

UEyes: An Eye-Tracking Dataset across User Interface Types

Yue Jiang

yue.jiang@aalto.fi

Aalto University

Finland

Hamed R. Tavakoli

hamed.rezazadegan_tavakoli@nokia.com

Nokia Technologies

Finland

Luis A. Leiva

name.surname@uni.lu

University of Luxembourg

Luxembourg

Julia Kylmälä

julia.kylmala@aalto.fi

Aalto University

Finland

Paul R. B. Houssel

name.surname@uni.lu

University of Luxembourg

Luxembourg

Antti Oulasvirta

antti.oulasvirta@aalto.fi

Aalto University

Finland

ABSTRACT

Different types of user interfaces differ significantly in the number of elements and how they are displayed. To examine how such differences affect the way users look at UIs, we collected and analyzed a large eye-tracking-based dataset, *UEyes* (62 participants, 1,980 UI screenshots, near 20K eye movement sequences), covering four major UI types: webpage, desktop UI, mobile UI, and poster. Furthermore, we analyze and discuss the differences in important factors, such as color, location, and gaze direction across UI types, individual viewing strategies and potential future directions. This position paper is a derivative of our recent paper [9] with a particular focus on the *UEyes* dataset.

ACM Reference Format:

Yue Jiang, Luis A. Leiva, Paul R. B. Houssel, Hamed R. Tavakoli, Julia Kylmälä, and Antti Oulasvirta. 2023. *UEyes: An Eye-Tracking Dataset across User Interface Types*. In *Proceedings of ACM Conference (Conference'17)*. ACM, New York, NY, USA, 5 pages. <https://doi.org/10.1145/nnnnnnnn.nnnnnnn>

1 INTRODUCTION

The study of what captures users' attention when interacting with user interfaces (UIs) has been a longstanding area of interest in HCI research. This understanding is essential for designers seeking to direct users' attention, communicate important information, and prevent visual overload [17, 18]. Despite years of research on this topic, our understanding of how different types of UIs vary regarding visual saliency remains limited. For example, posters usually feature only a few images, whereas desktop and mobile UIs tend to include more components, organized as widgets. Understanding how these differences affect eye-movement patterns is essential.

In this paper, we present *UEyes*, a new eye-tracking dataset captured using a high-fidelity in-lab eye tracker on a large scale. While previous studies relied on mouse movements or manual annotations as proxies for eye movements, *UEyes* provides fine-grained ground-truth data on visual saliency. Our dataset includes around 20K multi-duration saliency maps and scanpaths of 62 users who

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

Conference'17, July 2017, Washington, DC, USA

© 2023 Association for Computing Machinery.

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM...\$15.00

<https://doi.org/10.1145/nnnnnnnn.nnnnnnn>

viewed a total of 1,980 different UIs, comprising 495 UIs each from desktop, mobile, webpage, and poster applications. Furthermore, we analyze and compare saliency-related tendencies across different types of UIs, addressing both bottom-up factors related to the visual primitives of the stimulus (e.g., color bias) and top-down (learned) factors related to the distribution of features in the dataset (e.g., location bias and scanpath direction). We present several previously unreported findings that illuminate what distinguishes particular UI types.

2 EXISTING VISUAL SALIENCY DATASETS

Most visual saliency datasets available today focus on specific types of designs and provide a relatively small number of saliency results. These datasets are often limited to a particular type of visual design, with data collected from a small group of participants in a restricted context. For example, datasets such as MASSVIS [3], iSUN [21], SALICON [8], MIT1003 [11], MIT300 [10], NUSeF [16], and Leiva et al. [14] typically encompass only one specific type of image, such as indoor and outdoor natural images, mobile user interfaces, visual flows in comics, webpages, and posters. While CAT2000 [2] includes 20 categories, all of them are classes of natural images, with additional augmented natural images such as sketches and cartoons, and noisy natural images like low-resolution scenes and Gaussian-noised images. However, saliency prediction models trained on specific image types are often limited in their ability to adapt to broader types of images and cannot generalize to predict visual saliency for a wide range of them. To address this limitation, we collected the *UEyes* dataset for this work, which includes eye-tracking data for four common categories of user interfaces and a wide variety of images, with a focus on diverse visual designs. By including a broad range of UIs, our dataset provides a valuable resource for researchers seeking to develop saliency prediction models that can adapt to different types of images and generalize across various domains.

While prior research has explored the use of crowdsourcing to collect saliency-related data (e.g., Imp1k [6] and SALICON [8]), these methods cannot provide the high-fidelity in-lab eye-tracking data that is crucial for accurate results. Proxy sensors, such as webcams and cursor movements, present their own issues, including accuracy issues with webcam-based methods [21] during facial landmark tracking, eye region extraction, and calibration, and slower, more deliberative cognitive processes with cursor-based approaches [1, 8, 12, 13] than those involved in eye movements. To address these limitations, we collected real-time eye-tracking data

via an eye tracker and focused on common UI types to systematically analyze and compare user viewing behavior.

3 DATASET COLLECTION

The UEyes dataset is composed of both the 1,980 UI screenshots and the associated metadata and eye-tracking logs from 62 viewers, collected in a laboratory by means of a modern eye tracker. This dataset contains 495 screenshots from each of the following UI types:

Webpage: We collected 494 webpage images from the Alexa 500 dataset [19], 1,507 images from the Visual Complexity and Aesthetics dataset [15], and 200 images from the Imp1k dataset [6]. We extended the breadth of the webpage image set by capturing 103 additional webpage screenshots.

Desktop UI: The desktop UI image set contains the Waltteri Github desktop UI dataset [4], representing 51 desktop UIs, and an additional 303 desktop UI images collected in line with the criteria presented below.

Mobile UI: We extracted a sample of 1,761 images from among the 46,064 mobile UI images from the RICO dataset [5]. We extended the set with 42 further mobile UI images.

Poster: The poster image set contains 200 ads and 198 infographics from the Imp1k dataset [6], along with 103 additional posters we collected.

To ensure a diverse and representative dataset, we added new images that were either substantially different from those in the pre-existing dataset or commonly used in day-to-day life. For mobile UI images, we collected apps from various categories, including school, library, music, and settings. We also added more desktop UI images to create a balanced final dataset. We filtered out any images containing pornography and randomly sampled images of each type to create 55 "image blocks" for user assessment. Each block consisted of nine images representing each UI type, for a total of 36 images per block. To mimic users' typical viewing experience, we adjusted the screen angle for each participant during the data collection process. We used the same visual angle for all UI types, including mobile UIs, to ensure a fair comparison. This consistent presentation across types prevented the tracking technology's accuracy limits from disproportionately affecting the mobile UI results, enabling us to conduct consistent data collection and analysis across all UI types.

Participants. We recruited a total of 66 participants, comprising 23 males and 43 females, through mailing lists and social media promotions. The average age was 27.25 (SD = 7.26). All participants had normal vision, with 43 of them having uncorrected vision and 23 wearing glasses or contact lenses for corrected-to-normal vision. None of the participants were colorblind. We excluded the gaze data of four participants due to inaccurate eye-tracking calibration. Each user participated in the study for an hour and received 30 EUR as compensation for their time and effort.

Experimental Design. Our system selected nine blocks randomly from a pool of 55 blocks for each user, resulting in a total of 324 images shown to each user. We tried to make sure each block has been shown to about the same number of users. Each block contained nine images for each UI type. The order of presentation for the images within each block was randomized.

Apparatus. We presented the images on a 24-inch HP Compaq LA2405wg desktop monitor with dimensions of 32.5×52 cm and a resolution of 1920×1200 px. To collect high-quality gaze data, we used a Gazepoint GP3 eye tracker with a sampling rate of 60 Hz. The eye tracker was positioned under the screen and tilted upwards, with its angle adjusted for each participant. The eye-tracking software, Gazepoint Control, provided feedback to ensure that the participant was seated at a distance of approximately 50–65 cm from the tracker.

Procedure. The tracker was calibrated using Gazepoint Control's nine-point calibration and tested on the calibration test screen. To ensure the quality of the data in the post-processing stage, each participant was then shown three images of grids with different sizes and instructed to look at the corners of the grids, starting from the top left and moving clockwise. Following the calibration, each participant completed nine blocks with self-managed breaks. In each block, the participant was presented with UI images and asked to examine them for seven seconds, imagining they were in a corresponding real-world situation. No specific task was assigned, following the methodology of other bottom-up saliency studies. After the last block of UI images, the participant filled out a demographics questionnaire.

Data Processing. We double-checked the collected data to guarantee the dataset's quality, and we removed any user data exhibiting inaccurate calibration or duplicate results. Accordingly, the final dataset contains 94.86% of the raw data collected. Fixations beyond image boundaries (6.8% of the fixations) were not considered for analysis.

3.1 Components of the Dataset

In the following we describe the key features of our dataset.

Design images: 495 images of each UI category, 1,980 images in total (55 blocks of 36 images each).

Eye-tracking logs: 554 raw logs from the eye tracker in CSV format. Each log includes eye movement data for one participant and one image block.

Image types: Categorization of each image in a separate CSV file. Each image can belong to only one category.

Multi-duration saliency maps: Saliency maps for 1 s, 3 s, and 7 s of free-viewing. Each fixation is weighted by the time duration of each fixation.

Scanpaths: Sequences of fixations for participants. Figure 1 shows examples of saliency maps and scanpaths.

Segmentation information: A JSON file for each UI with bounding boxes of detected images, texts, and faces. We modified the UIED model [20] by (1) solving a model limitation that ignored text detection and (2) integrating face detection with OpenCV face detection approach using Haar feature-based cascade classifiers [7].

4 HOW PEOPLE LOOK AT UIS

We further analyzed our UEyes dataset and found that users tend to focus more on the upper-left region of a user interface, regardless of the type of UI, which is consistent with prior findings on mobile UIs [14]. Additionally, we have observed that when users make

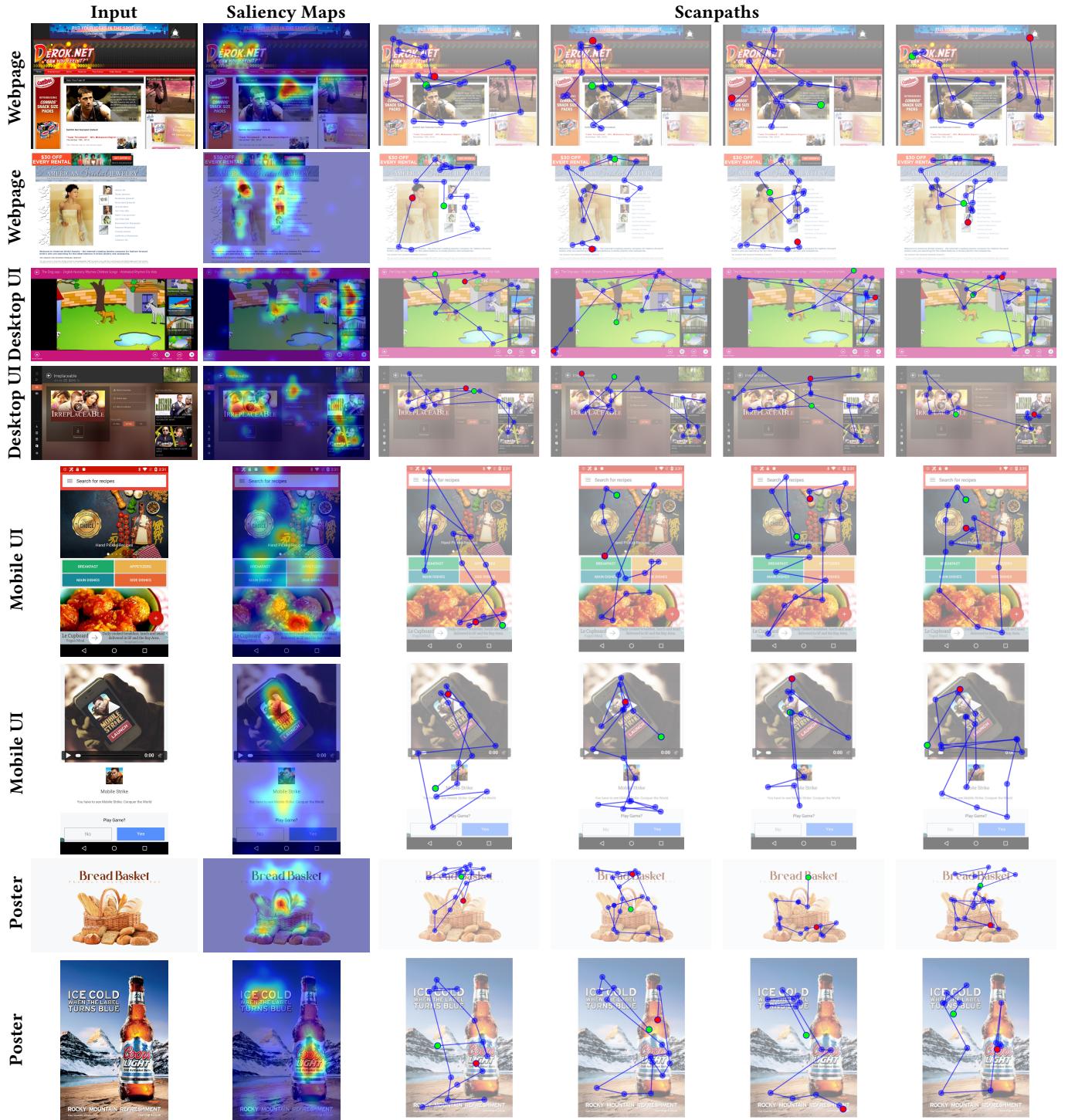


Figure 1: Examples of saliency maps and scanpaths in the UEyes dataset.

saccades, their gaze tends to move towards the right or bottom portion of the UI. We have also found that movements towards the

right tend to cover larger distances between consecutive points than movements towards the bottom.

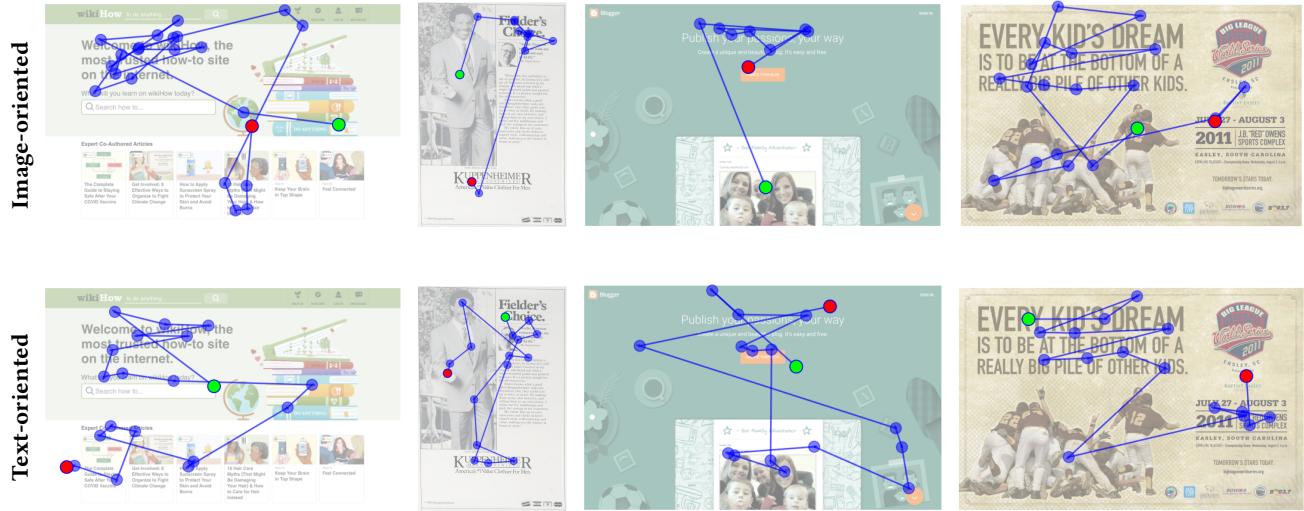


Figure 2: Different users have different viewing strategies on user interfaces. Image-oriented users often look at images before text, while text-oriented users prefer the opposite.

In our analysis, we discovered that text elements are more likely to capture users' attention than images. This observation explains why saccades tend to move from left to right, rather than the opposite direction. It is worth noting, however, that this result could be due to our dataset, which primarily consisted of English-language interfaces that require left-to-right reading. Despite this, the ratio of images to text does not appear to influence the distribution of visited and revisited elements within these two categories. Additionally, we found that saccades towards the right-hand side of the interface tend to cover greater distances between consecutive points than those towards the bottom. This finding provides evidence that user interfaces are not viewed in the same way as natural scenes [14]. Instead of a center bias, we observed a top-left bias. Our data allow a deeper dive into various subtle differences among the types of UIs examined. Several distinctions exemplify this:

Webpage: When looking at webpages, participants showed a preference for scanning from left to right, resulting in larger distances between consecutive fixations compared to other types of user interfaces.

Desktop UI: Unlike in other types of user interfaces, fixations on desktop UIs are not evenly spread over the top-left quadrant. Instead, salient areas are divided into two regions: one right above the center and the other around the top-left corner.

Mobile UI: Compared to other types of user interfaces, mobile UIs have lower visit and revisit ratios. This suggests that users tend to concentrate more on a few elements of the UI that attract their attention while ignoring others. Moreover, there is less likelihood of revisiting the same elements.

Poster: Compared to desktop and mobile UIs, participants exhibited a much stronger tendency to scan from left to right, with only a small proportion of saccades directed from top

to bottom. Additionally, the distances between consecutive fixation points showed more significant variation in this type of user interface than in others.

5 DISCUSSION AND FUTURE WORK

Our study sheds new light on the eye-movement behavior that occurs with specific UI types. Here, we discuss potential future directions.

5.1 Predictive Models

The dataset can inform the assessment and improvement of computational models for visual saliency. Given a UI as input, a saliency model predicts saliency maps or scanpaths, which simulate how users perceive that UI. These models assist UI designers by predicting where users are likely to fix their gaze within a given design, enabling the designers to update it to emphasize the important areas better. Additionally, such models may help designers 'reflow' UI designs and create versions that maintain the desired visual emphasis across various screen sizes. In our recent work, we presented some improvements on predictive models [9]. However, there is still ample room for improvement particularly in the accuracy of scanpath prediction.

5.2 Individual Viewing Strategies

We asked participants to self-report their viewing strategies. We found that some participants preferred to start by looking at images, while others focused on text or titles first. Participants also described different strategies for getting an overall idea of the UI. Some preferred to check the whole picture first, while others relied on titles to grab the general ideas and used images to understand the content better. Some participants indicated that they focused on the center of the page first, particularly when the layout was

center-aligned, before scanning from top to bottom and left to right. Figure 2 show the different individual viewing strategies visually in our dataset. These findings provide insights into the various viewing strategies that individuals use when looking at UIs and can inform the development of more personalized predictive models.

To accurately predict gaze behavior, it is essential to consider the varying viewing strategies that individuals use when looking at user interfaces. Future research should focus on understanding and modeling these differences across UI types and at the individual level. Personalized predictive models that account for these differences can help improve the accuracy of gaze behavior predictions.

5.3 Individual Element Influence

further investigations can be conducted to explore the potential impact of images on the viewing path of a user interface (UI). For instance, if we replace the food image in the recipe app with a different image, such as a human face, would users still follow the same viewing path? It is worth noting that images have inherent visual saliency that is distinct from the UI itself, and replacing them may significantly alter user perception and viewing trajectory. To gain a deeper understanding of the effects of individual UI components, future research could aim to disentangle the effects of various elements, such as photos, text, and buttons, rather than considering the entire UI as an indivisible entity.

ACKNOWLEDGMENTS

We appreciate the active discussion with Yao (Marc) Wang. This work was supported by Aalto University's Department of Information and Communications Engineering, the Finnish Center for Artificial Intelligence (FCAI), the Academy of Finland through the projects Human Automata (grant 328813) and BAD (grant 318559), the Horizon 2020 FET program of the European Union (grant CHISTERA-20-BCI-001), and the European Innovation Council Pathfinder program (SYMBIOTIK project, grant 101071147).

REFERENCES

- [1] Roman Bednarik and Markku Tukiainen. 2007. Validating the restricted focus viewer: A study using eye-movement tracking. *Behavior research methods* 39, 2 (2007).
- [2] Ali Borji and Laurent Itti. 2015. CAT2000: A Large Scale Fixation Dataset for Boosting Saliency Research. *CVPR 2015 workshop on "Future of Datasets"* (2015). arXiv preprint arXiv:1505.03581.
- [3] Michelle A Borkin, Zoya Bylinskii, Nam Wook Kim, Constance May Bainbridge, Chelsea S Yeh, Daniel Borkin, Hanspeter Pfister, and Aude Oliva. 2015. Beyond memorability: Visualization recognition and recall. *IEEE transactions on visualization and computer graphics* 22, 1 (2015).
- [4] Desktop UI Dataset. 2020. . <https://github.com/walteri/desktop-ui-dataset>
- [5] Biplob Deka, Zifeng Huang, Chad Franzen, Joshua Hirschman, Daniel Afergan, Yang Li, Jeffrey Nichols, and Ranjith Kumar. 2017. Rico: A Mobile App Dataset for Building Data-Driven Design Applications. In *Proceedings of the 30th Annual Symposium on User Interface Software and Technology (UIST '17)*.
- [6] Camilo Fosco, Vincent Casser, Amish Kumar Bedi, Peter O'Donovan, Aaron Hertzmann, and Zoya Bylinskii. 2020. Predicting visual importance across graphic design types. In *Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology*. 249–260.
- [7] Kruti Goyal, Kartikey Agarwal, and Rishi Kumar. 2017. Face detection and tracking: Using OpenCV. In *2017 International conference of Electronics, Communication and Aerospace Technology (ICECA)*, Vol. 1. IEEE, 474–478.
- [8] 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. <https://doi.org/10.1109/CVPR.2015.7298710>
- [9] Yue Jiang, Luis A. Leiva, Paul R. B. Houssel, Hamed R. Tavakoli, Julia Kyrmälä, and Antti Oulasvirta. 2023. Ueyes: Understanding Visual Saliency across User Interface Types. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) (*CHI '23*). Association for Computing Machinery, New York, NY, USA. <https://doi.org/10.1145/3544548.3581096>
- [10] Tilke Judd, Frédéric Durand, and Antonio Torralba. 2012. A Benchmark of Computational Models of Saliency to Predict Human Fixations. In *MIT Technical Report*.
- [11] Tilke Judd, Krista Ehinger, Frédéric Durand, and Antonio Torralba. 2009. Learning to predict where humans look. In *2009 IEEE 12th international conference on computer vision*. IEEE, 2106–2113.
- [12] Nam Wook Kim, Zoya Bylinskii, Michelle A Borkin, Krzysztof Z Gajos, Aude Oliva, Frédéric Durand, and Hanspeter Pfister. 2017. BubbleView: an interface for crowdsourcing image importance maps and tracking visual attention. *ACM Transactions on Computer-Human Interaction (TOCHI)* 24, 5 (2017). <https://doi.org/10.1145/3131275>
- [13] Nam Wook Kim, Zoya Bylinskii, Michelle A Borkin, Aude Oliva, Krzysztof Z Gajos, and Hanspeter Pfister. 2015. A crowdsourced alternative to eye-tracking for visualization understanding. In *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems*. 1349–1354.
- [14] Luis A Leiva, Yunfei Xue, Avya Bansal, Hamed R Tavakoli, Tuğçe Körödü, Jingzhou Du, Niraj R Dayama, and Antti Oulasvirta. 2020. Understanding visual saliency in mobile user interfaces. In *22nd International conference on human-computer interaction with mobile devices and services*. 1–12.
- [15] Aliaksei Miniukovich and Maurizio Marchese. 2020. Relationship between visual complexity and aesthetics of webpages. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–13.
- [16] Subramanian Ramanathan, Harish Katti, Nicu Sebe, Mohan Kankanhalli, and Tat-Seng Chua. 2010. An eye fixation database for saliency detection in images. In *European conference on computer vision*. Springer, 30–43.
- [17] Ruth Rosenholtz, Amal Dorai, and Rosalind Freeman. 2011. Do Predictions of Visual Perception Aid Design? *ACM Trans. Appl. Percept.* 8, 2 (2011).
- [18] Jeremiah D. Still and Christopher M. Mascioccchi. 2010. A Saliency Model Predicts Fixations in Web Interfaces. In *Proc. MDDAU Workshop*.
- [19] Alexa Top 500 Websites. 2022. <https://www.expireddomains.net/alexa-top-websites/>.
- [20] Mulong Xie, Sidong Feng, Zhenchang Xing, Jieshan Chen, and Chunyang Chen. 2020. UIED: A Hybrid Tool for GUI Element Detection. In *Proceedings of the 28th ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering (Virtual Event, USA) (ESEC/FSE 2020)*. Association for Computing Machinery, New York, NY, USA, 1655–1659. <https://doi.org/10.1145/3368089.3417940>
- [21] Pingmei Xu, Krista A Ehinger, Yinda Zhang, Adam Finkelstein, Sanjeev R Kulkarni, and Jianxiong Xiao. 2015. Turkergaze: Crowdsourcing saliency with webcam based eye tracking. *arXiv preprint arXiv:1504.06755* (2015).