.848	0.894	0.862	**0.927**	0.891	0.901	0.868	0.894	**0.924**	0.909	0.878	0.876	0.921	0.855	0.925	**0.939**
Frogmouth	0.896	0.945	0.941	0.932	0.926	0.947	**0.951**	0.950	0.936	0.941	0.935	0.945	**0.958**	0.953	0.937	0.941	0.947	**0.955**
Gecko	0.825	0.830	0.861	**0.868**	0.857	0.858	0.822	0.828	0.867	0.848	**0.871**	0.847	0.865	0.831	0.909	0.886	0.908	**0.929**
GhostPipefish	0.779	0.819	0.821	0.823	0.808	0.823	**0.832**	**0.832**	0.831	0.817	0.826	0.813	**0.830**	0.811	0.870	0.856	0.872	**0.883**
Giraffe	0.773	0.821	0.809	0.788	0.763	0.810	0.794	**0.850**	0.779	0.784	0.770	0.833	**0.866**	0.817	0.863	0.808	0.846	**0.881**
Grasshopper	0.801	0.808	0.821	0.830	0.814	0.831	0.846	**0.850**	0.829	0.831	0.843	0.835	**0.851**	0.838	0.867	0.853	0.872	**0.892**
Grouse	0.927	**0.937**	0.917	0.933	0.893	0.921	0.924	0.936	0.924	0.921	0.929	0.931	**0.958**	0.950	0.958	0.936	0.942	0.956
Heron	0.775	0.820	0.831	0.830	0.777	0.814	**0.837**	0.835	0.803	0.814	0.815	0.827	**0.854**	0.844	0.852	0.834	0.846	**0.871**
Human	0.740	0.780	0.787	0.766	**0.796**	0.794	0.766	0.776	0.789	0.815	0.776	0.799	**0.819**	0.778	0.824	0.784	0.796	**0.842**
Kangaroo	0.736	0.761	0.746	0.786	0.767	0.792	0.766	**0.832**	0.805	0.814	**0.852**	0.804	0.840	0.852	**0.886**	0.815	0.792	0.864
Katydid	0.771	0.797	0.797	0.811	0.793	0.809	0.828	**0.838**	0.823	0.809	0.831	0.824	**0.847**	0.825	0.852	0.835	0.851	**0.870**
LeafySeaDragon	0.576	0.625	0.654	0.641	0.654	**0.674**	0.628	0.641	0.633	0.670	0.640	0.637	**0.750**	0.659	0.719	0.672	**0.782**	0.742
Leopard	0.804	**0.833**	0.794	0.799	0.821	0.814	0.824	**0.843**	**0.830**	0.823	0.816	0.828	0.829	0.813	0.862	0.844	0.845	**0.874**
Lion	0.758	0.805	0.817	0.807	**0.826**	0.820	0.814	0.788	0.812	0.803	**0.818**	0.767	0.816	0.799	**0.858**	0.824	0.829	0.816
Lizard	0.798	0.817	0.818	0.827	0.811	0.827	0.831	**0.850**	0.821	0.826	0.835	0.834	**0.843**	0.825	0.865	0.847	0.849	**0.877**
Mantis	0.794	0.754	0.764	0.779	0.752	0.775	0.793	**0.804**	0.766	0.774	0.767	0.765	**0.809**	0.779	0.828	0.797	0.832	**0.843**
Mockingbird	0.720	0.783	0.814	0.812	0.784	0.796	**0.845**	0.841	0.819	0.818	0.835	0.794	**0.838**	0.807	0.854	0.846	0.859	**0.883**
Monkey	0.675	0.720	0.808	0.877	0.878	0.876	0.874	**0.895**	0.835	**0.888**	0.882	0.849	0.884	0.881	0.861	0.873	**0.913**	0.903
Moth	0.863	0.893	0.894	0.908	0.866	0.902	**0.921**	0.918	0.901	0.917	0.909	0.912	**0.927**	0.928	0.933	0.925	0.941	**0.954**
Octopus	0.843	0.863	**0.873**	0.865	0.872	0.867	**0.873**	0.866	0.869	0.888	**0.914**	0.892	0.894	0.893	0.903	0.900	0.885	**0.909**
Owl	0.827	0.846	0.859	0.851	0.852	0.874	0.868	**0.886**	0.867	0.867	0.875	0.859	**0.892**	0.880	0.889	0.887	0.875	**0.908**
Pagurian	0.643	0.640	0.682	0.678	0.675	0.661	0.709	**0.742**	0.650	**0.698**	0.683	0.620	0.673	0.677	0.692	0.686	0.710	**0.756**
Pipefish	0.725	0.769	0.777	0.773	0.753	0.783	**0.794**	0.789	0.785	0.776	0.799	0.784	**0.809**	0.799	0.822	0.788	0.817	**0.850**
Rabbit	0.801	0.834	0.829	0.832	0.819	0.828	0.837	**0.847**	0.834	0.834	**0.855**	0.832	0.853	0.834	0.882	0.872	0.879	**0.893**
Reccoon	0.737	0.769	0.744	0.775	0.772	0.749	0.735	**0.840**	0.787	0.762	0.789	0.803	**0.808**	0.780	0.823	0.809	0.789	**0.837**
Sciuridae	0.821	0.830	0.794	0.836	0.821	0.830	0.855	**0.874**	0.851	0.839	0.865	0.853	**0.882**	0.875	0.880	0.894	0.850	**0.906**
ScorpionFish	0.740	0.773	0.799	0.806	0.787	0.794	0.807	**0.821**	0.778	0.808	0.819	0.801	**0.822**	0.830	0.860	0.839	0.851	**0.888**
SeaHorse	0.798	0.797	0.823	0.825	0.801	**0.828**	0.826	0.819	0.824	0.822	0.811	0.806	**0.836**	0.811	0.832	0.835	0.850	**0.868**
Sheep	0.490	0.582	**0.750**	0.561	0.518	0.748	0.489	0.569	0.686	0.500	0.482	**0.754**	0.750	0.544	**0.748**	0.495	0.493	0.476
Shrimp	0.718	0.737	0.700	0.714	0.689	0.732	0.724	**0.756**	0.741	0.735	**0.765**	0.753	**0.782**	0.757	0.780	0.795	0.818	**0.831**
Slug	0.792	0.836	0.594	0.743	0.672	0.859	**0.861**	0.837	0.803	0.729	0.727	**0.851**	0.782	0.735	0.711	**0.841**	0.696	0.739
Snake	0.796	0.824	0.816	0.819	0.796	0.821	0.844	**0.855**	0.839	0.831	0.834	0.854	**0.846**	0.839	0.834	0.834	0.853	**0.882**
Spider	0.736	0.758	0.757	0.766	0.744	0.766	0.788	**0.797**	0.769	0.770	0.789	0.772	**0.799**	0.780	0.820	0.790	0.807	**0.845**
StarFish	0.799	0.852	0.856	0.903	0.840	0.883	0.913	**0.922**	0.892	0.890	0.870	**0.912**	**0.912**	**0.917**	0.858	0.917	0.889	**0.939**
StickInsect	0.657	0.696	0.741	0.722	0.709	0.735	0.725	**0.753**	0.728	0.695	**0.745**	0.724	0.741	0.689	0.764	0.727	0.758	**0.784**
Stingaree	0.724	0.785	0.789	0.781	0.801	0.808	**0.824**	0.816	0.791	0.815	0.827	0.801	**0.823**	0.814	0.792	0.829	0.881	**0.911**
Tiger	0.662	0.674	0.663	0.668	0.660	0.673	0.700	**0.716**	0.694	0.703	0.678	0.698	**0.733**	0.703	**0.752**	0.717	0.733	0.735
Toad	0.850	0.866	0.863	0.867	0.854	0.883	0.873	**0.887**	0.865	0.869	0.871	0.872	**0.893**	0.890	0.884	0.889	0.892	**0.912**
Turtle	0.774	0.838	0.823	0.822	0.779	**0.882**	0.801	0.840	0.785	0.760	**0.815**	0.795	0.810	0.788	0.825	0.768	**0.883**	0.871
Wolf	0.736	0.729	0.706	**0.753**	0.746	0.742	0.712	0.744	0.746	0.748	0.747	**0.777**	0.767	0.737	**0.831**	0.732	0.747	0.758
Worm	0.681	0.791	0.739	**0.804**	0.734	0.799	0.776	0.798	0.691	0.806	**0.827**	0.761	0.809	0.737	0.846	0.783	0.812	**0.855**

Table 2: Quantitative results on 66 subclasses from COD10K [7] datasets in terms of S_m. The best result is shown in **blod**.

Method	ECSSD (1000 images)						PASCAL-S (850 images)						HKU-IS (4447 images)						DUTS-TE (5019 images)					
	$\mathcal{M}\downarrow$	$F_\varphi^m\uparrow$	$F_\varphi^a\uparrow$	$F_\varphi^w\uparrow$	$S_m\uparrow$	$E_m\uparrow$	$\mathcal{M}\downarrow$	$F_\varphi^m\uparrow$	$F_\varphi^a\uparrow$	$F_\varphi^w\uparrow$	$S_m\uparrow$	$E_m\uparrow$	$\mathcal{M}\downarrow$	$F_\varphi^m\uparrow$	$F_\varphi^a\uparrow$	$F_\varphi^w\uparrow$	$S_m\uparrow$	$E_m\uparrow$	$\mathcal{M}\downarrow$	$F_\varphi^m\uparrow$	$F_\varphi^a\uparrow$	$F_\varphi^w\uparrow$	$S_m\uparrow$	$E_m\uparrow$
EGNet [50]	0.037	0.947	0.920	0.903	0.925	0.947	0.074	0.865	0.817	0.795	0.852	0.877	0.031	0.936	0.902	0.886	0.917	0.955	0.039	0.889	0.815	0.816	0.885	0.907
CPD [42]	0.037	0.939	0.917	0.898	0.918	0.950	0.071	0.860	0.820	0.794	0.848	0.887	0.033	0.927	0.895	0.879	0.908	0.952	0.043	0.866	0.805	0.795	0.868	0.904
PoolNet [22]	0.039	0.944	0.915	0.896	0.921	0.945	0.075	0.863	0.815	0.793	0.849	0.876	0.032	0.934	0.900	0.883	0.915	0.955	0.040	0.880	0.809	0.807	0.882	0.904
GateNet [51]	0.040	0.945	0.916	0.894	0.920	0.943	0.067	0.869	0.819	0.797	0.858	0.884	0.033	0.934	0.899	0.880	0.915	0.953	0.040	0.888	0.807	0.809	0.884	0.903
ITSD [52]	0.035	0.947	0.895	0.911	0.925	0.932	0.066	0.870	0.785	0.812	0.859	0.863	0.031	0.934	0.899	0.894	0.917	0.953	0.041	0.883	0.804	0.824	0.884	0.898
MINet [31]	0.034	0.948	0.924	0.911	0.925	0.953	0.064	0.867	0.829	0.809	0.856	0.898	0.029	0.935	0.909	0.897	0.918	0.960	0.037	0.884	0.828	0.825	0.883	0.917
RASNet [3]	0.036	0.944	0.916	0.905	0.920	0.941	0.068	0.862	0.820	0.804	0.853	0.883	0.030	0.932	0.906	0.892	0.914	0.956	0.039	0.880	0.825	0.822	0.879	0.913
DCN [43]	0.032	0.952	0.934	0.920	0.928	0.958	0.062	0.872	0.837	0.820	0.861	0.902	0.027	0.939	0.914	0.905	0.921	0.962	**0.035**	**0.894**	0.844	0.840	0.890	**0.923**
CSF [4]	0.033	0.951	0.925	0.911	0.930	0.954	0.069	0.874	0.823	0.807	0.862	0.884	-	-	-	-	-	-	0.038	0.890	0.823	0.823	0.889	0.914
DSPN [21]	0.039	0.945	0.910	0.891	0.924	0.942	0.067	0.874	0.819	0.801	0.864	0.883	0.035	0.932	0.893	0.873	0.915	0.951	0.043	0.883	0.791	0.794	0.881	0.892
SUCA [19]	0.036	0.939	0.915	0.906	0.919	0.948	0.067	0.861	0.818	0.803	0.849	0.886	0.031	0.928	0.897	0.890	0.912	0.955	0.044	0.859	0.803	0.802	0.868	0.903
DCENet [29]	0.035	0.947	0.926	0.913	0.921	0.952	0.061	0.880	0.845	0.825	0.862	0.903	0.029	0.933	0.908	0.898	0.915	0.957	0.038	0.884	0.842	0.834	0.881	0.922
DNA [25]	0.042	0.940	0.891	0.883	0.915	0.935	0.079	0.855	0.790	0.772	0.837	0.861	0.035	0.928	0.863	0.864	0.905	0.937	0.046	0.873	0.747	0.765	0.859	0.864
RCSBNet [16]	0.034	0.944	0.927	0.916	0.922	0.948	0.059	0.876	**0.848**	0.826	0.860	0.907	**0.027**	0.938	**0.924**	**0.909**	0.918	0.959	**0.035**	0.889	**0.856**	0.840	0.879	0.920
ICON-R [56]	0.032	0.950	0.928	0.918	0.929	0.954	0.064	0.876	0.833	0.818	0.861	0.893	0.029	0.939	0.910	0.902	0.920	0.959	0.037	0.892	0.838	0.837	0.887	0.919
Ours-R50	**0.026**	**0.955**	**0.936**	**0.928**	**0.935**	**0.963**	**0.058**	**0.882**	0.844	**0.832**	**0.867**	**0.910**	**0.027**	**0.940**	0.913	**0.909**	**0.923**	**0.962**	0.037	0.892	0.847	**0.847**	**0.892**	**0.923**
VST [23]	0.033	0.951	0.920	0.910	0.932	0.957	0.061	0.876	0.829	0.816	0.871	0.902	0.029	0.942	0.900	0.897	0.928	0.960	0.037	0.890	0.818	0.828	0.895	0.916
GLST [33]	0.025	0.961	0.935	0.930	0.942	0.965	0.052	0.897	0.855	0.846	**0.886**	0.920	0.024	0.951	0.915	0.914	0.935	0.967	0.027	0.921	0.857	0.873	0.918	0.940
BBRF [27]	0.022	0.962	0.950	0.944	0.939	0.972	0.049	0.891	**0.869**	0.856	0.877	**0.925**	**0.020**	0.951	**0.937**	0.932	0.932	0.972	0.025	0.916	0.893	0.887	0.907	**0.953**
ICON-P [56]	0.024	0.959	0.936	0.933	0.940	0.962	0.051	0.893	0.854	0.847	0.882	0.915	0.022	0.952	0.925	0.924	0.934	0.970	0.026	0.923	0.868	0.882	0.916	0.942
Ours-Pvt	**0.020**	**0.967**	**0.953**	**0.949**	**0.949**	**0.973**	**0.048**	**0.898**	**0.869**	**0.860**	0.885	0.924	**0.020**	**0.954**	0.936	**0.934**	**0.938**	**0.974**	**0.024**	**0.925**	**0.894**	**0.897**	**0.921**	**0.953**

Table 3: Quantitative results on four SOD datasets, *i.e.*, ECSSD [44], PASCAL-S [20], HKU-IS [20], and DUTS-TE [38]. The best result is shown in **blod**. Note that "Ours-R50", "Ours-R2N", and "Ours-Pvt" present ResNet50 [12]/Res2Net [10]/PVTv2 [39] as backbone.

Methods	CVC-300(60 images)						Kvasir(100 images)						ETIS(196 images)						CVC-ColonDB(380 images)					
	$\mathcal{M}\downarrow$	$F_\varphi^m\uparrow$	$F_\varphi^a\uparrow$	$F_\varphi^w\uparrow$	$S_m\uparrow$	$E_m\uparrow$	$\mathcal{M}\downarrow$	$F_\varphi^m\uparrow$	$F_\varphi^a\uparrow$	$F_\varphi^w\uparrow$	$S_m\uparrow$	$E_m\uparrow$	$\mathcal{M}\downarrow$	$F_\varphi^m\uparrow$	$F_\varphi^a\uparrow$	$F_\varphi^w\uparrow$	$S_m\uparrow$	$E_m\uparrow$	$\mathcal{M}\downarrow$	$F_\varphi^m\uparrow$	$F_\varphi^a\uparrow$	$F_\varphi^w\uparrow$	$S_m\uparrow$	$E_m\uparrow$
UNet [34]	0.022	0.805	0.702	0.684	0.842	0.867	0.055	0.876	0.832	0.794	0.858	0.901	0.036	0.497	0.394	0.366	0.682	0.645	0.059	0.621	0.562	0.491	0.710	0.758
UNet++ [54]	0.018	0.817	0.706	0.687	0.838	0.884	0.048	0.904	0.853	0.808	0.862	0.907	0.035	0.554	0.465	0.390	0.681	0.704	0.061	0.623	0.555	0.467	0.692	0.759
SFA [9]	0.065	0.558	0.353	0.341	0.640	0.604	0.075	0.801	0.715	0.670	0.782	0.828	0.109	0.367	0.255	0.231	0.557	0.515	0.094	0.554	0.393	0.366	0.629	0.634
ACSNet [49]	0.013	0.907	0.782	0.825	0.922	0.916	0.032	**0.938**	0.902	0.882	0.920	0.944	0.059	0.654	0.536	0.530	0.750	0.774	0.039	0.771	0.703	0.697	0.829	0.861
PraNet [8]	0.010	0.905	0.824	0.843	0.924	0.938	0.030	0.929	0.897	0.885	0.915	0.943	0.031	0.675	0.602	0.600	0.791	0.792	0.043	0.765	0.718	0.699	0.820	0.847
SANet [40]	0.008	0.881	0.823	0.859	0.927	**0.948**	0.028	0.931	0.903	0.892	0.914	0.950	0.015	0.748	0.656	0.685	0.843	0.835	0.043	0.808	0.731	0.726	0.836	0.855
EUNet [32]	0.015	0.851	0.792	0.805	0.903	0.915	0.028	0.929	0.900	0.893	**0.916**	0.945	0.067	0.712	0.639	0.636	0.788	0.807	0.045	0.798	0.738	0.730	0.831	0.867
DCRNet [46]	0.010	0.904	0.804	0.830	0.921	0.934	0.035	0.929	0.881	0.868	0.910	0.931	0.096	0.600	0.492	0.506	0.732	0.741	0.052	0.744	0.693	0.684	0.820	0.840
Ours-R50	**0.007**	**0.917**	**0.844**	**0.863**	**0.930**	0.947	0.025	0.925	0.900	0.883	0.914	0.953	0.018	0.780	0.679	0.691	0.842	0.860	0.033	0.825	0.760	0.747	0.852	0.881
Ours-Pvt	**0.009**	0.904	0.843	0.856	**0.930**	0.945	**0.027**	0.929	**0.908**	**0.893**	0.914	**0.954**	**0.013**	**0.831**	**0.751**	**0.762**	**0.880**	**0.898**	**0.021**	**0.854**	**0.813**	**0.807**	**0.881**	**0.919**

Table 4: Quantitative results on four PS datasets, including CVC-300 [2], Kvasir [14], ETIS [35], and CVC-ColonDB [37]. The best result is shown in **blod**. Note that "Ours-R50", "Ours-R2N", and "Ours-Pvt" present ResNet50 [12]/Res2Net [10]/PVTv2 [39] as backbone.

train the proposed FSEL model on the DUTS-TR [38] dataset, which contains 10,553 training images from natural scenes.

PS Datasets. The proposed FSEL model on PS is based on the four public polyp segmentation datasets, including CVC-300 [2] (60 images), ETIS [35] (96 images), Kvasir [14] (100 images), and CVC-ColonDB [37] (380 images). We exploit the same 900 and 548 images from ClinicDB [1] and Kvasir [14] datasets to train the proposed FSEL method.

Evaluation metrics. Similar to COD, we utilize six evaluation metrics to test the performance of our FSEL model in SOD and PS tasks, that is, Mean Absolute Error ($\mathcal{M}$), Maximum F-measure (F_φ^m), Average F-measure (F_φ^a), Weighted F-measure (F_φ^w), S-measure (S_m), and E-measure (E_m).

2.2 Experimental results

SOD. We compare the proposed FSEL method with 19 state-of-the-art SOD methods, including EGNet [50], CPD [42], PoolNet [22], GateNet [51], ITSD [52],

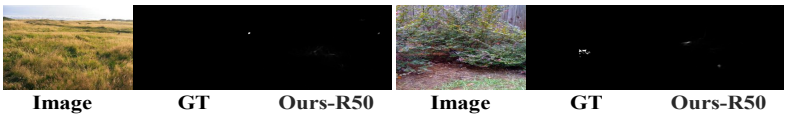

Fig. 1: Failed prediction results."GT" denotes the ground truth.

MINet [31], LDF [41], DCN [43], CSF [4], DSPN [21], SUCA [19], DCENet [29], DNA [25], RCSBNet [16], ICON-R [56], VST [23], GLST [33], BBRF [27], and ICON-P [56]. From Table 3, it can be observed that the proposed FSEL model achieves superior performance in the SOD task under different evaluation metrics, whether based on Transformer or ResNet backbone.

PS. Table 4 give the quantitative and qualitative results of the proposed FSEL method and other eight methods in the PS task, including UNet [34], UNet++ [54], SFA [9], ACSNet [49], PraNet [8], SANet [40], EUNet [32], and DCRNet [46], and it can be seen that our FSEL method with different backbones is highly competitive. These results significantly demonstrate that our FSEL model has a strong generalization ability in the more general binary segmentation task. This indicates that deeply exploring frequency and spatial domain features, and fully leveraging their global-local information, is beneficial for understanding image content and subsequently assisting in object segmentation.

3 Limitations and discussions

We present some failure cases of our FSEL in the Fig. 1. When the camouflaged object is extremely small and is occluded in complex scenes, achieving accurate detection with our method is also difficult. We assume there are several potential solutions: 1) Increasing the number of challenging images during training. 2) Adopting edge supervision and multi-stage segmentation strategies. 3) Considering the introduction of new data sources (such as infrared images). Apart from the task itself, we believe that frequency features with image-level receptive fields are immensely valuable for image understanding, however, optimizing these features is not straightforward. Moreover, frequency features generated using the Fourier Transform are complex numbers (*i.e.*, real and imaginary parts), suggesting they contain richer information. In our view, how to efficiently explore and utilize meaningful frequency features is a very promising research.

References

1. Bernal, J., Sánchez, F.J., Fernández-Esparrach, G., Gil, D., Rodríguez, C., Vilariño, F.: Wm-dova maps for accurate polyp highlighting in colonoscopy: Validation vs. saliency maps from physicians. CMIG **43**, 99–111 (2015)
2. Bernal, J., Sánchez, J., Vilarino, F.: Towards automatic polyp detection with a polyp appearance model. PR **45**(9), 3166–3182 (2012)

3. Chen, S., Tan, X., Wang, B., Lu, H., Hu, X., Fu, Y.: Reverse attention-based residual network for salient object detection. TIP **29**, 3763–3776 (2020)
4. Cheng, M.M., Gao, S.H., Borji, A., Tan, Y.Q., Lin, Z., Wang, M.: A highly efficient model to study the semantics of salient object detection. TPAMI **44**(11), 8006–8021 (2022)
5. Cong, R., Sun, M., Zhang, S., Zhou, X., Zhang, W., Zhao, Y.: Frequency perception network for camouflaged object detection. In: ACM MM (2023)
6. Fan, D.P., Ji, G.P., Cheng, M.M., Shao, L.: Concealed object detection. TPAMI **44**(10), 6024–6042 (2022)
7. Fan, D.P., Ji, G.P., Sun, G., Cheng, M.M., Shen, J., Shao, L.: Camouflaged object detection. In: CVPR. pp. 2777–2787 (2020)
8. Fan, D.P., Ji, G.P., Zhou, T., Chen, G., Fu, H., Shen, J., Shao, L.: Pranet: Parallel reverse attention network for polyp segmentation. In: MICCAI. pp. 263–273 (2020)
9. Fang, Y., Chen, C., Yuan, Y., Tong, K.y.: Selective feature aggregation network with area-boundary constraints for polyp segmentation. In: MICAAI. pp. 302–310 (2019)
10. Gao, S.H., Cheng, M.M., Zhao, K., Zhang, X.Y., Yang, M.H., Torr, P.: Res2net: A new multi-scale backbone architecture. TPAMI **43**(2), 652–662 (2019)
11. He, C., Li, K., Zhang, Y., Tang, L., Zhang, Y., Guo, Z., Li, X.: Camouflaged object detection with feature decomposition and edge reconstruction. In: CVPR. pp. 22046–22055 (2023)
12. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: CVPR. pp. 770–778 (2016)
13. Huang, Z., Dai, H., Xiang, T.Z., Wang, S., Chen, H.X., Qin, J., Xiong, H.: Feature shrinkage pyramid for camouflaged object detection with transformers. In: CVPR. pp. 5557–5566 (2023)
14. Jha, D., Smedsrud, P.H., Riegler, M.A., Halvorsen, P., de Lange, T., Johansen, D., Johansen, H.D.: Kvasir-seg: A segmented polyp dataset. In: MM. pp. 451–462 (2020)
15. Jia, Q., Yao, S., Liu, Y., Fan, X., Liu, R., Luo, Z.: Segment, magnify and reiterate: Detecting camouflaged objects the hard way. In: CVPR. pp. 4713–4722 (2022)
16. Ke, Y.Y., Tsubono, T.: Recursive contour-saliency blending network for accurate salient object detection. In: WACV. pp. 2940–2950 (2022)
17. Li, A., Zhang, J., Lv, Y., Liu, B., Zhang, T., Dai, Y.: Uncertainty-aware joint salient object and camouflaged object detection. In: CVPR. pp. 10071–10081 (2021)
18. Li, G., Yu, Y.: Visual saliency based on multiscale deep features. In: CVPR. pp. 5455–5463 (2015)
19. Li, J., Pan, Z., Liu, Q., Wang, Z.: Stacked u-shape network with channel-wise attention for salient object detection. TMM **23**, 1397–1409 (2021)
20. Li, Y., Hou, X., Koch, C., Rehg, J.M., Yuille, A.L.: The secrets of salient object segmentation. In: CVPR. pp. 280–287 (2014)
21. Liansheng, W., Rongzhen, C., Lei, Z., Haoran, X., Xiaomeng, L.: Deep sub-region network for salient object detection. TCSVT **31**(2), 728–741 (2021)
22. Liu, J.J., Hou, Q., Cheng, M.M., Feng, J., Jiang, J.: A simple pooling-based design for real-time salient object detection. In: CVPR. pp. 3917–3926 (2019)
23. Liu, N., Zhang, N., Wan, K., Shao, L., Han, J.: Visual saliency transformer. In: ICCV. pp. 4722–4732 (2021)
24. Liu, W., Shen, X., Pun, C.M., Cun, X.: Explicit visual prompting for low-level structure segmentations. In: CVPR. pp. 19434–19445 (2023)
25. Liu, Y., Cheng, M.M., Zhang, X.Y., Nie, G.Y., Wang, M.: Dna: Deeply supervised nonlinear aggregation for salient object detection. TCYB **52**(7), 6131–6142 (2021)

26. Lv, Y., Zhang, J., Dai, Y., Li, A., Liu, B., Barnes, N., Fan, D.P.: Simultaneously localize, segment and rank the camouflaged objects. In: CVPR. pp. 11591–11601 (2021)
27. Ma, M., Xia, C., Xie, C., Chen, X., Li, J.: Boosting broader receptive fields for salient object detection. TIP **32**, 1026–1038 (2023)
28. Mei, H., Ji, G.P., Wei, Z., Yang, X., Wei, X., Fan, D.P.: Camouflaged object segmentation with distraction mining. In: CVPR. pp. 8772–8781 (2021)
29. Mei, H., Liu, Y., Wei, Z., Zhou, D., Wei, X., Zhang, Q., Yang, X.: Exploring dense context for salient object detection. TCSVT **32**(3), 1378–1389 (2022)
30. Pang, Y., Zhao, X., Xiang, T.Z., Zhang, L., Lu, H.: Zoom in and out: A mixed-scale triplet network for camouflaged object detection. In: CVPR. pp. 2160–2170 (2022)
31. Pang, Y., Zhao, X., Zhang, L., Lu, H.: Multi-scale interactive network for salient object detection. In: CVPR. pp. 9413–9422 (2020)
32. Patel, K., Bur, A.M., Wang, G.: Enhanced u-net: A feature enhancement network for polyp segmentation. In: RV. pp. 181–188 (2021)
33. Ren, S., Wen, Q., Zhao, N., Han, G., He, S.: Unifying global-local representations in salient object detection with transformer. arXiv preprint arXiv:2108.02759 (2021)
34. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: MICCAI. pp. 234–241 (2015)
35. Silva, J., Histace, A., Romain, O., Dray, X., Granado, B.: Toward embedded detection of polyps in wce images for early diagnosis of colorectal cancer. IJCARS **9**, 283–293 (2014)
36. Sun, Y., Chen, G., Zhou, T., Zhang, Y., Liu, N.: Context-aware cross-level fusion network for camouflaged object detection. In: IJCAI. pp. 1025–1031 (2021)
37. Tajbakhsh, N., Gurudu, S.R., Liang, J.: Automated polyp detection in colonoscopy videos using shape and context information. TMI **35**(2), 630–644 (2015)
38. Wang, L., Lu, H., Wang, Y., Feng, M., Wang, D., Yin, B., Ruan, X.: Learning to detect salient objects with image-level supervision. In: CVPR. pp. 136–145 (2017)
39. Wang, W., Xie, E., Li, X., Fan, D.P., Song, K., Liang, D., Lu, T., Luo, P., Shao, L.: Pvt v2: Improved baselines with pyramid vision transformer. CVM **8**(3), 415–424 (2022)
40. Wei, J., Hu, Y., Zhang, R., Li, Z., Zhou, S.K., Cui, S.: Shallow attention network for polyp segmentation. In: MICCAI. pp. 699–708 (2021)
41. Wei, J., Wang, S., Wu, Z., Su, C., Huang, Q., Tian, Q.: Label decoupling framework for salient object detection. In: CVPR. pp. 13025–13034 (2020)
42. Wu, Z., Su, L., Huang, Q.: Cascaded partial decoder for fast and accurate salient object detection. In: CVPR. pp. 3907–3916 (2019)
43. Wu, Z., Su, L., Huang, Q.: Decomposition and completion network for salient object detection. TIP **30**, 6226–6239 (2021)
44. Yan, Q., Xu, L., Shi, J., Jia, J.: Hierarchical saliency detection. In: CVPR. pp. 1155–1162 (2013)
45. Yang, F., Zhai, Q., Li, X., Huang, R., Luo, A., Cheng, H., Fan, D.P.: Uncertainty-guided transformer reasoning for camouflaged object detection. In: ICCV. pp. 4146–4155 (2021)
46. Yin, Z., Liang, K., Ma, Z., Guo, J.: Duplex contextual relation network for polyp segmentation. In: ISBI. pp. 1–5. IEEE (2022)
47. Zhai, Q., Li, X., Yang, F., Chen, C., Cheng, H., Fan, D.P.: Mutual graph learning for camouflaged object detection. In: CVPR. pp. 12997–13007 (2021)
48. Zhang, M., Xu, S., Piao, Y., Shi, D., Lin, S., Lu, H.: Preynet: Preying on camouflaged objects. In: ACM MM. pp. 5323–5332 (2022)

49. Zhang, R., Li, G., Li, Z., Cui, S., Qian, D., Yu, Y.: Adaptive context selection for polyp segmentation. In: MICAAI. pp. 253–262 (2020)
50. Zhao, J.X., Liu, J.J., Fan, D.P., Cao, Y., Yang, J., Cheng, M.M.: Egnet: Edge guidance network for salient object detection. In: ICCV. pp. 8779–8788 (2019)
51. Zhao, X., Pang, Y., Zhang, L., Lu, H., Zhang, L.: Suppress and balance: A simple gated network for salient object detection. In: ECCV. pp. 35–51 (2020)
52. Zhou, H., Xie, X., Lai, J.H., Chen, Z., Yang, L.: Interactive two-stream decoder for accurate and fast saliency detection. In: CVPR. pp. 9141–9150 (2020)
53. Zhou, T., Zhou, Y., Gong, C., Yang, J., Zhang, Y.: Feature aggregation and propagation network for camouflaged object detection. TIP **31**, 7036–7047 (2022)
54. Zhou, Z., Rahman Siddiquee, M.M., Tajbakhsh, N., Liang, J.: Unet++: A nested u-net architecture for medical image segmentation. In: MIA. pp. 3–11 (2018)
55. Zhu, H., Li, P., Xie, H., Yan, X., Liang, D., Chen, D., Wei, M., Qin, J.: I can find you! boundary-guided separated attention network for camouflaged object detection. In: AAAI. vol. 36, pp. 3608–3616 (2022)
56. Zhuge, M., Fan, D.P., Liu, N., Zhang, D., Xu, D., Shao, L.: Salient object detection via integrity learning. TPAMI **45**(3), 3738–3752 (2023)