Symposium: 22nd International Command and Control Research and Technology Symposium

Topic 7: Human Information Interaction

Title: Predicting where people look in information graphics

Authors:
Andre V. Harrison
U.S. Army Research Lab, RDRL-CII-B
2800 Powder Mill Rd, Adelphi, MD 20783
andre.v.harrison2.civ@mail.mil

Adrienne J. Raglin
U.S. Army Research Lab, RDRL-CII-B
2800 Powder Mill Rd, Adelphi, MD 20783
adrienne.j.raglin.civ@mail.mil

Mark A. Livingston
U.S. Naval Research Laboratory,
4555 Overlook Ave SW, Washington, DC 20375
mark.a.livingston@us.navy.mil

Predicting where people look in information graphics

Abstract:

Military command and control (C2) refers to dominating by reason of location. Gathering and analyzing data from multiple aspects is important to achieving and maintaining that authority in the environment in which the operation will be conducted. High level understanding of the battlefield conducted in Command Centers is made with maps and images but also with information presented in plots, charts, and graphs. Graphs display data as imagery to convey a particular interpretation of that data as information. Treating graphs as images, we use visual saliency models to try and predict where people will look in a graph. Modeling graphs in this way, we aim to better understand the interplay between memorability and comprehension in graphs and other information graphics. This work is an initial step in the development of computational models that can accurately predict the salience, memorability, and clarity of information graphs. Throughout this paper we utilize the Massachusetts Visualization dataset for the ground truth values of where people look when looking at information graphics. Using this dataset, we analyze the predictive accuracy of three different saliency models when people look at information graphics.

1. Introduction:

Establishing dominance in a modern command and control (C2) environment relies on gathering and analyzing data from multiple sources and from multiple locations concerning different aspects of the environment in which the operation will be conducted. That analyzed data must be processed and filtered to convey the relevant information to commanders in a way that is easily understood and remembered. This often is done through the use of different types of visualizations. High level understanding of the battlefield conducted in command centers is made with maps and images but also with information presented in plots, charts, graphs, and other information graphics. The type of visualization used can be tightly coupled by the type of information to be presented, the patterns to infer, and conclusions (decisions) to be reached [1–9]. Beyond that, the type of visualization and the design of each type of visualization can impact how easily the presented information will be understood or how well it will stand out and be remembered. There are several papers on what type of visualization should be used in what circumstances and how the choice and design of an information graphic can affect the comprehension or the memorability of an information graphic [2, 3, 5, 10, 11]. However, these papers mainly provide qualitative guidelines and are not easily automated, since the factors that impact the memorability and ease of understanding depend on contextual information, understanding what information and patterns are present in the graph, and the principles of human vision. To contribute to the eventual development of a more automated visualization pipeline, we focus on the last element in the previous list, by trying to predict where people will look when viewing an information graphic. This may also help more precisely

answer how the design choices of information graphics affects memorability and comprehension, as they have sometimes been noted to be opposing elements in the design of an information graphic [10, 11]. Our work is based off of long held theories and models of visual search in natural images [12, 13]. Our prior work in modeling where people looked when they see visualizations [14], focused on modeling visual search within statistical graphs (SG), which is a much simpler type of data visualization. Statistical graphs like line graphs, bar charts, and scatter plots are a specific type of data visualization that are used to visualize quantitative information and high order relationships.

In this paper, we plan to extend our modeling efforts and apply it to more generic information graphics that have a mix of just statistical information and more natural objects. By studying graphs as imagery, we hope to eventually develop a better understand of the interplay between memorability and comprehension in graphs and other information graphics. This work represents an initial step in the development of computational models that can accurately predict the salience, memorability, and clarity of data visualizations.

In section 2, we describe some relevant work in the design of SG, similar studies in the perception of SG, as well as a brief background in the modeling of visual salience. In section 3, we discuss the types of analyses we did using the eye tracking data from the MASSVIS dataset. In section 4, we discuss the results of the analysis on the patterns of eye gaze and the prediction of eye gaze using different saliency models. In the last section, we provide a short conclusion of the analysis and modeling done in this work.

2. Background:

2.1 The Design of information graphics

The focus in the selection and design of information and data visualization methods has largely been assessed in terms of their ease of understanding and memorability. In military C2 situations, system utility is often the main focus in interface design, as highly specialized and complex data needs to be presented in a way that is most useful to the trained system user so they may complete their mission. This can result in C2 information graphics that are not easily understood by novices, and may not be very memorable or visually appealing. However, human factors guidelines are often not utilized in the design of visualizations. Most human factors guidelines on the design and choice of visualizations are typically focused on comprehension; as such they often prioritize sparsity over aesthetic quality or memorability [2, 3]. The reasoning being that only the relevant information should be shown in a visualization and it should be displayed in the simplest layout possible. Any other information, diagrams, text, graphics, and other types of "visual clutter" are discouraged as it may serve as a distraction from the core message of the graphic. Some studies on visualizations and information graphics, however, have also found that prioritizing understanding can have a negative impact on memorability [10, 11]. Those elements of visual clutter that may interfere with understanding the message of the graphic may help in making the overall visualization more memorable [10, 11].

2.2 Modeling where people look

There is an obvious link between comprehending and remembering the content of an information graphic and visual attention. Visual attention is a required first step before either can occur. Visual fixation is often used as a measure of visual attention and there have been several papers that have studied the relationship between fixations and how well an information graphic was understood or how likely it would be remembered. Models of where people fixate in an image are known as visual saliency models. These

models typically consider 1 or 2 types of factors that influence where someone will direct their fixation. These are known as top-down factors or bottom-up factors.

Top-Down factors are goal or mission oriented factors that can drive fixation/attention to a particular location, feature, or object in an image. These are typically factors that are independent of the image or scene being viewed. For instance, looking for a green line in a plot or looking for the title of an information graphic are top down factors. Visual saliency models do not always take these kind of factors into account as they can be subjective, are independent of the image, and typically require training/machine learning and all of the complexities that come with that.

Bottom-up factors are stimulus driven factors that attract a person's attention or gaze to certain locations in an image. These models are all based on the visual perception concept known as feature integration theory (FIT) [12]. FIT states that when a person views a scene the basic features that make up that scene are extracted all at once by the visual system. However, a person can only direct their attention to a specific location in the scene at any given moment in time. Thus, attention must move around a scene in order for a person to fully perceive it. Bottom-up factors typically direct a person's gaze/attention to locations that have unique, unexpected, or locally different features. Bottom-up factors such as a colored location in an otherwise black and white graph or a line that is vertically higher than all other lines in that graph are unique features that are likely to draw someone's attention.

Current theories of visualization design exist as more qualitative suggestions of what might attract the viewer's attention based on the intent of the creator of the information graphic [2–4]. Developments in computational models of fixation for information graphics might enable the creation and design of information graphics to be a bit more automated where the model feedback along with knowledge about the type of data, content of that data, and the mission requirements may enhance the design of information graphics so that they are more consistent with human factors guidelines. Our prior work in modeling visual fixation when viewing statistical graphics has had moderate success in predicting fixation using only bottom-up saliency models, but was also hampered by noisy ground-truth data. But at the same time well-known patterns of fixation that are observed when viewing natural images were noticeably different or entirely absent for statistical graphics [14].

2.3 The MASSVIS dataset:

The MASSachusetts (MASSive) VISualization (MASSVIS) dataset is a dataset of >1000 information graphic images from various civilian domains and of various complexity, **Figure 1**. The dataset has been used to study topics on the memorability and recall of information graphics under different time scales [10, 11, 15]. This paper uses data collected from one of the MASSVIS memorability studies where the eye fixation of participants was recorded while they conducted a memorization task [15]. In that study 33 participants viewed 100+ images, from a randomly selected subset of 393 images, for ten seconds each. As a result, each image was viewed by 10 – 22 participants.

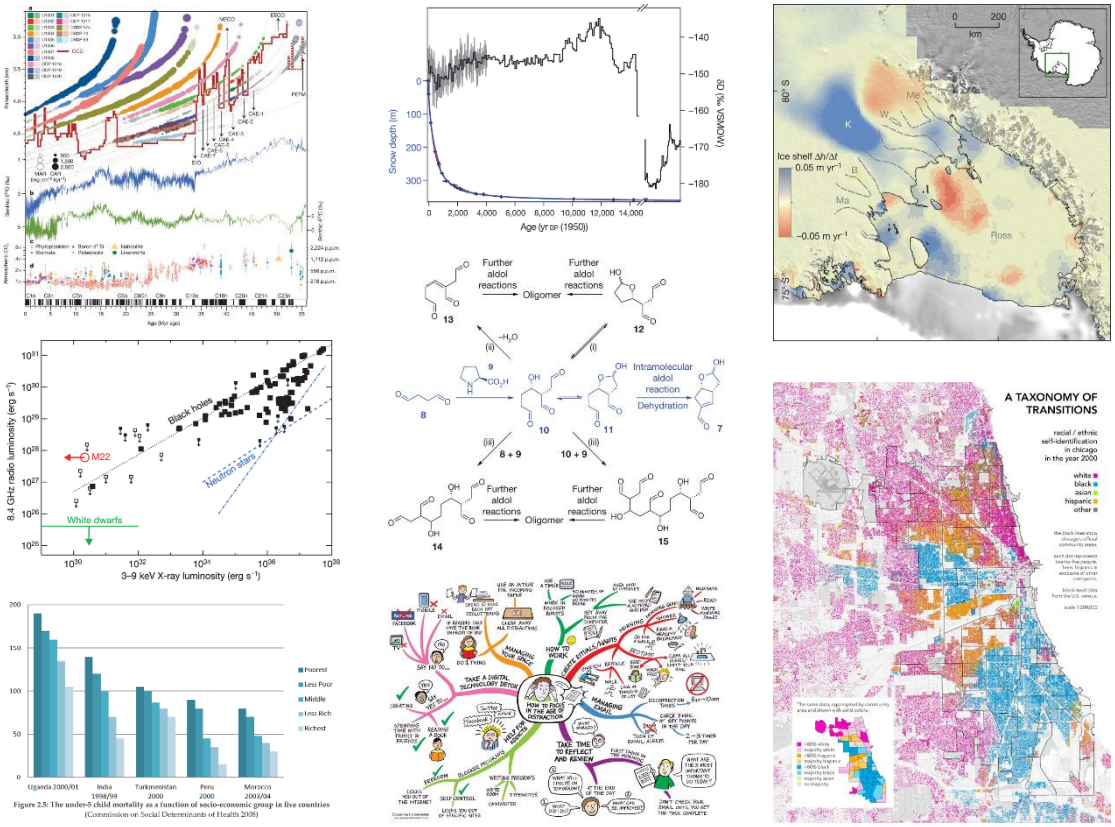

Figure 1. *Examples of the types of visualizations present in the MASSVIS dataset including different types of statistical graphs shown in all rows of the leftmost column and the top image in the second column* [10, 11, 15].

3. Analysis:

We analyzed all 393 images using 3 well-known bottom-up only saliency models [13, 16, 17]. Our prior work in predicting where people fixate when they view statistical graphics found that the central bias typically found in eye fixation maps of natural images did not seem to be present [14]. The MASSVIS dataset has a wide range of visualization categories including statistical graphs, so we have analyzed the eye fixation patterns as well as the predictions made by different saliency models. Within the MASSVIS dataset we pay special attention to visualizations that are statistical graphs vs. those that are diagrams. These groupings are very similar to the MASSVIS groups of highly memorable graphics vs. least memorable graphics.

For our analysis, we followed the standard procedure to assess the performance of saliency models. We compared the predictive accuracy of saliency maps produced by each saliency model against the eye fixations for each image. To establish a ceiling of the highest possible predictive accuracy we use the inter-subject consistency [16]. The inter-subject consistency takes into account that there is an inherent variability in the viewing patterns of people when they are looking at an image. If 10 people look at the same image they are all likely to have different patterns of fixation. Thus, the best performance any saliency model can be expected to reach is to predict the eye fixation of one person as accurately as is possible when using the eye fixation patterns of the other 9 individuals as a saliency map. The average value is known as the inter-subject consistency.

There are several possible metrics that can be used to compare a saliency map with patterns of eye fixation. We have chosen to use three different metrics, the area under the curve of the receiver operating characteristic (ROC), the normalized scan-path saliency (NSS), and the Kullback-Liebler divergence (KL). Each metric focuses on a different aspect of the output of each saliency model and matches the eye fixations to the saliency predictions in slightly different ways [18, 19]. The ROC metric measures the area under the curve of the true positive rate vs. the false positive rate of how well eye fixations can be predicted using the saliency map. Thus, the ROC can be thought to evaluate the ability of a saliency model to rank the most and least salient locations in an image in the same order as in the eye fixation map. The NSS metric converts the values in a saliency map so that it has zero mean and unit variance. The average value of the locations in the normalized saliency map of where people looked is calculated. The higher the value the better the accuracy. So, the NSS metric evaluates a saliency map based on the relative saliency value given to fixation locations vs. non-fixation locations. The KL-divergence compares the probability distribution of the values in the saliency map at locations where people looked vs. a random sampling of the saliency map. So rather than looking at ranking or relative value, the KL divergence metric looks at the distribution of values in the saliency map. Using the output from all of the metrics together provides a richer comparison of the different saliency models and it is a stronger result if a single saliency model always has the best performance across each metric.

For the ROC, NSS, and KL divergence values of 50%, 0, and 0, respectively, would be assumed to indicate random chance. However, the central bias of natural images has typically enabled a better than chance level of predictive accuracy simply by using a Gaussian kernel centered in the middle of the image. Even with the possible lack of central bias in statistical graphs the predictive accuracy of the Gaussian kernel serves as a best guess of where people may look without knowing what is contained in the image. Also, if there truly is no central bias the Gaussian kernel should have a predictive accuracy no better than random chance.

4. Results

Comparing the maps of eye gaze against the different saliency maps created by each of the saliency models shows that there is a fairly consistent difference in what the models identify as salient vs. what the participants perceived to be salient. The maps of eye gaze in **Figure 2** show that for simple visualizations like the statistical graphs in the top two rows of **Figure 2** participants tended to fixate on the text in each image and at locations where there is information describing the category or organization of the visualization, as was observed by Borkin et al. [15]. This can conflict with the type of patterns that saliency models typically look for. Typically, bottom-up saliency models look for spatial locations with unique patterns in size, shape, color, or orientation. This is evidenced by the locations highlighted by most of the saliency models in **Figure 2**. Specifically, in the case of the SG images, the points in the scatter plot and the line in the line graph are identified as a salient location or the most salient location, which conflicts with the eye gaze maps as the participants rarely fixated on these locations. It is of note that the ioM highlights more areas as salient and its saliency maps are more detailed, in that the structure of the original image is visible. Its saliency maps are more detailed because it extracts higher spatial frequency patterns relative to the GBVS and Itti saliency models. This may be one reason why the Itti and GBVS models never highlight any of the locations with text in the images as salient. Most of the text in visualizations have a high spatial frequency in relation to the overall image and as such are being filtered out by these two models.

A more quantitative evaluation of the predictive strength of the bottom-up saliency models that we tested found that even though the saliency models can fail to predict seemingly important locations of eye fixation in visualizations they are able to predict where people look. For most of the metrics we used each saliency model can even predict where people look with an accuracy above chance **Table *1***. From prior work in assessing the performance of saliency models we continue to use the Gaussian kernel, centered in the middle of the image, as an effective measure of chance. The results of this work confirm the results from an earlier paper on eye fixation when viewing SG [14]. We see that a central bias is not present in the eye fixation patterns of people when they look at the information graphics in the MASSVIS dataset as the results of the Gaussian kernel for all tested metrics is at or near the theoretical random choice value for each metric. The differences between the predictive strength of the tested saliency models vs. the Gaussian Kernel and the Inter-subject consistency shows that:

1.) For ROC and NSS metrics, the saliency models do better than chance in predicting where people look by rank and value, respectively. However, the results from the KL divergence metric find that that the distribution of saliency values for the Itti and GBVS model are not better than effective chance.
2.) There is still a big difference between the performance of the bottom-up saliency models and where people actually look as the inter-subject consistency is so much higher than the performance of the different saliency models.
3.) Across all three metrics the ioM consistently performs better than all of the other tested saliency models. It is also the only tested saliency model that had an accuracy above effective chance using the KL divergence metric. This may be due to the fact that the ioM consistently identifies text regions in each image as salient, **Figure 2**, which are often fixated on by the participants.

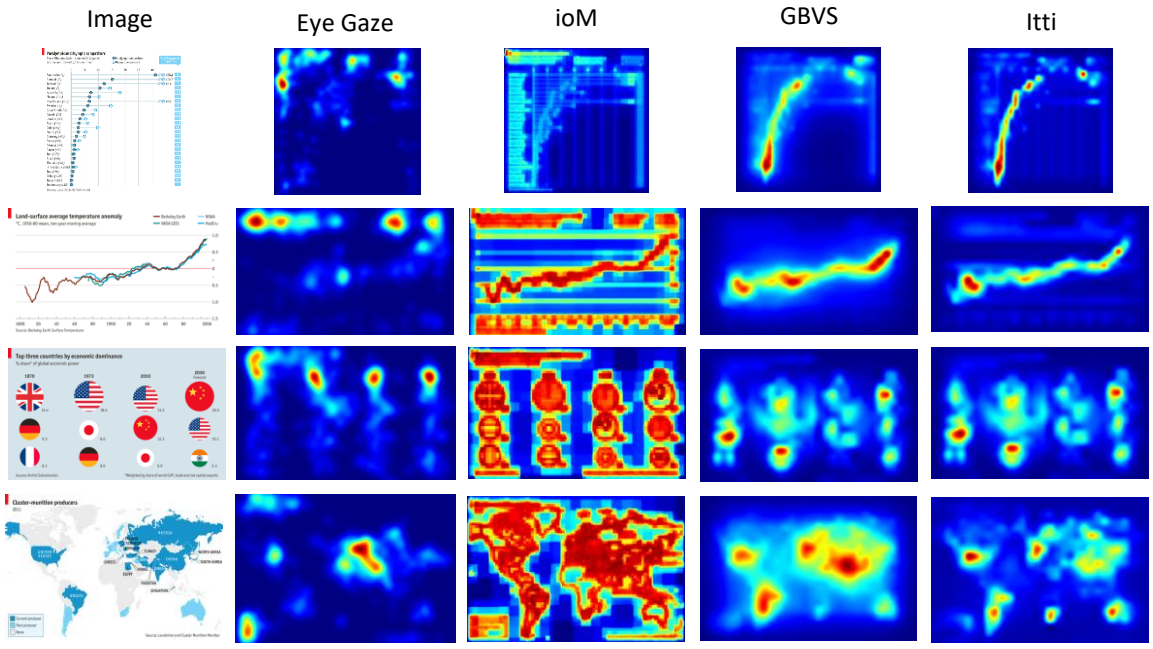

Figure 2. Qualitative comparison of the saliency maps from the different saliency models. The 1st column, the leftmost column, shows the original images used within the study. The 2nd column shows the aggregated eye fixations for each image. The 3rd, 4th and 5th columns shows the saliency maps generated by the ideal observer model (ioM) [17] the graph based visual saliency (GBVS) model [16], and the Itti model [13].

Table 1. Comparison of the predictive strength of the Gaussian Kernel, the inter-subject consistency, and the ioM, GBVS, and Itti saliency models using the ROC, NSS, and KL-divergence.

	ROC	NSS	KL
Gaussian Kernel	0.51815	0.05395	0.43425
Itti [13]	0.5976	0.1980	0.3882
GBVS [16]	0.58435	0.16585	0.40125
ioM [17]	0.5990	0.32105	0.4731
Inter-subject consistency	0.83065	1.58985	5.16665

5. Conclusion:

The goal of information visualization is to transform data into information so that the viewer can use it to complete some objective. Commanders, particularly at higher echelons, must be able to obtain the right information at the right time, but the right information also needs to be in the right form. For military command centers visualization is key in providing commanders with the information needed for decision making. Command centers strive to provide commanders with information that allows them the ability to understand the battlefield and what actions may need to be taken. This "picture" must show the current state of the battlefield, but also incorporate the commander's interpretation of the previous state and possible new state(s) as the operation changes. Additionally, this picture must facilitate the communication of the battlefield to the commander's staff and to units at other echelons. The "picture" can consist of images, maps, plots, charts, or graphs. The ultimate goal of these graphs is to provide clear and accurate information concisely. Studying how graphs, plots, and charts as images convey information can improve the "picture" that the commander needs.

The work presented in this paper builds upon prior work in predicting eye gaze when people look at visualizations during a memory task. Early work focused exclusively on SG and used a less accurate eye tracker, while this work has analyzed eye gaze from multiple types of visualizations. The results from this paper add to the understanding gained in the prior effort; confirming the lack of central bias in SG, but also extending that finding to many types of data visualizations. The work presented in this paper also attempted to predict eye gaze patterns on multiple types of visualizations, which we were able to do with an accuracy above chance.

From this work we have shown that, while current saliency models can predict eye gaze in different visualizations, they do so only just slightly better than chance. In order to improve upon a saliency model's ability to predict where people look in different visualizations, future models will need to process more than just low spatial frequency patterns; high spatial frequency information, like text, will be need to be extracted and processed as well. As a result, the down sampling many saliency models perform on the input image, presumably to improve the computational speed, may need to be eliminated, reduced, or selectively applied. In the development of saliency models for predicting where people look in images, face detection algorithms have often been added to saliency models, as faces are a strong indicator of where people will look in an image [20, 21]. Similarly, for predicting where people look in SG and other visualizations, adding a text detector to identify locations in an image where text is present may improve

the predictive accuracy of future saliency models. Future work in this area of predicting where people look in visualizations would benefit from the development of top-down saliency models that could recognize the typical features of a graph (title, axes, axis labels, key chart, category labels, etc).

6. References:

1. Aumer-Ryan, P.: Visual Rating System for HFES Graphics: Design and Analysis. In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting. pp. 2124–2128 (2006).
2. Fausset, C.B., Rogers, W.A., Fisk, A.D.: Visual Graph Display Guidelines. , Atlanta, GA (2008).
3. Gillan, D.J., Wickens, C.D., Hollands, J.G., Carswell, C.M.: Guidelines for Presenting Quantitative Data in HFES Publications. Hum. Factors J. Hum. Factors Ergon. Soc. 40, 28–41 (1998).
4. Petkosek, M.A., Moroney, W.F.: Guidelines for constructing graphs. In: Society, Human Factors ergonomics Meeting, Annual. pp. 1006–1010 (2004).
5. Vessey, I., Galletta, D.: Cognitive Fit: An Empirical Study of Information Acquisition. Inf. Syst. Res. 2, 63–84 (1991).
6. Vessey, I.: Cognitive Fit: A Theory-Based Analysis of the Graphs Versus Tables Literature. Decis. Sci. 22, 219–240 (1991).
7. Acartürk, C.: Towards a systematic understanding of graphical cues in communication through statistical graphs. J. Vis. Lang. Comput. 25, 76–88 (2014).
8. Greenberg, R.A.: Graph Comprehension: Difficulties, Individual Differences, and Instruction, (2014).
9. Halford, G.S., Baker, R., McCredden, J.E., Bain, J.D.: How Many Variables Can Humans Process? Psychol. Sci. 16, 70–76 (2005).
10. Borkin, M.A., Bylinskii, Z., Kim, N.W., Bainbridge, C.M., Yeh, C.S., Borkin, D., Pfister, H., Member, S., Oliva, A.: Beyond Memorability : Visualization Recognition and Recall. IEEE Trans. Vis. Comput. Graph. 22, 519–528 (2016).
11. Borkin, M.A., Vo, A.A., Bylinskii, Z., Isola, P., Sunkavalli, S., Oliva, A., Pfister, H.: What Makes a Visualization Memorable? IEEE Trans. Vis. Comput. Graph. 19, 2306–2315 (2013).
12. Treisman, A.M., Gelade, G.: A feature-integration theory of attention. Cogn. Psychol. 12, 97–136 (1980).
13. Itti, L., Koch, C., Niebur, E.: A model of saliency-based visual attention for rapid scene analysis. IEEE Trans. Pattern Anal. Mach. Intell. 20, 1254–1259 (1998).
14. Harrison, A., Livingston, M.A., Brock, D., Decker, J., Perzanowski, D., Van Dolson, C., Mathews, J., Lulushi, A., Raglin, A.: The Analysis and Prediction of Eye Gaze When Viewing Statistical Graphs. In: D., S. and C., F. (eds.) Augmented Cognition. pp. 148–165. Springer, Cham (2017).
15. Bylinskii, Z., Borkin, M.A.: Eye Fixation Metrics for Large Scale Analysis of Information Visualizations. In: ETVIS Workshop on Eye Tracking and Visualization (2015).
16. Harel, J., Koch, C., Perona, P.: Graph-based visual saliency. In: Advances in neural information processing systems. pp. 545–552 (2007).
17. Harrison, A., Etienne-Cummings, R.: An entropy based ideal observer model for visual saliency. In: 2012 46th Annual Conference on Information Sciences and Systems (CISS). pp. 1–6. IEEE (2012).
18. Borji, A., Sihite, D.N., Itti, L.: Quantitative Analysis of Human-Model Agreement in Visual Saliency Modeling: A Comparative Study. IEEE Trans. Image Process. 22, 55–69 (2013).
19. Zhang, L., Tong, M.H., Marks, T.K., Shan, H., Cottrell, G.W.: SUN: A Bayesian framework for saliency using natural statistics. J. Vis. 8, 32.1-20 (2008).
20. Zhao, Q., Koch, C.: Learning a saliency map using fixated locations in natural scenes. J. Vis. 11, 1–15 (2011).
21. Zhao, Q., Koch, C.: Learning Visual Saliency. Conf. Inf. Sci. Syst. 1–6 (2011).