Supplemental Material

Filtering MTurk Data

To guard against extreme outliers, we employed a number of quality control methods. First, we filtered workers who participated in the same HIT more than once. We also filtered bubbles (per image) if duplicate description is found for the image stimulus. We also removed bubbles (per image) if the number of bubbles is too small; for the description task, we use a threshold of 10, while a threshold of 2 is used for the free-viewing task. The small number of bubbles often indicates either the participant submitted a low-quality description or did not pay attention to the task. Similarly, bubbles that fall outside of an image area were also filtered; this is rather a technical issue related to the HTML canvas. All these exceptional cases rarely happened, however. Finally, we employed the interquartile range (IQR)-based outlier removal procedure from Komarov et al. in order to exclude bubbles (per image) whose size is more than 3xIQR higher than the third quartile, or one that is more than 3xIQR lower than the first quartile; this approach filters much less data compared to the ones based on mean +/- 2 standard deviations.

Steven Komarov, Katharina Reinecke, and Krzysztof Z Gajos. 2013. Crowdsourcing performance evaluations of user interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 207–216.

Metric calculations

Given two distributions, P and Q^D, and their covariance σ(P,Q^D), their cross-correlation (CC) score is calculated as:

$$CC(P, Q^D) = \frac{\sigma(P, Q^D)}{\sigma(P) \times \sigma(Q^D)}$$

Given a distribution P and a binary map of fixated locations Q^B (which has a unity value only at fixated pixels, and zero elsewhere), the normalized scanpath saliency (NSS) score is calculated as:

$$NSS(P, Q^B) = \frac{1}{N} \sum_i \overline{P_i} \times Q_i^B$$

$$\text{where } N = \sum_i Q_i^B \text{ and } \overline{P} = \frac{P - \mu(P)}{\sigma(P)}$$

Here, i indexes the i-th pixel, and N is the total number of fixated pixels.

Jiang et al. used a different evaluation metric: sAUC. However, due to the discussion in Bylinskii et al. we chose to use CC instead. The CC metric is symmetric in its treatment of false positives and false negatives and is better behaved than sAUC. We report IOC using NSS, since it is a location-based metric that is highly related to CC (Bylinskii et al.).

Ming Jiang, Shengsheng Huang, Juanyong Duan, and Qi Zhao. 2015. SALICON: Saliency in Context. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 1072–1080. DOI:http://dx.doi.org/10.1109/CVPR.2015.7298710

Zoya Bylinskii, Tilke Judd, Aude Oliva, Antonio Torralba, and Fre ́do Durand. 2016. What do different evaluation metrics tell us about saliency models? *CoRR* abs/1604.03605 (2016). http://arxiv.org/abs/1604. 03605

Additional experimental analyses and results

Experiment 1: Information Visualizations

Note about data: From the 393 target images on the MASSVIS dataset, we filtered out 181 resized images (~47\%) that are too difficult for a layperson to understand (e.g., images from scientific sources) or when the image contents are too small to be recognized (e.g., illegible paragraphs or data labels), resulting in the final set of 202 images. We originally selected 204 images but later found two duplicate images, resulting in the 202 images.

Exp. 1.1: set 1 with 51 visualizations

NUMBER OF FIXATIONS (official): M=39.33, SD=1.86
OBSERVERS (official): M=16.65, SD=2.24 (min: 12)
(per image)

BUBBLE RADIUS	#OF CLICKS	#OF USERS	DESC LENGTH	TIME/IMAGE	FILTER RATE
16	M=102.63, SD=25.71	M=43.29, SD=2.33 (min: 38)	M=239.83, SD=91.90	M=3.38, SD=1.89	M=1.99, SD=2.97
24	M=80.00, SD=19.81	M=42.25, SD=1.65 (min: 39)	M=229.81, SD=82.98	M=3.03, SD=1.47	M=2.32,, SD=2.43
32	M=64.77, SD=17.95	M=42.53, SD=1.35 (min: 40)	M=229.32, SD=82.46	M=2.80, SD=1.45	M=1.21, SD=1.86

Exp. 1.2: set 2 with 51 visualizations

NUMBER OF FIXATIONS (official): M=38.98, SD=2.06
OBSERVERS (official): M=16.69, SD=1.99 (min: 11)
(per image)

BUBBLE RADIUS	#OF CLICKS	#OF USERS	DESC LENGTH	TIME/IMAGE	FILTER RATE
24	M=69.61, SD=22.67	M=23.47, SD=1.16 (min: 20)	M=245.35, SD=106.31	M=3.48, SD=2.56	M=1.84, SD=2.72
32	M=63.99, SD=20.65	M=23.22, SD=1.71 (min: 18)	M=249.38, SD=106.62	M=3.43, SD=2.36	M=1.88, SD=2.96
40	M=55.34, SD=19.67	M=13.98, SD=1.03 (min: 11)	M=240.83, SD=96.26	M=3.17, SD=2.46	M=1.77, SD=4.17

Using a one-way anova with bubble size as the factor, we found a significant effect of bubble radius size on number of clicks [F(2,150)=5.96, p<0.01]. Post hoc paired t-tests showed a significant difference between the number of clicks between bubble radius sizes of 24 and 40. Participants compensate for a smaller bubble size by clicking more on the image to expose more regions.

Exp. 1.3: set 3 with 102 visualizations

NUMBER OF FIXATIONS (official): M=39.16, SD=2.09
OBSERVERS (official): M=16.68, SD=2.05 (min: 11)
(per image)

BUBBLE RADIUS	#OF CLICKS	#OF USERS	DESC LENGTH	TIME/IMAGE	FILTER RATE
32	M=65.68, SD=19.16	M=13.70, SD=1.06	M=246.72, SD=105.14	M=3.65, SD=3.02	M=1.98, SD=4.81

Experiment 2: Natural Images

NUMBER OF FIXATIONS (official): M=9.30, SD=0.65
OBSERVERS (official): M=16.68, SD=2.05 (min: 11)
(per image)

TASK (TIME)	BUBBLE RADIUS	#OF CLICKS	#OF USERS	FILTER RATE

Mouse clicks
(10 sec)
	30	M=12.23,
		SD=1.36
			M=58.33, SD=2.34,
			min=54
				M=1.85, SD=0.92

Mouse movements
(5 sec)
	30	M=165.07,
		SD=7.16
			M=57.18, SD=3.75,
			min=49
				M=2.94, SD=2.19

Experiment 3: Static webpages

Exp. 3.1: free-viewing task

NUMBER OF FIXATIONS (official): M=17.93, SD=0.70

OBSERVERS (official): M=11.00, SD=0.00 (min: 11)

(per image)

	TIME	BUBBLE RADIUS	#OF CLICKS	#OF USERS	FILTER RATE
	10	30	M=17.77, SD=4.48	M=14.78, SD=0.46 (min: 13)	M=1.56, SD=3.35
	10	50	M=15.39, SD=3.29	M=15.80, SD=0.45 (min: 14)	M=1.33, SD=3.04
	10	70	M=12.68, SD=1.30	M=19.92, SD=1.32 (min: 17)	M=3.82, SD=3.47
	30	30	M=46.14, SD=3.35	M=14.92, SD=0.84 (min: 14)	M=0.56, SD=2.39
	30	50	M=39.86, SD=5.20	M=14.43, SD=0.64 (min: 13)	M=1.73, SD=3.48
	30	70	M=28.53, SD=3.44	M=18.59, SD=0.64 (min: 17)	M=2.34, SD=3.64

NSS scores for BubbleView similarity to eye fixations, averaged over all 17 images per type (text, pictorial, or mixed):

TIME	BUBBLE RADIUS	Text (IOC = 1.97)	Pictorial (IOC = 1.77)	Mixed (IOC = 1.80)
10	30	1.14	1.32	1.27
10	50	1.43	1.34	1.37
10	70	1.40	1.37	1.35
30	30	1.50	1.44	1.45
30	50	1.51	1.47	1.39
30	70	1.46	1.36	1.28
Unlimited (description task)	30	1.50	1.45	1.43

Exp. 3.2: description task

NUMBER OF FIXATIONS (official): M=17.93, SD=0.70
OBSERVERS (official): M=11.00, SD=0.00 (min: 11)
(per image)

BUBBLE RADIUS	#OF CLICKS	#OF USERS	DESC LENGTH	TIME/IMAGE	FILTER RATE
30	M=85.71, SD=24.91	M=13.71, SD=0.94 (min: 12)	M=240.64, SD=104.17	M=2.83, SD=2.08	M=1.50, SD=3.05

Experiment 4: Graphic designs

Note about data: We segmented the 51 graphic designs into elements. Bounding boxes were manually constructed around distinct elements in each design for an average of 6-7 elements segmented per image (SD: 4). Elements include text boxes, distinct text like a title, photographs, logos, etc.

BUBBLE RADIUS	#OF CLICKS	#OF USERS	FILTER RATE
30	M=15.17, SD=2.19	M=14.86, SD=0.34 (min=14.00)	M=0.92, SD=2.29

Experiment 5: natural images from SALICON

Exp. 5.1: BubbleView with clicks

NUMBER OF MOUSE LOCS (official): M=121.36, SD=14.36
OBSERVERS (official): M=58.06, SD=3.17 (min: 52)
(per image)

IMAGE BLUR	BUBBLE RADIUS	#OF CLICKS	#OF USERS	FILTER RATE
30	30	M=11.41, SD=2.15	M=15.29, SD=1.24 (min: 13)	M=2.69, SD=4.83
30	50	M=14.03, SD=2.93	M=14.45, SD=0.86 (min: 12)	M=1.71, SD=4.35
30	70	M=9.43, SD=1.25	M=14.94, SD=0.73 (min: 13)	M=2.78, SD=4.08
50	30	M=13.38, SD=1.44	M=15.16, SD=0.64 (min: 14)	M=1.24, SD=2.68
50	50	M=14.86, SD=1.76	M=15.00, SD=0.00	M=0.00, SD=0.00

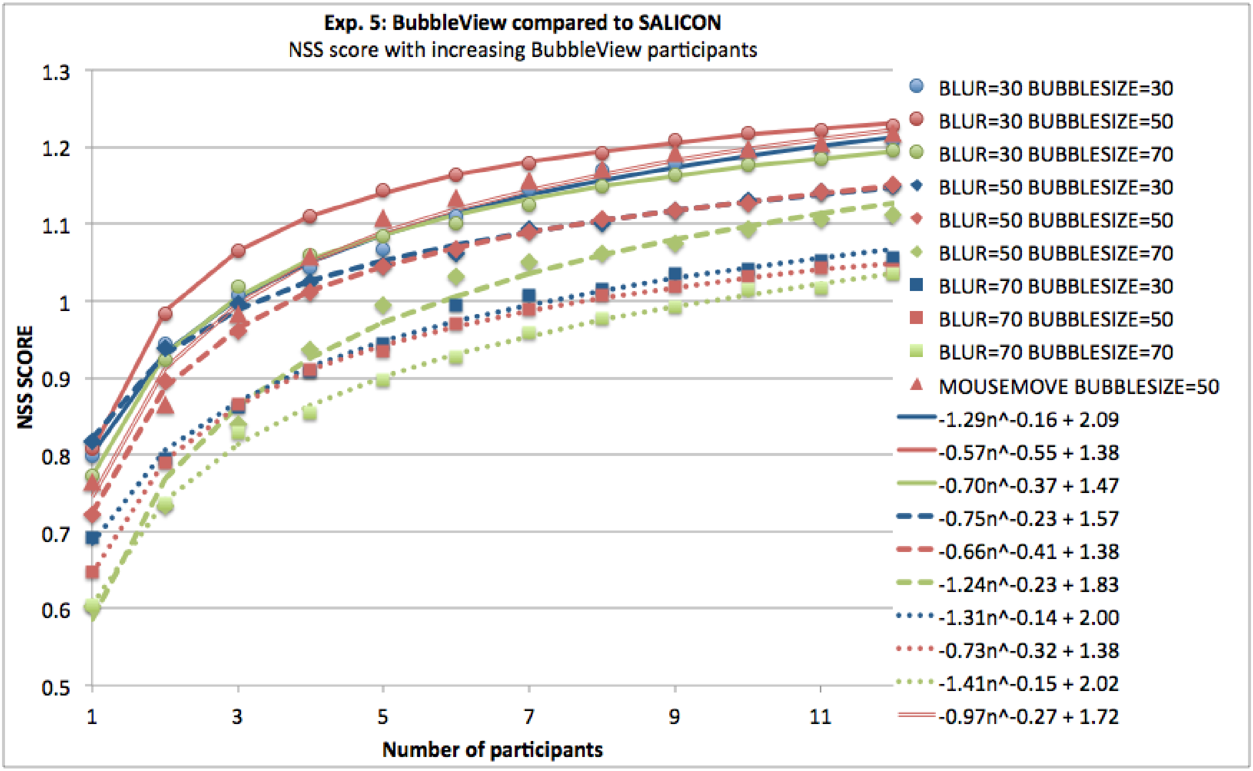
			(min: 15)	
50	70	M=13.25, SD=4.19	M=15.02, SD=1.05 (min: 13)	M=4.65, SD=5.43
70	30	M=12.40, SD=2.11	M=15.18, SD=0.68 (min: 14)	M=3.40, SD=4.90
70	50	M=15.63, SD=1.92	M=15.61, SD=0.63 (min: 13)	M=0.44, SD=2.33
70	70	M=13.80, SD=2.89	M=15.73, SD=0.96 (min: 14)	M=1.88, SD=4.13

Exp. 5.2: BubbleView with moving-window

Note about data: We maintained a sampling rate of 100 Hz when a mouse cursor was moving on an image, which is a simple approximation to the normalized sampling rate of 100 Hz used in the SALICON experiment.

NUMBER OF MOUSE LOCS (official): M=121.36, SD=14.36
OBSERVERS (official): M=58.06, SD=3.17 (min: 52)
(per image)

BUBBLE RADIUS	#OF CLICKS	#OF USERS	FILTER RATE
30	M=169.33, SD=12.00	M=13.86, SD=1.06 (min: 11)	M=1.17, SD=3.45
50	M=158.77, SD=17.11	M=15.31, SD=0.47 (min: 15)	M=0.13, SD=0.92


Exp. 5: BubbleView compared to SALICON
NSS score with increasing BubbleView participants
NSS SCORE
1.3
1.2
1.1
1
0.9
0.8
0.7
0.6
0.5
1
3
5
7
9
11
Number of participants
BLUR=30 BUBBLESIZE=30
BLUR=30 BUBBLESIZE=50
BLUR=30 BUBBLESIZE=70
BLUR=50 BUBBLESIZE=30
BLUR=50 BUBBLESIZE=50
BLUR=50 BUBBLESIZE=70
BLUR=70 BUBBLESIZE=30
BLUR=70 BUBBLESIZE=50
BLUR=70 BUBBLESIZE=70
MOUSEMOVE BUBBLESIZE=50
-1.29n^-0.16 + 2.09
-0.57n^-0.55 + 1.38
-0.70n^-0.37 + 1.47
-0.75n^-0.23 + 1.57
-0.66n^-0.41 + 1.38
-1.24n^-0.23 + 1.83
-1.31n^-0.14 + 2.00
-0.73n^-0.32 + 1.38
-1.41n^-0.15 + 2.02
-0.97n^-0.27 + 1.72

Exp. 5: natural scenes (ground-truth IOC: 1.50)	IMAGE BLUR	BUBBLE RADIUS	CC	NSS	Normalized NSS
Clicks (Exp. 5.1)	30	30	0.84	1.21	81%
Clicks (Exp. 5.1)	30	50	0.86	1.23	82%
Clicks (Exp. 5.1)	30	70	0.84	1.20	80%
Clicks (Exp. 5.1)	50	30	0.84	1.15	77%
Clicks (Exp. 5.1)	50	50	0.84	1.15	77%
Clicks (Exp. 5.1)	50	70	0.84	1.11	74%
Clicks (Exp. 5.1)	70	30	0.78	1.06	71%
Clicks (Exp. 5.1)	70	50	0.80	1.04	69%
Clicks (Exp. 5.1)	70	70	0.79	1.04	69%
Moving-window (Exp. 5.2)	30	30	0.87	1.21	81%
Moving-window (Exp. 5.2)	30	50	0.88	1.24	83%