AGILE: GIScience Series, 5, 51, 2024. https://doi.org/10.5194/agile-giss-5-51-2024
Proceedings of the 27th AGILE Conference on Geographic Information Science, 4–7 Sept. 2024.
Editors: Alison Heppenstall, Mingshu Wang, Urska Demsar, Rob Lemmens, and Jing Yao.
This contribution underwent peer review based on a full paper submission.

EyeCatchingMaps, a Dataset to Assess Saliency Models on Maps

Laura Wenclik [1] and Guillaume Touya [1]

[1]LASTIG, Univ Gustave Eiffel, IGN-ENSG, F-77420 Champs-sur-Marne, France

Correspondence: Laura Wenclik (laura.wenclik@ign.fr)

Abstract. Saliency models try to predict the gaze behaviour of people in the first seconds of their observation of an image. To assess how much these models can perform to predict saliency in maps, we lack a ground truth to compare to. This paper proposes EyeCatchingMaps, an open dataset that can be used to benchmark saliency models for maps. The dataset has been obtained by recording the gaze of participants looking at different maps for 3 seconds with an eye-tracker. The use of EyeCatchingMaps is demonstrated by comparing two different saliency models from the literature to the real saliency maps derived from people's gaze.

Keywords. cartography, maps, saliency, eye-tracking

1 Introduction

Maps usually are visual stimuli with a clear hierarchy, with foreground elements that attract attention first, and background elements that can be explored when the map is read with care. However, it is not always easy or clear for a map designer what the foreground salient elements of the maps are. Saliency models developed in the past 25 years in computer vision try to model and predict this visual hierarchy in different kinds of images and now achieve great success for natural scene images (Borji et al., 2019). A recent experience with paintings shows that the saliency models designed for photographs do not predict as well the saliency for other types of images (Le Meur et al., 2020). Though the importance of research on saliency in maps has been identified to guide map design (Fabrikant et al., 2010), there has been only one attempt to measure or model the salient parts of a map (Krassanakis et al., 2013). In this work, saliency was measured with an eye-tracker, which is an important tool to understand how maps are used, as we can see for instance in the two recent reviews on the use of eye-tracking in research on cartography (Keskin and Kettunen, 2023; Fairbairn and Hepburn, 2023).

The first challenge to study how to model visual saliency in maps is to build a ground truth that models can then try to predict as much as possible. This ground truth should be a dataset of map images with their saliency measured by eye-tracking, similarly to MIT300[1] (Judd et al., 2012), the main benchmark for visual saliency. We address this challenge with the EyeCatchingMaps dataset, a set of 322 maps with their measured saliency maps. The paper is structured as follows: Section 2 reviews past saliency models and benchmarks. Section 3 describes how EyeCatchingMaps was experimentally produced. Section 4 shows how several saliency models from the literature perform on predicting saliency on EyeCatchingMaps images.

2 Visual Saliency Models

A saliency model predicts for a pixel (x,y) the probability to be fixated by someone looking briefly at the image (Kümmerer et al., 2018). From the seminal Itti-Koch model (Itti et al., 1998) until now, many saliency models were proposed, as shown by two recent literature review papers (Borji et al., 2019; Cong et al., 2019). while the first models were based on unsupervised image processing techniques, the recent models are all based on supervised deep neural networks. From these saliency models, the usual output is a saliency map, i.e. a grayscale image showing the salient pixels of the image (Figure 1).

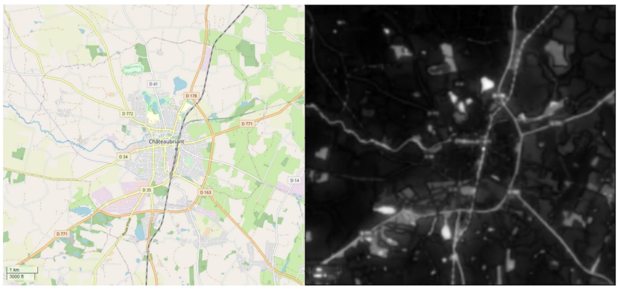

Figure 1. A map extract on the left (©OpenStreetMap contributors), and the derived saliency map using the Itti-Koch saliency model, on the right.

Besides saliency maps, research on visual saliency is also interested in the detection of salient objects from saliency

[1] http://saliency.mit.edu/home.html

models (Borji et al., 2019). As shown in the example of Figure 2, the detection of salient objects mixes the saliency map with an object detection problem to find the complete objects that stand out in the image, which makes even more sense in a map composed of separated symbols. Some models also allow the prediction of scan paths (Le Meur et al., 2020), i.e. the path of the most probable gaze during the first seconds of looking at the image, which somehow orders the salient regions of the images.

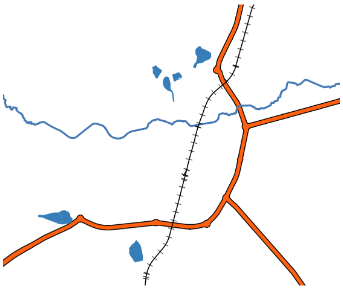

Figure 2. Salient objects detected on the previous map extract following the same saliency model, with a threshold on the probability.

During the past 15 years, research on visual saliency has been fostered by the datasets compiled by the researchers in the domain. From the initial MIT300 (Judd et al., 2012) to the current MIT/Tuebingen Saliency Benchmark (Kümmerer et al., 2018), there are now dozens of datasets giving ground truth data for different types of images, but none is related to maps.

3 EyeCatchingMaps

3.1 Apparatus and participants

To create this dataset, we use an eye tracker to follow a person's gaze. The eye tracker used is the Pupil Core from Pupil Labs. The configurations used are those of the basic pupil capture software, i.e. the time for a fixation point is 80 ms to 220 ms. The maps were displayed in a local web application on a 23.8" screen with a resolution of 1920x1080 pixels.

The participants in this survey were mainly members of our institution. A few students from the university also took part in the survey. A call for participation was sent out via internal mailing lists. There were no external participants. Participants without sight problems were privileged to facilitate the calibration of the eye tracker. There were 44 participants (31 male, 13 female), but in some cases, one or more sessions could not be used due to the loss of the calibration of the eye-tracker during the session. However, each map has been seen at least 20 times. The age distribution is presented in Table 1

The experiment was conducted according to the principles expressed in the Declaration of Helsinki: participants were instructed on the experiment goal and gave consent to participate in the experiment, by validating a consent form. Participant's names were never recorded and eye-tracking data were analyzed anonymously.

	-24	**25-34**	**35-44**	**45-54**	**55+**
nb of participant	17	23	1	1	2

Table 1. distribution of the ages

city	14	16	18	
country	14	16	18	
river	**14**	**16**	18	
mountain	**12**	14	16	
seaside	10	12	14	16
monument	**16**	**18**		
country and city	8	10	12	**14**

Table 2. Distribution of the zoom levels where maps were extracted for each type of landscape. The bold numbers correspond to scales where we use a map centred and a map decentred.

3.2 Stimuli

The dataset consists exclusively of maps. It is made up of 322 different maps. These different maps cover different dimensions of map types. There are several landscapes (city, country, with a large river, mountain, seaside, with a salient monument, country+city), three topographic styles (OpenStreetMap, Google Maps, Plan IGN), and different scales ranging from zoom level 8 to zoom level 18. Table 2 lists the different zoom levels at which maps were extracted for each type of landscape.

All these different elements are represented in two image formats: large format 1704x856 pixels, to simulate maps looked at on computer screens, and small format 290x553 pixels to simulate maps looked at on a smartphone screen. At some scales and for specific landscapes, we selected two maps, one where the landscape element (e.g. the river, the city or the monument) is located in the centre of the map, and one where this landscape element is located in a corner of the image (see bold scales in the table). We make this distinction because visual saliency is known to be skewed towards the centre of the image, and we wanted to verify this bias with maps (Figure 3).

In addition, to complete the dataset, we randomly selected around forty maps from the three base maps and a Google Maps dataset of 37 existing maps (Keskin et al., 2023). Finally, thematic maps were extracted from open geography textbook maps[2]. The legend has been removed from these maps because that is not what we want to test in this benchmark.

[2] http://www.cartolycee.net/

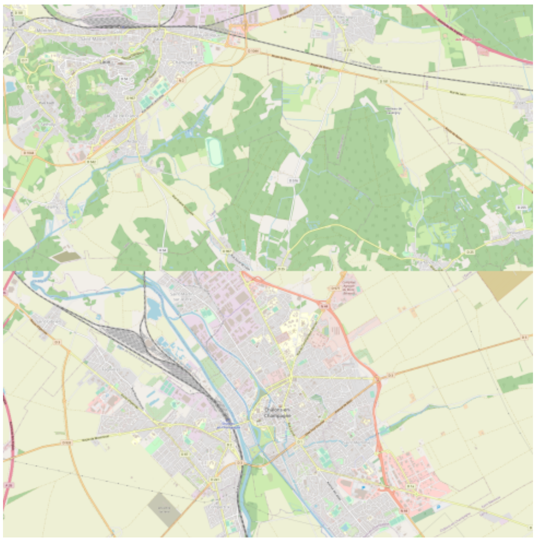

Figure 3. Example of a dataset map with centred and non-centred elements. The first map shows the non-centred city and the second shows the centred city.

3.3 Procedure

The dataset has been split into two parts to make the experiment shorter for the participants. In addition, looking at more than 150 maps consecutively is cognitively intense for the participant, so it is preferable to divide the survey into four sessions of around 40 maps to make sure the participants get some rest between each session and to check the eye-tracker settings between each session. This also prevents the user from losing concentration during the session. The four sessions are composed of a thematic map session, a portable format map session, a full-screen session and the last one is either a full-screen or Google Maps-only session. The four sessions are displayed randomly, but the maps in each session are in a fixed order.
Based on the various benchmark protocols already in use, the maps are displayed on screen for 3 seconds on a grey background, with a 0.5-second pause between each map. The grey screen avoids being stimulated between two images. Each session lasts approximately 2 minutes and 40 seconds.

During the experiment, the instructions given to the participants were the following: "Look at the different maps as if you were discovering a map for the first time, there is no need to remember the elements of the map. There is no question at the end."

The survey begins with a consent form, after which the principle of salience and the eye tracker are explained to the participants, along with the reasons for the survey.

The eye tracker is calibrated and recorded using the Pupil Capture tool from Pupil Labs. Instructions are then given to the participant. The participant is notified when they reach the halfway point in the experiment and when the experiment is about to end.

3.4 Post-Processes

Once the survey has been completed, the first step in post-processing the raw data is to locate the fixation points (i.e. when the gaze is fixed between 80-220ms in one place) on the different maps. There are several stages to this. Firstly, you need to locate the fixation points on the screen. The eye tracker used does not allow the fixation points on the screen to be determined directly. To do this, we need to place QR codes on the corners of the screen so that in post-processing we can detect the screen in this scene. After this first step, we obtain the screen coordinates of the fixation points. We also recorded the display time of each map. This means that for each fixation point, we can determine which image was displayed at that time. Just as we know the position and dimensions of each map, we can now determine the map coordinates of each fixation point.

To determine the area the user is looking at, we need to know how accurate our data is. To do this, we first need to determine the distance between the eye-tracker and the screen and thus the accuracy of the eye-tracker on the screen. From the data given by the eye-tracker, we get an accuracy angle for one session, and we can derive the precision of the eye-tracker in screen pixels for Equation 1 where pixel-size $= 0.02$, and where mean-distance is the distance between the eyes of the participant and the screen. This distance is derived from the raw data provided by the eye-tracker and from the size of the QR codes surrounding the screen (2.3 cm in our experiment).

We then use the following formula to obtain the distance in pixels for the precision of the eye tracker:

$$\text{pixel-screen} = \frac{\tan(\text{mean-angle-accuracy}) * \text{mean-distance}}{\text{pixel-size}} \tag{1}$$

After obtaining the distance in pixels, we use a Gaussian filter to calculate the saliency map. To determine the sigma parameter of the Gaussian filter, we use the precision calculated in Equation 1. Specifically, we divide the distance by 2.355, which is derived from the relationship between the standard deviation (sigma) and the maximum width at half-height of the Gaussian distribution.

3.5 Description of the Final Dataset

For each map, an initial file containing all the fixation points from all participants is generated, named 'coord_fixation_name_map.png.csv' in CSV format with the following elements:

- **world_index**: the id of the fixation;
- **id_fixation**: the id of the fixation point;
- **time**: the time of the fixation point based on the time of the survey;

- **x**: x coordinate of the fixation point on the image;
- **y**: y coordinate of the fixation point on the image;
- **dispersion**: distance between all gaze locations during a fixation, in degree;
- **accuracy**: in degree;
- **precision**:in degree;
- **participant**: Participant ID;
- **distance**: distance between eye-tracker and screen in cm;
- **time_to_map**: the time of the fixation point starting at the time the current map was first displayed;
- **height** : height of the map
- **width** : width of the map

It is thus possible to produce a scan path for each file, which retraces the path of each participant's gaze (Figure 4).

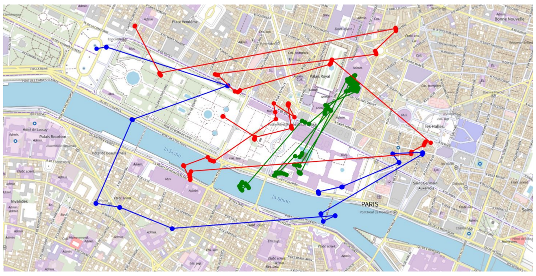

Figure 4. Example of the scan path of three participants on one of the maps in the dataset obtained using eye-tracker. Each colour represents a participant. The scan is composed of points representing each of the user's fixation points.

We also produce a heatmap file for each map named 'heatmap_name_map.png' (Figure 5).

4 Evaluation of the Dataset

As EyeCatchingMaps aims to benchmark existing saliency models, we tried to compare two models from the literature to verify that our dataset can serve this purpose. We chose two unsupervised methods to model saliency, namely Covsal (Erdem and Erdem, 2013) and Fast and Efficient Saliency (FES) (Rezazadegan Tavakoli et al., 2011), and we tested them on 16 of our maps. CovSal uses region covariance descriptors to better capture the local features of images that may explain saliency. FES searches for portions in the image where there is a significant contrast between the pixels of the region and the pixels surrounding this region. These two methods were chosen because they were both easy to reuse and efficient on the MIT300 benchmark. To compare these two methods to the ground truth from our dataset, we analysed the saliency maps generated by these methods. We compared them with the corresponding heatmap in our dataset. Among the metrics existing to compare saliency models, we selected the Pearson Correlation Coefficient (CC) (Riche et al., 2013) that is close to 1 when the saliency map matches the ground truth heatmap. The results are compiled in Table 3. The mean CC value for CovSal on the 16 maps is 0.724, which is higher than the performance of CovSal on the MIT300 benchmark (0.500). Similarly, FES performs better on these 16 maps than with MIT300, with a 0.664 mean CC value (0.483 on MIT300). Those results tend to show that maps in EyeCatchingMaps are images where saliency is easier to predict than in the natural scene images contained in MIT300.

id of map	CC for CovSal	CC for FES
1	0.875	0.899
2	0.486	0.930
3	0.932	0.937
4	0.336	0.130
5	0.491	0.610
6	0.800	0.796
7	0.700	0.672
8	0.724	0.579
9	0.817	0.659
10	0.932	0.904
11	0.582	0.363
12	0.741	0.345
13	0.709	0.649
14	0.685	0.434
15	0.814	0.852
16	0.960	0.858

Table 3. Comparison between EyeCatchingMaps ground truth and the saliency maps derived from CovSal (Erdem and Erdem, 2013) and FES (Rezazadegan Tavakoli et al., 2011) saliency models. We use Pearson correlation coefficient (CC) (Riche et al., 2013) to compare the two heatmaps.

Figure 6 shows how both methods similarly predict the saliency of this smartphone-format map. In this case, the most salient part of the map is the plaza with the city hall, in the centre of the map, and both methods predict this plaza as a salient hot spot, even though they slightly disagree with the saliency of the surroundings of the plaza. Figure 7 shows a map where the saliency maps from both methods differ greatly. In this map of the world extracted from a geography textbook, the ground truth from eye tracking shows that the salient regions are mainly the ones in dark purple, with a significant bias around the Mediterranean Sea. FES correctly predicts this saliency, but CovSal predicts that the salient part is mainly the centre of Africa, which is not consistent with ground truth. This example confirms that different saliency models may have different predictions on different types of maps.

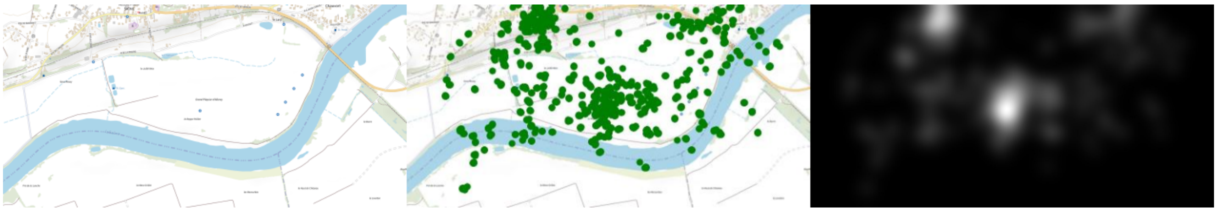

Figure 5. The visual outputs of the EyeCatchingMaps dataset: on the left, one of the 322 initial maps, The same map with the fixation points of all participants in the middle, and the ground truth saliency map on the right.

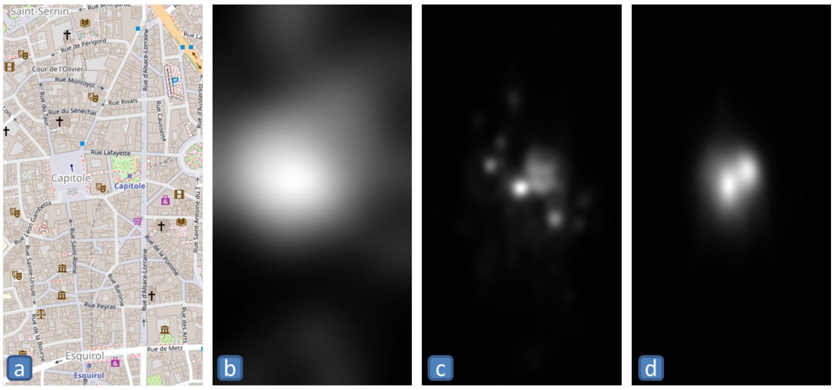

Figure 6. Results of CovSal and FES methods on image 15 from EyeCatchingMaps: (a) the map; (b) the heatmap from eye-tracking; (c) CovSal saliency map; (d) FES saliency map.

4.1 Data and Software Availability

The data and the code used in this research can be found on Zenodo (https://zenodo.org/doi/10.5281/zenodo.10619512).

5 Conclusion and Future Research

To conclude, EyeCatchingMaps is a dataset that can help researchers or cartographers compare, calibrate and choose the visual saliency models of the literature. It can also serve as a basis for the development of a saliency model that better predicts visual saliency in maps.

Besides this straightforward use of the dataset to benchmark saliency models, we think that it would be interesting to analyse the dataset in terms of map design. what are the map features that are more frequently found in salient regions of a map? What are the most salient objects in maps? What are the visual variables that characterise the salient regions? How does map saliency change with map style? Are saliency models very different for small-scale and large-scale maps? All these questions and many others can be explored just by analysing the eye-tracking data from EyeCatchingMaps, and this is what we plan to do in the future.

Finally, the best generic saliency models are now deep models trained on pictures of natural scenes, and these models can be fine-tuned to achieve very good results on specific types of images such as paintings (Le Meur et al., 2020). EyeCatchingMaps could also be used to fine-tune such models and make them more successful on maps.

Author contributions. Laura Wenclik: Conceptualization, Methodology, Software, Data curation, Writing- Original draft preparation. Guillaume Touya: Writing- Original draft preparation, Reviewing and Editing, Conceptualization, Methodology, Supervision, Funding acquisition, Project administration.

Competing interests. The authors declare that no competing interests are present.

Acknowledgements. This research has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No. 101003012). The authors would like to thank all the participants of these projects, particularly for all the sessions of brainstorming that guided this research.

References

Borji, A., Cheng, M.-M., Hou, Q., Jiang, H., and Li, J.: Salient object detection: A survey, Computational Visual Media, 5, 117–150, https://doi.org/10.1007/s41095-019-0149-9, 2019.

Cong, R., Lei, J., Fu, H., Cheng, M.-M., Lin, W., and Huang, Q.: Review of Visual Saliency Detection With Comprehensive Information, IEEE Transactions on Circuits and Systems for Video Technology, 29, 2941–

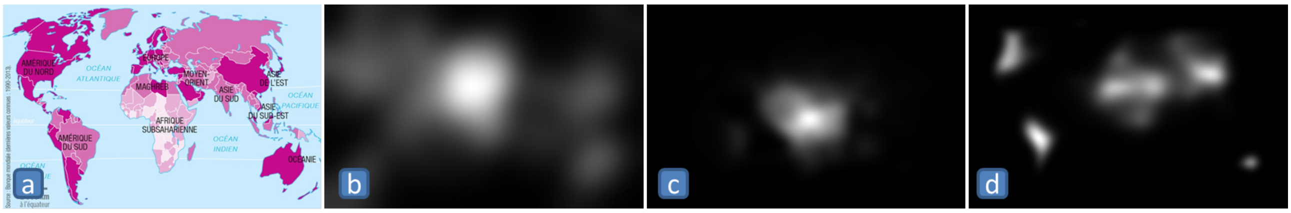

Figure 7. Results of CovSal and FES methods on image 2 from EyeCatchingMaps: (a) the map; (b) the heatmap from eye-tracking; (c) CovSal saliency map; (d) FES saliency map.

2959, https://doi.org/10.1109/TCSVT.2018.2870832, conference Name: IEEE Transactions on Circuits and Systems for Video Technology, 2019.

Erdem, E. and Erdem, A.: Visual saliency estimation by nonlinearly integrating features using region covariances, Journal of Vision, 13, 11, https://doi.org/10.1167/13.4.11, 2013.

Fabrikant, S. I., Hespanha, S. R., and Hegarty, M.: Cognitively Inspired and Perceptually Salient Graphic Displays for Efficient Spatial Inference Making, Annals of the Association of American Geographers, 100, 13–29, https://doi.org/10.1080/00045600903362378, 2010.

Fairbairn, D. and Hepburn, J.: Eye-tracking in map use, map user and map usability research: what are we looking for?, International Journal of Cartography, 9, 231–254, https://doi.org/10.1080/23729333.2023.2189064, 2023.

Itti, L., Koch, C., and Niebur, E.: A Model of Saliency-Based Visual Attention for Rapid Scene Analysis, IEEE Trans. Pattern Anal. Mach. Intell., 20, 1254–1259, https://doi.org/10.1109/34.730558, 1998.

Judd, T., Durand, F., and Torralba, A.: A Benchmark of Computational Models of Saliency to Predict Human Fixations, in: MIT Technical Report, 2012.

Keskin, M. and Kettunen, P.: Potential of eye-tracking for interactive geovisual exploration aided by machine learning, International Journal of Cartography, 9, 150–172, https://doi.org/10.1080/23729333.2022.2150379, 2023.

Keskin, M., Krassanakis, V., and Çöltekin, A.: Visual Attention and Recognition Differences Based on Expertise in a Map Reading and Memorability Study, ISPRS International Journal of Geo-Information, 12, 21, https://doi.org/10.3390/ijgi12010021, number: 1 Publisher: Multidisciplinary Digital Publishing Institute, 2023.

Krassanakis, V., Lelli, A., Lokka, I.-E., Filippakopoulou, V., and Nakos, B.: Searching for salient locations in topographic maps, in: SAGA 2013: 1st International Workshop on Solutions for Automatic Gaze Data Analysis, Bielefeld, Germany, https://doi.org/https://doi.org/10.2390/biecoll-saga2013_11, 2013.

Kümmerer, M., Wallis, T. S. A., and Bethge, M.: Saliency Benchmarking Made Easy: Separating Models, Maps and Metrics, in: Computer Vision – ECCV 2018, edited by Ferrari, V., Hebert, M., Sminchisescu, C., and Weiss, Y., Lecture Notes in Computer Science, pp. 798–814, Springer International Publishing, 2018.

Le Meur, O., Le Pen, T., and Cozot, R.: Can we accurately predict where we look at paintings?, PLOS ONE, 15, e0239 980, https://doi.org/10.1371/journal.pone.0239980, publisher: Public Library of Science, 2020.

Rezazadegan Tavakoli, H., Rahtu, E., and Heikkilä, J.: Fast and Efficient Saliency Detection Using Sparse Sampling and Kernel Density Estimation, in: Image Analysis, edited by Heyden, A. and Kahl, F., Lecture Notes in Computer Science, pp. 666–675, Springer, Berlin, Heidelberg, https://doi.org/10.1007/978-3-642-21227-7_62, 2011.

Riche, N., Duvinage, M., Mancas, M., Gosselin, B., and Dutoit, T.: Saliency and Human Fixations: State-of-the-Art and Study of Comparison Metrics, in: 2013 IEEE International Conference on Computer Vision, pp. 1153–1160, https://doi.org/10.1109/ICCV.2013.147, iSSN: 2380-7504, 2013.