Utrecht University 2020-2021
Applied Cognitive Psychology

Master Thesis, 27.5 ECT

23 Jul 2021

Identifying scanpath starting point in structured web images at group level

Comparing mouse and eye tracking with saliency map

Ava Tse
o.k.tse@students.uu.nl
2022060

Alpha.One Supervisors	Coert van Gemeren *coert@alpha.one* Ingrid Nieuwenhuis *ingrid@alpha.one*
UU Supervisor	Samson Chota *s.chota@uu.nl*
UU Second Assessor	Stefan van der Stigchel *s.vanderstigchel@uu.nl*

Universiteit Utrecht

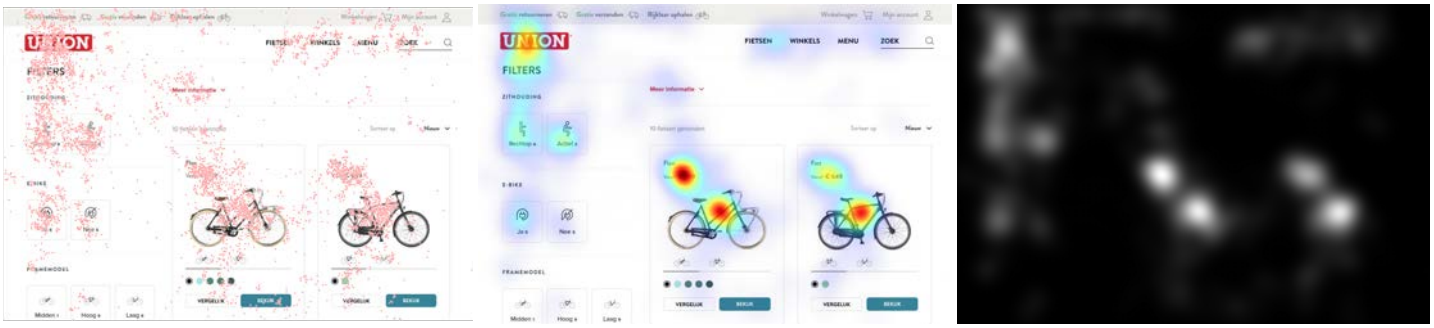
Alpha.One

Abstract

Understanding where and when people look on webpages is essential to web creators. However, collecting gaze data with traditional eye tracking (ET) is expensive and time-consuming. Alpha.One, a neural marketing company, aims to predict the gaze sequence of viewing on webpages, using deep learning and generative adversarial neural networks (GANs). The models are trained on salience data which is aggregated from mouse tracking (MT) experiments on Amazon's Mechanical Turk. The experiments are conducted via a psychophysical paradigm known as the mouse-contingent multi-resolutional paradigm (Jiang et al., 2015). The hypothesis of this study is that the shifts of viewing order are initiated toward the salient intensity level (Henderson, 2003; Itti, 2005; Tseng & Howes, 2008; Underwood, 2009). This research presents a novel approach to (a) determine the starting point of where users are most likely to look at first on a webpage and (b) produce a general scanpath. The ET heat maps are compared to the starting point in general viewing order generated from ET and MT data. The results show the starting point usually is not in the most salient area of the ET heat maps, and the hypothesis that the first element to be looked at is in the most salient area is disproved. This indicates that the viewing order cannot be simply deduced from the salient intensity levels of the heat map.

Keywords: Web viewing, eye tracking, mouse tracking, saliency model, scanpath analysis, heat map analysis

Contents



(a) (b) (c)

Figure 1 From left to right, (a) Original raw gaze points (b) Colored heat map (Red zones represent higher density designate where the viewers focused their gaze with a higher frequency) and (c) Kernel density estimation heat map (similar to colored heat map, the brighter the region, the higher the gaze frequency in that area)

1. Introduction

1.1. Eye tracking and webpage design

People make snap judgments. It takes less than a second to get the first impression of a person. Webpages are no different. Users establish an opinion about a webpage in about 50 milliseconds (ms) (Lindgaard et al., 2006), and 94% of first impressions are design-related (Sillence et al., 2004). Understanding how users allocate their gaze while viewing webpages is important for web creators. To support web designers and researchers in identifying which visual elements draw attention on webpages, eye tracking products typically provide heat maps (*Tobii Pro Lab User Manual*, 2021). Figure 1 illustrates an example of gaze points visualization on a webpage. A heat map (also referred to as density maps) displays the spatial distribution of gaze data on a stimulus, which can be aggregated over time for one or multiple participants.

1.2. Saliency model to predict human gaze pattern

Visual saliency refers to the perceptual quality that makes an object or location stand out from its surroundings, and thus attracts our attention. Visual attention can be driven by either stimulus-driven (bottom-up) or goal-oriented (top-down) factors (Pinto et al., 2013). In stimulus-driven attention, visual saliency is purely driven by visual data itself. In top-down factors, high-level information like the goal and preferences of the viewers can modulate and guide the deployment of attention. It is much harder to predict top-down attention, because this depends on the viewer's goal. Therefore, predicting top-down attention is out of scope of the current work.

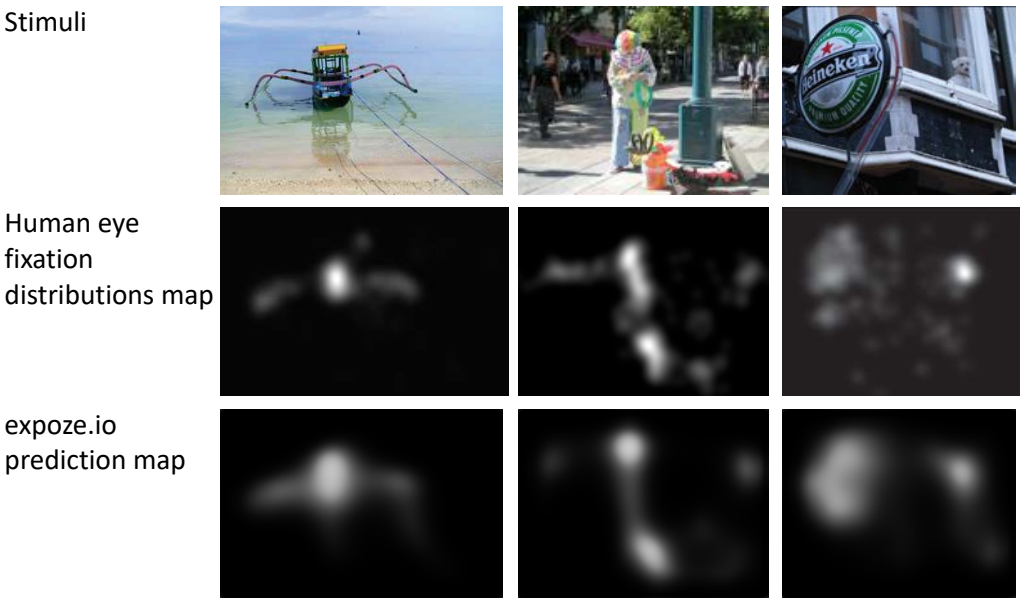


Figure 2 Some examples of human eye fixation maps and expoze.io prediction maps with natural images ("MIT Saliency Benchmark," 2012)

However, collecting gaze data with traditional eye tracking is expensive and time-consuming, making it difficult for web designers to benefit from the insight of eye movements. Using computer vision and machine learning to predict visual attention has been an area of active research in recent years. Alpha.One, a neural marketing company, successfully developed the application platform expoze.io[1], which utilizes Generative Adversarial Networks (GANs) to predict where people look at natural images in a free-viewing condition reported accuracy compared to actual human performance (Expoze.io, 2021), some examples shown in Figure 2.

1.3. The evolution of salience modeling

Classic computer vision model

Judd et al. (2009) developed a saliency model classifier, along with semantic-level features, and multiplying center bias. Preattentive elements including intensity, orientation and color contrast are classified as "low-level features". "High-level features" include faces, animals, text, objects, and social interaction. The work of Tseng et al. (2009) shows people gaze fixations are biased toward the center of the natural scene stimuli. The model of Judd introduced the center prior of which indicates the distance to the center for each pixel.

Deep learning salience predictions

The success of convolutional neural networks on large scale object recognition has brought along a new wave of saliency models that perform markedly better than traditional saliency models based on handcrafted features (Borji, 2021). expoze.io is based on the saliency model SalGAN (Pan et al., 2017) which utilizes Generative Adversarial Networks (GAN). The GAN is called "generative" because it generates new data with the same statistics as the training set, the framework of GAN as shown in Figure 3. The generator produces new salience data that is derived from the learned probability distribution. The discriminator acts like a judge. It decides if its input comes from the generator or from the ground truth training set. The generator is trying to maximize the probability of making the discriminator mistake its input for the ground truth data. The discriminator forces the generator to produce more realistic images. The GAN is a zero-sum game between generator and discriminator, and the optimization goal is to reach Nash equilibrium (Ratliff et al., 2013), where the generator would capture the general training data distribution from human saliency maps, and the discriminator would always be unsure whether the map is an artificial map or from human training data.

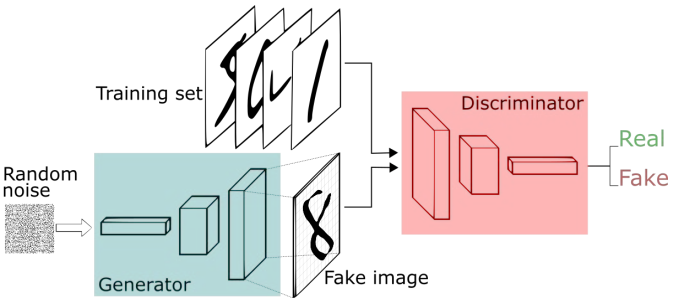
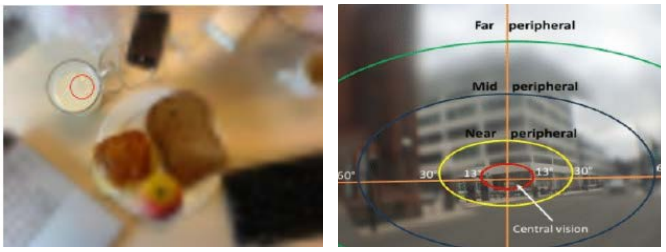


Figure 3 Generative Adversarial Networks framework (*An Intuitive Introduction to Generative Adversarial Networks (GANs)*, 2018)

1.4. Utilizing large scale mouse data on saliency model training

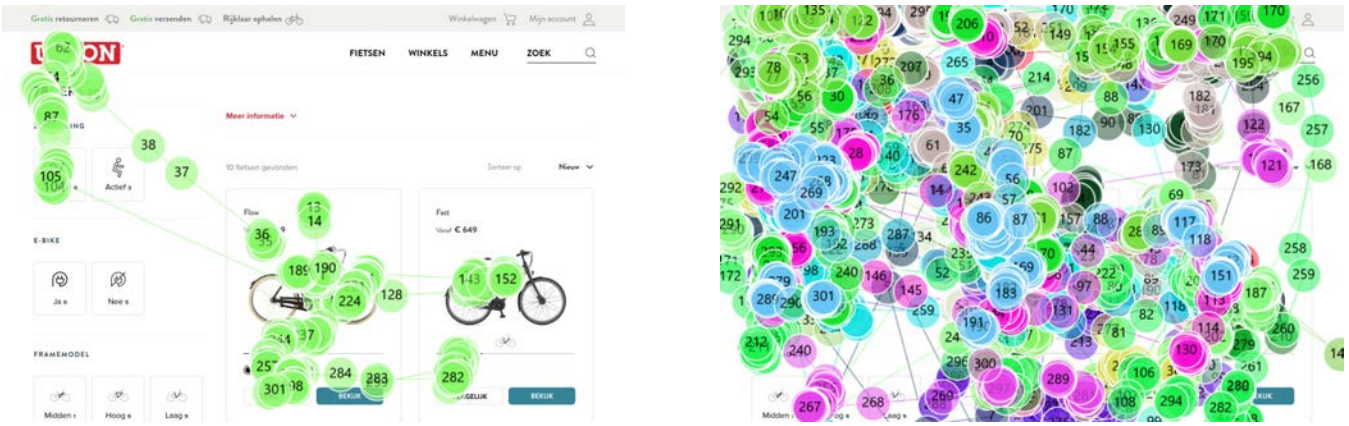
As deep learning saliency models rely heavily on large scale data, crowd-sourcing mouse tracking data is a potential alternative to lab-based eye trackers. Jiang et al. (2015) designed a *mouse-contingent multi-resolutional paradigm* relying on neurophysiological and psychophysical studies of peripheral vision. Blurring the image outside the center of the mouse aimed at simulating natural viewing conditions of humans (shown as Figure 4). It is assumed that the viewer points the mouse intuitively to the elements that, despite the blur, stand out most from the background, reflecting the saliency of the various image elements to the viewer.



(a) (b)

Figure 4 Jiang et al. (2015) develop (a) mouse tracking paradigm with a 'moving window' to mimic (b) human field of view (Younis et al., 2019)

[1] Visual attention prediction platform developed by Alpha.One https://www.expoze.io/



(a) (b)

Figure 5 Scanpath(s) from (a) one and (b) group level of 25 participants

1.5. Going beyond a heat map to a starting point and scanpath

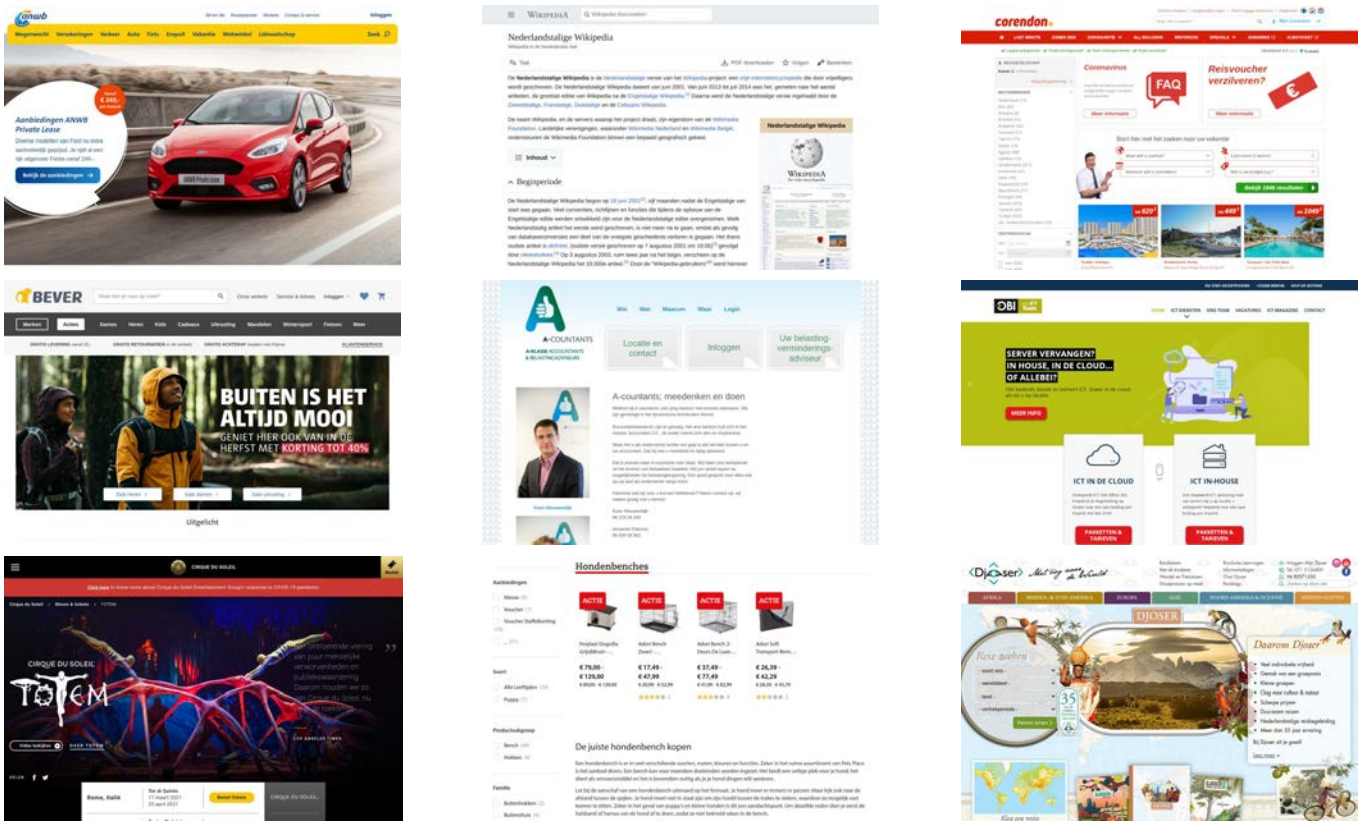
The high-level and long-term goal of Alpha.One is to train a neural network to predict the scanpath on marketing material and webpages. A crucial goal of any visual design is to communicate the relative importance of different design elements, so that the viewer knows where to focus attention and how to interpret the design. Although heat maps can provide a valuable overview of important elements of interest on a stimulus, they are not designed to illustrate the scanpath, which reveals the time sequence of viewing. However, studies on multiple users' scanpath are problematic, as the scanpath may differ a lot between individuals. Scanpaths overlap and the visualization becomes cluttered and hard to interpret, as shown in Figure 5.

Some neural-computational saliency models (Henderson, 2003; Itti, 2005; Tseng & Howes, 2008; Underwood, 2009) suggest the shifts of attention and subsequent saccadic eye movements are sequentially executed towards locations with ascending salience intensity levels. The first fixation should be directed towards the most salient location in the visual field, the second fixation to the second most salient location, etc. However, those models focused on less complex natural scenes, and the validity of rich visual element contents like webpages is still unknown.

1.6. Problem statement and outline of study

"Starting point" is the first visual element that users are most likely to look at first on a webpage (Drusch et al., 2014), which is also considered the first point in the scanpath. Understanding the common viewing order, and especially the starting point, is particularly important to web designers. As the first information we receive is usually anchored with the primacy effect that people tend to recall items at the beginning of a sequence more than other items (Foulsham & Kingstone, 2013; Interaction Design Foundation, 2020; Wiswede et al., 2007). The underlying motivation for this project emerged from the following questions: How can we define the starting point? Is the starting point also in the most salient area on the heat map? To answer these questions, this study has the following three objectives:

(a) To quantify how similar the heat maps of eye tracking, mouse tracking and expoze.io are.
(b) To define a method to find the starting point in group level data and compare the starting point from both mouse tracking and eye tracking data.
(c) To define a method to find a scanpath in group level data from both mouse tracking and eye tracking data.



(a) Pictorial (b) Text (c) Mixed

Figure 6 Examples of webpages in dataset

2. Webpage dataset collection

There is no publicly available eye and mouse tracking dataset on real webpages. Fortunately, a dataset of webpage stimuli was collected at Alpha.one in the past. The dataset was created by collecting mouse tracking data from 27 participants and eye tracking data from 25 participants on 51 webpage stimuli during a free-viewing task.

2.1. Webpage stimuli

51 screenshots of Dutch webpages were created. The screenshots were made by grabbing the viewport of a web browser above the fold without any further scrolling or clicking. The images were collected from various sources on the internet at a resolution of 1080 x 720 pixels (px). These webpages were categorized as pictorial, text, and mixed according to the composition of text and pictures (Example of webpages shown in Figure 6).

2.2. Eye tracking data collection

Illustration of computer screen for ET experiment

Participants	25
Device	Tobii Pro Nano, Monitor with resolution of 1920 x 1020 px
Sampling rate	60 Hz
Task	Free-viewing
Viewing Time	5s
Setup and procedure	Participants were seated in front of the computer screen at a distance of 65-75 cm. Calibration was done using the 9-point grid method. For each stimulus, an image was presented in random order. Participants were informed to free-view the webpage for 5s. Participants were asked to fixate their eyes at the center fixation cross for 2s before each trail.

2.3. Mouse tracking data collection

Although the work of Jiang et al. (2015) demonstrates a high level of success of eye-mouse tracking saliency map with nearly 90% of similarity utilizing the mouse-contingent multi-resolutional paradigm, the paper does not give implementation details of the blurring mask they have used. Therefore, Alpha.One implemented its own mouse-contingent paradigm, similar to Jaing et al. (2015), to be used on Amazon Mechanical Turk (AMT) with the following parameters:

1. The whole image of the webpage was blurred in real time using Gaussian Blurring with SD=7. Figure 7 helps to understand the effect of SD value to blurring.
2. The location of the mouse representing the fovea was unblurred.
3. The unblurred view of the image had a diameter of 110 pixels (~5◦ visual angle) and a soft edge with a SD=7.
4. The appearance of the high-resolution area to the viewer was delayed by 200 ms with consideration of saccadic latency (Vencato & Madelain, 2020).

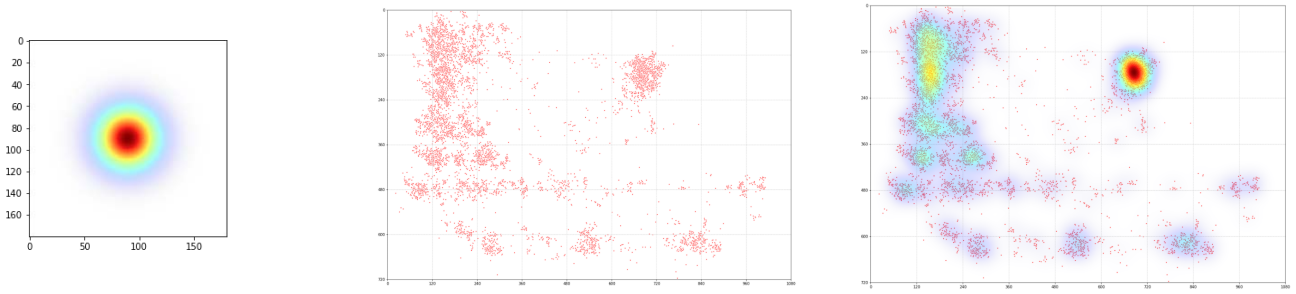
We discuss the potential adjustment of parameters in Section 6.1. Areas for enhancements.

Figure 7 The effect of different standard deviation value of Gaussian blurring

Illustration of computer screen for MT experiment	
Participants	27
Device	Mouse & desktop or laptop computer
Sampling rate	60 Hz
Task	Free-viewing
Viewing Time	10s, considered cursor moves slower than eye (Jiang et al., 2015)
Setup and procedure	Participants proceeded the experiment on the AMT online platform with their own devices. A training round with three trails of blurred words in different locations allowed users to familiarize with the setup. For each stimulus, an image was presented in a random order. Participants were informed to free-view the webpage for 10s. Participants must had to move the cursor onto the fixation cross before each trail.

2.4. Data analysis and manipulation

Although Tobii Pro Lab software and some open source toolbox like PyGaze (Dalmaijer et al., 2014) provide a high-level analysis and plotting program for eye tracking data, there is no integrated platform for eye-mouse tracking data. Therefore, we developed our own code at Alpha.One based on Python for data analysis and visualization in this project. Due to the difference in hardware and software setting, the mouse tracking and eye tracking data have different sampling rates. Eye tracker collects data at a sampling rate of 60 Hz, while mouse data sampling is only triggered with cursor movement (no data is collected if no mouse event) as restricted by the AMT environment. The mouse tracking data is resampled in 60 Hz, which matches the data in position.



(a) Colored Gaussian Kernel Filter (b)Raw position data dots (c) Data applied with Gaussian Kernel Filter

Figure 8 Heat map visualizations

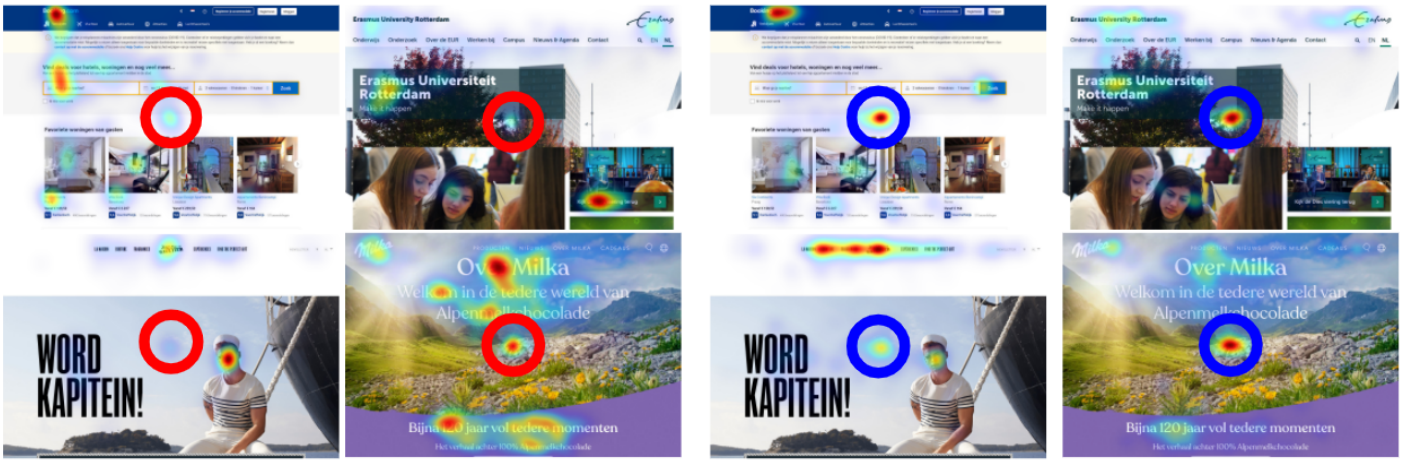
2.4.1. Heat map generation

The heat map presented in this project is raw position data and does not classify the data into fixations, saccades or other eye or mouse movements (*Tobii Pro Lab User Manual*, 2021) to avoid uncertainties influenced by the implemented fixation detection algorithm (Hessels et al., 2018). There is no concrete study or systematic research done on investigating the correlation between time spent and fixation duration on mouse tracking and eye tracking behavior.

The heat map is generated by convolving a 2D Gaussian kernel filter (*Generate a Heatmap in MatPlotLib Using a Scatter Data Set*, 2017) on position points gathered in the dataset. In this work, the total kernel is fixed at 180 px with a standard deviation of 15 px and is used to smooth the point and generate a map (example shown as Figure 8).

2.4.2. Reducing the effect of position bias

A strong position bias is found within the collected dataset. Examples in Figure 9 illustrated unusual viewing patterns around the center area of stimuli, even though no visual design elements exist there. It is believed the phenomena is mainly due to the center fixation cross displayed at the beginning of the trial. Position bias can be problematic when comparing webpage salience if content is more concentrated in the center of the scene. Thus, we attempt to reduce the effect of position bias with post-hoc analysis.



(a) Eye tracking heat map (b) Mouse tracking heat map

Figure 9 Position bias, especially in the center area of stimuli, is observed with the collected dataset

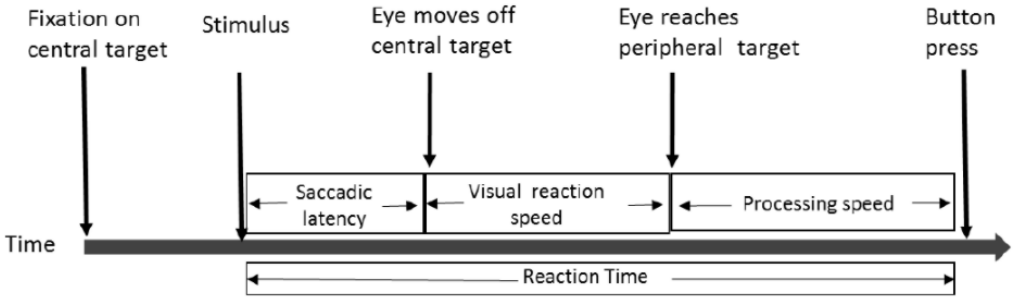


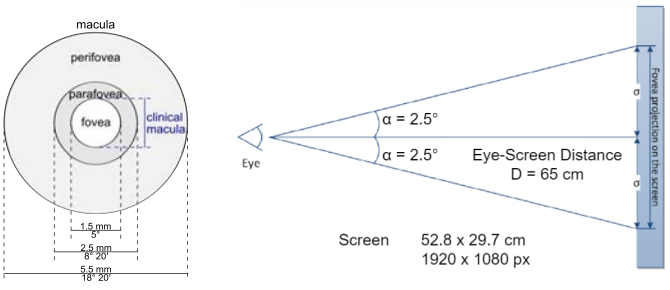
Figure 10 Breakdown of events and time intervals related to the measurement of reaction time (Lange et al., 2018).

Visually guided pointing movements from one visual target to another are usually preceded by corresponding saccadic eye movements, which are preceded by several brain processes preparing the two movements. (Lange et al., 2018) (2018) adopted the visual and motor response times measurement from Magill & Anderson (2010) and suggested a breakdown of events with eye-hand coordination considering the present of fixation cross (shown as Figure 10).

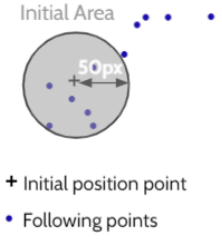
Although there are many studies on eye-hand coordination behaviors in computer tasks, few work focus on the reaction time during free-viewing on webpages, and none of them are in the blurred MT setting. Thus, we cannot directly implement the findings from other studies. Upon this, a spatial and temporal analysis is conducted in the dataset by identifying the initial area around the first location point and reviewing the time spent in the initial area.

Identifying the initial area

(a) Fovea projection on the screen



(b) Initial area



+ Initial position point

• Following points

Figure 11 Define an initial area for analysis saccadic latency time of eye moves off the central target.

Figure 10 illustrates the initial area with the first position points and the surrounding region in a radius of 50 px. This size approximates the size of the human's eye foveal (5°) projection on the screen. Under the consideration of participants' viewing distance away from the screen (~65 cm), screen size (52.8 x 29.7 cm) and screen resolution (1920 x 1080 px), as mentioned Section 2.2 Eye tracking data collection.

Reviewing the time spent in the initial area

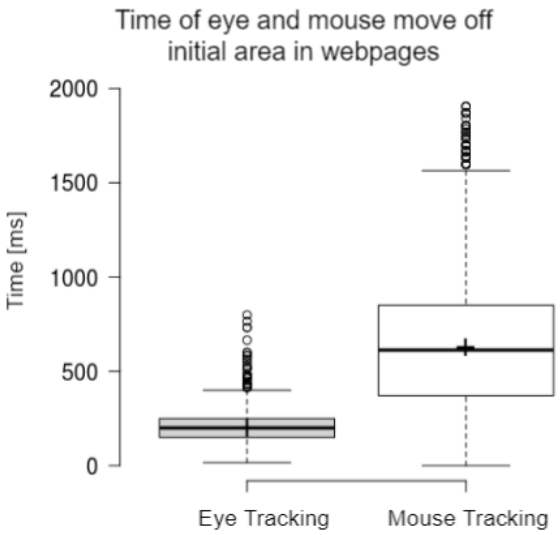


Figure 12 Comparison of eye-mouse tracking time spent inside the initial area

Figure 12 shows the time spent inside the initial area of all participants on 51 stimuli. The mean time in eye tracking data is 201 ms (SD = 101.6) and 629 ms (SD=421.9) in mouse tracking data. The finding converges with the eye behavior described by Salthouse & Ellis (1980) that the minimum pause time of the eye is estimated to be about 200ms without any stimulus processing. Huang et al. (2012) found that delay from gaze to cursor actions between 250-700 ms.

(a) ET - Original (b) ET - Positional bias filter (c) MT - Original (d) MT - Positional bias filter

Figure 13 The effect of Spatial & Temporal Filter on heat maps

Filtering the data

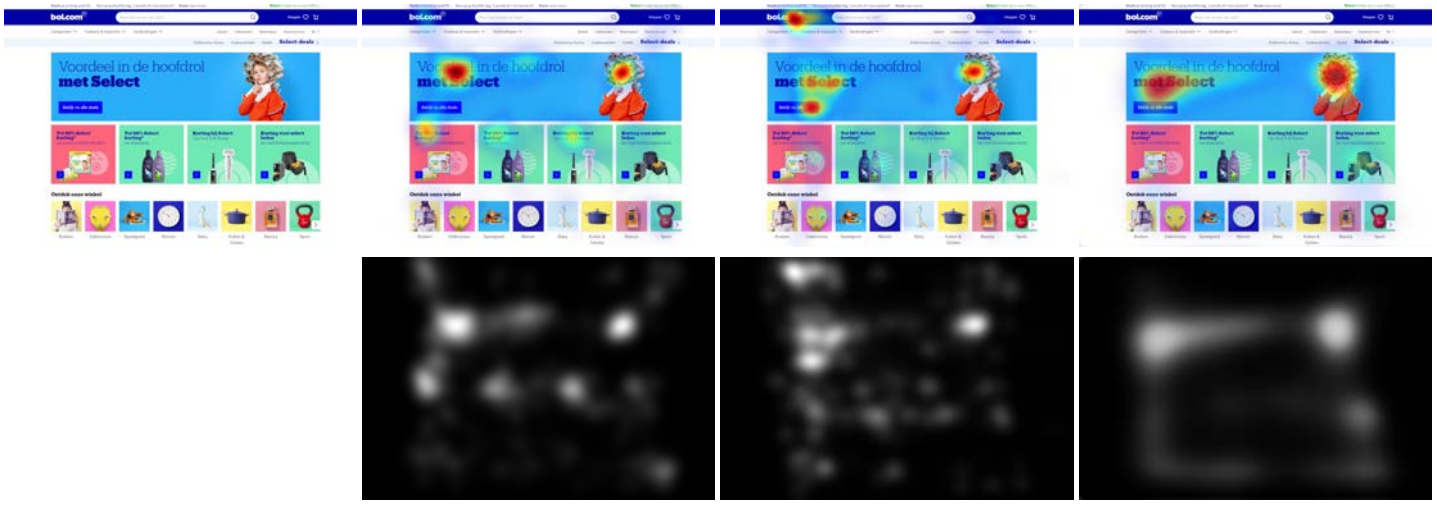
Spatial and temporal filtering are applied on the datasets to reduce the effect of position bias with the following criteria:

1. Remove the subject's data from the stimulus' dataset if a subject stays in the initial area longer than 2000 ms. This indicates the subject does not actively participate in that trail or data lost during the experiment.
2. For eye tracking data, only consider viewing data points outside the initial area for the first 250 ms (upper quartile of time spent in the initial area with eye tracking data as shown in Figure 12), with the consideration of saccadic latency time of eye moves off the central target.
3. For mouse tracking data, only consider viewing data points outside the initial area for the first 700 ms (upper quartile of time spent initial area with mouse tracking data as shown in Figure 12), with the consideration of delay from gaze to cursor actions.

Figure 13 demonstrates how the spatial-temporal filter reduces the effect of position bias on the heat map.

3. Objective 1: Comparing maps similarity

The work of Jiang et al. (2015) shows that lab eye tracking and AMT mouse tracking can generate heat maps with high similarity. They had successfully trained a computational model utilizing mouse maps on AMT on natural scenes (Bylinskii et al., 2015). This suggests mouse tracking can be useful for model training, aimed at predicting gaze patterns. How well can computational model expoze.io predict saliency in webpages? Is it possible to reach human level accuracy by AMT mouse data at current experimental settings? In order to help us understand the models, it is important to evaluate the similarity between eye tracking (ET) maps, AMT-mouse tracking (MT) maps and prediction maps. Figure 14 shows an example of produced maps.



(a) Original (b) Eye tracking map (c) AMT-mouse tracking map (d) expoze.io prediction map

Figure 14 Example of heat map from eye and mouse tracking experiment and computational model

3.1. Metrics to measure the agreement between two maps

There are several indices for evaluation metrics to measure the agreement between heat maps. In general, the matrices of saliency evaluation are divided into location-based and distribution-based metrics. Location-based metrics consider saliency map values at discrete fixation locations, while distribution-based metrics treat both ground truth fixation maps and saliency maps as continuous distributions. For easier interpretation of the result, we chose shuffled AUC (sAUC) in the first category and histogram intersection similarity (SIM) in the second category.

3.1.1. Location-based metric: Shuffled AUC (sAUC)

The most widely used method to evaluate and compare saliency models is Area under ROC Curve (AUC) as illustrated in Figure 15. The saliency map is treated as a binary classifier to separate positive from negative samples at various thresholds. The true positive (TP) rate is the proportion of the saliency map's values above the threshold of fixation locations. The false positive (FP) rate is the proportion of the saliency map's values that occur above the threshold sampled from random pixels sampled at a fixed step size. The shuffled AUC (sAUC) (Borji et al., 2013) compensate for center bias, which is common in saliency datasets, by sampling FPs from fixations from other images. sAUC reduces the center-bias to ensure an unbiased comparison of saliency models (Bylinskii et al., 2019). The range of sAUC is from 0 to 1. The higher the value, the more accurate the saliency model predicts human eye movements.

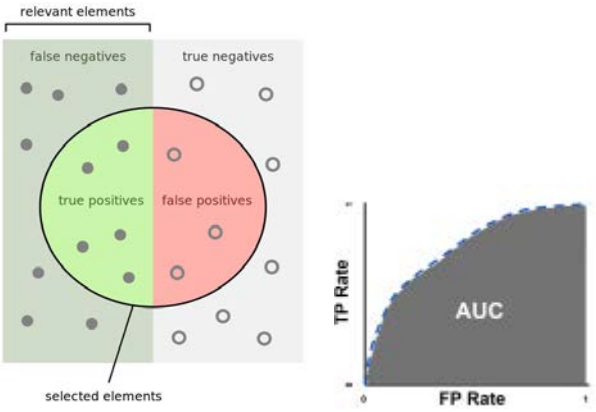


Figure 15 Area under curve

3.1.2. Distribution-based metric: Similarity (SIM)

This similarity metric is also called histogram intersection and measures the similarity of two discrete probability distributions (histograms). SIM is calculated as the sum of the minimum values of each pixel as:

$$SIM = \sum min\,(S\,i = 1\,(i), G\,(i))$$

Where $\mathbf{S}$ and $\mathbf{G}$ are the normalized saliency map and the fixation map, respectively. A similarity score between zero (no overlap) and one (the distributions are the same). Figure 16 shows an illustration of an example histogram intersection similarity method.

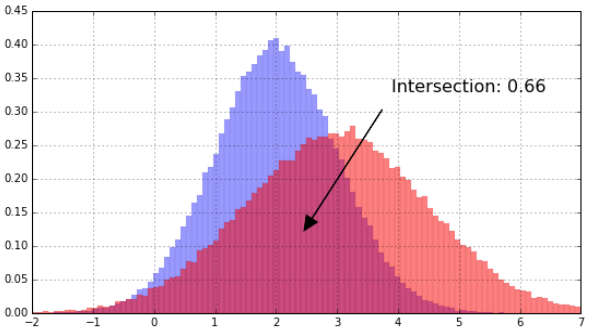


Figure 16 Histogram intersection similarity method

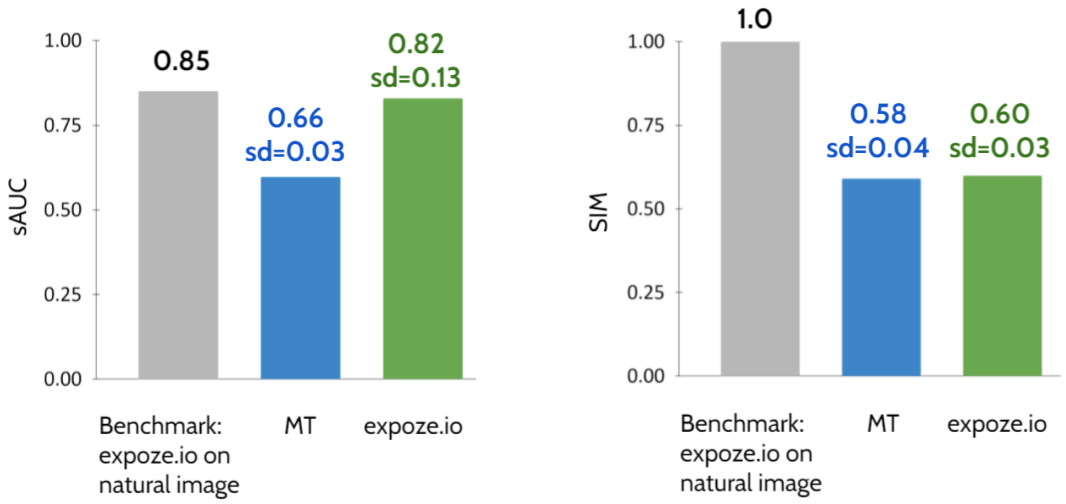


Figure 17 The performance of across eye tracking, mouse tracking and prediction model

3.2. Analysis on produced heat maps

Produced heat maps of 51 stimuli from ET, MT and expoze.io can be found in Appendix A.

3.2.1. The performance across eye tracking, mouse tracking and prediction model

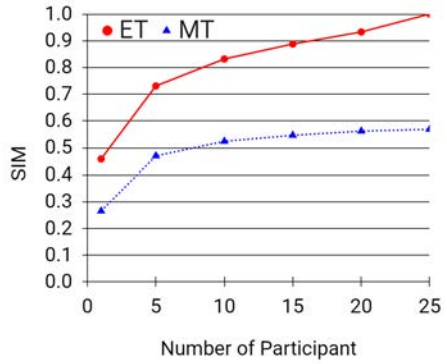
Figure 17 reports the result of SIM and sAUC score of MT and expoze.io prediction map in comparison to ET map, with the benchmark of natural image ("MIT Saliency Benchmark," 2012).

The results in Figure 17 shows that both MT and Expoze.io did not produce heat maps comparable to ET maps on webpage stimuli up to the standard on natural scenes. Surprisingly, although expoze.io is trained on less complex natural scenes, it is able to generate prediction maps with sAUC score 0.82 with acceptable level of overlapping to human eye results.

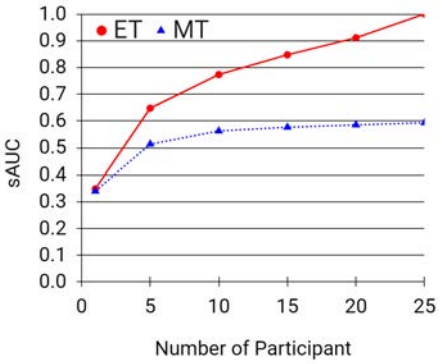
On the other hand, the performance of MT is even lower then computational network without specifically trained on webpage datasets. Figure 18 presents the images with high and low sAUC scores in MT. While the mouse-eye agreement is high in pictorial webpages, it is generally lower in more complex mixed layout with texts and pictures.

Figure 18 presents the image with high and low sAUC scores in MT.

Figure 18 Examples with high and low eye-mouse evaluated with sAUC. Colored heat maps are overlaid. expoze.io prediction map as reference to sAUC score on the same stimuli



(a) SIM score



(b) sAUC score

Figure 19 The number of ET and MT participants to achieve ET performance

3.2.2. The performance of varying numbers of participants

We suspected the number of participants is a potential cause of the performance gap between ET and MT. To more carefully examine whether increasing the number of MT participants can improve the maps agreement with ET, we measure how well different numbers of ET or MT participants can approximate the ground-truth ET maps. To evaluate the performance as a function of the number of participants, the participants are randomly chosen from the ET and MT dataset to produce a heat map, and the mean performance is reported. The result shows in Figure 19.

In Figure 19 (a) and (b), the slopes flattened after 20 participants (slope > 0.002). This indicates increasing the number of MT participants with the current setting is less likely to reduce the performance gap to ET. Even 25 participants contributing mouse data can not achieve the ET performance of 5 observers.

3.2.3. The most salient location comparison

To understand the gap between MT map, expoze.io and ET map, the most salient location on each map is evaluated. The visual comparisons are affected by the range and scale of saliency values, and are driven by the most salient locations, while small values are not as perceptible and don't enter into the visual calculations (Bylinskii et al., 2019). The brightness intensity of a pixel on a saliency map indicating the saliency. The saliency value is normalized from 0 (pure black) to 255 (pure white). The most salient location is deduced from the maximum value on the saliency map. Table 2 shows whether the most salient location on MT and expoze.io prediction map agree with ET map.

	The most salient location agree with ET map (stimuli = 51)
Human MT	27.41%
expoze.io	41.18%

Table 2 The most salient location comparison of ET, MT, expoze.io prediction maps

The result shows the agreement between the ET-MT heat map on the most salient location is explicitly low with merely 27% of the most salient location in the ET map is observed in the ET map. The accuracy for expoze.io predicting the most salient area in ET maps is only 48%

3.3. Discussion

In this section, we investigate how well the current expoze.io model performs in predicting saliency on webpages and whether crowdsourced mouse tracking can be an alternative to eye tracking data to train a neural network.

Although expoze.io is trained on a natural image dataset, the results indicate that expoze.io is able to produce saliency maps at an acceptable level to humans (SIM=0.60, sAUC=0.82). However, there is still nearly 60% false prediction of the most salient location. There is still room for improvement to enhance the accuracy by training the GAN model on webpage dataset.

We investigate the utility of using mouse position data collected on AMT with current parameters to mimic the human field of vision as an alternative for eye tracking data. The agreement between MT to ET (SIM=0.58, sAUC=0.66) is even worse than expoze.io to ET (SIM=0.60, sAUC=0.82). Adding the number of MT participants is less likely to reduce the performance gap. The difference between MT and ET may be caused by the ill-defined mouse tracking parameter, as mentioned in Section 2.3. Mouse tracking data collection. The values for the blurring time, intensity and unblurring time need to be experimentally verified and optimized. The finding adversely affects the reliability of mouse tracking ground truth as an alternative to eye tracking for model evaluation under the current settings.

4. Objective 2: Determining and comparing the starting point

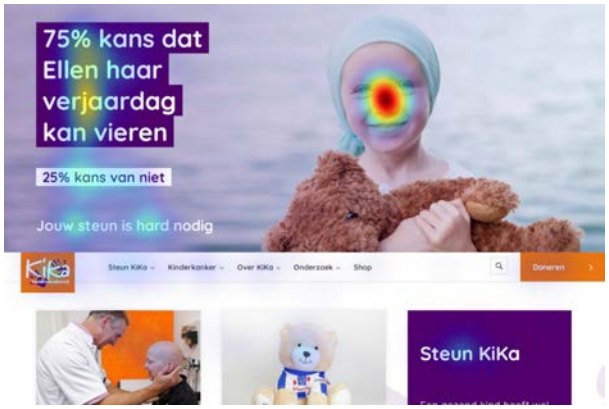


Figure 20 Face is the most salient area in the example webpage. Is the face also the starting point?

Some neural-computational saliency models (Henderson, 2003; Itti, 2005; Tseng & Howes, 2008; Underwood, 2009) suggested that the shifts of attention and saccadic eye movements are initiated toward the salient intensity level. If this model succeeded, a general scanpath can be produced based on the resulting heat map. The starting point, which is defined as the first visual element that users are most likely to look at first on a webpage, plays a critical role in viewing sequence analysis. The motivation questions in this section are listed as the following:

- How can we determine the starting point from ET and MT data at the group level?
- Is the starting point always located in the most salient area on the ET heat map?
- Are the starting points of ET & MT matching?

4.1. Methodology

4.1.1. Temporal binning

It is assumed that most people look into certain locations at the initial time due to stimulus driven orienting, and the gaze point becomes more dispersed over time (Sutcliffe & Namoun, 2012). Thus, the initial time of viewing (referred as the first temporal bin) has to be defined.

Fixations are the times when the eyes essentially stop scanning, holding the central foveal vision in place so that the visual system can take in detailed information about what is being looked at (*Tobii Pro Lab User Manual*, 2021). Fixation time has to be long enough to enable encoding of the visual information around the fixation point. Although there is no explicit definition of fixation and fixation duration, fixation duration is considered as 250 ms within the range of webpage eye tracking studies (Djamasbi, 2014; Tullis & Albert, 2008) for computational approach.

To extend the approach as described in Section 2.4.2. Reducing the effect of position bias, the first temporal bin for eye tracking data is 0-500 ms by considering 250 ms saccadic latency time of eye moves off the central target, plus 250 ms fixation time for engagement with visual information. The first temporal bin for mouse tracking data is 0-700 ms by considering the cursor tends to lag behind gaze by 700 ms (Huang et al., 2012). Figure 21 shows the example of gaze point in the first temporal bin for starting point analysis.

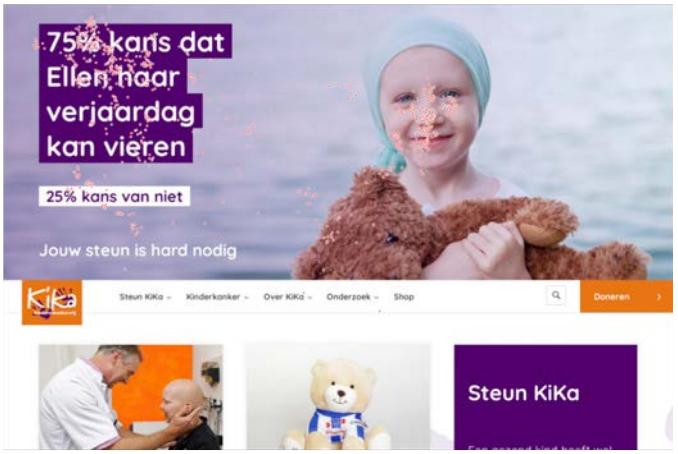
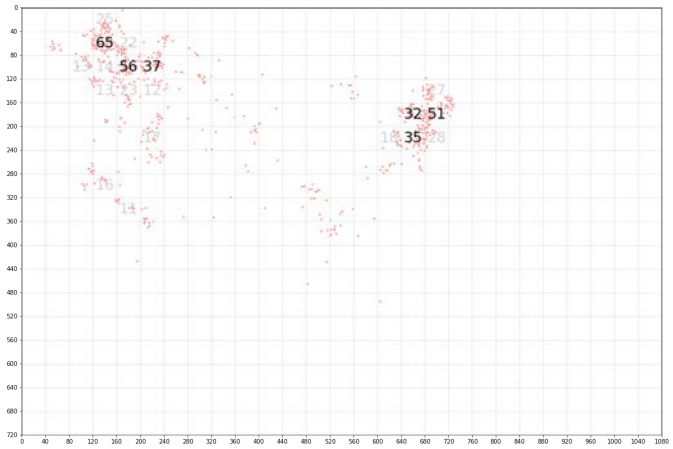
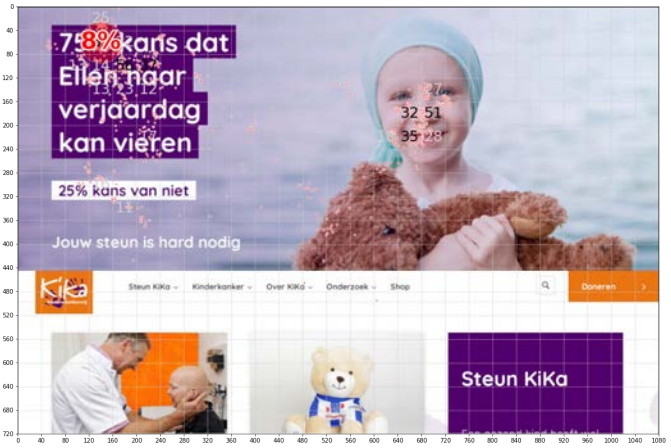


Figure 21 Raw eye tracking gaze point in the first temporal bin of 0-500 ms.



(a) The cell in column 4 and row 2 contains the highest number of gazes.



(b) Starting point (red circle) is annotated with the ratio of gaze point inside the cell.

Figure 22 Example of determining starting point by grid method.

4.1.2. Grid method

To determine the starting point, the first simple approach is to apply a grid-layout segmentation with grid size a 40 x 40 px. The cell containing the maximum amount of gaze point is annotated as the starting point, as illustrated in Figure 22. The percentage value shown inside the dot is the ratio of gaze point inside the cell over alldata points within the temporal bin.

Limitation of the grid method

To verify whether the grid method can generate a consistent starting point, different grid dimensions are applied. It is found the method is sensitive to the size of cells in some stimuli, as an example demonstrates in Figure 24. As the grid is structured in fixed size, without associating with the position of actual visual stimuli or other data points in the neighboring cells, a miss leading starting point may be produced. Figure 23 illustrates an example of two clustered gaze points (in color green and blue) with a mesh applied, and location F4 is considered the starting point with the grid method. Although the green cluster contains more gaze than the blue one, the points are split evenly, and the number of counts in the cell does not represent the actual condition.

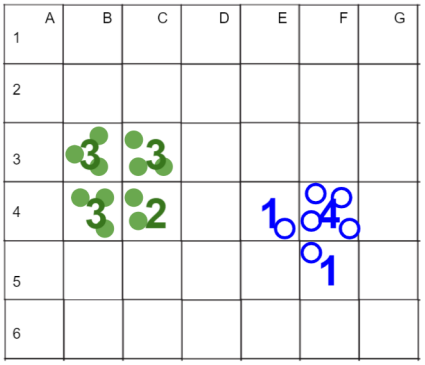


Figure 23 Miss leading starting point (location F4) produced with grid method.

In view of the limitations brought by the grid method, the methodology of identifying the densest area is revisited by considering the distance between individual points, and will be discussed in the next section.

(a) Grid size 20 x 20 px

(b) Grid size 40 x 40 px

(c) Grid size 120 x 120px

Figure 24 Different grid sizes could vary the starting point (red circles).

4.1.3. Density map method: Kernel density estimation with Gaussian kernel

Kernel density estimation (KDE) is a non-parametric way to estimate the probability density function of a random variable. The contribution of each data point is smoothed out from a single point into a region of space surrounding it. Figure 25 illustrates the KDE with a Gaussian kernel visualization on data points. The parameter of map generation can be found in Section 2.4.1. Heat map generation.

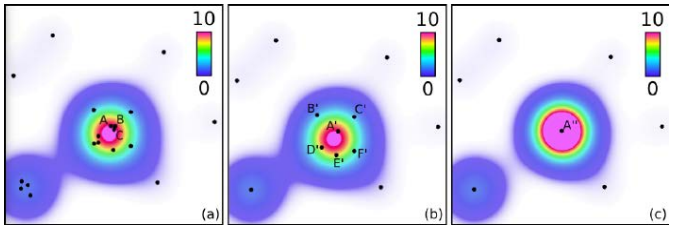


Figure 25 Visualization of KDE with a Gaussian kernel computed on data points (Perrot et al., 2015). Image (a) presents the density function visualization of the original data point. On image (b) The closest points have been merged into a single point whose weight is the sum of the weights of the merged points. For instance, points A, B and C have been merged into A'. The weight of A' is the sum of the weights of A, B and C. Image (c) shows the result of another merging pass. Points B' through F' have been merged into A".

The density map method is able to produce a more accurate starting point location compared to grid method, reflecting the densest area of MT or ET participants located at. By considering the time interval mentioned in Section 4.1.1. Temporal binning, density maps are produced individually from ET and MT data, and the densest areas on the maps are considered as the starting points. An example of identifying the starting point in ET and MT density maps, shown in Figure 27.

The random distribution of position data points on a pixel is 0.002%. The starting point of ET contains 7.0% data point (SD = 2.42) and the starting point of MT contains 6.9% data point (SD = 1.96) in the average of 51 stimuli. This indicates the methodology of the temporal bin and density map method can generate a convincing starting point from the dataset.

4.2. Starting point analysis

To test the hypothesis whether the starting point is always located in the most salient area on the ET heat map, the analysis of the starting point on the ET and MT heat maps is carried out. Figure 26 demonstrates a plot of starting points in ET and MT data from the first temporal bin, and overlays the ET heat map (output from 0-5 s viewing time). In this example, the most salient area is the girl's face on the right-hand side but the starting point of ET is generally located on the large text box on the top left area, and the MT starting point agrees with ET data. This means most ET and MT participants looked at the text box first instead of the most salient face area in this stimulus.

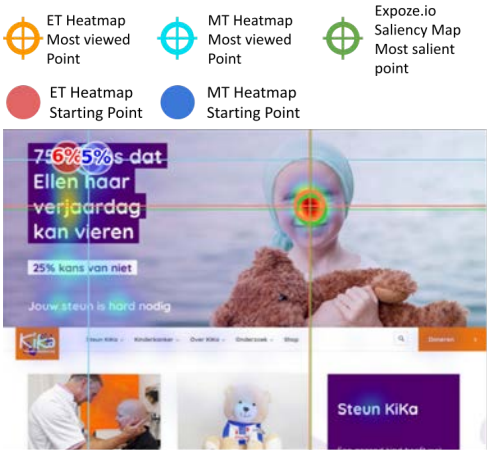
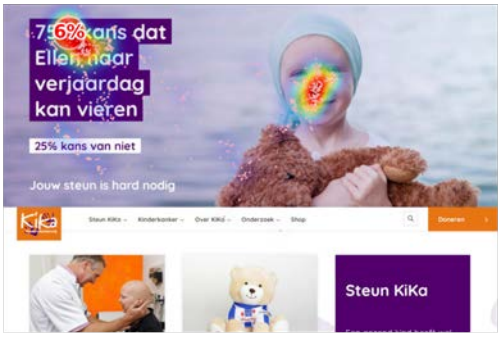
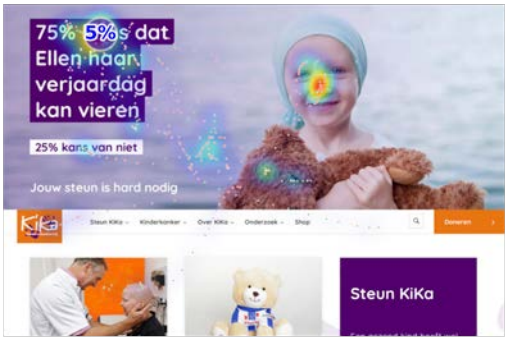


Figure 26 A plot of ET and MT starting points on ET heat map produced from 5s free viewing.



(a) ET starting point (in color red)



(b) MT starting point (in color blue)

Figure 27 ET density map (a) created from the first temporal bin 0-500 ms. MT's density map (b) created from the first temporal bin 0-700 ms.

51 stimuli plotted with the starting point of ET and MT, and the most salient point in ET heat map, are examined to understand whether the ET starting point shares the same visual element as the most salient point, and if ET and MT starting point agree with each other. As logos appear in every webpage, we want to know whether a logo is likely to be viewed as the first place in ET and MT settings. A classification of logos is carried out on the data set as well. All plotted stimuli can be found in Appendix B and the table 1 shows the result of starting point analysis.

ET starting point share the most salient point on ET heat map	41.18%
ET-MT staring point agree	28.41%
MT starting point is in the logo area	70.59%
ET starting point is in the logo area	9.80%

Table 1 Starting point analysis

4.3. Discussion

Three questions are raised at the beginning of this section: (1) How can we determine the starting point from ET and MT data at the group level? (2) Is the starting point always located in the most salient area on the ET heat map? (3) Are the starting points of ET & MT matching?

The density map method is able to generate a convincing starting point from ET and MT data with the first temporal at the group level. The initial hypothesis of the starting point is it is always located in the most salient point on the ET heat map. However, the hypothesis is disproved, with only 41% of starting points located in the most salient area. This indicates that the viewing ordering cannot be simply deduced from heat map's salient intensity levels.

There is a huge gap between the agreement of ET-MT starting point, with only 28% of them sharing the same area. This shows the MT data at the current setting is unreliable to use as a training dataset for starting point prediction in the GAN model. An unexpected finding is that over 70% of the starting points are in the logo area with MT tracking condition, while only 10% ET starting points are observed in the logo area. This reflects that there is a distinctive viewing behavior in ET and MT study under the current setting. The behavioral difference may be caused by the ill-defined mouse tracking parameter, as mentioned in Section 2.3. Mouse tracking data collection. The values for the blurring time, intensity and unblurring time need to be experimentally verified and optimized.

5. Objective 3: Determining a general scanpath

Is it possible to find a scanpath from both mouse tracking and eye tracking data at the group level? Although the hypothesis of generating a general scanpath based on the heat map's salient intensity level is disproved, the methodology of identifying the starting point from the density map in a temporal bin provides insight for producing a general scanpath by a shift of time.

5.1. Methodology and result

By extending the methodology in Section 4.1.1. Temporal binning, the following temporal bins are 250 ms after the starting point in ET data, considering the fixation time for engagement with visual information (Djamasbi, 2014; Tullis & Albert, 2008). Similar configuration on MT data with the following temporal bins set as 700 ms, by considering the cursor tends to lag behind gaze by 700 ms (Huang et al., 2012).

(a) 0-500ms (b) 500-750ms ... (c) 2500-2750ms (d)2750-3000ms

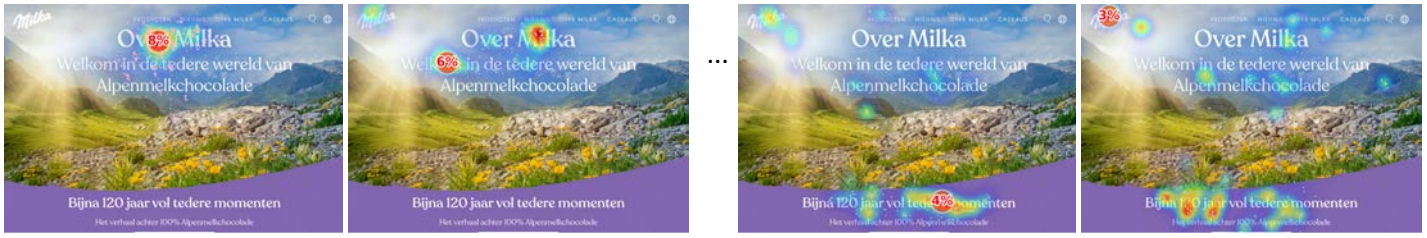


Figure 28 The most commonly viewed point (circle in color red) shifts in each temporal bin of ET data

A density map is generated in each temporal bin, and the densest point on the map is considered the most commonly viewed point within that time interval. Figure 28 demonstrates the location of the densest point shifts over time. By connecting the point in each temporal bin, a general viewing order can be produced across spatial-temporal aspects of the data. 11 temporal bins are resulting in 11 most commonly viewed points in time sequence. It is expected there will be two continuous points very close to each other if an attractive element exists. If the two continuous points are within the distance of 40 px, the following point will be merged in the previous point to avoid overcrowding in the scanpath. The size of the merged circle will be larger to indicate the viewing duration of that point. An ascending sequential numbering annotated in each of the most commonly viewed points indicates the viewing order. Figure 29 shows an example of a general scanpath produced from ET and MT data by the temporal bin shift method at group level. All general scanpath produced can be found in Appendix B.

(a) ET general scanpath

(b) MT general scanpath

Figure 29 Example of general scanpaths produced from temporal bin shift method

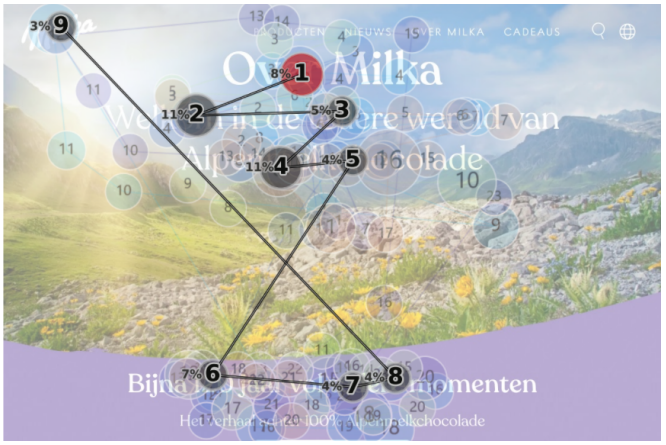


Figure 30 ET general scanpath overlay on scanpaths from 5 individual participants with Tobii Fixation Filter

Figure 30 shows a comparison of ET general scanpath generated by temporal bin shift method to the scanpaths of 5 individual participants generated by Tobii Fixation I-VT Filter[2], the default setting of eye tracking software Tobii Pro Lab to generate scanpath from a participant (*Tobii Pro Lab User Manual*, 2021). The comparison demonstrates that the general scanpath is able to reproduce the viewing order from multiple participants, by starting at the title, then scanning the text from left-to-right and top-to-bottom as general gaze sweep behavior (Malcolm et al., 2018).

5.2. Discussion

A question is raised in the beginning of this section: Is it possible to find a scanpath from both mouse tracking and eye tracking data at group level? The comparison in the previous section shows that it is possible to deliver a reasonable general scanpath with the novel approach of the temporal bin shift method. Unlike the heat map, there is no ground-truth for scanpath from all participants, and it is hard to verify the validity of the general scanpath produced. Thus, we can only address the potential limitations of the temporal bin shift method:

1. **Time validity**: Later the time point, more likely the top-down attention will take over (Connor et al., 2004), the gaze is then not visual driven, but influenced by personal preference. It is expected the density map in the later temporal bins is less indicative, as the gazes are scattered in different elements on the page driven by personal viewing characteristics

[2] The I-VT fixation filter is set to define the minimum fixation duration to 60 ms, with a velocity threshold of 30°/s.

2. **Insufficient dynamic in bin size**: The temporal time bin is fixed by considering the mean fixation and latencies of eye-hand coordination, referencing the averaged result from the data set and finding from other studies. This approach assumes the participant's performance, the cognitive load on webpages, and time spent on each visual element are the same. The variance between each participant and stimuli is not considered.

6. Conclusion

In this project, we investigate if the starting point, the first element most commonly looked at on a webpage, can be deduced from the most salient area in an eye tracking map. A new concept is introduced to cluster starting points by producing density maps at fixed viewing duration. 51 structured web images with mouse tracking data, eye tracking data, expoze.io prediction saliency maps are reviewed.

MT does not concur with ET at the current setting:
The study of comparing map similarities found that the level of agreement between MT and ET is even lower than expoze.io and ET. The finding adversely affects the reliability of mouse tracking ground truth as an alternative to eye tracking for model evaluation under the current setting. The parameters of the mouse-contingent multi-resolutional paradigm are not well investigated in the setting to mimic the human field of vision on the web viewing.

The starting point is not necessary on the highest saliency peak:
In the study of determining and comparing the starting point, no strong association is observed between the starting point and the most salient point on heat maps. The hypothesis of the starting point in the most salient area is disproved.

Produce a general scanpath with temporal bin shift method:
The study of determining a general scanpath shows that it is possible to produce a reasonable viewing order by the shift of temporal bin. Time validity and insufficient dynamic in bin size are the limitations of the proposed method.

6.1. Areas for enhancements

Center fixation cross
Although we proposed a data-driven approach to reduce positional bias, it is a post hoc analysis by removing gaze or mouse position data points around fixation cross area at early viewing time. If a salient element locates in the fixation cross area, the early gaze on it will be considered as noise. Salient elements in the center may be neglected due to the loss of spiritual and temporal information. This may influence the order of the elements in the produced scanpath, especially for the starting point.

Some studies (Peacock et al., 2020; Rothkegel et al., 2016, 2017; Trukenbrod et al., 2019) show that moving the initial fixation cross from the center to a random location can reduce central fixation bias in free viewing priorly. We suggest randomizing the fixation cross in both ET and MT studies to minimize the positional bias issue.

Mouse tracking setting to mimic peripheral vision
Although there are multiple success cases (Anwyl-Irvine et al., 2021; Jiang et al., 2015; Kim et al., 2017; Lio et al., 2019; Sidorov et al., 2020) that mouse tracking can serve as an approximation of eye-tracking data by mimicking peripheral vision, the parameter of the blurring filter and the mechanism revealing unblurred area varies between studies. Table 2 provides an overview of blurring parameters in our current study and others' works.

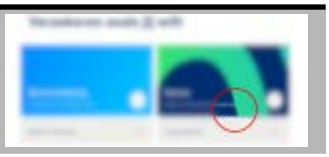
Related work	Aperture diameter to mimic foveal (px)	Blurred edge of aperture (px)	Blur sigma (px)	Unblurring mechanism	Sample of blurring setting
Our MT setting	110	28	28	200ms delay on a fixed location	
Anwyl-Irvine et al. (2021)	5% of viewing screen size (~96 px on a 1920*1080 px screen)	10	20	Reveals unblurred area along cursor	
Sidorov et al. (2020)	400	No (sharp edge)	15	Hole the mouse button to deblur (max. deblurring time for 4s)	
Lio et al. (2019)	110	Yes but parameter not stated	40	Reveals unblurred area along finger	
Kim et al. (2017)	60-100	No (sharp edge)	30-50	Click to unblur	

Table 2 Comparison of parameters to mimic peripheral vision in different studies

Table 2 shows that the aperture diameter and burring sigma at the current setting are generally within the benchmark range. However, the diverse unblurring mechanisms across different studies indicate more detailed study on how mouse movement reveals gaze behavior is required.

6.2. General scanpath prediction in future

Our work shows a novel method to produce a general scanpath from multi-duration saliency maps. On the premise of harvesting high quality data, the scanpath prediction model could be trained on saliency information with spatial and temporal dimensions.

The recent work of Xia et al. (2020) provided an overview of different metrics to compare consistency between two scanpaths which could help to evaluate the agreement of general scanpath across individual participant's scanpath in objective metrics.

Conia et al. (2018) designed a method to predict saliency by incorporating an attention mechanism based on the combination of Long Short Term Memory (LSTM) (Hochreiter & Schmidhuber, 1997) and convolutional networks. LSTM is a type of recurrent neural network capable of learning order dependence in sequence prediction problems. Figure 31 shows a LSTM unit, three different gates: input gate, forget gate and output gate that regulate information flow in an LSTM cell. LSTM uses memory to store the relevant contextual information and then add/modify/delete contextual information based on new Input. This helps the network to predict well for new input when context from very old input will be required to be referred.

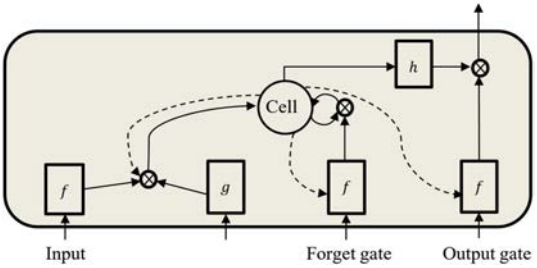


Figure 31 LSTM unit structure diagram (Zaroug et al., 2020)

The underlying idea of using LSTM is to predict the current salient region according to the previous ones to increase the sequential dependence between salient regions. The conventional saliency map is a two-dimensional topographic map that encodes saliency values. By considering the timestamp associated with gaze point, the information of temporal dimension stack on spatial dimension and a salience map in three-dimensional space is created (as shown in Figure 32). The values of each temporal slice can be normalized, converting the slice into a probability map that represents the probability of each pixel being looked at by viewers at each timestep. Scanpath can also be extracted from the 3D saliency map.

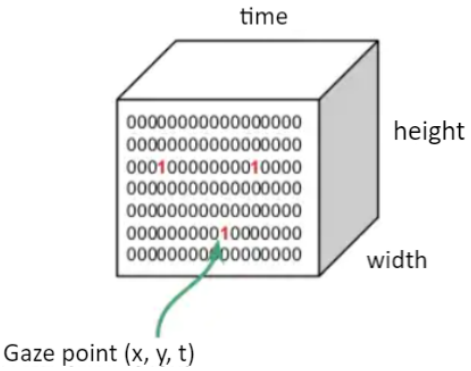


Figure 32 Three-dimensional saliency map

On the premise of harvesting high quality data, the scanpath prediction model could possibly be trained with 3D saliency maps with spatial and temporal dimensions on the LSTM neural network.

7. Acknowledgement

Special thanks to Coert van Gemeren and Ingrid Nieuwenhuis from Alpha.One, and Samson Chota from Utrecht University for providing expertise, support and encouragement throughout the research project. And to all the participants in the study for contributing their time and effort. Without all of you, this study could not have happened.

8. Reference

An intuitive introduction to Generative Adversarial Networks (GANs). (2018). freeCodeCamp.org. https://www.freecodecamp.org/news/an-intuitive-introduction-to-generative-adversarial-networks-gans-7a2264a81394/

Anwyl-Irvine, A. L., Armstrong, T., & Dalmaijer, E. S. (2021). *MouseView.js: Reliable and valid attention tracking in web-based experiments using a cursor-directed aperture*. https://doi.org/10.31234/osf.io/rsdwg

Borji, A. (2021). Saliency Prediction in the Deep Learning Era: Successes and Limitations. In *IEEE Transactions on Pattern Analysis and Machine Intelligence* (Vol. 43, Issue 2, pp. 679–700). https://doi.org/10.1109/tpami.2019.2935715

Borji, A., Tavakoli, H. R., Sihite, D. N., & Itti, L. (2013). Analysis of scores, datasets, and models in visual saliency prediction. *Proceedings of the IEEE International Conference on Computer Vision*, 921–928. https://www.cv-foundation.org/openaccess/content_iccv_2013/html/Borji_Analysis_of_Scores_2013_ICCV_paper.html

Bylinskii, Z., Judd, T., Borji, A., Itti, L., Durand, F., Oliva, A., & Torralba, A. (2015). *Mit saliency benchmark*. http://saliency.mit.edu/results_mit300.html

Bylinskii, Z., Judd, T., Oliva, A., Torralba, A., & Durand, F. (2019). What Do Different Evaluation Metrics Tell Us About Saliency Models? In *IEEE Transactions on Pattern Analysis and Machine Intelligence* (Vol. 41, Issue 3, pp. 740–757). https://doi.org/10.1109/tpami.2018.2815601

Connor, C. E., Egeth, H. E., & Yantis, S. (2004). Visual attention: bottom-up versus top-down. *Current Biology: CB*, *14*(19), R850–R852. https://doi.org/10.1016/j.cub.2004.09.041

Cornia, M., Baraldi, L., Serra, G., & Cucchiara, R. (2018). Predicting Human Eye Fixations via an LSTM-based Saliency Attentive Model. *IEEE Transactions on Image Processing: A Publication of the IEEE Signal Processing Society*. https://doi.org/10.1109/TIP.2018.2851672

Dalmaijer, E. S., Mathôt, S., & Van der Stigchel, S. (2014). PyGaze: an open-source, cross-platform toolbox for minimal-effort programming of eyetracking experiments. *Behavior Research Methods*, *46*(4), 913–921. https://doi.org/10.3758/s13428-013-0422-2

Djamasbi, S. (2014). Eye Tracking and Web Experience. *AIS Transactions on Human-Computer Interaction*, *6*(2), 37–54. https://aisel.aisnet.org/thci/vol6/iss2/2/

Drusch, G., Bastien, J., & Paris, S. (2014). *Analysing eye-tracking data: From scanpaths and heatmaps to the dynamic visualisation of areas of interest*. https://www.semanticscholar.org/paper/d5dfacbfe406c9c1c135e838893efd88e279d044

Expoze.io. (2021, June). *White paper - The tech behind expoze.io*. https://www.expoze.io/uploads/75/91/445e3fc2c638df338a13ffaff89e820e.pdf

Foulsham, T., & Kingstone, A. (2013). Fixation-dependent memory for natural scenes: an experimental test of scanpath theory. *Journal of Experimental Psychology. General*, *142*(1), 41–56. https://doi.org/10.1037/a0028227

Generate a heatmap in MatPlotLib using a scatter data set. (2017, October). Stackoverflow. https://stackoverflow.com/questions/2369492/generate-a-heatmap-in-matplotlib-using-a-scatter-data-set

Henderson, J. M. (2003). Human gaze control during real-world scene perception. *Trends in Cognitive Sciences*, *7*(11), 498–504. https://doi.org/10.1016/j.tics.2003.09.006

Hessels, R. S., Niehorster, D. C., Nyström, M., Andersson, R., & Hooge, I. T. C. (2018). Is the eye-movement field confused about fixations and saccades? A survey among 124 researchers. *Royal Society Open Science*, *5*(8), 180502. https://doi.org/10.1098/rsos.180502

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, *9*(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

Huang, J., White, R., & Buscher, G. (2012). User see, user point: gaze and cursor alignment in web search. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 1341–1350). Association for Computing Machinery. https://doi.org/10.1145/2207676.2208591

Interaction Design Foundation. (2020, August 18). *Serial position effect: How to create better user interfaces*. https://www.interaction-design.org/literature/article/serial-position-effect-how-to-create-better-user-interfaces

Itti, L. (2005). Models of bottom-up attention and saliency. In *Neurobiology of attention* (pp. 576–582). Elsevier. https://www.sciencedirect.com/science/article/pii/B9780123757319500987

Jiang, M., Huang, S., Duan, J., & Zhao, Q. (2015). Salicon: Saliency in context. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 1072–1080. https://www.cv-foundation.org/openaccess/content_cvpr_2015/html/Jiang_SALICON_Saliency_in_2015_CVPR_paper.html

Judd, T., Ehinger, K., Durand, F., & Torralba, A. (2009). Learning to predict where humans look. *2009 IEEE 12th International Conference on Computer Vision*, 2106–2113. https://doi.org/10.1109/ICCV.2009.5459462

Kim, N. W., Bylinskii, Z., Borkin, M. A., Gajos, K. Z., Oliva, A., Durand, F., & Pfister, H. (2017). BubbleView: An Interface for Crowdsourcing Image Importance Maps and Tracking Visual Attention. *ACM Trans. Comput.-Hum. Interact.*, *24*(5), 1–40. https://doi.org/10.1145/3131275

Lange, B., Hunfalvay, M., & Murray, N. (2018). *Reliability of computerized eye-tracking reaction time tests in non-athletes, athletes, and individuals with traumatic brain injury*. researchgate.net. https://www.researchgate.net/profile/C-M-Roberts/publication/322937197_Reliability_of_computerized_eye-tracking_reaction_time_tests_in_non-athletes_athletes_and_individuals_with_traumatic_brain_injury/links/5a78804945851541ce5c6c6a/Reliability-of-computerized-eye-tracking-reaction-time-tests-in-non-athletes-athletes-and-individuals-with-traumatic-brain-injury.pdf

Lindgaard, G., Fernandes, G., Dudek, C., & Brown, J. (2006). Attention web designers: You have 50 milliseconds to make a good first impression! *Behaviour & Information Technology*, *25*(2), 115–126. https://doi.org/10.1080/01449290500330448

Lio, G., Fadda, R., Doneddu, G., Duhamel, J., & Sirigu, A. (2019). Digit-tracking as a new tactile interface for visual perception analysis. In *Nature Communications* (Vol. 10, Issue 1). https://doi.org/10.1038/s41467-019-13285-0

Magill, R., & Anderson, D. (2010). *Motor learning and control*. McGraw-Hill Publishing New York. https://www.academia.edu/download/62953224/Motor_Learning_and_Control__Concepts_and_-_Anderson__David20200414-28877-18hp4j2.pdf

Malcolm, G. L., Silson, E. H., Henry, J. R., & Baker, C. I. (2018). Transcranial Magnetic Stimulation to the Occipital Place Area Biases Gaze During Scene Viewing. *Frontiers in Human Neuroscience*, *12*, 189. https://doi.org/10.3389/fnhum.2018.00189

MIT Saliency Benchmark. (2012). [Data set]. In *A Benchmark of Computational Models of Saliency to Predict Human Fixations*. http://saliency.mit.edu/datasets.html

Pan, J., Ferrer, C. C., McGuinness, K., O'Connor, N. E., Torres, J., Sayrol, E., & Giro-i-Nieto, X. (2017). SalGAN: Visual Saliency Prediction with Generative Adversarial Networks. In *arXiv [cs.CV]*. arXiv. http://arxiv.org/abs/1701.01081

Peacock, C. E., Hayes, T. R., & Henderson, J. M. (2020). Center Bias Does Not Account for the Advantage of Meaning Over Salience in Attentional Guidance During Scene Viewing. *Frontiers in Psychology*, *11*, 1877. https://doi.org/10.3389/fpsyg.2020.01877

Perrot, A., Bourqui, R., Hanusse, N., Lalanne, F., & Auber, D. (2015). *Large Interactive Visualization of Density Functions on Big Data Infrastructure*. https://doi.org/10.1109/LDAV.2015.7348077

Pinto, Y., van der Leij, A. R., Sligte, I. G., Lamme, V. A. F., & Scholte, H. S. (2013). Bottom-up and top-down attention are independent. *Journal of Vision*, *13*(3), 16. https://doi.org/10.1167/13.3.16

Ratliff, L. J., Burden, S. A., & Sastry, S. S. (2013). Characterization and computation of local Nash equilibria in continuous games. *2013 51st Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, 917–924. https://doi.org/10.1109/Allerton.2013.6736623

Rothkegel, L. O. M., Trukenbrod, H. A., Schütt, H. H., Wichmann, F. A., & Engbert, R. (2016). Influence of initial fixation position in scene viewing. *Vision Research*, *129*, 33–49. https://doi.org/10.1016/j.visres.2016.09.012

Rothkegel, L. O. M., Trukenbrod, H. A., Schütt, H. H., Wichmann, F. A., & Engbert, R. (2017). Temporal evolution of the central fixation bias in scene viewing. *Journal of Vision*, *17*(13), 3. https://doi.org/10.1167/17.13.3

Salthouse, T. A., & Ellis, C. L. (1980). Determinants of eye-fixation duration. *The American Journal of Psychology*, *93*(2), 207–234. https://doi.org/10.2307/1422228

Sidorov, O., Pedersen, M., Shekhar, S., & Kim, N. W. (2020). Are All the Frames Equally Important? *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*, 1–7. https://doi.org/10.1145/3334480.3382980

Sillence, E., Briggs, P., Fishwick, L., & Harris, P. (2004). Trust and mistrust of online health sites. In *Proceedings of the 2004 conference on Human factors in computing systems - CHI '04*. https://doi.org/10.1145/985692.985776

Sutcliffe, A., & Namoun, A. (2012). Predicting User Attention in Complex Web Pages. *Behaviour & Information Technology*, *31*(7). https://doi.org/10.1080/0144929X.2012.692101

Tobii Pro Lab User Manual (Version 1.171). (2021). https://www.tobiipro.com/siteassets/tobii-pro/user-manuals/Tobii-Pro-Lab-User-Manual/

Trukenbrod, H. A., Barthelmé, S., Wichmann, F. A., & Engbert, R. (2019). Spatial statistics for gaze patterns in scene viewing: Effects of repeated viewing. *Journal of Vision*, *19*(6), 5. https://doi.org/10.1167/19.6.5

Tseng, Y.-C., & Howes, A. (2008). The adaptation of visual search strategy to expected information gain. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1075–1084. https://doi.org/10.1145/1357054.1357221

Tullis, T., & Albert, B. (2008). Behavioral and Physiological Metrics. In *Measuring the User Experience* (pp. 167–189). https://doi.org/10.1016/b978-0-12-373558-4.00007-8

Underwood, G. (2009). Cognitive processes in eye guidance: Algorithms for attention in image processing. *Cognitive Computation*, *1*(1), 64–76. https://doi.org/10.1007/s12559-008-9002-7

Vencato, V., & Madelain, L. (2020). Perception of saccadic reaction time. *Scientific Reports*, *10*(1), 17192. https://doi.org/10.1038/s41598-020-72659-3

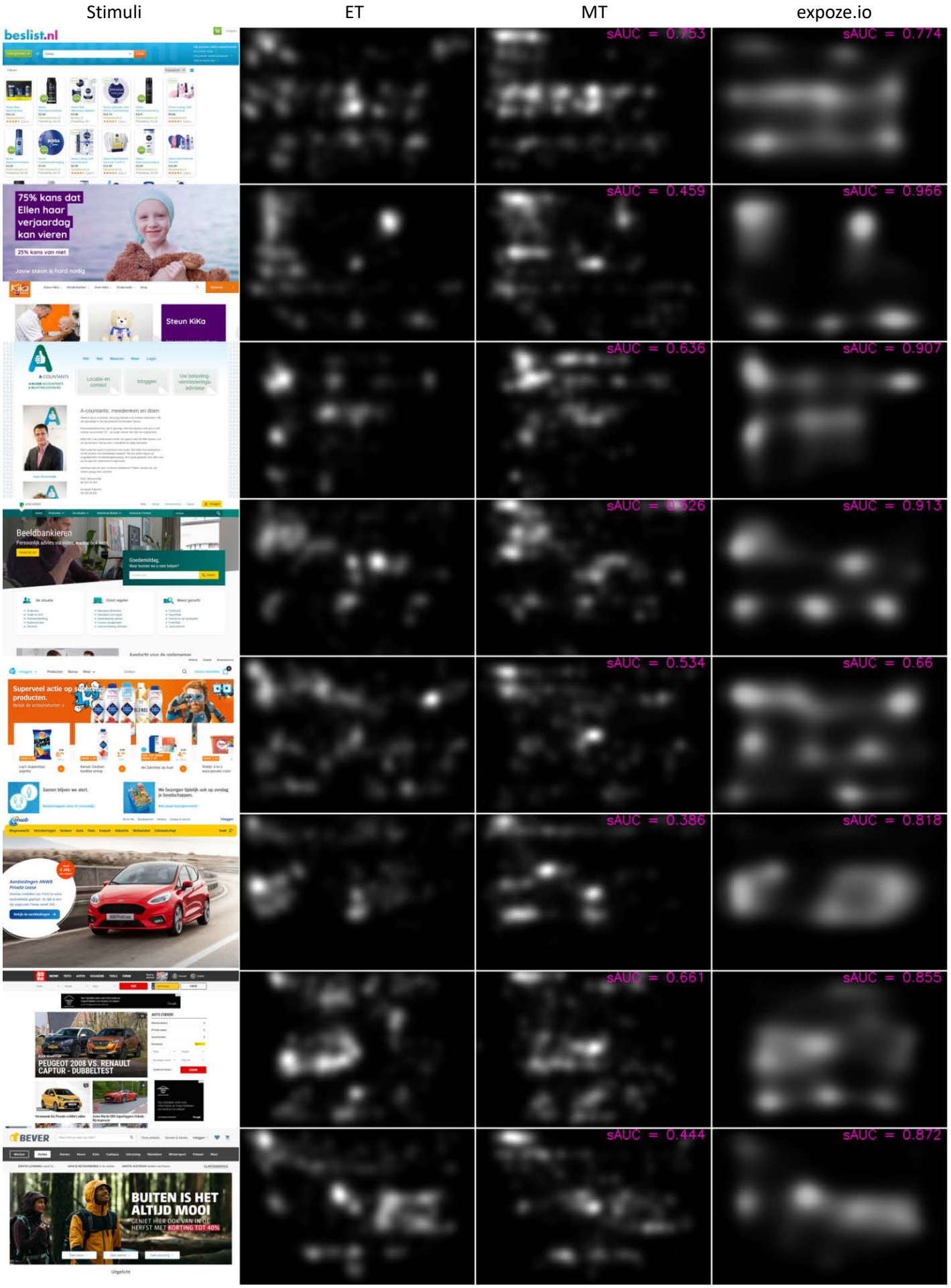
Wiswede, D., Rüsseler, J., & Münte, T. F. (2007). Serial position effects in free memory recall--An ERP-study. *Biological Psychology*, *75*(2), 185–193. https://doi.org/10.1016/j.biopsycho.2007.02.002

Xia, C., Han, J., & Zhang, D. (2020). Evaluation of Saccadic Scanpath Prediction: Subjective Assessment Database and Recurrent Neural Network Based Metric. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *PP*. https://doi.org/10.1109/TPAMI.2020.3002168

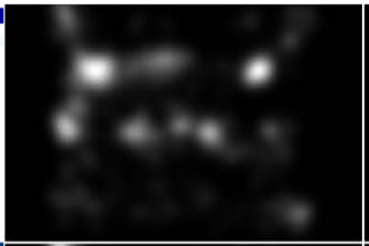
Younis, O., Al-Nuaimy, W., Alomari, M. H., Rowe, F., Choi, W., Paik, S.-B., Yang, C.-Y., Chen, P.-Y., Wen, T.-J., Jan, G. E., Li, X., Zhao, K., Lu, C., & Wang, Y. (2019). *Fig. 1. Human field of view (FOV) for both eyes showing different*. https://www.researchgate.net/figure/Human-field-of-view-FOV-for-both-eyes-showing-different-levels-of-peripheral-vision_fig1_331409336

Zaroug, A., Lai, D. T. H., Mudie, K., & Begg, R. (2020). *Lower Limb Kinematics Trajectory Prediction Using Long Short-Term Memory Neural Networks*. https://doi.org/10.3389/fbioe.2020.00362

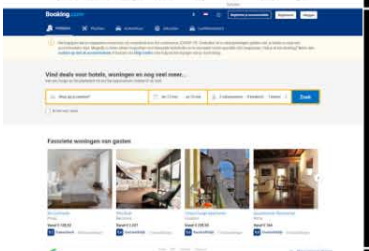
Appendix A: Saliency map comparison

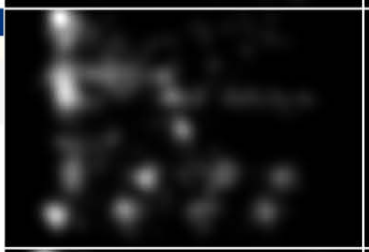



Voordeel in de hoofdrol
met Select


sAUC = 0.627

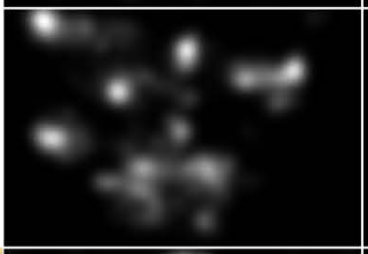

sAUC = 0.827


Booking.com

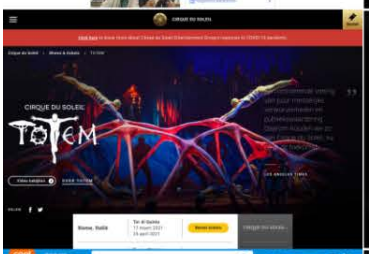

sAUC = 0.475

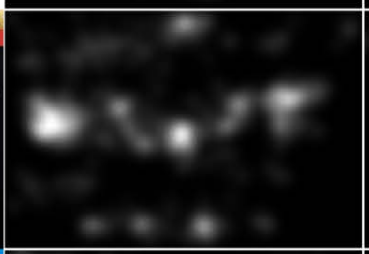

sAUC = 0.556


Coronavirus
Waar bent u naar op zoek?


sAUC = 0.472

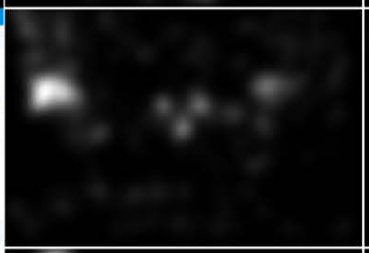

sAUC = 0.692


CIRQUE DU SOLEIL
TOTEM


sAUC = 0.61

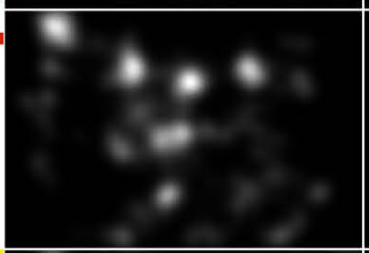

sAUC = 0.905


LAADPALEN.
Gooi in maar volt.

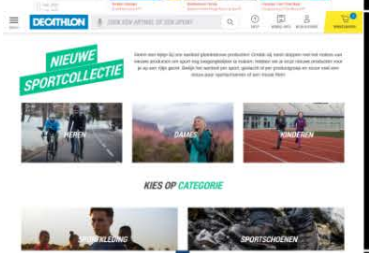

sAUC = 0.586

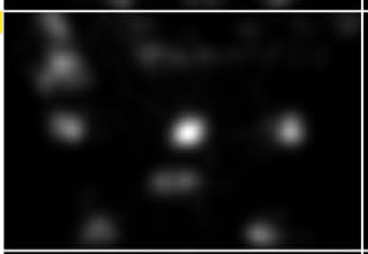

sAUC = 0.993


corendon
Coronavirus
FAQ
Reisvoucher verzilveren?


sAUC = 0.589

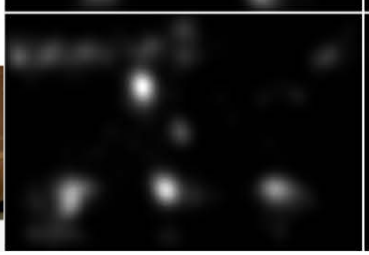

sAUC = 0.869


DECATHLON
NIEUWE SPORTCOLLECTIE
HEREN
DAMES
KINDEREN
KIES OP CATEGORIE
SPORTSCHOENEN


sAUC = 0.839

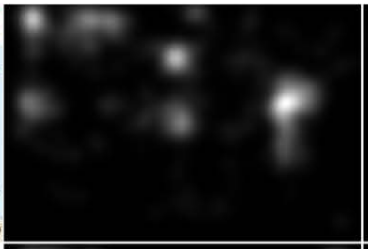

sAUC = 0.796


Laat zien wie je bent
Machtigen
Code ontvangen


sAUC = 0.67

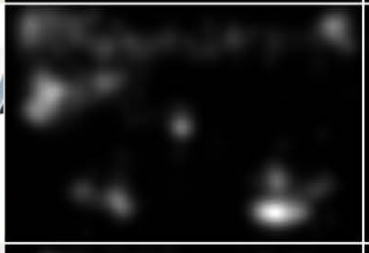

sAUC = 0.987


DJOSER
Dwarrym Djoser


sAUC = 0.302

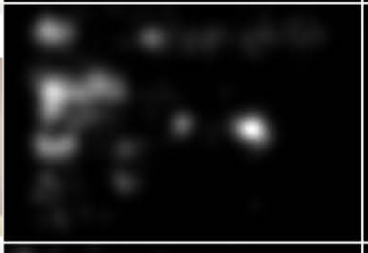

sAUC = 0.822


Erasmus University Rotterdam
Erasmus Universiteit Rotterdam
Make it happen

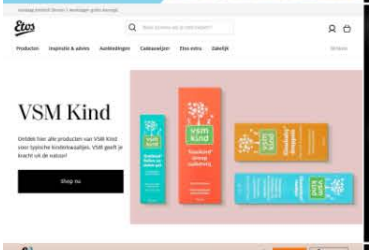

sAUC = 0.322

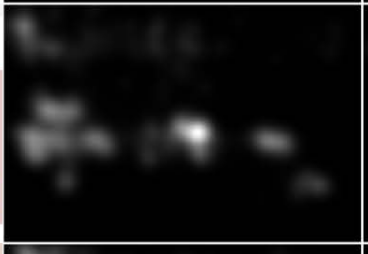

sAUC = 0.613

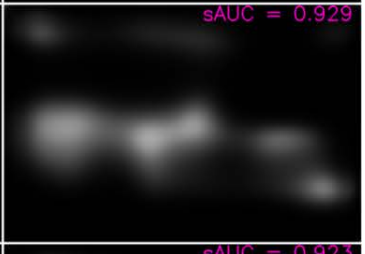

Erasmus MC
Dringend advies: draag mondneusmasker in het Erasmus MC
Bereid uw bezoek goed voor:

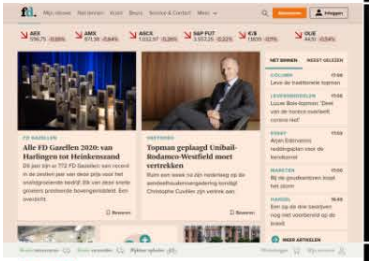

sAUC = 0.549

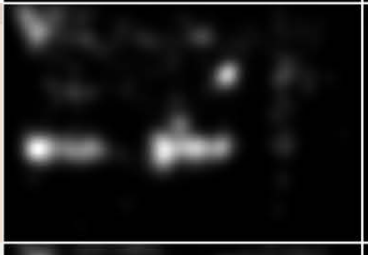

sAUC = 0.907

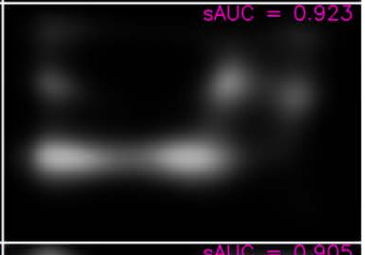

Etos
VSM Kind
Shop nu

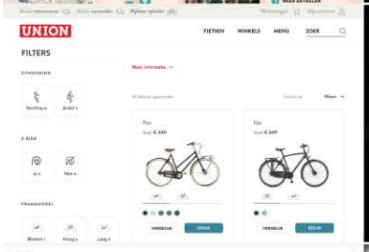

sAUC = 0.75

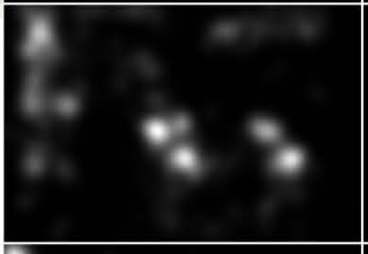

sAUC = 0.929

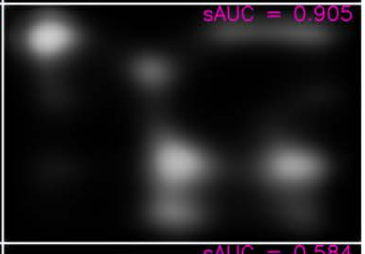

Alle FD Gazellen 2020: van Harlingen tot Heinkenszand
Topman geplaagd Unibail-Rodamco-Westfield moet vertrekken

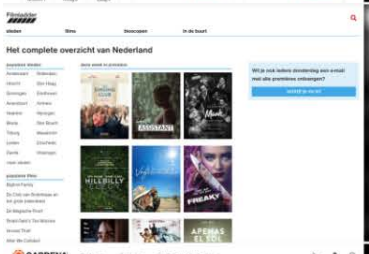

sAUC = 0.36

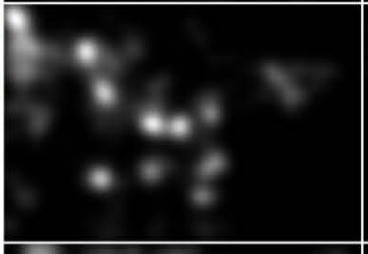

sAUC = 0.923

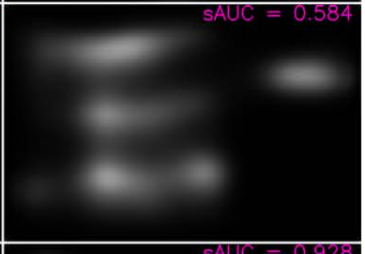

UNION
FILTERS

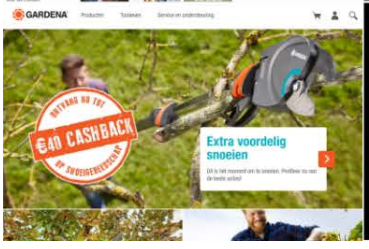

sAUC = 0.589


sAUC = 0.905


Het complete overzicht van Nederland


sAUC = 0.535

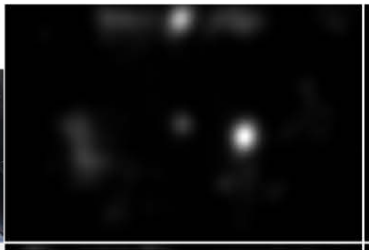

sAUC = 0.584

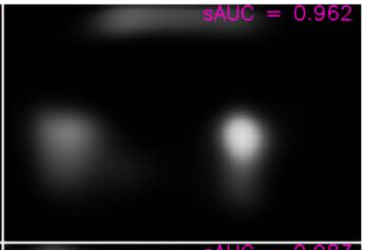

GARDENA
€40 CASHBACK
Extra voordelig snoeien

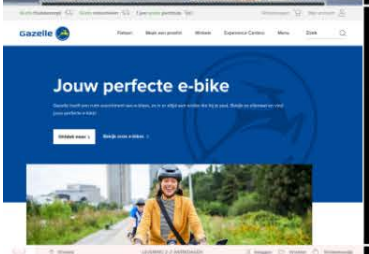

sAUC = 0.443


sAUC = 0.928


WORD
KAPITEIN!


sAUC = 0.914

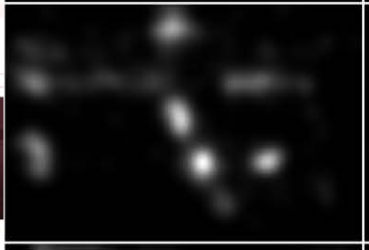

sAUC = 0.962


Gazelle
Jouw perfecte e-bike

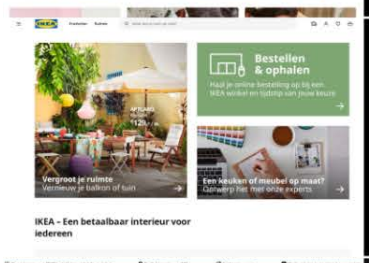

sAUC = 0.709

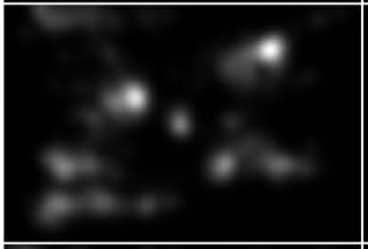

sAUC = 0.983


hunkemöller
Alle Bh's


sAUC = 0.921

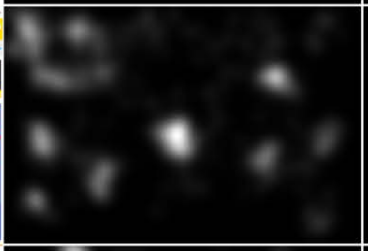

sAUC = 0.95


Bestellen
& ophalen
Vergroot je ruimte
Een keuken of meubel op maat?
IKEA - Een betaalbaar interieur voor
iedereen

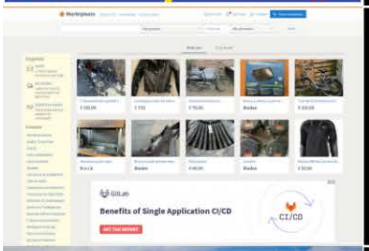

sAUC = 0.596

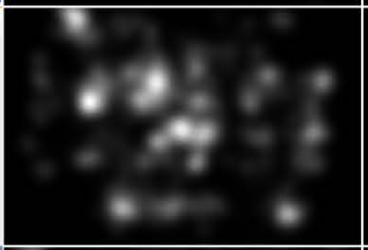

sAUC = 0.776


Intertoys
01:06:06:45
BLACK
FRIDAY
30%
KORTING
MEGA
VOORDEEL
20%
KORTING
25%
KORTING
2 voor
15:-
UNO


sAUC = 0.712

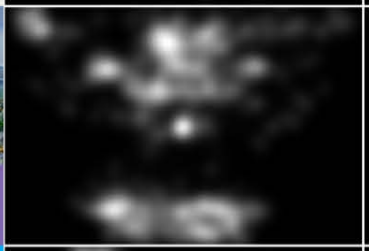

sAUC = 0.829


Benefits of Single Application CI/CD
CI/CD

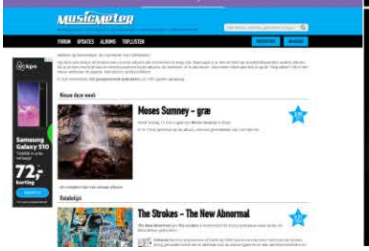

sAUC = 0.728

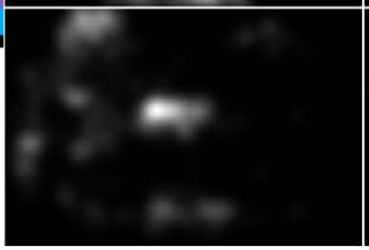

sAUC = 0.68


Milka
Over Milka
Welkom in de tedere wereld van
Alpenmelkchocolade
Bijna 120 jaar vol tedere momenten
Het verhaal achter 100% Alpenmelkchocolade

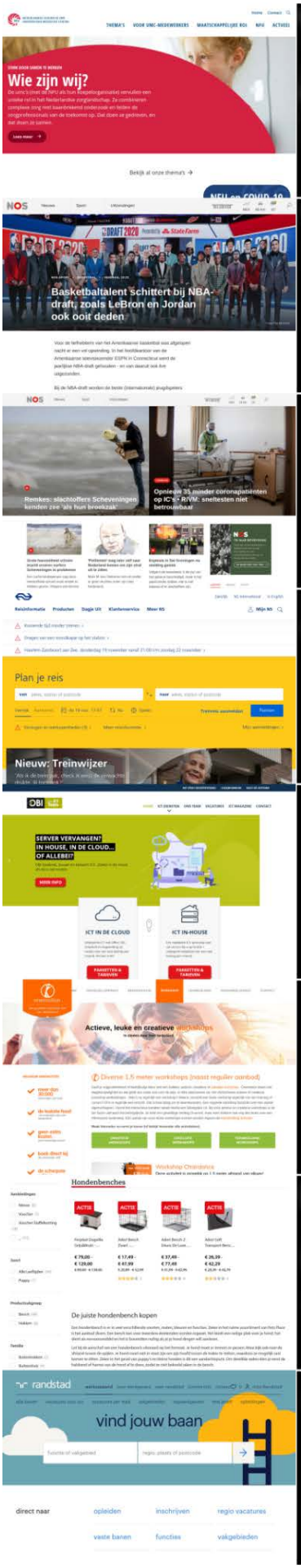

sAUC = 0.668


sAUC = 0.851


musicmeter
Samsung
Galaxy S10
72,-
Moses Sumney – græ
The Strokes – The New Abnormal

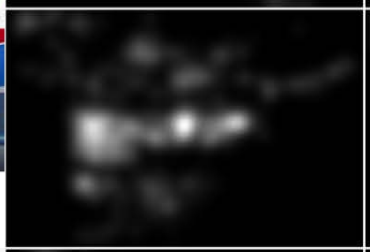

sAUC = 0.865


sAUC = 0.959


Wie zijn wij?
Basketbaltalent schittert bij NBA-draft, zoals LeBron en Jordan ook ooit deden
Remkes: slachtoffers Scheveningen kenden zee 'als hun broekzak'
Opnieuw 35 minder coronapatiënten op IC's • RIVM: sneltesten niet betrouwbaar
Plan je reis
Nieuw: Treinwijzer
SERVER VERVANGEN? IN HOUSE, IN DE CLOUD... OF ALLEBEI?
ICT IN DE CLOUD
ICT IN-HOUSE
Actieve, leuke en creatieve
Hondenbenches
De juiste hondenbench kopen
vind jouw baan
direct naar
opleiden
inschrijven
regio vacatures
vaste banen
functies
vakgebieden


sAUC = 0.704


sAUC = 0.855


sAUC = 0.706

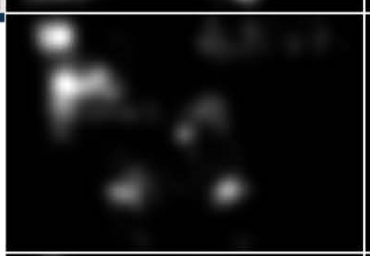

sAUC = 0.837


sAUC = 0.653

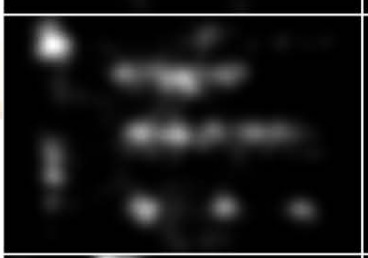

sAUC = 0.644


sAUC = 0.383

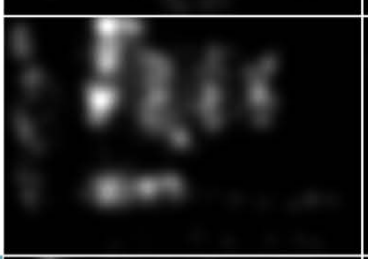

sAUC = 0.734


sAUC = 0.693

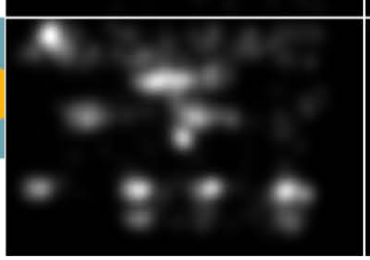

sAUC = 0.886


sAUC = 0.579


sAUC = 0.77

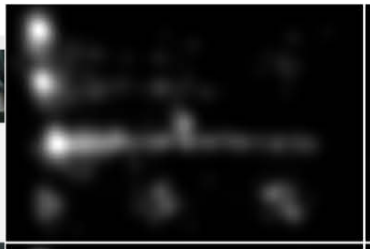

sAUC = 0.692


sAUC = 0.735

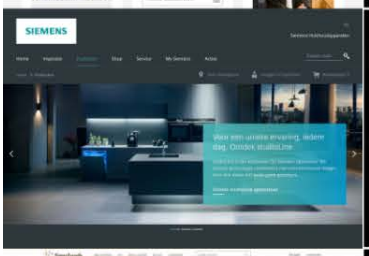

sAUC = 0.55

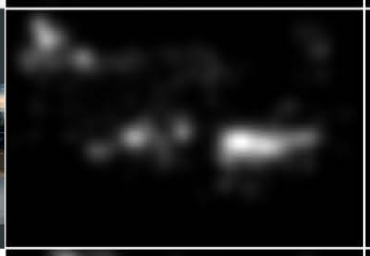

sAUC = 0.771


Rechtspraak


sAUC = 0.49

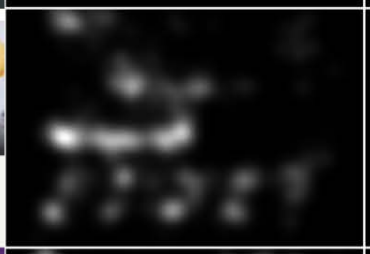

sAUC = 0.94


SIEMENS


sAUC = 0.565

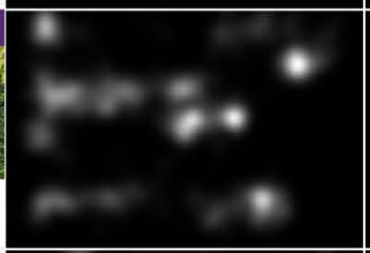

sAUC = 0.969


Bananen mug cake met bosbessen

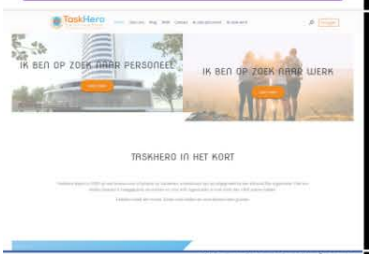

sAUC = 0.529

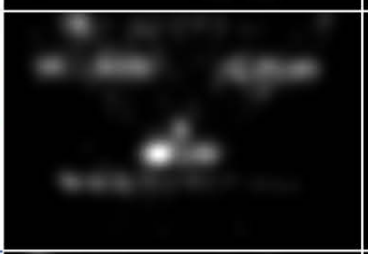

sAUC = 0.62


SNS
Keuze uit meerdere aanbieders voor jouw hypotheek
Waar kunnen we je mee helpen?


sAUC = 0.587

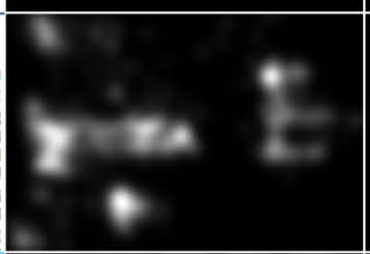

sAUC = 0.798


TaskHero
IK BEN OP ZOEK NAAR PERSONEEL
IK BEN OP ZOEK NAAR WERK
TASKHERO IN HET KORT

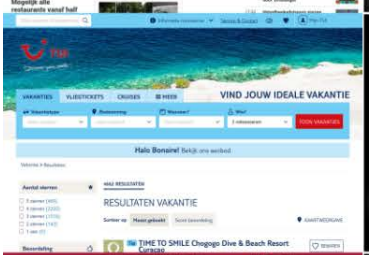

sAUC = 0.763


sAUC = 0.966

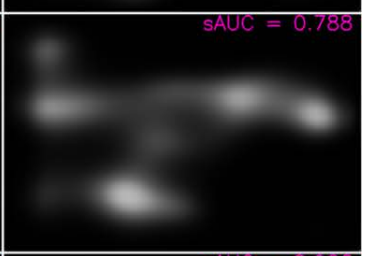

De Telegraaf
ZORGBAAS: 'WE ZIEN STEEDS MEER GUNSTIGE TEKENEN'
Onverzekerde gelovige draait op voor schade dodelijk ongeval


sAUC = 0.438

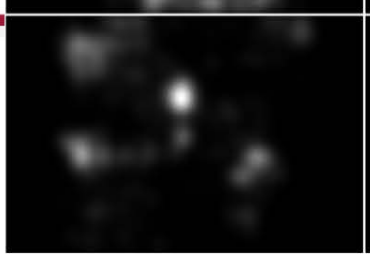

sAUC = 0.683


VIND JOUW IDEALE VAKANTIE
RESULTATEN VAKANTIE
TIME TO SMILE Chogogo Dive & Beach Resort Curaçao

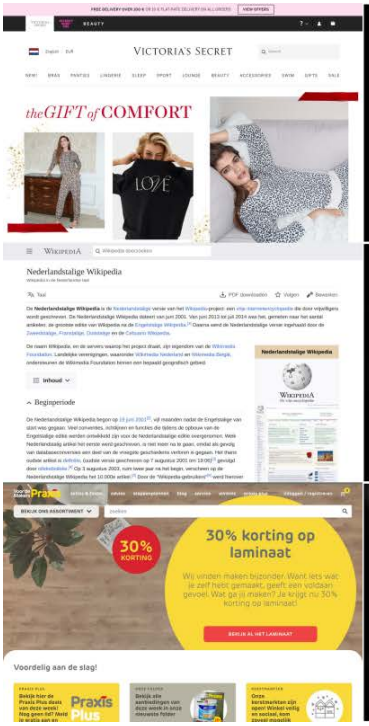

sAUC = 0.711

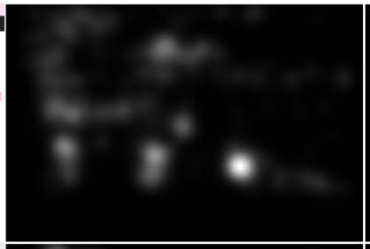

sAUC = 0.788


Best Buy

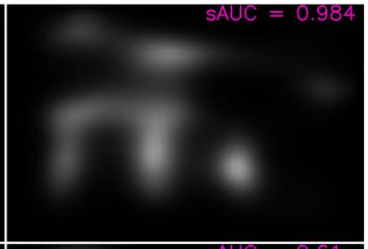

sAUC = 0.795


sAUC = 0.926


VICTORIA'S SECRET
theGIFT of COMFORT
LOVE




sAUC = 0.429

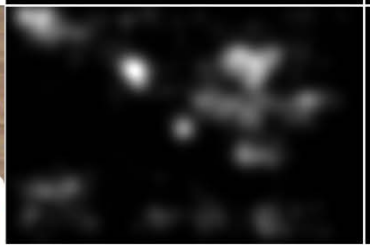

sAUC = 0.984

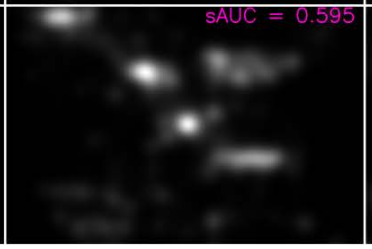
Wikipedia
Nederlandstalige Wikipedia

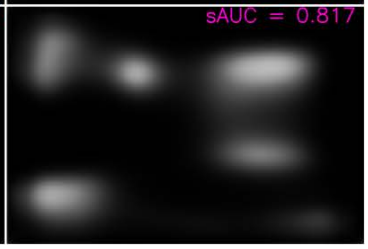

sAUC = 0.325


sAUC = 0.61

30% KORTING
30% korting op laminaat
Voordelig aan de slag!




sAUC = 0.595


sAUC = 0.817

Appendix B: Staring point and general scanpath

ET Heatmap Most viewed Point

MT Heatmap Most viewed Point

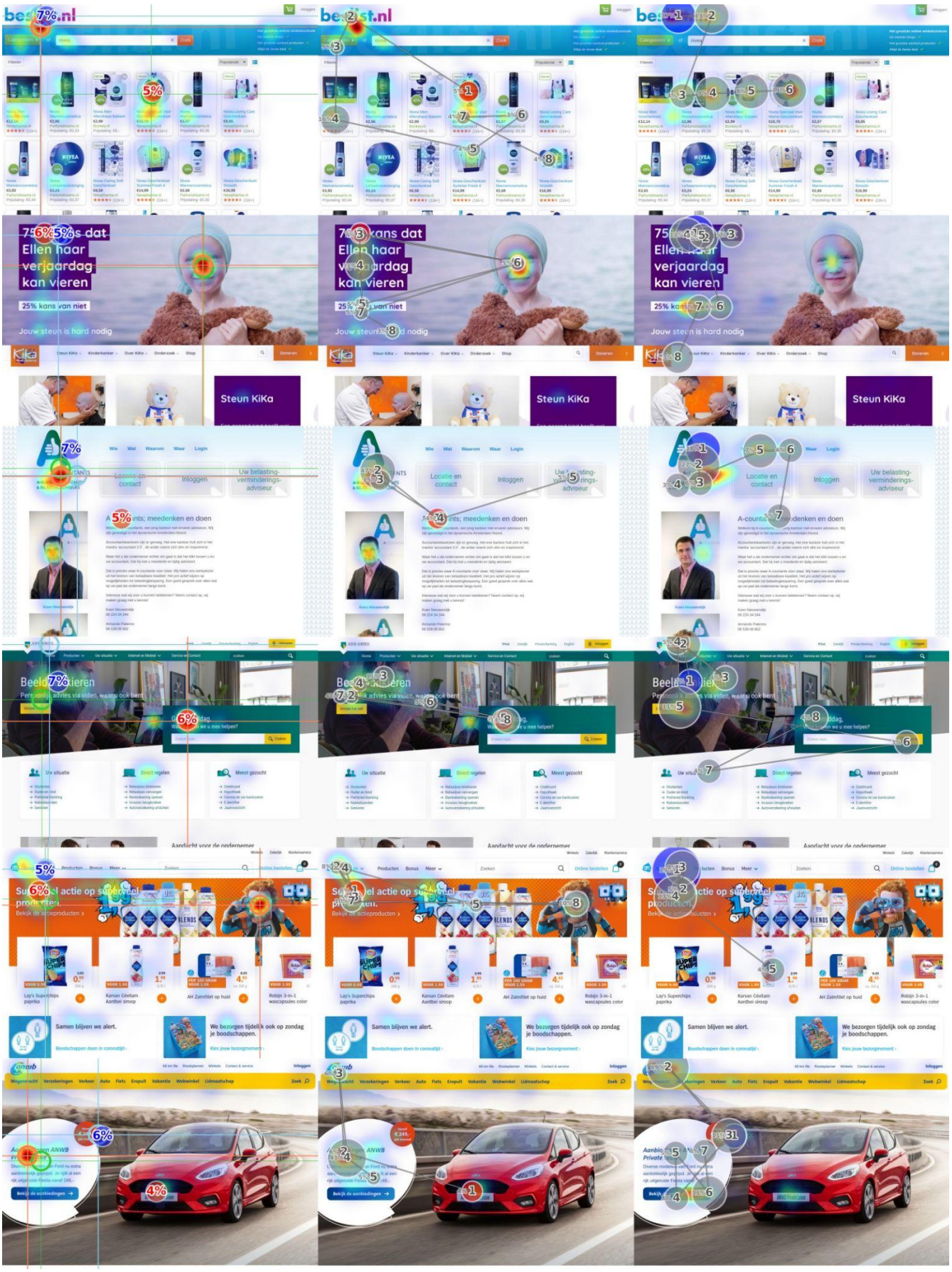
Expoze.io Saliency Map Most salient point

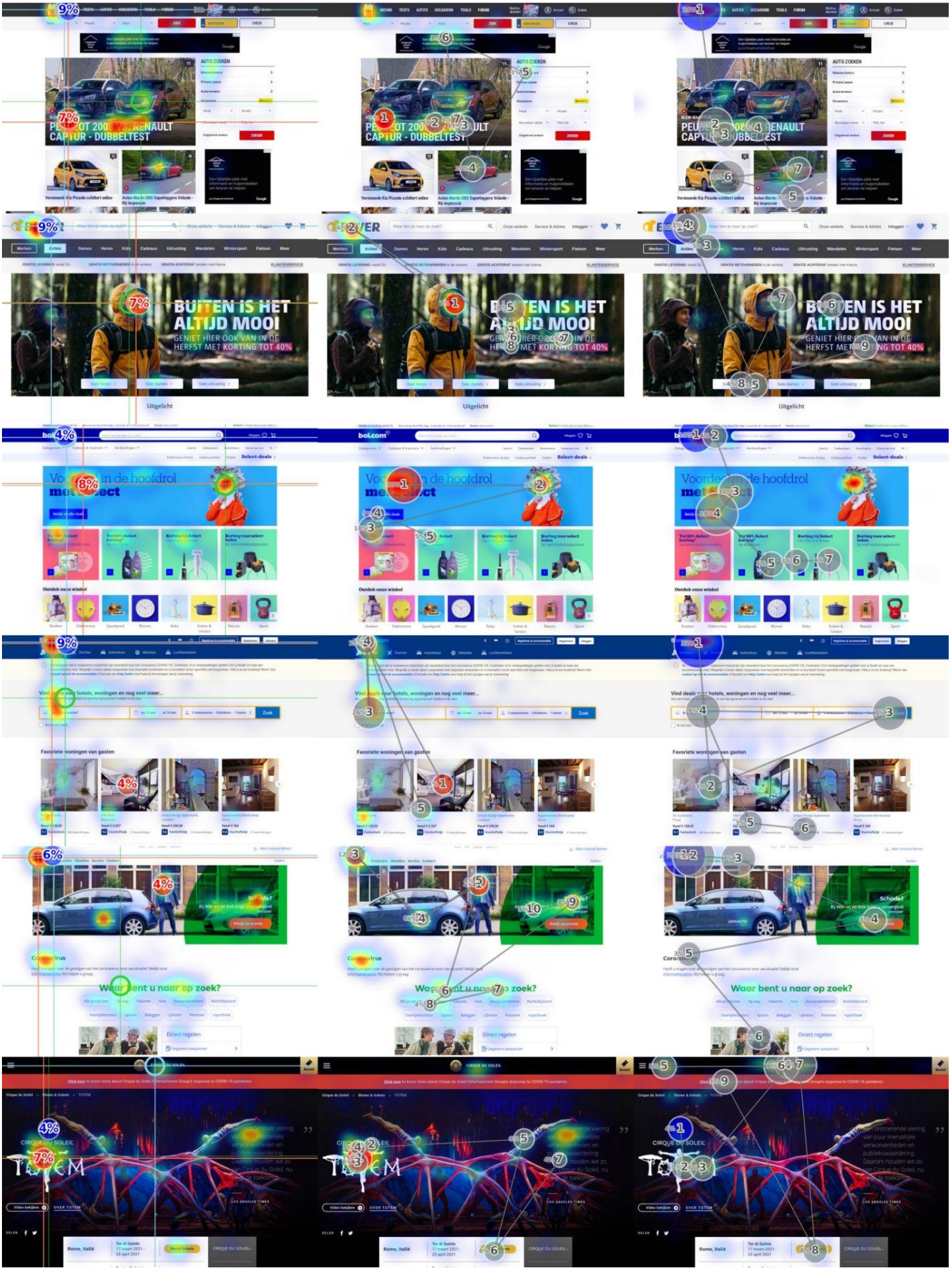
ET Heatmap Starting Point

MT Heatmap Starting Point

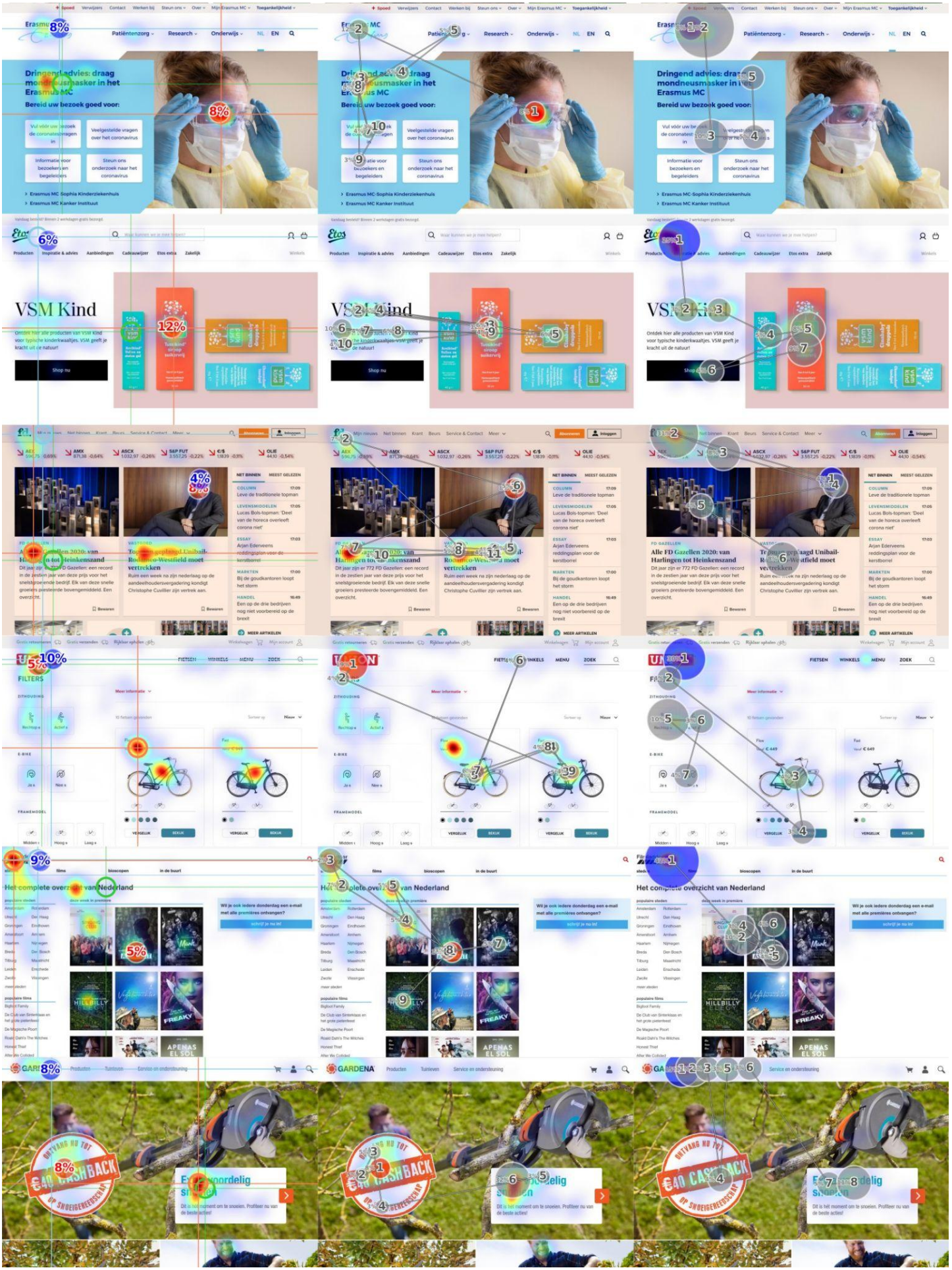
ET general scanpath
Participants = 25

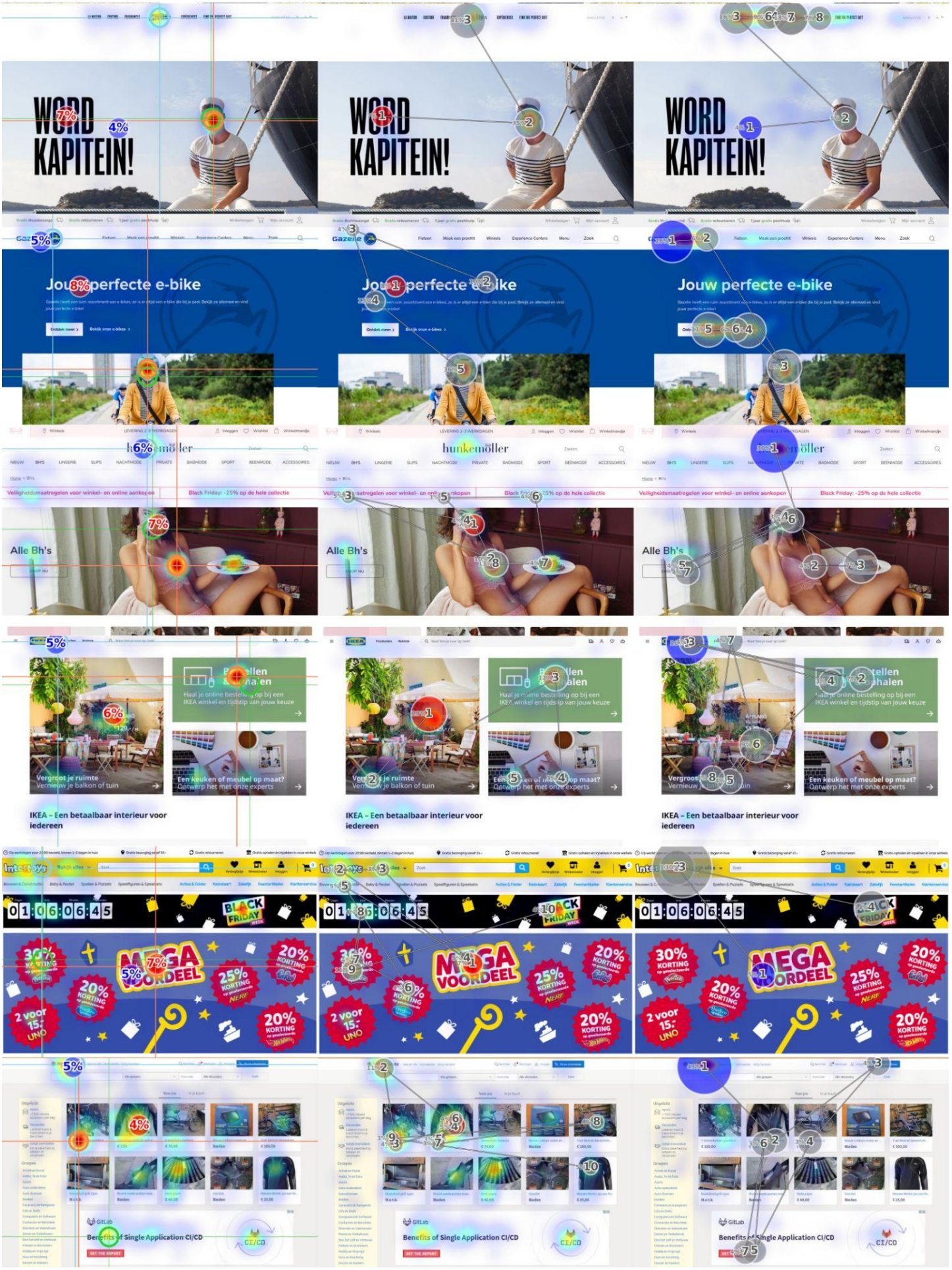
MT general scanpath
Participants = 27

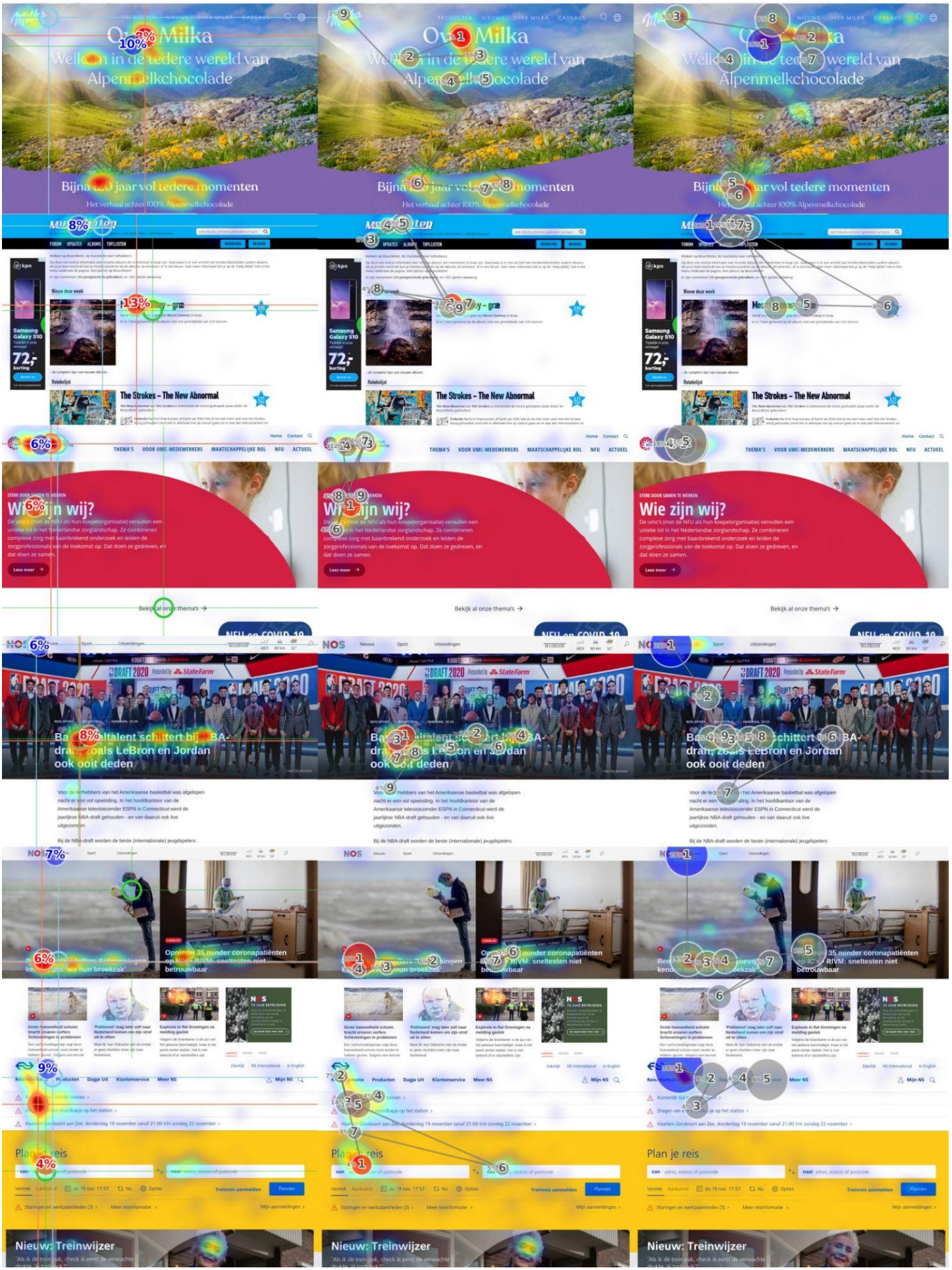



AUTO ZOEKEN
CAPTUR - DUBBELTEST
BUITEN IS HET ALTIJD MOOI
GENIET HIER OOK VAN IN DE HERFST MET KORTING TOT 40%
Uitgelicht
Select-deals
Favoriete woningen van gasten
Waar bent u naar op zoek?
CIRQUE DU SOLEIL
TOTEM


Blader door onze categorieën
Coronavirus
Reisvoucher verzilveren?
Bekijk 1946 resultaten
NIEUWE SPORTCOLLECTIE
KIES OP CATEGORIE
SPORTKLEDING
SPORTSCHOENEN
Laat zien wie je bent
Aanvragen of activeren
Aanvragen
Code invullen
Code ontvangen
Vul uw activeringscode, herstelcode, e-mailcode of machtigingscode in.
Reis zoeken
Daarom Djoser
Erasmus University Rotterdam
Onderwijs
Onderzoek
Over de EUR
Werken bij
Campus
Nieuws & Agenda
Contact
Make it happen


Dringend advies: draag mondneusmasker in het Erasmus MC
Bereid uw bezoek goed voor:
Vul vóór uw bezoek de coronatestvragen in
Veelgestelde vragen over het coronavirus
Informatie voor bezoekers en begeleiders
Steun ons onderzoek naar het coronavirus
Erasmus MC-Sophia Kinderziekenhuis
Erasmus MC Kanker Instituut
VSM Kind
Ontdek hier alle producten van VSM Kind voor typische kinderkwaaltjes. VSM geeft je kracht uit de natuur!
Shop nu
Alle FD Gazellen 2020: van Harlingen tot Heinkenszand
Dit jaar zijn er 772 FD Gazellen: een record in de zesden jaar van deze prijs voor het snelstgroeiende bedrijf. Elk van deze snelle groeiers presteerde bovengemiddeld. Een overzicht.
Rijnmond-Westfield moet vertrekken
Ruim een week na zijn nederlaag op de aandeelhoudersvergadering kondigt Christophe Cuvillier zijn vertrek aan.
FILTERS
FIETSEN
WINKELS
MENU
ZOEK
Het complete overzicht van Nederland
films
bioscopen
in de buurt
GARDENA
Producten
Tuinleven
Service en ondersteuning


WORD KAPITEIN!
Jouw perfecte e-bike
hunkemöller
Alle Bh's
Veiligheidsmaatregelen voor winkel- en online aankopen
Black Friday: -25% op de hele collectie
IKEA – Een betaalbaar interieur voor iedereen
Vergroot je ruimte
Een keuken of meubel op maat?
MEGA VOORDEEL
20% KORTING
25% KORTING
2 voor 15:-
30% KORTING
Benefits of Single Application CI/CD
GitLab
CI/CD


Over Milka
Welkom in de tedere wereld van Alpenmelkchocolade
Bijna 120 jaar vol tedere momenten
Het verhaal achter 100% Alpenmelkchocolade
Nieuw deze week
Samsung Galaxy S10
72,- korting
The Strokes – The New Abnormal
THEMA'S
VOOR UMC-MEDEWERKERS
MAATSCHAPPELIJKE ROL
NFU
ACTUEEL
STERK DOOR SAMEN TE WERKEN
Wie zijn wij?
De umc's (met de NFU als hun koepelorganisatie) vervullen een unieke rol in het Nederlandse zorglandschap. Ze combineren complexe zorg met baanbrekend onderzoek en leiden de zorgprofessionals van de toekomst op. Dat doen ze gedreven, en dat doen ze samen.
Lees meer
Bekijk al onze thema's
NOS
Sport
Uitzendingen
DRAFT 2020
Basketbaltalent schittert bij NBA-draft, zoals LeBron en Jordan ook ooit deden
Voor de liefhebbers van het Amerikaanse basketbal was afgelopen nacht er een vol opwinding. In het hoofdkantoor van de Amerikaanse televisiezender ESPN in Connecticut werd de jaarlijkse NBA-draft gehouden - en van daaruit ook live uitgezonden.
Bij de NBA-draft worden de beste (internationale) jeugdspelers
Opnieuw 35 minder coronapatiënten op IC's, RIVM: sneltesten niet betrouwbaar
Grote hoeveelheid schuim bracht ervaren surfers Scheveningen in problemen
'Politiemol' mag later zelf naar Nederland komen om zijn straf uit te zitten
Explosie in flat Groningen na melding gaslek
Plan je reis
Nieuw: Treinwijzer