Topic: Are click explorations a valid alternative to eye tracking?

Human attention is a crucial cognitive process in various aspects of our lives. In the field of human-computer interaction (HCI), for many applications, such as eye tracking/gaze detection, it is essential to understand where humans focus their attention in a scene [1]. Eye tracking involves capturing and analyzing eye movement data to understand where an individual looks at the scene. By analysing these eye movement data, one can detect the gaze position in the surrounding environment. However, eye tracking is not always an option. Eye trackers are expensive and do not scale for large data collection. Thus, as an alternative, many studies [2][3][4][5] have shown to employ a computer mouse to indicate eye movements. In [6][7], the relationship between the mouse-click and eye movements has also been explored while navigating web pages. The click-contingent paradigm encourages mouse movements to reveal intriguing objects in the peripheral vision, likewise how humans shift their gazes to bring objects of interest into the fovea. However, the question arises: are these click explorations a valid alternative to eye tracking?

In this master thesis, our primary goal is to validate if click is a valid alternative and how it correlates with actual gaze. To study this, our focus will be on task-driven visual attention behavior. People's initial focus is mostly guided by visual characteristics that capture their attention under task-free situations, and this tends to be consistent among various individuals. Nonetheless, their focus quickly shifts in different directions if their attention gets influenced by some tasks [2]. An example of such tasks is image captioning, i.e. describing the content of an image scene [3]. In [8], study indicates that the way human attention behaves, varies between free-viewing and image description tasks, where individuals have to provide a verbal description of the scene content while their eye movement data is being recorded. Thus, investigating task-driven human attention behavior, and its interplay with language, holds potential research interest.

In this work, we collect data from different users while engaged in providing textual descriptions of different images. We will conduct real-time experiments in two sessions, and each session will contain a series of 20 distinct images, selected from the public MIT1003 [1] dataset. The images will be categorized into animate and inanimate groups, each featuring both simple and complex designs. Users will be shown each set of images, asking for the text captions of each image through the web application introduced in [2]. The images, displayed in two sessions, will be presented in pseudo-randomized order using a free gaze exploration and click-contingent approach respectively. To record real-time gaze data and avoid a conventional restricted eye tracking environment setup found in most research [1][5][8], here we use a flexible deep learning-powered wearable eye-tracker called Neon [9]. Throughout both sessions, we will record gaze and fixation data and during the click-contingent experiment, we will simultaneously capture click data and eye movement data to understand the exploration strategy of users while they are using clicks.

In terms of extracting the eye movement data from Neon, we will utilize the marker mapper enrichment [9] feature, which enables tracking of where an individual will be looking in a particular area by positioning markers in a flexible surrounding environment. This approach will also allow us to actively monitor real-time gaze data within a flexible environment, enabling a thorough comparison with the click data. Additionally, to evaluate the gaze correlation between free-viewing and captioning within an unconstrained eye tracking environment, we aim to compare the new gaze data acquired during captioning with the existing eye-tracking data collected under free-viewing scenarios in the public MIT1003 [1] dataset.

The proposed work consists of the following parts:

- Literature research of relevant work for click-contingent and free gaze human attention explorations.
- Configure the web application in preparation for the experimental procedures.
- Conducting real-time experiments with Neon eye tracker and introducing a dataset with synchronously recorded gaze data, click data, and the scene description provided by the subjects.

- Analysis of gaze and clicks during the click-based exploration, to understand the visual strategy of subjects while they are using clicks. We compute NSS between gaze fixations and click-based saliency maps.
- Analysing the correlation between gaze data and click data using saliency metric (NSS, AUC) to compare the spatial distribution of fixation points against a random baseline, and scanpath metrics (String-edit distance) to know if with gaze and clicks subjects use the same exploration strategy.
- Comparing the scanpaths between free-viewing from [1] and the scanpaths collected in our captioning experiment using the same metrics mentioned above.

The research project must contain a detailed description of all developed and used algorithms, as well as a profound result evaluation and discussion. The implemented code has to be documented and provided. An extended research on literature, existing patents and related work in the corresponding areas has to be performed.

Advisors:	Dr. Dario Zanca
	Naga Venkata Sai Jitin Jami, M. Sc.
	Prof. Dr. Bjoern Eskofier
Student:	Moumita Chakraborty
Start – End:	1st February 2024 - 1st August 2024

References

[1] Judd, T., Ehinger, K., Durand, F., Torralba, A.: *Learning to predict where humans look*. 2009. IEEE 12th international conference on computer vision, 2106–2113. IEEE

[2] Zanca, D., Zugarini, A., Dietz, S., Altstidl, T. R., Ndjeuha, M. A. T., Schwinn, L., Eskofier, B.: *Contrastive Language-Image Pretrained Models are Zero-Shot Human Scanpath Predictors*. 2023. arXiv preprint arXiv:2305.12380

[3] Jiang, M., Huang, S., Duan, J., Zhao, Q.: *Salicon: Saliency in context*. 2015. Proceedings of the IEEE conference on computer vision and pattern recognition, 1072–1080

[4] Egner, S., Reimann, S., Hoeger, R., Zangemeister, W. H. *Attention and information acquisition: Comparison of mouse-click with eye-movement attention tracking*. 2018. Journal of Eye Movement Research, 11(6). European Group for Eye Movement Research

[5] Kim, N. W., Bylinskii, Z., Borkin, M. A., Oliva, A., Gajos, K. Z., Pfister, H.: *A crowdsourced alternative to eye-tracking for visualization understanding*. 2015. Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems, 1349–1354

[6] Chen, M. and Anderson, J. R., Sohn, M.: *What can a mouse cursor tell us more? Correlation of eye/mouse movements on web browsing*. 2001. CHI'01 extended abstracts on Human factors in computing systems, 281–282

[7] Rodden, K., Fu, X.: *Exploring how mouse movements relate to eye movements on web search results pages*. 2007

[8] He, S., Tavakoli, H. R., Borji, A., Pugeault, N.: *Human attention in image captioning: Dataset and analysis*. 2019. Proceedings of the IEEE/CVF International Conference on Computer Vision, 8529–8538

[9] *Neon-Home-Pupil Labs* [online] https://docs.pupil-labs.com/neon/