Supplementary Material for
Taking a Deeper Look at Co-Salient Object Detection

Deng-Ping Fan[1,2,*] Zheng Lin[1,*] Ge-Peng Ji[3] Dingwen Zhang[4] Huazhu Fu[2] Ming-Ming Cheng[1] ✉
[1] TKLNDST, College of CS, Nankai University [2] Inception Institute of Artificial Intelligence
[3] School of Computer Science, Wuhan University [4] Xidian University
(* Equal contributions) http://dpfan.net/CoSOD3k/

Abstract

In this supplementary material, we provide more details on our benchmark, datasets, metrics, and results.

- ***Benchmark.*** *Here, we provide details on the motivation behind this benchmark.*
- ***Dataset.*** *We provide several important characteristics in terms of dataset comparisons, mixture-specific category masks, and quality control.*
- ***Metrics.*** *We provide the details of evaluation metrics including S-measure [5] (S_α), max E-measure [6] (E_ξ), max F-measure [1] (F_β), MAE [15] (M), and PR curve [3].*
- ***Results.*** *We provide more qualitative comparisons on the proposed CoSOD3k. Five visual results of 10 state-of-the-art models are shown in Fig. 10∼Fig. 14.*

To promote growth in the co-salient object detection (CoSOD) community, we will make our benchmark toolbox and results publicly available upon publication.

1. Benchmark

Similar to RGB-D salient object detection (SOD) [7], co-salient object detection (CoSOD) also suffers from the similar limitations (convenience, sufficiency, completeness, and fairness) which is a big hindrance to this field.

(a) Convenience. The current CoSOD datasets have various input formats (*e.g.*, *.bmp, *.png, *.jpg), making it challenging to evaluate models. We provide a convenient benchmark toolbox to integrate various publicly available CoSOD datasets with uniform input/output formats (*i.e.*, *.jpg for image; *.png for GT and predicted map).

(b) Sufficiency. In previous works [4, 10, 17, 18, 18], only a limited number of datasets (1∼3) are evaluated. This prevents an accurate assessment of the generalization ability of

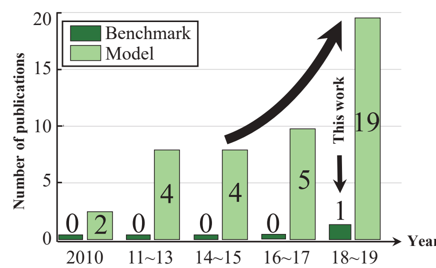

Figure 1: Statistics on the number of publications in different years. Note that the statistics only focus on top-tier conferences (*e.g.*, CVPR, ICCV, AAAI, IJCAI, MM, *etc*) or journals (TPAMI, IJCV, TNNLS, TIP, *etc*).

models. We make the first attempt to assess the models on all existing datasets (Image Pair [11], MSRC [19], iCoSeg [2], and CoSal2015 [23]) as well as the proposed *CoSOD3k*.

(c) Completeness. F-measure, and MAE are two widely used metrics in previous model evaluations. As discussed in [14], these metrics have their inherent limitations. To provide thorough and reliable conclusions, we introduce two more accurate metrics (*i.e.*, S-measure [5] and E-measure [6]).

(d) Fairness. To evaluate the F-measure, the first step is to binarize a saliency map into a set of foreground maps using different threshold values. There are many binarization strategies [3], such as adaptive threshold, fixed threshold and so on. Different strategies will result in different F-measure performances. In some previous works, they do not provide the details on the binarization strategy, easily resulting in inconsistent F-measures for different researchers.

(e) Flourish. As can be seen from the trend depicted in Fig. 1, the co-salient object detection task has been rapidly growing since 2018.

Based on the aforementioned reasons, we argue that integrating various publicly available CoSOD algorithms, datasets, and metrics, and then providing a complete benchmark, is highly desired. The 160 average performances of per sub-class are listed in Tab. 1, 2 & 3. The complete evaluation results can be found in our online benchmark.

2. Dataset

2.1. Dataset Comparison

As can be seen in Fig. 4, the proposed *CoSOD3k* is more challenging than previous datasets.

2.2. Mixture-specific Category Masks

In Fig. 8 & 9, we show the masks for 160 mixture-specific categories. It can be observed that some categories with unique shapes (e.g., airplane, zebra, and bicycle) often present shape-bias maps, while the categories with non-rigid or convex shapes (e.g., goldfish, bird, and bus) may have no clear shape-bias.

2.3. Quality Control

To provide high-quality annotations, we introduced a series of strict annotation guidelines during the labeling process. Examples of cases that either passed or were rejected are shown in Fig. 7.

3. Evaluation Metrics

3.1. Recently Released

S-measure [5] (S_α) was proposed by Fan *et al.* to evaluate the structure similarity between a saliency map (S_i) and ground-truth (G_i):

$$S_\alpha(S_i, G_i) = (1-\alpha) * S_o(S_i, G_i) + \alpha * S_r(S_i, G_i), \quad (1)$$

where α is a balance factor between the region-aware similarity S_r and object-aware similarity S_o. We use the default setting (α=0.5) as described in the original paper.

E-measure [6] (E_ξ) is another recently released metric that evaluates both local and global similarity between the predicted map and ground-truth simultaneously. The formulation is defined as:

$$E_\xi(S_i, G_i) = \frac{1}{W \times H} \sum_{x=1}^{W} \sum_{y=1}^{H} \phi(S_i, G_i), \quad (2)$$

where W and H are the width and height of the ground-truth map, respectively. $\phi(\cdot)$ is the enhanced alignment matrix described in [6].

3.2. Widely-Used

Precision-Recall (PR) Curve [3] is defined as:

$$Recall(t) = \frac{|S_i^t \cap G_i|}{|G_i|}, Precision(t) = \frac{|S_i^t \cap G_i|}{|S_i^t|}, \quad (3)$$

where S_i^t is a binarized foreground map after using a threshold $t \in [0, 255]$. $|\cdot|$ is the total number of pixels. The PR curve can be plotted according to a 255 threshold t.

F-measure [1] (F_β) is a widely used metric and is defined as:

$$F_\beta = \frac{(1+\beta^2)Precision \cdot Recall}{\beta^2 \cdot Precision + Recall}, \quad (4)$$

where β^2 is set to 0.3 as default. We compute a set of precisions and recalls using (Eq. 3) then plot the PR curve.

MAE [15] (M) is defined as:

$$M = \frac{1}{N}|S_i - G_i|, \quad (5)$$

where N is the total number of pixels in G_i.

4. Experiment Results

In our *manuscript*, we present some experimental results for two groups of images. Here, we provide more comparison results (see Fig. 10∼Fig. 14) on our *CoSOD3k*.

Mid-Class/Sub-class	CBCS [8]	CSHS [13]	ESMG [12]	CODR [21]	DIM [22]	UMLF [9]	IML [16]	CSMG [24]	CPD [20]	EGNet [25]
Animal/ant	0.594	0.611	0.607	0.667	0.526	0.551	0.679	0.604	0.756	**0.776**
Animal/antelope	0.543	0.602	0.543	0.705	0.582	0.57	0.802	0.722	0.879	**0.901**
Animal/armadillo	0.532	0.548	0.565	0.652	0.538	0.704	0.855	0.81	0.904	**0.916**
Animal/bear	0.637	0.702	0.69	0.776	0.687	0.734	0.885	0.902	**0.927**	0.926
Animal/bee	0.489	0.489	0.594	0.547	0.468	0.555	0.575	0.574	0.647	**0.7**
Animal/bird	0.559	0.56	0.494	0.614	0.517	0.644	0.841	0.83	**0.922**	0.922
Animal/butterfly	0.55	0.627	0.551	0.681	0.599	0.732	0.842	0.827	**0.931**	0.929
Animal/camel	0.541	0.572	0.558	0.65	0.598	0.66	0.788	0.819	0.822	**0.835**
Animal/cattle	0.588	0.566	0.617	0.662	0.605	0.674	0.797	0.817	**0.847**	0.831
Animal/centipede	0.546	0.52	0.493	0.559	0.499	0.501	0.659	0.631	**0.718**	0.718
Animal/dog	0.586	0.604	0.565	0.679	0.608	0.703	0.854	0.831	**0.884**	0.877
Animal/domestic_cat	0.566	0.579	0.519	0.666	0.507	0.722	0.766	0.799	**0.843**	0.83
Animal/dragonfly	0.492	0.523	0.472	0.666	0.537	0.618	0.701	0.659	0.736	**0.75**
Animal/elephant	0.481	0.654	0.603	0.753	0.593	0.725	0.854	0.865	**0.882**	0.858
Animal/fox	0.525	0.622	0.582	0.718	0.609	0.737	0.883	0.87	0.92	**0.927**
Animal/frog	0.535	0.584	0.532	0.707	0.593	0.691	0.848	0.794	**0.877**	0.874
Animal/giant_panda	0.497	0.556	0.525	0.699	0.574	0.679	0.751	0.841	0.803	**0.85**
Animal/goldfish	0.573	0.634	0.633	0.698	0.619	0.593	0.776	0.671	**0.783**	0.777
Animal/hamster	0.443	0.603	0.644	0.671	0.577	0.751	0.73	0.714	0.798	**0.816**
Animal/hippopotamus	0.511	0.617	0.548	0.722	0.589	0.686	0.812	0.855	**0.894**	0.891
Animal/horse	0.511	0.516	0.5	0.651	0.552	0.633	0.777	**0.827**	0.806	0.795
Animal/isopod	0.638	0.567	0.558	0.672	0.62	0.641	0.789	0.796	0.838	**0.855**
Animal/jellyfish	0.589	0.653	0.589	0.764	0.612	0.625	0.802	0.681	0.759	**0.809**
Animal/koala_bear	0.424	0.398	0.529	0.691	0.49	0.753	0.738	0.834	**0.848**	0.839
Animal/ladybug	0.652	0.611	0.638	0.722	0.611	0.68	0.814	0.71	0.885	**0.888**
Animal/lion	0.47	0.567	0.51	0.626	0.621	0.718	0.848	0.864	**0.894**	0.883
Animal/lizard	0.515	0.516	0.489	0.619	0.519	0.583	0.739	0.738	0.829	**0.858**
Animal/lobster	0.54	0.565	0.672	0.616	0.608	0.634	0.665	0.621	**0.677**	0.596
Animal/monkey	0.562	0.577	0.536	0.66	0.578	0.72	0.829	0.839	0.869	**0.879**
Animal/otter	0.555	0.614	0.571	0.69	0.585	0.726	0.854	0.871	**0.908**	0.905
Animal/person	0.494	0.527	0.461	0.572	0.439	0.622	0.644	0.752	0.736	**0.756**
Animal/porcupine	0.532	0.566	0.581	0.733	0.588	0.782	0.841	0.85	0.876	**0.891**
Animal/rabbit	0.559	0.701	0.714	0.734	0.664	0.762	0.876	0.864	0.896	**0.902**
Animal/ray	0.601	0.637	0.503	0.693	0.573	0.703	0.854	0.808	**0.89**	0.885
Animal/red_panda	0.608	0.523	0.601	0.658	0.619	0.663	0.777	0.822	**0.842**	0.84
Animal/scorpion	0.604	0.617	0.545	0.683	0.54	0.597	0.745	0.684	0.792	**0.805**
Animal/seal	0.564	0.623	0.642	0.694	0.603	0.651	0.811	0.853	**0.894**	0.861
Animal/sheep	0.537	0.631	0.538	0.745	0.609	0.668	0.845	0.841	0.887	**0.893**
Animal/skunk	0.624	0.663	0.659	0.749	0.601	0.704	0.774	0.822	0.837	**0.852**
Animal/snail	0.596	0.663	0.636	0.76	0.654	0.736	0.82	0.754	**0.883**	0.874
Animal/snake	0.498	0.519	0.491	0.605	0.493	0.53	0.698	0.676	0.815	**0.846**
Animal/squirrel	0.51	0.572	0.551	0.616	0.533	0.673	0.732	0.823	0.868	**0.896**
Animal/starfish	0.602	0.617	0.537	0.687	0.598	0.648	0.839	0.721	0.823	**0.851**
Animal/swine	0.491	0.626	0.55	0.75	0.591	0.745	0.855	0.878	**0.923**	0.916
Animal/tick	0.65	0.701	0.648	0.741	0.622	0.614	0.839	0.703	**0.88**	0.874
Animal/tiger	0.494	0.594	0.567	0.672	0.59	0.693	0.826	0.808	**0.871**	0.854
Animal/turtle	0.485	0.581	0.474	0.647	0.53	0.675	0.815	0.757	**0.857**	0.847
Animal/whale	0.653	0.667	0.667	0.729	0.617	0.635	0.81	0.776	0.902	**0.911**
Animal/zebra	0.451	0.615	0.547	0.732	0.572	0.66	0.805	0.794	0.878	**0.886**
Ball/baseball	0.486	0.555	0.482	0.614	0.567	0.521	0.698	0.671	0.706	**0.726**
Ball/basketball	0.538	0.427	0.483	0.469	0.423	0.422	0.523	**0.651**	0.509	0.505
Ball/croquet_ball	0.59	0.507	0.429	0.521	0.491	0.513	0.646	0.613	**0.709**	0.702
Ball/golf_ball	0.613	0.715	0.575	0.768	0.599	0.78	0.869	**0.911**	0.893	0.909
Ball/rugby_ball	0.493	0.547	0.466	0.574	0.517	0.622	0.654	**0.741**	0.658	0.674
Ball/soccer_ball	0.552	0.511	0.481	0.584	0.505	0.595	0.633	**0.75**	0.641	0.64
Ball/tennis_ball	0.604	0.415	0.429	0.489	0.505	0.516	0.627	**0.659**	0.618	0.623

Table 1: **Results of Structure-measure [5] (S_α) for each sub-class in our *CoSOD3k* dataset (1/3).** The highest score of each subclass is marked in **bold**.

Sub-class	CBCS [8]	CSHS [13]	ESMG [12]	CODR [21]	DIM [22]	UMLF [9]	IML [16]	CSMG [24]	CPD [20]	EGNet [25]
Nece/bow_tie	**0.472**	0.438	0.441	0.454	0.401	0.443	0.44	0.375	0.39	0.425
Nece/brassiere	**0.638**	0.535	0.535	0.565	0.575	0.542	0.58	0.572	0.548	0.558
Nece/cream	0.444	0.497	0.349	0.617	0.578	0.713	0.733	0.71	0.806	**0.817**
Cosm/face_powder	0.481	0.643	0.5	0.633	0.601	0.72	0.74	0.62	**0.749**	0.736
Cosm/hair_spray	0.529	0.534	0.433	0.611	0.553	0.566	0.649	0.668	0.706	**0.728**
Cosm/lipstick	0.544	0.583	0.533	0.633	0.591	0.592	0.716	0.592	0.79	**0.817**
Cosm/perfume	0.464	0.516	0.469	0.532	0.376	0.584	**0.657**	0.649	0.632	0.631
Elec/computer_keyboard	0.438	0.555	0.535	0.605	0.526	0.592	0.552	**0.807**	0.581	0.607
Elec/computer_mouse	0.487	0.482	0.47	0.528	0.451	0.543	0.627	0.725	0.734	**0.738**
Elec/electric_fan	0.456	0.565	0.502	0.651	0.568	0.752	0.798	0.765	**0.858**	0.831
Elec/hair_dryer	0.554	0.499	0.462	0.56	0.511	0.559	0.622	**0.647**	0.633	0.634
Elec/iPod	0.57	0.641	0.494	0.713	0.573	0.736	0.785	0.735	**0.882**	0.862
Elec/lamp	0.619	0.592	0.537	0.609	0.562	0.592	0.741	0.605	**0.775**	0.766
Elec/laptop	0.559	0.541	0.539	0.501	0.583	0.539	0.609	**0.738**	0.687	0.691
Elec/remote_control	0.457	0.597	0.576	0.633	0.526	0.607	0.668	0.706	0.703	**0.739**
Elec/tape_player	0.461	0.684	0.525	0.742	0.604	0.787	0.791	0.8	**0.848**	0.811
Food/bagel	0.449	0.47	0.637	0.685	0.68	0.741	0.719	**0.803**	0.769	0.797
Food/hamburger	0.557	0.675	0.662	0.747	0.683	0.77	0.791	**0.87**	0.836	0.858
Food/pizza	0.384	0.574	0.481	0.692	0.647	0.779	0.727	**0.805**	0.795	0.745
Food/popsicle	0.623	0.476	0.465	0.497	0.496	0.498	0.545	**0.626**	0.525	0.573
Food/pretzel	0.465	0.55	0.521	0.608	0.625	0.655	0.683	**0.764**	0.735	0.756
Fruit/apple	0.657	0.674	0.739	0.769	0.71	0.753	0.799	**0.875**	0.84	0.844
Fruit/banana	0.573	0.525	0.635	0.612	0.578	0.637	0.712	**0.752**	0.717	0.731
Fruit/cucumber	0.564	0.612	0.68	0.613	0.535	0.594	0.611	**0.713**	0.674	0.65
Fruit/fig	0.532	0.657	0.488	0.729	0.588	0.751	0.864	0.802	**0.914**	0.909
Fruit/lemon	0.622	0.702	0.743	0.728	0.75	0.724	0.798	0.803	**0.865**	0.85
Fruit/orange	0.697	0.815	0.82	0.85	0.861	0.88	**0.936**	0.877	0.912	0.924
Fruit/pineapple	0.486	0.504	0.517	0.572	0.588	0.599	0.715	0.635	0.748	**0.765**
Fruit/pomegranate	0.595	0.626	0.521	0.682	0.695	0.727	0.801	0.718	0.833	**0.856**
Fruit/strawberry	0.696	0.599	0.698	**0.706**	0.656	0.614	0.627	0.634	0.586	0.596
Inst/banjo	0.482	0.442	0.486	0.523	0.438	0.509	**0.58**	0.566	0.502	0.511
Inst/cello	0.535	0.496	0.613	0.547	0.54	0.412	0.616	**0.638**	0.525	0.612
Inst/chime	0.491	0.497	0.49	0.593	0.461	0.638	0.687	0.732	0.763	**0.796**
Inst/drum	0.485	0.508	0.542	0.525	0.569	0.526	**0.634**	0.62	0.557	0.559
Inst/flute	**0.465**	0.44	0.464	0.451	0.422	0.389	0.434	0.434	0.391	0.419
Inst/french_horn	0.531	0.509	0.522	0.585	0.63	0.72	**0.795**	0.701	0.633	0.7
Inst/guitar	0.526	0.481	0.441	0.571	0.557	0.579	**0.679**	0.566	0.608	0.644
Inst/harmonica	0.445	0.554	0.473	0.58	0.471	0.591	0.584	**0.672**	0.563	0.567
Inst/harp	0.488	0.555	0.613	0.626	0.54	**0.629**	0.626	0.603	0.547	0.484
Inst/maraca	0.542	0.643	0.566	0.726	0.658	0.694	0.807	0.675	0.845	**0.852**
Inst/oboe	0.535	0.536	0.574	0.586	0.527	0.447	**0.591**	0.552	0.577	0.57
Inst/piano	0.655	0.678	0.7	0.599	0.639	0.676	**0.86**	0.727	0.798	0.769
Inst/trumpet	0.497	0.518	0.512	0.559	0.482	0.465	0.633	0.491	0.647	**0.664**
Inst/violin	0.546	0.637	0.628	0.592	0.617	0.556	**0.651**	0.592	0.599	0.55
Kitch/bowl	0.547	0.502	0.39	0.569	0.488	0.804	0.886	0.9	**0.928**	0.928
Kitch/frying_pan	0.519	0.542	0.59	0.645	0.606	0.646	**0.8**	0.676	0.794	0.768
Kitch/microwave	0.571	0.555	0.544	0.509	0.485	0.605	0.612	**0.867**	0.724	0.73
Kitch/refrigerator	0.416	0.432	0.362	0.483	0.402	0.555	0.495	**0.785**	0.649	0.642
Kitch/salt_or_pepper_shaker	0.536	0.65	0.475	0.711	0.547	0.772	0.85	0.781	**0.911**	0.909
Kitch/spatula	0.575	0.623	0.606	0.638	0.521	0.578	0.767	0.625	0.775	**0.826**
Kitch/strainer	0.476	0.576	0.518	0.659	0.538	0.647	**0.75**	0.615	0.714	0.728
Kitch/toaster	0.483	0.636	0.482	0.643	0.603	0.703	0.776	0.811	**0.866**	0.843
Kitch/washer	0.422	0.473	0.458	0.65	0.563	0.725	0.605	0.698	0.806	**0.825**
Nece/band_aid	0.577	0.492	0.465	0.577	0.511	0.535	0.568	0.579	0.614	**0.645**
Nece/beaker	0.414	0.487	0.377	0.53	0.488	0.678	0.687	0.633	0.761	**0.788**

Table 2: **Results of Structure-measure [5]** (S_α) **for each sub-class in our *CoSOD3k* dataset (2/3).** The highest score of each subclass is marked in **bold**.

Sub-class	CBCS [8]	CSHS [13]	ESMG [12]	CODR [21]	DIM [22]	UMLF [9]	IML [16]	CSMG [24]	CPD [20]	EGNet [25]
Nece/cup_or_mug	0.477	0.513	0.41	0.573	0.494	0.676	0.709	0.674	**0.795**	0.773
Nece/digital_clock	0.644	0.725	0.671	0.709	0.676	0.736	0.769	0.813	0.831	**0.853**
Nece/milk_can	0.503	0.603	0.518	0.572	0.512	0.76	0.776	0.829	**0.887**	0.874
Nece/pitcher	0.485	0.595	0.521	0.658	0.577	0.719	0.8	0.829	**0.867**	0.854
Nece/soap_dispenser	0.529	0.528	0.429	0.599	0.585	0.685	0.741	0.644	0.766	**0.773**
Nece/water_bottle	0.498	0.469	0.392	0.561	0.518	0.633	0.696	0.701	0.746	**0.772**
Nece/wine_bottle	0.396	0.477	0.392	0.62	0.493	0.659	0.663	0.631	**0.779**	0.752
Other/baby_bed	0.471	0.529	0.454	0.614	0.557	0.677	**0.747**	0.728	0.709	0.736
Other/bench	0.462	0.536	0.491	0.612	0.45	0.551	0.626	**0.716**	0.671	0.636
Other/binder	0.587	0.678	0.475	0.585	0.587	0.606	0.648	0.533	0.65	**0.715**
Other/bookshelf	0.369	0.463	0.416	0.603	0.529	**0.668**	0.551	0.506	0.575	0.583
Other/crutch	0.464	0.487	0.449	0.526	0.407	0.409	0.54	0.486	0.598	**0.615**
Other/dumbbell	0.453	0.461	0.474	0.506	0.478	0.436	**0.541**	0.523	0.508	0.509
Other/flower_pot	0.547	0.564	0.544	0.54	0.591	0.524	0.596	**0.658**	0.629	0.618
Other/hat_with_a_wide_brim	0.504	0.469	0.441	0.526	0.503	0.55	**0.6**	0.561	0.563	0.554
Other/pencil_box	0.445	0.716	0.572	0.781	0.652	0.819	0.766	0.603	**0.926**	0.904
Other/punching_bag	0.477	0.437	0.522	0.604	0.444	0.605	0.638	0.714	0.737	**0.753**
Other/purse	0.522	0.54	0.415	0.72	0.586	0.719	0.785	0.756	**0.817**	0.814
Other/rubber_eraser	0.547	0.513	0.564	0.585	0.553	0.534	0.637	0.622	0.674	**0.718**
Other/sofa	0.526	0.284	0.455	0.356	0.486	0.558	0.632	**0.791**	0.662	0.648
Other/stethoscope	0.55	**0.58**	0.533	0.519	0.529	0.375	0.527	0.499	0.496	0.532
Other/syringe	0.466	0.485	0.451	0.531	0.412	0.418	0.575	0.606	0.626	**0.636**
Other/tie	0.51	0.448	0.433	0.46	0.439	0.484	**0.538**	0.461	0.474	0.52
Other/traffic_light	0.567	0.581	0.58	0.673	0.594	0.563	0.647	**0.719**	0.718	0.719
Tool/axe	0.484	0.429	0.438	0.489	0.426	0.442	0.543	0.549	0.566	**0.571**
Tool/bow	0.563	0.566	0.581	0.552	0.495	0.36	0.574	0.53	**0.603**	0.586
Tool/can_opener	0.619	0.745	0.644	0.789	0.689	0.758	0.835	0.698	**0.894**	0.892
Tool/chain_saw	0.591	0.568	0.554	0.564	0.569	0.518	**0.636**	0.6	0.631	0.618
Tool/cocktail_shaker	0.449	0.663	0.506	0.768	0.595	0.707	0.783	0.742	**0.83**	0.829
Tool/hammer	0.503	0.463	0.483	0.595	0.538	0.474	0.655	0.582	0.687	**0.699**
Tool/helmet	0.49	0.485	0.485	0.468	0.485	0.444	**0.611**	0.463	0.598	0.584
Tool/nail	0.473	0.522	0.449	0.502	0.458	0.452	0.64	0.543	**0.784**	0.78
Tool/power_drill	0.516	0.556	0.507	0.578	0.546	0.57	0.684	0.633	0.674	**0.739**
Tool/screwdriver	0.543	0.484	0.487	0.548	0.458	0.441	0.592	0.6	0.607	**0.66**
Tool/snowplow	0.521	0.637	0.556	0.692	0.662	0.705	0.825	0.799	**0.88**	0.869
Traf/airplane	0.568	0.628	0.561	0.713	0.579	0.647	0.863	0.827	**0.918**	0.911
Traf/bicycle	0.471	0.51	0.498	0.572	0.51	0.544	0.6	**0.681**	0.665	0.647
Traf/bus	0.499	0.586	0.464	0.729	0.642	0.784	0.828	0.868	**0.895**	0.879
Traf/car	0.523	0.574	0.421	0.678	0.573	0.705	0.832	0.786	0.905	**0.914**
Traf/cart	0.521	0.502	0.536	0.495	0.575	0.607	0.633	0.586	**0.669**	0.622
Traf/golfcart	0.455	0.557	0.534	0.669	0.538	0.741	0.791	0.76	**0.873**	0.853
Traf/motorcycle	0.431	0.543	0.509	0.638	0.547	0.657	0.727	0.776	**0.802**	0.767
Traf/snowmobile	0.568	0.615	0.55	0.638	0.575	0.629	0.741	0.734	**0.845**	0.829
Traf/train	0.517	0.602	0.605	0.639	0.567	0.732	0.786	**0.893**	0.838	0.823
Traf/watercraft	0.567	0.629	0.587	0.717	0.584	0.644	0.818	0.786	0.826	**0.841**
Vege/artichoke	0.485	0.6	0.467	0.603	0.599	0.703	0.79	0.638	0.782	**0.792**
Vege/bell_pepper	0.625	0.471	0.495	0.62	0.572	0.732	0.786	0.782	0.87	**0.883**
Vege/head_cabbage	0.317	0.315	0.311	0.575	0.475	0.609	**0.626**	0.368	0.558	0.609
Vege/mushroom	0.62	0.696	0.679	0.728	0.727	0.799	0.865	0.79	0.867	**0.895**
Overall	0.525	0.560	0.520	0.623	0.561	0.642	0.714	0.700	0.752	0.758
Ranking	9	8	10	6	7	5	3	4	2	1

Table 3: **Results of Structure-measure [5] (S_α) for each sub-class in our *CoSOD3k* dataset (3/3).** The highest score of each subclass is marked in **bold**. The overall score is based on 160 sub-classes.

Figure 2: Objects from our *CoSOD3k*. Each object is sampled per sub-class (Fig. 3).

ant	antelope	armadillo	bear	bee	bird	butterfly	camel	cattle	centipede
dog	domestic_cat	dragonfly	elephant	fox	frog	giant_panda	goldfish	hamster	hippopotamus
horse	isopod	jellyfish	koala_bear	ladybug	lion	lizard	lobster	monkey	otter
person	porcupine	rabbit	ray	red_panda	scorpion	seal	sheep	skunk	snail
snake	squirrel	starfish	swine	tick	tiger	turtle	whale	zebra	baseball
basketball	croquet_ball	golf_ball	rugby_ball	soccer_ball	tennis_ball	face_powder	hair_spray	lipstick	perfume
computer_keyboard	computer_mouse	electric_fan	hair_dryer	iPod	lamp	laptop	remote_control	tape_player	bagel
hamburger	pizza	popsicle	pretzel	apple	banana	cucumber	fig	lemon	orange
pineapple	pomegranate	strawberry	banjo	cello	chime	drum	flute	french_horn	guitar
harmonica	harp	maraca	oboe	piano	trumpet	violin	bowl	frying_pan	microwave
refrigerator	salt_or_pepper_shaker	spatula	strainer	toaster	washer	band_aid	beaker	bow_tie	brassiere
cream	cup_or_mug	digital_clock	milk_can	pitcher	soap_dispenser	water_bottle	wine_bottle	baby_bed	bench
binder	bookshelf	crutch	dumbbell	flower_pot	hat_with_a_wide_brim	pencil_box	punching_bag	purse	rubber_eraser
sofa	stethoscope	syringe	tie	traffic_light	axe	bow	can_opener	chain_saw	cocktail_shaker
hammer	helmet	nail	power_drill	screwdriver	snowplow	airplane	bicycle	bus	car
cart	golfcart	motorcycle	snowmobile	train	watercraft	artichoke	bell_pepper	head_cabbage	mushroom

Figure 3: The 160 sub-classes of our *CoSOD3k*.

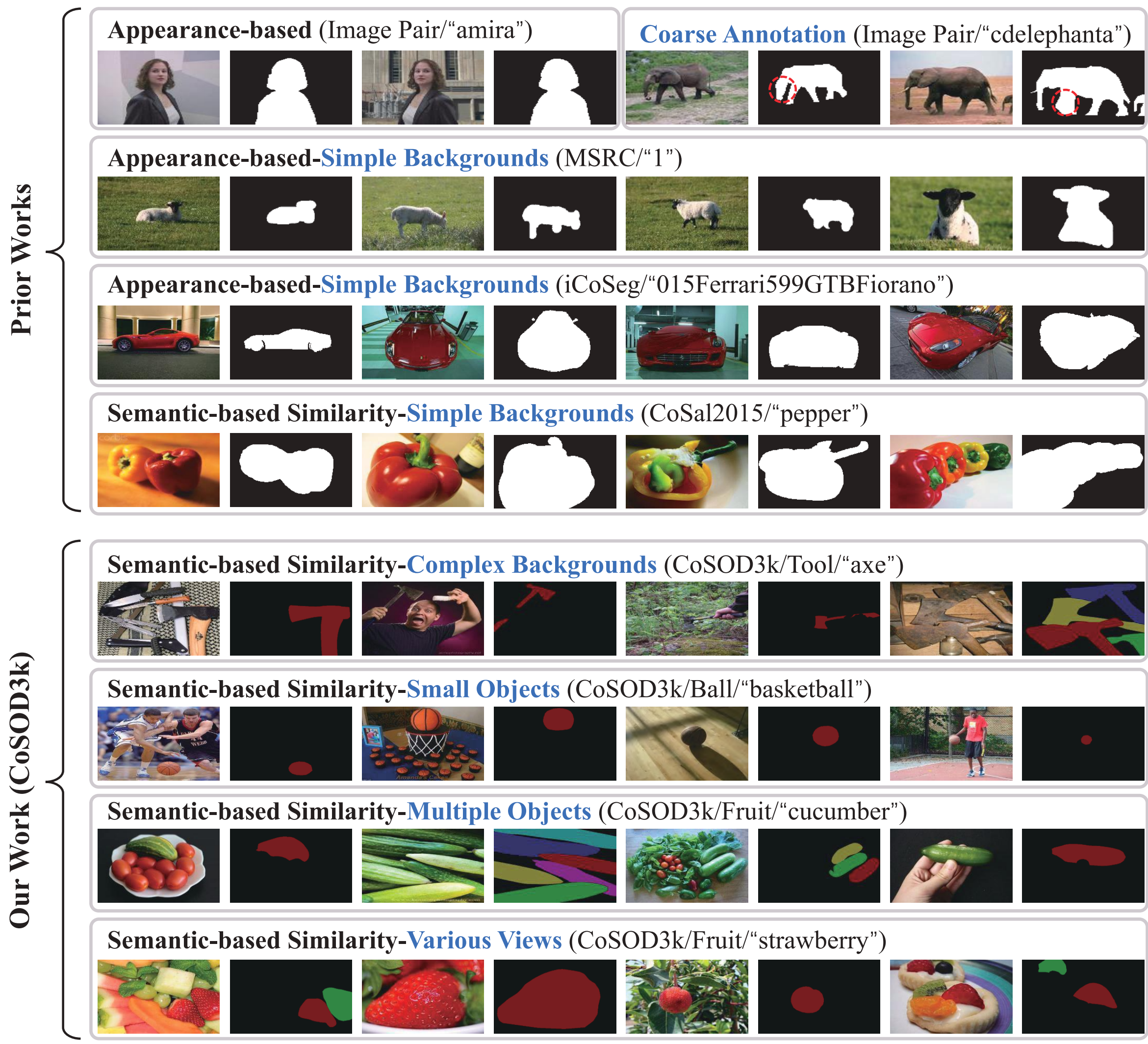

Figure 4: Examples of image groups from the existing datasets and the proposed *CoSOD3k*. Different from previous works (*e.g.*, appearance-aware or semantic-aware with simple backgrounds), the image group contained in our *CoSOD3k* are semantic-aware with complex backgrounds, small objects, multiple objects and various views.

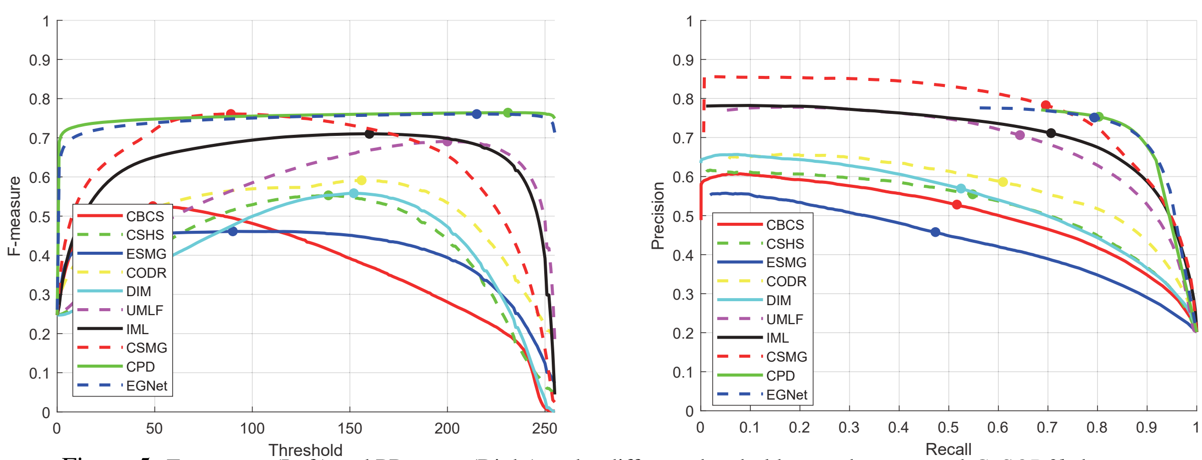

Figure 5: F-measure (Left) and PR curve (Right) under different thresholds t on the proposed *CoSOD3k* dataset.

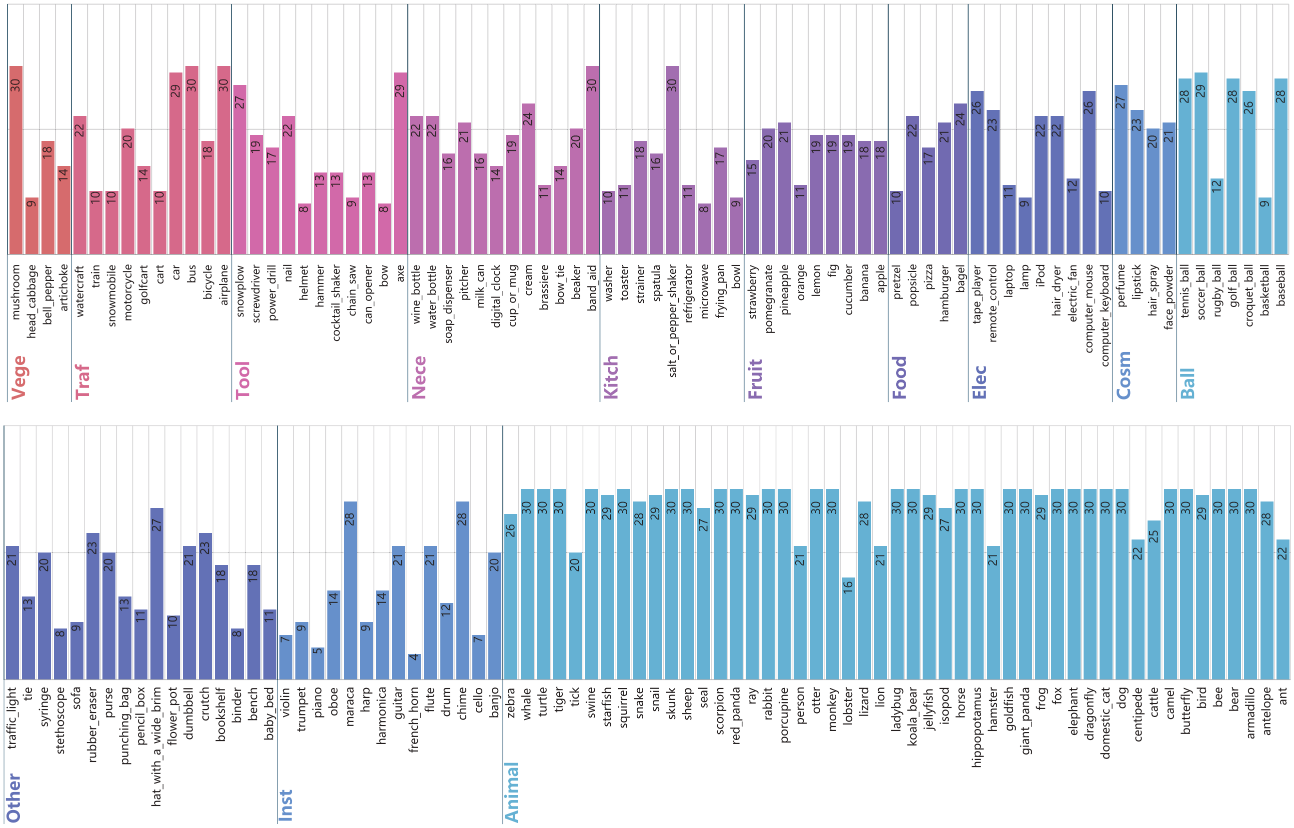

Figure 6: Number of images in 160 sup-classes.

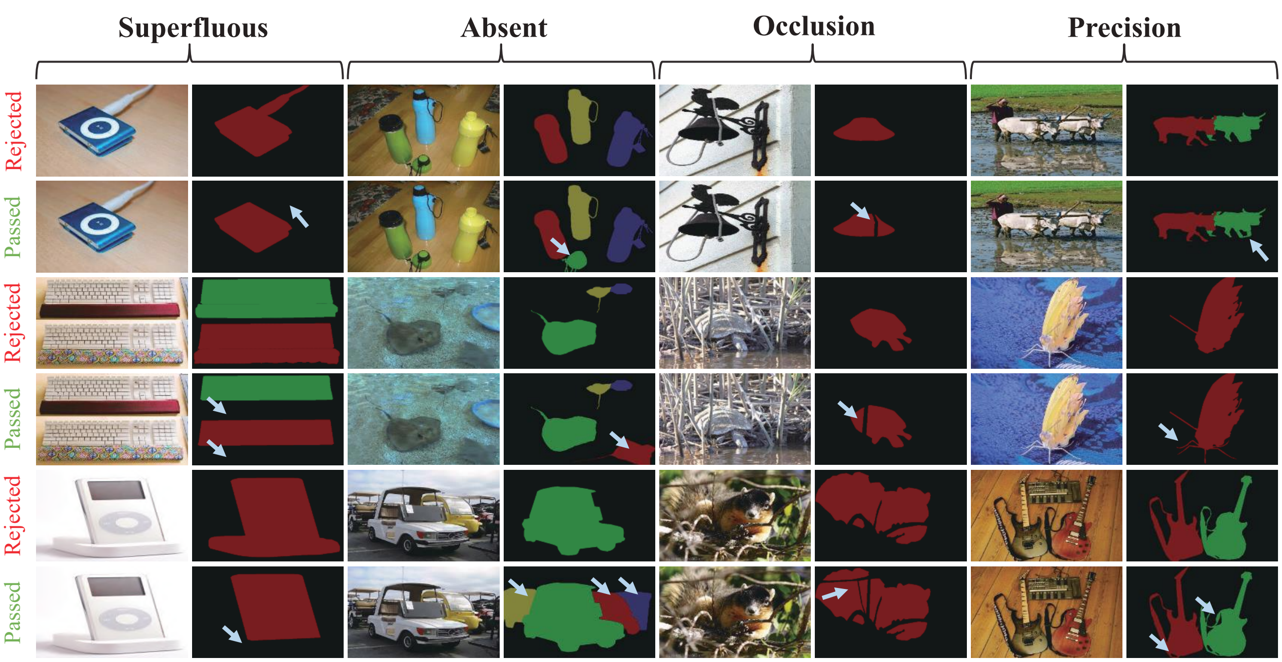

Figure 7: Some passed and rejected cases (*e.g.*, superfluous, absent, occlusion, precision) in our *CoSOD3k*.

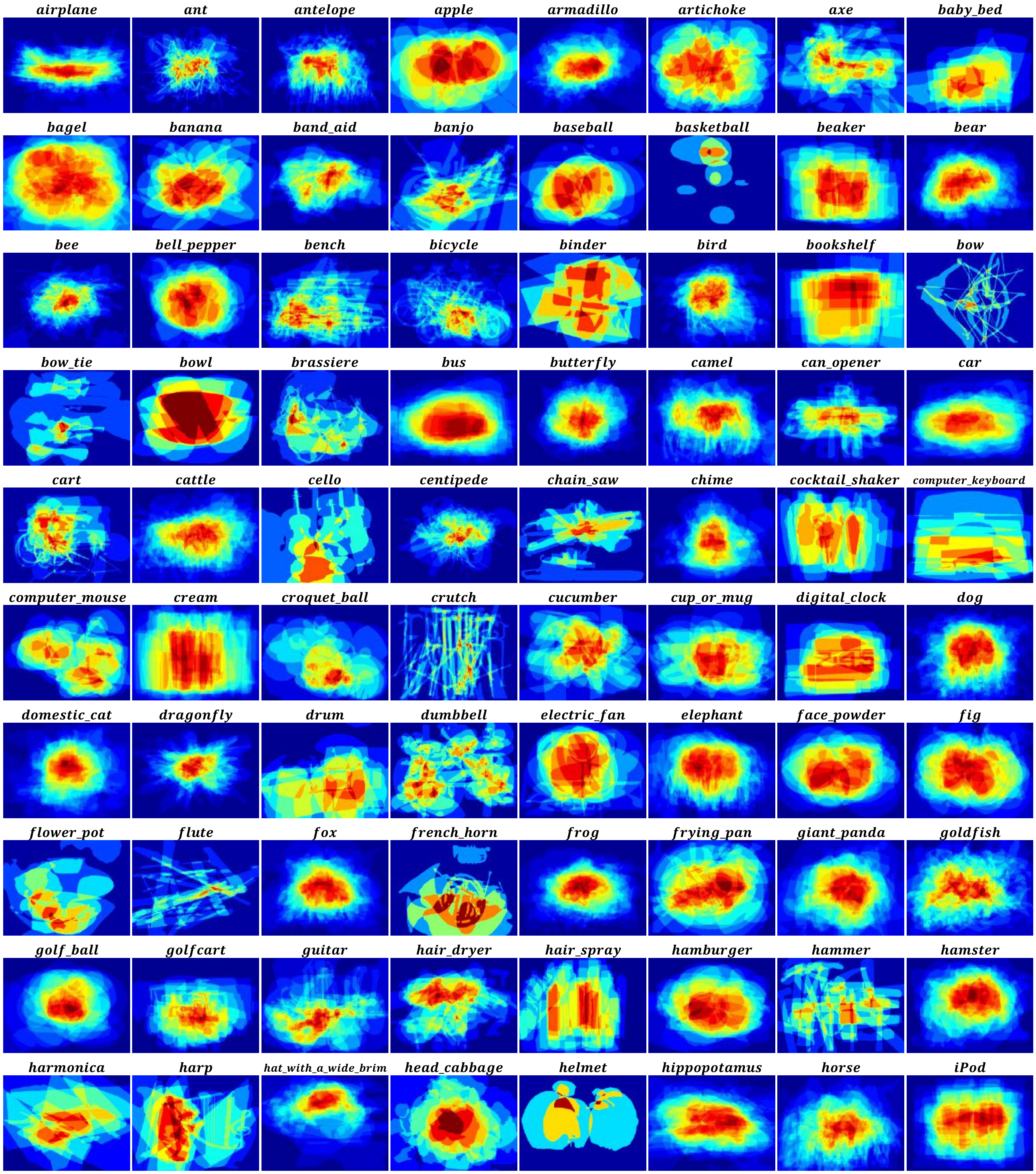

Figure 8: Visualization of overlap masks for 160 mixture-specific categories (1/2).

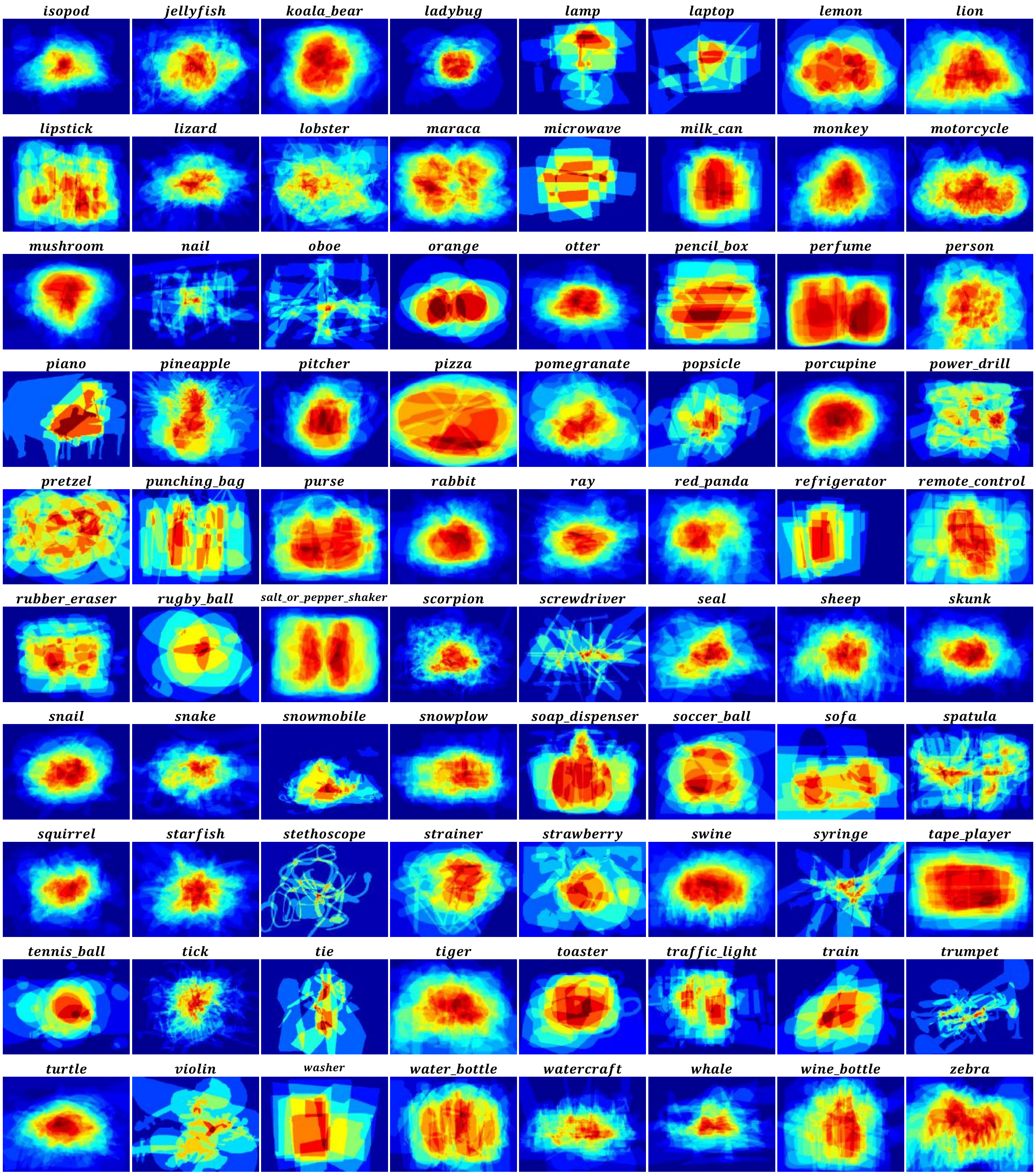

Figure 9: Visualization of overlap masks for 160 mixture-specific categories (2/2).

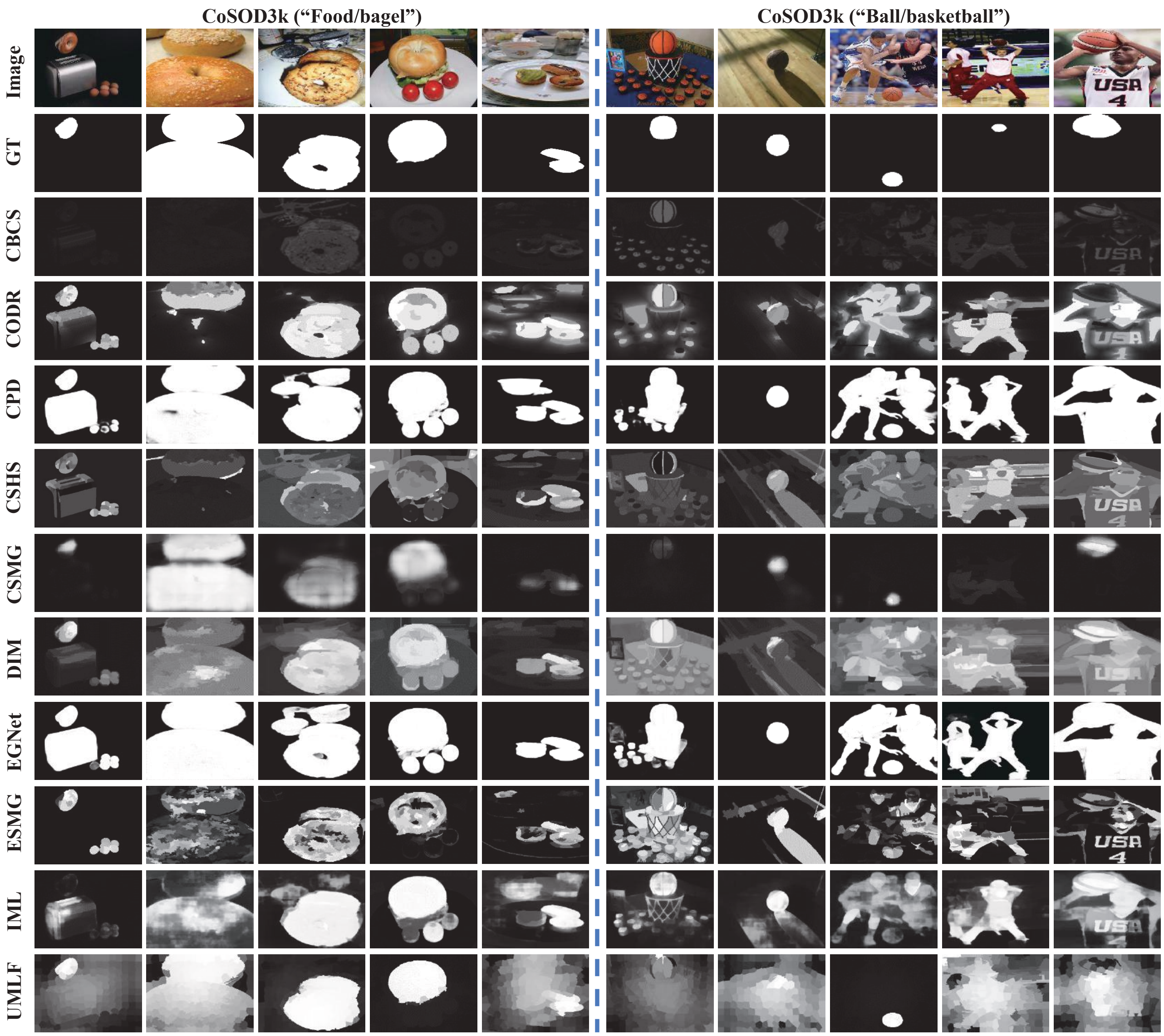

Figure 10: Qualitative examples of the top-10 existing models on *CoSOD3k*. From top to bottom, each row represents the original image group (Image), ground-truth (GT), EGNet [25], CPD [20], IML [16], CSMG [24], UMLF [9], CODR [21], DIM [22], CSHS [13], CBCS [8], and ESMG [12], respectively. Zoom-in for the best view (1/5).

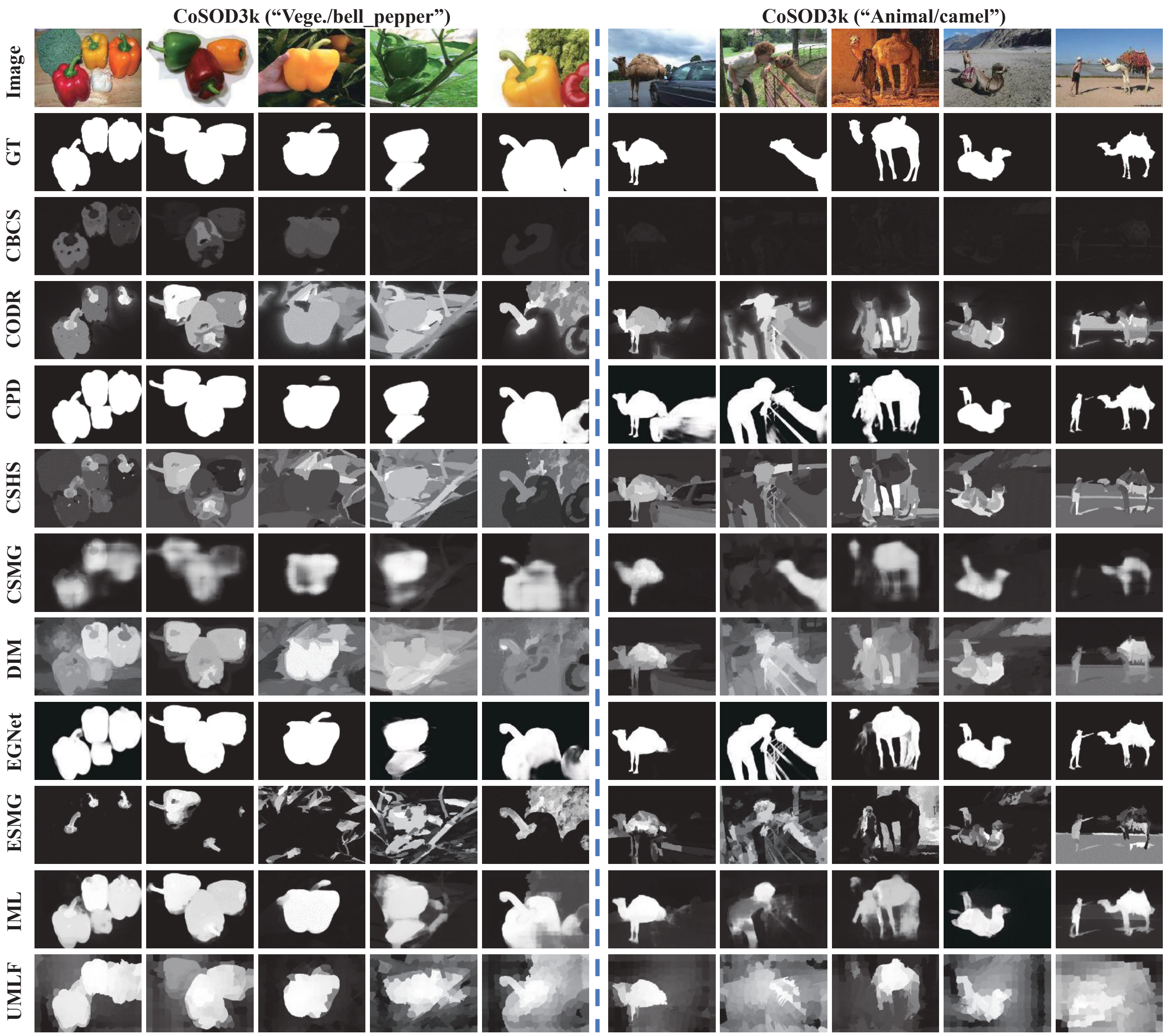

Figure 11: Qualitative examples of the top-10 existing models on *CoSOD3k*. From top to bottom, each row represents the original image group (Image), ground-truth (GT), EGNet [25], CPD [20], IML [16], CSMG [24], UMLF [9], CODR [21], DIM [22], CSHS [13], CBCS [8], and ESMG [12], respectively. Zoom-in for the best view (2/5).

Figure 12: Qualitative examples of the top-10 existing models on *CoSOD3k*. From top to bottom, each row represents the original image group (Image), ground-truth (GT), EGNet [25], CPD [20], IML [16], CSMG [24], UMLF [9], CODR [21], DIM [22], CSHS [13], CBCS [8], and ESMG [12], respectively. Zoom-in for the best view (3/5).

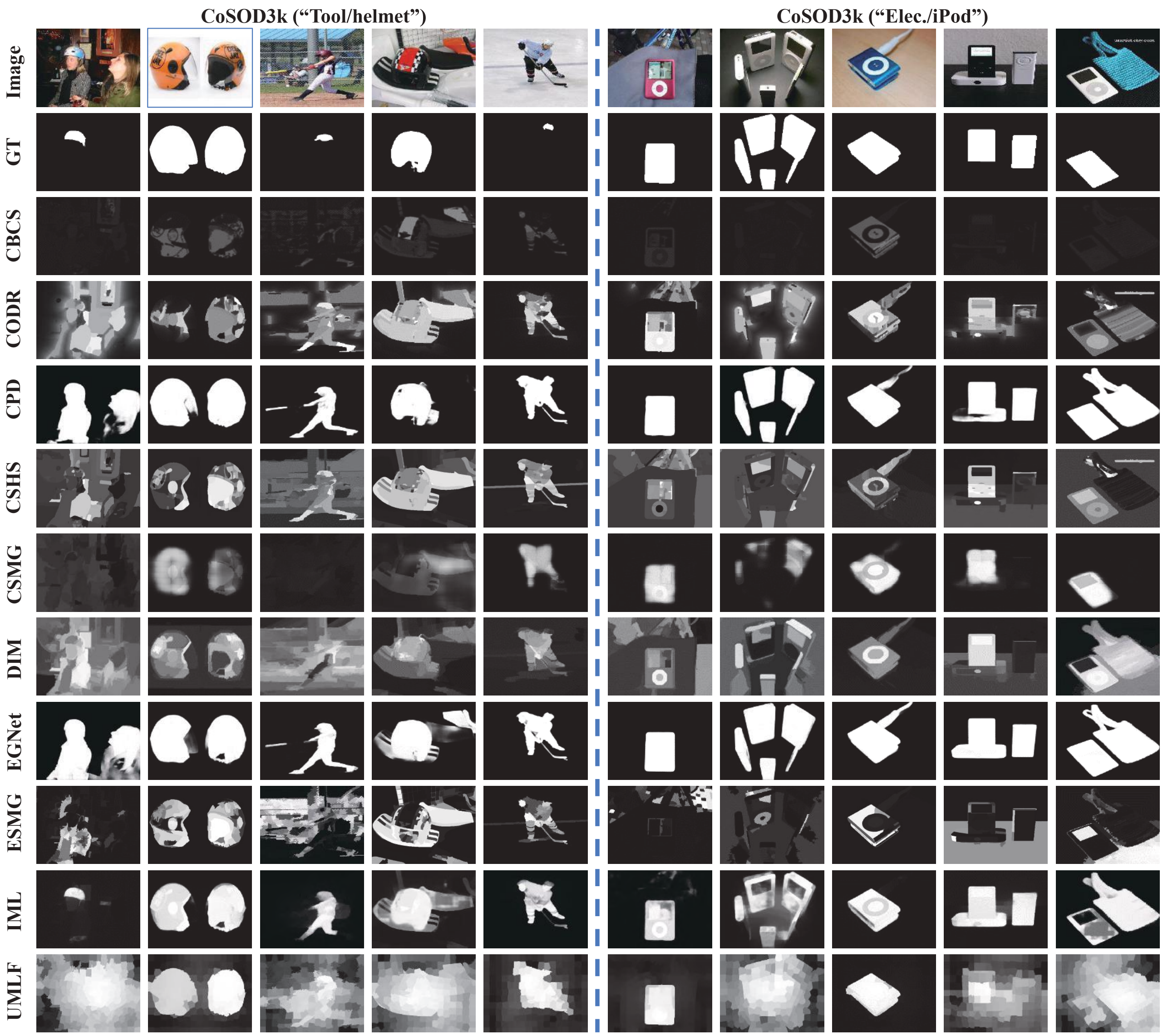

Figure 13: Qualitative examples of the top-10 existing models on *CoSOD3k*. From top to bottom, each row represents the original image group (Image), ground-truth (GT), EGNet [25], CPD [20], IML [16], CSMG [24], UMLF [9], CODR [21], DIM [22], CSHS [13], CBCS [8], and ESMG [12], respectively. Zoom-in for the best view (4/5).

Figure 14: Qualitative examples of the top-10 existing models on *CoSOD3k*. From top to bottom, each row represents the original image group (Image), ground-truth (GT), EGNet [25], CPD [20], IML [16], CSMG [24], UMLF [9], CODR [21], DIM [22], CSHS [13], CBCS [8], and ESMG [12], respectively. Zoom-in for the best view (5/5).

References

[1] Radhakrishna Achanta, Sheila Hemami, Francisco Estrada, and Sabine Süsstrunk. Frequency-tuned salient region detection. In *IEEE CVPR*, number CONF, pages 1597–1604, 2009. 1, 2

[2] Dhruv Batra, Adarsh Kowdle, Devi Parikh, Jiebo Luo, and Tsuhan Chen. icoseg: Interactive co-segmentation with intelligent scribble guidance. In *IEEE CVPR*, pages 3169–3176, 2010. 1

[3] Ali Borji, Ming-Ming Cheng, Huaizu Jiang, and Jia Li. Salient object detection: A benchmark. *IEEE TIP*, 24(12):5706–5722, 2015. 1, 2

[4] Xiaochun Cao, Zhiqiang Tao, Bao Zhang, Huazhu Fu, and Wei Feng. Self-adaptively weighted co-saliency detection via rank constraint. *IEEE TIP*, 23(9):4175–4186, 2014. 1

[5] Deng-Ping Fan, Ming-Ming Cheng, Yun Liu, Tao Li, and Ali Borji. Structure-measure: A New Way to Evaluate Foreground Maps. In *IEEE ICCV*, pages 4548–4557, 2017. 1, 2, 3, 4, 5

[6] Deng-Ping Fan, Cheng Gong, Yang Cao, Bo Ren, Ming-Ming Cheng, and Ali Borji. Enhanced-alignment Measure for Binary Foreground Map Evaluation. In *IJCAI*, pages 698–704, 2018. 1, 2

[7] Deng-Ping Fan, Zheng Lin, Jia-Xing Zhao, Yun Liu, Zhao Zhang, Qibin Hou, Menglong Zhu, and Ming-Ming Cheng. Rethinking RGB-D salient object detection: Models, datasets, and large-scale benchmarks. *arXiv preprint arXiv:1907.06781*, 2019. 1

[8] Huazhu Fu, Xiaochun Cao, and Zhuowen Tu. Cluster-based co-saliency detection. *IEEE TIP*, 22(10):3766–3778, 2013. 3, 4, 5, 11, 12, 13, 14, 15

[9] Junwei Han, Gong Cheng, Zhenpeng Li, and Dingwen Zhang. A unified metric learning-based framework for co-saliency detection. *IEEE TCSVT*, 28(10):2473–2483, 2017. 3, 4, 5, 11, 12, 13, 14, 15

[10] Bo Li, Zhengxing Sun, Quan Wang, and Qian Li. Co-saliency detection based on hierarchical consistency. In *ACM MM*, pages 1392–1400, 2019. 1

[11] Hongliang Li and King Ngi Ngan. A co-saliency model of image pairs. *IEEE TIP*, 20(12):3365–3375, 2011. 1

[12] Yijun Li, Keren Fu, Zhi Liu, and Jie Yang. Efficient saliency-model-guided visual co-saliency detection. *IEEE SPL*, 22(5):588–592, 2014. 3, 4, 5, 11, 12, 13, 14, 15

[13] Zhi Liu, Wenbin Zou, Lina Li, Liquan Shen, and Olivier Le Meur. Co-saliency detection based on hierarchical segmentation. *IEEE SPL*, 21(1):88–92, 2013. 3, 4, 5, 11, 12, 13, 14, 15

[14] Ran Margolin, Lihi Zelnik-Manor, and Ayellet Tal. How to evaluate foreground maps? In *IEEE CVPR*, pages 248–255, 2014. 1

[15] Federico Perazzi, Philipp Krähenbühl, Yael Pritch, and Alexander Hornung. Saliency filters: Contrast based filtering for salient region detection. In *CVPR*, pages 733–740, 2012. 1, 2

[16] Jingru Ren, Zhi Liu, Xiaofei Zhou, Cong Bai, and Guangling Sun. Co-saliency detection via integration of multi-layer convolutional features and inter-image propagation. *Neurocomputing*, 2019. 3, 4, 5, 11, 12, 13, 14, 15

[17] Chung-Chi Tsai, Weizhi Li, Kuang-Jui Hsu, Xiaoning Qian, and Yen-Yu Lin. Image co-saliency detection and co-segmentation via progressive joint optimization. *IEEE TIP*, 28(1):56–71, 2018. 1

[18] Lina Wei, Shanshan Zhao, Omar El Farouk Bourahla, Xi Li, Fei Wu, and Yueting Zhuang. Deep group-wise fully convolutional network for co-saliency detection with graph propagation. *IEEE TIP*, 2019. 1

[19] John Winn, Antonio Criminisi, and Tom Minka. Object categorization by learned universal visual dictionary. In *IEEE ICCV*, pages 1800–1807, 2005. 1

[20] Zhe Wu, Li Su, and Qingming Huang. Cascaded partial decoder for fast and accurate salient object detection. In *IEEE CVPR*, pages 3907–3916, 2019. 3, 4, 5, 11, 12, 13, 14, 15

[21] Linwei Ye, Zhi Liu, Junhao Li, Wan-Lei Zhao, and Liquan Shen. Co-saliency detection via co-salient object discovery and recovery. *IEEE SPL*, 22(11):2073–2077, 2015. 3, 4, 5, 11, 12, 13, 14, 15

[22] Dingwen Zhang, Junwei Han, Jungong Han, and Ling Shao. Cosaliency detection based on intrasaliency prior transfer and deep intersaliency mining. *IEEE TNNLS*, 27(6):1163–1176, 2015. 3, 4, 5, 11, 12, 13, 14, 15

[23] Dingwen Zhang, Junwei Han, Chao Li, and Jingdong Wang. Co-saliency detection via looking deep and wide. In *IEEE CVPR*, pages 2994–3002, 2015. 1

[24] Kaihua Zhang, Tengpeng Li, Bo Liu, and Qingshan Liu. Co-saliency detection via mask-guided fully convolutional networks with multi-scale label smoothing. In *CVPR*, pages 3095–3104, 2019. 3, 4, 5, 11, 12, 13, 14, 15

[25] Jia-Xing Zhao, Jiang-Jiang Liu, Deng-Ping Fan, Yang Cao, Jufeng Yang, and Ming-Ming Cheng. Egnet: Edge guidance network for salient object detection. In *IEEE ICCV*, Oct 2019. 3, 4, 5, 11, 12, 13, 14, 15