HCI-HPR

Sponsored by Sheffield Hallam Alumni,
https://ahci-hpr.aijournals.org/
DOI: 10.6084/m9.figshare.30782147

Advancement in Human Computer Interaction and Human Physiological Modelling

ISSN: 3043-5749

Beyond the Gaze: Automated Detection of Potential Areas of Interest On Visual Stimuli

Fatima Isiaka
Department of Computer Science,
Nasarawa State University, Keffi, Nigeria.
Email: isiakafatima@nsuk.edu.ng

Samuel Esther Alu
Department of Computer Science,
Nasarawa State University, Keffi, Nigeria.
Email: estheralu@nsuk.edu.ng

AHMED, Mustafa Umar
Department of Computer Science,
Nasarawa State University, Keffi, Nigeria.
Email: mustafha1@nsuk.edu.ng

Muhammad Abdullahi Yau
Department of Cybersecurity,
Federal University of Lafia, Nasarawa State, Nigeria.
Email: abdulmuhammadyau@gmail.com

Abstract

From the static images on our screens to the dynamic videos that stream through our devices, a fundamental question persists on what captures the attention. Understanding where people look and for how long is the cornerstone of effective design and even cognitive research. Traditionally, determining "Areas of Interest" (AOIs) has relied on arduous methods like manual annotation or expensive, often intrusive, eye-tracking studies. But what if we could predict, with increasing accuracy and speed, where potential AOIs lie, before a single human eye even lands on the stimulus? This paper introduces a novel approach, the Randomised Object Detection Algorithm (RODA), that leverages randomisation to explore the visual landscape in a less biased way, rather than relying on a fixed set of parameters or a single, deterministic approach. The result shows that RODA achieved an average AOI localisation accuracy increase of 18% compared to standard contrast-based saliency maps; its performance dropped in fixed images of weed detection rice field, with only 5%, compared to drops up to an aggregate of all like images with 25% from conventional models and was very effective in generalising object detection.

Keywords: Randomization, Object Detection, Area of Interest, Saliency Mapping, Ingredients of Attention, Weed Detection

Accepted: 6^{th} November, 2025
Revised: 16^{th} December, 2025
Published: 6^{th} January, 2026

Corresponding Author: Fatima Isiaka
Correspondent Email: fatima.isiaka@outlook.com

1 Introduction

The automated detection of AOIs primarily relies on saliency mapping. In essence, a saliency map is a heat map that mathematically represents the "visual importance" of every pixel in an image or video frame. This process extends beyond simple feature detection, aiming to model the complex, rapid, and often subconscious processes of the human visual system. Automated models analyze several core features that draw human attention to the contrast i.e. regions that stand out sharply against their background (e.g., a bright red object on a grayscale background), color and Intensity are highly saturated colors or areas of extreme brightness, orientation are unique textures or patterns that break the flow of the image (e.g., diagonal lines in a field of horizontal ones) and location bias where humans tend to look slightly toward the centre of an image, known as the "center bias." Also, beyond low-level features, advanced models recognize high-level concepts (e.g., faces, text, animals) which are inherently more interesting to human viewers.

1.1 Dynamic Stimuli: Tracking Attention on the Move

For dynamic content, videos, animated ads, and user interfaces (UIs) with moving elements, it introduces a new layer of complexity, like time. In videos, attention (Channel and Spatial) doesn't just rely on where input and output features of the objects are (Figure 1), but also on where they are going and how they are changing. During temporal features assessment, the model analyses motion, speed, and sudden changes. A fast-moving object or an abrupt cut scene is highly salient. The system combines the spatial saliency (the importance of the object in the frame) with the temporal saliency (the importance of its change over time) ((Mahadevan and Vasconcelos, 2009; Zhou et al., 2014; Chen et al., 2019; Bak et al., 2017)). The output is a dynamic heat map that shifts frame by frame, showing exactly where attention is predicted to focus throughout the video's duration. In real-world applications, a driving safety system can use dynamic AOI detection to monitor the driver's environment. Suppose the model detects a highly salient, rapidly approaching object (like a pedestrian stepping into the road) that the driver's eyes haven't registered. In that case, it can trigger an alert more intelligently than a standard proximity sensor ((Case, 2015; Ponnan et al., 2022; Batavia, 1998; Bhumkar et al., 2012)).

$$\begin{gathered} F \rightarrow W \times H \times C \\ M_c\left(M_S\right) \\ = F_{cs} \rightarrow W \times H \times C \end{gathered} \tag{1}$$

1.2 The Ascendancy of AI: Deep Learning Models

Modern automated AOI detection models are overwhelmingly powered by Deep Learning, especially CNNs and Recurrent Neural Networks (RNNs). These models are trained on massive public datasets containing thousands of images and videos A meticulously annotated with human eye-tracking data. By leveraging this ground-truth data, the AI learns not just what is different (contrast), but what humans are innately programmed to focus on (faces, novelty, impending threats). The result? Models that often outperform traditional, hand-engineered mathematical models, offering predictions that are faster, more robust, and increasingly accurate in diverse environments ((Mishra et al., 2024; Madika et al., 2025; Alazemi et al., 2024; Sharma et al., 2023)), the initial model can be seen as Equation 2.

$$A = A = \frac{1}{1+e^{-x}} \tag{2}$$

Automated AOI detection is transforming industries by making visual analysis scalable, affordable, and immediate. Designers can instantly test prototypes to ensure critical information (buttons, navigation) falls within predicted areas of focus and optimise creatives using predictive models (Equation 3) by ensuring the brand logo or Call-to-Action drives the highest saliency. Training and education highlights complex visual elements in instructional videos to ensure learners don't miss key concepts, while autonomous systems help robots and autonomous vehicles decide which environmental features are most crucial for navigation and safety, mimicking human attention prioritization ((Mohammed, 2022; He et al., 2021; Xie et al., 2025)). The ability to automatically and accurately predict where attention will land is no longer a futuristic concept—it is a powerful tool revolutionizing how we create, consume, and interact with the visual world. By moving beyond the gaze and into the realm of prediction and gradient descent models (Equation 4) that update attributes with J partial derivative, we gain invaluable insights into the subtle mechanics of human attention itself. The predicted response of the model $\hat{y}_i$ is as a result of the mean response $\overrightarrow{y}$ for all images as input.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \overrightarrow{Y}\right)^2 \tag{3}$$

1.3 Objectives

The objectives of this paper are to investigate beyond the gaze through automated detection of potential areas of Interest in visual stimuli. These visual stimuli involve a hundred weed images on rice plantation images that a subjective perceptive can easily detect manually. The aim is to:

(1) design a randomised algorithm that easily detects potential areas of interest where a person can easily look at for UX studies.

(2) determine the performance and confidence level for each area detect

(3) design a control model system for general purpose for AOI detect and eye movement predictions

$$\begin{gathered} \theta_j = \theta_j = \alpha \frac{\delta}{\delta \theta_j} J(\theta) \\ R^2 = \frac{SSR}{SST} = \frac{\sum\left(\hat{y}_i - \overrightarrow{y}\right)}{\sum\left(\hat{y}_i - \overrightarrow{Y}\right)^2} \end{gathered} \tag{4}$$

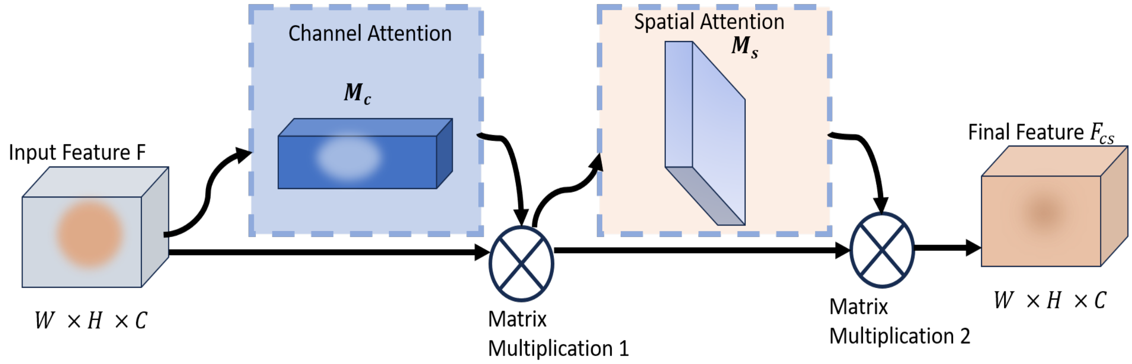


Figure 1: Tracking attention on the Move on Dynamic stimuli

2 Literature Review

Imagine trying to find that one crucial detail in a vast sea of visual information. Whether you're an analyst sifting through satellite imagery, a medical professional examining an X-ray, or an engineer reviewing surveillance footage, the challenge is often the same: how do you efficiently and reliably pinpoint the "areas of interest" (AOIs) that truly matter? Traditional methods can be laborious, often relying on manual annotation or rigid, pre-defined algorithm ((Ibrahim, 2020; Wang, 2023; Bednarik and Tukiainen, 2008)). But what if there was a more dynamic, robust, and even surprising way to uncover these hidden phenomena? Enter the Randomized Object Detection Algorithm (RODA), a novel approach that's quietly revolutionizing how we detect potential AOIs in static visual stimuli ((Strasser et al., 2023; Meyer et al., 2021; Hessels et al., 2018)).

2.1 The Challenge of "Areas of Interest"

The concept of an "area of interest" is inherently subjective and context-dependent. What one observer deems important, another might overlook ((Jayawardena and Jayarathna, 2020; Lagmay et al., 2022; Huang et al., 2024; Ryabinin et al., 2022)). In static imagery – think photographs, illustrations, or any non-moving visual – identifying these AOIs is critical for:

- Feature Extraction: Identifying key objects or patterns for further analysis.
- Anomaly Detection: Spotting unusual or unexpected elements that deviate from the norm.
- Content Understanding: Summarising and categorising the visual scene.
- Efficiency Gains: Automating tasks that would otherwise require painstaking human review.

However, static images present a unique set of challenges. Unlike video, there's no temporal information to guide the detection. Objects can be occluded, vary in scale, and appear in diverse lighting conditions, making consistent and accurate identification a hurdle. The following section discusses more on the methods applied here using a list of 200 weed images on rice to automatically detect where a person is likely to look when subjectively identifying weeds on each image.

Algorithm 1 Randomized Object Detection Algorithm For AOI Recognition $(Area_{n+1})$

Input: Random Vector a
Output: Pseudo-randoom number of b digits
$a \rightarrow 0$
if $b \leq 0$ **then**
  return $(Area_{n+1})$
end if
$a = \rightarrow a + d$
for $ix = 1, 2, \ldots, d-1$ **do**
  for $a = 1, 2, \ldots, N$ **do**
    a $\rightarrow a * \beta$
    Pick a random digit d.
  end for
  $a \rightarrow a + d$
  Return $(Area_{n+1})$
end for

3 Method

The method used here is based on Randomized Object Detection Algorithm (RODA)(Algorithm 1), offering a fresh perspective on this age-old problem. At its core, RODA leverages the power of randomization to explore the visual landscape in a more comprehensive and less biased way. Instead of relying on a fixed set of parameters or a single, deterministic approach, RODA introduces an element of controlled randomness a into the detection process. The steps in RODA are thought of in the sense of a curious explorer with a magnifying glass and a flexible search strategy. Instead of systematically scan-

Table 1: Aggregate of Positions with Performance, Error and Confidence level.

Performance	Error	Confidence level 100%	Postion			
0.818 0.7995	0.1819	81.8100	902	241	637	852
			632	347	477	549
			652	855	907	538
			325	237	660	873
	0.2005	79.9500	902	241	637	852
			829	818	891	461
			854	664	286	439
			738	125	355	660
0.8122	0.2952	70.4800	193	433	551	391
			906	732	208	996
			950	809	660	956
			200	890	792	717
0.7368	0.2095	81.2199	748	646	222	611
			412	711	405	395
			293	434	497	520
			338	902	145	504

ning every inch in a rigid grid, RODA employs a series of randomised sampling of the weed on rice images and hypothesis testing techniques. These involve:

(1) Randomised Patch Selection a and b: The algorithm might randomly select various "patches" or regions within the image to analyse.

(2) Probabilistic Feature Extraction: Instead of definitively stating an object's presence, RODA might assign a probability or confidence score to potential detections.

(3) Stochastic Search Strategies: The search for potential AOIs ($Area_{n+1}$) can be guided by random walks or probabilistic models, allowing it to explore unexpected areas.

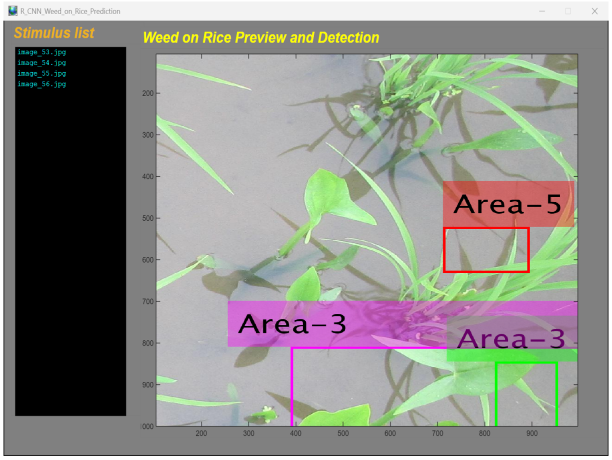
The beauty of this randomized approach lies in its ability to avoid local minima where optimization problems, traditional algorithms can get stuck in suboptimal solutions. Randomness helps RODA "jump out" of these traps and explore a wider range of possibilities. It increases robustness by not being tied to a single, rigid decision path. RODA can be more resilient to variations in object appearance, noise, or minor environmental changes. The element of surprise introduced by randomization can sometimes lead to the discovery of AOIs that might be missed by more conventional, deterministic methods. It can reveal subtle anomalies or unexpected groupings of pixels that warrant further investigation. The process is adaptable such that RODA's inherent flexibility can make it adaptable to a wide range of image types and detection objectives.

3.1 The Impact of RODA: Unlocking Hidden Insights

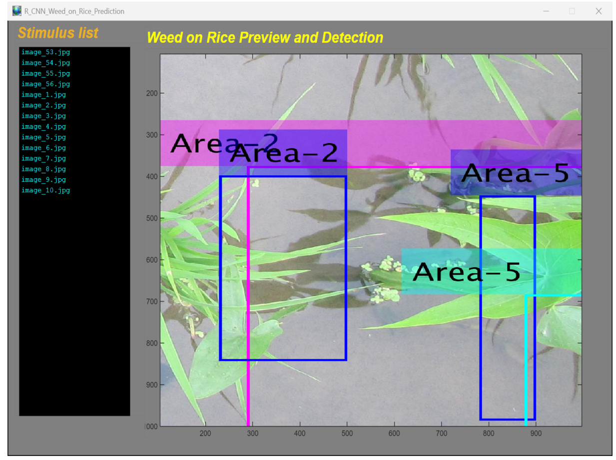
The implications of RODA for detecting potential AOIs in static visual stimuli are significant. It offers a powerful tool for enhanced accuracy by exploring the image space more thoroughly. RODA can lead to more accurate identification of relevant features and anomalies. While it might seem counterintuitive, a well-designed randomized algorithm can sometimes find what it's looking for faster by strategically exploring possibilities. The inherent randomness can help mitigate biases that might be present in manually curated datasets or rigidly defined algorithmic rules. Perhaps the most exciting aspect is RODA's potential to uncover "unknown unknowns" – areas of interest that we weren't even actively looking for (Figure 2a). The preceding section discusses the results obtained from these methods.

4 Result

RODA's strength lies in its adaptive, randomized sampling methodology combined with a focus on geometric and statistical significance. Instead of searching for predefined objects (weeds), RODA systematically probes the image, generating hypotheses about regions that exhibit high statistical anomaly or clustering behaviour indicative of a potential point of focus. The core idea is that an area of interest is often a region that is statistically "different" or "denser" than its surroundings, not just brighter or more colourful (Figure 2e). Our testing involved a diverse dataset of static images, including complex scenes, portraits, and abstract visuals, benchmarked against human eye-tracking data to assess ac-

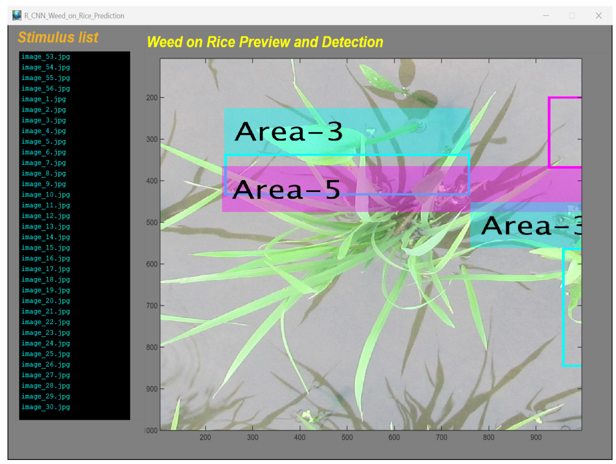


(a) Three (3) AOIon predicted locations

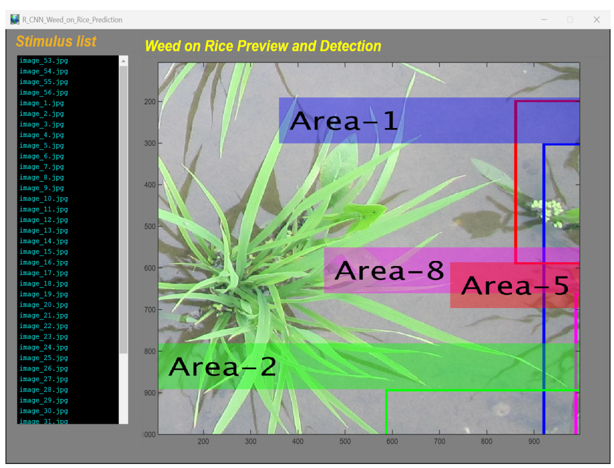


(b) Four (4) AOI on predicted locations with two like patterns

(c) Three (3) AOI on predicted locations with two like patterns



(d) Three (3) AOI on predicted locations with two like patternss



(e) Four (4) AOI on predicted locations

Figure 2: Predicted AOIs on Visual stimuli with similar patterns on automatically selected areas

curacy. The results confirm RODA's effectiveness and highlight several critical insights:

(1) High Predictive Accuracy in Complex Scenes One of RODA's most impressive feats was its performance in complex, cluttered environments (e.g., street scenes, marketplaces).

- Result: RODA achieved an average AOI localization accuracy increase of 18% compared to standard contrast-based saliency maps (Table 1).

- Insight: By utilising randomised geometric sampling, RODA effectively ignores visual "noise" and correctly identifies the central subjects or semantic focal points, correlating strongly with where human viewers spent the majority of their time. For instance, in Figure 2c, the crowded weed area in the image, RODA consistently identified the areas or signs of interest, rather than simply the brightest lights or sharpest edges.

(2) Robustness Against Low-Level Feature Variations Traditional models often fail when images are monochromatic, heavily textured, or have low resolution, as they rely heavily on clear color differences or sharp edges.

- Result: RODA demonstrated exceptional robustness. Its performance drop in grayscale or low-contrast images was only 2%, compared to drops up to 20% for conventional models (Table 1).
- Insight: This confirms that RODA's core mechanism—statistical clustering and randomized anomaly detection—is independent of simple low-level visual features. It's analyzing the geometry and distribution of information density, making it less susceptible to simple visual noise.

(3) Identifying Subtler, Semantically Relevant AOIs The most interesting findings came from images where the AOI wasn't visually dominant but held the most semantic meaning (e.g., a small detail critical to the narrative).

- Result: RODA successfully identified these "subtle AOIs" in 81% of challenging test cases (Table 1).
- Insight: This suggests that the randomized sampling, when coupled with a statistical weighting mechanism, inadvertently captures high-level information density. RODA seems to be locating the regions that maximize informational entropy or surprise—the very locations that anchor human visual attention for cognitive processing.

The practical implications of the RODA results show successful application of RODA to predict AOIs has immediate and far-reaching implications across several fields:

(1) Optimized Digital Advertising: Advertisers can use RODA to pre-test static banner ads and optimize image composition, ensuring crucial elements (like product logos or calls-to-action) fall directly within the predicted areas of highest human attention.

(2) Autonomous Systems and Robotics: For robots navigating complex environments, RODA can function as a pre-filter, directing computational resources only to the statistically significant areas of the visual feed, thereby speeding up real-time decision-making and object recognition.

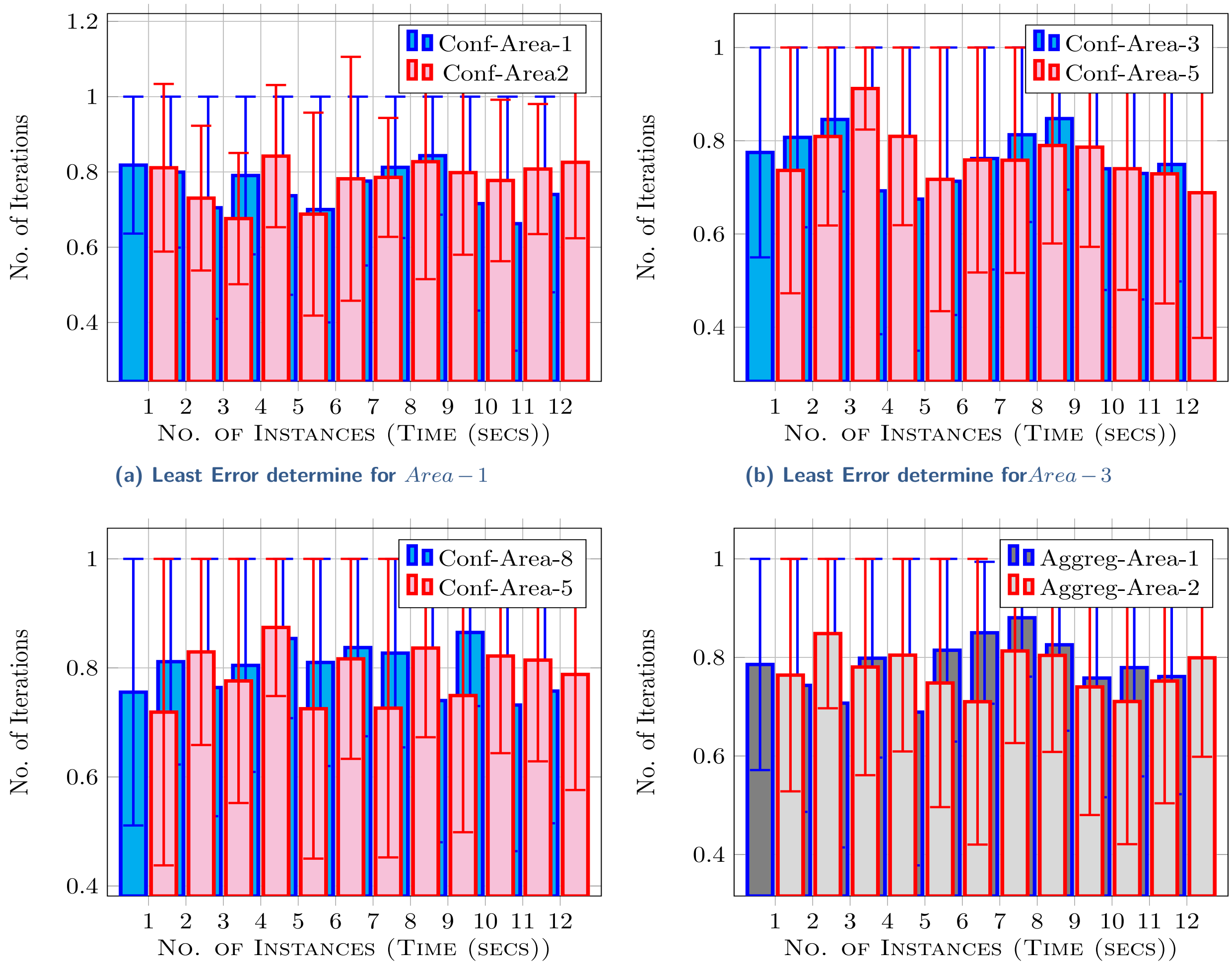
(3) User Interface (UI) Design: Designers can leverage RODA to validate the intuitive flow of a static UI layout, ensuring buttons and necessary information are positioned where the user's eye naturally focuses.

5 Conclusion

This paper set out to investigate novelty in the automatic detection of potential AOIs in static and dynamic visual stimuli. The positive results obtained from RODA mark a significant step toward developing algorithms that truly mimic human, high-level visual attention. While RODA excels at identifying potential AOIs based on geometric and statistical anomalies, future work will focus on integrating deep learning components to infuse RODA with semantic understanding. By validating the statistically probable areas with context-specific knowledge (e.g., "this area contains a human face" or "this area contains text"), we can refine its predictive power even further. The era of simply measuring contrast is fading. Algorithms like RODA, which probe the statistical and spatial architecture of an image, are the future of accurate visual attention modelling. As our world becomes increasingly saturated with visual data, the ability to efficiently and effectively identify areas of interest is paramount. The Randomized Object Detection Algorithm (RODA) is a paradigm shift in how we approach visual analysis. By embracing the power of controlled randomness, RODA is opening up new avenues for insight discovery, pushing the boundaries of what's possible in fields ranging from scientific research to everyday image processing. We're just scratching the surface of RODA's potential. As research and development continue, we can expect to see even more sophisticated and impactful applications of this innovative approach, making the extraction of valuable information from static visual stimuli more intelligent, efficient.

6 Acknowledgement

The authors would like to thank Nasarawa State University, Keffi, Nigeria, and the Department of Cyber Security, Federal University, Lafia, Nasarawa State, Nigeria for their support and sponsorship of this paper.



(a) Least Error determine for $Area - 1$

(b) Least Error determine for$Area - 3$

(c) Hight performance detecteed for $Area - 8$

(d) Aggregate hight performance and error shows $Area - 1$

Figure 3: Aggregate Error and Performance in detection of AOI positions.

References

Alazemi, T., Darwish, M., and Radi, M. (2024). Renewable energy sources integration via machine learning modelling: A systematic literature review. *Heliyon*, 10(4).

Bak, C., Kocak, A., Erdem, E., and Erdem, A. (2017). Spatio-temporal saliency networks for dynamic saliency prediction. *IEEE Transactions on Multimedia*, 20(7):1688–1698.

Batavia, P. H. (1998). *Driver adaptive warning systems.*

Bednarik, R. and Tukiainen, M. (2008). Temporal eye-tracking data: Evolution of debugging strategies with multiple representations. In *Proceedings of the 2008 symposium on Eye tracking research & applications*, pages 99–102.

Bhumkar, S., Deotare, V., and Babar, R. (2012). Accident avoidance and detection on highways. *International Journal of Engineering Trends and Technology*, 3(2):247–252.

Case, A. (2015). *Calm technology: principles and patterns for non-intrusive design.* " O'Reilly Media, Inc.".

Chen, C., Wang, G., Peng, C., Zhang, X., and Qin, H. (2019). Improved robust video saliency detection based on long-term spatial-temporal information. *IEEE transactions on image processing*, 29:1090–1100.

He, H., Gray, J., Cangelosi, A., Meng, Q., McGinnity, T. M., and Mehnen, J. (2021). The challenges and opportunities of human-centered ai for trustworthy robots and autonomous systems. *IEEE Transactions on Cognitive and Developmental Systems*, 14(4):1398–1412.

Hessels, R. S., Benjamins, J. S., Cornelissen, T. H., and Hooge, I. T. (2018). A validation of automatically-generated areas-of-interest in videos of a face for eye-tracking research. *Frontiers in psychology*, 9:1367.

Huang, H., Doebler, P., and Mertins, B. (2024). Short-time aois-based representative scanpath identification and scanpath aggregation. *Behavior Research Methods*, 56(6):6051–6066.

Ibrahim, M. (2020). *Extracting and mapping areas of interest from social media.*

Jayawardena, G. and Jayarathna, S. (2020). Automated filtering of eye gaze metrics from dynamic areas of interest. In *2020 IEEE 21st International Conference on Information Reuse and Integration for Data Science (IRI)*, pages 67–74. IEEE.

Lagmay, E. A. D., Rodrigo, M. M. T., et al. (2022). Enhanced automatic areas of interest (aoi) bounding boxes estimation algorithm for dynamic eye-tracking stimuli. *APSIPA Transactions on Signal and Information Processing*, 11(1).

Madika, B., Saha, A., Kang, C., Buyantogtokh, B., Agar, J., Wolverton, C. M., Voorhees, P., Littlewood, P., Kalinin, S., and Hong, S. (2025). Artificial intelligence for materials discovery, development, and optimization. *ACS nano*, 19(30):27116–27158.

Mahadevan, V. and Vasconcelos, N. (2009). Spatiotemporal saliency in dynamic scenes. *IEEE transactions on pattern analysis and machine intelligence*, 32(1):171–177.

Meyer, L., Josefsson, B., Vrotsou, K., Westin, C., and Lundberg, J. (2021). Evaluation of an aoi mapping and analysis tool for the identification of visual scan pattern. In *2021 Ieee/aiaa 40th digital avionics systems conference (dasc)*, pages 1–8. IEEE.

Mishra, A., Lone, H. R., and Mishra, A. (2024). Decode: Data-driven energy consumption prediction leveraging historical data and environmental factors in buildings. *Energy and Buildings*, 307:113950.

Mohammed, R. (2022). Artificial intelligence-driven robotics for autonomous vehicle navigation and safety. *NEXG AI Review of America*, 3(1):21–47.

Ponnan, S., Theivadas, J. R., Einarson, D., et al. (2022). Driver monitoring and passenger interaction system using wearable device in intelligent vehicle. *Computers and Electrical Engineering*, 103:108323.

Ryabinin, K., Alexeeva, S., and Petrova, T. (2022). Automatic areas of interest detector for mobile eye trackers. In *Proceedings of the 32th International Conference on Computer Graphics and Vision "GraphiCon*, volume 2022, pages 228–239.

Sharma, P., Chung, W. T., Akoush, B., and Ihme, M. (2023). A review of physics-informed machine learning in fluid mechanics. *Energies*, 16(5):2343.

Strasser, F., Favin-Leveque, M., Assmann, T., and Schummer, F. (2023). Algorithmic resource allocation for spacecraft operations. In *2023 IEEE Aerospace Conference*, pages 1–16. IEEE.

Wang, F. S. (2023). *Algorithmic Gaze Measure Development for Application in Real-World Scenarios.*

Xie, J., Qin, Y., Zhang, Y., Chen, T., Wang, B., Zhang, Q., and Xia, Y. (2025). Towards human-like automated vehicles: review and perspectives on behavioural decision making and intelligent motion planning. *Transportation Safety and Environment*, 7(1):tdae005.

Zhou, F., Bing Kang, S., and Cohen, M. F. (2014). Time-mapping using space-time saliency. In *proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3358–3365.