Saliency-Aware Subtle Augmentation Improves Human Visual Search Performance in VR

Olga Lukashova-Sanz [1,*] and Siegfried Wahl [1,2]

1   Institute for Ophthalmic Research, University of Tübingen, 72076 Tübingen, Germany;
    siegfried.wahl@uni-tuebingen.de
2   Carl Zeiss Vision International GmbH, 73430 Aalen, Germany
*   Correspondence: olga.lukashova@uni-tuebingen.de; Tel.: +49-7071-29-83710

Abstract: Visual search becomes challenging when the time to find the target is limited. Here we focus on how performance in visual search can be improved via a subtle saliency-aware modulation of the scene. Specifically, we investigate whether blurring salient regions of the scene can improve participant's ability to find the target faster when the target is located in non-salient areas. A set of real-world omnidirectional images were displayed in virtual reality with a search target overlaid on the visual scene at a pseudorandom location. Participants performed a visual search task in three conditions defined by blur strength, where the task was to find the target as fast as possible. The mean search time, and the proportion of trials where participants failed to find the target, were compared across different conditions. Furthermore, the number and duration of fixations were evaluated. A significant effect of blur on behavioral and fixation metrics was found using linear mixed models. This study shows that it is possible to improve the performance by a saliency-aware subtle scene modulation in a challenging realistic visual search scenario. The current work provides an insight into potential visual augmentation designs aiming to improve user's performance in everyday visual search tasks.

Keywords: visual search; virtual reality; subtle visual augmentation; realistic visual scenes

1. Introduction

Visual search is one of the most common tasks in everyday life, be it when a person is looking for a friend in a crowd or when a doctor is analyzing an optical coherence tomography (OCT) scan from a patient [1]. Search becomes more challenging when the time to find the target is limited. For example, when a person is searching for the keys right before leaving the house, or when a surgeon is performing a meticulous manual task during the surgery using digital surgical microscope images [2]. In this study we focus on how performance in a visual search under limited time conditions can be improved.

An extensive amount of research has been done to investigate how people are searching for a target among distractors, and which neural mechanisms are laying behind, where numerous search task paradigms have been implemented (for reviews, see [3–5]). The difficulty of the visual search task depends on various factors, including how similar the target and the background are, how distinct the target is from the distractors, how complex the scene is, whether the observer has seen the scene already before, and many other aspects [3–5]. The human capacity to process visual content is limited, and mainly, in complex searches, it is crucial to select and prioritize visual information to complete the task. Attention is one mechanism that supports visual search and enables the searcher to find the target more efficiently [6–8]. The shift of attention is associated with eye movements and fixating attended locations [5,9]. Among numerous aspects impacting guiding of attention in visual search is stimulus-driven saliency of different elements of the visual scene. Specifically, the observer's attention can be attracted by salient distractors, even though from a goal-oriented perspective, they are irrelevant. To which extent saliency

plays a role when it comes to visual search strategy has been long debated. Some studies showed that top-down mechanisms primarily drive visual search strategy, where fixation density can be explained by saliency only for the first few fixations [10–12].

Other studies demonstrated that salient goal-irrelevant distractors could attract the observer's attention, slowing down the search [13–16]. The large variability of the results reported in the literature supports the notion that a combination of factors affects human attention guiding when looking for a target [17]. Here we approach visual search assuming that salient regions can attract attention in visual search.

When it comes to saliency, there are many different definitions that one meets in the literature. A salient region, in its broad context, is an area of the visual scene. It has a high contrast with its surroundings in one or multiple feature dimensions, be it color, shape, spatial frequency, speed of motion, contextual meaning, location within the visual field of view. Even though saliency is often associated with bottom-up low-level feature contrast [18,19], both aspects are connected, but not identical [20]. Therefore, a part of a scene is considered salient if it is likely to attract the observer's attention, which is usually accompanied by fixating on that region.

The vast majority of existing knowledge on visual search is based on experimental studies conducted on conventional 2D-screens often using not-realistic synthetic stimuli as search arrays in a controlled environment. The search behavior in such artificial conditions can differ from real-life scenarios. A group of studies investigated human's visual behavior in a search task using naturalistic scenes displayed in a 2D-screen (e.g., [12,21–25]). The fast development of modern technologies such as Virtual Reality (VR) and VR eye tracking enabled researchers to study visual search in more realistic 3D environments [26,27]. In order to increase the level of immersion of the experimental paradigm, in the current study, we focus on real-world static visual scenes displayed in virtual reality, ensuring free body and head movement.

New technological tools, such as augmented reality (AR), enable the use of additional visual cues to purposefully guide user's attention and improve user's performance [28–30]. On the other hand, one drawback of augmenting visual input with additional content is that it introduces a trade-off of the potential benefit between performance and overlaying real visual scene with an additional layer of information. That, in turn, captures attentional resources, which essentially becomes a bottleneck for the design of visual augmentation [31]. In this study, we hypothesize that by subtly modifying the visual scene, it is possible to drive the observer's attention away from salient distracting locations, enabling the user to find the target faster. Previously, some attempts were made to apply subtle visual content modification for gaze guidance, where color, luminance, spatial frequency, and other domains were modulated [32–41]. In this work, blur was selected as a domain for modifying visual content, as it was previously shown that blur, although with limitations, can be used for gaze guidance to an extent where the observer does not even notice the modification [42–44].

Furthermore, the idea of slightly defocusing parts of the scene for triggering the observer to fixate on more clear locations is widely used in photography and cinematography [45–47]. Also, Sitzmann et al. [48] proposed blurring images based on saliency to downsample the resolution of the non-salient regions for further image compression. In contrast to their approach, the strongest blur was applied to the most salient regions in this study.

This study investigates whether blurring salient regions of the visual scene, which would otherwise likely attract the observer's attention, can improve visual search task's performance, enabling the observer to find the target faster. Using a psychophysical approach, we evaluate observers' ability to locate the target within a limited amount of time. Eye-tracking data was recorded to support the results. Specifically, fixations were analyzed.

2. Materials and Methods

2.1. Participants

Twenty naive participants (14 female and 6 male), with normal or corrected to normal vision were tested. Participants were aged between 20 and 38 years old. The study was conducted according to the guidelines of the Declaration of Helsinki. The study was approved by the ethics committee of the Faculty of Medicine at the University of Tübingen with a corresponding ethical approval identification code 138/2017b02. Signed informed consent was obtained from each subject before the measurements. All data were stored and analyzed in full compliance with the principles of the Data Protection Act GDPR 2016/679 of the European Union.

2.2. Experimental Setup

2.2.1. Hardware Specifications

The visual content was displayed to the participant using HTC Vive Pro Eye (HTC Corporation, Taoyuan, Taiwan) virtual reality headset running on a Windows 10 PC with NVIDIA GeForce GTX 1070 graphics card (NVIDIA Corporation, Santa Clara, CA, USA). The field of view of the headset and the refresh rate reported by the manufacturer are 110° and 90 Hz, respectively. The participant interacted with the environment via the HTC Vive controller. The position and rotation of the headset and the controller were tracked via the HTC base stations 2.0. The eye-tracking data was collected using a built-in eye tracker at a frequency of 120 Hz.

2.2.2. Software Specifications

The experimental paradigm was generated using the Unity Game engine [49], Unity version 2019.3.15.f1. The eye movement data was collected using Unity package SRanipal version 1.1.0.1. Recording of the eye movement data at a maximum sampling rate 120 Hz was realized by means of using a separate thread parallel to the main script execution [50]. The data analysis was performed using Python 3.6 packages *NumPy* [51] version 1.19.1, *SciPy* [52] version 1.5.2 and *Pandas* [53] version 1.1.3. The statistical analysis was conducted using R [54] version 3.6.1, in particular, package *lme*4 [55]. The data visualization was performed using Python packages *Matplotlib* version 3.3.1 [56] and *Seaborn* [57] version 0.11.0.

2.3. Virtual Environment and Stimuli

2.3.1. Real-World Scenes

The virtual environment was composed of omnidirectional images displayed in virtual reality (VR) by back-projecting it to the Skybox sphere (Figure 1C). An omnidirectional image is a 360-degree panoramic image. An equirectangular projection is one way to represent an omnidirectional image, where the aspect ratio of the projection is 2:1. The horizontal and vertical coordinates of the projection are polar (ϕ) and azimuthal (θ) angles, respectively, where ϕ can be in the range from 0° to 360°, and θ ranges from 0° to 180° (Figure 1A). For the main experiment, 24 scenes were selected from the Salient360! Training dataset with diverse content: indoor/outdoor, day/night time, containing people/not containing people, etc. (see full set of images in Supplementary Information Figures S3 and S4). The resolution of the images ranged from 3000×1500 to $10,000 \times 5000$ pixels with mean resolution of $(5878 \pm 2443) \times (2939 \pm 1222)$ pixels. For the training phase, five scenes from the Salient360! Training dataset were selected, different from the scenes of the main experiment (see Supplementary Information Figure S2).

A)

B)

C)
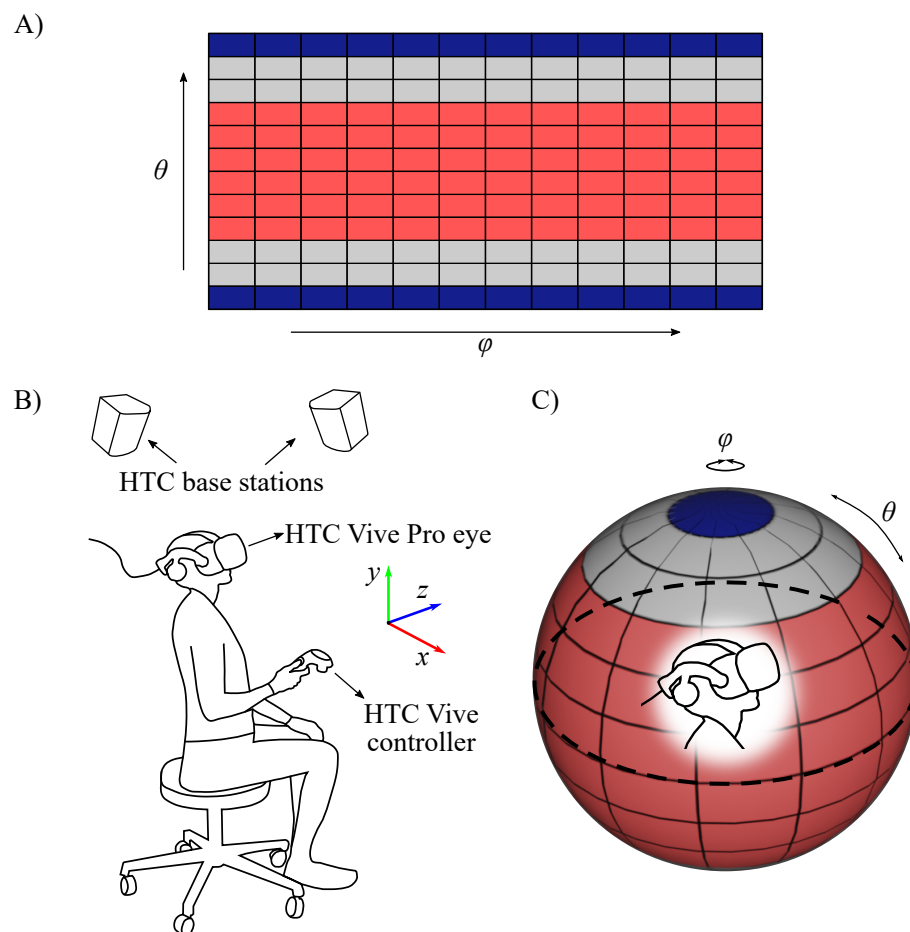
Figure 1. (**A**) A schematic representation of equirectangular projection. Each rectangle is 30° wide and 15° high. (**B**) The experimental setup. The experimental paradigm was implemented in Unity software and displayed in a VR headset—HTC Vive Pro Eye with a built-in eye tracker. Position and rotation tracking of the headset was realized via the HTC base stations 2.0. The participant performed the experiment in a seated position on a rotating stool. As soon as the search target was located, the participant pressed a button on the HTC Vive controller. Additionally, a left-hand coordinate system is shown, which is used to set the Unity world coordinates. For details on the experimental procedure see Section 2.4.1. (**C**) A schematic representation of a sphere to which the equirectangular projections (omnidirectional images) were back-projected in Unity as a Skybox. The colors indicate like-colored locations on the equirectangular projection in (**A**).

2.3.2. Saliency Maps

To evaluate which regions of the scenes are likely to attract attention, saliency prediction models are widely used [58]. In recent years interest in saliency models for 360-degree omnidirectional images largely grew [59]. In the current study, the saliency maps used for spatial modulation of the omnidirectional images were obtained using the method described in [60] which won in the "Head and Eye Movement Prediction" category of "Salient360!" Grand Challenge at ICME'2017 [61]. Authors of the method proposed composition of the continuity-aware and the cube map approaches using a combination of saliency predictors, with applied equator bias (for details, see [60]). The gray-scale saliency maps used for the main experiment can be found in Supplementary Information Figures S5 and S6. The pixel color values of the saliency maps ranging from 0 (black) to 1 (white) are referred to as saliency values. In Figure 2 an example of one scene together with its saliency map is shown.
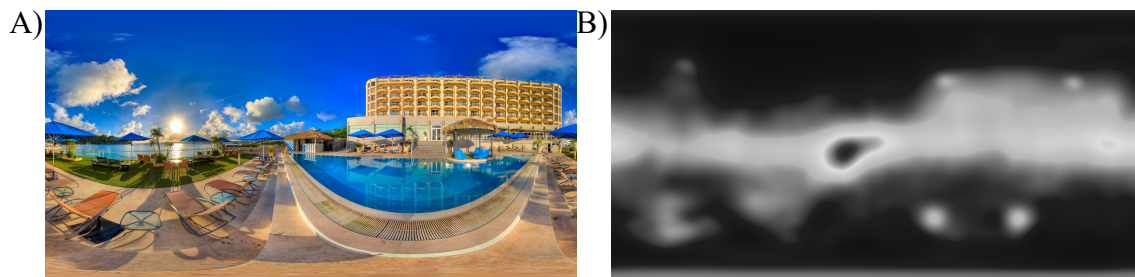
Figure 2. (**A**) Example of one omnidirectional image used in the main experiment. (**B**) The corresponding saliency map generated using the approach in [60]. White regions correspond to the most salient areas, whereas black color indicates the least salient locations.

2.3.3. Blurred Images

The omnidirectional images were modified by applying blur. The strength of blur was spatially varied based on the corresponding saliency map. In particular, each blurred image resulted from a convolution of the respective original image and a two-dimensional Gaussian kernel. The kernel's size was fixed to one degree, whereas the standard deviation σ of the kernel was varied, determining the blur strength. A set of blurred 360-degree images using two different values of the Gaussian kernel standard deviation was generated. In particular, $\sigma_1 = 17\%$ of kernel size, and $\sigma_2 = 34\%$ of the kernel size, were used defining two of the experimental conditions (Figure 3). At a given experimental condition, the blur strength for each pixel was weighted with its saliency value. Specifically, the standard deviation of the Gaussian kernel for each pixel was computed by multiplying the standard deviation corresponding to the experimental condition (σ_1 or σ_2) with the respective saliency value of that pixel (values from 0 to 1, see Section 2.3.2). The size of the kernel in pixels, as well as the standard deviation, were scaled considering the individual image size in pixels. An example of a blurred part of an image is shown in Figure 3.
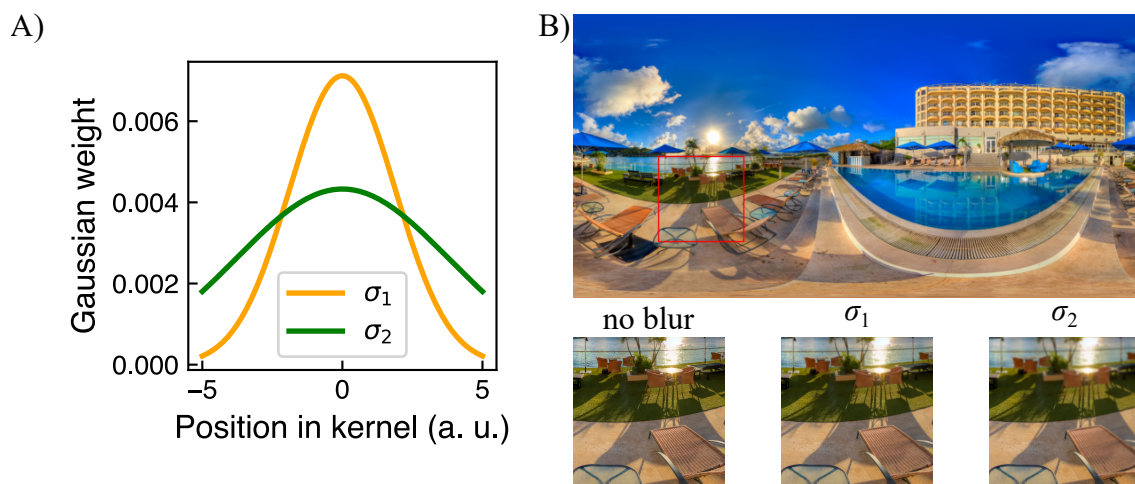

Figure 3. (**A**) A cross-section of a 2D Gaussian kernel used to blur images, with two standard deviations: $\sigma_1 = 17\%$ of kernel size, and $\sigma_2 = 34\%$ of the kernel size. The kernel size was fixed to a number of pixels corresponding to 1° visual angle. For schematic illustration, the kernel size in the plot equals 10 a. u. (**B**) Example of blurred part of one scene generated as a convolution of respective 2D Gaussian kernel and original image, considering its saliency map. In the lower row of (**B**), a part of the original image which is indicated by a red rectangular is shown in three blur conditions: no blur, Gaussian blur with maximum standard deviation σ_1, and Gaussian blur with maximum standard deviation σ_2.

2.3.4. Search Target

As a search target, a *Gabor cross* was used. Similar to [12], a Gabor cross is a sum of vertical and horizontal Gabor stimuli, where each of the Gabors is a product of 8 cyc/° cosine function with Gaussian envelope with standard deviations of 0.06° and 0.32° (Figure 4).

The total width and height of the cross was approximately 3° visual angle. The target was overlaid with the visual scene by positioning it in Unity at a pseudorandom location on top of the omnidirectional image (see Section 2.3.5). To blend the target with the omnidirectional image, the transparency of the cross was decreasing towards the edges of the Gabor cross at a rate of the Gaussian envelope used to generate the cross.
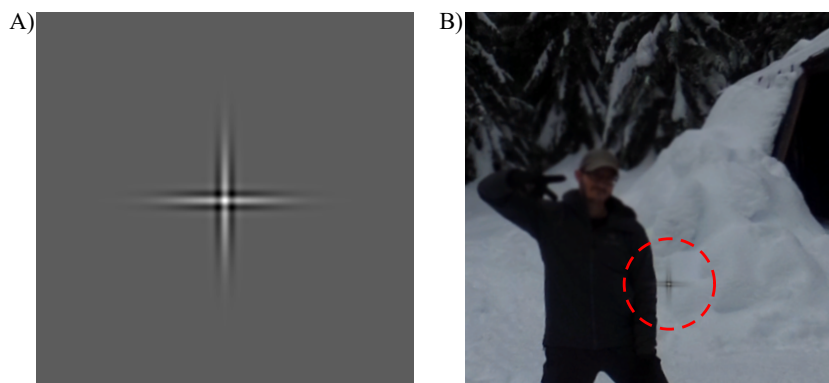
Figure 4. (**A**) The Gabor cross was used as a search target in the visual search task. It is a sum of vertical and horizontal Gabor stimuli, where each of the Gabors is a product of 8 cyc/° cosine function with Gaussian envelope with standard deviations of 0.06° and 0.32°. (**B**) An example of the search target is overlaid with a visual scene at a pseudorandom location. For simplicity, the cross is highlighted by a surrounding red dashed line. In the experiment, the red line was not present.

2.3.5. Search Target Locations

In each trial, the search target was positioned at a pseudorandom location defined by spherical coordinates (r, ϕ, θ). Specifically, r was set to a fixed value, whereas ϕ and θ were varied. Prior to the experiment, a set of nine possible target locations (ϕ, θ) was generated for each visual scene, where ϕ could be in the range $[0°, 135°]$ and $[225°, 360°]$, and θ in the range $[45°, 135°]$. This range was selected to ensure that subjects do not have to rotate their head in an uncomfortable position trying to look too much up or down. Additionally, taking into account that in two experimental conditions the visual scenes were spatially blurred based on the corresponding saliency values, the regions with low saliency were selected as possible search target locations to avoid displaying the target on very differently blurred backgrounds in different experimental conditions. Specifically, possible (ϕ, θ) were limited to regions with low saliency for each visual scene—locations with saliency values under 17% of maximum saliency value on corresponding saliency map. During the experiment, each participant had the same set of pseudorandom locations for corresponding visual scenes and respective conditions. The size of the search target was was kept constant in visual angle in all visual scenes. The distribution of the final set of possible search target locations for all visual scenes can be seen in Supplementary Information Figure S1.

2.4. Experimental Procedure

2.4.1. General Procedure

Each participant performed a visual search task wherein each trial they had to search for the search target located at a pseudorandom location within a limited amount of time set to 20 s. Participants were instructed to find the target as fast as possible, naturally moving their head and gaze. Once the target was found, participants were asked to fixate the target and press a button on the controller. As soon as the button was pressed, the trial was terminated. Participants performed the experiment in a seated position on a rotating stool enabling free head movement (Figure 1B). The experiment was performed in a single one-and-a-half-hour experimental session. First, each participant had a training phase where they got acquainted with the virtual environment and the search target. Next, the main part of the experiment was conducted. Every time the participant put on the headset, a five-

point built-in calibration procedure of the eye tracker was performed. Each scene was explored starting from the same spatial point $\phi = 180°, \theta = 90°$. To do so, before starting a new trial, participants had to fixate a reference fixation point of size 1° visual angle on a gray background located in the corresponding direction. The gaze position was controlled using the eye tracker. Once the gaze position condition was satisfied, the trial was started. That is, the scene with the search target was onset.

2.4.2. Training Phase

The training phase consisted of 10 trials with no time limit. In each trial, one of five different visual scenes was shown with the search target overlaid on top of the scene. Thus, each scene was presented twice. The scenes used in the training phase were different from the ones used in the main experiment (see Supplementary Information Figure S2). By the end of the training phase, all participants verbally responded that the task was straightforward and that they felt comfortable to continue with the main experiment.

2.4.3. Main Experiment

In the main experiment, participants were instructed to perform the search task the same as during the training phase. Furthermore, they were informed that some scenes are partly blurred, but that it does not affect the task instructions. Participants were also informed that each trial's time is limited to 20 s, but that in some trials, it is too difficult to find the target in the background. Thus, it is normal not to find the target within the given time in some trials. Each trial finished pressing a button if the target was found or after the time limit was reached. The fixation on the target was not controlled during the trial. The main experiment was split into three blocks with short three-to-five minutes breaks. During each break, the participant took off the headset and rested. Before starting each next block, the eye tracker calibration procedure was conducted. Each block consisted of 72 trials. All 24 scenes were presented in each block three times: as an original image with no blur, blurred with standard deviation σ_1, and blurred with standard deviation σ_2. The blur strength determined three experimental conditions: Condition 0 (no blur), Condition 1 (blurred with σ_1), and Condition 2 (blurred with σ_2). During each block, all three conditions were presented in random order. Each block's duration varied depending on how long it took the participant to find the target in each trial, ranged between 10 and 20 min. During the main experiment, each participant performed a total of 216 trials divided into three blocks, with each block containing the three conditions above-mentioned for every scene (24). For each scene, the search target location was always different selected from the set of nine possible locations (see Section 2.3.5).

2.5. Analysis

The behavioral data as well as the eye movement data were analyzed using Python as well as *lme*4 library of R, and compared using linear mixed model analysis [62].

2.5.1. Behavioral Performance Metrics

The *search time* and the *proportion of missed trials* were selected as the main behavioral performance metrics. The search time is defined as the time since a trial started until the button was pressed. Thus, it is a continuous variable ranging between 0 ms and 20,000 ms. Only trials where the target was found were used to compute the search time. The proportion of missed trials is computed as the number of trials where the target was not found within 20 s divided by the total amount of trials in the respective experimental condition. As for each condition, there were 72 trials from each participant. The proportion of missed trials is a discrete variable with a minimum step of $1/72 = 0.014$. Better performance is defined by a shorter search time as well as a lower proportion of missed trials. The impact of blur was estimated by fitting the linear mixed models to the data, fitting a separate model for each of the metrics, where *search time* and *proportion of missed trials* are dependent variables, *blur condition* is fixed effects, and *subject* is a random

factor. Furthermore, to evaluate learning throughout the experiment, the *block number* was introduced as an additional fixed effect in the model. To further explore whether the differences between blur conditions evolve with time, an extension of the model could be done where one could account for possibly different slopes in data subgroups at a given blur condition and given block number. However, the limited amount of data points collected in this study does not support a more complex model. Therefore, a proposed linear mixed model with two fixed effects and one random factor was selected. An extensive quantitative analysis of performance evolution over the course of time is out of the scope of the present work and is a subject of future studies.

2.5.2. Eye Movement Metrics

As a secondary set of metrics to characterize visual search performance, the number of fixations until the target was found (*NumOfFix*), the proportion of fixations within the area of interest (*PropFixInAOI*), and duration of fixations (*FixDuration*) were computed. The area of interest (AOI) is defined, similar as in [48], as 5% most salient pixels of the original image based on its saliency map (see example in Figure 5). The proportion of fixations within the area of interest is defined as the number of fixations within the area of interest divided by the total number of fixations for each trial. The number of fixations until the target was found was defined as the total number of fixations in each trial where the target was found. Finally, the fixation duration was computed as the duration of fixations in each trial. The effect of blur condition was evaluated for each metric by fitting a separate linear mixed model where *NumOfFix*, *PropFixInAOI*, and *FixDuration* are dependent variables, *blur condition* is fixed effect, and *subject* is a random factor. To account for possible learning throughout the experiment, the *block number* was introduced as an additional fixed effect in the model.
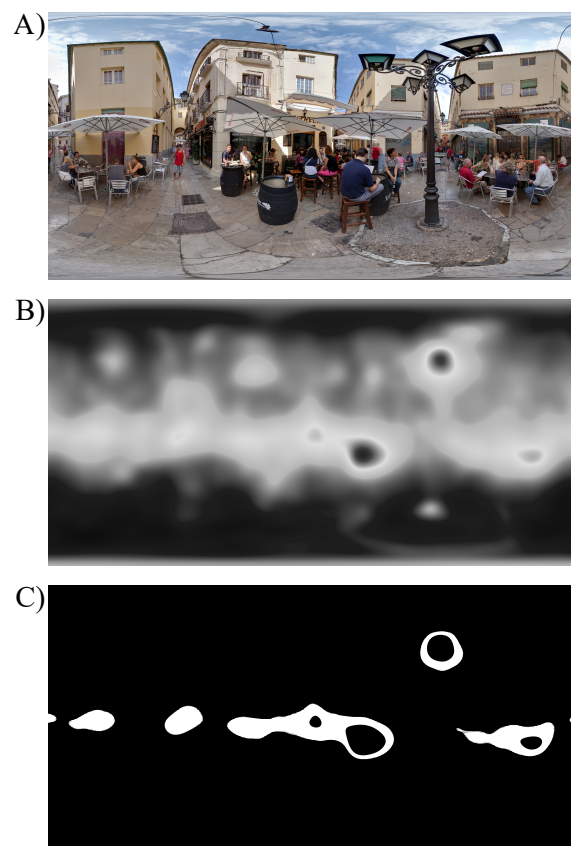
Figure 5. (**A**) Example of one omnidirectional image, (**B**) its saliency map in gray scale, and (**C**) selected 5% most salient pixels indicated by white color. The area of interest is the white region in (**C**).

2.5.3. Eye Movement Raw Data Pre-Processing

The eye movement data were recorded at a frequency of 120 Hz. The gaze position data was accessed using a customized written Unity script utilizing the HTC SRanipal SDK package functions [63]. The time variable was taken using the system time, as it was previously shown that the time function from SRanipal package returns not always reliable values [64]. In Table 1 the main recorded variables are described. All variables were recorded for left and right eyes.

Table 1. Main eye- and head-movement-related raw variables recorded during the experiment.

Variable	Units	Meaning
Timestamp	any integer number	The system time in ms at the moment of sample recording.
Eye data validity bit mask	an integer from 0 to 31	Indicates the validity of the data. A value of 31 indicates the highest validity of the recorded data. This parameter is used to filter the raw data where the eye tracker lost the pupil, including filtering blinks.
Gaze normalized direction vector	A three-coordinates vector (x, y, z) with each coordinate ranging from −1 to 1	A gaze vector indicating the direction of gaze in the headset right-hand coordinate system. To convert it to the left-hand coordinate system (Figure 1B), the x coordinate was multiplied by −1.
Head rotation	a rotation quaternion (x, y, z, w) of head	A quaternion describing the rotation of the headset in Unity world coordinates. The position of the headset was always fixed to the origin (0, 0, 0).

To prepare the data for further processing, first, similar to [64], the raw data were filtered based on the eye data validity bit mask value, which represents the bits containing all validity for the current frame. After the filtering, only the data where the eye data validity bit mask had value 31 for both eyes, were selected. Doing so, the data where the eye tracker partly or completely lost the pupil (including blinks) was filtered out. Next, the gaze position was calculated in spherical coordinates. In particular, the polar ϕ and azimuthal θ angles were computed using Equations (1) and (2). In Unity, the z-axis corresponds to the depth dimension.

$$\phi = \arctan \frac{x}{z}, \tag{1}$$

$$\theta = \arctan2\left(y, \sqrt{x^2 + z^2}\right), \tag{2}$$

where (x, y, z) are coordinates of normalized gaze directional vector in headset coordinates. Note that SRanipal returns the gaze direction vector in the right-handed coordinate system. To convert the coordinates in the left-hand coordinate system (same as Unity world coordinate system, see Figure 1B), the x-coordinate was multiplied by −1. To compute the gaze position in Unity world coordinate system, the gaze position in headset coordinate system was multiplied by the head rotation quaternion. In Figure 6 an example of gaze position for one subject in one trial in spherical coordinates (ϕ, θ) is shown.

### 2.5.4. Fixation Detection Algorithm—I-VT

Fixations were identified using *velocity threshold algorithm for fixation identification* (I-VT) [65]. The algorithm was implemented following the description in [66,67]. The gaze velocity v was computed in °/s between each two consecutive samples (Equation (3)).

$$v = \frac{\sqrt{(\phi_i - \phi_{i-1})^2 + (\theta_i - \theta_{i-1})^2}}{t_i - t_{i-1}}, \tag{3}$$

where (ϕ_i, θ_i) and $(\phi_{i-1}, \theta_{i-1})$ are consecutive gaze positions in degrees visual angle in headset coordinates, and t_i and t_{i-1} are respective time stamps. To reduce the noise level of the data, a running average filter was applied with the window size of three samples which is ~25 ms. An eye movement was considered to be a fixation if the gaze velocity did not exceed a threshold 60 °/s [68]. Two fixations were merged in a single fixation if the time between them was under 75 ms [69], and the angular distance was under 1° [69,70]. Too short fixation with a duration under 60 ms were filtered out [69,70]. In Figure 7 the eye movement data processing algorithm is summarized in a flow chart.
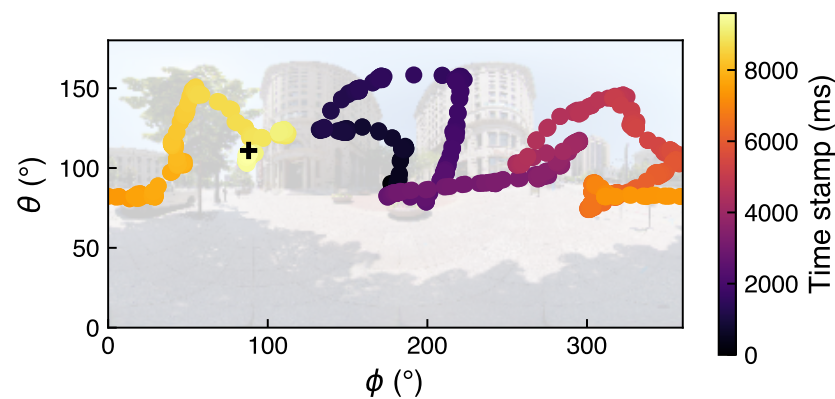
Figure 6. Example of raw gaze position in spherical coordinates for one subject in one trial. The background image is the visual scene shown in this particular trial—for easier visualization of the gaze positions, the omnidirectional image is washed out in the figure. The black cross with coordinates $(88°, 111°)$ indicates position of the search target in this trial. The color of the scatter points range from dark violet to light yellow and indicates the time stamp of each gaze position sample starting from the beginning of the trial: here, the participant started the search from the center of the image and ended around the target position. This trial lasted approximately ten seconds and the search target was successfully found by the participant.
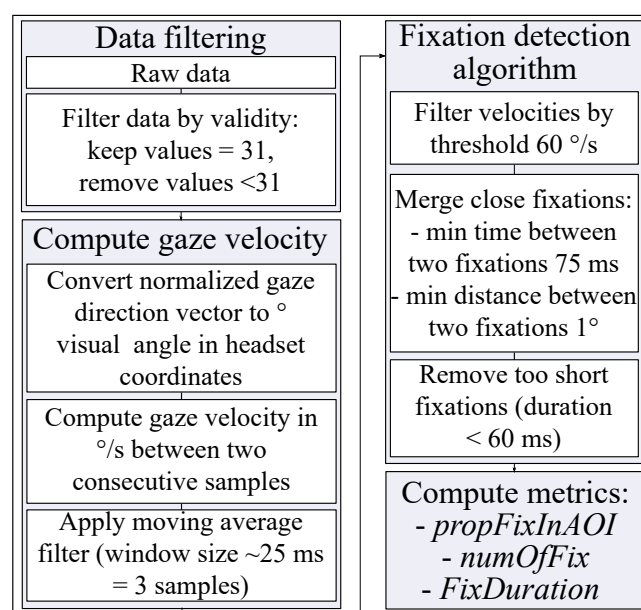

Figure 7. Eye movement data processing algorithm. For details see Section 2.5.4.

3. Results

3.1. Behavioral Data

In Figure 8 the mean search time and the proportion of trials where the target was not found, estimated across all subjects, are shown. The mean value of the search time considering trials where the target was found was $8423\,\text{ms} \pm 1358\,\text{ms}$, $8146\,\text{ms} \pm 1258\,\text{ms}$, and $7670\,\text{ms} \pm 1550\,\text{ms}$ for conditions 0, 1, and 2, respectively. The mean value of the proportion of trials where the target was not found, was 0.17, 0.11, and 0.10 for conditions 0, 1, and 2, respectively. The individual data for each subject can be found in Supplementary Information Section S5. From linear mixed model analysis (see Section 2.5.1) over the course of all trials, a significant effect of blur was found for both behavioral metrics. Specifically, for the search time, a significant difference between the no-blur and σ_2-blur conditions was found with $p < 0.001$. The difference between the no-blur and σ_1-blur conditions was close to but not significant with $p = 0.09$. In terms of the proportion of trials where the target was not found, significant differences between the no-blur and σ_1-blur, as well as between no-blur and σ_2-blur conditions, were found with $p < 0.001$ for both conditions.

To check for the learning effect, the mean search time and proportion of missed trials were evaluated over the course of the experiment (see Supplementary Information Section S7, Figure S12A,B). A significant effect of block number was found for mean search time as well as for proportion of missed trials with $p < 0.001$. This can also be observed as a downtrend of both behavioral metrics for all three blur conditions indicating a general learning of the task. This is expected, as with time participants get more familiar with the VR headset, as well as get more acquainted with the visual scenes since the same images were presented in each experimental block. Despite general learning, the difference in performance metrics between the no-blur condition and other two blur conditions was observed already in the first block of the experiment, most prominently, for the proportion of missed trials.
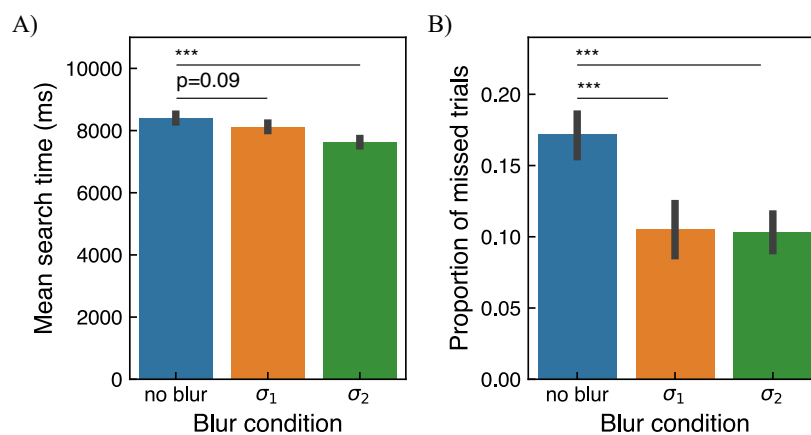

Figure 8. (**A**) An estimate of mean search time in each blur condition. The mean search time was computed for each subject by averaging the search time over the number of trials where the target was found. (**B**) An estimate of the proportion of trials where the target was not found in each blur condition. The bar plots represent data from 20 subjects. The error bars show the standard error of the mean. The indicators of significant differences obtained from the linear mixed model analysis are *: $p < 0.05$, **: $p < 0.005$, ***: $p < 0.001$. For non-significant differences, the p-value is shown.

3.2. Eye Movement Data

In Figure 9A distribution of fixation duration is visualized as a kernel density estimate plot which is an alternative to a histogram plot. The plot represents all fixations computed for all subjects and all trials regardless of whether the target was found or not in a respective trial. The total number of detected fixations was 21,504, 18,059, and 16,092 for the Conditions 0 (no blur), 1 (σ_1), and 2 (σ_2), respectively. The mean duration of fixation averaged over all subjects across all trials is $133\,\text{ms} \pm 76\,\text{ms}$. This value is comparable with

typically reported fixation duration considering that in a visual search task, fixations are usually shorter than in a free-viewing task [71]. In Figure 9B an estimate of the number of fixations until the target was found (*NumOfFix*) is shown. To estimate *NumOfFix* only trials where the target was found were considered. The mean values of *NumOfFix* are 11.9, 10.8, and 9.6 for Conditions 0, 1, and 2, respectively. Figure 9C demonstrates an estimate of the proportion of fixations within the area of interest (*PropFixInAOI*). To estimate *PropFixInAOI*, all trials were considered regardless of whether the target was found or not in the respective trial. The mean values of *PropFixInAOI* are 0.23, 0.20, and 0.19 for Conditions 0, 1, and 2, respectively. Finally, in Figure 9D an estimate of fixation duration (*FixDuration*) is shown. To compute the estimate of *FixDuration* all fixations were considered regardless of whether the target was found or not and whether a fixation was within AOI or not. See details of metrics' definitions in Section 2.5.2. The individual data for each subject can be found in Supplementary Information Section S6.
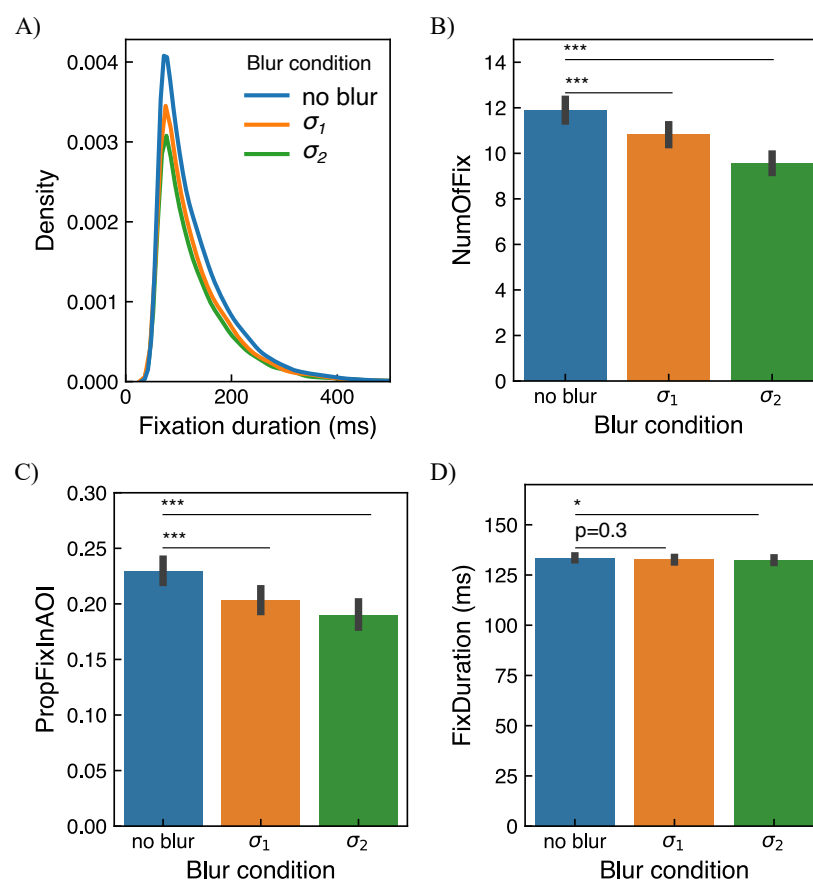
Figure 9. (**A**) A distribution of fixation duration as a kernel density estimate plot. Only fixations with duration up to 500 ms are shown to better resolve differences between different blur conditions. The mean duration of fixation averaged over all subjects across all trials is 133 ms ± 76 ms. The plot represents all fixations computed for all subjects and all trials. Each curve is normalized to the number of observations such that the total area under all densities sums to 1. (**B**) An estimate of the number of fixations until the target was found in each blur condition. The bar plots represent data from 20 subjects considering only trials where the target was found. (**C**) An estimate of the proportion of fixations within the area of interest in each blur condition. The bar plots represent data from 20 subjects considering all trials. (**D**) An estimate of fixation duration in each blur condition. The bar plots represent data from 20 subjects considering all trials. The error bars show the standard error of the mean. The indicators of significant differences obtained from the linear mixed model analysis are *: $p < 0.05$, **: $p < 0.005$, ***: $p < 0.001$. For non-significant differences, the *p*-value is shown.

From the linear mixed model analysis (see Section 2.5.2) over the course of individual trials and fixations, a significant effect of blur was found for $NumOfFix$ and $PropFixInAOI$ for all blur conditions. Specifically, for both metrics ($NumOfFix$ and $PropFixInAOI$), a significant difference between the no-blur and σ_1-blur, as well as between the no-blur and σ_2-blur conditions, were found with $p < 0.001$ for both conditions. For $FixDuration$ a significant difference was found between the no-blur and σ_2-blur conditions with $p < 0.05$. The $FixDuration$ difference between the no-blur and σ_1-blur conditions was not significant with $p = 0.3$. The differences in $FixDuration$ only for fixation within AOI were not significant with $p = 0.61$ and $p = 0.16$ for Condition 0 vs Conditions 1, and Condition 0 vs Condition 2, respectively.

Evolution of $PropFixInAOI$ was assessed across the time course of the experiment to check for learning (see Supplementary Information Section S7 Figure S12C). A significant effect of block number was found with $p < 0.001$. This indicates a general learning effect across three blocks of measurements which as mentioned in the previous section, can be expected due to familiarizing of subjects with the setup and the scenes. Nonetheless, the proportion of fixations in the AOI in the blurred conditions compared to the non-blur condition appears to be smaller already in the first block of the experiment.

4. Discussion

In a visual search task, it was investigated whether blurring salient regions of the visual scene can drive the observer's attention away from those regions and enable the observer to find the target faster. Using a psychophysics approach implemented in VR, we evaluated the observer's ability to find the target in a real-world scene within a limited amount of time, as well as how long it took the observer to find the target. The eye movement data were evaluated by accounting for the number of fixations within the most blurred areas and the total number of fixations until the target was found. Similarly, the mean duration of fixations were computed for a non-blur and two blur conditions.

Overall, the experimental paradigm captured well a challenging visual search task in a realistic 3D environment. The possibility of freely moving head and gaze brought the controlled experimental setting closer to a real-life scenario compared to traditional screen-based paradigms.

The significant decrease found in the search time in conditions when salient regions were blurred shows that modulating the visual scene by spatially applying saliency-aware blur can lead to a faster locating of the target. In particular, in the condition with the strongest blur (σ_2), the gain in search time was almost 10% compared to the no-blur condition. A more noticeable difference was found for the proportion of trials where the target was not found. Specifically, compared to the no-blur condition, in both blur conditions (σ_1 and σ_2), participants missed fewer trials with a significant drop of 40%. These results indicate a possibility to improve performance in a challenging visual search task via partial scene modulation using blur.

Further analysis of the eye movement data, in particular, fixations, revealed a significant decrease of both: the number of fixations until the target was found, and the proportion of fixations within the area of interest. These results correlate well with the trend observed in the behavioral metrics. Specifically, participants needed to make fewer fixations to find the target in blurred conditions than the non-blur condition, resulting in shorter search times and fewer missed trials. The fact that participants fixated fewer times within the salient areas when those areas were blurred supports our hypothesis that the observer's gaze was driven away from the scenes' blurred locations, contributing to a more successful search in terms of search time and ability to find the target. One possibility is that participants learned over the course of the experiment that the target is located in not blurred areas, which could result in a tendency to search non-blurred regions. However, the emerging difference in performance metrics and $PropFixInAOI$ between the no-blur and two blur conditions already in the first block of the experiment illustrates that, although some learning took place indicated by a significant effect of block number,

participants did not simply learn to search only non-blurred regions. These findings are in line with existing knowledge on eye movements during real-world scene viewing. Several authors [23,72,73] showed that when a low-pass spatial frequency filter is applied across the visual field, saccades are preferentially initiated to unfiltered scene regions, both in free-viewing memorization task and visual search task. Precisely, the authors showed that when a low-pass filter, or blur, is applied to the central field of view, saccadic amplitudes tend to increase, and the number of fixations decreases. Accordingly, our results also indicate that fewer saccades landed in blurred areas, and subsequently, fewer fixations fell in those regions. In accordance with the literature available, our results support the notion that eye movements are adjusted to increase the potential usefulness of inspected visual information, moreover when foveating higher spatial frequency regions of the scene [72–79].

Fixation duration, however, did not change much across different blur conditions. Cajar et al. [72] suggested that, concerning the central field of view, fixations were longer when the available spatial frequencies matched the necessary function of foveal vision, which is the analysis of details, compared to the low-pass filtered central field. On the other hand, Nuthmann [77] argued that fixations become longer when the task requires higher processing capacity, for example, when the image is slightly blurred, and it is more difficult to distinguish details. Contrarily, Cajar et al. [72], found that fixations increase in duration when the visual task becomes more difficult due to spatial-frequency filtering only when the task complexity is moderate, that is when the viewer still can make use of fixating on some locations of the scene for a longer time. In the current study, different parts of the scene were blurred to a different extent. We suggest that observers fixated mostly on more sharp areas, as is also evident from the data (*PropFixInAOI*), and did not need to adjust their fixation duration significantly. It is important to note that the task nature and the type of search target also play a role in fixation duration [12,71,80]. In the current study, the search target was a simple Gabor-cross in contrast to some sophisticated objects. The target was always known and was constant throughout the whole experiment. Therefore, the fact that fixation duration did not change much across the different experimental conditions indicates that the task difficulty did not vary significantly, but participants just needed fewer fixations to locate the target.

When studying visual search in VR, several challenges arise, producing some limitations in the current study. One aspect is the variability in the target visibility across the different combinations of scene-target location. It is easier to find the target in some trials because it is more visible in the background, whereas it might be more difficult in others. The target's visibility depends on many factors such as contrast, spatial frequency, and retinal eccentricity, as also discussed by Rothkegel et al. [12]. Simultaneously, the difficulty of finding the target represented by how long it took the participant to locate it in each trial depends not only on the visibility of the target but also in which direction (clockwise or counterclockwise) the observer started the search relative to the target location. In the current study, we did not control for the target visibility. Furthermore, by keeping the search target abstract rather than a meaningful object, we avoided spatial scene-contextual bias of the search where observers would likely fixate only the scene's locations containing anchors for the searched object [26]. However, in real life, we search for real objects. Thus, it is of interest to test our approach on realistic objects as the search target, although that may induce further bias, as some objects are expected in specific locations but not others. Also, in our study a high spatial frequency target was used. By blurring some parts of the image the distribution of the spatial frequencies of visual scene shifts towards lower frequencies. Considering feature-based guidance, this could cause the searchers to fixate less on low pass filtered locations resulting in a better performance in blurred conditions. Testing more search targets also of low spatial frequencies could provide further insight into the relation between the spatial frequency distribution of the scene and the target. In the present study, however, by selecting a certain range of possible target locations, the immediate proximity of the target was not altered by blurring. Based on the results of the study, blurring salient

locations potentially can improve visual search performance at least when the search target is of high spatial frequency. Another particularity of the current study is the variability of visual scenes. While the different scenes eliminate the spatial bias of the scenes' search and learning, it introduces a scene dependence of the salient areas. Nonetheless, it is reasonable to use multiple diverse, realistic scenes for the visual search task [12,21], as that is what we observe in real life. One more aspect to keep in mind is that we based the spatial blur pattern on a saliency map computed by a model in this study. The ultimate goal of using saliency maps, as aforementioned, is to know which locations are likely to be fixated by the viewer. The vast majority of fixation prediction models, including those we used in this study, are developed for a free-viewing task and not a search task. Therefore, for further improvement of our method, it would be interesting to generate saliency maps using more fitting models, among which can be deep-neural-network based models trained not only on ground-truth spatial locations of the fixations but also taking into account the temporal dimension of fixations [59,81]. In addition to fixation analysis, it is of interest to evaluate saccadic behavior. But compared to a screen-based paradigm, it becomes more challenging to detect saccades in the eye-tracking data stream due to a high noise level caused by the headset and rapid head movement. Thus, further development of methods for a more reliable saccade detection in VR eye-tracking data is necessary such as, for example, proposed by Diaz et al. [82]. Also, when applying a blur to the visual scene, one has to keep in mind that the visual information is partly impaired in blurred regions. This study shows that even a small blur (σ_1) can improve visual search performance. However, when the task is more demanding in processing small details, the blur strength may affect the task. Future studies analyzing how much blur can be applied improving performance in search tasks while not disturbing other performance are required. Finally, it is worth mentioning that the levels of blur used in this study were selected based on a subjective judgment of the experimenters. Specifically, a relatively small (σ_1) and a medium (σ_2) blur levels were chosen. Although no systematic analysis of subjective blur perception was conducted, some subjects reported that they noticed some parts of the scenes to be blurred more than others. Beyond the scope of the present work, future studies on subjective blur awareness can be conducted (e.g., [42]).

In the present study, a setting was used, where the target was located in not salient, and consequently, not blurred regions. This way it was intended to avoid a potential pop-out effect of the target when blurring the background. In the scope of this study, the results support the notion that blurring most salient regions can make the search more efficient at least when the target is situated in not salient locations. For subsequent studies it would be interesting to investigate a broader range of settings such as blurring non-salient regions, or blurring locations not based on saliency.

Regarding the guiding mechanisms in visual search, there are multiple factors which affect the search to a different extent, including bottom-up saliency, top-down goal-oriented feature guidance, scene guidance, recent search history of the observer as well as effects of value [4]. Various models of attentional guidance in visual search have been proposed describing the interplay of those factors (for review see [83]). A commonly used framework to describe the neurophysiological mechanism of combining the guidance elements is a priority map which is a weighted average of different factors. The priority map is considered to combine the representation of bottom-up object's distinctiveness and its top-down relevance to the observer [84,85]. The impact of saliency in visual search guidance has been debated, where the contribution of saliency is apparent in some tasks at hand [13–16], but not in others [10–12]. The neurophysiological studies also demonstrate a varied saliency contribution in the priority map building across different tasks [84]. In the present study, by demonstrating more efficient search upon suppression of salient regions, our findings, among others, implicitly show that saliency can indeed play a role in visual search strategy not only in case of synthetically generated search sets, but also in real-world visual scenes. Importantly, results of the current work illustrate saliency impact in a 360-degree scenario where a natural free head rotation is enabled as well the visual field of view is significantly

larger in contrast to screen-based studies. Even so, from previous studies it is clear that the role of saliency in visual search guidance is limited [4]. Such, if the search target would be more naturalistic and context-dependent, a smaller impact of saliency and a larger contribution of goal-oriented factors such as scene context, search history or value-based guidance can be expected. Nevertheless, we believe that through a 360-degree setting, multiple diverse visual scenes and use of a recent promising saliency model, our study contributes to understanding of saliency role in a real-world scenario of visual search.

This study served as a proof-of-principle of using a scene modulation by blurring salient regions to make the search more efficient. On a long term, the concept could potentially be implemented using, for example, see-through augmented reality head-mounted displays where the real-time video content could be modified. Another potential realization of the approach could be implementation into VR experiences such as VR simulations for professional training, e.g., in surgical procedures [86], driving simulators [87], or flight simulators [88].

5. Conclusions

To conclude, this work shows that, in a challenging realistic visual search scenario, it is possible to improve the task's performance by a saliency-aware scene modulation, specifically, partial blur. The approach's subtle nature is prone to support the user's search strategy and not be disturbing. This study provides insight into potential visual augmentation designs aiming to improve user's performance in challenging everyday visual search tasks.

Conflicts of Interest: We declare that Siegfried Wahl is a scientist at the University of Tübingen and employee of Carl Zeiss Vision International GmbH, as detailed in the affiliations. There is no conflict of interests regarding the current study.

Abbreviations

The following abbreviations are used in this manuscript:

VR   Virtual reality
AR   Augmented reality
AOI   Area of interest

References

1. Branchini, L.; Regatieri, C.V.; Flores-Moreno, I.; Baumann, B.; Fujimoto, J.G.; Duker, J.S. Reproducibility of choroidal thickness measurements across three spectral domain optical coherence tomography systems. *Ophthalmology* **2012**, *119*, 119–123. [CrossRef] [PubMed]
2. Wahl, S.; Dragneva, D.; Rifai, K. Digitalization versus immersion: Performance and subjective evaluation of 3D perception with emulated accommodation and parallax in digital microsurgery. *J. Biomed. Opt.* **2019**, *24*, 1. [CrossRef] [PubMed]
3. Chan, L.K.; Hayward, W.G. Visual search. *Wiley Interdiscip. Rev. Cogn. Sci.* **2013**, *4*, 415–429. [CrossRef]
4. Wolfe, J.M. Visual Search: How Do We Find What We Are Looking For? *Annu. Rev. Vis. Sci.* **2020**, *6*, 2.1–2.24. [CrossRef]
5. Verghese, P. Visual Search and Attention: A Signal Detection Theory Approach. *Neuron* **2001**, *31*, 523–535. [CrossRef]
6. Posner, M.I. Orienting of attention. *Q. J. Exp. Psychol.* **1980**, *32*, 3–25. [CrossRef]
7. Ferrante, O.; Patacca, A.; Di Caro, V.; Della Libera, C.; Santandrea, E.; Chelazzi, L. Altering spatial priority maps via statistical learning of target selection and distractor filtering. *Cortex* **2018**, *102*, 67–95. [CrossRef]
8. Jiang, Y.V. Habitual versus goal-driven attention. *Cortex* **2018**, *102*, 107–120. [CrossRef]
9. Borji, A.; Sihite, D.N.; Itti, L. What/Where to Look Next? Modeling Top-Down Visual Attention in Complex Interactive Environments. *IEEE Trans. Syst. Man Cybern. Syst.* **2014**, *44*, 523–538. [CrossRef]
10. Chen, X.; Zelinsky, G.J. Real-world visual search is dominated by top-down guidance. *Vis. Res.* **2006**, *46*, 4118–4133. [CrossRef] [PubMed]
11. Henderson, J.M.; Brockmole, J.R.; Castelhano, M.S.; Mack, M. Visual saliency does not account for eye movements during visual search in real-world scenes. In *Eye Movements*; Elsevier Ltd.: Amsterdam, The Netherlands, 2007; pp. 537–562. [CrossRef]
12. Rothkegel, L.O.; Schütt, H.H.; Trukenbrod, H.A.; Wichmann, F.A.; Engbert, R. Searchers adjust their eye-movement dynamics to target characteristics in natural scenes. *Sci. Rep.* **2019**, *9*, 1635. [CrossRef]
13. Jung, K.; Han, S.W.; Min, Y. Search efficiency is not sufficient: The nature of search modulates stimulus-driven attention. *Atten. Percept. Psychophys.* **2019**, *81*, 61–70. [CrossRef] [PubMed]
14. Bertleff, S.; Fink, G.R.; Weidner, R. Attentional capture: Role of top-down focused spatial attention and the need to search among multiple locations. *Vis. Cogn.* **2017**, *25*, 326–342. [CrossRef]
15. Foulsham, T.; Underwood, G. If Visual Saliency Predicts Search, Then Why? Evidence from Normal and Gaze-Contingent Search Tasks in Natural Scenes. *Cogn. Comput.* **2011**, *3*, 48–63. [CrossRef]
16. Theeuwes, J. Top-down search strategies cannot override attentional capture. *Psychon. Bull. Rev.* **2004**, *11*, 65–70. [CrossRef] [PubMed]
17. Wolfe, J.M.; Horowitz, T.S. Five factors that guide attention in visual search. *Nat. Hum. Behav.* **2017**, *1*, 0058. [CrossRef]
18. Itti, L.; Koch, C. *A Saliency-Based Search Mechanism for Overt and Covert Shifts of Visual Attention*; Vision Research; Pergamon: Oxford, UK, 2000; Volume 40, pp. 1489–1506. [CrossRef]
19. Bahmani, H.; Wahl, S. Distorted Low-Level Visual Features Affect Saliency-Based Visual Attention. *Front. Comput. Neurosci.* **2016**, *10*, 124. [CrossRef] [PubMed]
20. Schütt, H.H.; Rothkegel, L.O.; Trukenbrod, H.A.; Engbert, R.; Wichmann, F.A. Disentangling bottom-up versus top-down and low-level versus high-level influences on eye movements over time. *J. Vis.* **2019**, *19*, 1. [CrossRef]
21. Nuthmann, A. On the visual span during object search in real-world scenes. *Vis. Cogn.* **2013**, *21*, 803–837.[CrossRef]
22. Li, C.L.; Aivar, M.P.; Kit, D.M.; Tong, M.H.; Hayhoe, M.M. Memory and visual search in naturalistic 2D and 3D environments. *J. Vis.* **2016**, *16*, 9. [CrossRef]
23. Cajar, A.; Engbert, R.; Laubrock, J. How spatial frequencies and color drive object search in real-world scenes: A new eye-movement corpus. *J. Vis.* **2020**, *20*, 8. [CrossRef]

24. Drewes, J.; Trommershauser, J.; Gegenfurtner, K.R. Parallel visual search and rapid animal detection in natural scenes. *J. Vis.* **2011**, *11*, 20. [CrossRef] [PubMed]

25. Eckstein, M.P.; Koehler, K.; Welbourne, L.E.; Akbas, E. Humans, but Not Deep Neural Networks, Often Miss Giant Targets in Scenes. *Curr. Biol.* **2017**, *27*, 2827–2832.e3. [CrossRef] [PubMed]

26. Boettcher, S.E.; Draschkow, D.; Dienhart, E.; Võ, M.L. Anchoring visual search in scenes: Assessing the role of anchor objects on eye movements during visual search. *J. Vis.* **2018**, *18*, 11. [CrossRef]

27. Olk, B.; Dinu, A.; Zielinski, D.J.; Kopper, R. Measuring visual search and distraction in immersive virtual reality. *R. Soc. Open Sci.* **2018**, *5*, 172331. [CrossRef]

28. Dey, A.; Billinghurst, M.; Lindeman, R.W.; Swan, J.E. A Systematic Review of 10 Years of Augmented Reality Usability Studies: 2005 to 2014. *Front. Robot. AI* **2018**, *5*, 37. [CrossRef]

29. Coughlan, J.M.; Miele, J. AR4VI: AR as an Accessibility Tool for People with Visual Impairments. In Proceedings of the 2017 IEEE International Symposium on Mixed and Augmented Reality (ISMAR-Adjunct), Nantes, France, 9–13 October 2017; Volume 2017, pp. 288–292. [CrossRef]

30. Booth, T.; Sridharan, S.; McNamara, A.; Grimm, C.; Bailey, R. Guiding attention in controlled real-world environments. In Proceedings of the ACM Symposium on Applied Perception—SAP '13, Dublin, Ireland, 22–23 August 2013; ACM Press: New York, NY, USA, 2013; p. 75. [CrossRef]

31. Raja, V.; Calvo, P. Augmented reality: An ecological blend. *Cogn. Syst. Res.* **2017**, *42*, 58–72. [CrossRef]

32. Gatys, L.A.; Kümmerer, M.; Wallis, T.S.A.; Bethge, M. Guiding human gaze with convolutional neural networks. *arXiv* **2017**, arXiv:1712.06492.

33. Lu, W.; Duh, B.L.H.; Feiner, S. Subtle cueing for visual search in augmented reality. In Proceedings of the 2012 IEEE International Symposium on Mixed and Augmented Reality (ISMAR), Altanta, GA, USA, 5–8 November 2012; pp. 161–166. [CrossRef]

34. Grogorick, S.; Stengel, M.; Eisemann, E.; Magnor, M. Subtle gaze guidance for immersive environments. In Proceedings of the SAP 2017, ACM Symposium on Applied Perception, Cottbus, Germany, 16–17 September 2017; Association for Computing Machinery, Inc: New York, NY, USA, 2017; pp. 1–7. [CrossRef]

35. Biocca, F.; Owen, C.; Tang, A.; Bohil, C. Attention issues in spatial information systems: Directing mobile users' visual attention using augmented reality. *J. Manag. Inf. Syst.* **2007**, *23*, 163–184. [CrossRef]

36. Lu, W.; Duh, H.B.L.; Feiner, S.; Zhao, Q. Attributes of subtle cues for facilitating visual search in augmented reality. *IEEE Trans. Vis. Comput. Graph.* **2014**, *20*, 404–412. [CrossRef] [PubMed]

37. Bailey, R.; McNamara, A.; Sudarsanam, N.; Grimm, C. Subtle gaze direction. *ACM Trans. Graph.* **2009**, *28*, 1–14. [CrossRef]

38. Danieau, F.; Guillo, A.; Dore, R. Attention guidance for immersive video content in head-mounted displays. In Proceedings of the 2017 IEEE Virtual Reality (VR), IEEE Computer Society, Los Angeles, CA, USA, 18–22 March 2017; pp. 205–206. [CrossRef]

39. Pomarjanschi, L.; Dorr, M.; Barth, E. Gaze guidance reduces the number of collisions with pedestrians in a driving simulator. *ACM Trans. Interact. Intell. Syst.* **2012**, *1*, 1–14. [CrossRef]

40. Sridharan, S.; Pieszala, J.; Bailey, R. Depth-Based Subtle Gaze Guidance in Virtual Reality Environments. In Proceedings of the ACM SIGGRAPH Symposium on Applied Perception, Tübingen, Germany, 13–14 September 2015; ACM: New York, NY, USA, 2015.

41. Lin, Y.C.; Chang, Y.J.; Hu, H.N.; Cheng, H.T.; Huang, C.W.; Sun, M. Tell me where to look: Investigating ways for assisting focus in 360-degree video. In Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, Denver, CO, USA, 6–11 May 2017; Association for Computing Machinery: New York, NY, USA, 2017; pp. 2535–2545. [CrossRef]

42. Hata, H.; Koike, H.; Sato, Y. Visual guidance with unnoticed blur effect. In Proceedings of the Workshop on Advanced Visual Interfaces AVI, Bari, Italy, 7–10 June 2016; Association for Computing Machinery: New York, NY, USA, 2016; pp. 28–35. [CrossRef]

43. Ueda, T.; Iwai, D.; Sato, K. IlluminatedFocus: Vision augmentation using spatial defocusing. In Proceedings of the SIGGRAPH Asia 2019 Emerging Technologies, SA 19, Brisbane, QLD, Australia, 17–20 November 2019; Association for Computing Machinery, Inc.: New York, NY, USA, 2019; pp. 21–22. [CrossRef]

44. Khan, R.A.; Dinet, E.; Konik, H. Visual attention: Effects of blur. In Proceedings of the 2011 18th IEEE International Conference on Image Processing, Brussels, Belgium, 11–14 September 2011; pp. 3289–3292. [CrossRef]

45. Enns, J.T.; MacDonald, S.C. The role of clarity and blur in guiding visual attention in photographs. *J. Exp. Psychol. Hum. Percept. Perform.* **2013**, *39*, 568–578. [CrossRef]

46. Yamaura, H.; Tamura, M.; Nakamura, S. Image blurring method for enhancing digital content viewing experience. In *Lecture Notes in Computer Science*; including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics; Springer: Berlin/Heidelberg, Germany, 2018; Volume 10901 LNCS, pp. 355–370._29. [CrossRef]

47. Huynh-Thu, Q.; Vienne, C.; Blondé, L. *Visual Storytelling in 2D and Stereoscopic 3D Video: Effect of Blur on Visual Attention*; Human Vision and Electronic Imaging XVIII; Rogowitz, B.E., Pappas, T.N., de Ridder, H., Eds.; SPIE: Bellingham, WA, USA, 2013; Volume 8651, p. 865112. [CrossRef]

48. Sitzmann, V.; Serrano, A.; Pavel, A.; Agrawala, M.; Gutierrez, D.; Masia, B.; Wetzstein, G. Saliency in VR: How Do People Explore Virtual Environments? *IEEE Trans. Vis. Comput. Graph.* **2018**, *24*, 1633–1642. [CrossRef] [PubMed]

49. Unity Technologies. *Unity*; Unity Technologies: San Francisco, CA, USA, 2019.

50. Getting VerboseData at the Fastest Rate Possible. Vive Eye Tracking SDK—Community Forum. Available online: https://forum.vive.com/topic/5897-getting-verbosedata-at-the-fastest-rate-possible/ (accessed on 30 June 2019).

51. Van Der Walt, S.; Colbert, S.C.; Varoquaux, G. The NumPy array: A structure for efficient numerical computation. *Comput. Sci. Eng.* **2011**, *13*, 22–30. [CrossRef]
52. Virtanen, P.; Gommers, R.; Oliphant, T.E.; Haberland, M.; Reddy, T.; Cournapeau, D.; Burovski, E.; Peterson, P.; Weckesser, W.; Bright, J.; et al. SciPy 1.0: Fundamental algorithms for scientific computing in Python. *Nat. Methods* **2020**, *17*, 261–272. [CrossRef] [PubMed]
53. Mckinney, W. *Data Structures for Statistical Computing in Python*; AQR Capital Management, LLC: Greenwich, CT, USA, 2010.
54. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2020.
55. Bates, D.; Mächler, M.; Bolker, B.; Walker, S. Fitting Linear Mixed-Effects Models Using lme4. *J. Stat. Softw.* **2015**, *67*, 1–48. [CrossRef]
56. Hunter, J.D. Matplotlib: A 2D graphics environment. *Comput. Sci. Eng.* **2007**, *9*, 99–104. [CrossRef]
57. Waskom, M.; Botvinnik, O.; O'Kane, D.; Hobson, P.; Lukauskas, S.; Gemperline, D.C.; Augspurger, T.; Halchenko, Y.; Cole, J.B.; Warmenhoven, J.; et al. mwaskom/seaborn: V0.8.1 (September 2017). 2017. Available online: https://seaborn.pydata.org/ (accessed on 30 June 2019). [CrossRef]
58. Borji, A. Saliency Prediction in the Deep Learning Era: Successes and Limitations. *IEEE Trans. Pattern Anal. Mach. Intell.* **2019**, *43*, 1. [CrossRef]
59. Xu, M.; Li, C.; Zhang, S.; Callet, P.L. State-of-the-Art in 360-degree Video/Image Processing: Perception, Assessment and Compression. *IEEE J. Sel. Top. Signal Process.* **2020**, *14*, 5–26. [CrossRef]
60. Startsev, M.; Dorr, M. 360-aware saliency estimation with conventional image saliency predictors. *Signal Process. Image Commun.* **2018**, *69*, 43–52. [CrossRef]
61. ICME'17 | Salient360!—Visual Attention Modeling for 360-Degree Content. Available online: https://salient360.ls2n.fr/grand-challenges/icme17/ (accessed on 30 July 2020).
62. Baayen, R.H.; Davidson, D.J.; Bates, D.M. Mixed-effects modeling with crossed random effects for subjects and items. *J. Mem. Lang.* **2008**, *59*, 390–412. [CrossRef]
63. VIVE SRanipal SDK. Available online: https://hub.vive.com/en-US/download (accessed on 30 July 2020).
64. Imaoka, Y.; Flury, A.; de Bruin, E.D. Assessing Saccadic Eye Movements With Head-Mounted Display Virtual Reality Technology. *Front. Psychiatry* **2020**, *11*, 1. [CrossRef] [PubMed]
65. Salvucci, D.D.; Goldberg, J.H. Identifying fixations and saccades in eye-tracking protocols. In Proceedings of the Eye Tracking Research and Applications Symposium 2000, Palm Gardens, FL, USA, 6–8 November 2000; Association for Computing Machinery (ACM): New York, NY, USA, 2000; pp. 71–78. [CrossRef]
66. Kübler, T.C. The perception engineer's toolkit for eye-tracking data analysis. In Proceedings of the Eye Tracking Research and Applications Symposium (ETRA), Stuttgart, Germany, 2–5 June 2020; Association for Computing Machinery: New York, NY, USA, 2020; pp. 1–4. [CrossRef]
67. Olsen, A. The Tobii I-VT Fixation Filter Algorithm Description; Technical Report; Tobii Technology (USA): Reston, VA, USA, 2012.
68. Leube, A.; Rifai, K.; Wahl, S. Sampling rate influences saccade detection in mobile eye tracking of a reading task. *J. Eye Mov. Res.* **2017**, *10*. [CrossRef]
69. Komogortsev, O.V.; Gobert, D.V.; Jayarathna, S.; Koh, D.H.; Gowda, S.M. Standardization of automated analyses of oculomotor fixation and saccadic behaviors. *IEEE Trans. Biomed. Eng.* **2010**, *57*, 2635–2645. [CrossRef]
70. Over, E.A.; Hooge, I.T.; Vlaskamp, B.N.; Erkelens, C.J. Coarse-to-fine eye movement strategy in visual search. *Vis. Res.* **2007**, *47*, 2272–2280. [CrossRef]
71. Nuthmann, A.; Smith, T.J.; Engbert, R.; Henderson, J.M. CRISP: A Computational Model of Fixation Durations in Scene Viewing. *Psychol. Rev.* **2010**, *117*, 382–405. [CrossRef]
72. Cajar, A.; Engbert, R.; Laubrock, J. Spatial frequency processing in the central and peripheral visual field during scene viewing. *Vis. Res.* **2016**, *127*, 186–197. [CrossRef]
73. Cajar, A.; Schneeweiß, P.; Engbert, R.; Laubrock, J. Coupling of attention and saccades when viewing scenes with central and peripheral degradation. *J. Vis.* **2016**, *16*, 8. [CrossRef] [PubMed]
74. Laubrock, J.; Cajar, A.; Engbert, R. Control of fixation duration during scene viewing by interaction of foveal and peripheral processing. *J. Vis.* **2013**, *13*. [CrossRef]
75. Shioiri, S.; Ikeda, M. Useful resolution for picture perception as a function of eccentricity. *Perception* **1989**, *18*, 347–361. [CrossRef]
76. Ojanpää, H.; Näsänen, R. Utilisation of spatial frequency information in face search. *Vis. Res.* **2003**, *43*, 2505–2515. [CrossRef]
77. Nuthmann, A. How do the regions of the visual field contribute to object search in real-world scenes? Evidence from eye movements. *J. Exp. Psychol. Hum. Percept. Perform.* **2014**, *40*, 342–360. [CrossRef] [PubMed]
78. Najemnik, J.; Geisler, W.S. Optimal eye movement strategies in visual search. *Nature* **2005**, *434*, 387–391. [CrossRef] [PubMed]
79. Najemnik, J.; Geisler, W.S. Eye movement statistics in humans are consistent with an optimal search strategy. *J. Vis.* **2008**, *8*, 4. [CrossRef] [PubMed]
80. Becker, S.I.; Atalla, M.; Folk, C.L. Conjunction search: Can we simultaneously bias attention to features and relations? *Atten. Percept. Psychophys.* **2020**, *82*, 246–268. [CrossRef] [PubMed]

81. Assens, M.; Giro-i Nieto, X.; McGuinness, K.; O'Connor, N.E. PathGAN: Visual Scanpath Prediction with Generative Adversarial Networks. In *Lecture Notes in Computer Science*; (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics); Springer: Berlin/Heidelberg, Germany, 2018; Volume 11133 LNCS, pp. 406–422.

82. Diaz, G.; Cooper, J.; Kit, D.; Hayhoe, M. Real-time recording and classification of eye movements in an immersive virtual environment. *J. Vis.* **2013**, *13*, 5. [CrossRef]

83. Frintrop, S.; Rome, E.; Christensen, H.I. Computational visual attention systems and their cognitive foundations: A survey. *ACM Trans. Appl. Percept.* **2010**, *7*, 1–39. [CrossRef]

84. Fecteau, J.H.; Munoz, D.P. Salience, relevance, and firing: A priority map for target selection. *Trends Cogn. Sci.* **2006**, *10*, 382–390. [CrossRef] [PubMed]

85. Liesefeld, H.R.; Müller, H.J. Distractor handling via dimension weighting. *Curr. Opin. Psychol.* **2019**, *29*, 160–167. [CrossRef] [PubMed]

86. Lewis, T.M.; Aggarwal, R.; Rajaretnam, N.; Grantcharov, T.P.; Darzi, A. Training in surgical oncology—The role of VR simulation. *Surg. Oncol.* **2011**, *20*, 134–139. [CrossRef] [PubMed]

87. Goedicke, D.; Li, J.; Evers, V.; Ju, W. VR-OOM: Virtual reality on-road driving simulation. In Proceedings of the Conference on Human Factors in Computing Systems, Florence, Italy, 5–10 April 2018; Association for Computing Machinery: New York, NY, USA, 2018; pp. 1–11. [CrossRef]

88. Oberhauser, M.; Dreyer, D. A virtual reality flight simulator for human factors engineering. *Cogn. Technol. Work* **2017**, *19*, 263–277. [CrossRef]