

LARGE-SCALE BENCHMARK

1. MORE DATASET COMPARISONS AND ILLUSTRATIONS

Table 1. Supplemental dataset comparison (Part 1).

Dataset Name	Publication	Number	Category	Image Size	IPQ	Annotation	SAQ (Ours)	URL
Salient Object Detection								
ECSSD [1]	CVPR'13	1000	N/A	375±48 × 311±56	3.260	Binary	0.885±0.046	Link
HKU-IS [2]	CVPR'15	4447	N/A	387±37 × 292±51	4.408	Binary	0.875±0.047	Link
DUT-OMRON [3]	CVPR'13	5168	N/A	377±46 × 321±54	4.079	Binary	0.847±0.055	Link
MSRA10K [4]	TPAMI'10	10000	N/A	370±50 × 325±56	2.538	Binary	0.894±0.052	Link
MSRA-B [5]	IJCV'17	5000	N/A	371±51 × 322±56	2.891	Binary	0.881±0.055	Link
SED [6]	TIP'15	195	N/A	293±23 × 264±79	4.133	Binary	0.880±0.064	Link
THUR15K [7]	VC'14	6232	5	399±80 × 329±58	3.114	Binary	0.869±0.049	Link
DUTS [8]	CVPR'17	15570	N/A	375±47 × 322±54	3.374	Binary	0.865±0.054	Link
SOD [9]	POCV'10	298	N/A	432±72 × 367±72	4.744	Binary	0.869±0.049	Link
SOC [10]	TPAMI'22	3000	N/A	435±72 × 367±72	2.601	Instance	0.865±0.054	Link
High-Resolution Salient Object Detection								
UHRSD [11]	CVPR'22	5920	N/A	4938±1026 × 3722±850	10.138	Binary	0.874±0.053	Link
HRSOD [12]	ICCV'19	2010	N/A	3412±1408 × 2713±1042	5.852	Binary	0.859±0.052	Link
HRS10K [13]	ACM MM'23	10500	N/A	4041±1303 × 3678±1307	8.950	Binary	0.867±0.051	Link
RGB-Depth Salient Object Detection								
DES [14]	ICIMCS'14	135	N/A	640±0 × 480±0	3.710	Binary	0.761±0.061	Link
LFSD [15]	CVPR'14	100	N/A	360±0 × 360±0	2.558	Binary	0.883±0.046	Link
SIP [16]	TNN'20	929	1	779±86 × 957±86	4.834	Binary	0.826±0.025	Link
SSD [17]	ICCV'17	80	N/A	960±0 × 1080±0	4.514	Binary	0.838±0.052	Link
STERE [18]	CVPR'12	1000	N/A	676±256 × 675±204	3.086	Binary	0.877±0.042	Link
NJU2K [19]	ICIO'14	1985	N/A	704±226 × 485±68	3.858	Binary	0.858±0.039	Link
NLPR [20]	ECCV'14	1000	N/A	592±73 × 528±73	3.262	Binary	0.852±0.068	Link
DUTLF-Depth [21]	ICCV'19	1200	N/A	256±0 × 256±0	2.662	Binary	0.888±0.055	Link
ReDWeb-S [22]	TPAMI'21	3179	N/A	471±215 × 410±116	7.719	Binary	0.821±0.045	Link
COME15K [23]	ICCV'21	15625	N/A	1017±320 × 597±199	2.970	Instance	0.834±0.050	Link
RGB-Thermal Salient Object Detection								
VT723 [24]	Arxiv'22	723	N/A	640±0 × 480±0	2.328	Binary	0.759±0.044	Link
VT821 [25]	IGTA'18	821	N/A	635±0 × 476±0	2.513	Binary	0.830±0.071	Link
VT1000 [26]	TMM'19	1000	N/A	640±0 × 480±0	1.854	Binary	0.831±0.075	Link
VT5000 [27]	TMM'22	5000	N/A	640±0 × 480±0	2.469	Binary	0.840±0.717	Link
Co-Salient Object Detection								
CoCA [28]	ECCV'20	1295	80	617±63 × 456±80	5.109	Instance	0.831±0.057	Link
CoSal2015 [29]	IJCV'16	2015	50	475±144 × 395±125	3.256	Binary	0.885±0.049	Link
CoSOD3K [30]	CVPR'20	3316	160	489±200 × 423±163	3.120	Instance	0.828±0.053	Link
iCoSeg [31]	CVPR'10	643	38	462±65 × 385±74	7.689	Binary	0.889±0.048	Link
ImagePair [32]	TIP'11	210	105	131±29 × 105±43	2.547	Binary	0.877±0.063	Link
MSRC [33]	ICCV'05	233	7	308±34 × 225±34	4,452	Binary	0.732±0.035	Link

Table 2. Supplemental dataset comparison (Part 2). SR: Scientific Reports; UM: Unpublished manuscript.

Dataset Name	Publication	Number	Category	Image Size	IPQ	Annotation	SAQ (Ours)	URL
Few-Shot Segmentation								
FSS-1000 [34]	CVPR'20	10000	1000	224±0 × 224±0	1.755	Binary	0.867±0.024	Link
Dichotomous Image Segmentation								
DIS5K [35]	ECCV'22	5470	N/A	3111±1360 × 2513±1053	107.611	Binary	0.845±0.068	Link
Thin Object Segmentation								
ThinObject5K [36]	CVPR'21	5748	392	1325±958 × 1186±910	26.587	Binary	0.886±0.059	Link
COIFT [37]	ICIP'16	280	102	600±79 × 488±92	11.882	Binary	0.873±0.063	Link
Interactive Image Segmentation								
GrabCut [38]	TOG'04	50	N/A	494±106 × 452±96	2.863	Binary	0.886±0.059	Link
Berkeley [39]	ICCV'01	100	N/A	436±72 × 366±72	3.463	Binary	0.873±0.063	Link
High-Resolution Segmentation								
R-PASCAL [40]	CVPR'20	501	N/A	470±60 × 384±65	4.438	Binary	0.873±0.063	Link
BIG [40]	CVPR'20	150	N/A	3672±1129 × 2801±890	11.944	Binary	0.863±0.048	Link
Semantic/Instance Segmentation								
PASCAL VOC [41]	IJCV'10	2913	20	472±57 × 385±64	6.003	Instance	0.742±0.053	Link
SBD [42]	ICCV'11	100	20	470±59 × 386±64	2.331	Instance	0.791±0.045	Link
Cityscapes [43]	CVPR'16	3475	33	2048±0 × 1024±0	21.919	Instance	0.792±0.013	Link
ADE20K [44]	CVPR'17	27574	150	933±686 × 749±536	4.030	Semantic	0.789±0.028	Link
MSCOCO [45]	ECCV'14	1449	80	640±0 × 480±0	2.408	Instance	0.786±0.039	Link
LVIS [46]	CVPR'19	110914	1203	577±92 × 484±97	2.761	Instance	0.834±0.042	Link
RGB-Depth Semantic Segmentation								
NYUD [47]	ECCV'12	1449	40	640±0 × 480±0	5.223	Semantic	0.770±0.022	Link
SUNRGBD [48]	CVPR'15	10335	37	646±72 × 481±47	5.504	Semantic	0.757±0.030	Link
RGB-Thermal Semantic Segmentation								
MFNet [49]	IROS'17	1555	8	640±0 × 480±0	4.898	Semantic	0.761±0.041	Link
Underwater Object Segmentation								
SUIM [50]	IROS'20	1488	7	703±195 × 498±82	7.019	Semantic	0.793±0.046	Link
DeepFish [51]	SR'20	310	N/A	1920±0 × 1080±0	2.284	Binary	0.795±0.031	Link
Camouflaged/Concealed Object Detection								
COD10K [52]	CVPR'20	10000	78	870±197 × 669±170	15.096	Instance	0.802±0.040	Link
CHAMELEON [53]	UM'23	76	N/A	981±465 × 742±452	15.580	Binary	0.796±0.040	Link
NC4K [54]	CVPR'21	4121	N/A	709±199 × 530±158	7.277	Binary	0.802±0.043	Link
CAMO [55]	CVIU'19	1250	N/A	653±573 × 509±414	4.039	Binary	0.781±0.046	Link
Video Object Segmentation								
DAVIS2017 [56]	ArXiv'17	6208	90	2412±878 × 1345±482	12.934	Instance	0.827±0.042	Link
YouTube-VOS [57]	ArXiv'18	95829	3978	1258±120 × 717±39	3.292	Instance	0.753±0.052	Link
Human Object Segmentation								
Fashionpedia [58]	ECCV'20	46781	13	755±147 × 986±107	3.829	Instance	0.819±0.033	Link
LIP [59]	CVPR'17	40462	19	231±132 × 325±122	2.426	Instance	0.861±0.027	Link
PennFudanPed [60]	ACCV'07	170	1	492±127 × 398±62	7.961	Binary	0.849±0.025	Link
Other Object Segmentation								
Trans10K [61]	ECCV'20	10428	2	2941±1203 × 3321±1366	4.286	Semantic	0.838±0.060	Link
LeedsButterfly [62]	BMVC'09	832	1	632±423 × 478±303	11.979	Binary	0.880±0.039	Link

Table 3. Supplemental dataset illustration (Part 1).

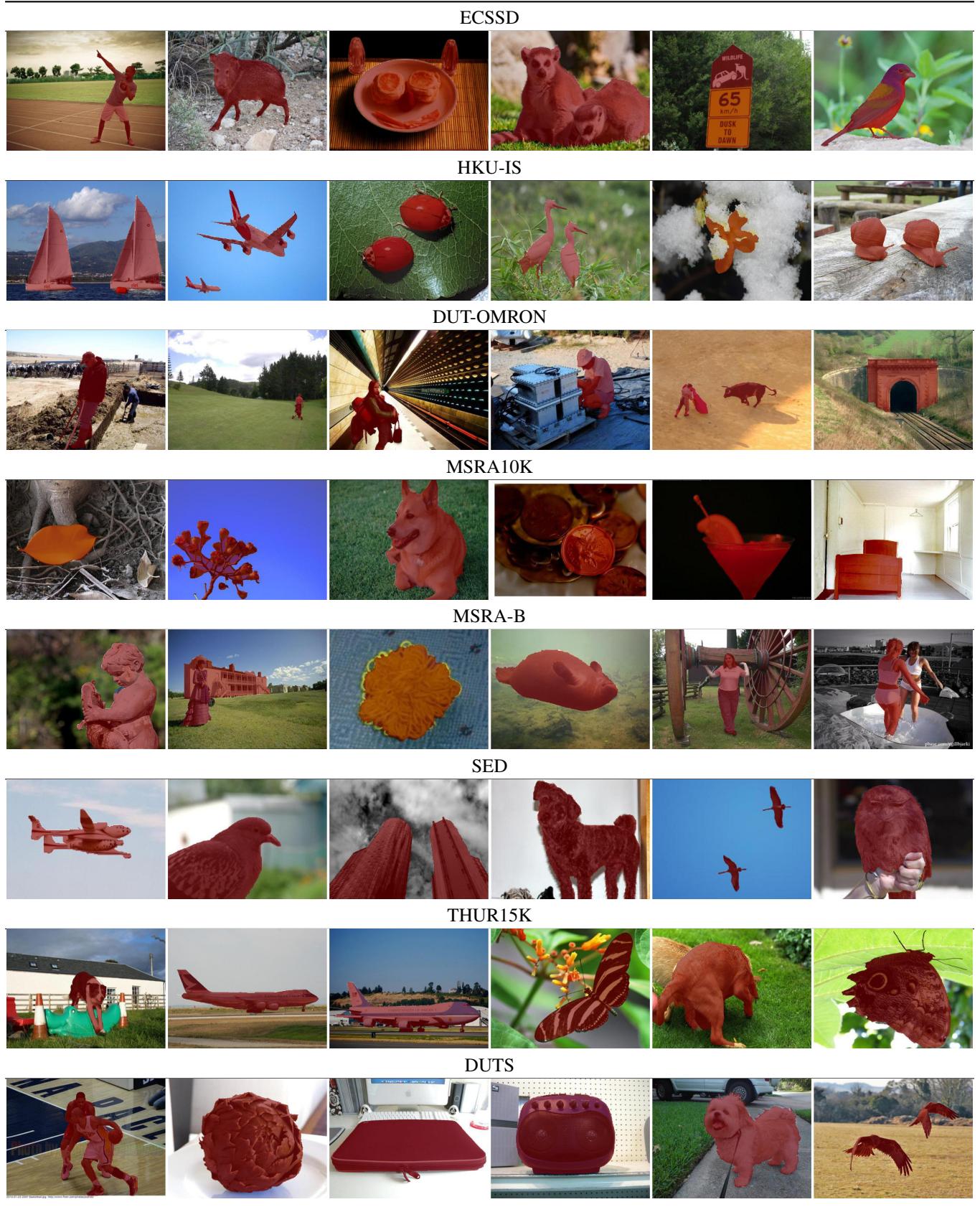


Table 4. Supplemental dataset illustration (Part 2).

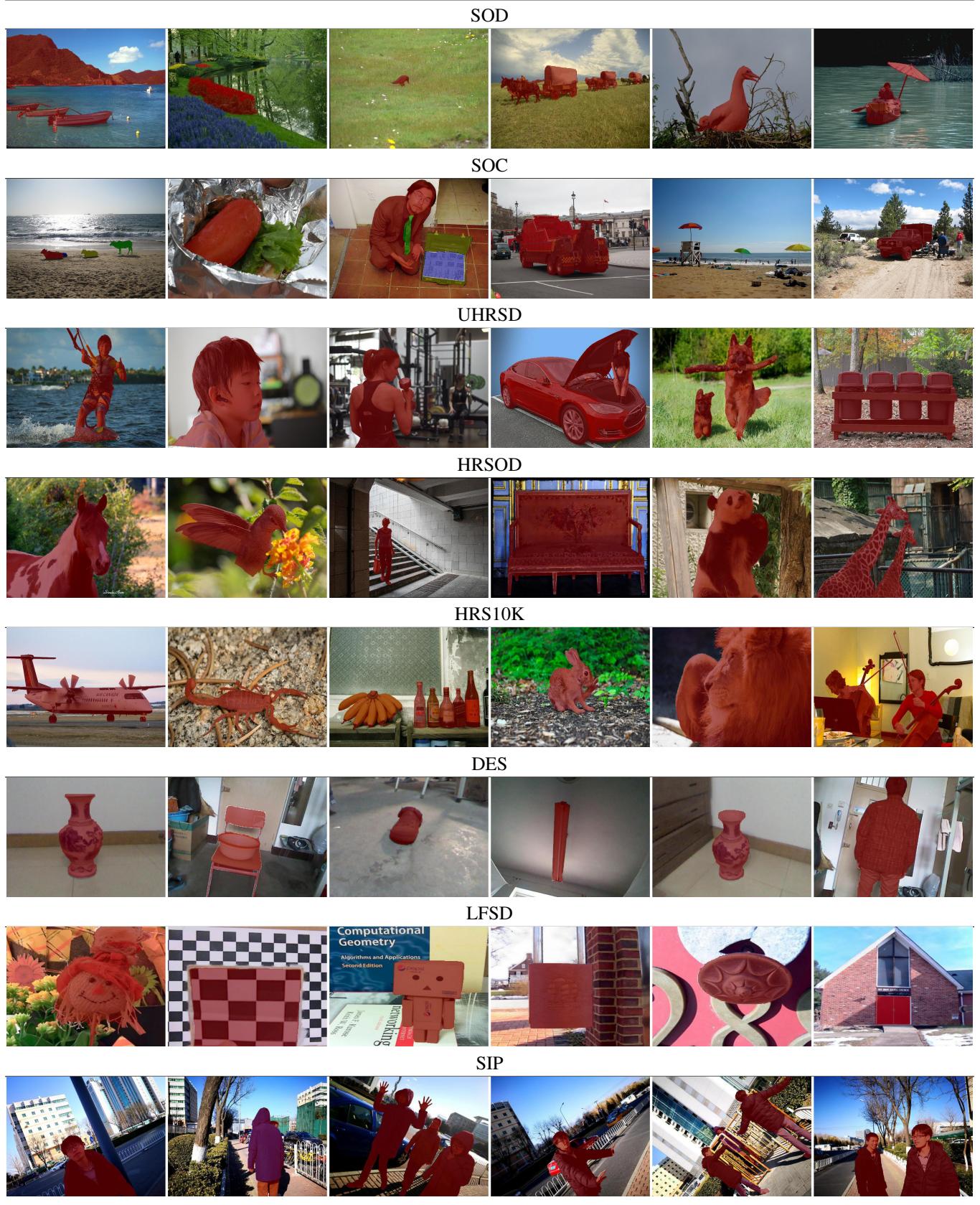


Table 5. Supplemental dataset illustration (Part 3).

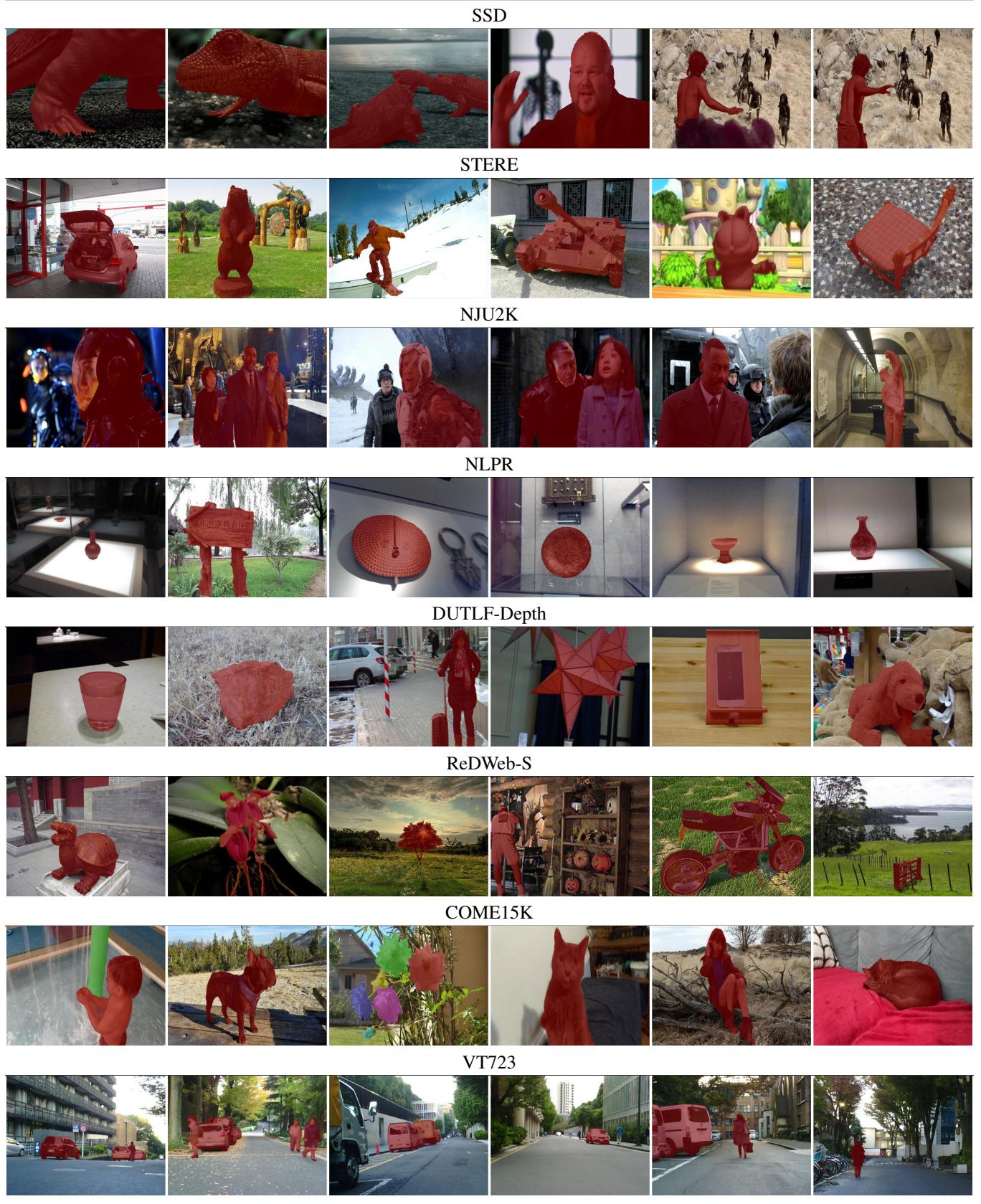


Table 6. Supplemental dataset illustration (Part 4).

VT821



VT1000



VT5000



CoCA



CoSal2015



CoSOD3k



iCoSeg



ImagePair

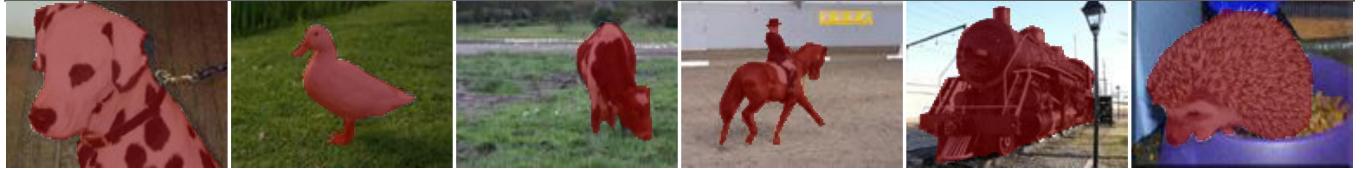


Table 7. Supplemental dataset illustration (Part 5).

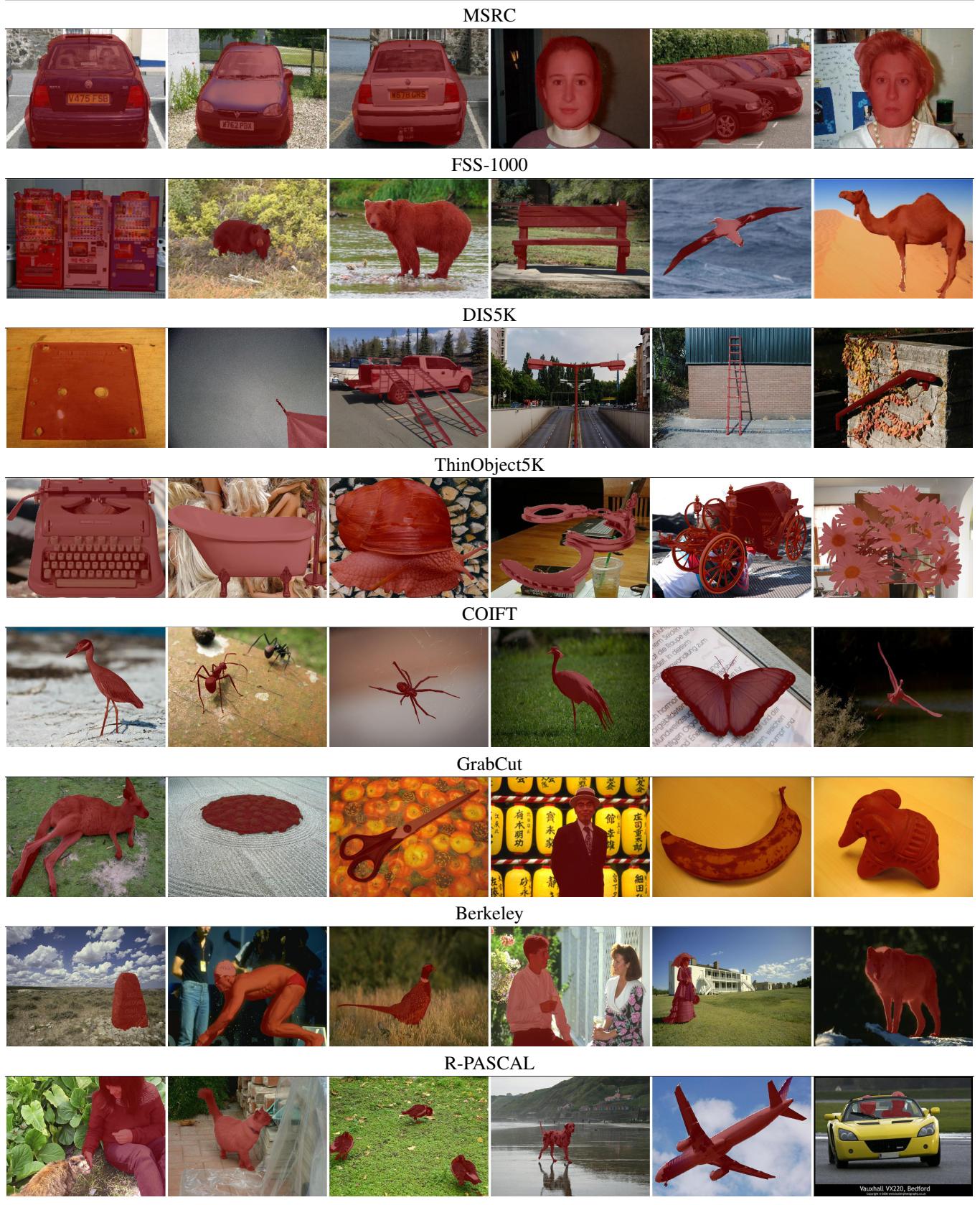


Table 8. Supplemental dataset illustration (Part 6).

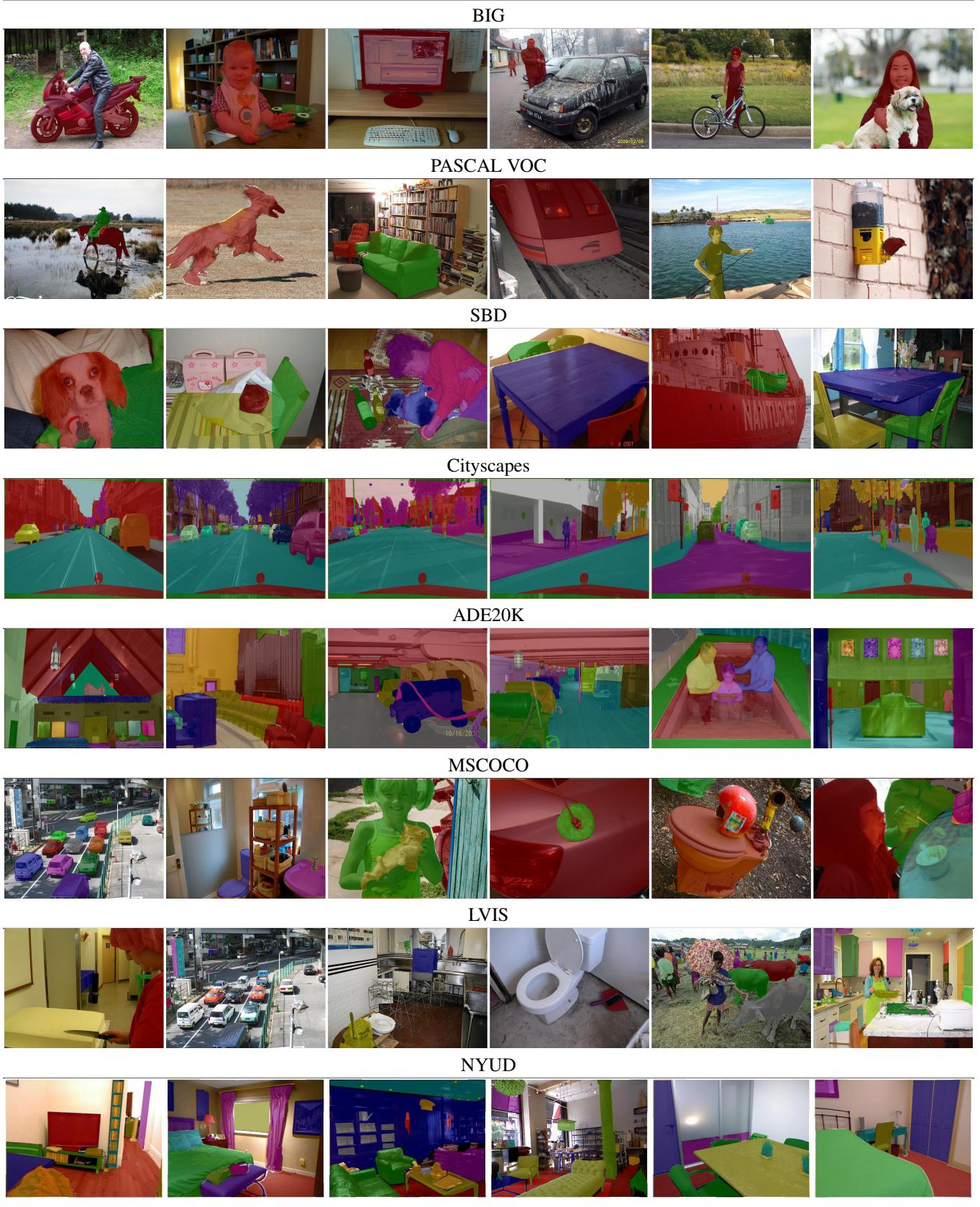


Table 9. Supplemental dataset illustration (Part 7).

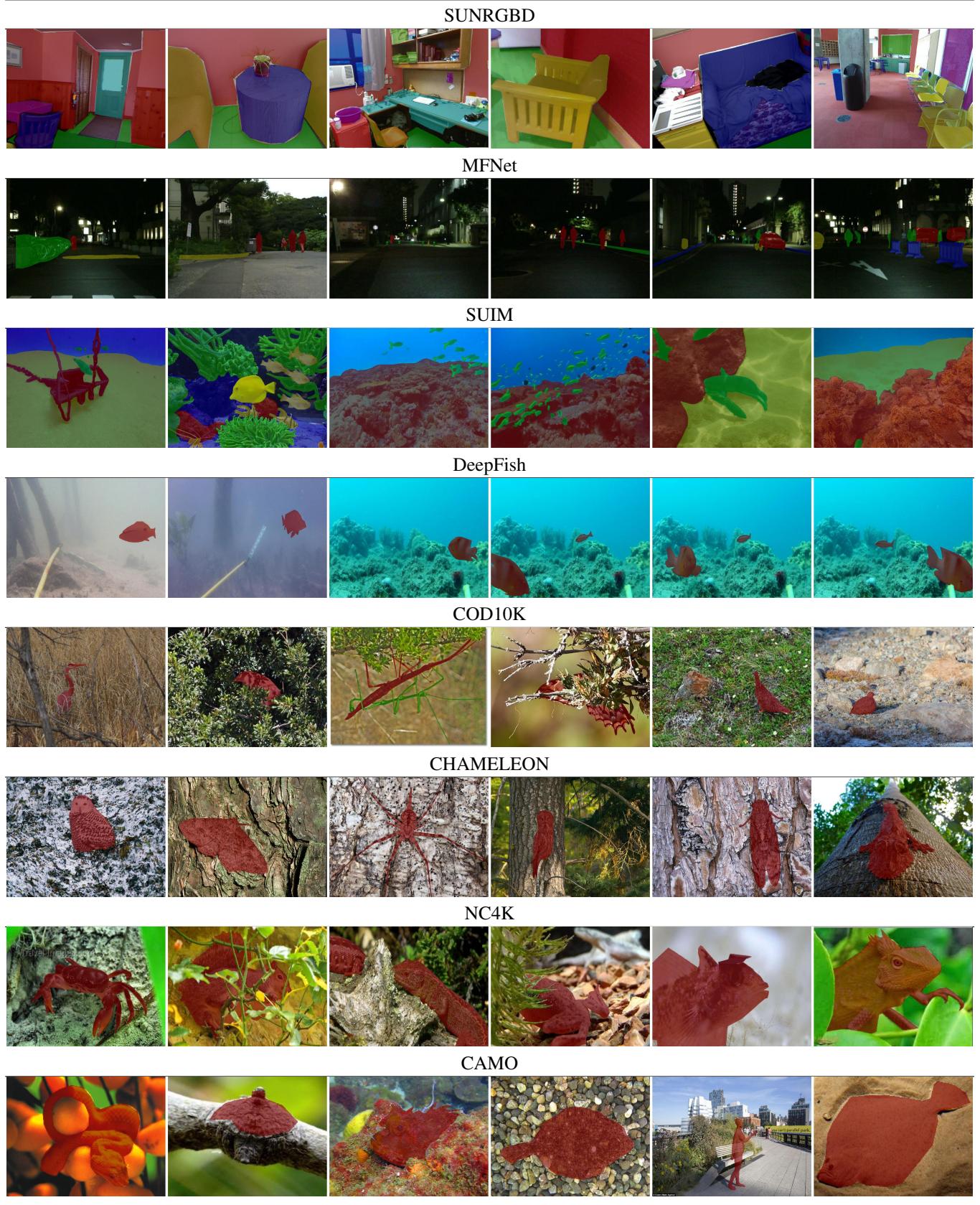


Table 10. Supplemental dataset illustration (Part 8).

DAVIS2017



YouTube-VOS



Fashionpedia



LIP



PennFudanPed



Trans10K



LeedsButterfly



2. REFERENCES

- [1] Qiong Yan, Li Xu, Jianping Shi, and Jiaya Jia, “Hierarchical saliency detection,” in *CVPR*, 2013. [1](#)
- [2] Guanbin Li and Yizhou Yu, “Visual saliency based on multiscale deep features,” in *CVPR*, 2015. [1](#)
- [3] Chuan Yang, Lihe Zhang, Huchuan Lu, Xiang Ruan, and Ming-Hsuan Yang, “Saliency detection via graph-based manifold ranking,” in *CVPR*, 2013. [1](#)
- [4] Tie Liu, Zejian Yuan, Jian Sun, Jingdong Wang, Nanning Zheng, Xiaou Tang, and Heung-Yeung Shum, “Learning to detect a salient object,” *IEEE TPAMI*, 2010. [1](#)
- [5] Jingdong Wang, Huaizu Jiang, Zejian Yuan, Ming-Ming Cheng, Xiaowei Hu, and Nanning Zheng, “Salient object detection: A discriminative regional feature integration approach,” *IJCV*, 2017. [1](#)
- [6] Ali Borji, Ming-Ming Cheng, Huaizu Jiang, and Jia Li, “Salient object detection: A benchmark,” *IEEE TIP*, 2015. [1](#)
- [7] Ming-Ming Cheng, Niloy J Mitra, Xiaolei Huang, and Shi-Min Hu, “Salientshape: group saliency in image collections,” *The Visual Computer*, 2014. [1](#)
- [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. [1](#)
- [9] Vida Movahedi and James H Elder, “Design and perceptual validation of performance measures for salient object segmentation,” in *CVPRW*, 2010. [1](#)
- [10] Deng-Ping Fan, Jing Zhang, Gang Xu, Ming-Ming Cheng, and Ling Shao, “Salient objects in clutter,” *IEEE TPAMI*, 2022. [1](#)
- [11] Chenxi Xie, Changqun Xia, Mingcan Ma, Zhirui Zhao, Xiaowu Chen, and Jia Li, “Pyramid grafting network for one-stage high resolution saliency detection,” in *CVPR*, 2022. [1](#)
- [12] Yi Zeng, Pingping Zhang, Jianming Zhang, Zhe Lin, and Huchuan Lu, “Towards high-resolution salient object detection,” in *ICCV*, 2019. [1](#)
- [13] Xinhao Deng, Pingping Zhang, Wei Liu, and Huchuan Lu, “Recurrent multi-scale transformer for high-resolution salient object detection,” in *ACM MM*, 2023. [1](#)
- [14] Yupeng Cheng, Huazhu Fu, Xingxing Wei, Jiangjian Xiao, and Xiaochun Cao, “Depth enhanced saliency detection method,” in *ICIMCS*, 2014. [1](#)
- [15] Nianyi Li, Jinwei Ye, Yu Ji, Haibin Ling, and Jingyi Yu, “Saliency detection on light field,” in *CVPR*, 2014. [1](#)
- [16] Deng-Ping Fan, Zheng Lin, Zhao Zhang, Menglong Zhu, and Ming-Ming Cheng, “Rethinking rgb-d salient object detection: Models, data sets, and large-scale benchmarks,” *IEEE TNNLS*, 2020. [1](#)
- [17] Chunbiao Zhu and Ge Li, “A three-pathway psychobiological framework of salient object detection using stereoscopic technology,” in *ICCVW*, 2017. [1](#)
- [18] Yuzhen Niu, Yujie Geng, Xueqing Li, and Feng Liu, “Leveraging stereopsis for saliency analysis,” in *CVPR*, 2012. [1](#)
- [19] Ran Ju, Ling Ge, Wenjing Geng, Tongwei Ren, and Gangshan Wu, “Depth saliency based on anisotropic center-surround difference,” in *ICIP*, 2014. [1](#)
- [20] Houwen Peng, Bing Li, Weihua Xiong, Weiming Hu, and Rongrong Ji, “Rgbd salient object detection: A benchmark and algorithms,” in *ECCV*, 2014. [1](#)
- [21] Yongri Piao, Wei Ji, Jingjing Li, Miao Zhang, and Huchuan Lu, “Depth-induced multi-scale recurrent attention network for saliency detection,” in *ICCV*, 2019. [1](#)
- [22] Nian Liu, Ni Zhang, Ling Shao, and Junwei Han, “Learning selective mutual attention and contrast for rgbd saliency detection,” *IEEE TPAMI*, 2021. [1](#)
- [23] Jing Zhang, Deng-Ping Fan, Yuchao Dai, Xin Yu, Yiran Zhong, Nick Barnes, and Ling Shao, “Rgbd saliency detection via cascaded mutual information minimization,” in *ICCV*, 2021. [1](#)
- [24] Xiurong Jiang, Lin Zhu, Yifan Hou, and Hui Tian, “Mirror complementary transformer network for rgbd-thermal salient object detection,” *arXiv preprint arXiv:2207.03558*, 2022. [1](#)
- [25] Guizhao Wang, Chenglong Li, Yunpeng Ma, Aihua Zheng, Jin Tang, and Bin Luo, “Rgbd saliency detection benchmark: Dataset, baselines, analysis and a novel approach,” in *IGTA*, 2018. [1](#)
- [26] Zhengzheng Tu, Tian Xia, Chenglong Li, Xiaoxiao Wang, Yan Ma, and Jin Tang, “Rgbd image saliency detection via collaborative graph learning,” *IEEE TMM*, 2019. [1](#)
- [27] Zhengzheng Tu, Yan Ma, Zhun Li, Chenglong Li, Jieming Xu, and Yongtao Liu, “Rgbd salient object detection: A large-scale dataset and benchmark,” *IEEE TMM*, 2022. [1](#)
- [28] Zhao Zhang, Wenda Jin, Jun Xu, and Ming-Ming Cheng, “Gradient-induced co-saliency detection,” in *ECCV*, 2020. [1](#)
- [29] Dingwen Zhang, Junwei Han, Chao Li, Jingdong Wang, and Xuelong Li, “Detection of co-salient objects by looking deep and wide,” *IJCV*, 2016. [1](#)

- [30] Deng-Ping Fan, Zheng Lin, Ge-Peng Ji, Dingwen Zhang, Huazhu Fu, and Ming-Ming Cheng, “Taking a deeper look at co-salient object detection,” in *CVPR*, 2020. [1](#)
- [31] Dhruv Batra, Adarsh Kowdle, Devi Parikh, Jiebo Luo, and Tsuhan Chen, “icoseg: Interactive co-segmentation with intelligent scribble guidance,” in *CVPR*, 2010. [1](#)
- [32] Hongliang Li and King Ngi Ngan, “A co-saliency model of image pairs,” *IEEE TIP*, 2011. [1](#)
- [33] John Winn, Antonio Criminisi, and Thomas Minka, “Object categorization by learned universal visual dictionary,” in *ICCV*, 2005. [1](#)
- [34] Xiang Li, Tianhan Wei, Yau Pun Chen, Yu-Wing Tai, and Chi-Keung Tang, “Fss-1000: A 1000-class dataset for few-shot segmentation,” in *CVPR*, 2020. [2](#)
- [35] Xuebin Qin, Hang Dai, Xiaobin Hu, Deng-Ping Fan, Ling Shao, and Luc Van Gool, “Highly accurate dichotomous image segmentation,” in *ECCV*, 2022. [2](#)
- [36] Jun Hao Liew, Scott Cohen, Brian Price, Long Mai, and Jiashi Feng, “Deep interactive thin object selection,” in *WACV*, 2021. [2](#)
- [37] Lucy AC Mansilla, Paulo AV Miranda, and Fábio AM Cappabianco, “Oriented image foresting transform segmentation with connectivity constraints,” in *ICIP*, 2016. [2](#)
- [38] Carsten Rother, Vladimir Kolmogorov, and Andrew Blake, ““ grabcut” interactive foreground extraction using iterated graph cuts,” *ACM TOG*, 2004. [2](#)
- [39] David Martin, Charless Fowlkes, Doron Tal, and Jitendra Malik, “A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics,” in *ICCV*, 2001. [2](#)
- [40] Ho Kei Cheng, Jihoon Chung, Yu-Wing Tai, and Chi-Keung Tang, “Cascadeps: Toward class-agnostic and very high-resolution segmentation via global and local refinement,” in *CVPR*, 2020. [2](#)
- [41] Mark Everingham, Luc Van Gool, Christopher KI Williams, John Winn, and Andrew Zisserman, “The pascal visual object classes (voc) challenge,” *IJCV*, 2010. [2](#)
- [42] Bharath Hariharan, Pablo Arbeláez, Lubomir Bourdev, Subhransu Maji, and Jitendra Malik, “Semantic contours from inverse detectors,” in *ICCV*, 2011. [2](#)
- [43] Marius Cordts, Mohamed Omran, Sebastian Ramos, Timo Rehfeld, Markus Enzweiler, Rodrigo Benenson, Uwe Franke, Stefan Roth, and Bernt Schiele, “The cityscapes dataset for semantic urban scene understanding,” in *CVPR*, 2016. [2](#)
- [44] Bolei Zhou, Hang Zhao, Xavier Puig, Sanja Fidler, Adela Barriuso, and Antonio Torralba, “Scene parsing through ade20k dataset,” in *CVPR*, 2017. [2](#)
- [45] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick, “Microsoft coco: Common objects in context,” in *ECCV*, 2014. [2](#)
- [46] Agrim Gupta, Piotr Dollar, and Ross Girshick, “Lvis: A dataset for large vocabulary instance segmentation,” in *CVPR*, 2019. [2](#)
- [47] Nathan Silberman, Derek Hoiem, Pushmeet Kohli, and Rob Fergus, “Indoor segmentation and support inference from RGBD images,” in *ECCV*, 2012. [2](#)
- [48] Shuran Song, Samuel P. Lichtenberg, and Jianxiong Xiao, “SUN RGB-D: A RGB-D scene understanding benchmark suite,” in *CVPR*, 2015. [2](#)
- [49] Qishen Ha, Kohei Watanabe, Takumi Karasawa, Yoshitaka Ushiku, and Tatsuya Harada, “Mfnet: Towards real-time semantic segmentation for autonomous vehicles with multi-spectral scenes,” in *IROS*. IEEE, 2017. [2](#)
- [50] Md Jahidul Islam, Chelsey Edge, Yuyang Xiao, Peigen Luo, Muntaqim Mehtaz, Christopher Morse, Sadman Sakib Enan, and Junaed Sattar, “Semantic segmentation of underwater imagery: Dataset and benchmark,” in *IROS*, 2020. [2](#)
- [51] Alzayat Saleh, Issam H Laradji, Dmitry A Konovalov, Michael Bradley, David Vazquez, and Marcus Sheaves, “A realistic fish-habitat dataset to evaluate algorithms for underwater visual analysis,” *Scientific Reports*, 2020. [2](#)
- [52] Deng-Ping Fan, Ge-Peng Ji, Guolei Sun, Ming-Ming Cheng, Jianbing Shen, and Ling Shao, “Camouflaged object detection,” in *CVPR*, 2020. [2](#)
- [53] Przemysław Skurowski, Hassan Abdulameer, J Błaszczyk, Tomasz Depta, Adam Kornacki, and P Kozieł, “Animal camouflage analysis: Chameleon database,” *Unpublished manuscript*, 2018. [2](#)
- [54] Yunqiu Lv, Jing Zhang, Yuchao Dai, Aixuan Li, Bowen Liu, Nick Barnes, and Deng-Ping Fan, “Simultaneously localize, segment and rank the camouflaged objects,” in *CVPR*, 2021. [2](#)
- [55] Trung-Nghia Le, Tam V Nguyen, Zhongliang Nie, Minh-Triet Tran, and Akihiro Sugimoto, “Anabanch network for camouflaged object segmentation,” *CVIU*, 2019. [2](#)
- [56] Jordi Pont-Tuset, Federico Perazzi, Sergi Caelles, Pablo Arbeláez, Alex Sorkine-Hornung, and Luc Van Gool, “The 2017 davis challenge on video object segmentation,” *arXiv preprint arXiv:1704.00675*, 2017. [2](#)

- [57] Ning Xu, Linjie Yang, Yuchen Fan, Dingcheng Yue, Yuchen Liang, Jianchao Yang, and Thomas Huang, “Youtube-vos: A large-scale video object segmentation benchmark,” *arXiv preprint arXiv:1809.03327*, 2018. [2](#)
- [58] Menglin Jia, Mengyun Shi, Mikhail Sirotenko, Yin Cui, Claire Cardie, Bharath Hariharan, Hartwig Adam, and Serge Belongie, “Fashionpedia: Ontology, segmentation, and an attribute localization dataset,” in *ECCV*, 2020. [2](#)
- [59] Ke Gong, Xiaodan Liang, Dongyu Zhang, Xiaohui Shen, and Liang Lin, “Look into person: Self-supervised structure-sensitive learning and a new benchmark for human parsing,” in *CVPR*, 2017. [2](#)
- [60] Liming Wang, Jianbo Shi, Gang Song, and I-fan Shen, “Object detection combining recognition and segmentation,” in *ACCV*, 2007. [2](#)
- [61] Enze Xie, Wenjia Wang, Wenhai Wang, Mingyu Ding, Chunhua Shen, and Ping Luo, “Segmenting transparent objects in the wild,” in *ECCV*, 2020. [2](#)
- [62] Josiah Wang, Katja Markert, and Mark Everingham, “Learning models for object recognition from natural language descriptions.,” in *BMVC*, 2009. [2](#)