Modeling Human Comprehension of Data Visualizations

Michael J. Haass(✉), Andrew T. Wilson, Laura E. Matzen, and Kristin M. Divis

Sandia National Laboratories, Albuquerque, NM, USA
mjhaass@sandia.gov

Abstract. A critical challenge in data science is conveying the meaning of data to human decision makers. While working with visualizations, decision makers are engaged in a visual search for information to support their reasoning process. As sensors proliferate and high performance computing becomes increasingly accessible, the volume of data decision makers must contend with is growing continuously and driving the need for more efficient and effective data visualizations. Consequently, researchers across the fields of data science, visualization, and human-computer interaction are calling for foundational tools and principles to assess the effectiveness of data visualizations. In this paper, we compare the performance of three different saliency models across a common set of data visualizations. This comparison establishes a performance baseline for assessment of new data visualization saliency models.

Keywords: Visual saliency · Visualization · Modeling · Visual search

1 Introduction

A critical challenge in data science is conveying the meaning of data to human decision makers. While working with visualizations, analysts or decision makers are engaged in a visual search for information to support their reasoning process. As sensors proliferate and high performance computing becomes increasingly accessible, the volume of data that analysts must contend with is growing continuously. The resulting bloom of data and derived data products is driving the need for more efficient and effective means of presenting data to human analysts and decision makers. Consequently, researchers across the fields of data science, visualization, and human-computer interaction are calling for foundational tools and principles to assess the effectiveness of data visualizations [9]. In this paper, we describe the need for a computational model of bottom-up, stimulus-driven

Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000. SAND2016-1685 C.

S. Lackey and R. Shumaker (Eds.): VAMR 2016, LNCS 9740, pp. 125–134, 2016.
DOI: 10.1007/978-3-319-39907-2_12

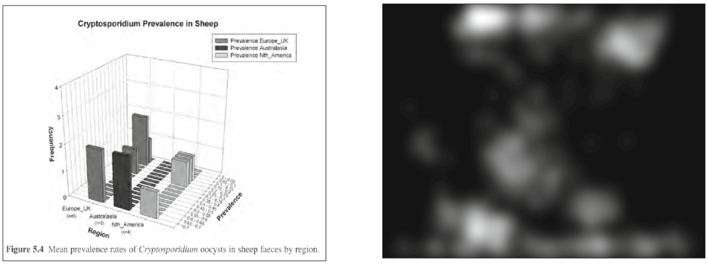
visual saliency that is appropriate for abstract data visualization. We compare the performance of three different saliency models across a common set of data visualizations to establish a performance baseline for assessment of new data visualization saliency models.

Human visual processing is guided by two parallel processes: bottom-up and top-down visual attention [16]. When viewing an image, a person's eye movements are guided by both the visual properties of the image that capture bottom-up attention (e.g. color, contrast, motion) and top-down processes such as task goals, prior experience, and use of search strategies [8]. Many bottom-up models are based on the neurophysiology of human and primate visual systems [1]. These models construct a number of features from the image data and then highlight differences in the features across multiple scales of image resolution. The chosen features are based on the response of neurons in the visual processing system to certain image characteristics such as luminance, hue, contrast and orientation. Various models have explored the use of different visual features at different scales to predict where humans will look in natural scene imagery.

Maps of bottom-up visual saliency have been valuable tools for studying how people process information in natural scenes, and could also be useful for evaluating the effectiveness of data visualizations. Ideally, the most important information in a data visualization would also have high visual saliency. This evaluation approach has been demonstrated with scene-like data visualizations [12], but it is unclear whether or not it is applicable to abstract data visualizations. In addressing this question, it is important to consider how visual search may differ between natural scene visualizations and abstract data visualizations. For the latter, viewers are engaged in drawing conclusions about causality, efficacy or consequences rather than identifying objects or properties of objects. The visual appearance of their target (information) may not be well defined or known ahead of time. The vast majority, if not all, existing computational models were developed and optimized to predict visual saliency for image-like, or natural, scenes and may not perform as well when applied to abstract data visualizations. In fact one published taxonomy of visual stimuli used in studies of gaze direction lists only three types of stimuli: psychophysics laboratory stimuli, static natural scenes, and dynamic natural scenes [15]. To date, we have been unable to find any published examples of bottom-up saliency models designed explicitly for data visualizations. In the following sections, we compare the performance of three high performing natural scene saliency models across a common set of data visualizations.

2 Method

The MIT Saliency Benchmark [7] is an online source of saliency model performance and datasets. The site scores and reports performance on author-contributed saliency models on datasets where the human fixation positions are not public. This approach prevents model performance inflation due to overfitting of the test dataset. We selected three saliency models, described below, listed on the MIT Saliency Benchmark site that span a range of performance on natural scenes when measured on standard stimuli with a common set of human gaze data. For baseline performance on natural scenes for each model, we used results for the cat2000 data set [4] because it is the most recent (introduced Jan 2015). MATLAB or Python code for each model was downloaded from saliency.mit.edu and saliency maps were constructed with each model on a set of data visualizations. We measured the performance of each model for the data visualizations using the same eight metrics used for the saliency benchmark project. We selected 184 example data visualizations from the Massachusetts (Massive) Visualization Dataset [6] with corresponding eye-movement data [5] from 33 viewers (average 16 viewers per visualization, minimum of 11, maximum of 22). Figure 1 shows an example data visualization and corresponding human fixation map. The MASSVIS samples were selected from infographic blogs, government reports, news media websites and scientific journals.



(a) (b)

Fig. 1. Example data visualization (a) and human fixation map (b).

2.1 Saliency Models

Itti, Koch and Nieber. Numerous saliency prediction models have been developed in recent years, taking a variety of approaches to predict which parts of an image are likely to draw a viewer's attention. Several of these approaches involve the creation of feature maps that are weighted, combined, and filtered to produce a visual saliency map. The most prominent of these models, the Itti, Koch and Niebur model [11], is based on the properties of the human visual system. The model detects changes in low-level features such as color, intensity, and orientation at varying spatial scales. It then weights those features and uses an iterative spatial competition process to create feature maps that are then summed to produce the saliency map. More recently, other researchers have developed new approaches to create saliency maps. When compared using the MIT Saliency Benchmark, two visual saliency models that consistently perform well with images of natural scenes are the Boolean Map based Saliency model [20,21] and the Ensembles of Deep Networks model [19].

Boolean Map Based Saliency. The Boolean Map based Saliency model (BMS) [20] creates a set of Boolean maps to characterize images. It relies on the Gestalt principle of figure-ground segregation and the idea that visual attention will be drawn to the figures in an image rather than the background. The model randomly thresholds an image's feature maps, such as the color map, to generate a set of Boolean maps. For each Boolean map, the model uses the feature of surroundedness [21] (a connected region with a closed outer contour) to identify figures within the image and to create an attention map. The attention maps are then normalized and combined to form the full-resolution attention map. This approach differs from many other saliency models because it utilizes scale-invariant information about the topological structure of the images. It does not use multi-scale processing, center-surround filtering, or statistical analysis of features. Thus, it is a relatively simple model that focuses on identifying figures within images.

Ensebles of Deep Networks. Like the classic Itti and Koch model, the ensembles of Deep Networks (eDN) model is hierarchical with operations that are based on the known mechanisms of the human visual cortex. However, rather than hand-selecting visual features of interest, a guided search procedure is used to optimize the model for identifying salient features. In other words, the saliency prediction task is a supervised learning problem in which the model is optimized for predicting where humans will look in natural scenes. Multiple high-performing models are identified and the combination of the models is optimized. Center bias and Gaussian smoothing are used to create the final saliency maps from the model outputs. For this comparison, the eDN model coefficients provided by Vig et al., learned using natural scene stimuli rather than data visualizations, were used to illustrate the difference in feature sensitivity across the two stimuli types. Future comparisons of learned model coefficients across the stimuli types could inform the development of saliency models for data visualizations. Figure 2 shows examples of each saliency model applied to the data visualization shown in Fig. 1.

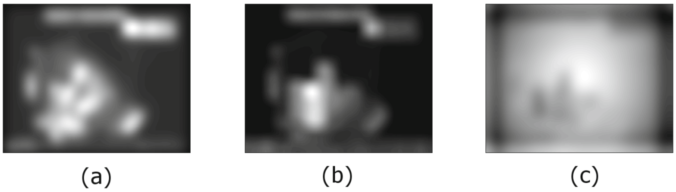


Fig. 2. Example saliency maps, (a) Itti, (b) BMS, (c) eDN, for data visualization shown in Fig. 1.

2.2 Comparison Metrics

Many different metrics have been proposed for comparing saliency and fixation maps. Riche et al. provide a thorough review and taxonomy of published comparison metrics [17]. The authors use a two-dimensional taxonomy to organize the various metrics. Along one dimension, they categorize the metrics as "value-based," "location-based" or "distribution-based." Along the other dimension, they categorize the metrics as "common," "hybrid" or "specific." Metrics categorized as common are generalized and were not originally designed for saliency comparisons. Metrics categorized as hybrid are adapted from other fields to work with saliency and fixation data. Metrics categorized as specific were developed directly for application to saliency comparisons. In order to compare model performance on natural scenes and data visualizations, we elected to use the eight comparison metrics used by the MIT Saliency Benchmark project. Of the eight metrics, one was value-based, three were location-based, and four are distribution-based, as described in more detail below.

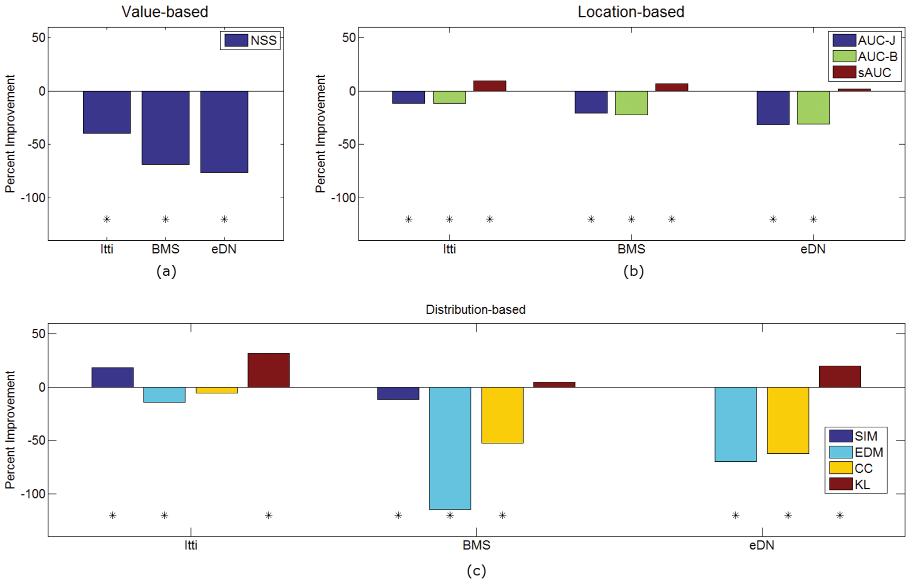
Value-Based Metric. The normalized scanpath saliency metric (NSS)[2] first standardizes saliency values to have zero mean and unit standard deviation, then computes the average saliency value at human fixation locations. When NSS is greater than one, the saliency map exhibits significantly higher values at fixation locations compared to other locations.

Location-Based Metrics. Three of the comparison metrics are based on the receiver-operator characteristic (ROC). For these metrics, the human gaze positions are considered positive examples and all other points are considered negative examples. The saliency map is treated as binary classifier to separate the positive and negative example sets at various thresholds and the area under the resulting ROC curve (AUC) is computed. As the saliency map and fixation map become more similar, AUC values approach one. Random chance agreement results in an AUC value of 0.5. For all three implementations, the true positive rate is the proportion of saliency values above the threshold at all fixation locations. For the AUC-Judd implementation the false positive rate is the proportion saliency values above the threshold at non-fixated locations and the thresholds are sampled from the saliency map values [17]. For the AUC-Borji implementation, the false positive rate is based on saliency values sampled uniformly from all image pixels and the thresholds are sampled with a fixed stepsize [3]. For the shuffled AUC implementation, the false positive rate is based on saliency values sampled uniformly from fixation locations on a random set of other images [3,22].

Distribution-Based Metrics. The similarity score (SIM) is a histogram intersection measure. Each distribution is scaled so that its sum is one. Similarity is the sum of the minimum value between the two scaled distributions at each point. When SIM equals one, the distributions are the same and when SIM equals zero, there is no overlap between the two distributions. The earth mover's distance (EMD)[18] is based on the minimal cost to transform one distribution (the saliency map) into the other distribution (the fixation map). Smaller values of EMD represent better agreement between the saliency map and the fixation map and when EMD equals zero, the two distributions are identical. The linear correlation coefficient (CC) is a measure of the linear relationship between a fixation map and a saliency map [2]. When CC is close to one, the linear relationship between the saliency map and the fixation map is nearly perfect. The Kullback Leibler divergence (KL)[10] is a measure of the information lost when the saliency map is used to approximate the fixation map. KL ranges from zero, when the two maps are identical, to infinity.

3 Experimental Results

Figure 3 shows the performance of the three models on the natural scenes and data visualizations. The results are displayed in the form of a percent difference score that is negative when the models performed better on natural scenes and positive when the models performed better on data visualizations. The corresponding numerical values are shown in Table 1. Table 2 shows the effect size, using Glass's delta across natural scenes and data visualization.



(a)

(b)

(c)

Fig. 3. Model Comparison Across Stimuli Type and Metric. (a) Value-based metric, (b) Location-based Metrics, (c) Distrbution-based Metrics. Results are displayed in the form of a difference score that is negative when the models performed better on natural scenes and positive when the models performed better on data visualizations.

Table 1. Model Comparison Across Stimulus Type. First value in each pair is sample mean; second value is standard error of the mean (SEM). Bold font indicates significant differences between mean values for natural scenes and visualizations ($p < 0.05$).

	Itti		BMS		eDN	
	Nat	Vis	Nat	Vis	Nat	Vis
AUC-J	**0.77 ± 0.002**	**0.68 ± 0.006**	**0.85 ± 0.001**	**0.67 ± 0.006**	**0.85 ± 0.001**	**0.58 ± 0.009**
SIM	**0.48 ± 0.002**	**0.57 ± 0.006**	**0.61 ± 0.002**	**0.54 ± 0.005**	0.52 ± 0.002	0.52 ± 0.005
EMD	**3.44 ± 0.016**	**3.92 ± 0.11**	**1.95 ± 0.013**	**4.19 ± 0.12**	**2.64 ± 0.013**	**4.48 ± 0.12**
AUC-B	**0.76 ± 0.002**	**0.67 ± 0.006**	**0.84 ± 0.001**	**0.65 ± 0.006**	**0.84 ± 0.001**	**0.58 ± 0.009**
sAUC	**0.59 ± 0.002**	**0.64 ± 0.007**	**0.59 ± 0.002**	**0.63 ± 0.006**	0.55 ± 0.002	0.56 ± 0.009
CC	0.42 ± 0.004	0.40 ± 0.017	**0.67 ± 0.002**	**0.32 ± 0.014**	**0.54 ± 0.002**	**0.20 ± 0.020**
NSS	**1.06 ± 0.012**	**0.64 ± 0.030**	**1.67 ± 0.012**	**0.52 ± 0.025**	**1.30 ± 0.006**	**0.30 ± 0.032**
KL	**0.92 ± 0.006**	**0.63 ± 0.019**	0.83 ± 0.012	0.79 ± 0.021	**0.97 ± 0.006**	**0.78 ± 0.018**

Table 2. Glass's Delta Effect Size for Model Comparison Across Stimulus Type. Bold font indicates significant differences between mean values for natural scenes and visualizations ($p < 0.05$). For normalization of Glass's delta, the natural scenes were treated as the control group.

	AUC-J.	SIM	EMD	AUC-B.	sAUC	CC	NSS	KL
Itti	**−0.98**	**1.23**	**0.66**	**−0.98**	**0.69**	−0.14	**−0.79**	**−1.16**
BMS	**−3.58**	**−1.00**	**3.79**	**−3.77**	**0.58**	**−3.21**	**−2.09**	−0.07
eDN	**−5.34**	−0.04	**3.07**	**−5.18**	0.14	**−4.22**	**−3.43**	**−0.67**

Generally, the models had poorer performance for data visualizations than for natural scenes. All three models performed worse on visualizations than on natural scenes as measured by four of the eight metrics: the value-based metric NSS, two location-based metrics, AUC-Judd and AUC-Borji, and the distribution-based metric EMD. For these metrics, the effect sizes were largest for the BMS and eDN models. The performance of the eDN model was not significantly different for visualizations and natural scenes when measured by the location-based metric sAUC and the distribution-based metric SIM. Similarly, the performance of the Itti model was not significantly different for visualizations and natural scenes when measured by the distribution-based metric CC. However, for the distribution-based metric KL, both the Itti and eDN models performed significantly better for data visualizations than natural scenes. This is consistent with the finding of Riche et al. [17] that the KL metric is quite different from the other metrics. Because the KL metric does not take absolute location into account, but considers only the statistical distribution of the map, two maps having similar distributions can have very different location properties. The performance of the BMS model was not significantly different for visualizations and natural scenes when assessed by the KL metric. For this metric, the effect size was largest for the Itti model followed by the eDN model, while the effect size for the BMS model was close to zero. Of note, for the metrics where the performance of all three models was significantly different between visualizations and natural scenes, the Itti model performed better on visualizations than either the BMS model or the eDN model. This is contrary to the general trend in performance on natural scenes for these metrics where eDN is the best performing saliency model.

4 Discussion and Conclusion

The visualizations used in this comparison study are all highly curated, employing text and graphic design principles to help viewers identify the most important results. The Itti model may perform best on these data visualizations because of its close ties to the human visual processing system, while other models have been designed and optimized for natural scenes, placing less emphasis on faithful representation of neural processes. The natural scene models may also under perform on data visualizations, since many graphical elements used in visualization have smaller spatial extent than objects that typically appear in natural scenes. The finer resolution graphical elements result in higher frequency components to which natural scene models maybe insensitive. Another factor that may limit the applicability of natural scene models is the use of text in data visualizations. Text plays a significant role in human attentional allocation and the resulting direction of eye movements. The process of reading text in a visualization would result in a higher density of fixations around text elements. Future work should leverage a taxonomy of visualization elements such as the one described in Munzner's book [13]. Our future research will focus on data visualization techniques for two-dimensional representation of high-dimensional data.

This comparison study has established a baseline that can be used to assess the performance of new saliency models for data visualizations. The current trend towards better model performance on natural scenes seems to come at the expense of performance on data visualizations. This inverse relationship between model performance on natural scenes and on data visualizations supports our position that new saliency models are needed to aid development of generalized theories of visual search for data visualizations. In future work, we will expand on existing models of visual saliency to address these issues and investigate the role of top-down visual attention in viewers' navigation of abstract data visualizations. Developing general models of top-down sense-making has proven to be quite difficult [14]. Knowledge elicitation techniques have been used to identify top-down goals and strategies and the resulting influence on eye movements. Other approaches have applied machine learning techniques to eye movement data collected as experts perform a given task. The resulting models can predict expert attention allocation for new stimuli, but it is often difficult to use these models to understand why experts allocate attention to certain content and not to other content. Because of this difficulty, we advocate the combination of computational models of bottom-up saliency with empirical studies of eye movements to identify tacit sense-making strategies.

As this work progresses, we will also explore the role of expertise in visual processing of data visualizations. Expertise is a crucial factor in top-down visual attention, and its impact may be even greater with abstract visualizations, where users cannot rely on their prior experience with real-world scenes to guide their search. Visual search tasks using abstract data visualizations can be contrasted with visual search tasks in complex decision making domains. For example, airport luggage screeners search x-ray imagery for prohibited items. In this domain, as in many abstract visualizations, the visual appearance of the target is often not known in advance and furthermore the target may be obscured by overlapping items. However, the users' knowledge about the image features may be quite different. Luggage screening personnel have extensive training and experience in how to search through images, but may have little expertise on the physics of the image formation process. In contrast, experts such as scientists and engineers who work with abstract data are likely to have very deep knowledge of the physical properties driving the content of visualizations. These differences should be considered as top-down factors are identified.

References

1. Borji, A., Itti, L.: State-of-the-art in visual attention modeling. IEEE Trans. Pattern Anal. Mach. Intell. **85**, 185–207 (2013)
2. 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)
3. Borji, A., Tavakoli, H.R., Sihite, D.N., Itti, L.: Analysis of scores, datasets, and models in visual saliency prediction. In: IEEE International Conference on Computer Vision (ICCV) (2013)
4. Borji, A., Itti, L.: Cat 2000: A large scale fixation dataset for boosting saliency research. In: CVPR 2015 Workshop on "Future of Datasets" (2015). arXiv:1505.03581
5. Borkin, M., Bylinskii, Z., Kim, N., Bainbridge, C.M., Yeh, C., Borkin, D., Pfister, H., Oliva, A.: Beyond memorability: visualization recognition and recall. IEEE Trans. Vis. Comput. Graph. **22**, 519–528 (2015). (Proceedings of InfoVis)
6. Borkin, M., Bylinskii, Z., Krzysztof, G., Kim, N., Oliva, A., Pfister, H.: Massachusetts (massive) visualization dataset. massvis.mit.edu
7. Bylinskii, Z., Judd, T., Borji, A., Itti, L., Durand, F., Oliva, A., Torralba, A.: Mit Saliency Benchmark
8. Connor, C.E., Egeth, H.E., Yantis, S.: Visual attention: bottom-up versus top-down. Curr. Biol. **14**(19), 850–852 (2004)
9. Green, T.M., Ribarsky, W., Fisher, B.: Building and applying a human cognition model for visual analytics. Inf. Vis. **8**, 1–13 (2009)
10. Itti, L., Baldi, P.: A principled approach to detecting surprising events in video. In: IEEE Computer Society Conference on Computer Vision and pattern Recognition (CVPR) (2005)
11. 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)
12. Matzen, L.E., Haass, M.J., Tran, J., McNamara, L.A.: Using eye tracking metrics and visual saliency maps to assess image utility. Paper Presented at the IS and T International Symposium on Electronic Imaging: Human Vision in Electronic Imaging, San Francisco, CA, USA (2016)

13. Munzner, T.: Visualization Analysis and Design. CRC Press, Boca Raton (2014)
14. Navalpakkam, V., Itti, L.: Modeling the influence of task on attention. Vis. Res. **45**, 205–231 (2005)
15. Peters, R.J., Itti, L.: Beyond bottom-up: incorporating task-dependent influences into a computational model of spatial attention. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2007)
16. Pinto, Y., van der Leij, A.R., Sligte, I.G., Lamme, V.A.F., Scholte, H.S.: Bottom-up and top-down attention are independent. J. Vis. **13**(3) (2013)
17. Riche, N., Duvinage, M., Mancas, M., Gosselin, B., Dutoit, T.: Saliency and human fixations: state-of-the-art and study of comparison metrics. In: IEEE International Conference on Computer Vision (ICCV) (2013)
18. Rubner, Y., Tomasi, C., Guibas, L.J.: The earth mover's distance as a metric for image retrieval. Int. J. Comput. Vis. **40**, 99–121 (2000)
19. Vig, E., Dorr, M., Cox, D.: Large-scale optimization of hierarchical features for saliency prediction in natural images. In: IEEE Computer Vision and Pattern Recognition (CVPR) (2014)
20. Zhang, J., Sclaroff, S.: Saliency detection: a Boolean map approach. In: Proceedings of the IEEE International Conference on Computer Vision (ICCV) (2013)
21. Zhang, J., Sclaroff, S.: Exploiting surroundedness for saliency detection: a Boolean map approach. IEEE Trans. Pattern Anal. Mach. Intell. (TPAMI) **38**, 889–902 (2015)
22. Zhang, L., Tong, M.H., Marks, T.K., Shan, H., Cottrell, G.W.: Sun: a Bayesian framework for saliency using natural statistics. J. Vis. **16**, 1–20 (2008)