Salient Object Detection: A Benchmark Supplementary Material

Ali Borji, Dicky N. Sihite, and Laurent Itti

Department of Computer Science, University of Southern California
http://ilab.usc.edu

1 Introduction

This supplementary material accompanies the paper "Salient Object Detection: A Benchmark" at ECCV 2012. It provides more details on comparison of saliency detection models. It also contains some illustrative figures.

In this work, as the initial seed, we focus on those models that are easily accessible, attained good accuracies, or have been highly referred. Software for some models was already available online. For others, we contacted their creators for the code; the authors then either sent us the source code to compile or the executables. Some authors, however, preferred to run their models on our stimuli and hence send us back the saliency maps. In order to achieve a thorough model comparison, we intend to open an online challenge where modelers could contribute by submitting their results.

We compare three categories of models: 1) those aiming to detect and segment the most salient object in a scene (emphasized more here), 2) active segmentation approaches, and 3) models that address fixation prediction. Table 1 shows the list of models from the first two categories, and table 2 shows category 3.

Fig. 1 shows accuracy of a combined model built from the best two models (here CBSal and SVO) versus the combined model built from the three best models (RC, SVO, and CBSal). As it can be seen there is not much difference between two types of combinations (best two and best three) indicating effectiveness of our combination approach.

Fig. 2 shows precision-recall curves of all models (from both categories) over small (ascending) and large (descending) objects for all datasets mentioned in the paper.

Fig. 3 shows all four scores of models over ASD dataset for small (ascending) and large (descending) objects.

Fig. 4 shows precision-recall curves of models over least consistent and most consistent images for all datasets with multiple annotations (all except ASD which has only one annotation). Since SED datasets have 100 images, we sorted stimuli for each dataset (SED1 and SED2) and calculated scores for first and last 50 images.

Fig. 5 shows all four scores of models with saliency maps smoothed with two Gaussian sizes (5 and 15). G=0 means no Gaussian smoothing (original saliency maps). Fig. 6 shows F-measure and AUC scores of models with three

#	**Acronym** (Model)	**Ref.**	**Pub/Year**	**Code**	**Resolution**	**DB**	Avl.
1	**IO**: Inter Observer model	-	-	M	$w \times h$	All	✓
2	**MAP**: Mean Annotation Map	-	-	M	500×500	All	✓
3	**MZ**: Ma and Zhang	[52]	ACM-M/2003	S	$w \times h$	ASD	✓
4	**LC**: Zhai and Shah	[18]	ACM-M/2006	C	$w \times h$	All	✓
5	**salLiu**: Liu *et al.*	[33]	CVPR/2007	M	max 200	All	✓
6	**AC**: Achanta *et al.*	[14]	ICVS/2008	M	$w \times h$	All	✓
7	**MSSS**: Achanta and Sušstrunk	[56]	ICIP/2009	M	$w \times h$	All	✓
8	**FTS**: Achanta *et al.*	[16]	CVPR/2009	M	$w \times h$	All	✓
9	**EDS**: Rosin	[19]	PR/2009	C	$w \times h$	All	✓
10	Gopalakrishnan *et al.*	[34]	CVPR/2009	-	-	-	-
11	Marchesotti *et al.*	[35]	ICCV/2009	-	-	-	-
12	**Valenti**: Valenti *et al.*	[40]	ICCV/2009	S	$w \times h$	ASD/MSRA	✓
13	**Goferman**: Goferman *et al.*	[15]	CVPR/2010	M	max 250	All	✓
14	**PMehrani**: Mehrani and Veksler	[23]	BMVC/2010	S	$w \times h$	ASD/SED1	✓
15	Rahtu *et al.*	[29]	ECCV/2010	-	-	-	-
16	Khuwuthyakorn *et al.*	[28]	ECCV/2010	-	-	-	-
17	Zhang *et al.*	[21]	IEEE TOM/2010	-	-		-
18	**JiaLiSal**: Jia Li *et al.*	[36]	IJCV/2010	S	$(w \times h)/16$	ASD/MSRA	✓
19	**LiuICIP**: Liu *et al.*	[54]	ICIP/2010	S	$w \times h$	ASD	✓
20	**MichalGazit**: Gazit *et al.*	[37]	ECCV-W/2010	M	$w \times h$	All	✓
21	**DAKlein**: Klein and Frintrop	[25]	ICCV/2011	S	$w \times h$	All	✓
22	**MengW**: M. Wang *et al.*	[18]	CVPR/2011	S	$w \times h$	ASD	✓
23	Feng *et al.*	[22]	ICCV/2011	-	-		-
24	Deng and Luo	[39]	OE/2011	-	-	-	-
25	Lu *et al.*	[24]	ICCV/2011	-	-	-	-
26	L. Wang *et al.*	[26]	ICCV/2011	-	-	-	-
27	**SVO**: Chang *et al.*	[27]	ICCV/2011	M	$w \times h$	All	✓
28	**CBsal**: Jiang *et al.*	[31]	BMVC/2011	M	$w \times h$	All	✓
29	**RC**: M.M. Cheng *et al.*	[13]	CVPR/2011	C	$w \times h$	All	✓
30	**HC**: M.M. Cheng *et al.*	[13]	CVPR/2011	C	$w \times h$	All	✓
31	**Materias**: Li *et al.*	[36]	BMVC/2011	M	$w \times h$	All	✓
32	**LiuIETIP**: Liu *et al.*	[43]	IEEE TIP/2011	S	$w \times h$	ASD	✓
33	**Mishra**: Mishra *et al.*	[50]	PAMI/2011	C	$w \times h$	All	✓
34	**SRS1**: Siagian and Koch	[51]	Submitted.	C	$w \times h$	All	✓

Table 1. Compared salient object detection models (checked) sorted chronologically. Abbreviations: {M: Matlab, C: C/C++, E: Executables, S: Sent saliency maps}. w and h: image width/height.

#	**Acronym** (Model)	**Ref.**	**Pub/Year**	**Code**	**Resolution**	**DB**	Avl.
1	**ITTI**: Itti *et al.*	[2]	PAMI/1998	C	$w/16 \times h/16$	All	✓
2	**ITTI98**: Itti *et al.* (maxNorm)	[2]	PAMI/1998	C	$w/16 \times h/16$	All	✓
3	**AIM**: Bruce and Tsotsos	[4]	NIPS/2005	M	$w/2 \times h/2$	All	✓
4	**GBVS**: Harel *et al.*	[3]	NIPS/2006	M	$w \times h$	All	✓
5	**HouCVPR**: Hou and Zhang	[5]	CVPR/2007	M	64×64	All	✓
6	**HouNIPS**: Hou and Zhang	[6]	NIPS/2008	M	$w \times h$	All	✓
7	**SUN**: Zhang *et al.*	[10]	JoV/2008	M	$w/2 \times h/2$	All	✓
8	**PQFT**: Guo and Zhang	[57]	TiP/2009	M	400×400	All	✓
9	**SEO**: Seo and Milanfar	[8]	JoV/2009	M	$w \times h$	All	✓
10	**AWS**: Diaz *et al.*	[7]	ACIVS/2009	M	$w/2 \times h/2$	All	✓
11	**Judd**: Judd *et al.*	[1]	ICCV/2009	M	$w \times h$	All	✓

Table 2. Compared saliency models originally developed for eye fixation prediction.

cases: 1) non-smoothed saliency maps (Gauss = 0), 2) saliency maps smoothed with Gauss = 5, and 3) maps smoothed with Gauss = 15. While there are slight changes in scores across different Gaussian sizes, model rankings and thus our qualitative conclusions in the paper stay the same (are not affected).

Fig. 7 shows results of our analysis of model similarity over all datasets. First column shows the 2D space layout for models. Please note what matters here is the distance between models and not exact locations. Second column shows the amount of variance in data that could be explained by retaining x dimensions (x-axis). Third column shows the 3D similarity representation. Fourth column shows similarity matrix among models.

Fig. 8 shows consistency analysis (for 100 least and 100 most consistent images in terms of annotation) over the MSRA dataset.

Fig. 9 shows analysis of object size over model scoring for the MSRA dataset.

Fig. 10 shows consistency analysis results over the SED2 dataset (50 most and 50 least).

Fig. 11 shows object-size analysis over SED2 dataset (for 50 smallest and 50 biggest objects)..

Fig. 12 shows object-size analysis over the SED1 single-object dataset (for 50 smallest and 50 biggest objects).

Fig. 13 shows consistency analysis over SOD dataset. There are 300 images in SOD dataset and we chose 100 least and 100 most consistent images in terms of annotation.

Fig. 14 shows object-size analysis over the SOD (for 100 smallest and 100 biggest objects).

Fig. 15 shows Gaussian smoothing of saliency maps for six models (CBsal, SVO, HC, RC, SRS1, and Mishra). Results are shown for all four scores (PR, F-measure, ROC, and AUC) over the ASD dataset.

Fig. 16 show the 5 easiest and 5 most difficult stimuli for 11 best models according to the AUC score for images of ASD dataset. Figs. 17 and 18 show human annotations and saliency maps of models, respectively.

References

1. T. Judd, K. Ehinger, F. Durand and, A. Torralba. Learning to predict where humans look. *ICCV*, 2009.
2. L. Itti, C. Koch, and E. Niebur. A model of saliency-based visual attention for rapid scene analysis. *IEEE Trans. PAMI*, 20(11):1254-1259, 1998.
3. J. Harel, C. Koch, and P. Perona. Graph-based visual saliency. *NIPS*, 2006.
4. N.D.B. Bruce and J.K. Tsotsos. Saliency based on information maximization. *NIPS*, 2005.
5. X. Hou and L. Zhang. Saliency detection: A spectral residual approach. *CVPR*, 2007.
6. X. Hou and L. Zhang. Dynamic attention: Searching for coding length increments. *NIPS*, 2008.

7. A. Garcia-Diaz, X. R. Fdez-Vidal, X. M. Pardo, and R. Dosil. Decorrelation and distinctiveness provide with human-like saliency. *ACIVS*, 5807, 2009.
8. H.J. Seo and P. Milanfar. Static and space-time visual saliency detection by self-resemblance. *Journal of Vision*, 9:1-27, 2009.
9. D. Walther and C. Koch. Modeling attention to salient proto-objects. *Neural Net.*, 2006.
10. L. Zhang, M. H. Tong, T. K. Marks, H. Shan, and G. W. Cottrell. SUN: A Bayesian framework for saliency using natural statistics. *JOV*, 2008.
11. B.W. Tatler. The central fixation bias in scene viewing: selecting an optimal viewing position independently of motor bases and image feature distributions. *J. Vision*, 14(7): 2007.
12. Y. Zhai and M. Shah. Visual attention detection in video sequences using spatiotemporal cues. *ACM Multimedia*, 2006.
13. M.M. Cheng, G.X. Zhang, N.J. Mitra, Xiaolei Huang, and S.M. Hu. Global contrast based salient region detection. *CVPR*, 2011.
14. R. Achanta, F. Estrada, P. Wils, and S. Süsstrunk. Salient region detection and segmentation. *ICVS*, 2008.
15. S. Goferman, L. Zelnik-Manor, and A. Tal. Context-aware saliency detection. *CVPR*, 2010.
16. R. Achanta, S. Hemami, F. Estrada, and S. Süsstrunk. Frequency-tuned salient region detection. *CVPR*, 2009.
17. J. Wang, J. Sun, L. Quan, X. Tang, and H.Y Shum. Picture collage. *CVPR*, 2006.
18. M. Wang, J. Konrad, P. Ishwar, Y. Jing, and H. Rowley. Image saliency: from intrinsic to extrinsic context. *CVPR*, 2011.
19. P.L. Rosin. A simple method for detecting salient regions. *Pattern Rec.*, 2009.
20. S. Goferman, A. Tal, L. Zelnik-Manor. Puzzle-like collage. *EuroGraphics*, 2010.
21. W. Zhang, Q.M.J. Wu, G. Wang, and H. Yin. An adaptive computational model for salient object detection. *IEEE Trans. On Multimedia*, 12(4), 2010.
22. J. Feng, Y. Wei, L. Tao, C. Zhang, and J. Sun. *ICCV*, 2011.
23. P. Mehrani and O. Veksler.Saliency segmentation based on learning and graph cut.*BMVC*,2010.
24. Y. Lu,W. Zhang,H. Lu,and X. Xue. Salient object detection using concavity context.*ICCV*,2011.
25. D.A. Klein and S. Frintrop. Center-surround divergence of feature statistics for salient object detection. *ICCV*, 2011.
26. L. Wang, J. Xue, N. Zheng, and G. Hua. Automatic Salient object extraction with contextual cue. *ICCV*, 2011.
27. K.Y. Chang, T.L Liu, H.T. Chen, and S.H. Lai. Fusing generic objectness and visual saliency for salient object detection. *ICCV*, 2011.
28. P. Khuwuthyakorn, A. Robles-Kelly, and J. Zhou. *ECCV*, 2010.
29. E. Rahtu, J. Kannala, M. Salo, and J. Heikkil a. Segmenting salient objects from images and videos. *ECCV*, 2010.
30. C. Kanan and G. Cottrell. Robust classification of objects, faces, and flowers using natural image. *CVPR*, 2010.
31. H. Jiang, J. Wang, Z. Yuan, T. Liu, N. Zheng, and S. Li. Automatic salient object segmentation based on context and shape prior. *BMVC*, 2011.

32. J. Li, Y. Tian, T. Huang, and W. Gao. A dataset and evaluation methodology for visual saliency in video. *Int. Conf. on Multimedia and Expo*, 442-445, 2009.
33. T. Liu, J. Sun, N Zheng, X. Tang, and H. Shum.Learning to detect a salient object. *CVPR*,2007.
34. V. Gopalakrishnan, Y. Hu, and D. Rajan. *CVPR*, 2009.
35. L. Marchesotti, C. Cifarelli, and G. Csurka. A framework for visual saliency detection with applications to image thumbnailing. *ICCV*, 2009.
36. J. Li, M.D. Levine, X. An, and H. He. Saliency detection based on frequency and spatial domain analysis. *BMVC*, 2011.
37. M. Holtzman-Gazit, L. Zelnik-Manor, and I. Yavneh. Salient edges: A multi scale approach. *ECCV, Workshop on Vision for Cognitive Tasks, 2010.*
38. Y. Luo, J. Yuan, P. Xue, and Q. Tian. Saliency density maximization for object detection and localization. *ACCV*, 2010.
39. Q. Deng and Y. Luo. Edge-based method for detecting salient objects. *Opt. Eng.*, 50, 2011.
40. R. Valenti, N. Sebe, and T. Gevers. Image saliency by isocentric curvedness and color. *ICCV*, 2009.
41. T.Z. Huang, K.Y. Cheng, and Y.Y. Chuang. A collaborative benchmark for region of interest detection algorithms. *CVPR*, 2009.
42. Q. Zhang, H. Liu, J. Shen, G. Gu, and H. Xiao. An improved computational approach for salient region detection. *Journal of Computers*, 2010.
43. H. Li and K.N. Ngan. A co-saliency model of image pairs. *IEEE Trans. Image Process.*, 2011.
44. F. Ge and S. Wang. New benchmark for image segmentation evaluation. *Journal of Electronic Imaging*, 16(3), 2007.
45. F.J. Estrada and A.D. Jepson. Benchmarking image segmentation algorithms. *IJCV*, 2009.
46. C.O. Ancuti, C. Ancuti, and P. Bekaert. *CVPR*, 2011.
47. V. Movahedi and J. H. Elder, Design and perceptual validation of performance measures for salient object segmentation. *POCV*, 2010.
48. S. Alpert, M. Galun, R. Basri, and A. Brandt. Image segmentation by probabilistic bottom-up aggregation and cue integration. *CVPR*, 2007.
49. F. Ge, S. Wang, and T. Liu, Image-segmentation evaluation from the perspective of salient object extraction. *CVPR*, 2006.
50. A. K. Mishra, Y. Aloimonos, C. L. Fah, and A. Kassim. Active visual segmentation. *IEEE Trans. PAMI*, 2011.
51. C. Siagian and C. Koch, Salient segmentation using contours and region growing. *Submitted*.
52. Y.F. Ma and H.J. Zhang, Contrast-based image attention analysis by using fuzzy growing. *ACM Multimedia*, 374-381, 2003.
53. Z. Liu, Y. Xue, H. Yan, and Z. Zhang, Efficient saliency detection based on Gaussian models. *IET Image Processing*, 5(2): 122-131, 2011.
54. Z. Liu, Y. Xue, L. Shen, and Z. Zhang. Nonparametric saliency detection using kernel density estimation. *ICIP*, 253-256, 2010
55. J. Li, Y. Tian, T. Huang, and W. Gao. Probabilistic multi-task learning for visual saliency estimation in video. *IJCV*, 90(2):150-165, 2010.

56. R. Achanta and S. Susstrunk. Saliency detection for content-aware image resizing. *ICIP*, 2009.
57. C. Guo and L. Zhang. A novel multiresolution spatiotemporal saliency detection model and Its applications in image and video compression. *IEEE Trans. on Image Processing*, 2010.
58. T.H Huang, K.Y Cheng, and Y.Y Chuang, A collaborative benchmark for region of interest detection algorithms. *CVPR*, 2009.
59. C.M Masciocchi, S. Mihalas, D. Parkhurst, E. Niebur. Everyone knows what is interesting: salient locations which should be fixated. *Journal of Vision*, 2009.
60. L. Itti. Automatic Foveation for Video Compression using a neurobiological model of visual attention. *IEEE Trans. Image Process.*, 2004.
61. Y. Ma, X. Hua, L. Lu, and H. Zhang. A generic framework of user aattention model and its application in video summarization. *IEEE trans. Multimedia*, 2005.
62. A. Toet. Computational versus psychophysical image saliency: a comparative evaluation study. *IEEE Trans. Pattern Anal. Mach. Intell.*, 2011.
63. D. Gao and N. Vasconcelos. Discriminant saliency for visual recognition from cluttered scenes. *NIPS*, 2004.
64. W. Kienzle, M.O. Franz, B. Schölkopf, and F.A. Wichmann. Center-surround patterns emerge as optimal predictors for human saccade targets. *J. Vision*, 2009.
65. T. Judd. Understanding and predicting where people look, Phd Thesis, MIT, 2011.
66. L. Itti, N. Dhavale, and F. Pighin. *SPIE*, 2003.
67. K. Koehler, F. Guo, S. Zhang, M. Eckstein. *Vision Science Society*, 2011.
68. Authors. Quantitative analysis of human-model agreement in visual saliency modeling: a comparative study. *Under review.*

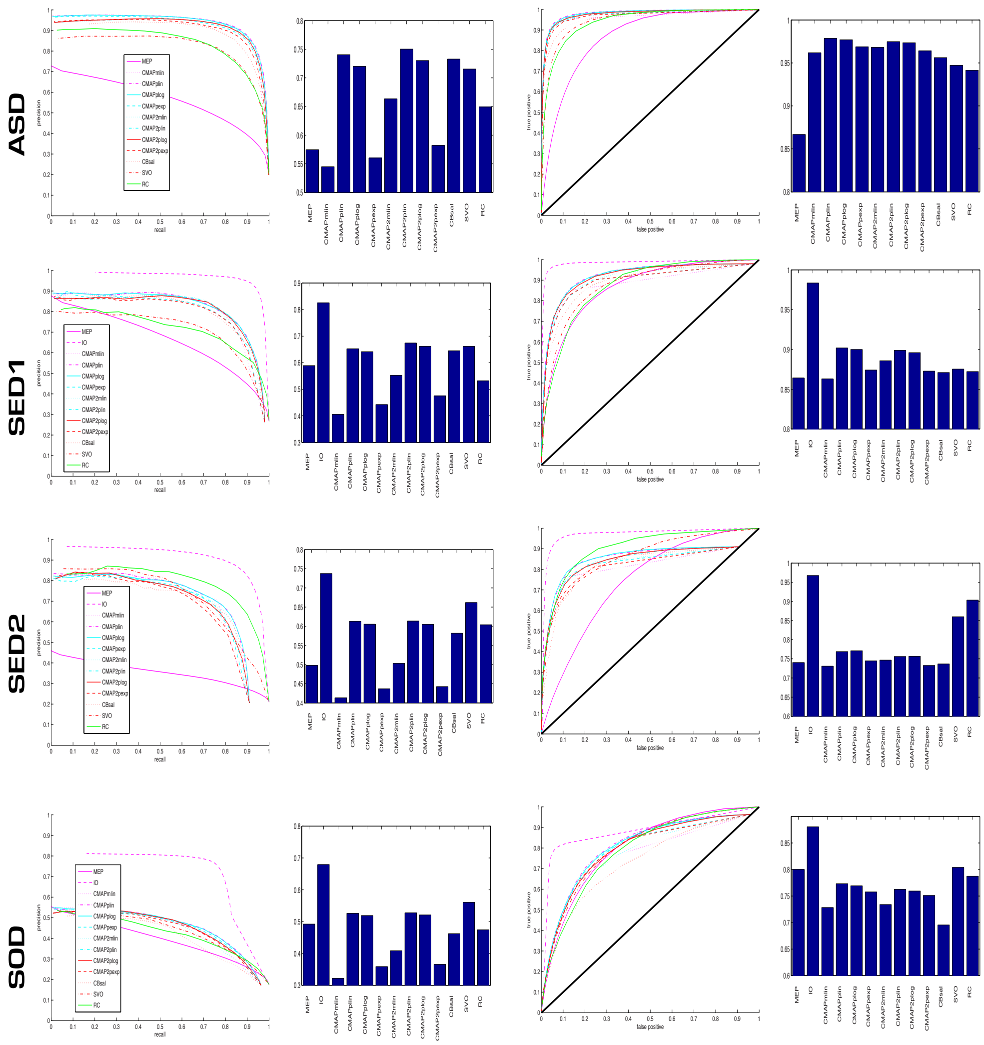

Fig. 1. Accuracy of models combined from the best two models (CBSal and SVO) versus models combined of (RC, SVO, and CBSal) over all datasets.

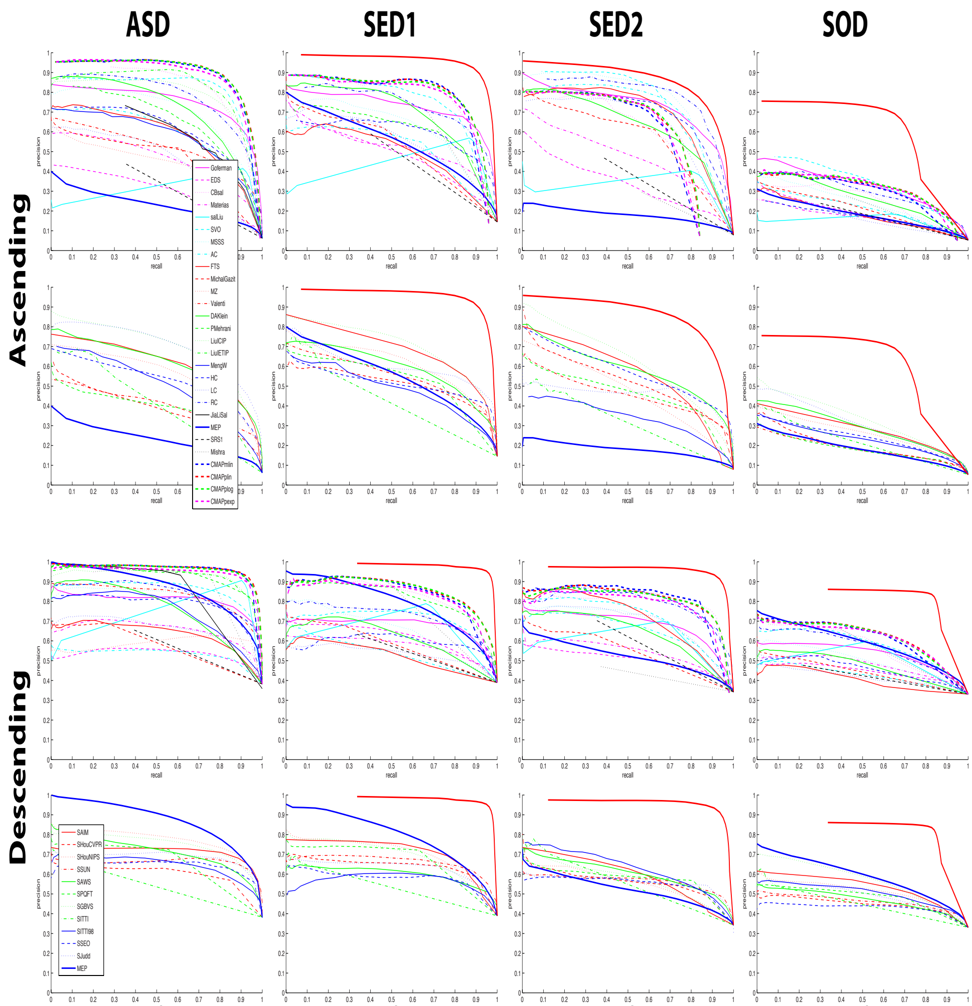

Fig. 2. Accuracy of models over small (ascending) and large (descending) objects for all datasets. First row for each dataset shows accuracy of object segmentation models and second row shows saliency estimation models.

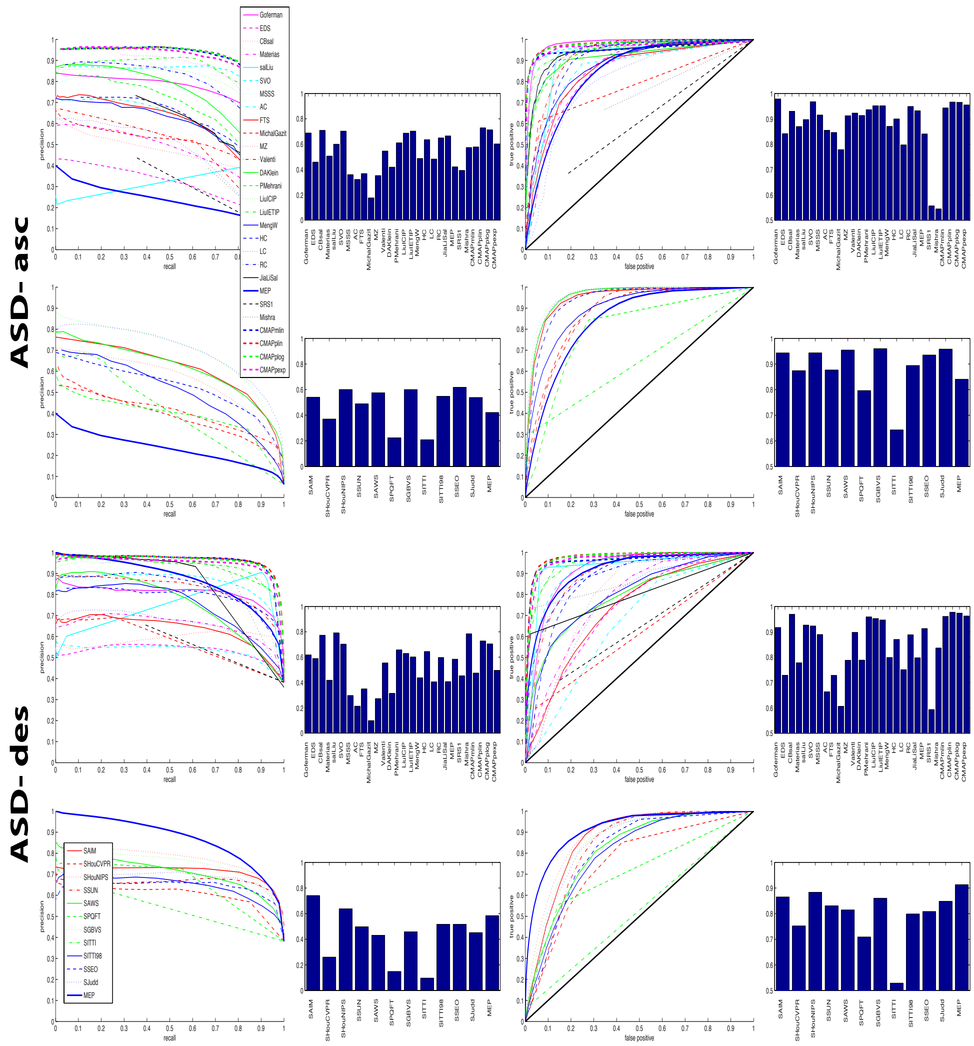

Fig. 3. All four scores of models over ASD dataset for small (ascending) and large (descending) objects.

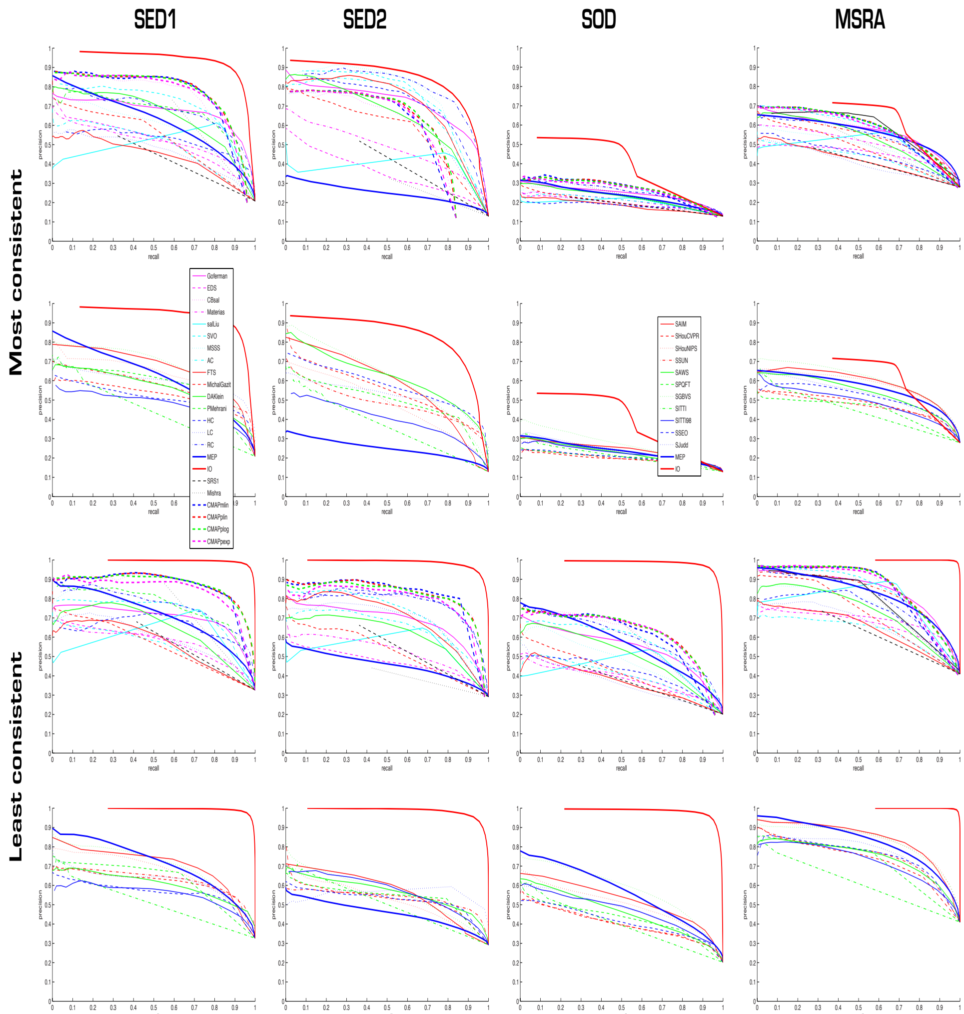

Fig. 4. Model accuracies (PR) for models over least and most consistent images foe each dataset.

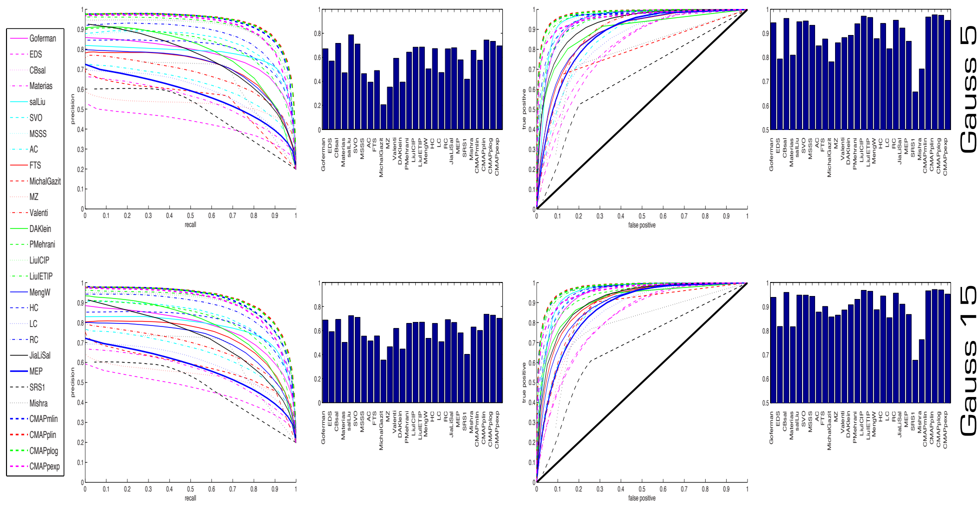

Fig. 5. Accuracy of models with smoothed (convolved) saliency maps with Gaussian kernels of two sizes.

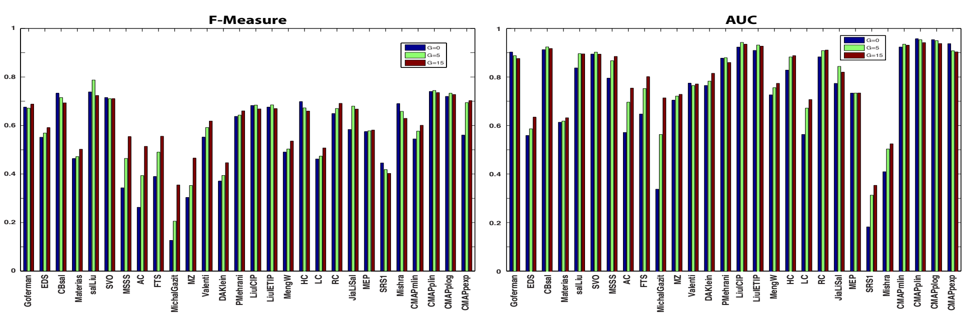

Fig. 6. F-measure (left) and AUC (right) scores of smoothed saliency maps of models.

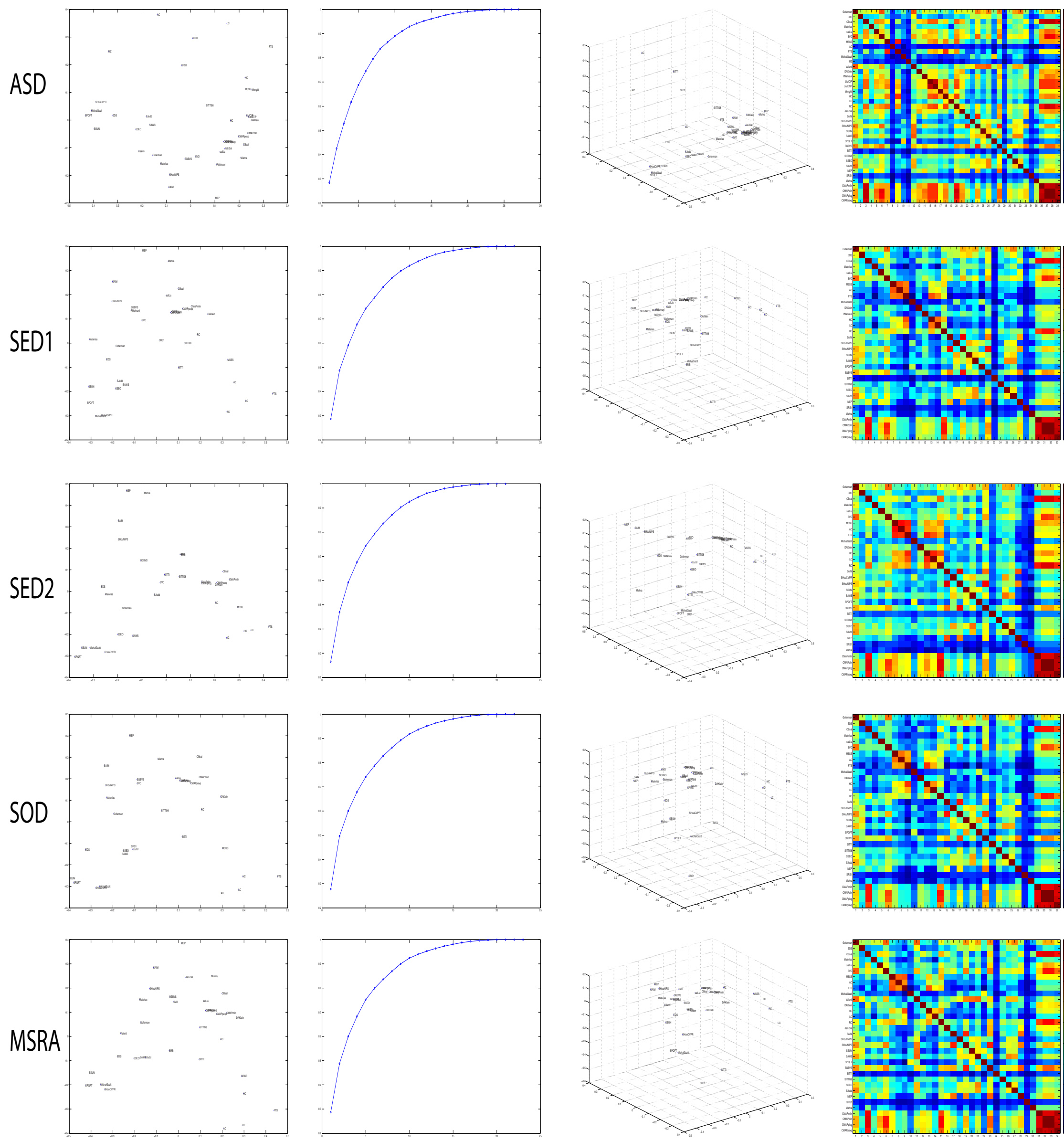

Fig. 7. Analysis of model similarity over all datasets.

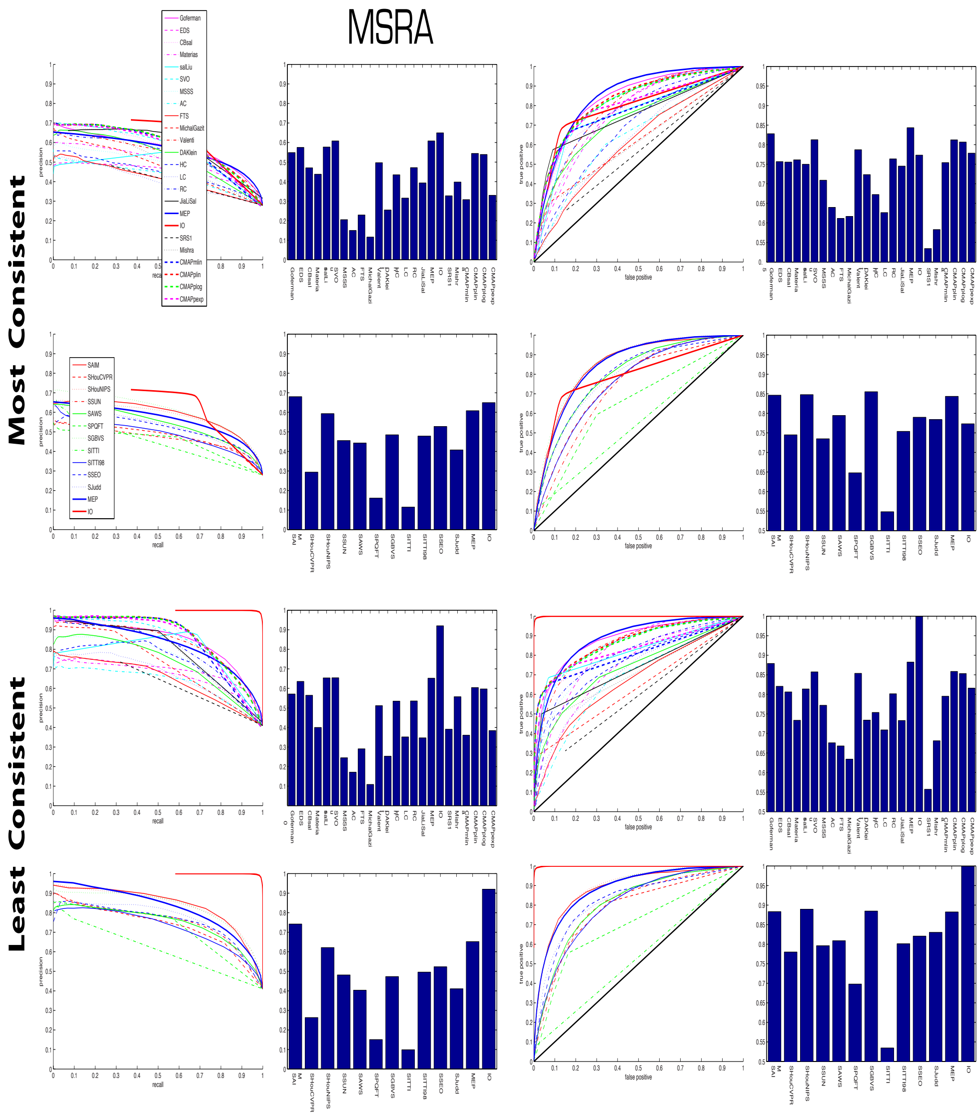

Fig. 8. Analysis of annotation consistency on model scoring over the MSRA dataset.

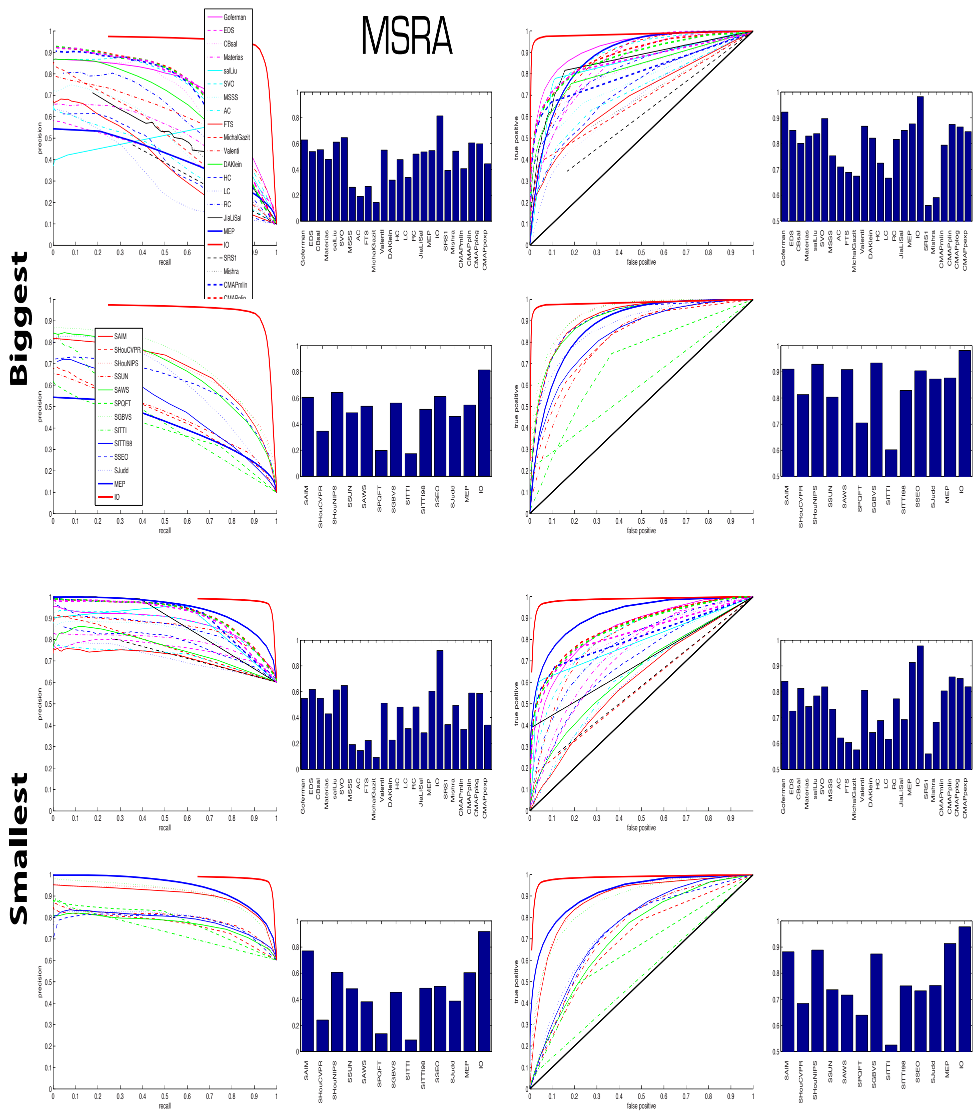

Fig. 9. Analysis of object-size over accuracy of models for the MSRA dataset.

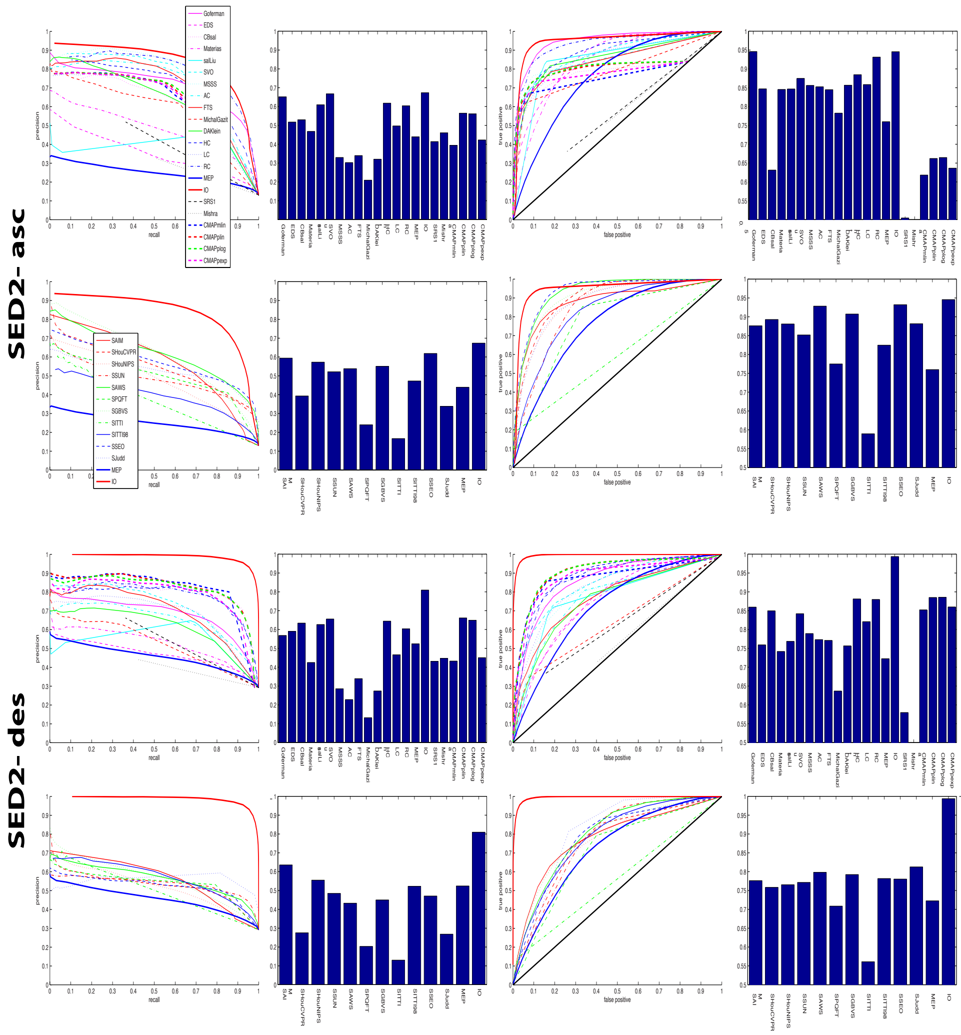

Fig. 10. Analysis of annotation consistency on models scoring over the SED2 dataset. (asc = least consistent and des = most consistent.)

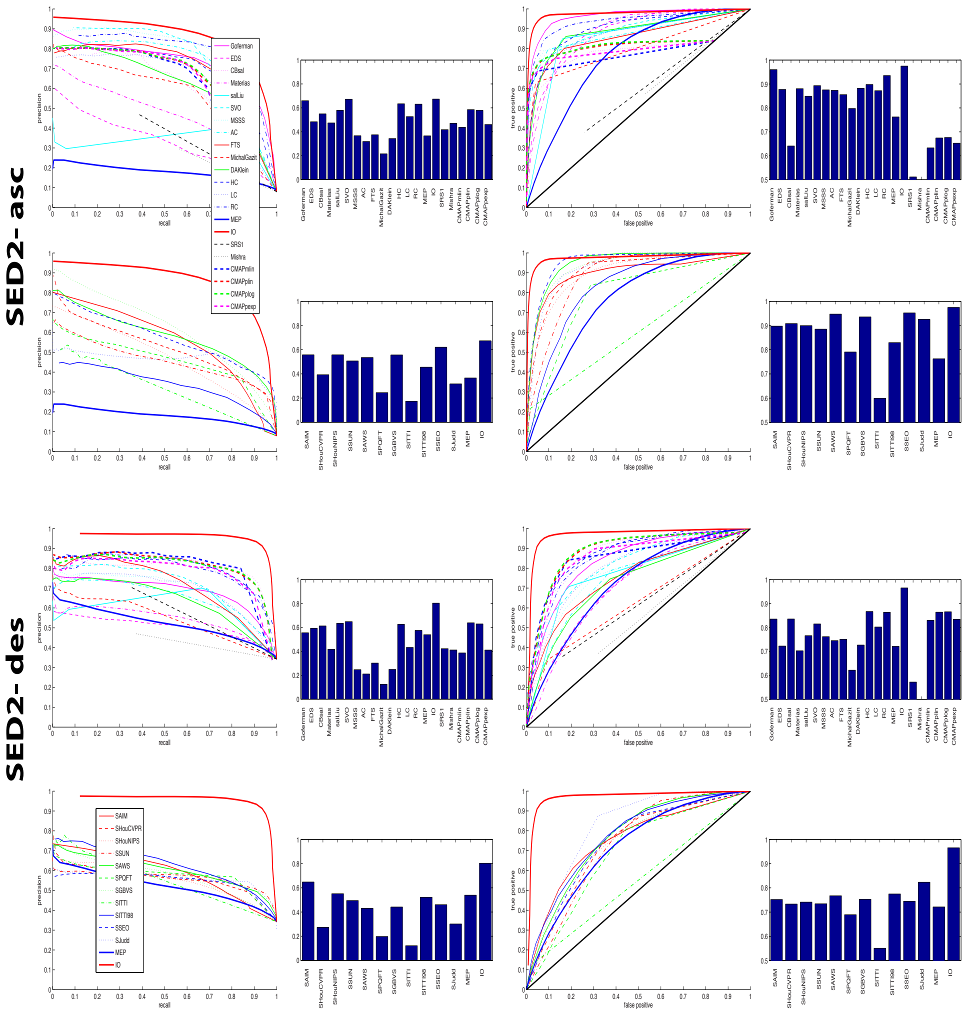

Fig. 11. Analysis of object size on model scoring over the SED2 dataset. (asc = smallest and des = biggest objects.)

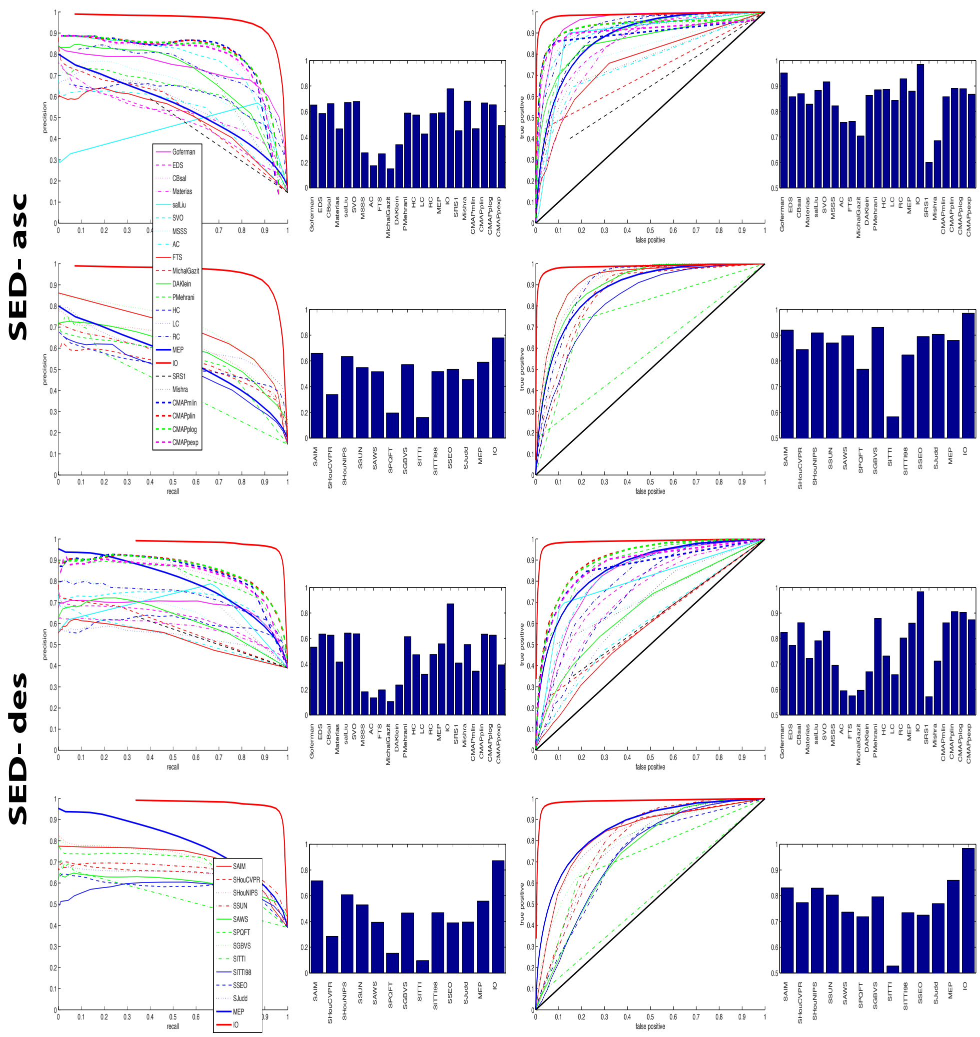

Fig. 12. Analysis of object size on model scoring over the SED1 single-object dataset. (asc = smallest and des = biggest objects.)

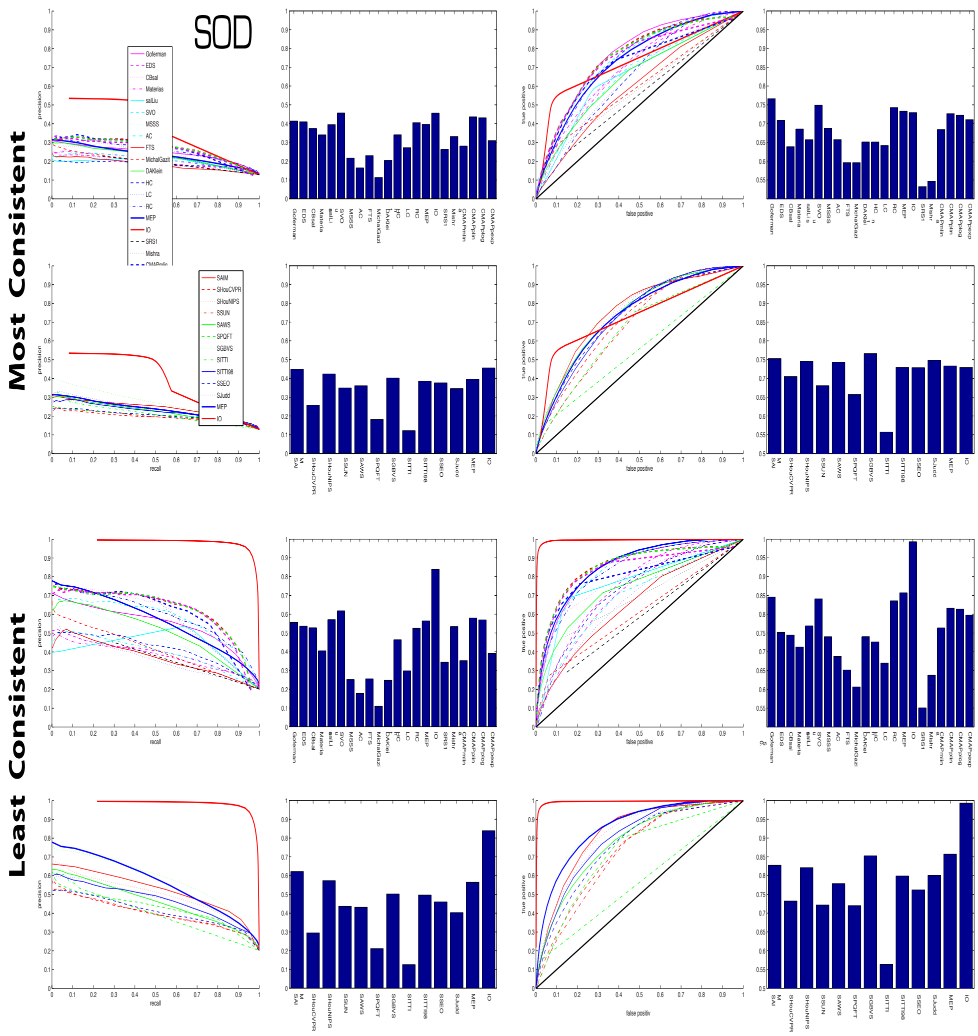

Fig. 13. Model rankings on least and most consistent images over the SOD dataset.

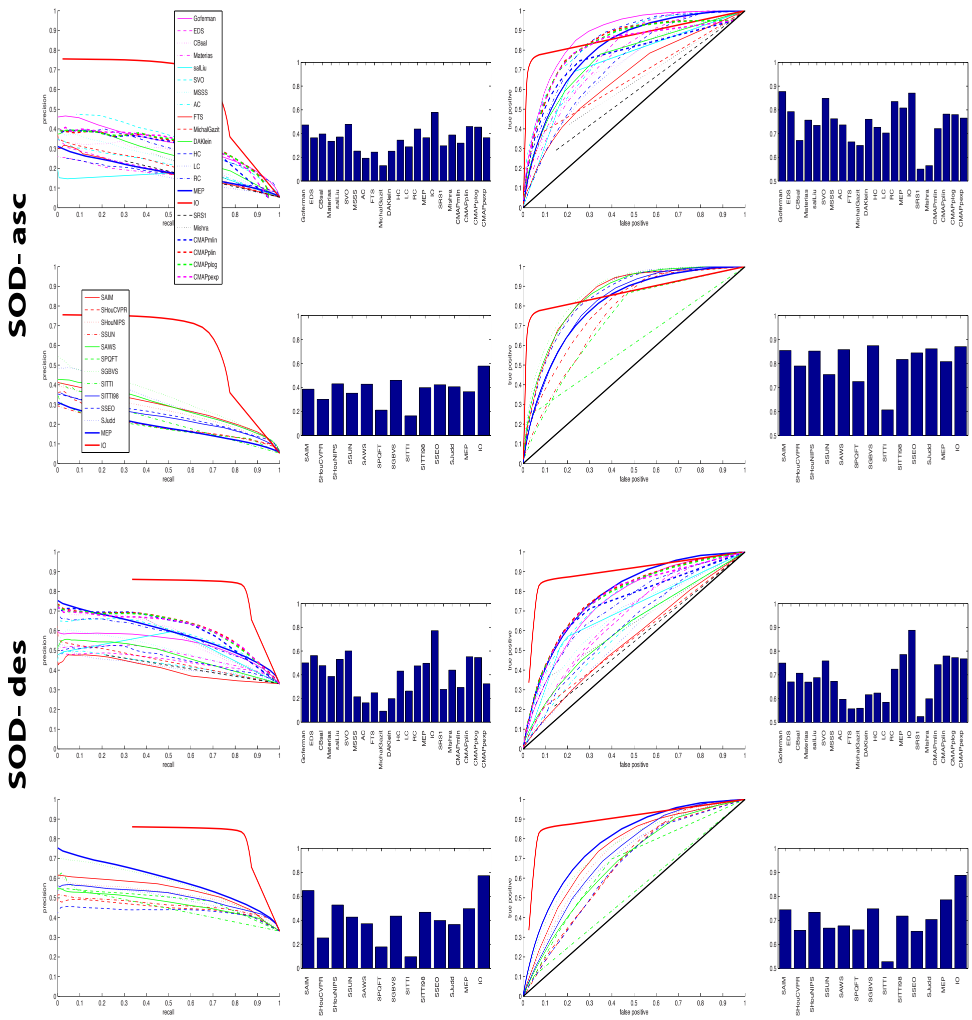

Fig. 14. Analysis of object size on model scoring over the SOD (asc = smallest and des = biggest objects.)

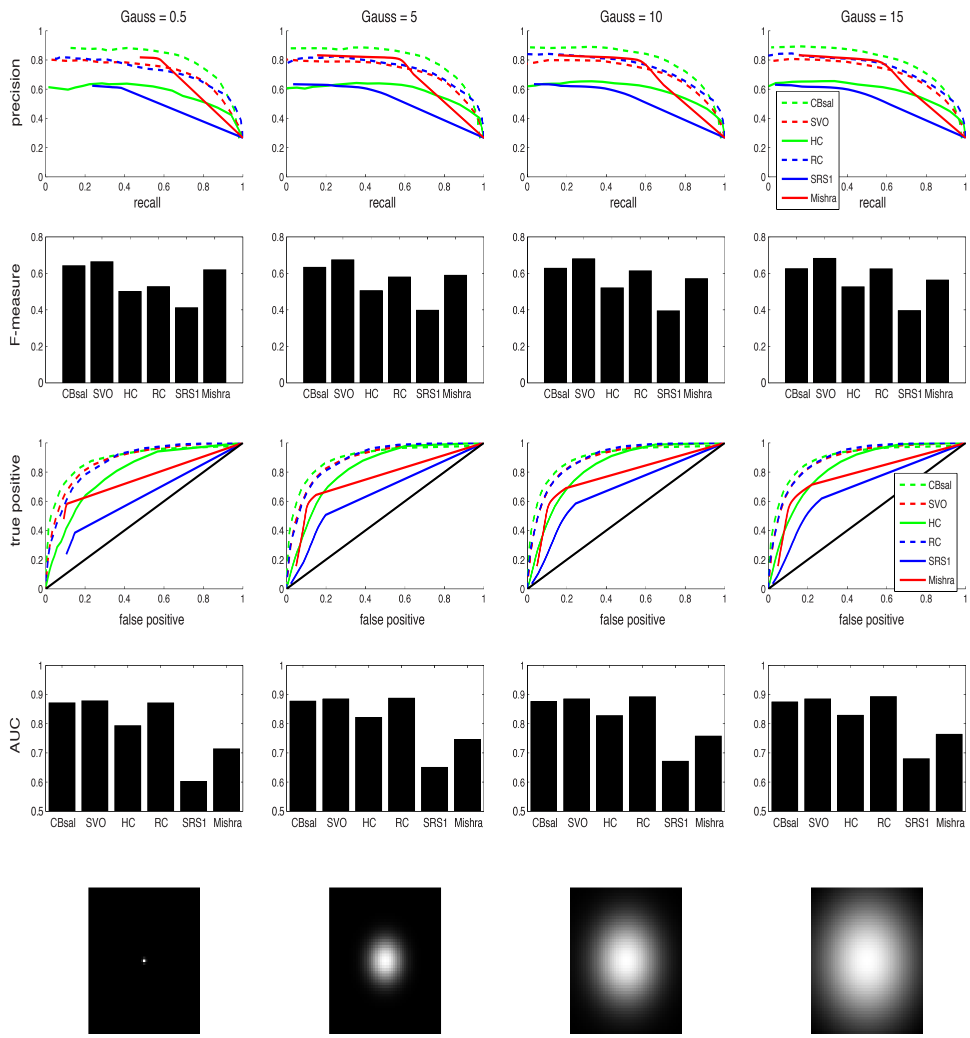

Fig. 15. Results of Gaussian smoothing of saliency maps for six models (CBsal, SVO, HC, RC, SRS1, and Mishra). Results are shown for all four scores (PR, F-measure, ROC, and AUC) over the ASD dataset.

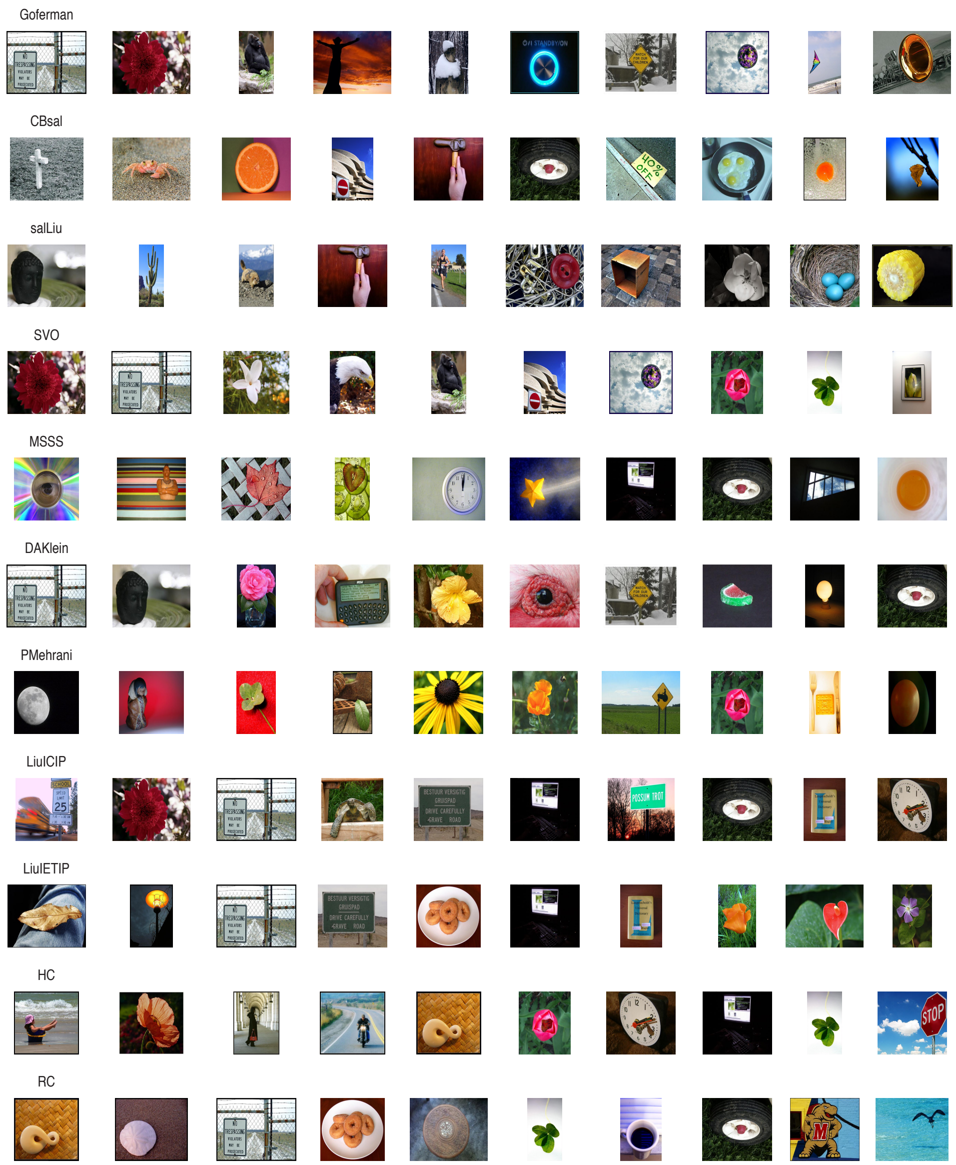

Fig. 16. First 5 images are the easiest for models in terms of AUC score. The second 5 images in each row are the most difficult stimuli for each model.

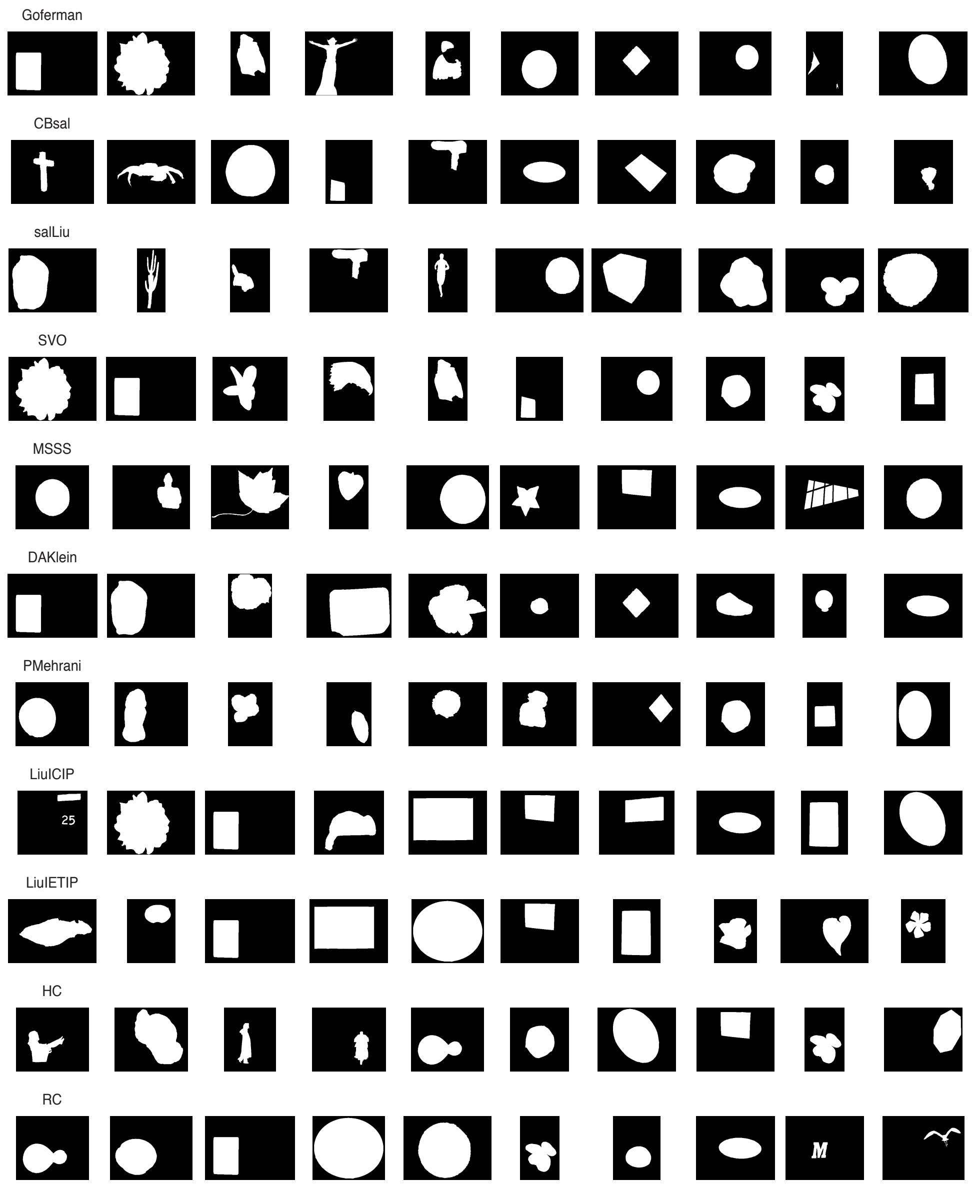

Fig. 17. Ground truth annotations for the images shown in 16.

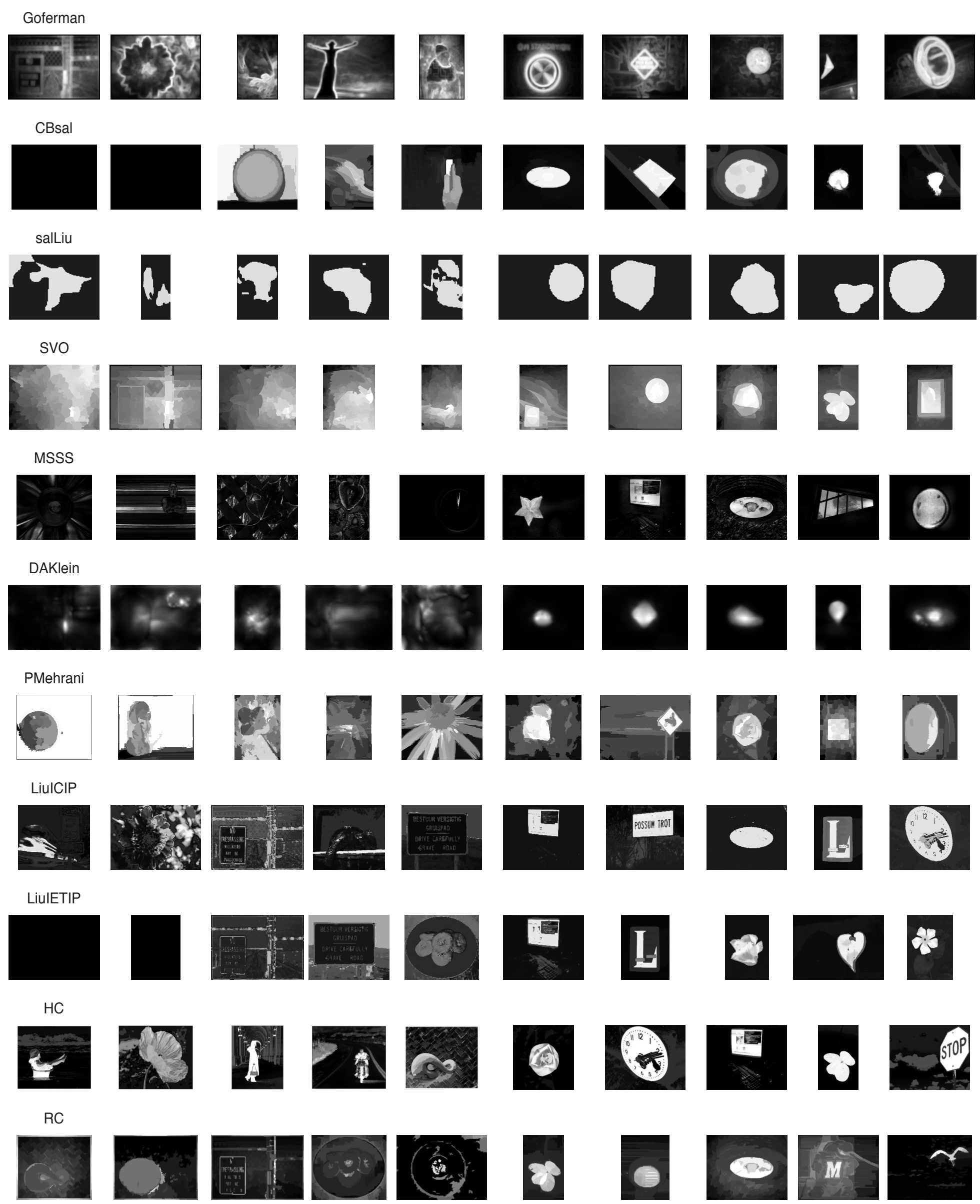

Fig. 18. Saliency maps of models for images shown in 16.