

## Article

# Evaluating Human Expert Knowledge in Damage Assessment Using Eye Tracking: A Disaster Case Study

Muhammad Rakeh Saleem <sup>1</sup>, Robert Mayne <sup>2</sup> and Rebecca Napolitano <sup>1,\*</sup><sup>1</sup> Department of Architectural Engineering, Pennsylvania State University, University Park, PA 16802, USA; rakeh@psu.edu<sup>2</sup> Department of Mathematics, Chariho Regional High School, Richmond, RI 02894, USA; robert.mayne@chariho.k12.ri.us

\* Correspondence: nap@psu.edu

**Abstract:** The rising frequency of natural disasters demands efficient and accurate structural damage assessments to ensure public safety and expedite recovery. Human error, inconsistent standards, and safety risks limit traditional visual inspections by engineers. Although UAVs and AI have advanced post-disaster assessments, they still lack the expert knowledge and decision-making judgment of human inspectors. This study explores how expertise shapes human–building interaction during disaster inspections by using eye tracking technology to capture the gaze patterns of expert and novice inspectors. A controlled, screen-based inspection method was employed to safely gather data, which was then used to train a machine learning model for saliency map prediction. The results highlight significant differences in visual attention between experts and novices, providing valuable insights for future inspection strategies and training novice inspectors. By integrating human expertise with automated systems, this research aims to improve the accuracy and reliability of post-disaster structural assessments, fostering more effective human–machine collaboration in disaster response efforts.

**Keywords:** eye tracking; fixation maps; damage assessment; disaster reconnaissance; saliency maps; visual features



**Citation:** Saleem, M.R.; Mayne, R.; Napolitano, R. Evaluating Human Expert Knowledge in Damage Assessment Using Eye Tracking: A Disaster Case Study. *Buildings* **2024**, *14*, 2114. <https://doi.org/10.3390/buildings14072114>

Academic Editor: Hugo Rodrigues

Received: 21 June 2024

Revised: 6 July 2024

Accepted: 8 July 2024

Published: 10 July 2024



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Over the past 50 years, there has been an increase in the frequency of disaster scenarios, which is a major cause of concern due to its detrimental effects on infrastructure, loss of life, and social and economic instability [1]. While most natural disasters have a significant immediate economic impact, in certain instances, they can also result in long-term economic losses [2]. In the aftermath of natural disasters (e.g., earthquakes, tornadoes, landslides, floods, tsunamis, etc.), immediate and accurate structural damage assessment is crucial for public safety and recovery efforts [3]. Traditional practices in damage assessment rely heavily on the expertise of engineers who perform visual inspections according to stringent guidelines [4]. These conventional methods, however, are often hampered by the inherent limitations of manual inspections, such as human error, inconsistent assessment standards, and the significant safety risks posed to inspectors in unstable post-disaster environments [5]. The urgency and complexity of assessing disaster-prone infrastructures demand more efficient and reliable inspection methods to ensure the integrity of buildings and other structures [6]. There has been a significant shift towards using interdisciplinary approaches and techniques to address these challenges.

With advancements in robots and drone technology, unmanned aerial vehicles (UAVs) have gained popularity and become more readily available solutions to assist in search and rescue (SAR) missions [7,8] and enable aerial post-disaster damage assessment [9]. UAVs offer a safer and often more comprehensive vantage point for assessing damage, capable of accessing dangerous or impossible areas for human inspectors to reach. These

aerial platforms have been effectively utilized in various inspection contexts, including building inspections [10], railway inspections [11,12], and evaluating structural components through both imagery and LIDAR technologies [13,14]. In addition to these, researchers have noticed a rise in artificial intelligence (AI), which is widely used in structural damage detection and assessment [15,16]. It has propelled the development of sophisticated tools that can analyze structural data with greater precision, reducing the reliance on human inspectors while aiming to maintain high levels of accuracy [17,18]. Machine learning (ML) has made significant progress in computer vision applications [19]. Structural monitoring and damage identification using deep learning theory are important directions of industrial intelligence development.

Despite the advancements in UAV and AI applications for disaster reconnaissance, significant challenges remain, particularly concerning integrating human expertise with automated systems. Current UAV-based methods include (1) direct visual inspections by inspectors on-site, (2) UAVs conducting pre-programmed paths that capture images without real-time adaptation, and (3) remotely controlled UAVs that rely heavily on the quality and timeliness of video feedback. These methods cannot often make proactive, informed decisions during inspections as they cannot fully replicate the nuanced judgment of experienced human inspectors [20]. Although supervised strategies can provide reasonably accurate damage assessment results, they necessitate a proper and systematic model training process, which requires large datasets that encompass both the undamaged structure and the structure with various damage conditions. This requirement is challenging to fulfill and satisfy when dealing with built-environment structures [21]. To overcome this, we use eye tracking technology to facilitate the information-sharing process for efficient human–machine collaboration.

Previous studies [22,23] utilizing eye tracking technology across various fields provide valuable insights into visual attention and decision-making processes. For instance, experienced drivers demonstrate more efficient and flexible visual search patterns, responding to hazards earlier than novices. This suggests the need for focused training in visual search strategies and hazard anticipation [24,25]. Research indicates that experts, compared to novices, show distinct eye movement patterns that are more efficient and focused, with fewer fixations, shorter fixation durations, and more strategic saccades when examining critical features in visual scenes [26]. In medical imaging [27], experts show more focused visual search patterns, reducing diagnostic errors compared to novices [28]. Understanding these eye movement patterns can help develop targeted training programs for medical professionals [29]. A pilot study on cardiologists interpreting angiograms found that experts made more efficient diagnostic decisions with shorter review times and fewer fixations [30]. This suggests the potential of eye tracking in enhancing medical training for dynamic image interpretation.

Eye tracking is a technique in which eye movements are recorded while inspectors look at a visual stimulus, such as a building image. By capturing building inspectors' gaze data, we can investigate more specifically how expert and novice building inspectors interact with buildings differently by focusing on attention patterns and gaze paths. Up to now, the potential of eye tracking technology in the realm of the AEC industry is limited to virtual reality (VR)-based methods for inspector's training [31], and not much work has been conducted for disaster inspections concerning human–building interaction [32]. Therefore, this work's main research question is *how expert and novice building inspectors differ in their visual attention to structural and non-structural elements during post-disaster damage assessments and their implications on human–robot interaction strategies*.

The research question more specifically investigates and identifies how experts and novices interact with buildings differently. It focuses on their visual attention and gaze patterns during disaster assessments and addresses what types of insights can be taken from them. By doing this, we can either help novices learn to be better inspectors or teach robots (UAVs or UGVs) to perform inspections like humans do to promote collaboration and facilitate inspection and post-disaster reconnaissance missions. This work includes

a series of hypotheses that will be tested to validate and address the research question. Further, these findings and claims will set the base moving forward to train either a robot to perform structural inspections or human inspectors to acquire the necessary training to perform damage assessments. The main contributions of this work are (1) entailment of a screen-based method for visual inspection of a damaged site in an indoor-controlled environment, (2) analysis of eye tracking metrics to understand the inspector's visual attention, (3) implementation of the statistical methods to interpret differences between expert and novice building inspectors, (4) training a machine learning model to predict saliency maps using gaze information, and (5) assessing the predictive performance of a saliency model for disaster dataset.

The rest of the paper is organized as follows: Section 2 provides a literature review and describes prior work related to this study. Section 3 presents the overall methodology of the paper, which is further subdivided and discusses the experimental design setup, experimental procedure, evaluation approach, statistical modeling methods, and saliency mapping. Following this, Section 4 discusses the results and talks about qualitative and quantitative methods followed by machine learning results. Lastly, Section 5 concludes the paper, summarizing the research implications and future lines of research.

## 2. Related Work

### 2.1. Eye Tracking and Visual Attention

Eye tracking technology has become vital for understanding human visual attention and cognitive processes. People often struggle to accurately report the cues they use for making judgments because these assessments are frequently based on intuitive and implicit cognitive processes [33]. Eye tracking provides an external measure of behavior, offering insights into these underlying processes [34,35]. By measuring eye movements and capturing gaze focus, researchers can determine where a person is looking at any given time and the sequence of their gaze movements [36]. Eye tracking has been extensively used to study problem-solving, decision-making, reaction times, reasoning, mental states, and search strategies [37–39]. In human–computer interaction (HCI) research, software usability is evaluated, and design decisions are made by measuring attributes like fixation count and fixation time to assess visibility, meaningfulness, and placement of user interface elements [40]. Fixations, or stable gazes, represent information acquisition and processing.

Due to limited cognitive resources, the human visual system must select relevant information from a cluttered visual environment [41–43]. Attention enhances important information while inhibiting irrelevant data, with visual attention often shifting through physical eye movements [44,45]. Overt attention mechanisms involving direct eye fixation can be effectively measured with eye trackers to define what viewers attend to on visual media [46]. Eye tracking has also been used to model the salient features of images for fixation location prediction [47]. The relationship between visual stimuli features and attention patterns helps predict how people observe visual media [48].

### 2.2. Immersive Techniques for Training

Virtual reality (VR) has emerged as a valuable tool for disaster training [32,49,50], offering immersive environments for various safety training applications. VR has been tested for fire safety [51], tsunami and earthquake training [52,53], aviation training [54], and counter-terrorism safety training [55]. Studies have shown that VR training tools can better enhance knowledge acquisition and retention than traditional methods [56]. VR can simulate realistic disaster scenarios, helping communities in high-risk areas prepare for emergencies by increasing risk perception and promoting early evacuation decisions [57]. For instance, VR frameworks have been developed to create realistic 3D gaming environments that increase public awareness of flood risks [58]. Other applications focus on co-design methodologies to identify the contents of VR simulations for river flood emergency training [56]. These preliminary studies demonstrate the feasibility of using VR for disaster safety training. However, VR has not yet been widely applied to assess shelter

choices during tsunami evacuations or to optimize building locations for vertical evacuation. The potential of VR to enhance community preparedness and response to disasters through improved self-reliance is promising and warrants further exploration.

### *2.3. AI and Computer Vision in Infrastructure*

Artificial intelligence (AI) has significantly advanced structural damage detection and assessment. Deep learning (DL), particularly convolutional neural networks (CNNs), has made significant strides in computer vision applications, such as structural monitoring and damage identification [59,60]. CNNs can discover abstract features and complex classifier boundaries, enhancing model performance by learning from large datasets without extensive manual feature engineering [61]. AI-driven methods have been applied in automated bridge deck crack detection, providing crucial information for infrastructure maintenance and timely repairs [62]. Automated systems using UAVs for infrastructure inspection have shown the potential to reduce costs, increasing inspection speed, accuracy, and safety [63]. UAVs equipped with high-definition cameras can capture high-quality inspection data, and hybrid solutions combining ground and aerial robots have been developed for challenging inspection tasks [64]. These robotic systems can significantly enhance infrastructure inspections, although their development level still lags behind other areas, necessitating further advancements [65]. Integrating AI and UAV technologies promises to enhance the efficiency and accuracy of structural inspections, addressing the limitations of traditional methods.

### *2.4. Challenges in Developing Human-Centered AI/ML Decision Support*

AI techniques offer scalable and affordable solutions for automated knowledge discovery and reasoning. Machine learning (ML) algorithms improve through data, with tasks categorized into regression, classification, and clustering under supervised, unsupervised, and reinforcement-learning paradigms [66]. Expert systems provide an alternative approach, offering increased explainability compared to 'black box' ML applications [67]. AI has been applied to various tasks in healthcare, such as blood glucose prediction and lifestyle support [68,69]. Despite the potential of AI in various fields, integrating human expertise with AI systems remains a challenge. Ensuring model interpretability and optimizing performance requires a comprehensive understanding of the underlying cognitive processes and effective collaboration between humans and machines [70]. Integrating eye tracking, VR, AI, and UAV technologies holds great promise for enhancing post-disaster structural inspections. Eye tracking provides valuable insights into human visual attention and cognitive processes, informing AI models that replicate expert behavior. VR training tools offer immersive environments for disaster preparedness, while AI and UAV technologies enable efficient and accurate damage assessments [32]. Despite the current challenges, continued advancements in these fields will likely lead to more effective human-machine collaboration, improving disaster response and recovery efforts [40].

## **3. Methodology**

This study analyzes how expertise shapes human-building interaction during disaster inspections and how this can provide insights for training a machine learning (ML) model for efficient structural inspection and damage assessment, facilitating human-robot collaboration. It helps explore human attention patterns and reasonings to understand how an inspector (expert or novice) perceives a disaster site and performs a damage assessment. Once the eye tracking data are collected, metrics such as fixations, visits, and gaze paths will be considered. To obtain meaningful information from the collected data, our goal in this study was threefold as follows:

1. First, to understand the differences in visual attention between expert and novice building inspectors, eye tracking features were explored in depth, and statistical analysis was conducted.

2. Second, to model visual attention patterns, a task-specific ML model was trained leveraging inspectors' gaze data to predict saliency maps and subsequently compare and discern the distinct attention patterns of experts and novices.
3. Third, saliency metrics analysis was conducted to evaluate the ML model's performance in predicting visual attention.

To this end, we have conducted a screen-based case study to collect eye tracking data of building inspectors on a disaster site for post-disaster damage assessment and structural inspection. Furthermore, we have trained an ML model for predicting saliency maps to highlight the damaged parts of a structure. These saliency maps serve multiple purposes. Firstly, they allow for the visualization and quantification of areas of interest within an inspection scene, making it possible to identify specific structural and non-structural elements that attract more or less attention from each group. By comparing saliency maps generated from experts' gaze data to those from novices, we can pinpoint the key features or elements that experts consistently focus on but novices may overlook. This information is invaluable for training novice inspectors to develop a more expert-like approach to damage assessment. Additionally, understanding visual attention patterns can inform the design of collaborative inspection strategies, where robots can be programmed to highlight or focus on areas that experts typically prioritize, thus guiding novices during their assessments. By providing actionable insights, these saliency maps enhance training and human–robot interaction strategies, directly addressing our research question and bridging the gap between understanding the inspector's visual attention and applying this knowledge to practical scenarios in post-disaster damage assessments.

### 3.1. Experimental Design Setup

#### 3.1.1. Participants

Twelve participants (8 male and 4 female) were recruited from the architectural engineering (AE) department at the Pennsylvania State University (PSU) for the study. All participants had normal or corrected-to-normal vision and had no reported color blindness. Eleven participants ranged in age from 28 to 37 years old ( $M = 32$  years,  $SD = 2.8$ ), and one participant was 70+ years old. Participants were divided into groups: Experts—those with a deep understanding of structural assessment and damage detection with prior field experience in structural assessment and disaster reconnaissance and recovery missions; and novices—those with basic knowledge of structural assessment and damage detection with no expertise or field experience in disaster reconnaissance missions.

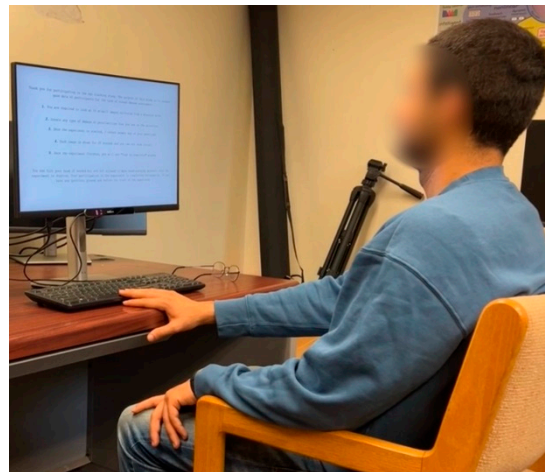
The Federal Emergency Management Agency (FEMA) provides a preliminary damage assessment guide, providing detailed information about the experience and training required for inspectors involved in preliminary damage assessments (PDAs) [5]. The document emphasizes the importance of experience in conducting PDAs, indicating that the PDA team leads should be senior program specialists or specialists involved in PDAs multiple times. The experience is critical for ensuring consistency, accuracy, and successful outcomes in PDAs. Further, it highlights the importance of training for damage assessment personnel to ensure they are familiar with damage assessment plans, standard operating procedures (SOPs), and protocols. We speculated that by comparing these knowledgeable novices and expert building inspectors, we could identify and clarify the relationships between inspectors' knowledge and visual assessment strategies. The number of participants and the dependent variables for the analysis were pre-defined based on the results of a pilot experiment in a real built environment [10]. The study followed the ethical guidelines of Human Subjects Research and was approved by the Institutional Review Board (IRB) for the Human Research Protection Program (HRPP) of PSU (STUDY00016625). The participants gave informed consent before the start of the experiment.

#### 3.1.2. Apparatus

Eye movement data were recorded using Tobii Pro Nano, a screen-based eye tracking equipment developed by Tobii Technology [71] (60 Hz sampling rate; 0.3 deg accuracy,



0.1 deg precision; and gaze recovery time of 250 ms). It uses pupil corneal reflection and binocular dark pupil acquisition technology to track participants' eye movements. The calibration was conducted using Tobii's proprietary SDK [72], which allows for a 5-point calibration of eye gaze data. The experimental setup is shown in Figure 1. During the calibration, participants were allowed to sit in front of the screen-based eye tracker with no head movement and were to focus on the calibration marker shown on the screen. Once the calibration is completed, a validation prompt appears on the screen to confirm and use the calibration for collecting gaze data. Pro Lab software was used for the data analysis, which provides a visual and functional user interface for complex experiment design, viewing individual recordings, and aggregating data for analysis and visualization.



**Figure 1.** Experimental setup for data collection.

The instructional prompt given to the participant before the beginning of the experiment is as follows:

*Thank you for participating in the eye tracking study. This study aims to collect gaze data from participants to assess visual damage.*

- *You are required to look at the images collected from a disaster site.*
- *Each image will be shown for 20 sec, and you cannot zoom in/out.*
- *Locate any type of damage or peculiarities that you see on the structure.*
- *You will not be evaluated based on your skills; this study will not impact you.*
- *You can tilt your head slightly, but you should not move your head.*
- *Once the experiment is started, I cannot answer any of your questions.*
- *At the end of the experiment, you will see the “Task is completed” window.*

*Your participation in the experiment is completely voluntary. If you have any questions, please ask before the start of the experiment.*

### 3.2. Experimental Procedure

#### 3.2.1. Data Collection

The eye tracking data were collected from a post-disaster site in Mayfield, Kentucky, a historic district following the Midwest tornadoes in December 2021. The prevalent building typologies and most common structural typologies consisted of one-story, two-story, multiple-story, and unique structures separated from other structural and cultural heritage buildings [73]. These include the (1) Merit Clothing Mill, a two-storied structure with flat brick plasters and symmetric windows; (2) Hall Hotel, a five-storied structure with a brick-infilled concrete frame, termed the tallest structure in Mayfield; and (3) American Legion, the first brick-infilled steel-frame structure in town.

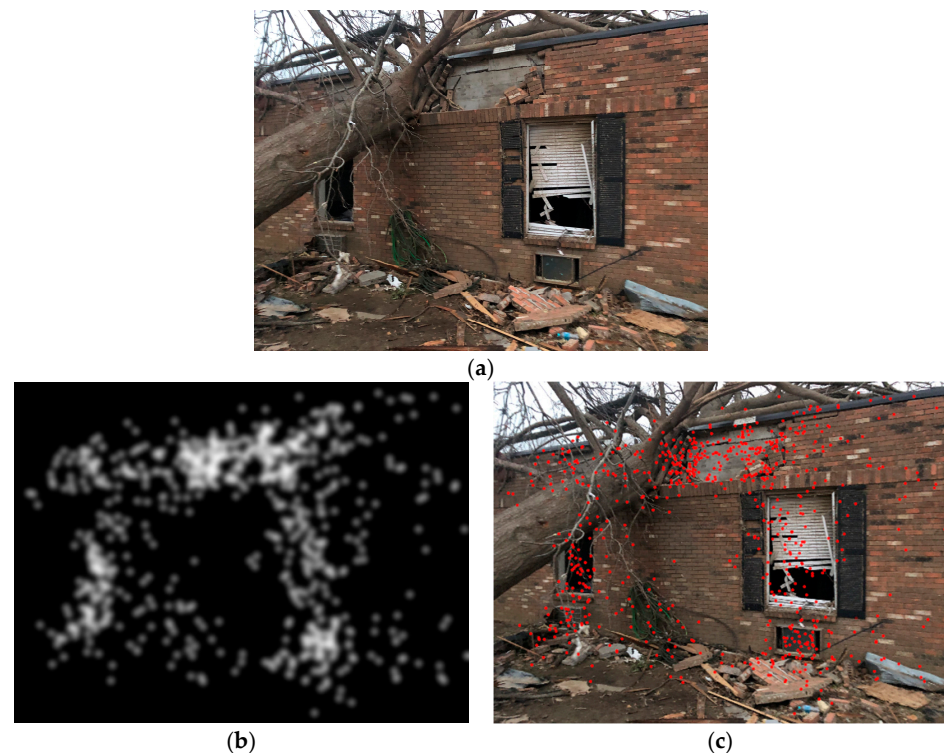
Additionally, the data collected consisted of images captured from an expert's mobile phone camera. The photos were taken by an expert inspector and a member of StEER's

(structural extreme events reconnaissance) field assessment structural teams (FAST). They have participated in multiple field inspections and have over five years of experience visually inspecting built infrastructure, including buildings, bridges, and disaster sites. However, to maintain the objectivity of our study, the expert inspector who collected data was not included in the experiment. We anticipated including their data could introduce bias and subjectivity into the results. Thus, we decided to exclude them from the study to avoid any potential bias, ensure we did not double-check the expert's work, and maintain the integrity of the collected field images.

### 3.2.2. Data Processing

Similar to any data, eye tracking data encounters variability in viewing patterns due to differences among participants, research fields, and visual contexts, making it challenging to establish universally optimal parameter settings. This work seeks to understand the relationship between parameter settings and the features derived from them. Prior to feature extraction, we explored the velocity-threshold identification (I-VT) [74] filter for fixation detection and attention to improve the fixation classification. The choice of algorithm depends on the specific eye movement that is of interest. Since this case study is focused on a screen-based experiment with no head-and-body movement, we used the fixation filter to classify individual gaze points. Since the study of the I-VT filter is not part of this study, we used the default settings suggested by Tobii [72] for efficient feature classification.

The collected gaze data will serve as the ground truth fixation map for metric analysis and training a ML model. Before model training and metric analysis, the collected gaze dataset was prepared for data denormalization, image analysis, and rectification to remove incorrect or missing data. Following this, the model was fine-tuned to predict saliency maps using gaze information. An important step was generating discrete and continuous fixation maps for saliency metrics. Some metrics compute similarity scores based on continuous fixation maps, and others are based on discrete fixation locations. Figure 2 shows an example of a distribution-based fixation map. The results are compared between fine-tuned saliency maps and ground truth fixations based on the gaze data.



**Figure 2.** (a) Original input image, (b) ground truth fixations location ( $n = 12$ ), and (c) overlay of fixations data (red dots) on top of the actual image.

### 3.3. Eye Tracking Metrics and Analysis

Eye movements include where a person is looking, how long they are fixated, and their gaze pattern; these all define how their attention switches between different visual scenes. The eye metrics that will be considered in this study are fixations and visits [22,28]; *fixations* are when eyes are stable and stationary, looking at a particular region of interest in the scene, and *visits* inform us about the back-and-forth fixations that occur when a person looks from one area of interest (AOI) to another, particularly structural elements. Once the data is collected, feature extraction is performed to capture and analyze these metrics. This work will consider the fixation duration, fixation count, visit duration, and visit count. These eye tracking metrics will form the basis for understanding gaze patterns for structural inspection and damage detection. In addition, metrics such as pupil dilation are also important in informing the inspector's focus and cognitive load under which they are completing a task.

For qualitative analysis, heat maps and gaze plots provide fast and accurate data visualizations that are important in understanding aspects of visual behavior. An example of a "Heat Map" is shown in Figure 4 for detail. The heat map describes the overall distribution of the inspector's vision over a particular stimulus. The areas with red color indicate the high intensity of an inspector's fixations compared to low-intensity areas indicated by light green. Heat maps can effectively document visual attention in a scene and consider all the fixations of an inspector; this can be used to understand decision-making processes. In addition, we classified the stimuli (building images) into *structure* and *non-structure* categories to understand the differences between experts and novices. The structure category represents the structural regions with potential damage related to cracks, connection joints, façade, roofs, and windows. The non-structural category represents the surroundings or debris of the structure containing structural rubble, such as broken bricks, steel frames and beams, and general debris. Section 4.1.2 explains the image-labeling process and area of interest (AOI) tag classification in detail.

### 3.4. Statistical Modeling Methods

In addition to the visualization of the eye tracking metrics, statistical analysis was performed to determine if there was a significant difference in the data. The dataset was categorized into two categories: (1) level of experience, i.e., experts and novices, and (2) damage category, i.e., structures and non-structures. We tried to test three hypotheses to obtain further insights into how expertise shapes building inspectors' assessment strategies and their key differences. First, we tested whether there is a significant difference between experts detecting damage for fixation and visit metrics compared to novices (*Hypothesis 1*), as reported by [28]. Second, regarding the localization of damage to the structure or surroundings, is there a difference in structural and non-structural damage location (*Hypothesis 2*)? Third, regarding visual attention behavior, we identified that damage typology could critically influence experts' and novices' fixation times. Therefore, we were keenly interested in understanding whether damage localization and typology differ significantly among inspectors (*Hypothesis 3*).

Before data analysis, the normality condition must be checked using the Shapiro–Wilk test or Wilcoxon test, and the homogeneity of variance must be checked through Levene's test, respectively. The Shapiro–Wilk test is a more formal test to check the normalcy condition. Levene's test assesses the homogeneity of variances for a variable; these tests of normality and homogeneity give quantitative insight into the nature of the data. In statistics, a parametric test such as paired *t-test* or ANOVA is used for comparing paired samples when they meet the condition of normality, and a non-parametric test such as the Wilcoxon signed-rank test or Spearman's rank correlation coefficient test is used for comparing paired samples when the differences are non-uniformly distributed. For our case study, all three metrics, fixation count, visit count, and visit duration, satisfy the normality condition for both expert and novice groups, except for fixation duration for the novice group. Therefore, it can be assumed that all but one group's dataset is not normally



distributed. In our statistical modeling, we used a 2-sample *t-test* as the parametric method and a Wilcoxon signed-rank test as a non-parametric method.

The statistical analyses inform significant differences in visual attention and decision-making processes between expert and novice building inspectors, directly addressing our main research question. These findings highlight areas where inspection strategies diverge based on experience and identify specific elements within visual scenes that are critical for effective damage assessment. By analyzing these statistically significant differences, we are informed about the critical aspects of visual attention that need to be emphasized in training and technology development. With this foundation, we aim to utilize machine learning to bridge the gap between human cognitive capabilities and automated systems.

### 3.5. Machine Learning Approach

Subsequent machine learning applications aim to take advantage of these statistical insights by developing models that can predict and replicate the saliency maps of visual attention observed in building inspectors. This step is crucial as it allows us to (1) create tools that can potentially train novice inspectors to adopt more expert-like visual strategies and improve their efficiency and (2) implement them in a robot to assist in human-machine collaboration and automated inspection in disaster scenarios. Furthermore, by comparing the machine-generated saliency maps with the ground truth data collected from expert and novice inspectors, we can validate the effectiveness of our ML approach. This supports the findings from our statistical analysis and contributes to refining our understanding of how visual attention can be optimized in the context of structural inspections post-disaster. The integration of these findings with AI-driven tools can, therefore, facilitate more informed and strategic human-machine collaboration, directly enhancing the inspection process and safety measures following natural disasters.

#### 3.5.1. Datasets

For the success of any ML model, a prerequisite is having a good dataset for training purposes. Despite the abundance of many large-scale eye tracking datasets [75], the absence of building inspection datasets with an effort to eye tracking data remains a significant hurdle. Therefore, this study utilizes the disaster dataset developed by Kaushal et al. [73], and the dataset was filtered to remove redundant and noisy images from the tornado site. The total number of images gathered from the disaster site was 70; however, we performed screening to remove images with noise and obstructions in visual scenes. Each image consisted of accumulated gaze measurements for all twelve participants and contained fixation points ( $x,y$ ) for each inspector used for training. The input images had different resolutions, such as (1)  $3024 \times 4032$ , (2)  $2268 \times 4032$ , and (3)  $3840 \times 2160$ .

#### 3.5.2. Model Architecture

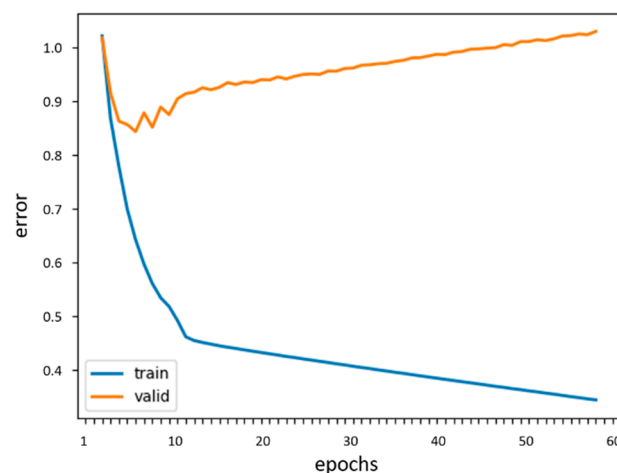
Our model architecture was adapted from Kroner et al.'s [76] work, which utilizes a popular VGG16 [77] architecture as an image encoder-decoder by reusing the pre-trained convolution layers to extract complex features along its hierarchy. On the encoding side, convolution layers were dilated at a rate of 2, increasing the receptive field to compensate for a higher resolution. It is important for us since our images have 2K and 4K image resolutions. This approach effectively solves saliency prediction problems [78,79]. We chose a decoder architecture to restore the original image resolution for better approximation and saliency prediction.

#### 3.5.3. Implementation Details and Training

The present study utilized a desktop with an Intel i9-9900K CPU @ 3.60 GHz, 64 GB RAM, and NVIDIA GeForce RTX 2080 GPU for training and testing the models. For training, the final data prepared to collect eye tracking data contained 33 images of different visual scenes at the disaster site. The gaze information of 12 inspectors was collected, and their fixations were accumulated together. We split the model training into 25 images for training

and 8 images for validation with mixed visual scenes of the disaster site. These images were shuffled randomly for uniform distribution, containing simpler building images and a mix of partially destroyed buildings. For testing, we have considered 7 images from the bigger dataset, i.e., originally containing 73 images, to test the model's predictive performance. These images were part of the single-story nursing home building taken from different angles, showing different pieces of information and structural details.

The deep learning network was trained using the stochastic gradient descent optimizer with a learning rate of  $1 \times 10^{-6}$  and a batch size of 1. The batch size was set to 1 due to the smaller dataset, and all images were resized to  $1080 \times 720$  pixels during training to facilitate efficient processing. The encoding layer was based on the VGG16 architecture pre-trained on the ImageNet dataset. To train the model, we used stimulus images and their corresponding fixations data saved as binary maps as input, and the model's output was a 2D saliency map. We trained our model for 50 epochs and used the best-performed checkpoint for inference. The results of our training are shown in Figure 3, which indicates the model was learning until epoch 10, and the error loss was decreasing. Following this, the model suffers from overfitting, and we chose checkpoint 4 for inference and documenting the results.



**Figure 3.** Training results.

### 3.6. Saliency Metrics Analysis

In this research, we have computed the metrics score to test which saliency metrics are more reasonable and whether disaster data will impact the metrics score. This will help us define and create a baseline for the dataset used in this work. To illustrate the distinct behaviors of saliency evaluation metrics, we compare the ground truth fixation map results with the predicted saliency map. The goal is to use the model for saliency prediction, compute similarity scores for various metrics, and generate a saliency map using eye gaze data. These metrics generate scores to evaluate the accuracy of saliency predictions based on their closeness to the ground truth. We reviewed the five predominant saliency evaluation metrics outlined in Table 1.

**Table 1.** Metrics used for evaluating saliency maps.

Metrics	Location-Based	Distribution-Based
Similarity	Area under ROC curve [80–82]	Similarity [83]
	Normalized scanpath saliency [80,82,84–86]	Pearson's correlation coefficient [80,81,84]
Dissimilarity	—	Kullback–Leibler divergence [83,85]

The MIT saliency benchmark interprets these metrics and evaluates different methodologies, treating saliency maps simply as intensity maps without specific format restrictions. If a metric requires probability distributions, we normalize the saliency maps accordingly without further changes. These metrics also differ in how they utilize ground truth: some assess saliency based on the exact fixation points (location-based), whereas others consider the overall distribution (distribution-based). Equation (1) illustrates how these metrics relate saliency maps to their ground truth values, aiming to maximize the similarity scores and minimize the dissimilarity scores:

$$S = f(\text{Sal}, \text{GT}) \quad (1)$$

where  $S$  denotes the similarity score, and  $f$  represents the scoring function that assesses the saliency map ( $\text{Sal}$ ) against the ground truth ( $\text{GT}$ ). This paper examines these five metrics independently of their input specifics, using an eye tracking dataset for analysis. We differentiate the metrics based on whether they use discrete fixation points or continuous fixation maps as ground truth, categorizing them accordingly. We discuss the strengths and weaknesses of each metric and provide visual examples of their computational methods.

## 4. Results

### 4.1. Eye Tracking Measures

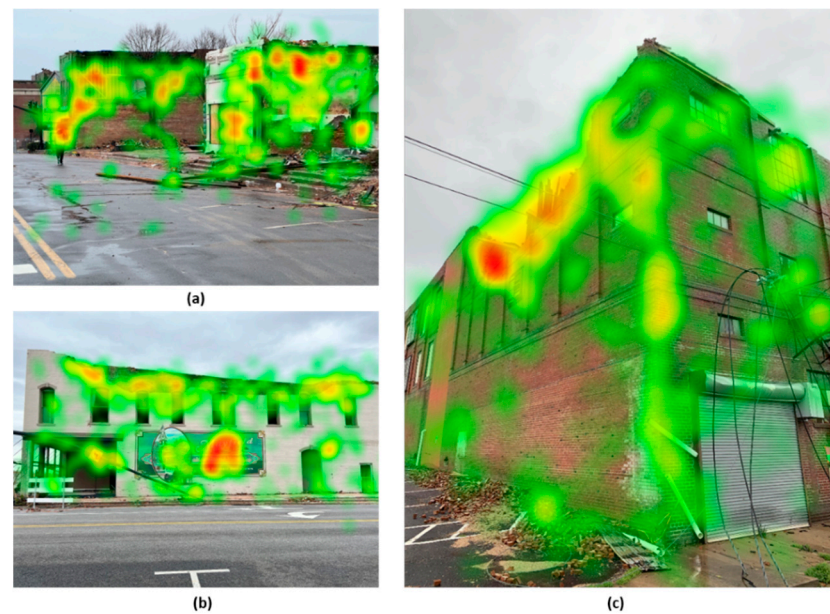
#### 4.1.1. Qualitative Analysis

Different techniques, such as heat maps and scan paths, provide fast and accurate data visualization for understanding inspector gaze patterns and visual attention. Figure 4 shows heat maps for multiple buildings, consisting of the aggregated data for all the inspectors. The buildings include (1) two-story row building structures (see Figure 4a), (2) a two-story corner building (see Figure 4b), and (3) a multi-story clothing mill structure (see Figure 4c). The distribution of the inspector's fixation shows they were more interested in the façade of the structure. For all three structures, the heat maps indicate a higher intensity on the outer envelope of the structure, followed by the debris surrounding the structure. In Figure 4a, the inspectors were more interested in the façade of the structure, and they had higher fixations on the completely damaged wall, followed by a human inspector who was walking in the visual scene. Figure 4b indicates that the inspectors had higher fixations at the central region at the poster. Following this, we noticed a higher gaze distribution at the top of the structure, where the roof is more damaged and detached from the wall. Similarly, for Figure 4c, the inspectors paid more visual attention to the top of the structure where the wall was broken, and they distributed their overall attention to the façade envelope.

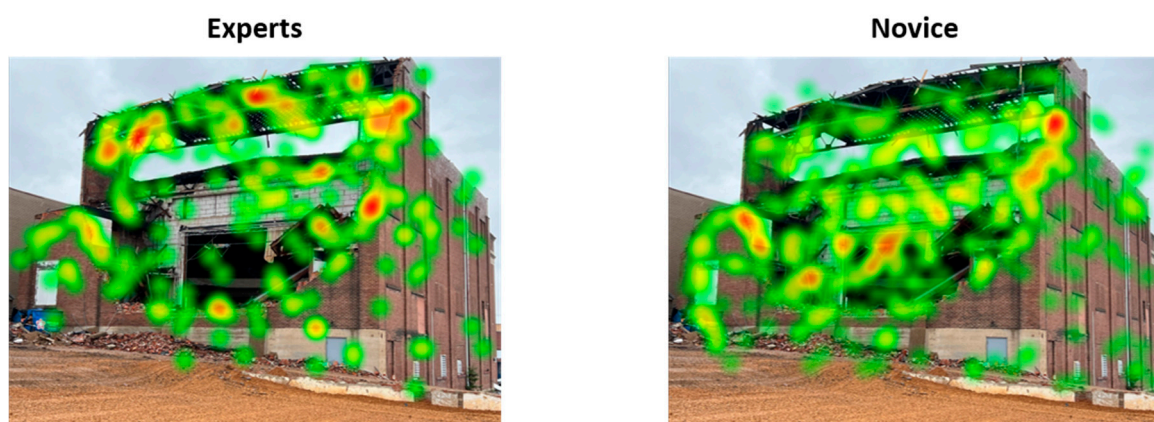
We also compared multi-story tall structures such as the American Legion and two-story row buildings among expert and novice inspections. Figure 5 shows the distribution of visual attention for experts and novice inspectors. The results show that the experts had fewer fixations on the building façade than the novices. Similarly, the experts looked at the steel-frame structures and the broken walls. At the same time, the novices had more fixations only on the building façade and walls, with their fixations distributed all over the structure. Figure 6 shows a scene of row buildings in the downtown area with a person looking at the structure in the surroundings. It is indicated that expert inspectors fixate on the façade of the masonry structure and the timber portion in the wall compared to novice inspectors, who gaze at the rubble of the structure or a person walking on the road. Novice inspectors' attention is distributed more and entirely compared to experts who fixate on the structural elements of the buildings.

To make deductions beyond those drawn from the heat maps, the inspectors were asked to rate their fatigue level on a scale of 0 to 10 and provide feedback/comments, if any, after completing the eye tracking experiment. Inspectors with higher fatigue levels provided feedback that reflects difficulty in maintaining concentration, such as comments on the task's intensity and the desire for a transition window to avoid a fixated gaze. This could indicate that the task's more demanding or less intuitive aspects contribute to higher fatigue levels. On the other hand, lower fatigue levels are associated with more constructive

feedback about the task rather than the task's effect on the participant. It could suggest that the participants with lower fatigue levels can better cope with the task's demands or find it less cognitively taxing.



**Figure 4.** Heat maps for qualitative comparison among (a) a two-story row building, (b) a two-story corner building, and (c) a multi-story building structure.



**Figure 5.** Qualitative comparison for American Legion.



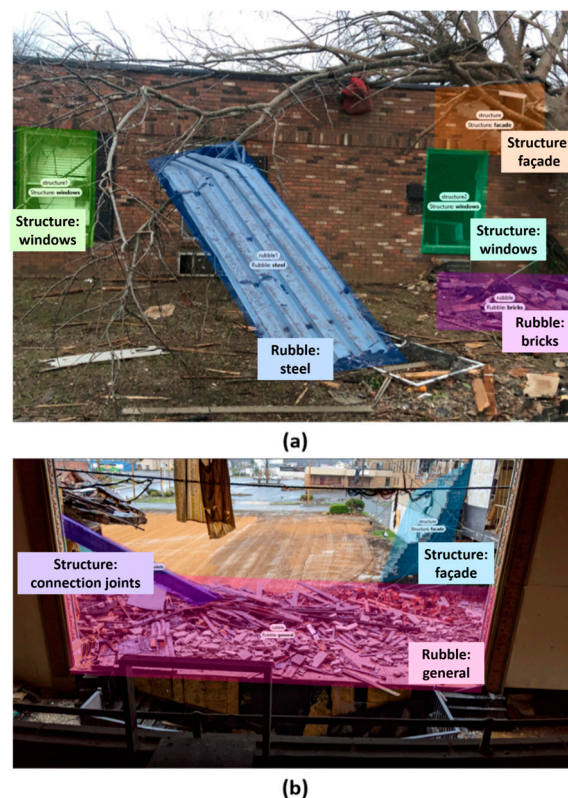
**Figure 6.** Qualitative comparison for row buildings.



In summary, experts offer more targeted feedback that could enhance the inspection process, whereas novices often express a need for more guidance and support during the task. Higher fatigue levels appear to be associated with feedback reflecting the cognitive or visual strain of the task. This information could be invaluable for refining the inspection process and training programs, particularly by addressing the specific challenges and needs identified by different experience levels and in relation to the fatigue reported by participants. It also underscores the importance of clear task instructions and the potential value of rest periods or a more intuitive interface, especially for tasks perceived as cognitively demanding.

#### 4.1.2. Quantitative Analysis

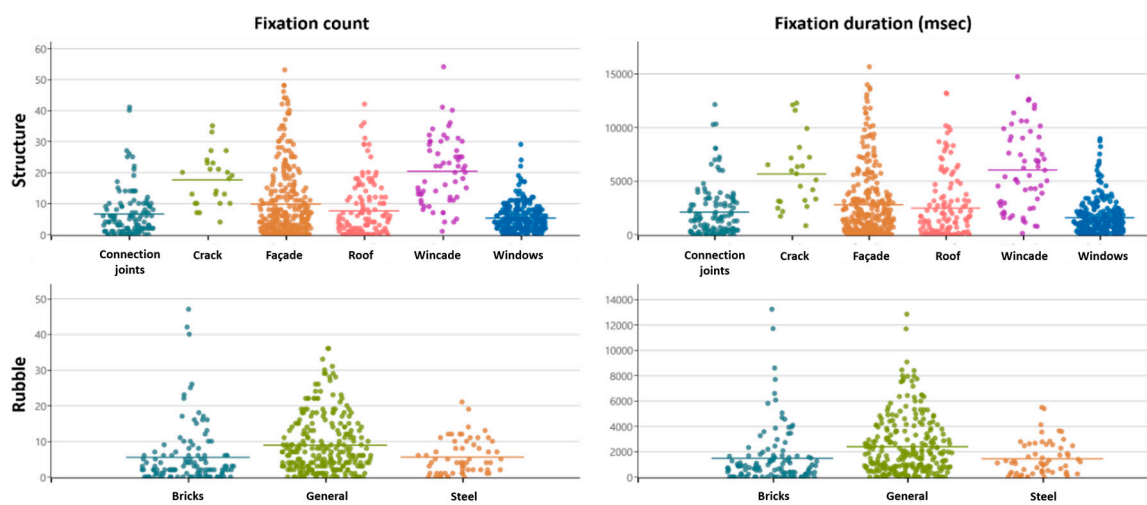
Unlike qualitative analysis, eye tracking data analyzes experts' and novice inspectors' behavior quantitatively using the metrics mentioned in Section 4.1. The stimulus images were marked with areas of interest (AOI) and classified into structure and non-structure categories. These labels were marked to compare the performance of inspectors based on their knowledge and experience. The *structural* category represents damage annotated based on different damage regions such as connection points, cracks, façade, roofs, and windows. The *non-structural* category represents non-damaged regions that are part of the building rubble (bricks, steel, and general debris) or the structure's surroundings. Figure 7 provides an illustration of the AOI tags that were marked for the analysis. The AOIs are marked according to the structural and non-structural tags and further labeled into individual damaged categories.



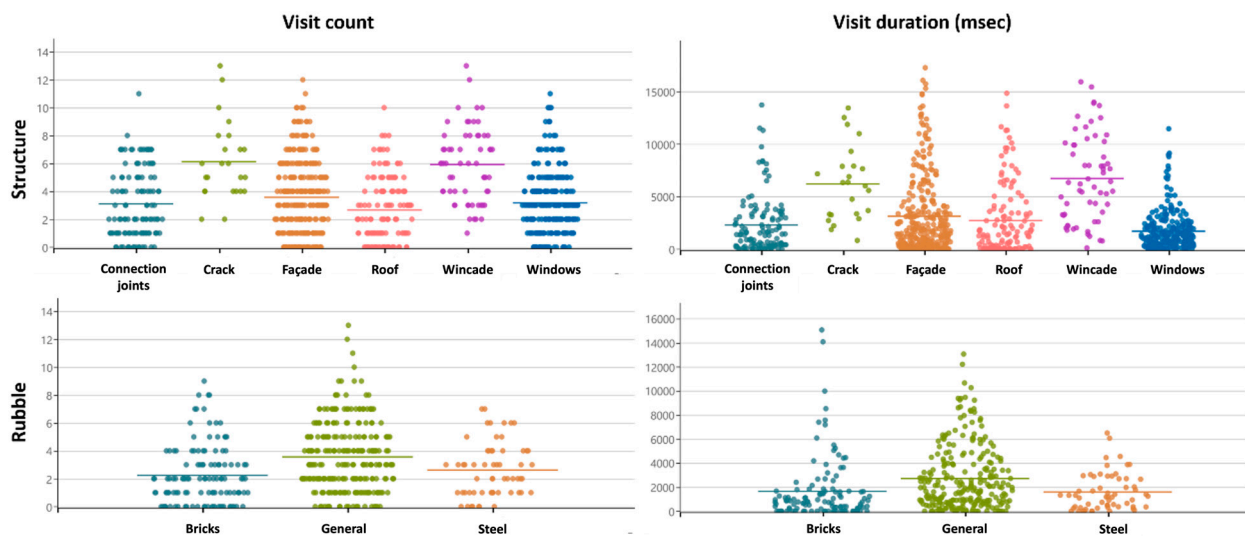
**Figure 7.** Example of area of interest (AOI)-based damage labels for (a) a nursing home structure (exterior) and (b) American legion building (interior).

As mentioned in the previous section, metrics considered for analysis include the fixation duration, fixation count, visit duration, and visit count. Figures 8 and 9 provide illustrations for the fixations and visit metrics. Figure 8 provides the fixation metrics dissected into the fixation count and fixation duration, offering a contrast between both

frequency and temporal aspects of visual attention. The horizontal solid line represents the median, often referred to as the second quartile, that divides the data into two equal parts. From the number of fixations (left plot, fixation count), a higher concentration of data points around the median line for façade, windows, and connection joints elements suggests that these areas command more frequent visual attention than others. It indicates that inspectors may be instinctively drawn to focus on these elements while the relative spread of data points for crack, steel, and bricks is more distributed, implying a more varied attention pattern among participants. In terms of fixation duration (right plots), the clusters illustrated for façade and general debris show longer fixation times. This implies that when participants look at these areas, they do so for extended periods, potentially reflecting the complexity or significance of these elements in the context of the task at hand. The long tail of data points extending for the façade and bricks suggests that certain fixations on this feature are particularly lengthy, perhaps due to its intricacy or relevance in the visual evaluation.



**Figure 8.** Comparison of fixation count and fixation duration metrics (median is represented by a horizontal solid bar).



**Figure 9.** Comparison of visit count and visit duration metrics (median is represented by a horizontal solid bar).

Similarly, Figure 9 represents the visitation data broken down into visit count and duration, and the horizontal solid line represents the median, often referred to as the

second quartile, which divides the data into two equal parts. For the ‘structural’ element category, the visit count shows a relatively even distribution across connection joints, facades, roofs, and windows, and there is a slightly higher count for the crack and wincade elements. This evenness suggests that the inspector’s gaze was not exclusively fixated on a single feature but distributed among various aspects of the structure. It is interesting to note that although the crack and wincade elements display slightly higher visitation, the façade element exhibits prolonged engagement with façade fixations, notably having an extended range of durations. This suggests that these elements may be visually complex or have significant importance in the assessment context, necessitating more extended examination periods. Additionally, in the rubble category, the visit count and duration are more concentrated on general debris, indicating the inspector’s tendency to direct their gaze more generically on the surrounding damaged structures.

These plots offer a detailed breakdown of visual attention allocation across various structural and material elements for visual inspection in disaster reconnaissance and recovery missions. Such data are pivotal in elucidating the cognitive priorities and visual strategies employed by inspectors when presented with complex scenes (disaster sites) requiring detailed, fast, and accurate visual analysis. Elements likely to be more task-relevant, such as facades in structures and general debris in rubble, tend to attract more fixations and longer examination times. This may be due to inspectors prioritizing these areas as they may hold more information necessary for decision-making within the task context. Assuming that the inspectors’ level of expertise could vary, one could hypothesize that experts might show fewer but longer fixations on complex elements, denoting efficient information gathering. In comparison, novices might demonstrate shorter fixations, indicating a search for relevant cues. This, however, would need validation with additional data on the inspector’s level of expertise.

#### 4.2. Statistical Modeling

This work investigates inspectors’ visual attention patterns and performs statistical analyses of their gaze data. As discussed in the previous section, to address the main research question of this study, several hypotheses were proposed and anticipated to address the goal of this paper. These hypotheses were built upon the main research question, “*how expert and novice building inspectors differ in their visual attention to structural and non-structural elements during post-disaster damage assessments, and what are their implications on human-robot interaction strategies*”. By statistically validating these differences, we set the stage for applying machine learning models to synthesize these complex patterns into practical training and operational enhancement tools. The results from these analyses help highlight the specific areas where machine learning can assist, particularly in standardizing the damage detection strategies across varying levels of inspector experience and optimizing the interaction between human inspectors and robot teams.

We have divided our dataset into two variables: (1) level and (2) category, where *level* indicates the experience of inspectors, which is expert and novice, while *category* indicates the type of damage, which is structure and non-structure. To perform statistical modeling based on the level of experience and structural components, we first checked the condition of normalcy to help choose between parametric and non-parametric tests. When comparing the *level* of the inspectors, we will use a 2-sample *t-test*; however, when comparing the *category* of damage, we will use a paired analysis as the variables are dependent on the inspector. The results are summarized below as follows:

**Hypothesis 1.** *This hypothesis evaluates whether there is a significant difference in damage detection between experts and novices based on fixation and visit metrics. The conditions for the hypothesis are as follows:*

$$H_0: \mu_{\text{difference}} = 0$$

$$H_A: \mu_{\text{novices}} > \mu_{\text{experts}}$$

We find that there is insufficient evidence to claim a difference in damage detection between experts and novices based on the fixation count ( $t_{10}$ ,  $t = 0.13$ ,  $p = 0.8965$ ), visit duration ( $t_{10}$ ,  $t = -0.60$ ,  $p = 0.5598$ ), and visit count ( $t_{10}$ ,  $t = 0.04$ ,  $p = 0.9721$ ). Since the fixation duration for novices did not meet the condition for normality, a Wilcoxon rank-sum test will be conducted. This test determines if two independent samples come from the same distribution. The Wilcoxon test provides a  $p$ -value just like the  $t$ -tests performed earlier. If the  $p$ -value is less than the significance level of 0.05, we reject the null hypothesis that the independent samples have the same distribution.

Based on the Wilcoxon test, we found that there is no significant difference in damage detection between experts and novices based on fixation and visit metrics ( $Statistic = 19.5$ ,  $Z = -1.0208$ ,  $p = 0.3073$ ).

**Hypothesis 2.** *This hypothesis evaluates if there is a difference between fixation metrics for structural and non-structural damage categories. The conditions for the hypothesis are as follows:*

$$H_0: \mu_{\text{difference}} = 0$$

$$H_A: \mu_{\text{difference}} \neq 0$$

Since each participant has fixation metrics for structural and non-structural damages, these data are paired and not independent. However, each pairing is independent of every other pairing. The only other condition to check would be whether the differences between structural and non-structural fixation durations and fixation counts are normally distributed. A Wilcoxon signed-rank test will be performed for the paired fixation duration data and a paired  $t$ -test for the fixation counts. There appears to be strong evidence of a difference between structural and non-structural categories based on the fixation duration ( $S = -39$ ,  $p = 0.0005$ ) and fixation counts ( $t_{11}$ ,  $t = -8.27$ ,  $p < 0.0001$ ).

**Hypothesis 3.** *This hypothesis evaluates whether experts and novices have differences based on damage typologies and whether they spend longer fixations on structural elements (connection points, façade, and roof joints) than non-experts. The conditions for the hypothesis are as follows:*

$$H_0: \mu_{\text{difference}} = 0$$

$$H_{A1}: \mu_{\text{novices}} < \mu_{\text{experts}}$$

$$H_{A2}: \mu_{\text{structural}} > \mu_{\text{non-structural}}$$

We find that the fixation counts of novices are the only group of data that is not normally distributed ( $W = 0.8117$ ,  $p = 0.0381$ ). Therefore, a one-tailed, 2-sample  $t$ -test will be conducted for fixation duration, and a Wilcoxon rank-sum test will be conducted for the fixation count.

The results indicate that experts do not have longer fixation durations on structural elements compared to novice participants ( $t_{10}$ ,  $t = -1.21$ ,  $p = 0.1266$ ). However, there appears to be strong evidence to suggest experts have higher fixation counts on structural elements compared to novices ( $Statistic = 37$ ,  $Z = 1.7833$ ,  $p = 0.0373$ ).

Similarly, we noticed that these data are not independent, as the participants who are experts have measurements for structural elements and non-structural elements. Therefore, the interest is in the paired differences. Based on the Shapiro–Wilk test, both the differences between fixation durations ( $W = 0.9609$ ,  $p = 0.7846$ ) and differences in fixation counts ( $W = 0.9293$ ,  $p = 0.5903$ ) are normally distributed. Therefore, it is appropriate to perform a one-tailed, paired  $t$ -test for the differences in fixation duration and fixation count between structural elements and non-structural elements for the experts.

Strong evidence supports the claim that experts spend longer fixation durations on structural elements compared to non-structural elements ( $t_3$ ,  $t = 15.60$ ,  $p = 0.0006$ ).



There is also strong evidence to support the claim that experts have higher fixation counts on structural elements than non-structural elements ( $t_3$ ,  $t = 10.91$ ,  $p = 0.0016$ ).

## Discussion

**Hypothesis 1.** *Significant differences between experts and novices in detecting damages.*

Experts and novices often exhibit significant differences in visual search strategies due to developing domain-specific skills through experience. According to cognitive psychology, experts develop more efficient and effective visual search strategies, allowing them to quickly identify relevant information and ignore irrelevant details. This phenomenon is well-documented in medical imaging and airport security screening, where experts perform better than novices. For instance, studies in medical imaging have shown that radiologists, due to their extensive experience, can detect anomalies in medical images more rapidly and accurately than novices [87]. Similarly, in airport security screening, expert screeners can identify prohibited items in luggage more efficiently than those with less experience [88]. These differences arise from the experts' ability to develop sophisticated mental schemas and visual patterns that guide their attention to the most pertinent areas of an image.

In the context of damage detection, experts are likely to exhibit shorter fixation durations and fewer fixations on irrelevant areas compared to novices. Their visits to different regions of an image are more strategic, focusing primarily on critical structural elements. This results in a more efficient scanning pattern, enabling them to identify damage more effectively.

**Hypothesis 2.** *Differences in Structural and Non-Structural Damage Localization.*

The ability to distinguish between task-relevant and task-irrelevant information is critical to visual inspection tasks. Experienced inspectors are better equipped to prioritize their attention toward structural elements crucial for damage assessment. This is supported by theories of visual attention and cognitive load, which suggest that experts can manage their cognitive resources more effectively than novices. Research in visual attention indicates that experts can focus their attention on task-relevant information while ignoring distractions [89]. Cognitive load theory further explains that experts have a higher working memory capacity for domain-specific tasks, allowing them to process relevant information more efficiently without becoming overwhelmed by irrelevant details [90].

**Hypothesis 3.** *Influence of Damage Typology on Visual Attention Behavior.*

The influence of damage typology on visual attention behavior is closely related to theories of perceptual learning and expertise development. Perceptual learning theory posits that through repeated exposure and practice, individuals develop more refined mental representations of specific visual stimuli [91]. This process enables experts to recognize and categorize different types of damage more accurately and efficiently. Studies in medicine have demonstrated that experts in fields such as dermatology or pathology develop a heightened sensitivity to subtle visual cues associated with specific conditions [92]. These refined mental representations allow experts to allocate their attention more strategically, focusing on areas that are more likely to contain critical information based on the typology of the damage observed.

The significant differences in visual attention behavior between experts and novices, influenced by damage typology, underscore the critical role of perceptual learning in expertise development. Experts develop detailed mental representations of various damage types, enabling them to allocate their attention more effectively and accurately assess damage. This ability highlights the importance of targeted training and extensive experience in honing the skills necessary for reliable visual inspections. In summary, the observed differences across the three hypotheses—experts' efficient visual search strategies, refined attention control, and detailed mental representations—demonstrate that expertise is culti-

vated through experience and practice, emphasizing the need for specialized training to achieve superior damage detection and assessment performance.

#### 4.3. Saliency Maps Generation

Following our statistical exploration of eye tracking metrics, we utilized machine learning techniques to generate saliency maps. These maps are crucial, as they visually represent the predicted areas of interest based on the inspector's aggregated gaze data collected. By creating these maps, we aim to visually synthesize and highlight the key differences in gaze patterns between expert and novice inspectors, as revealed by our statistical analysis. This visualization validates our statistical findings and practically demonstrates how AI can simulate and enhance human visual inspection capabilities. Saliency maps are instrumental in training novice inspectors by visually guiding them towards the most salient features of a scene, as commonly focused on by expert inspectors. This is directly aligned with our research question, as it provides a way to bridge the experience gap between different levels of inspectors. Moreover, these maps can be integrated into robotic systems, enhancing their utility in assisting human inspectors by directing their attention to critical areas, thus facilitating a more efficient and accurate inspection.

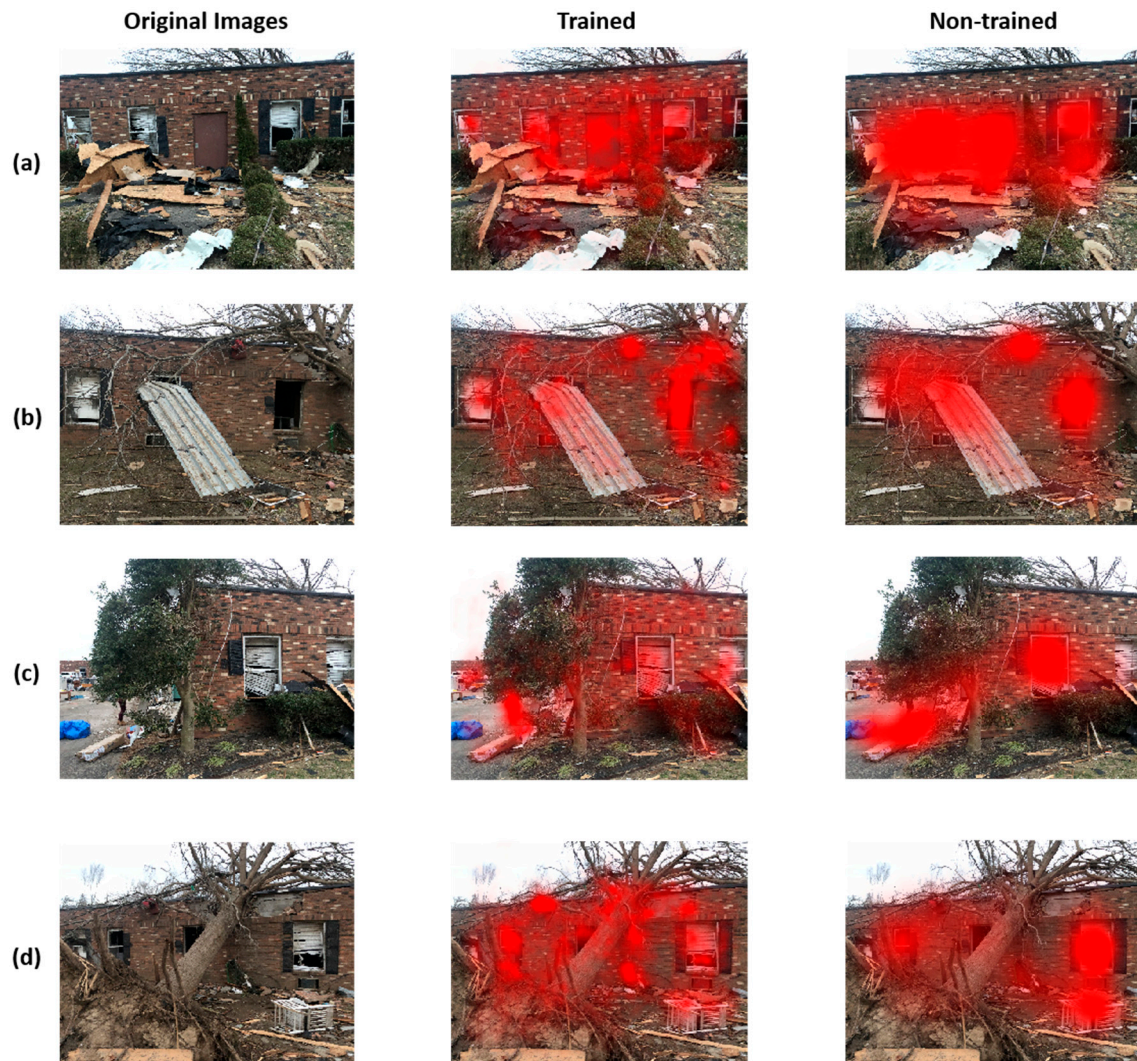
In this study, we utilized the encoder–decoder model by Kroner et al. [76] and trained it on our eye tracking dataset containing disaster site images to characterize and predict the model's output. We then qualitatively compared the results on an independent test dataset to distinguish how well our trained model generalizes to unseen images. We are interested in estimating saliency predictions based on the inspector's gaze data to identify which saliency features are important from the inspectors' viewpoint. By comparing the performance of the non-trained conventional saliency predictor with our trained saliency predictor, we find that the inspector's eye movements hold significant information and details.

Figure 10 indicates the visual illustrations of saliency prediction on test images collected from a disaster site. The test images were taken from the nursing home building affected by the tornado, and the images under consideration were completely unseen by our model. The goal of our saliency predictor is to highlight the damage areas that were relevant or considered important by the building inspectors instead of localizing the salient features of an image that might not be considered crucial in assessing the structure's condition. Figure 10a indicates the model output, and the salient features are the windows and façade of the structure, with surrounding areas and building rubble slightly highlighted. On the contrary, the non-trained model predicted the central window's region and the wooden plywood as the most salient region in the image without impacting structural stability or integrity. Following this, we notice in Figure 10b that the predicted model highlights the windows region, the steel plate hanging on a tree branch, and the tree in the top right corner that fell off the structure, damaging the roof and façade. We notice that the non-trained model cannot identify the tree's impact on the structure, while windows are considered the most important area of interest.

We tested various images with different vantage points to see how the model would respond to the change in view or angle. Figure 10c indicates the building corner side, and the model's performance indicates higher saliency near the tree due to the presence of a person, followed by the building façade with lower saliency. However, the non-trained model shows higher saliency in the window and surrounding regions, with debris on the ground. Lastly, Figure 10d provides a full view of the fallen tree and the damage it caused to the structure. The trained model utilizing the inspector's gaze information predicts saliency at the windows, trees, and façade damage caused. In contrast, the untrained model only highlights the window regions and rubble of the structure.

Interestingly, the trained model shows saliency prediction with varied intensity levels on the structure with different regions such as broken windows, façade damage due to a fallen-off tree and steel plates, etc. We notice that in practical scenarios, these are the damages that capture the inspector's attention, and they are interested in knowing how

these different elements tie together and the connection between them. However, the non-trained model only predicts the more obvious regions, such as windows and building rubble in the surroundings. It can be argued that conventional saliency predictors rely on simple pixel distributions in an image and changes in its intensity due to lighting or sharp color enhancement compared to understanding the scene's context. This could explain why Windows has higher saliency for non-trained models since they have higher changes in pixel distribution or edge detection techniques [93,94].



**Figure 10.** Comparison of saliency predictor performance on the original test set (a–d: unseen images from disaster site) for the ML model trained on gaze data and the non-trained SALICON model [75].

#### 4.4. Saliency Metrics Evaluation

Evaluating the performance of these saliency maps is the next step in assessing how well the saliency predictions match the actual fixation data (ground truth). By doing so, we gain insights into the effectiveness and accuracy of our machine learning models. These evaluations suggest areas for further research and development, particularly in improving the AI models to better capture the nuances of human visual attention in disaster assessment scenarios. Different methods have been used in the literature to assess the predictive performance of saliency maps. As mentioned in Table 1, we used five commonly used metrics to evaluate the performance of trained saliency maps. The results are summarized in Appendix A of the paper, which contains detailed findings for all the images and a comparison among different metrics.



The AUC–Borji metric assesses the predictive accuracy of a saliency map by determining how effectively it captures ground truth fixations across various threshold levels. The saliency map functions as a binary classifier for fixations at multiple thresholds to compute the AUC, generating a ROC curve. Each thresholding operation on the saliency map creates a level set, with the true-positive rate defined by the proportion of fixations within the level set and the false-positive rate defined by the proportion of non-fixated image pixels within the level set. The AUC score is then derived from the area under this ROC curve. Judd et al. [84] introduced an AUC variant, known as AUC–Borji, which is visualized alongside the ROC curve to evaluate the saliency map’s classification performance. For example, the highest score for the dataset is 0.58, which is a marginally acceptable score. Conversely, models predicting only the center achieve lower scores, around 0.5, due to their inability to differentiate true positives from false positives.

NSS (normalized scanpath saliency) evaluates the similarity between a saliency map and a fixation map (human ground truth) based on fixation locations. A higher NSS score indicates greater similarity, akin to the AUC metric. NSS normalizes the saliency map by the standard deviation of its values. NSS achieves 0.69, the highest score among all images, showing great results and a higher compute score. SIM is widely used in the saliency community to compare saliency maps by computing the sum of pixel minimums between the predicted saliency map and the ground truth human fixations. A similarity score of 1 indicates perfect similarity between the predicted map and the ground truth. The CC metric, on the other hand, measures the linear correlation between saliency and fixation maps, with scores ranging from  $-1$  to  $+1$ . While CC treats false positives and negatives symmetrically, SIM emphasizes false negatives more than false positives. Consequently, the saliency maps often exhibit low SIM and CC scores, indicating a negative correlation with the ground truth fixation map. Unlike SIM and CC, the KLdiv metric assesses dissimilarity between the saliency map and ground truth data and is particularly sensitive to false negatives.

## 5. Discussion and Limitations

The differences between expert and novice building inspectors are significant, particularly in their approaches to identifying and assessing structural issues. *Expert* inspectors spend considerable time analyzing the building, focusing on major structural principles and looking for patterns that indicate potential problems. They organize information based on key concepts and frequently reevaluate their findings to ensure accuracy. Experts typically move from known issues to unknowns, building on their extensive knowledge base to explore new areas of concern. However, they may struggle with inflexibility and resistance to new methods or technologies due to their established expertise. On the other hand, *Novice* inspectors often begin their assessments by taking immediate action, focusing on tactical issues and quick solutions. They tend to highlight specific details without understanding the broader context, relying heavily on their initial impressions and findings. Novices move from unknowns to givens, gradually building their understanding through hands-on experience and adaptation to new procedures. Their flexibility and willingness to adapt can be advantageous, especially in dynamic or unfamiliar inspection environments.

Both groups are susceptible to overconfidence, which can lead to oversight and errors. Experts might become overly reliant on their experience and established methods, while novices might overestimate their understanding or the effectiveness of their quick solutions. Recognizing these tendencies is crucial for both groups to mitigate potential pitfalls. By acknowledging these differences, building inspection teams can leverage experts’ thorough analysis and deep knowledge alongside novices’ adaptability and fresh perspectives. This balanced approach can enhance the overall effectiveness and accuracy of building inspections, ensuring that structural issues are identified and addressed comprehensively. Encouraging continuous learning and the integration of new methods and technologies can further improve inspection outcomes, fostering a culture of excellence and innovation in building inspection.



Despite the findings, this study has several limitations that must be addressed. Although safe and useful for data collection, the controlled, screen-based inspection method does not fully capture the complexities and dynamics of real-world disaster sites. This method fails to replicate the multi-sensory and chaotic environment of actual disaster scenarios, where factors such as noise, physical obstructions, and environmental hazards play critical roles. This limitation is crucial, especially considering the significant differences in planning and problem-solving approaches between expert and novice inspectors in real scenarios. The controlled environment may also influence the inspectors' behavior and decision-making processes, potentially skewing the results. Incorporating field conditions will allow researchers to observe inspectors' behaviors and decision-making processes in realistic settings, providing a more accurate representation of their capabilities and limitations. To mitigate these variables, we carefully designed the experimental setup by selecting a diverse set of disaster images representing various structural damage types. Inspectors were given standardized instructions to minimize variability in task understanding, and the eye tracking equipment was calibrated to ensure consistent data quality. Additionally, we accounted for individual differences in visual attention and cognitive load through detailed statistical analysis. Metrics such as fixation duration, fixation count, visit duration, and visit count were used to compare expert and novice inspectors. These measures were analyzed using parametric and non-parametric statistical tests to discern patterns indicative of experience level and independent of the controlled environment. For example, the use of the Shapiro–Wilk and Levene's tests helped ensure the normality and homogeneity of our data, respectively.

We also addressed the potential impact of environmental variables by ensuring a consistent testing environment. All participants performed the tasks under the same lighting conditions and used the same equipment setup to eliminate external factors that could influence visual attention. The visual stimuli were presented in a randomized order to each participant to prevent order effects. The controlled environment may also influence the inspectors' behavior and decision-making processes, potentially skewing the results. Incorporating field conditions will allow researchers to observe inspectors' behaviors and decision-making processes in realistic settings, providing a more accurate representation of their capabilities and limitations. While our controlled setup provided a robust framework for this proof-of-concept study, we acknowledge its limitations in replicating the complexities of real-world disaster scenarios. Future research will incorporate more dynamic and varied environments to simulate real-world conditions better. By expanding our research to include real-world assessments, we aim to validate our current findings and enhance the applicability of our model in actual disaster scenarios. This approach will help ensure that the technology and methodologies we develop are effective and reliable in practical, on-the-ground applications.

Additionally, the relatively small sample size might affect the generalizability of the findings. A limited sample size restricts the diversity of the data, potentially leading to biased conclusions that do not accurately reflect the broader population of building inspectors. This limitation is particularly pertinent when examining the varied approaches of experts and novices, as a more extensive and diverse sample could reveal a wider range of strategies and behaviors. The findings may, therefore, lack robustness and applicability across different geographic regions, building types, and inspection conditions.

## 6. Conclusions

In conclusion, this research successfully demonstrated the application of eye tracking technology to enhance the understanding and methodologies of structural inspections following disasters. By analyzing the gaze patterns of building inspectors at a disaster site in Mayfield, Kentucky, we captured and quantified the cognitive processes involved during visual assessments. Our findings highlight significant differences in visual attention strategies between experienced and novice inspectors, indicating that expertise plays a crucial role in the efficiency and focus of visual inspections. Specifically, experts displayed

more targeted and sustained fixations, suggesting a strategic approach to identifying critical issues quickly. At the same time, novices exhibited a broader and more exploratory gaze pattern, indicating less efficiency in pinpointing salient features. The potential application of this work can be used in training novice inspectors to teach how efficient structural assessment and damage detection can be done. Another potential avenue from an industry perspective is to design a course or exam that differentiates and finds whether a participant is an expert or novice, which is also important in many ways.

The integration of eye tracking data into traditional inspection methods holds substantial promise for refining and advancing the inspection process. By distinguishing between the visual engagement patterns of novices and experts, we can tailor training programs to better equip novice inspectors with the skills necessary for effective visual assessments. Moreover, the development of saliency maps based on expert gaze data offers a promising avenue for automating and enhancing the accuracy of damage assessments, particularly in post-disaster scenarios where rapid and reliable evaluations are crucial.

In future research, this work can be extended to involve larger and more diverse samples, incorporating real-life field conditions to validate and extend the findings. Engaging a broader set of expert building inspectors with diverse backgrounds can enhance the generalizability and reliability of the results. Exploring the potential of combining human expertise with automated systems is also crucial. By leveraging machine learning models trained on expert gaze data, we can develop automated systems capable of generating accurate saliency maps. This human-machine collaboration could significantly enhance the efficiency and effectiveness of post-disaster assessments. Automated systems can process large volumes of data quickly, identifying critical areas of damage that might be missed by human inspectors alone. This collaboration could facilitate rapid and reliable damage assessments, enabling faster response times and more informed decision-making during disaster recovery efforts.

Advancements in this field will contribute to more resilient infrastructure management practices. By integrating cutting-edge technology with expert knowledge, we can develop innovative solutions that improve public safety and disaster preparedness. Future research should focus on refining these automated systems, ensuring they can adapt to various disaster scenarios and seamlessly integrate with human inspectors' workflows. This approach will ultimately enhance the overall process of disaster response, ensuring timely and safe recovery efforts. Through continued exploration and innovation, we can build a more resilient and responsive infrastructure capable of withstanding the challenges posed by natural disasters.

**Author Contributions:** M.R.S. and R.N. contributed to the conception and the design of the study. M.R.S. performed the experiments, collected data, and prepared all figures. M.R.S. and R.M. carried out the statistical analysis. M.R.S. contributed to the interpretation of results and wrote the manuscript. M.R.S. and R.N. contributed to the revision of the manuscript. All authors have read and agreed to the published version of the manuscript.

**Funding:** This material is based upon work supported by the National Science Foundation under Grant No. BCS 2121909 and IIS 2123343.

**Data Availability Statement:** Data will be made available upon request.

**Acknowledgments:** Any opinions, findings, conclusions, or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation. The authors would also like to thank all the participants who obtained the data for this study and provided insights based on their experience and knowledge.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## Appendix A

**Table A1.** Evaluation report of saliency metrics.

Image	AUC_Borjii	CC	Kldiv	NSS	Similarity
'001.jpg'	0.5107	0.2985	0.9812	0.3780	0.4246
'002.jpg'	0.5083	0.6298	0.6118	0.3629	0.5886
'003.jpg'	0.5060	0.6963	0.6139	0.3319	0.6031
'004.jpg'	0.5111	0.6817	0.6302	0.4798	0.5826
'005.jpg'	0.5558	0.4323	0.9548	0.4666	0.4394
'006.jpg'	0.5117	0.4161	0.8816	0.3530	0.4748
'007.jpg'	0.5007	0.0194	1.1577	0.0961	0.4214
'008.jpg'	0.5022	0.0637	1.3890	0.1144	0.3859
'009.jpg'	0.5067	0.1239	0.7674	0.1349	0.5055
'010.jpg'	0.4978	−0.0172	0.8773	0.0487	0.4656
'011.jpg'	0.4976	−0.0749	1.1147	−0.0285	0.4133
'012.jpg'	0.5354	0.3722	0.8758	0.2810	0.4811
'013.jpg'	0.5420	0.4435	1.1080	0.6698	0.3795
'014.jpg'	0.5131	0.3354	1.0725	0.4300	0.3919
'015.jpg'	0.5572	0.3021	1.1867	0.5148	0.3664
'016.jpg'	0.5735	0.5193	0.8774	0.6189	0.4504
'017.jpg'	0.5034	0.7012	0.4577	0.4238	<b>0.6564</b>
'018.jpg'	0.5115	0.6130	<b>0.4120</b>	0.3658	0.6464
'019.jpg'	0.5388	0.7089	0.6023	0.6567	0.5988
'020.jpg'	0.5124	0.2791	0.7376	0.1814	0.5274
'021.jpg'	0.5065	0.1437	1.0380	0.2047	0.4418
'022.jpg'	0.5157	0.3005	1.1495	0.3141	0.3846
'023.jpg'	0.5040	0.3058	0.9354	0.1943	0.4497
'024.jpg'	0.5148	0.6162	0.5539	0.4713	0.5972
'025.jpg'	0.5116	<b>0.7192</b>	0.5380	<b>0.6929</b>	0.6460
'026.jpg'	0.5018	0.5980	0.7224	0.6270	0.5730
'027.jpg'	0.5178	0.5857	0.5507	0.5179	0.6009
'028.jpg'	0.5023	0.5821	0.5949	0.4383	0.5858
'029.jpg'	0.5445	0.3543	1.0044	0.4851	0.4652
'030.jpg'	0.5048	0.4651	1.0066	0.5149	0.4196
'031.jpg'	<b>0.5888</b>	0.4427	0.8703	0.5211	0.4601
'032.jpg'	0.5456	0.3815	1.0430	0.4713	0.3982
'033.jpg'	0.5313	0.2376	1.2683	0.5465	0.3608
Average	0.5208	0.4023	0.8662	0.3903	0.4905

Note: Bold letters indicate the best results and highest score for that saliency metric.

## References

1. Khan, A.; Gupta, S.; Gupta, S.K. Multi-hazard disaster studies: Monitoring, detection, recovery, and management, based on emerging technologies and optimal techniques. *Int. J. Disaster Risk Reduct.* **2020**, *47*, 101642. [CrossRef]
2. Benson, C.; Edward, J.C. Economic and Financial Impacts of Natural Disasters: An Assessment of Their Effects and Options for Mitigation. Economics and Environmental Science. Available online: <https://www.semanticscholar.org/paper/Economic-and-Financial-Impacts-of-Natural-an-of-and-Benson-Clay/a04c5f181b292050dddf011d50863872b7e52e6a> (accessed on 4 June 2024).
3. Chang, C.-M.; Lin, T.-K.; Moreu, F.; Singh, D.K.; Hoskore, V. Post Disaster Damage Assessment Using Ultra-High-Resolution Aerial Imagery with Semi-Supervised Transformers. *Sensors* **2023**, *23*, 8235. [CrossRef] [PubMed]
4. ATC-20. Available online: <https://www.atcouncil.org/atc-20> (accessed on 12 February 2024).
5. Preliminary Damage Assessments | FEMA.gov. Available online: <https://www.fema.gov/disaster/how-declared/preliminary-damage-assessments#report-guide> (accessed on 12 February 2024).
6. Varghese, S.; Hoskore, V. Unpaired image-to-image translation of structural damage. *Adv. Eng. Inform.* **2023**, *56*, 101940. [CrossRef]
7. Mishra, M.; Lourenço, P.B. Artificial intelligence-assisted visual inspection for cultural heritage: State-of-the-art review. *J. Cult. Herit.* **2024**, *66*, 536–550. [CrossRef]

8. McRae, J.N.; Nielsen, B.M.; Gay, C.J.; Hunt, A.P.; Nigh, A.D. Utilizing Drones to Restore and Maintain Radio Communication During Search and Rescue Operations. *Wilderness Environ. Med.* **2021**, *32*, 41–46. [\[CrossRef\]](#) [\[PubMed\]](#)
9. Zwegliński, T. The Use of Drones in Disaster Aerial Needs Reconnaissance and Damage Assessment—Three-Dimensional Modeling and Orthophoto Map Study. *Sustainability* **2020**, *12*, 6080. [\[CrossRef\]](#)
10. Saleem, M.R.; Mayne, R.; Napolitano, R. Analysis of gaze patterns during facade inspection to understand inspector sense-making processes. *Sci. Rep.* **2023**, *13*, 2929. [\[CrossRef\]](#) [\[PubMed\]](#)
11. Máthé, K.; Buşoniu, L. Vision and Control for UAVs: A Survey of General Methods and of Inexpensive Platforms for Infrastructure Inspection. *Sensors* **2015**, *15*, 14887–14916. [\[CrossRef\]](#) [\[PubMed\]](#)
12. Narazaki, Y.; Hoskere, V.; Chowdhary, G.; Spencer, B.F. Vision-based navigation planning for autonomous post-earthquake inspection of reinforced concrete railway viaducts using unmanned aerial vehicles. *Autom. Constr.* **2022**, *137*, 104214. [\[CrossRef\]](#)
13. Bolourian, N.; Hammad, A. LiDAR-equipped UAV path planning considering potential locations of defects for bridge inspection. *Autom. Constr.* **2020**, *117*, 103250. [\[CrossRef\]](#)
14. Mirzaei, K.; Arashpour, M.; Asadi, E.; Feng, H.; Mohandes, S.R.; Bazli, M. Automatic compliance inspection and monitoring of building structural members using multi-temporal point clouds. *J. Build. Eng.* **2023**, *72*, 106570. [\[CrossRef\]](#)
15. Xu, Y.; Brownjohn, J.M.W. Review of machine-vision based methodologies for displacement measurement in civil structures. *J. Civ. Struct. Health Monit.* **2018**, *8*, 91–110. [\[CrossRef\]](#)
16. Li, L.; Betti, R. A machine learning-based data augmentation strategy for structural damage classification in civil infrastructure system. *J. Civ. Struct. Health Monit.* **2023**, *13*, 1265–1285. [\[CrossRef\]](#)
17. Atha, D.J.; Jahanshahi, M.R. Evaluation of deep learning approaches based on convolutional neural networks for corrosion detection. *Struct. Health Monit.* **2018**, *17*, 1110–1128. [\[CrossRef\]](#)
18. Cha, Y.J.; Choi, W.; Büyüköztürk, O. Deep Learning-Based Crack Damage Detection Using Convolutional Neural Networks. *Comput. Aided Civ. Infrastruct. Eng.* **2017**, *32*, 361–378. [\[CrossRef\]](#)
19. Liu, C.; Man, J.; Liu, C.; Wang, L.; Ma, X.; Miao, J.; Liu, Y. Research on damage identification of large-span spatial structures based on deep learning. *J. Civ. Struct. Health Monit.* **2024**, *14*, 1035–1058. [\[CrossRef\]](#)
20. Moore, M.; Phares, B.; Graybeal, B.; Rolander, D.; Washer, G. Reliability of Visual Inspection for Highway Bridges. *McLean*. 2001. Available online: [https://www.researchgate.net/publication/273680040\\_Reliability\\_of\\_Visual\\_Inspection\\_for\\_Highway\\_Bridges](https://www.researchgate.net/publication/273680040_Reliability_of_Visual_Inspection_for_Highway_Bridges) (accessed on 2 February 2024).
21. Wang, Z.; Cha, Y.J. Unsupervised deep learning approach using a deep auto-encoder with a one-class support vector machine to detect damage. *Struct. Health Monit.* **2021**, *20*, 406–425. [\[CrossRef\]](#)
22. Keskin, M.; Ooms, K.; Dogru, A.O.; De Maeyer, P. Exploring the Cognitive Load of Expert and Novice Map Users Using EEG and Eye Tracking. *ISPRS Int. J. Geo-Inf.* **2020**, *9*, 429. [\[CrossRef\]](#)
23. Bruder, C.; Hasse, C. Differences between experts and novices in the monitoring of automated systems. *Int. J. Ind. Ergon.* **2019**, *72*, 1–11. [\[CrossRef\]](#)
24. Hosking, S.G.; Liu, C.C.; Bayly, M. The visual search patterns and hazard responses of experienced and inexperienced motorcycle riders. *Accid. Anal. Prev.* **2010**, *42*, 196–202. [\[CrossRef\]](#)
25. Silva, A.F.; Afonso, J.; Sampaio, A.; Pimenta, N.; Lima, R.F.; Castro, H.d.O.; Ramirez-Campillo, R.; Teoldo, I.; Sarmento, H.; Fernández, F.G.; et al. Differences in visual search behavior between expert and novice team sports athletes: A systematic review with meta-analysis. *Front. Psychol.* **2022**, *13*, 1001066. [\[CrossRef\]](#) [\[PubMed\]](#)
26. Takamido, R.; Kurihara, S.; Umeda, Y.; Asama, H.; Kasahara, S.; Tanaka, Y.; Fukumoto, S.; Kato, T.; Korenaga, M.; Hoshi, M.; et al. Evaluation of expert skills in refinery patrol inspection: Visual attention and head positioning behavior. *Heliyon* **2022**, *8*, e12117. [\[CrossRef\]](#) [\[PubMed\]](#)
27. Wang, S.; Ouyang, X.; Liu, T.; Wang, Q.; Shen, D. Follow My Eye: Using Gaze to Supervise Computer-Aided Diagnosis. *IEEE Trans. Med. Imaging* **2022**, *41*, 1688–1698. [\[CrossRef\]](#) [\[PubMed\]](#)
28. Brunyé, T.T.; Carney, P.A.; Allison, K.H.; Shapiro, L.G.; Weaver, D.L.; Elmore, J.G. Eye Movements as an Index of Pathologist Visual Expertise: A Pilot Study. *PLoS ONE* **2014**, *9*, e103447. [\[CrossRef\]](#)
29. Ritchie, M.J.; Parker, L.E.; Kirchner, J.A.E. From novice to expert: Methods for transferring implementation facilitation skills to improve healthcare delivery. *Implement. Sci. Commun.* **2021**, *2*, 39. [\[CrossRef\]](#) [\[PubMed\]](#)
30. Brunyé, T.T.; Nallamotheu, B.K.; Elmore, J.G. Eye-tracking for assessing medical image interpretation: A pilot feasibility study comparing novice vs expert cardiologists. *Perspect. Med. Educ.* **2019**, *8*, 65–73. [\[CrossRef\]](#) [\[PubMed\]](#)
31. Feng, Z.; González, V.A.; Amor, R.; Lovreglio, R.; Cabrera-Guerrero, G. Immersive virtual reality serious games for evacuation training and research: A systematic literature review. *Comput. Educ.* **2018**, *127*, 252–266. [\[CrossRef\]](#)
32. Hsu, E.B.; Li, Y.; Bayram, J.D.; Levinson, D.; Yang, S.; Monahan, C. State of virtual reality based disaster preparedness and response training. *PLoS Curr.* **2013**, *5*, ecurrents.dis.1ea2b2e71237d5337fa53982a38b2aff. [\[CrossRef\]](#) [\[PubMed\]](#)
33. Hartwig, M.; Bond, C.F. Why do lie-catchers fail? A lens model meta-analysis of human lie judgments. *Psychol. Bull.* **2011**, *137*, 643–659. [\[CrossRef\]](#)
34. Granhag, P.A.; Rangmar, J.; Strömwall, L.A. Small Cells of Suspects: Eliciting Cues to Deception by Strategic Interviewing. *J. Investig. Psychol. Offender Profiling* **2015**, *12*, 127–141. [\[CrossRef\]](#)



35. Dimoka, A.; Davis, F.D.; Gupta, A.; Pavlou, P.A.; Banker, R.D.; Dennis, A.R.; Ischebeck, A.; Müller-Putz, G.; Benbasat, I.; Gefen, D.; et al. On the use of neurophysiological tools in research: Developing a research agenda for neurois. *MIS Q* **2012**, *36*, 679–702. [\[CrossRef\]](#)
36. Sun, Z.K.; Wang, J.Y.; Luo, F. Experimental pain induces attentional bias that is modified by enhanced motivation: An eye tracking study. *Eur. J. Pain* **2016**, *20*, 1266–1277. [\[CrossRef\]](#) [\[PubMed\]](#)
37. Causse, M.; Lancelot, F.; Maillant, J.; Behrend, J.; Cousy, M.; Schneider, N. Encoding decisions and expertise in the operator's eyes: Using eye-tracking as input for system adaptation. *Int. J. Hum. Comput. Stud.* **2019**, *125*, 55–65. [\[CrossRef\]](#)
38. Guazzini, A.; Yoneki, E.; Gronchi, G. Cognitive dissonance and social influence effects on preference judgments: An eye tracking based system for their automatic assessment. *Int. J. Hum. Comput. Stud.* **2015**, *73*, 12–18. [\[CrossRef\]](#)
39. Li, J.; Li, H.; Umer, W.; Wang, H.; Xing, X.; Zhao, S.; Hou, J. Identification and classification of construction equipment operators' mental fatigue using wearable eye-tracking technology. *Autom. Constr.* **2020**, *109*, 103000. [\[CrossRef\]](#)
40. Seinfeld, S.; Feuchtner, T.; Maselli, A.; Müller, J. User Representations in Human-Computer Interaction. *Hum. Comput. Interact.* **2021**, *36*, 400–438. [\[CrossRef\]](#)
41. Egeth, H.E.; Yantis, S. Visual Attention: Control, Representation, and Time Course. *Annu. Rev. Psychol.* **1997**, *48*, 269–297. [\[CrossRef\]](#)
42. Kaspar, K. What Guides Visual Overt Attention under Natural Conditions? Past and Future Research. *Int. Sch. Res. Not.* **2013**, *2013*, 868491. [\[CrossRef\]](#)
43. Lavie, N. Perceptual Load as a Necessary Condition for Selective Attention. *J. Exp. Psychol. Hum. Percept. Perform.* **1995**, *21*, 451–468. [\[CrossRef\]](#)
44. Itti, L.; Koch, C. A saliency-based search mechanism for overt and covert shifts of visual attention. *Vis. Res.* **2000**, *40*, 1489–1506. [\[CrossRef\]](#)
45. Koch, C.; Ullman, S. Shifts in Selective Visual Attention: Towards the Underlying Neural Circuitry. *Hum. Neurobiol.* **1987**, *4*, 115–141. [\[CrossRef\]](#)
46. Geisler, W.S.; Cormack, L.K. Models of overt attention. In *The Oxford Handbook of Eye Movements*; Oxford Academic: Oxford, UK, 2011. [\[CrossRef\]](#)
47. Tilke, J.; Ehinger, K.; Durand, F.; Torralba, A. Learning to predict where humans look. In Proceedings of the IEEE International Conference on Computer Vision, Kyoto, Japan, 29 September–2 October 2009; pp. 2106–2113. [\[CrossRef\]](#)
48. Ramanathan, S.; Katti, H.; Huang, R.; Chua, T.S.; Kankanhalli, M. Automated localization of affective objects and actions in images via caption text-cum-eye gaze analysis. In Proceedings of the MM'09—Proceedings of the 2009 ACM Multimedia Conference, with Co-located Workshops and Symposia, Beijing, China, 19–24 October 2009; pp. 729–732. [\[CrossRef\]](#)
49. Takeichi, N.; Katagiri, T.; Yoneda, H.; Inoue, S.; Shintani, Y. Virtual Reality approaches for evacuation simulation of various disasters. *Collect. Dyn.* **2020**, *5*, 534–536. [\[CrossRef\]](#)
50. Lovreglio, R. Virtual and Augmented Reality for Human Behaviour in Disasters: A Review. In Proceedings of the Fire and Evacuation Modeling Technical Conference (FEMTC), Virtual, 9–11 September 2020; Available online: [https://www.researchgate.net/publication/343809101\\_Virtual\\_and\\_Augmented\\_Reality\\_for\\_Human\\_Behaviour\\_in\\_Disasters\\_A\\_Review](https://www.researchgate.net/publication/343809101_Virtual_and_Augmented_Reality_for_Human_Behaviour_in_Disasters_A_Review) (accessed on 13 June 2024).
51. Lovreglio, R.; Duan, X.; Rahouti, A.; Phipps, R.; Nilsson, D. Comparing the effectiveness of fire extinguisher virtual reality and video training. *Virtual Real.* **2021**, *25*, 133–145. [\[CrossRef\]](#)
52. Li, C.; Liang, W.; Quigley, C.; Zhao, Y.; Yu, L.F. Earthquake Safety Training through Virtual Drills. *IEEE Trans. Vis. Comput. Graph.* **2017**, *23*, 1275–1284. [\[CrossRef\]](#) [\[PubMed\]](#)
53. Kashiwayama, K.; Ling, G.; Matsumoto, J. Modeling and Simulation of Tsunami Using Virtual Reality Technology. Videos of Plenary Lectures presented at the VI International Conference on Coupled Problems in Science and Engineering (COUPLED PROBLEMS 2015). 2016. Available online: [https://www.scipedia.com/public/Contents\\_2016ag](https://www.scipedia.com/public/Contents_2016ag) (accessed on 13 June 2024).
54. Chittaro, L. Passengers' Safety in Aircraft Evacuations: Employing Serious Games to Educate and Persuade. In *Lecture Notes in Computer Science*; Springer: Berlin/Heidelberg, Germany, 2012; Volume 7284, pp. 215–226. [\[CrossRef\]](#)
55. Lovreglio, R.; Ngassa, D.-C.; Rahouti, A.; Paes, D.; Feng, Z.; Shipman, A. Prototyping and Testing a Virtual Reality Counterterrorism Serious Game for Active Shooting. *SSRN Electron. J.* **2021**. [\[CrossRef\]](#)
56. Gamberini, L.; Bettelli, A.; Benvegnù, G.; Orso, V.; Spagnoli, A.; Ferri, M. Designing 'Safer Water.' A Virtual Reality Tool for the Safety and the Psychological Well-Being of Citizens Exposed to the Risk of Natural Disasters. *Front. Psychol.* **2021**, *12*, 674171. [\[CrossRef\]](#) [\[PubMed\]](#)
57. Fujimi, T.; Fujimura, K. Testing public interventions for flash flood evacuation through environmental and social cues: The merit of virtual reality experiments. *Int. J. Disaster Risk Reduct.* **2020**, *50*, 101690. [\[CrossRef\]](#)
58. Sermet, Y.; Demir, I. Flood action VR: A virtual reality framework for disaster awareness and emergency response training. In Proceedings of the SIGGRAPH '19: ACM SIGGRAPH 2019 Posters, Los Angeles, CA, USA, 28 July–1 August 2019. [\[CrossRef\]](#)
59. Lecun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [\[CrossRef\]](#) [\[PubMed\]](#)
60. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet Classification with Deep Convolutional Neural Networks. In Proceedings of the Advances in Neural Information Processing Systems 25 (NIPS 2012), Lake Tahoe, NV, USA, 3–6 December 2012.
61. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778. [\[CrossRef\]](#)

62. Matikainen, L.; Lehtomäki, M.; Ahokas, E.; Hyypä, J.; Karjalainen, M.; Jaakkola, A.; Kukko, A.; Heinonen, T. Remote sensing methods for power line corridor surveys. *ISPRS J. Photogramm. Remote Sens.* **2016**, *119*, 10–31. [\[CrossRef\]](#)
63. Chen, S.; Laefer, D.F.; Mangina, E.; Zolanvari, S.M.I.; Byrne, J. UAV Bridge Inspection through Evaluated 3D Reconstructions. *J. Bridge Eng.* **2019**, *24*, 05019001. [\[CrossRef\]](#)
64. Murphy, R.R.; Stover, S. Rescue robots for mudslides: A descriptive study of the 2005 La Conchita mudslide response. *J. Field Robot.* **2008**, *25*, 3–16. [\[CrossRef\]](#)
65. Goodrich, M.A.; Schultz, A.C. Human–Robot Interaction: A Survey. *Found. Trends® Hum. Comput. Interact.* **2008**, *1*, 203–275. [\[CrossRef\]](#)
66. Flach, P. Machine Learning: The Art and Science of Algorithms that Make Sense of Data by Peter Flach. Cambridge University Press. Available online: <http://people.cs.bris.ac.uk/~flach/mlbook/> (accessed on 13 June 2024).
67. Rudin, C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nat. Mach. Intell.* **2019**, *1*, 206–215. [\[CrossRef\]](#)
68. Donsa, K.; Spat, S.; Beck, P.; Pieber, T.R.; Holzinger, A. Towards Personalization of Diabetes Therapy Using Computerized Decision Support and Machine Learning: Some Open Problems and Challenges. In *Lecture Notes in Computer Science*; Springer: Berlin/Heidelberg, Germany, 2015; Volume 8700, pp. 237–260. [\[CrossRef\]](#)
69. Sowah, R.A.; Bampoe-Addo, A.A.; Armoo, S.K.; Saalia, F.K.; Gatsi, F.; Sarkodie-Mensah, B. Design and Development of Diabetes Management System Using Machine Learning. *Int. J. Telemed. Appl.* **2020**, *2020*, 8870141. [\[CrossRef\]](#) [\[PubMed\]](#)
70. Mamykina, L.; Epstein, D.A.; Klasnja, P.; Spruijt-Metz, D.; Meyer, J.; Czerwinski, M.; Althoff, T.; Choe, E.K.; De Choudhury, M.; Lim, B. Grand Challenges for Personal Informatics and AI. In Proceedings of the CHI EA '22: Extended Abstracts of the 2022 CHI Conference on Human Factors in Computing Systems, New Orleans, LA, USA, 29 April–5 May 2022. [\[CrossRef\]](#)
71. Tobii, A.B. Tobii Pro Nano. Available online: [https://www.tobii.com/products/discontinued/tobii-pro-nano?creative=642408166205&keyword=tobii%20pro&matchtype=p&network=g&device=c&utm\\_source=google&utm\\_medium=cpc&utm\\_campaign=&utm\\_term=tobii%20pro&gad\\_source=1&gclid=CjwKCAjwvIWzBhAlEiwAHHWgvXxCQj1eg-gN4\\_615kH8Qk84Cru1ENPhQ25pZlqojwLO\\_JoL5BRWeBoCWToQAvD\\_BwE](https://www.tobii.com/products/discontinued/tobii-pro-nano?creative=642408166205&keyword=tobii%20pro&matchtype=p&network=g&device=c&utm_source=google&utm_medium=cpc&utm_campaign=&utm_term=tobii%20pro&gad_source=1&gclid=CjwKCAjwvIWzBhAlEiwAHHWgvXxCQj1eg-gN4_615kH8Qk84Cru1ENPhQ25pZlqojwLO_JoL5BRWeBoCWToQAvD_BwE) (accessed on 5 June 2024).
72. Tobii, A.B. Tobii Pro Lab. Computer Software. 2014. Available online: <http://www.tobiipro.com/> (accessed on 31 January 2022).
73. Kaushal, S.; Soto, M.G.; Napolitano, R. Understanding the Performance of Historic Masonry Structures in Mayfield, KY after the 2021 Tornadoes. *J. Cult. Herit.* **2023**, *63*, 120–134. [\[CrossRef\]](#)
74. Olsen, A. The Tobii IVT Fixation Filter. 2012, pp. 1–21. Available online: [http://www.vinis.co.kr/ivt\\_filter.pdf](http://www.vinis.co.kr/ivt_filter.pdf) (accessed on 31 January 2022).
75. Jiang, M.; Huang, S.; Duan, J.; Zhao, Q. SALICON: Saliency in Context. In Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015; pp. 1072–1080. [\[CrossRef\]](#)
76. Kroner, A.; Senden, M.; Driessens, K.; Goebel, R. Contextual encoder–decoder network for visual saliency prediction. *Neural Netw.* **2020**, *129*, 261–270. [\[CrossRef\]](#)
77. Simonyan, K.; Zisserman, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. *arXiv* **2014**, arXiv:1409.1556.
78. Liu, N.; Han, J. A Deep Spatial Contextual Long-Term Recurrent Convolutional Network for Saliency Detection. *IEEE Trans. Image Process.* **2018**, *27*, 3264–3274. [\[CrossRef\]](#) [\[PubMed\]](#)
79. Cornia, M.; Baraldi, L.; Serra, G.; Cucchiara, R. Predicting human eye fixations via an LSTM-Based saliency attentive model. *IEEE Trans. Image Process.* **2018**, *27*, 5142–5154. [\[CrossRef\]](#)
80. Borji, A.; Tavakoli, H.R.; Sihite, D.N.; Itti, L. Analysis of scores, datasets, and models in visual saliency prediction. In Proceedings of the IEEE International Conference on Computer Vision, Sydney, NSW, Australia, 1–8 December 2013; pp. 921–928. [\[CrossRef\]](#)
81. Engelke, U.; Liu, H.; Wang, J.; Le Callet, P.; Heynderickx, I.; Zepernick, H.-J.; Maeder, A. Comparative study of fixation density maps. *IEEE Trans. Image Process.* **2013**, *22*, 1121–1133. [\[CrossRef\]](#)
82. Le Meur, O.; Baccino, T. Methods for comparing scanpaths and saliency maps: Strengths and weaknesses. *Behav. Res. Methods* **2012**, *45*, 251–266. [\[CrossRef\]](#) [\[PubMed\]](#)
83. Riche, N.; Duvinage, M.; Mancas, M.; Gosselin, B.; Dutoit, T. Saliency and human fixations: State-of-the-art and study of comparison metrics. In Proceedings of the IEEE International Conference on Computer Vision, Sydney, NSW, Australia, 1–8 December 2013; pp. 1153–1160. [\[CrossRef\]](#)
84. Borji, A.; Sihite, D.N.; Itti, L. Quantitative analysis of human-model agreement in visual saliency modeling: A comparative study. *IEEE Trans. Image Process.* **2013**, *22*, 55–69. [\[CrossRef\]](#) [\[PubMed\]](#)
85. Wilming, N.; Betz, T.; Kietzmann, T.C.; König, P. Measures and Limits of Models of Fixation Selection. *PLoS ONE* **2011**, *6*, e24038. [\[CrossRef\]](#) [\[PubMed\]](#)
86. Zhao, Q.; Koch, C. Learning a saliency map using fixated locations in natural scenes. *J. Vis.* **2011**, *11*, 9. [\[CrossRef\]](#) [\[PubMed\]](#)
87. Nodine, C.F.; Mello-Thoms, C.; Weinstein, S.P.; Kundel, H.L.; Toto, L.C. Do subtle breast cancers attract visual attention during initial impression? In *Medical Imaging 2000: Image Perception and Performance*; SPIE: Bellingham, WA, USA, 2000; Volume 3981, pp. 156–159. [\[CrossRef\]](#)
88. McCarley, J.S.; Kramer, A.F.; Wickens, C.D.; Vidoni, E.D.; Boot, W.R. Visual skills in airport-security screening. *Psychol. Sci.* **2004**, *15*, 302–306. [\[CrossRef\]](#)
89. Goldstein, E.B.; Humphreys, G.W.; Shiffrar, M.; Yost, W.A. Blackwell Handbook of Sensation and Perception. In *Blackwell Handbook of Sensation and Perception*; Wiley: Hoboken, NJ, USA, 2008; pp. 1–788. [\[CrossRef\]](#)

90. Sweller, J. Cognitive Load Theory. In *Psychology of Learning and Motivation—Advances in Research and Theory*; Elsevier: Amsterdam, The Netherlands, 2011; Volume 55, pp. 37–76. [[CrossRef](#)]
91. Gibson, E.J. Principles of Perceptual Learning and Development. Available online: <https://psycnet.apa.org/record/1969-35014-00> (accessed on 16 June 2024).
92. Norman, G. Research in clinical reasoning: Past history and current trends. *Med. Educ.* **2005**, *39*, 418–427. [[CrossRef](#)]
93. Mostafaie, F.; Nabizadeh, Z.; Karimi, N.; Samavi, S. A General Framework for Saliency Detection Methods. 2019. Available online: <https://arxiv.org/abs/1912.12027v2> (accessed on 18 June 2024).
94. Subhash, B. Explainable AI: Saliency Maps. Medium. Available online: <https://medium.com/@bijil.subhash/explainable-ai-saliency-maps-89098e230100> (accessed on 18 June 2024).

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.