Computational Method for Predicting Visual Attention in Older Adults with Age-related Features.

LI, Xiangdong, SHI, Xinchi, GU, Haoyu, SHEN, Tianai, CHENG, Shiwei and WANG, Jing <http://orcid.org/0000-0002-5418-0217>

Multimodal Technologies and Interaction

MDPI

Article

Computational Method for Predicting Visual Attention in Older Adults with Age-Related Features

Xiangdong Li [1,*], Xinchi Shi [1], Haoyu Gu [1], Tianai Shen [2], Shiwei Cheng [3] and Jing Wang [4]

[1] College of Computer Science and Technology, Zhejiang University, 38 Zheda Road, Hangzhou 310027, China; ruazzm@zju.edu.cn (X.S.); 22321364@zju.edu.cn (H.G.)
[2] College of Software Technology, Zhejiang University, 38 Zheda Road, Hangzhou 310027, China; tianai.shen@zju.edu.cn
[3] College of Computer Science and Technology, Zhejiang University of Technology, Hangzhou 310023, China; swc@zjut.edu.cn
[4] College of Business, Technology and Engineering, Sheffield Hallam University, Sheffield S1 1WB, UK; jing.wang@shu.ac.uk
* Correspondence: axli@zju.edu.cn; Tel.: +86-0571-87952010

Abstract

Age-related changes in visual perception alter attentional deployment, yet computational models of visual attention have been validated almost exclusively on younger populations. This limits both the theoretical investigation of age-specific mechanisms and practical applications in age-inclusive design, where researchers depend on specialised eye-tracking equipment to observe such differences. Therefore, we present the Elderly Visual Attention Estimation (EVAE) model, a computational framework that predicts early visual attentional orienting in older adults by combining stimulus-driven image features with age-specific top-down priors. The framework models six dimensions of elderly visual attention from cross-age eye-tracking data: colour brightness sensitivity, centre bias, foreground–background differentiation, depth detection, early attentional prior, and sustained-attention spatial prior. On public datasets, EVAE achieves an AUC-Judd of 0.92, which outperforms existing saliency models and deep learning approaches such as DeepGaze II. The framework is optimised for an input resolution of 128 × 96 pixels, producing fixation probability maps that are upsampled to match the original stimulus resolution for practical interface evaluation. Cross-age validation confirms the model's specificity, as EVAE predicts attentional behaviour in older adults but does not generalise to younger adults. An ablation study shows that image features and top-down spatial priors each contribute independently to prediction accuracy, and that bottom-up saliency alone cannot account for age-related attentional patterns. Centre bias and early attentional prior are the strongest predictors, indicating that visual ageing involves greater reliance on spatial strategies and compensatory processing. As an alternative to hardware-based eye-tracking, EVAE widens the scope of empirical research into older adults' visual attention and informs the design of accessible digital interfaces.

Keywords: visual attention; older adults; visual attention prediction; visual attention model; eye tracking

Academic Editor: Arun K. Kulshreshth

Received: 20 March 2026
Revised: 6 May 2026
Accepted: 7 May 2026
Published: 1 June 2026

1. Introduction

Age-related changes in visual perception alter attention deployment patterns, as older adults experience declines in visual acuity and processing speed (Figure 1) [1], alongside

behavioural shifts including increased reliance on background information and altered scanning strategies [2]. These perceptual changes pose challenges for human–computer interaction, as understanding how older adults process visual information is critical for designing accessible systems. Yet, computational methods for visual attention, for example, tools widely used in interface evaluation, visualisation design, and graphics applications, remain calibrated predominantly to younger populations [3], which largely limit both theoretical understanding of age-related perceptual mechanisms and practical capacity for age-inclusive design.

Figure 1. Example of visual attention distributions across younger and older adults.

Previous studies have identified age-related features affecting visual attention, e.g., reduced colour sensitivity [4], increased centre bias [5], and altered foreground–background differentiation [2]; these features have been examined in isolation rather than as integrated components of attentional behaviour. How these features collectively influence visual attention in older adults remains poorly characterised. Existing visual attention models rely on bottom-up and top-down processing frameworks that are optimised for younger populations [6], which fail to account for the compensatory shift towards schema-driven processing that characterises cognitive ageing [7]. The lack of age-specific datasets [3,8] limits model development. Although eye-tracking provides ground truth data, it cannot provide a scalable solution for large-scale evaluation of age-inclusive interfaces or for integration into design workflows.

To fill the gap, we developed the Elderly Visual Attention Estimation (EVAE) model. EVAE is a computational method for predicting visual attention patterns in older adults from image features alone, without requiring eye-tracking equipment. The method inte-

grates six empirically validated features of age-related visual attention: colour brightness, centre bias, foreground–background differentiation, depth detection, early attentional prior, and sustained-attention spatial prior. These features are integrated within a unified predictive framework based on Probabilistic Gradient Boosting Machines (PGBM) combined with SHapley Additive exPlanations (SHAP) [9], together to enable accurate attention prediction and quantitative assessment of feature importance.

This work makes two contributions. (1) It identifies and validates the key determinants of visual attention in older adults. Feature analysis reveals that centre bias and early attentional prior exhibit the strongest influence on attention prediction. This advances theoretical understanding of attentional priorities in ageing and challenges assumptions embedded in models calibrated to younger adults. (2) We present a novel computational method that is specifically designed for older adults' visual attention. Evaluated against public datasets, EVAE achieves an AUC-Judd score of 0.92 in predicting the spatial distribution of initial fixations (see examples of predicted results in Figure 1), which consistently surpasses state-of-the-art models across multiple performance metrics. This provides researchers and designers with an efficient, scalable method for predicting visual attention in older adults, and establishes a foundation for evidence-based development of age-inclusive digital interfaces, visualisation systems, and interactive applications. The proposed method is optimised for downsampled inputs (128×96) while remaining scalable across spatial resolutions, with attention maps subsequently upsampled for integration into standard design workflows.

The paper is organised as follows. Section 2 reviews literature on visual attention mechanisms, age-related perceptual features, and computational models. Section 3 describes experimental datasets and methodology. Section 4 presents the EVAE framework, architectural design, and evaluation procedures. Section 5 analyses estimation results and feature importance findings. Section 6 discusses theoretical and practical implications, and Section 7 concludes with directions for future research.

2. Related Work

Understanding how older adults allocate visual attention requires knowledge from three domains, including the cognitive and neural mechanisms underlying age-related attentional changes, the specific perceptual features that distinguish elderly visual behaviour, and the computational approaches used to model these patterns. This section reviews each domain, then identifies gaps that motivate our work.

2.1. Visual Attention Mechanisms and Ageing

Visual attention operates through two complementary pathways [10], namely bottom-up processing, which is driven by stimulus salience such as colour contrast or motion, and top-down processing, which is guided by task goals and prior knowledge. In younger adults, these pathways interact dynamically to enable efficient visual search and scene comprehension. In contrast, ageing disrupts this balance.

Physiological changes in the visual system reduce contrast sensitivity, narrow useful visual fields, and slow neural transmission [1]. These sensory declines weaken bottom-up saliency signals that guide younger adults' attention. Concurrently, structural brain changes compromise top-down control, as reduced connectivity between prefrontal control regions and sensory cortices disrupts voluntary attention shifting and distractor suppression [11]. Although older adults have some capacity for stimulus-driven attention, their goal-directed attentional control declines [12], which reduces flexibility in dynamic visual environments.

Behavioural studies using eye-tracking reveal functional consequences. Older adults require longer viewing times to extract information from complex scenes [13,14], exhibit

altered fixation distributions compared to younger adults [15], and show different attentional responses to emotional content [14]. Cognitive training can partially slow these declines [16], which suggests that age-related attentional changes reflect reorganisation rather than simple deterioration.

Classical theories of visual attention [17,18] were developed using younger populations and emphasise bottom-up saliency and feature integration. These frameworks remain foundational, but they inadequately account for compensatory strategies observed in older adults, such as increased reliance on spatial priors and conservative exploration patterns. Understanding elderly attention therefore requires identifying which specific features drive these age-related behavioural shifts.

2.2. Perceptual Features Distinguishing Elderly Visual Attention

Visual attention models typically rely on hierarchical feature representations spanning low-level image properties (e.g., colour, brightness, edges), mid-level spatial organisation (e.g., segmentation, depth, object boundaries), and high-level semantic content (e.g., objects, context, meaning). Age-related changes alter both the processing efficiency and relative influence of features at each level.

At the low-level processing stage, reduced retinal sensitivity and neural efficiency eliminate older adults' responsiveness to colour brightness [4] and luminance contrast [19]. Texture and edge detection, which contribute to saliency computation in younger adults, exert weaker influence on elderly attention [20]. These declines suggest that models relying heavily on low-level saliency may poorly predict where older adults look.

At the mid-level of spatial organisation, older adults parse visual scenes differently than younger adults. They allocate more attention to background regions relative to foreground objects [2], struggle with depth perception tasks [21], and show reduced efficiency in segmenting overlapping objects [22,23]. These findings indicate that spatial organisation features, which structure scene representation, function differently in ageing vision.

At the semantic level, compensatory strategies emerge at higher processing levels. Older adults exhibit pronounced centre bias, which is characterised by a preference for fixating central screen regions regardless of peripheral saliency [5], and adopt conservative viewing patterns with longer fixation durations and fewer exploratory saccades [24,25]. These strategic adaptations likely compensate for reduced processing speed and attentional control, prioritising reliable central information over potentially distracting peripheral stimuli.

Individual studies have documented these features in isolation, yet how they collectively shape attention allocation remains unclear. Existing models weight features based on younger adults' behaviour, where low-level saliency strongly predicts fixations. For older adults, compensatory high-level strategies (centre bias, conservative exploration) may dominate, which change optimal feature weightings. Quantifying the relative importance of age-specific features represents a critical gap bridging perceptual research and computational modelling.

2.3. Computational Method for Visual Attention

Computational attention models predict spatial and temporal distributions of visual fixations from image or video input. Early models used manually selected features to compute saliency maps [26], combining colour, intensity, and orientation channels. Subsequent graph-based and probabilistic approaches improved prediction accuracy but retained focus on bottom-up saliency [27].

Deep learning architectures have advanced saliency prediction. Models such as DeepGaze II [28] leverage convolutional neural networks pre-trained on object recognition to extract rich feature representations, achieving high accuracy on standard bench-

marks. Transformer-based architectures [29] capture long-range dependencies through self-attention mechanisms, enabling more sophisticated scene understanding. TranSalNet [30] and hierarchical vision transformers [31] demonstrate that integrating low-level saliency with high-level semantic features improves prediction performance. These models find applications in interface design, content optimisation, and multimedia systems [32,33].

However, standard benchmarks such as MIT1003 and SALICON contain predominantly younger participants. Models trained on these datasets learn feature weightings and attentional biases specific to younger adults' viewing behaviour [34]. When applied to older adults, such models often fail because they cannot account for age-specific compensatory strategies. Additionally, state-of-the-art deep learning models impose substantial computational costs, limiting deployment in assistive technologies where real-time performance is necessary.

Researchers have recognised that older adults require specialised models other than saliency prediction approaches. Models for elderly users with mobility impairments incorporate cognitive capacity constraints [35,36] and altered gaze dynamics [37]. Recent work explores how cognitive load [38] and sensory impairments [39] modulate attention, emphasising accessibility and usability considerations [40]. Technical approaches include self-attention frameworks for elderly saliency prediction [41], hybrid convolutional-transformer architectures for pixel-level estimation [42,43], and gaze-tracking methods [44,45].

Furthermore, previous standard saliency models have demonstrated that integrating dataset-derived statistical priors, such as centre bias or temporal viewing tendencies, significantly improves prediction accuracy on novel images [46,47]. For older adults, who exhibit distinct compensatory strategies, relying on generic image features is insufficient. Therefore, extracting age-specific cognitive behavioural priors from a training corpus and formulating them as spatial templates is necessary to bridge the gap between bottom-up visual saliency and top-down cognitive ageing.

Despite these advances, elderly-specific models face two limitations. First, they address isolated aspects of age-related attention (e.g., contrast sensitivity in specific tasks) rather than systematically integrating the comprehensive set of perceptual and cognitive features that distinguish elderly behaviour. Second, they do not quantify which features most strongly influence attention in older adults. Without understanding feature importance, models remain black boxes, limiting both theoretical insight into age-related attentional mechanisms and practical guidance for interface design.

2.4. Research Gaps and Study Objectives

This preceding review highlights several research gaps. Firstly, mechanistic understanding remains incomplete. Although individual studies document features distinguishing elderly attention, e.g., reduced colour sensitivity, increased centre bias, altered depth processing, and conservative fixation patterns, these features have been examined in isolation. How they interact to produce observed attentional behaviour, and which features exert the strongest influence, remains implicit. Existing theories were developed for younger populations and inadequately explain compensatory strategies that characterise elderly visual behaviour.

Secondly, computational models lack age-specificity. Standard attention models are trained on younger adults and embed assumptions about feature processing that do not generalise to older populations. Elderly-specific models address narrow use cases without systematically incorporating the full range of age-related perceptual and cognitive features. Most critically, existing approaches treat models as black boxes, which fail to quantify which features drive predictions. This limits both theoretical understanding (which mechanisms matter most) and practical application (which design elements should be prioritised).

Thirdly, methodological constraints hinder research. Elderly-specific eye-tracking datasets are scarce [3,48], which limits model development. Eye-tracking itself, although providing ground truth data, cannot scale to large-scale interface evaluation or integrate into design workflows. Computational models offering efficient alternatives are needed, but only if they accurately capture age-specific attention patterns.

This study addresses these gaps through two objectives. We identify and validate perceptual features spanning low-level sensory processing (colour, contrast), mid-level spatial organisation (depth, segmentation), and high-level strategic adaptation (centre bias, sustained-attention spatial prior, early attentional prior), so as to quantify the relative importance of these features in predicting elderly attention, which identifies which spatial and visual factors contribute most strongly to the computational estimation of age-related viewing patterns. We develop a computational model integrating these features within an efficient predictive framework. This enables both a theoretical investigation, such as testing hypotheses about feature interactions and compensatory mechanisms, and a practical application in accessible interface design and assistive technologies.

Overall, this work integrates age-specific features and quantifies their predictive value, thereby connecting perceptual research with computational methods. It not only offers a clearer understanding of attention reorganisation in ageing, but also provides practical tools for age-inclusive design.

3. Method

3.1. Feature Selection and Theoretical Foundation

Through the preceding systematic review [2,5,19,22,49,50], we identified six intrinsic features as determinants of elderly visual attention, including colour brightness (C–B, brightness of areas of interest of colour images), centre bias (CB, favoured fixations of central screen regions), foreground–background differentiation (F–B D, attention distribution between salient objects and context), depth detection (DD, processing of spatial depth cues), early attentional prior (EAP, temporal delays in fixation initiation), and sustained-attention spatial prior (SASP, duration and spatial clustering of fixations). Three criteria were applied for feature selection.

- Features need to exhibit empirically documented age-related alterations with replicated evidence across independent studies [51];
- Features must be measurable through computational extraction from visual stimuli and eye-tracking data without manual annotation or subjective judgement;
- Features must remain relevant across diverse visual contexts (natural scenes, interfaces, multimedia) rather than being task-specific or domain-limited.

These criteria ensure features capture fundamental attentional mechanisms while also enabling automated, generalisable modelling.

We focus on intrinsic characteristics, which are properties inherent to visual stimuli (colour, depth, spatial composition) or attentional behaviour (centre bias, temporal dynamics, fixation patterns), that remain consistent across viewing conditions. We excluded extrinsic parameters such as viewing distance, display resolution, and ambient lighting. Although these factors influence viewing conditions, they introduce experimental variability unrelated to fundamental attentional mechanisms and would limit model generalisation to diverse real-world applications requiring different viewing setups.

Furthermore, these six features are not assumed independent. For example, colour brightness may interact with depth detection via shading cues, whilst centre bias could correlate with the sustained-attention spatial prior if central fixations are systematically prolonged. Disentangling these interactions is fundamental to characterising elderly vi-

sual attention. It remains to be determined whether compensatory high-level strategies (such as centre bias and sustained-attention spatial prior) dominate attentional allocation, or whether sensory-level features (including colour brightness and depth) preserve their predictive utility despite age-related sensory decline. Our methodology addresses two questions. First, which features most strongly predict elderly attention allocation? Second, how do features interact to produce observed attention patterns?

To answer these questions, we developed computational algorithms that extract quantitative measurements for each feature from visual stimuli and corresponding eye-tracking data. These feature representations are integrated into a unified gradient boosting framework that enables systematic quantification of individual feature contributions and interactive effects on elderly visual attention prediction.

3.2. Computational Feature Extraction

We extract quantitative measurements of six intrinsic features from visual stimuli and eye-tracking data. Computational algorithms transform RGB images and fixation coordinates into numerical representations encoding elderly-specific attentional characteristics. Features progress from sensory processing (colour brightness) through spatial organisation (centre bias, foreground–background differentiation, depth detection) to temporal dynamics (early attentional prior, sustained-attention spatial prior).

(1) Colour brightness

Colour constitutes a fundamental visual feature that influences both information extraction and the perception of other visual attributes [52]. Age-related physiological changes reduce colour brightness sensitivity in older adults [4]. We simulate this degradation by converting RGB images to greyscale using luminance-weighted transformation:

$$I_{grey} = 0.299Red + 0.587Green + 0.114Blue \quad (1)$$

Coefficients reflect human photopic sensitivity [4,19] with peak response in the green spectrum. This isolates brightness from chromatic content, preserving spatial luminance patterns critical for attention allocation.

Standard photometric coefficients are derived from young-adult photopic sensitivity and do not explicitly account for age-related ocular changes, such as lens yellowing, which attenuates short-wavelength light and alters perceived colour distributions [19]. While one could theoretically introduce manual physiological filters, such rigid parameterisation often fails to accommodate the significant inter-individual variability inherent in visual ageing. Consequently, rather than imposing a pre-defined physiological transformation, we utilise Equation (1) to extract scene luminance as a stable physical baseline. Under our data-driven framework, age-specific perceptual differences are not treated as hand-crafted corrections but are implicitly learnt through supervised training on age-stratified gaze data. This approach ensures that the 'age-specificity' of the model is not a post hoc adjustment, but an emergent property of the learnt conditional mappings between physical stimuli and elderly fixation behaviour, resulting in a more robust and statistically grounded estimation of attentional deployment.

The greyscale image (I_{grey}) eliminates chromatic information while preserving edge structures necessary for attention modelling (Figure 2). We apply this transformation to all images, which creates standardised luminance representations for subsequent extraction. In accordance with recent studies such as [53,54], a standard optical examination image was selected to demonstrate the computational results.

Figure 2. Example of colour brightness computation (**left**: original image, **right**: colour brightness image).

(2) Centre bias

Centre bias refers to systematic attention towards central regions, which intensifies with age [5]. Older adults exhibit stronger central preferences than younger populations, reflecting reduced peripheral acuity, narrowed useful field of view, and decreased exploratory scanning [49].

We compute centre bias through adaptive localisation integrating geometric layout and semantic content. Traditional methods assume uniform spatial weighting around the geometric centre (x_c, y_c). Older adult observers balance central fixation strategies with salient object locations. We identify the most salient object using pre-trained detection [51], compute its centroid (x_s, y_s), and calculate the effective centre as their midpoint:

$$CB_d = \frac{1}{\sigma_d\sqrt{2\pi}} e^{\frac{-(P-C_d)^2}{2\sigma_d^2}} \quad (2)$$

where CB_d is the Gaussian map value at image pixel P, C_d is the centre point of centre bias, and σ_d is the standard deviation of the 2D Gaussian function.

This accommodates diverse compositions. With centrally positioned salient objects, (x_{eff}, y_{eff}) converges towards the geometric centre, reinforcing bias. With off-centre objects, (x_{eff}, y_{eff}) shifts moderately, capturing elderly observers' compromise between central fixation and salient content.

We generate a 2D Gaussian centred at (x_{eff}, y_{eff}) with $\sigma_d = 0.3 \times min(width, height)$, creating smooth attention weighting decreasing with distance from centre (Figure 3). This spatial prior quantifies centre bias strength at each pixel.

Figure 3. Centre bias computation (**left**: original image, **right**: centre bias image).

(3) Foreground–background differentiation

Younger adults prioritise foreground objects through efficient figure-ground segregation; older adults allocate increased attention to background regions [2]. This reflects reduced contrast sensitivity impairing foreground extraction and compensatory reliance on contextual information.

We employ hierarchical matting [55] to segment foreground and background, generating a continuous alpha matte (α) where $\alpha = 1$ indicates foreground, $\alpha = 0$ indicates

background, and intermediate values capture transitions. The algorithm fuses semantic segmentation with detail-preserving matting through adaptive weighting:

$$\alpha = (1 - 2 \times |U - 0.5|) \times M + 2 \times |U - 0.5| \times U \quad (3)$$

where U represents unified foreground probability from semantic encoding; M represents refined matting output. The weighting term $(1 - 2 \times |U - 0.5|)$ balances contributions. When $U \approx 0.5$ (uncertain boundaries), the model emphasises M for detail preservation; when $U \approx 0$ or 1 (confident regions), the model prioritises U for semantic consistency.

Our elderly attention model inverts this representation, weighting background regions more heavily to reflect older adults' disproportionate allocation to contextual content (Figure 4). This design choice does not imply an active cognitive priority for the background, but rather serves as a behavioural representation of the compensatory scanning strategies resulting from age-related declines in figure-ground segregation [2].

Figure 4. Foreground–background differentiation computation (**left**: original image, **right**: white area is foreground region and black area is background region).

(4) Depth detection

Depth detection enables prioritisation of proximal objects and navigation [56]. Age-related degradation arises from reduced binocular disparity processing, diminished motion parallax sensitivity, and impaired monocular cue extraction [22]. These deficits alter attention allocation across depth planes, increasing reliance on foreground objects and reducing background exploration [21].

We employ transformer-based hierarchical depth estimation [57] comprising coarse estimation through multi-scale wavelet-decomposed feature extraction, followed by adaptive refinement via the AdaBins architecture (v1.0–weights relsease). AdaBins dynamically learns optimal depth bin boundaries rather than imposing fixed intervals, improving accuracy in scenes with non-uniform depth distributions.

Final depth values are computed through probability-weighted summation over N adaptive bins:

$$\tilde{d} = \sum_{k=1}^{N} c(b_k) p_k \quad (4)$$

where p_k is the predicted probability that the pixel belongs to depth bin k, and $c(b_k)$ is the centre value of bin k. The bin centres are determined by adaptive bin widths b_i, computed as: $b_i = \frac{b'_i + \epsilon}{\sum_{j=1}^{N} b'_j + \epsilon}$, $\epsilon = 10^{-3}$, $c(b_i) = d_{min} + (d_{max} - d_{min})(\frac{b_i}{2} + \sum_{j=1}^{i-1} b_j)$. Here, b'_i is the raw bin width output from a mini-Vision Transformer (ViT) encoder followed by an MLP head, normalised via the softmax-like operation to ensure $\sum_{i=1}^{N} b_i = 1$. The bin centre $c(b_i)$ is calculated as the midpoint of bin i within the depth range $[d_{min}, d_{max}]$, with cumulative width $\sum_{j=1}^{i-1} b_j$ positioning the bin along the depth axis. The small constant ε prevents numerical instability during normalisation.

As illustrated in Figure 5, the resulting depth map encodes relative spatial distance for each pixel, with darker values indicating proximity and lighter values indicating distance.

This depth representation serves as a feature input to our elderly visual attention model, where it interacts with other attention-guiding factors such as centre bias and foreground–background differentiation. Empirical analysis demonstrates that incorporating depth information significantly improves fixation prediction accuracy for elderly observers, particularly in scenes with pronounced depth structure where age-related depth processing deficits most strongly influence attention allocation.

Figure 5. Depth detection computation (**left**: original image, **right**: depth detection image).

(5) Early Attentional Prior

The previous four features rely exclusively on bottom-up image properties. However, image features alone are insufficient to model elderly attention, as age-related sensory decline forces older adults to rely heavily on top-down, compensatory oculomotor strategies [58]. To address this shift, we introduce the early attentional prior. This feature captures the inherent spatial exploration habits that older adults apply during the initial phase of scene observation.

Early attentional prior refers to temporal latency between stimulus onset and initial attention deployment, which systematically increases with age, reflects slowed visual pathway transmission, reduced parietal attention network efficiency, and diminished saccade programming control [51]. The temporal delay is accompanied by spatial consequences. Older adults exhibit more dispersed, less targeted initial fixations than younger adults' rapid convergence on salient regions [59].

We model this through a spatial prior encoding the statistical distribution of early-stage fixations from elderly observers, as elaborated in the following steps.

Training data aggregation. We isolate fixations occurring within the first two seconds post-stimulus onset from the training data. The initial ambient phase (typically the first 1.5 to 2 s) is characterised by rapid spatial orienting driven by global layout and top-down spatial priors [60,61]. Following this, visual behaviour shifts to focal processing for detailed semantic extraction. Because cognitive ageing slows visual processing and saccadic programming, this initial orienting phase is slightly extended in older adults [62]. Therefore, the two-second threshold provides a robust window to capture early attentional orienting while excluding sustained, semantic-driven attention phases.

Spatial discretisation. We partition each image into $N = 16$ uniform regions (R_k) in a 4×4 grid. This resolution captures coarse spatial preferences (upper-centre bias, peripheral avoidance) without overfitting to pixel-level noise, ensuring the prior generalises across diverse compositions.

Prior generation. For each region R_k, we accumulate total fixation duration E_k (milliseconds) from all initial fixations (0–2 s) across the training corpus. Duration-based weighting captures both fixation frequency and dwell time, which reflects attentional engagement strength. Regional weights are normalised to form a probability distribution:

$$RT_k = \frac{E_k}{\sum_{j=1}^{N} E_j} \tag{5}$$

where RT_k quantifies the proportion of early attentional resources elderly observers allocate to spatial zone R_k. This spatial prior functions as a temporal feature, encoding where elderly attention deploys during delayed initial orienting (Figure 6).

Figure 6. Early attentional prior computation (**left**: original image, **right**: early attentional prior image).

(6) Sustained-Attention Spatial Prior

Sustained-attention spatial prior characterises spatial distribution of sustained attention beyond initial orienting, reflecting stable viewing patterns during extended exploration. Elderly observers exhibit reduced saccadic amplitude (constraining exploration), prolonged fixation durations (indicating slower information extraction), and heightened central clustering due to peripheral vision degradation [25,63]. These characteristics define an elderly-specific 'attentional footprint' differing from younger adults' exploratory patterns.

We construct a global fixation prior that encodes the statistical spatial preferences exhibited by elderly observers during extended viewing. This sustained-attention spatial prior aggregates fixations across the entire viewing duration, reflecting stable, long-term exploration strategies. This temporal distinction is critical. Early orienting is dominated by bottom-up stimulus capture; sustained viewing integrates top-down factors such as scene comprehension goals and compensatory strategies [8].

This spatial prior is generated as follows.

Global data aggregation. We extract fixation data spanning complete viewing duration from all training images with no temporal truncation. This window captures cumulative attentional deployment, including exploratory scanning and revisitation patterns (repeated returns to central regions compensating for working memory limitations).

Regional weighting. Using the same 4×4 spatial discretisation ($N = 16$ regions R_k), we compute total dwell time T_k by summing fixation durations across training samples. Duration-based weighting ensures the prior reflects attentional engagement intensity: prolonged cumulative dwell indicates sustained processing resource allocation, consistent with characteristically longer elderly fixation durations.

Prior normalisation. Regional dwell times are normalised to yield spatial density of sustained elderly attention:

$$FB_k = \frac{T_k}{\sum_{j=1}^{N} T_j} \tag{6}$$

FB_k quantifies the proportion of total attentional resources elderly observers allocate to spatial zone R_k during sustained viewing. Higher FB_k values identify regions attracting prolonged attention across diverse content, revealing population-level spatial biases independent of scene semantics (Figure 7). The resulting map reveals sustained attention distribution, which exhibits strong central concentration and reduced peripheral weighting. This pattern reflects combined influence of reduced saccadic exploration and peripheral vision decline. The fixation prior functions as a spatial template modulating predictions. Regions with high FB_k receive elevated attention weights, which reflects that elderly observers systematically allocate disproportionate sustained attention to specific zones.

Figure 7. Sustained-attention spatial prior computation (**left**: original image, **right**: sustained-attention spatial prior image).

3.3. Feature Fusion

After extracting six complementary feature representations capturing distinct dimensions of elderly visual perception, we integrate these components into a unified computational framework to predict spatial attention allocation. The integration needs to model complex, non-linear interactions among features. Elderly visual attention does not simply sum individual contributions but emerges from synergistic interplay. Depth detection may amplify centre bias in scenes with foreground–background ambiguity; prolonged early attentional prior may modulate colour brightness influence on initial fixations. To capture these dependencies whilst maintaining interpretability, which are critical for understanding which perceptual factors govern elderly attention, we adopt a two-stage architecture comprising feature-level fusion followed by probabilistic ensemble learning.

The feature extraction pipeline processes each input image (normalised to 128 × 96 pixels for computational efficiency whilst preserving spatial structure) through six parallel pathways, which generate pixel-aligned feature maps: colour brightness (G_{cr}), centre bias (G_{cb}), foreground–background differentiation (G_{fd}), depth detection (G_{dd}), early attentional prior (G_{eap}), and sustained-attention spatial prior (G_{sasp}). Each map $G_i \in R^{(128\times96)}$ encodes the contribution of feature i to attention at every spatial location.

These maps are fused into a composite feature vector at each pixel location (x, y), yielding a six-dimensional representation:

$$g(x,y) = \left[G_{cr}(x,y), G_{cb}(x,y), G_{fd}(x,y), G_{dd}(x,y), G_{eap}(x,y), G_{sasp}(x,y)\right]^T \tag{7}$$

A naive linear combination would assume additive independence:

$$G(x,y) = \alpha_1 G_{cr}(x,y) + \alpha_2 G_{cb}(x,y) + \alpha_3 G_{fd}(x,y) + \alpha_4 G_{dd}(x,y) + \alpha_5 G_{eap}(x,y) + \alpha_6 G_{sasp}(x,y) \tag{8}$$

where α_i represents the weight of feature i. This combination assumes additive independence among features, implying that each perceptual factor contributes to visual attention in an isolated and proportional manner. However, this assumption does not hold for elderly visual perception, where attention emerges from complex and context-dependent interactions among multiple factors. For example, the contribution of centre bias may vary with scene structure, and the effect of colour or luminance contrast may be moderated by depth layout or foreground clarity. These interaction effects are inherently non-linear and cannot be adequately captured by a linear formulation. Therefore, relying on linear aggregation is insufficient for modelling the underlying mechanisms of elderly visual attention.

4. Model Development

To address the limitations of linear aggregation, the mapping from the multi-dimensional feature representation $g(x, y)$ to fixation probability $P_{\text{fix}}(x, y)$ is formulated as a non-linear supervised learning problem. The model aims to capture complex interactions and higher-order dependencies among heterogeneous perceptual features, whilst main-

taining interpretability for analysing the contribution of each factor in elderly attention mechanisms. Therefore, we adopt a probabilistic gradient boosting machine (PGBM) [64], an ensemble learning approach that models non-linear relationships through an additive sequence of decision trees. The final prediction is computed as:

$$P_{\text{fix}}(x, y) = \sigma(\sum_{t=1}^{T} \eta \cdot h_t(g(x, y))). \quad (9)$$

where T represents the number of boosting iterations, η represents the learning rate controlling ensemble convergence, and σ denotes the logistic sigmoid function mapping raw scores to fixation probabilities in [0, 1].

4.1. Model Structure

The EVAE model integrates six elderly-specific perceptual features into a unified framework designed for both accurate prediction and interpretable feature attribution. The architecture, illustrated in Figure 8, comprises four sequential modules that transform raw images into spatial attention predictions whilst maintaining transparency in feature contribution:

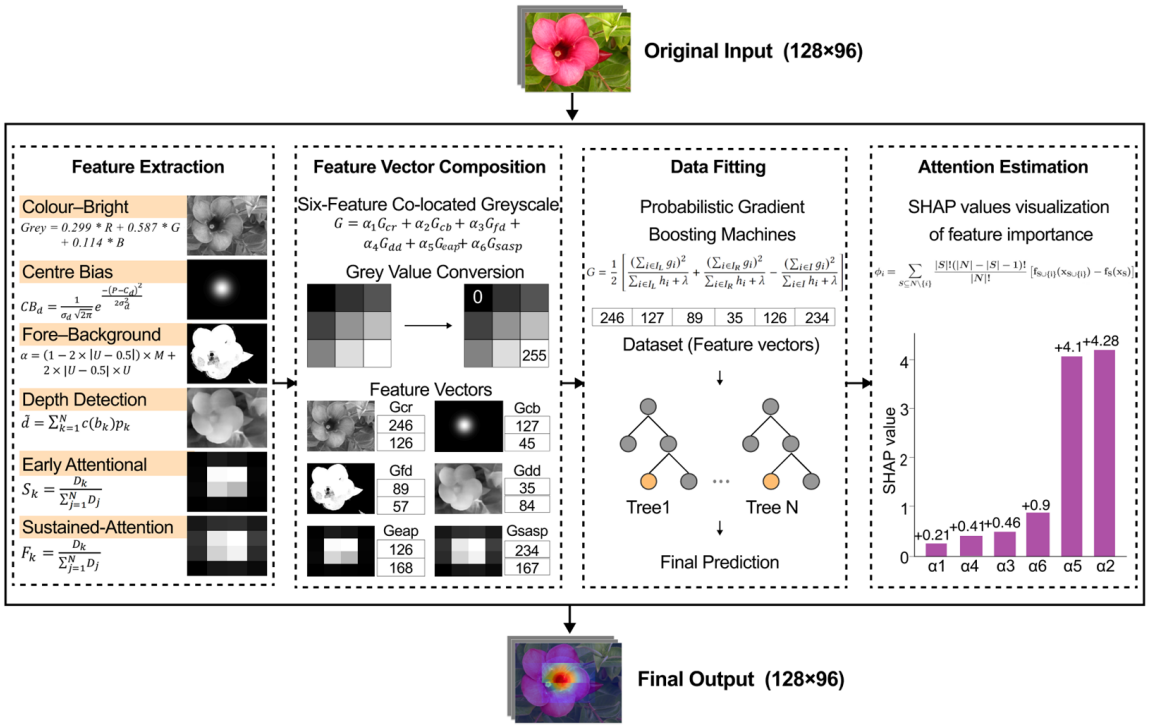

Figure 8. Overall structure of EVAE model.

- Module 1: Feature Extraction

This module generates the six complementary feature representations through parallel processing pipelines. Each feature employs specialised computational methods. Transformer-based networks extract depth information; adaptive segmentation isolates foreground–background regions; Gaussian kernels model centre bias tendencies; and statistical aggregation derives the early attentional and sustained-attention spatial priors from training fixation data. The module outputs six spatially registered feature maps, which capture both low-level visual properties (colour brightness, spatial structure) and high-level cognitive factors (attentional biases, temporal dynamics).

- Module 2: Feature Vector Composition

This module fuses heterogeneous feature maps into unified representations. It applies min–max normalisation to align all features to a [0, 1] scale, which preserves their relative spatial patterns. For each location (x, y), the normalised maps are combined as a six-dimensional vector $\left[G_{cr}(x,y), G_{cb}(x,y), G_{fd}(x,y), G_{dd}(x,y), G_{eap}(x,y), G_{sasp}(x,y)\right]^T$. This representation preserves both feature semantics and spatial context for subsequent learning.

- Module 3: Probabilistic Ensemble Learning

PGBM learns the non-linear mapping from feature vectors to fixation probabilities through iterative ensemble refinement. PGBM was selected for three critical properties. It naturally captures feature interactions, such as depth-amplifying centre bias in peripheral regions and early attentional prior modulating colour sensitivity, without requiring manual interaction term specification. It also handles heterogeneous feature types robustly; tree-based splits operate on raw feature orderings rather than Euclidean distances, eliminating sensitivity to scale variations. It provides feature importance quantification through permutation testing (measuring accuracy degradation when feature i is shuffled) and SHAP values (decomposing predictions into per-feature contributions with theoretical guarantees of local accuracy, missingness, and consistency). This interpretability directly addresses our research objective—quantifying which perceptual dimensions most strongly influence elderly attention allocation.

- Module 4: Attention Map Generation

This module transforms PGBM raw scores into normalised spatial attention maps. First, logistic transformation converts unbounded scores to probabilities [0, 1]. Second, divisive normalisation ensures valid probability distributions yielding pixel-wise fixation probabilities. The greyscale output format directly corresponds to empirical fixation density maps, which enables evaluation via standard saliency metrics (AUC-Judd, NSS, KL-divergence) and supports practical interface design applications where attention predictions guide visual element placement.

The four-module pipeline balances predictive accuracy through ensemble learning with scientific interpretability through feature attribution. It also maintains modularity for future methodological refinements. This architecture directly addresses the dual challenges of elderly visual attention modelling, namely, achieving high prediction performance on heterogeneous visual stimuli whilst providing actionable insights into the perceptual mechanisms governing attention allocation in ageing populations.

4.2. Model Procedure

(1) Feature extraction and ground truth generation

For each input image, we generated paired feature-target samples through parallel processing of visual content and empirical fixation data.

Feature Map Generation. Each source image was resized to 128 × 96 pixels using bicubic interpolation to maintain visual quality whilst standardising spatial dimensions. The resized images were then processed through the six feature extraction pipelines described in Section 3.2 to generate spatially aligned feature maps: G_{cr} (colour brightness), G_{cb} (centre bias), G_{fd} (foreground-background differentiation), G_{dd} (depth distribution), G_{eap} (early attentional prior distribution), and G_{sasp} (sustained-attention spatial prior). Each feature map represents a distinct perceptual or cognitive dimension relevant to elderly visual attention.

Ground Truth Derivation. The corresponding attention targets were derived from the eye-tracking dataset by Acik et al. [62], which recorded gaze data from 17 older adults

(M_{age} = 80.6 years) viewing 255 colour images (original resolution 1280 × 960 pixels). To specifically capture early attentional orienting, which is the pre-semantic phase where bottom-up salience and top-down biases interact, we aggregated fixation coordinates exclusively from the first two fixations of each trial, corresponding to approximately 1–2 initial saccades per image. Raw fixation points were convolved with a two-dimensional Gaussian kernel with standard deviation $\sigma = 1^\circ$ visual angle to produce continuous fixation density maps.

These maps were then normalised to [0, 1] range and downsampled to 128 × 96 resolution using bicubic interpolation, yielding greyscale heatmaps (Figure 9) that serve as prediction targets. Each pixel's intensity quantifies the empirical probability of fixation at that location, reflecting the actual spatial distribution of elderly visual attention across the image.

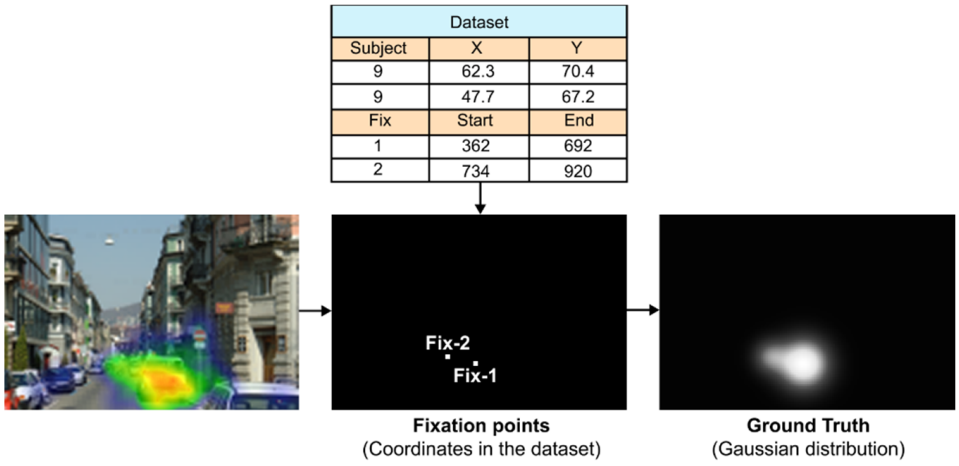

Figure 9. Procedure of ground truth preparation.

(2) Feature vector composition

This module fuses the six heterogeneous feature maps into a unified tabular representation suitable for supervised learning. To prevent scale imbalances between features with disparate value ranges, for instance, normalised depth values in [0, 1] vs. early attentional prior distributions in milliseconds, we first applied per-feature min–max normalisation ensuring all features occupy the [0, 1] range whilst preserving relative spatial patterns. For each spatial location, the normalised feature values were concatenated into a six-dimensional vector paired with the corresponding ground truth value from the fixation density map as its target label. This pixel-wise operation yields 12,288 training samples per image (128 × 96), each representing a specific spatial location with its associated feature characteristics and empirical fixation probability.

Applying this procedure across the training images produces a structured tabular dataset in CSV format with $N \times 12{,}288$ rows. Each row contains a six-dimensional feature vector $\mathrm{g}(x, y)$ and its associated fixation probability.

This tabular structure enables efficient batch processing during PGBM training (Figure 10) whilst maintaining explicit spatial correspondence for subsequent attention map reconstruction. The dataset size scales linearly with the number of training images.

(3) Probabilistic gradient boosting training

The data fitting module employs PGBM to learn the non-linear mapping from feature vectors to fixation probabilities.

Tree splits are optimised by maximising probabilistic gain. This criterion balances loss reduction with prediction confidence, which enhances robustness when handling diverse

features. For instance, it effectively accommodates depth maps with smooth gradients and early attentional prior distributions with sparse high-value regions (Figure 11).

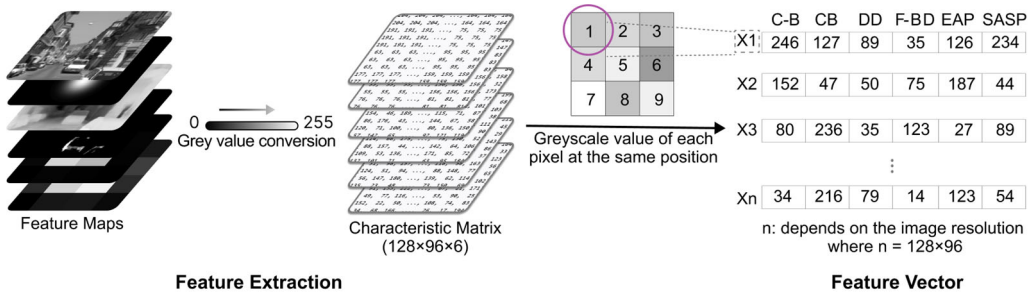

Figure 10. Procedure of feature fusion (CB: centre bias, EAP: early attentional prior, SASP: sustained-attention spatial prior, F–B D: foreground–background differentiation, DD: depth detection, C–B: colour brightness. The pink circle indicates that the same pixel position across all six feature maps is used to form a feature vector).

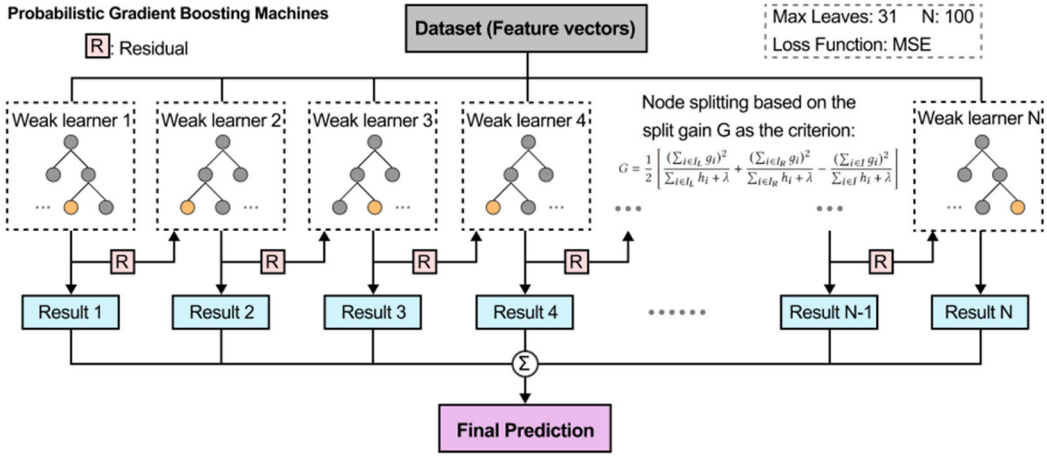

Figure 11. Procedure of model configuration.

4.3. Attention Map Generation and Visualisation

The final module applies the trained PGBM model to generate attention predictions for unseen test images. For each test image, the model processes all 12,288 feature vectors $g(x, y)$ through the ensemble, producing pixel-wise fixation probability estimates (Figure 12), where σ denotes the logistic function mapping unbounded PGBM scores to probabilities in [0, 1].

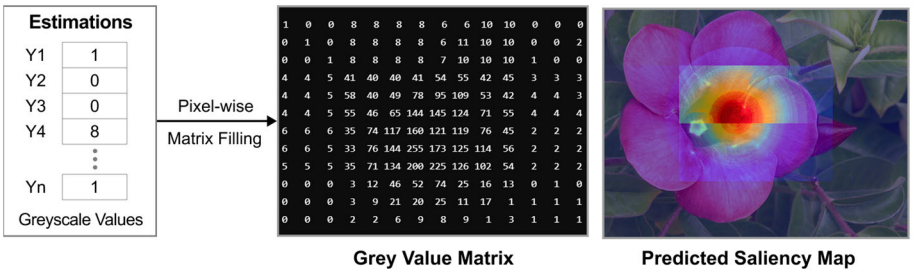

Figure 12. Procedure of attention estimation module.

The predicted heatmap can be upsampled to the original image resolution (1280 × 960 pixels) using bicubic interpolation and overlaid with semi-transparency (α = 0.6) to visualise estimated elderly attention distribution. This enables direct qualitative comparison with empirical fixation patterns for model validation, supports quantitative

evaluation via standard saliency metrics (AUC-Judd, NSS, KL-divergence), and facilitates practical applications in age-inclusive interface design, for instance, guiding placement of critical information within predicted high-attention regions or avoiding distractors in peripheral low-attention zones.

4.4. Model Evaluation

To assess the effectiveness of the proposed EVAE model, we conducted comprehensive evaluation using both qualitative and quantitative methods. The evaluation compared model predictions against ground truth fixation data across multiple image resolutions to provide robust assessment of prediction accuracy and computational scalability.

(1) Dataset preparation

We utilised a public age-stratified eye-tracking dataset validated by Wilming et al. [65] and originally collected by Acik et al. [62]. The dataset comprises gaze recordings from 58 participants across three age groups: children (n = 18, age 7–9, M = 7.6 years), young adults (n = 23, age 19–27, M = 22.1 years), and older adults (n = 17, age 72–88, M = 80.6 years). Participants viewed a subset of the stimulus pool (total 255 colour images) following a split-viewing protocol (e.g., each participant viewed approximately 128 images) across four categories: 64 naturals, 64 fractals, 64 manmades, and 63 pinks (Figure 13). Gaze data were recorded using an EyeLink 1000 eye-tracker (SR Research Ltd., Ottawa, ON, Canada) with 1000 Hz sampling rate. Following standard quality control during data preprocessing, valid trials were retained for subsequent analysis across all age cohorts.

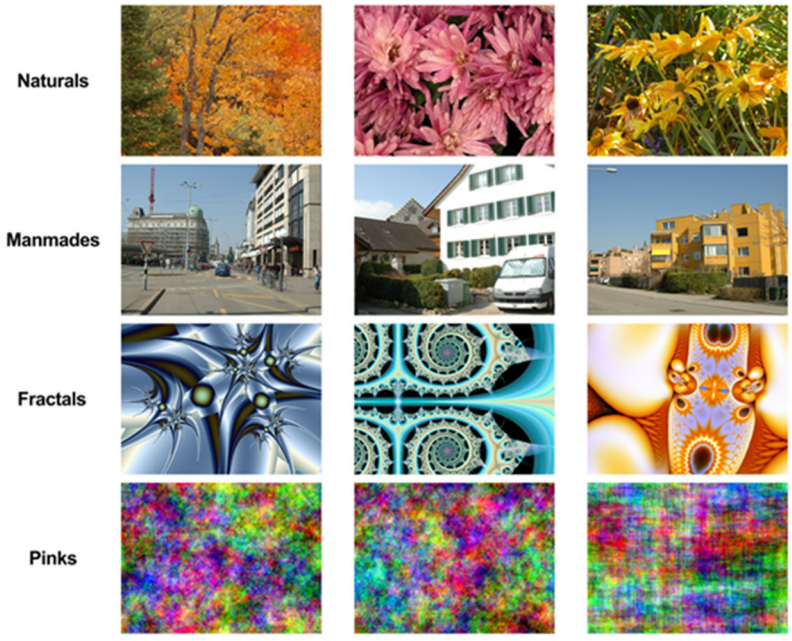

Figure 13. Sample images in four categories.

Given our model's focus on elderly visual attention, we exclusively analysed data from the older adult cohort (n = 17). For each image category, we randomly partitioned the dataset using an image-wise split into training (80%) and testing (20%) sets using stratified sampling to maintain category balance, which results in 204 training images and 51 testing images.

(2) Evaluation pipeline

Feature extraction. The six feature maps were extracted using the methods described in Section 3.3, then standardised to consistent spatial resolution through bicubic interpolation

and normalised to [0, 1] range via per-feature min–max scaling. For visualisation purposes, normalised values were mapped to eight-bit greyscale intensities (0–255) as shown in Figure 14.

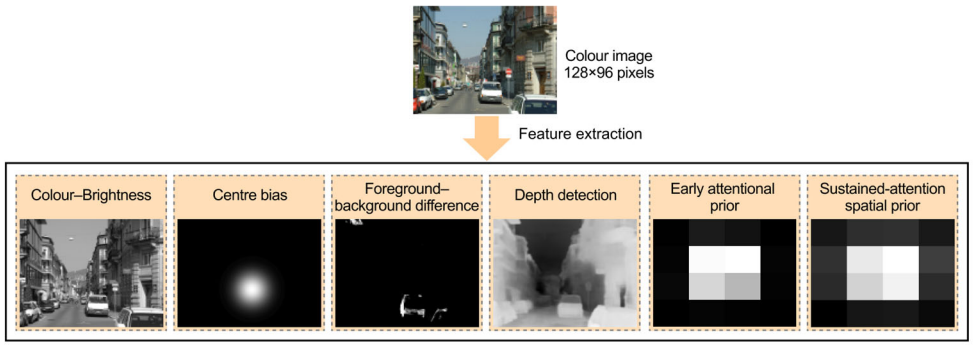

Figure 14. Feature extraction in evaluation.

To construct the dataset, feature vectors were extracted from all valid viewing trials in the training set. Aggregating pixel-level samples across all valid observer–image pairs yielded a total of approximately 26 million training samples (Figure 15). Each row in the constructed CSV file contains a six-dimensional feature vector $\mathrm{g}(x, y)$ paired with the corresponding fixation probability derived from the specific trial data.

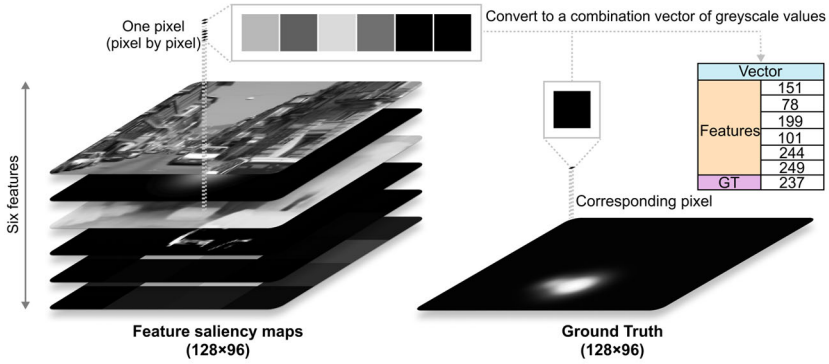

Figure 15. Feature vector construction.

Model Training. The model was implemented using the open-source pgbm Python library (version 2.3.0), using Mean Squared Error (MSE) as the loss function. To determine the optimal hyperparameter configuration, we performed a systematic grid search on the validation set over a constrained parameter space: learning rate $\eta \in \{0.01, 0.05, 0.1, 0.2\}$, maximum tree depth $d_{max} \in \{2, 4, 6, 8\}$, and L2 regularisation term $\lambda \in \{0, 0.01, 0.1\}$. The final configuration ($\eta = 0.1$, $d_{max} = 6$, $\lambda = 0.01$) was selected as it provided a stable trade-off between prediction accuracy and model complexity, with no significant performance gains observed beyond this range. We monitored the validation loss during training. Therefore, the number of boosting iterations was capped at 100 to prevent overfitting. For computational efficiency, training was conducted in batches of six million pixel-level samples on an NVIDIA A800 GPU (NVIDIA Corporation, Santa Clara, United States). To evaluate robustness across spatial scales, we trained separate models at eight resolution levels by systematically downsampling source images using bicubic interpolation: 8×6, 16×12, 32×24, 64×48, 128×96, 256×192, 512×384, and 1280×960 pixels.

Model Testing. The trained EVAE model was deployed on an NVIDIA A800 GPU for accelerated inference. For each of the 51 test images, the model generated pixel-wise fixation probability predictions following the pipeline. Predictions were upsampled to the

original resolution (1280 × 960 pixels) using bicubic interpolation and compared against ground truth fixation heatmaps using the metrics described below.

(3) Evaluation metrics

We employed a comprehensive metric suite [66,67] to evaluate complementary aspects of prediction quality.

Spatial Accuracy and Discrimination. AUC-Judd measures discrimination between fixated and non-fixated regions via area under the Receiver Operating Characteristic (ROC) curve, quantifying the model's ability to rank true fixation locations higher than non-fixated positions [64]. Values range from 0.5 (chance) to 1.0 (perfect discrimination). Normalised Scanpath Saliency (NSS) evaluates prediction strength at actual fixation points by measuring their z-score relative to mean saliency, where higher values indicate stronger alignment with empirical gaze patterns [28].

Distributional Similarity. Pearson Correlation Coefficient (CC) quantifies linear spatial correspondence between predicted and ground truth saliency maps, ranging from −1 to +1. Histogram Intersection (SIM) measures overlap in saliency value distributions, assessing probability mass similarity. Kullback–Leibler Divergence (KLD) quantifies distributional discrepancy via relative entropy, where lower values indicate closer alignment. Earth Mover's Distance (EMD) calculates the minimum 'work' required to transform one distribution into another, which captures spatial dissimilarity with units of pixel displacement.

These metrics collectively assess model performance across spatial accuracy (AUC-Judd, NSS), linear correspondence (CC), distributional fidelity (KLD, SIM), and geometric alignment (EMD), providing comprehensive characterisation of prediction quality.

5. Data Analysis and Results

5.1. Feature Influence

To reveal the contribution of each feature to attention prediction, we conducted feature importance analysis (SHAP values) using the SHAP Python library (version 0.47.0). SHAP provides an interpretation method that is not restricted to a specific model by computing each feature's marginal contribution to predictions [64,68,69], which enables transparent assessment of how visual features influence model outcomes (Figure 16).

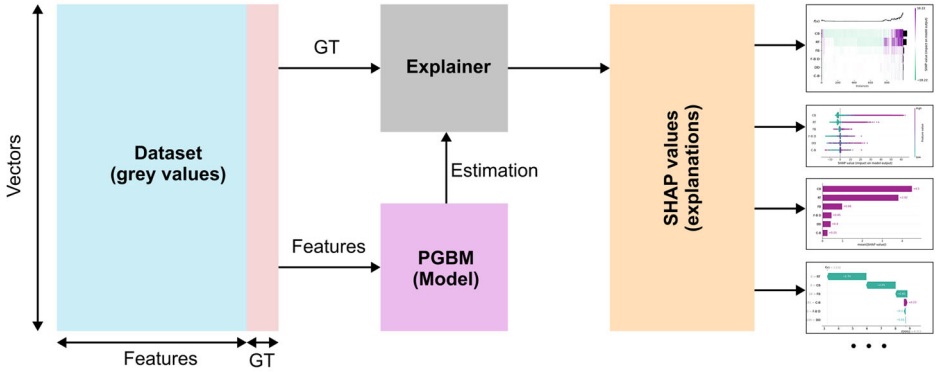

Figure 16. Analysis workflow of SHAP.

Figure 17 presents the mean absolute SHAP values across the six features. Centre bias and early attentional prior emerge as dominant contributors, with centre bias exhibiting the highest importance (SHAP = 4.28). This confirms the model's capacity to capture age-specific viewing behaviours, particularly the temporal dynamics characteristic of elderly fixation deployment. In contrast, spatial features, including foreground–background differentiation and depth detection, demonstrate relatively weaker impact and suggest that they are not primary determinants of elderly fixation patterns.

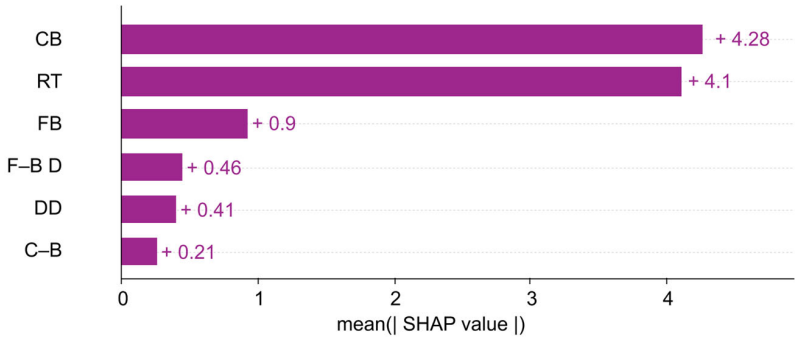

Figure 17. SHAP value of feature (CB: centre bias, EAP: early attentional prior, SASP: sustained-attention spatial prior, F–B D: foreground–background differentiation, DD: depth detection, C–B: colour brightness).

To examine spatial distribution of feature influence, we performed pixel-level SHAP analysis on images at 128 × 96 resolution. Figure 18 displays a SHAP heatmap for 1000 sampled pixels, where colour intensity represents feature impact magnitude. The heatmap reveals concentrated influence of centre bias and early attentional prior in central regions, which aligns with the documented centre-biased fixation patterns in elderly observers.

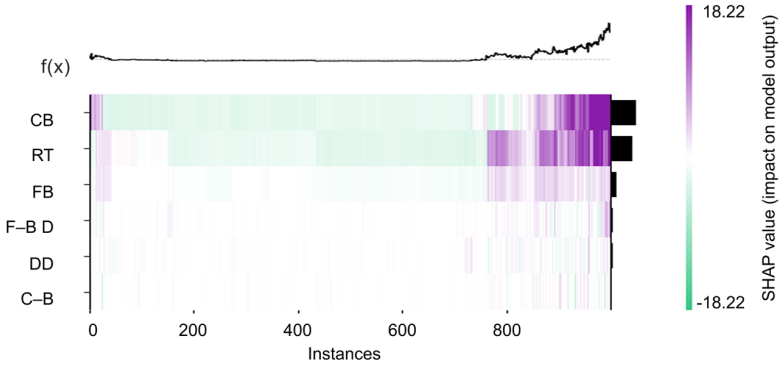

Figure 18. Heatmap result of spatial distribution of feature influence (CB: centre bias, EAP: early attentional prior, SASP: sustained-attention spatial prior, F–B D: foreground-background differentiation, DD: depth detection, C–B: colour brightness).

Figure 19 presents a scatter plot of SHAP values colour-coded by magnitude, which further reinforces the dominance of centre bias and early attentional prior through their concentration at high-saliency pixels. Sustained-attention spatial prior also exhibits notable contribution, which confirms the importance of spatial-temporal factors in elderly visual attention.

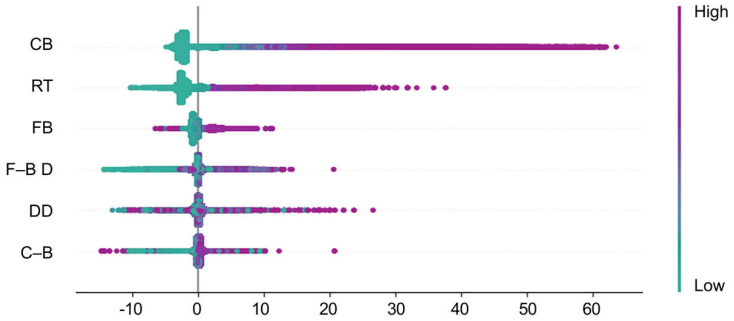

Figure 19. Scatter plot result of features' SHAP value distribution (CB: centre bias, EAP: early attentional prior, SASP: sustained-attention spatial prior, F–B D: foreground–background differentiation, DD: depth detection, C–B: colour brightness).

5.2. Qualitative Evaluation

Model effectiveness was assessed through alignment between predicted saliency maps and ground truth fixation distributions across scene categories and spatial resolutions. Figures 20 and 21 demonstrate that predictions consistently preserve high-density fixation regions, with minor deviations confined to low-saliency peripheral areas.

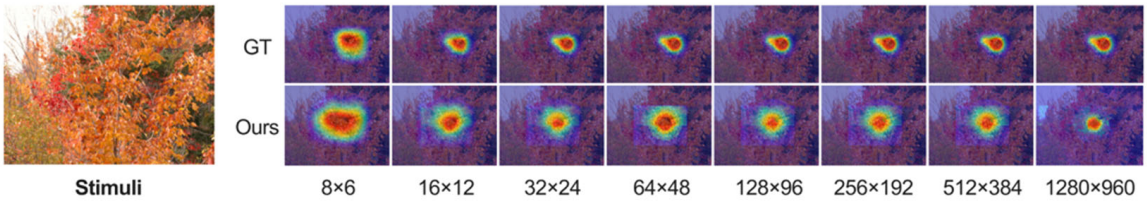

Figure 20. Comparison of ground truth and model-estimated saliency maps at different resolutions.

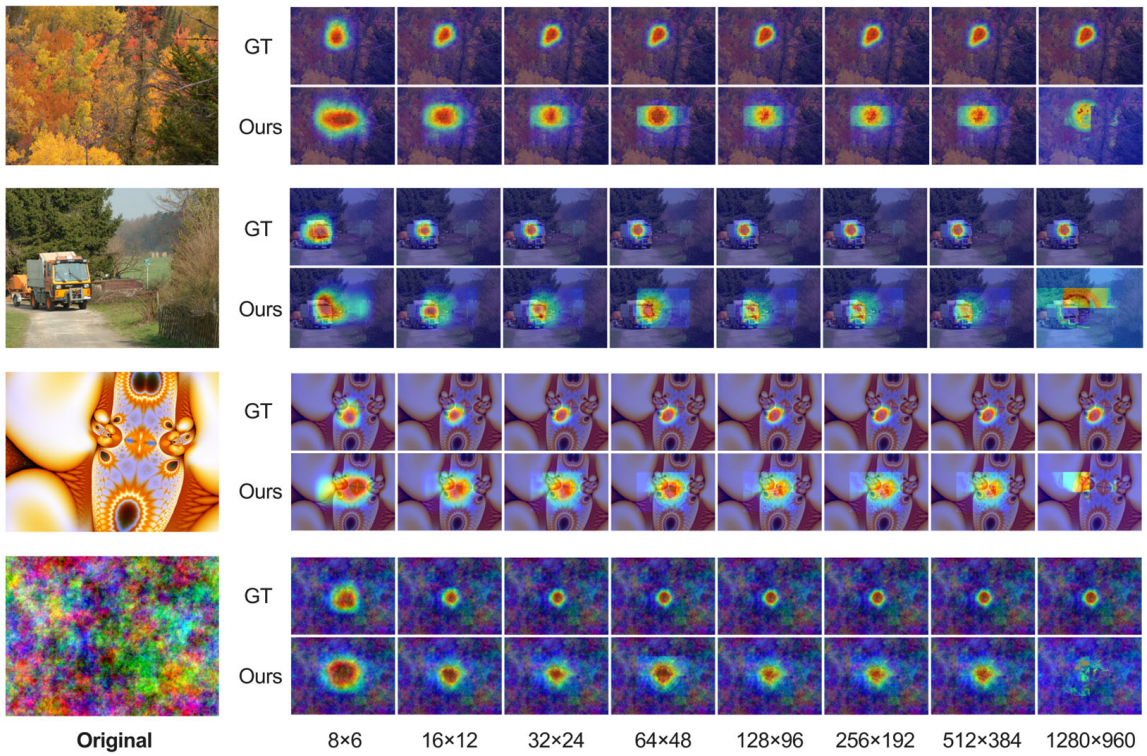

Figure 21. Comparative result of saliency map across image category model performance (GT: ground truth).

Qualitative analysis reveals robust category-specific adaptation.

- Naturals: Predictions align closely with ground truth at high-contrast regions such as textured foliage, object boundaries, and depth discontinuities, which indicates effective integration of colour, depth, and foreground–background.
- Manmades: Attention hotspots coincide with foreground structures, such as vehicles, roadways, and architectural elements, which demonstrates successful capture of spatial organisation and foreground–background separation through combined depth and centre bias signals.
- Fractals: The model accurately reproduces dispersed fixation patterns and density gradients characteristic of self-similar structures, which confirms robustness to statistical regularities absent in natural imagery.

- Pinks: Predictions exhibit pronounced centre concentration despite the absence of local salient features, which validates the centre bias module's effectiveness in encoding oculomotor preferences independent of bottom-up saliency.

This cross-category spatial pattern fidelity, which encompasses concentrated clusters (naturals), structured layouts (manmades), diffuse spreads (fractals), and centre-driven distributions (pinks), demonstrates effective integration of the six complementary attention factors. The qualitative observations are consistent with superior quantitative performance across all evaluation metrics.

5.3. Model Effectiveness

We evaluated the EVAE model using six established metrics: AUC-Judd (Area Under the Curve-Judd), NSS (Normalised Scanpath Saliency), CC (Correlation Coefficient), KLD (Kullback–Leibler Divergence), SIM (Similarity), EMD (Earth Mover's Distance), to compare its performance across resolutions and image categories. All evaluation metrics were calculated by comparing the model's pixel-wise predictions against the ground truth fixation maps. The quantitative results represent the mean performance scores averaged across all test images in each respective condition.

AUC-Judd. Table 1 shows consistently high scores (0.93) across most resolutions, with a slight decline at maximum resolution (1280 × 960). Pinks achieved highest performance (0.95) due to notable centre bias, while manmades showed marginally lower scores at fine scales, likely reflecting urban scene complexity. Results confirm strong fixation detection ability, particularly for structured attentional focus.

Table 1. Result of AUC-Judd analysis.

Resolution	All	Naturals	Manmades	Fractals	Pinks
8 × 6	0.93	0.94	0.93	0.94	0.95
16 × 12	0.93	0.93	0.92	0.93	0.95
32 × 24	0.93	0.93	0.92	0.93	0.94
64 × 48	0.90	0.90	0.89	0.88	0.92
128 × 96	0.92	0.92	0.90	0.91	0.94
256 × 192	0.91	0.91	0.90	0.90	0.93
512 × 384	0.91	0.91	0.89	0.90	0.93
1280 × 960	0.72	0.75	0.67	0.69	0.76

NSS. Table 2 demonstrates stable high scores at mid-range resolutions, peaking for pinks (2.93 at 512 × 384). Naturals and fractals maintained strong values, while manmades decreased at higher resolutions. Results indicate high sensitivity to fixation locations, especially in centre-focused content.

Table 2. Result of NSS analysis.

Resolution	All	Naturals	Manmades	Fractals	Pinks
8 × 6	1.87	1.97	1.75	1.93	2.11
16 × 12	2.25	2.17	1.96	2.21	2.65
32 × 24	2.33	2.21	1.95	2.24	2.88
64 × 48	2.18	2.10	1.79	2.12	2.71
128 × 96	2.33	2.19	1.96	2.26	2.90
256 × 192	2.31	2.18	1.95	2.24	2.89
512 × 384	2.32	2.18	1.94	2.25	2.93
1280 × 960	1.22	1.31	1.02	1.14	1.38

CC. Table 3 shows stable scores (0.44–0.54), peaking for pinks (0.60) and naturals (0.53) at coarse resolutions. Lower manmades scores suggest difficulty with structured artificial environments. Patterns confirm consistent spatial alignment with human fixations.

Table 3. Result of CC analysis.

Resolution	All	Naturals	Manmades	Fractals	Pinks
8 × 6	0.54	0.53	0.49	0.53	0.60
16 × 12	0.49	0.48	0.44	0.48	0.55
32 × 24	0.46	0.45	0.41	0.45	0.53
64 × 48	0.44	0.42	0.38	0.42	0.52
128 × 96	0.45	0.43	0.40	0.44	0.53
256 × 192	0.45	0.43	0.40	0.44	0.53
512 × 384	0.45	0.43	0.40	0.44	0.53
1280 × 960	0.25	0.25	0.21	0.23	0.29

KLD. Table 4 reveals minimal divergence for pinks (1.14 at 1280 × 960) with generally small values across resolutions. Occasional spikes in fractals and naturals at coarse scales reflect broader saliency spread. Results confirm high statistical similarity to human gaze data.

Table 4. Result of KLD analysis.

Resolution	All	Naturals	Manmades	Fractals	Pinks
8 × 6	5.59	5.37	5.72	5.87	5.43
16 × 12	2.45	2.31	2.18	2.23	3.05
32 × 24	2.43	2.36	2.23	2.30	2.84
64 × 48	8.95	8.81	9.07	9.33	8.64
128 × 96	2.16	2.06	2.14	1.90	2.51
256 × 192	2.13	1.97	2.07	1.93	2.54
512 × 384	1.98	1.82	1.98	1.83	2.29
1280 × 960	1.22	1.22	1.28	1.24	1.14

SIM. Table 5 shows highest overlap in pinks (0.48) and naturals (0.44) at low resolutions, reflecting strong spatial matching. Manmades and fractals show lower overlap at high resolution, indicating greater variation. Results support content-dependent spatial match quality.

Table 5. Result of SIM analysis.

Resolution	All	Naturals	Manmades	Fractals	Pinks
8 × 6	0.44	0.44	0.41	0.44	0.48
16 × 12	0.38	0.38	0.35	0.37	0.42
32 × 24	0.36	0.35	0.33	0.35	0.40
64 × 48	0.37	0.36	0.33	0.36	0.41
128 × 96	0.35	0.34	0.32	0.34	0.39
256 × 192	0.35	0.34	0.32	0.34	0.40
512 × 384	0.35	0.34	0.31	0.33	0.39
1280 × 960	0.18	0.19	0.17	0.18	0.19

EMD. Table 6 demonstrates lowest values for pinks (1.63 at 512 × 384), with higher values for fractals at high resolution. Mid-range resolutions maintained low EMD (1.77–1.98), indicating minimal spatial transformation required. Results confirm spatial accuracy across viewing conditions.

Table 6. Result of EMD analysis.

Resolution	All	Naturals	Manmades	Fractals	Pinks
8 × 6	1.78	1.76	1.76	1.76	1.76
16 × 12	1.77	1.76	1.87	1.80	1.64
32 × 24	1.78	1.78	1.88	1.82	1.64
64 × 48	1.98	1.98	2.08	2.04	1.80
128 × 96	1.78	1.77	1.90	1.80	1.64
256 × 192	1.78	1.78	1.89	1.83	1.65
512 × 384	1.77	1.76	1.89	1.81	1.63
1280 × 960	3.37	3.29	3.48	3.35	3.36

Age-Specificity Validation

To validate age-specificity, we applied the model to children (n = 18) and young adults (n = 23) at 128 × 96 resolution (Table 7). Results reveal substantial performance gaps. Older adults achieved AUC-J of 0.92 vs. 0.83 (children) and 0.81 (young adults). NSS for older adults (2.33) exceeded children (1.48) and young adults (1.28), indicating effective capture of characteristic high-intensity fixation clusters in elderly behaviour vs. exploratory scanning in younger groups. CC for older adults (0.45) showed stronger linear correspondence than other groups (0.36, 0.32). EMD was lowest for older adults (1.78), requiring minimal spatial transformation to align with ground truth. While KLD remained comparable across groups, consistent superiority in spatial accuracy metrics confirms intrinsic alignment with elderly physiological and cognitive visual mechanisms.

Table 7. Comparative performance of the EVAE model across different age groups (128 × 96).

Metric	AUC-J	NSS	CC	KLD	SIM	EMD
Children	0.83	1.48	0.36	2.13	0.30	2.03
Young Adults	0.81	1.28	0.32	2.24	0.28	2.16
Older Adults	0.92	2.33	0.45	2.16	0.35	1.78

Comparative Analysis

Table 8 compares EVAE against established saliency models and DeepGaze II, with sample predictions in Figure 22. Traditional bottom-up models (Itti, GBVS, Patch) yielded AUC-J scores of 0.54–0.57 with NSS near zero, demonstrating inadequacy of low-level visual features alone for elderly attention prediction. DeepGaze II showed poor performance (AUC-J = 0.41, NSS = −0.26) when applied at low resolution. This can be attributed to two factors. First, the resolution mismatch compromised its feature extraction capabilities; second, and more importantly, a fundamental domain shift exists between DeepGaze II's training target and our testing ground truth. Even when evaluated at native resolution (1280 × 960), DeepGaze II achieved an AUC-J of only 0.68, which confirms that generic saliency models do not generalise well to the specific spatial biases of the ageing population.

Table 8. Comparison result of EVAE and state-of-the-art saliency models (Older Adults, 128 × 96).

Model	AUC-J	NSS	CC	KLD	SIM	EMD
Itti [26]	0.57	0.21	0.05	1.46	0.14	3.02
GBVS [27]	0.54	0.13	0.03	1.59	0.13	3.08
Judd [46]	0.74	0.80	0.19	1.22	0.17	2.76
Patch [70]	0.57	0.19	0.05	1.58	0.14	3.06
DeepGaze II [28]	0.41	−0.26	−0.06	1.83	0.11	3.59
DeepGaze II (1280 × 960)	0.68	0.43	0.12	1.45	0.34	0.12
Ours	0.92	2.33	0.45	2.16	0.35	1.78

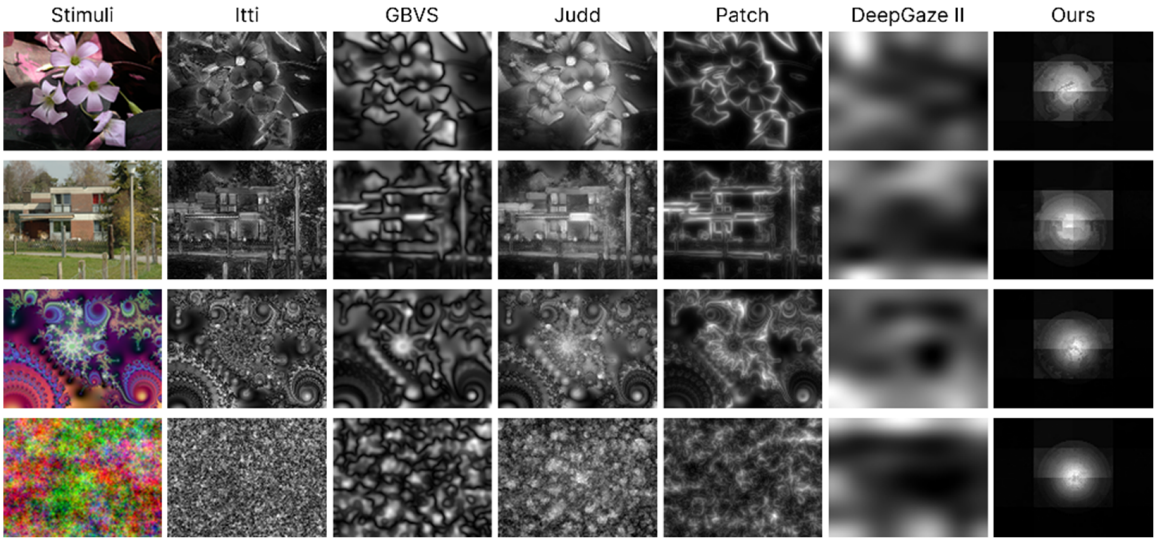

Figure 22. Comparison of saliency maps estimated by the EVAE model and the state-of-the-art models.

Ablation Study

To assess the relative contributions of top-down cognitive priors and bottom-up visual features, we performed an ablation analysis. To ensure a fair comparison, the PGBM model was retrained from scratch for each configuration using identical hyperparameters and the same train-test dataset splits. Each variant was then evaluated on the test set using the previously defined metrics. We compared five distinct feature configurations:

- Image-derived features only (colour brightness, foreground–background differentiation, depth detection);
- Image-derived features + centre bias;
- Image-derived features + gaze priors (early attentional prior and sustained-attention spatial prior);
- Gaze priors only (early attentional prior and sustained-attention spatial prior);
- Centre bias only.

As shown in Table 9, the image-derived-features-only configuration yielded the lowest performance, which confirms that purely bottom-up saliency is insufficient for modelling elderly attention. To evaluate whether the model's performance stems from a trivial overfitting to the centre bias, we introduced a centre-bias-only baseline. While this spatial prior achieves an AUC-J of 0.83, confirming the prevalence of central fixation in the elderly cohort, it fails to capture the precise distribution of attention, as evidenced by its high KLD (7.05) and lower NSS (1.65). The full EVAE model outperforms this baseline across all metrics, achieving an AUC-J of 0.92. The gaze priors only configuration established a strong baseline, which confirms the robust influence of top-down spatial strategies in ageing vision. The combinations of image-derived features with either centre bias or gaze priors yielded progressive improvements, while the full EVAE model (integrating all six features) achieved the highest overall accuracy. These results support our observation that optimal prediction of elderly visual attention requires combining both scene-specific visual structures and population-level cognitive priors.

Table 9 shows that the full model yields a higher KLD value than certain partial configurations, notably 'Image + Centre bias' (2.16 vs. 1.11). This does not indicate a failure of the model. KLD is sensitive to differences in distributional entropy, and penalises predictions that are more spatially concentrated than the ground truth fixation density. By integrating discretised spatial priors (EAP and SASP) through non-linear PGBM fusion, the full model produces sharper and more confident predictions. This directly contributes to

improved AUC-Judd and NSS, as well as the lowest EMD (1.78), indicating accurate spatial alignment of attention. The increased KLD therefore reflects a mismatch in distributional smoothness rather than an error in attentional localisation. This trade-off between spatial precision and distributional smoothness is a recognised characteristic of saliency models incorporating strong spatial or top-down priors and does not imply redundancy or conflict between the integrated features.

Table 9. Quantitative evaluation of ablation experiments (older adults, 128 × 96).

Model Configuration	AUC-J	NSS	CC	KLD	SIM	EMD
Image-derived features only	0.62	0.42	0.10	1.66	0.14	2.94
Image-derived features + centre bias	0.85	1.75	0.42	1.11	0.33	1.96
Image-derived features + gaze priors	0.87	1.51	0.37	1.15	0.29	1.89
Gaze priors only	0.86	1.43	0.35	1.19	0.30	1.92
Centre bias only	0.83	1.65	0.40	7.05	0.36	2.34
EVAE model (All six features)	0.92	2.33	0.45	2.16	0.35	1.78

6. Discussion

This study advances understanding of older adults' visual attention through systematic feature analysis and computational modelling. Our SHAP-based feature importance analysis identified centre bias and early attentional prior as dominant contributors to older adults' fixation patterns, with centre bias exhibiting the highest marginal contribution (SHAP = 4.28, Figure 17). Based on these insights, we developed the EVAE model, which achieves 92% prediction accuracy (AUC-Judd = 0.92, Table 1), establishing a specialised computational framework for elderly early visual attention estimation that combines age-specific behavioural priors with model interpretation analysis in ergonomics and human–computer interaction research. This approach provides a scalable, cost-effective alternative to labour-intensive eye-tracking studies and supports further research in age-inclusive interface design.

Our findings build upon existing research identifying colour brightness, foreground–background differentiation, depth detection, and sustained-attention spatial prior as relevant features [2,50]. Our contribution lies in systematic integration rather than isolated investigation. By modelling these features as an ensemble within a unified predictive framework, we empirically quantified their relative contributions and identified centre bias and early attentional prior as consistently dominant across diverse visual contexts (Figure 17). Our approach therefore moves beyond identifying relevant features to quantifying their relative contributions in a hierarchical manner.

The dominance of the early attentional prior and sustained-attention spatial prior in our SHAP analysis warrants careful interpretation. A critical concern might be whether the model simply learns a static 'elderly fixation template' rather than predicting attention based on image content. However, from a cognitive ageing perspective, this high reliance on global spatial templates is theoretically sound. Due to reduced peripheral acuity and slower processing speeds, older adults systematically rely more on top-down, conservative spatial strategies to compensate for weakened bottom-up sensory processing [49,62]. The global priors in our model effectively capture this physiological baseline.

The prominence of centre bias and early attentional prior provides insight into age-related shifts in attentional control. Existing human factors research documents older adults' visual attention variability stemming from physiological decline [71]. The model's performance reflects existing theories of compensatory mechanisms. As bottom-up, salience-driven processing weakens with age, both human observers and our predictive model increasingly rely on top-down, conservative spatial strategies. This interpretation is sup-

ported by cross-age validation (Table 7), where EVAE performance degraded substantially for younger adults (AUC-Judd 0.81–0.83 vs. 0.92 for older adults), indicating that older adults' attention operates under distinct mechanisms inadequately captured by generic saliency models.

Regarding the foreground–background feature, the foreground–background component should be interpreted with caution. Increased fixation on background regions in older adults may arise from compensatory scanning caused by less efficient foreground extraction, rather than from an active preference for contextual information. EVAE does not separate these mechanisms in this stage. Its purpose is to model the resulting spatial distribution of fixations, not to identify the causal perceptual process that generated them. By assigning relatively higher weights to background regions, the model reflects the more diffuse fixation patterns observed in ageing vision, which improves estimation accuracy without modelling the underlying failure recovery processes of the visual system.

Our use of SHAP analysis in addition to traditional ablation studies has methodological justification grounded in three principles of interpretable machine learning [72,73]. First, our feature set exhibits inherent spatial correlations, where centre bias and fixation density naturally co-vary due to shared spatial structure [5]. Ablation studies yield misleading importance estimates under multicollinearity, as models compensate for removed features through correlated proxies [69]. SHAP addresses this through Shapley value computation, which equitably distributes credit across correlated features [74]. Second, our objective is to explain the trained EVAE model's decision mechanism. Ablation requires retraining after feature removal, introducing distributional shifts that generate fundamentally different model instances, known as the "Rashomon effect" [75]. SHAP interprets existing model parameters without retraining confounds. Third, SHAP provides pixel-level granularity (Figures 18 and 19), enabling spatial visualisation of feature influence that is unattainable through global ablation metrics.

Our data reveal a pattern requiring careful interpretation. SHAP analysis identified strong centre bias influence (Figure 17), yet empirical fixation distributions for older participants showed greater spatial dispersion than younger adults across image categories. This apparent contradiction reflects the interaction between cognitive strategy and physiological constraint. While older observers exhibit stronger tendency towards central fixation, which is likely a compensatory strategy to reduce peripheral processing demands amid declining acuity [12], their realised attention patterns show greater instability and scatter.

Centre bias represents a top-down strategic preference, but age-related visual decline (reduced acuity, slower processing) disrupts execution, yielding dispersed fixation patterns. The model captures both phenomena, with centre bias representing a predictive feature (strategic component) and dispersed distributions as outcome (execution component). Future work could test this dual mechanism directly by manipulating scene complexity while independently controlling peripheral visual demands.

Centre bias persisted even in abstract fractal images lacking semantic content, which suggests that spatial preference constitutes a fundamental visual strategy rather than content-driven guidance. Whether this represents cognitive load reduction or default neural tendency in ageing visual systems requires further investigation.

Increased fixation duration variability in complex scenes indicates attentional instability, yet simplified stimuli (e.g., pink noise) elicited concentrated central fixations, which suggests that scene complexity modulates the balance between strategic bias and execution stability.

The superior performance of EVAE over DeepGaze II (Table 8) provides insight into model design for specialised populations. However, this comparison highlights the significant domain bias inherent in general-purpose models. DeepGaze II exhibits poor

performance (AUC-J = 0.41, NSS = −0.26) on the elderly dataset primarily because its internal representations are optimised for the visual behaviours of younger adults. This gap underscores the necessity of our age-specific approach. The superiority of EVAE in this context is attributed to its architectural suitability for the task. Unlike black-box deep models that require vast amounts of data to learn implicit features, EVAE leverages PGBM to fuse six features. Given our specialised dataset scale, PGBM provides a robust balance between predictive power and scientific interpretability. It demonstrates that even highly sophisticated deep learning architectures fail to generalise to the ageing population without explicit adaptation to age-related features.

While EVAE demonstrates robust performance across standard benchmarks, the results reveal certain boundary conditions where its predictive power may be constrained. One primary limitation manifests in scenarios with high semantic conflict, such as images containing small but critically important targets in peripheral regions. The model may under-predict fixations when a user's attention is captured by high-level meaning that overrides the stimulus-driven saliency.

The model exhibits sensitivity to extreme visual clutter. In environments with high visual entropy and dense textures, the foreground–background differentiation may provide less diagnostic gain, leading to more dispersed saliency maps. Furthermore, the current model is optimised for static images. In real-world interactive interfaces where dynamic elements (e.g., animations or pop-ups) are present, the features may limit its applicability.

Beyond its predictive capabilities, the EVAE model holds significant potential as a foundational component for generative design and adaptive user interfaces. By integrating EVAE's attention maps into automated design pipelines, the model can function as a spatial constraint or a fitness function in optimisation algorithms. Such a system could automatically rearrange interface elements positioning critical navigation cues or emergency notifications within the predicted high-attention regions. This transition would enable the creation of self-adjusting, age-inclusive digital environments that dynamically align with the physiological baselines of older users.

This study has limitations that should be acknowledged. First, reliance on a single public dataset limits generalisability across elderly cognitive and visual diversity. The current analysis primarily represents relatively healthy ageing; however, neurodegenerative conditions such as Alzheimer's disease (AD) are known to further impair oculomotor control, potentially leading to even more pronounced central fixation or reduced peripheral awareness. Future research should explore whether EVAE can be adapted as a non-invasive screening tool by identifying deviations in the attentional priors of AD patients. Furthermore, considering that visual scanning patterns are influenced by cultural cognitive styles, validating the model across diverse cultural backgrounds is an essential next step. Integrating these exploratory directions would ensure that age-inclusive design tools remain equitable and effective for a global, heterogeneous ageing population. Second, Gaussian modelling of fixation density, while standard, may inadequately represent elderly attention's increased spatial dispersion. Alternative probabilistic models (e.g., mixture distributions, kernel density estimation with adaptive bandwidths) warrant investigation. Additionally, a limitation of the current implementation is that age-related optical changes are not explicitly modelled at the image-transformation stage. Future work could combine EVAE with calibrated models of ageing-related colour appearance or with participant-level measures of visual acuity, contrast sensitivity, and lens density. Third, our focus on static images excludes dynamic stimuli prevalent in real-world interfaces. Extending EVAE to video and interactive content requires incorporating temporal prediction and action–attention coupling.

The choice of 128 × 96 as the internal processing resolution reflects a deliberate alignment between architectural design and behavioural scope. The EVAE model targets the early attentional orienting phase of interface viewing, during which fixation behaviour is predominantly governed by global scene structure, luminance salience, and spatial prior knowledge rather than fine-grained semantic content. At this temporal scale, a coarse spatial representation is empirically appropriate. However, this design introduces a clear constraint on applicability. Interactive elements with small spatial extent, including icons, form controls, and peripheral indicators, may be insufficiently represented at this resolution. This is consistent with our observation that the model systematically underpredicts fixations on small, peripherally located targets.

Accordingly, the practical contribution of EVAE should be precisely scoped. The model is best understood as a scalable tool for structural accessibility auditing, specifically for evaluating whether primary organisational regions of an interface align with the spatial attention characteristics of older adult users. It is not a substitute for hardware-based eye-tracking in tasks requiring element-level fixation precision. For design decisions involving fine spatial granularity, complementary approaches, including empirical user studies or high-resolution saliency models, remain necessary. Future work will investigate multi-scale architectures capable of integrating global structure with local detail, with the aim of extending predictive reliability to smaller interface elements without compromising computational efficiency in early-stage design evaluation.

7. Conclusions

We introduced the EVAE model, a computational method that integrates six empirically validated perceptual features to predict visual attention in older adults. Unlike generic saliency models, EVAE explicitly incorporates age-specific mechanisms, including reduced colour sensitivity and compensatory spatial strategies. Our key findings demonstrate that elderly visual attention is not only a degraded version of younger attention, but also a distinct process affected by compensatory strategies. Through SHAP-based interpretability analysis, the results indicate that centre bias and early attentional prior (representing early orienting spatial priors) are the dominant predictors of elderly fixation patterns. This confirms our hypothesis that as bottom-up sensory processing declines, older adults rely increasingly on top-down, conservative spatial strategies. The proposed method achieves strong performance (AUC-Judd = 0.92) on elderly datasets, which significantly outperforms deep learning baselines trained on younger populations. The implications of this work extend to both theoretical research and practical application. Theoretically, we provide computational evidence for the 'processing speed–accuracy trade-off' in ageing vision, quantifying how physiological constraints shape attentional deployment. In practice, EVAE offers designers and ergonomists a scalable, cost-effective tool to evaluate interface accessibility without requiring specialised eye-tracking equipment.

Author Contributions: All authors have agreed to this submission and have contributed to the research and writing process as outlined below: X.L.: Writing—review and editing, methodology, investigation, funding acquisition, and conceptualisation. X.S.: Writing—review and editing, writing—original draft, formal analysis, investigation, data curation, conceptualisation, and software programming. H.G.: Writing—review and editing. T.S.: Writing—review and editing. S.C.: Writing—review and editing. J.W.: Writing—review and editing. All authors have read and agreed to the published version of the manuscript.

Funding: The research work is supported by Key Research and Development Program of Zhejiang Province (grant number 2024C01210) and Ministry of Education Humanities and Social Sciences Project (grant number 24YJC760066).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data that support the findings of this study are available from Dryad | Data—An extensive dataset of eye movements during viewing of complex images, https://datadryad.org/dataset/doi:10.5061/dryad.9pf75 (accessed on 19 March 2026).

Acknowledgments: During the preparation of this manuscript/study, the authors used generative artificial intelligence tools for the purpose of language polishing.

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

Abbreviations

The following abbreviations are used in this manuscript:

EVAE	Elderly Visual Attention Estimation
PGBM	Probabilistic Gradient Boosting Machines
SHAP	SHapley Additive exPlanations
AUC-J	Area Under the Curve-Judd
NSS	Normalised Scanpath Saliency
CC	Correlation Coefficient
KLD	Kullback–Leibler Divergence
SIM	Similarity
EMD	Earth Mover's Distance

References

1. Nagarajan, N.; Assi, L.; Varadaraj, V.; Motaghi, M.; Sun, Y.; Couser, E.; Ehrlich, J.R.; Whitson, H.; Swenor, B.K. Vision impairment and cognitive decline among older adults: A systematic review. *BMJ Open* **2022**, *12*, e047929. [CrossRef]
2. Krishna, O.; Aizawa, K.; Irie, G. Computational attention system for children, adults and elderly. *arXiv* **2019**, arXiv:1904.12628. [CrossRef]
3. Pagano, T.P.; Loureiro, R.B.; Lisboa, F.V.N.; Peixoto, R.M.; Guimarães, G.A.S.; Cruz, G.O.R.; Araujo, M.M.; Santos, L.L.; Cruz, M.A.S.; Oliveira, E.L.S.; et al. Bias and unfairness in machine learning models: A systematic review on datasets, tools, fairness metrics, and identification and mitigation methods. *Big Data Cogn. Comput.* **2023**, *7*, 15. [CrossRef]
4. Saravanan, C. Color image to grayscale image conversion. In *Proceedings of the 2010 Second International Conference on Computer Engineering and Applications (ICCEA)*; IEEE: Piscataway, NJ, USA, 2010; pp. 196–199. [CrossRef]
5. Hayes, T.R.; Henderson, J.M. Center bias outperforms image salience but not semantics in accounting for attention during scene viewing. *Atten. Percept. Psychophys.* **2020**, *82*, 985–994. [CrossRef]
6. Strauch, C.; Hoogerbrugge, A.J.; Baer, G.; Hooge, I.T.C.; Nijboer, T.C.W.; Stuit, S.M.; Van der Stigchel, S. Saliency models perform best for women's and young adults' fixations. *Commun. Psychol.* **2023**, *1*, 34. [CrossRef] [PubMed]
7. Tsvetanov, K.A.; Mevorach, C.; Allen, H.; Humphreys, G.W. Age-related differences in selection by visual saliency. *Atten. Percept. Psychophys.* **2013**, *75*, 1382–1394. [CrossRef]
8. Zhang, X.; Jiang, Y.; Hou, W.; Jiang, N. Age-related differences in the transient and steady state responses to different visual stimuli. *Front. Aging Neurosci.* **2022**, *14*, 1004188. [CrossRef] [PubMed]
9. Yang, G.; Tang, H.; Ding, M.; Sebe, N.; Ricci, E. Transformer-based attention networks for continuous pixel-wise prediction. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*; IEEE: Piscataway, NJ, USA, 2021; pp. 16249–16259. [CrossRef]
10. Theeuwes, J. Top-down and bottom-up control of visual selection. *Acta Psychol.* **2010**, *135*, 77–99. [CrossRef]
11. Zito, G.A.; Cazzoli, D.; Scheffler, L.; Jäger, M.; Müri, R.M.; Mosimann, U.P.; Nyffeler, T.; Mast, F.W.; Nef, T. Street crossing behavior in younger and older pedestrians: An eye- and head-tracking study. *BMC Geriatr.* **2015**, *15*, 176. [CrossRef]
12. Failing, M.; Theeuwes, J. Selection history: How reward modulates selectivity of visual attention. *Psychon. Bull. Rev.* **2018**, *25*, 514–538. [CrossRef]
13. Chu, C.H.; Nyrup, R.; Leslie, K.; Shi, J.; Bianchi, A.; Lyn, A.; McNicholl, M.; Khan, S.; Rahimi, S.; Grenier, A. Digital ageism: Challenges and opportunities in artificial intelligence for older adults. *Gerontologist* **2022**, *62*, 947–955. [CrossRef] [PubMed]
14. Zhou, F.; Yang, X.J.; de Winter, J.C.F. Using eye-tracking data to predict situation awareness in real time during takeover transitions in conditionally automated driving. *IEEE Trans. Intell. Transp. Syst.* **2022**, *23*, 2284–2295. [CrossRef]

15. Pepper, J.L.; Nuttall, H.E. Age-related changes to multisensory integration and audiovisual speech perception. *Brain Sci.* **2023**, *13*, 1126. [CrossRef] [PubMed]
16. Wolinsky, F.D.; Weg, M.W.V.; Howren, M.B.; Jones, M.P.; Dotson, M.M. A randomized controlled trial of cognitive training using a visual speed of processing intervention in middle aged and older adults. *PLoS ONE* **2013**, *8*, e61624. [CrossRef]
17. Treisman, A.M.; Gelade, G. A feature-integration theory of attention. *Cogn. Psychol.* **1980**, *12*, 97–136. [CrossRef]
18. Koch, C.; Ullman, S. Shifts in selective visual attention: Towards the underlying neural circuitry. In *Matters of Intelligence: Conceptual Structures in Cognitive Neuroscience*; Vaina, L.M., Ed.; Springer: Dordrecht, The Netherlands, 1987; pp. 115–141. [CrossRef]
19. Owsley, C. Aging and vision. *Vis. Res.* **2011**, *51*, 1610–1622. [CrossRef]
20. Lanssens, A.; Desender, K.; Gillebert, C.R. Evidence for an age-related decline in feature-based attention. *Aging Neuropsychol. Cogn.* **2024**, *31*, 846–868. [CrossRef] [PubMed]
21. Krishna, O.; Aizawa, K. Age-adapted saliency model with depth bias. In *Proceedings of the ACM Symposium on Applied Perception*; Association for Computing Machinery: New York, NY, USA, 2017; pp. 1–8. [CrossRef]
22. Lass, J.W.; Bennett, P.J.; Peterson, M.A.; Sekuler, A.B. Effects of aging on figure-ground perception: Convexity context effects and competition resolution. *J. Vis.* **2017**, *17*, 15. [CrossRef]
23. Song, H.; Shim, W.M.; Rosenberg, M.D. Large-scale neural dynamics in a shared low-dimensional state space reflect cognitive and attentional dynamics. *eLife* **2023**, *12*, e85487. [CrossRef]
24. van der Laan, L.N.; Hooge, I.T.C.; de Ridder, D.T.D.; Viergever, M.A.; Smeets, P.A.M. Do you like what you see? The role of first fixation and total fixation duration in consumer choice. *Food Qual. Prefer.* **2015**, *39*, 46–55. [CrossRef]
25. Lawrence, R.K.; Edwards, M.; Goodhew, S.C. Changes in the spatial spread of attention with ageing. *Acta Psychol.* **2018**, *188*, 188–199. [CrossRef] [PubMed]
26. Itti, L.; Koch, C.; Niebur, E. A model of saliency-based visual attention for rapid scene analysis. *IEEE Trans. Pattern Anal. Mach. Intell.* **1998**, *20*, 1254–1259. [CrossRef]
27. Harel, J.; Koch, C.; Perona, P. Graph-based visual saliency. In *Advances in Neural Information Processing Systems 19 (NIPS 2006)*; MIT Press: Cambridge, MA, USA, 2006; pp. 545–552. Available online: https://proceedings.neurips.cc/paper_files/paper/2006/hash/4db0f8b0fc895da263fd77fc8aecabe4-Abstract.html.
28. Kümmerer, M.; Wallis, T.S.A.; Bethge, M. DeepGaze II: Reading fixations from deep features trained on object recognition. *Int. J. Comput. Vis.* **2019**, *127*, 1124–1142. Available online: https://arxiv.org/abs/1610.01563 (accessed on 19 March 2026).
29. Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S.; et al. An image is worth 16x16 words: Transformers for image recognition at scale. In Proceedings of the International Conference on Learning Representations (ICLR), Virtual, 3–7 May 2021. Available online: https://openreview.net/forum?id=YicbFdNTTy (accessed on 19 March 2026).
30. Lou, J.; Lin, H.; Marshall, D.; Saupe, D.; Liu, H. TranSalNet: Towards perceptually relevant visual saliency prediction. *Neurocomputing* **2022**, *494*, 455–467. [CrossRef]
31. Liu, Z.; Lin, Y.; Cao, Y.; Hu, H.; Wei, Y.; Zhang, Z.; Lin, S.; Guo, B. Swin transformer: Hierarchical vision transformer using shifted windows. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*; IEEE: Piscataway, NJ, USA, 2021; pp. 10012–10022. [CrossRef]
32. Rai, Y.; Le Callet, P. Visual attention, visual salience, and perceived interest in multimedia applications. In *Academic Press Library in Signal Processing*; Chellappa, R., Theodoridis, S., Eds.; Elsevier: Amsterdam, The Netherlands, 2018; Volume 6, pp. 113–161. [CrossRef]
33. Chen, A.T.; Teng, A.K.; Zhao, J.; Asirot, M.G.; Turner, A.M. The use of visual methods to support communication with older adults with cognitive impairment: A scoping review. *Geriatr. Nurs.* **2022**, *46*, 52–60. [CrossRef]
34. Hassanin, M.; Anwar, S.; Radwan, I.; Khan, F.S.; Mian, A. Visual attention methods in deep learning: An in-depth survey. *Inf. Fusion* **2024**, *108*, 102417. [CrossRef]
35. Machado, E.; Singh, D.; Cruciani, F.; Chen, L.; Hanke, S.; Salvago, F.; Kropf, J.; Holzinger, A. A conceptual framework for adaptive user interfaces for older adults. In *Proceedings of the IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops)*; IEEE: Piscataway, NJ, USA, 2018; pp. 782–787. [CrossRef]
36. de Freitas Pereira, M.L.G.; von Zuben de Arruda Camargo, M.; Bellan, A.F.R.; Tahira, A.C.; Santos, B.D.; Santos, J.D.; Machado-Lima, A.; Nunes, F.L.S.; Forlenza, O.V. Visual search efficiency in mild cognitive impairment and Alzheimer's disease: An eye movement study. *J. Alzheimer's Dis.* **2020**, *75*, 261–275. [CrossRef]
37. Khasnobish, A.; Gavas, R.; Chatterjee, D.; Rajput, V.; Naitam, S. EyeAssist: A communication aid through gaze tracking for patients with neuro-motor disabilities. In *Proceedings of the IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops)*; IEEE: Piscataway, NJ, USA, 2017; pp. 382–387. [CrossRef]
38. Matias, J.; Belletier, C.; Izaute, M.; Lutz, M.; Silvert, L. The role of perceptual and cognitive load on inattentional blindness: A systematic review and three meta-analyses. *Q. J. Exp. Psychol.* **2022**, *75*, 1844–1875. [CrossRef] [PubMed]

39. Slade, K.; Plack, C.J.; Nuttall, H.E. The effects of age-related hearing loss on the brain and cognitive function. *Trends Neurosci.* **2020**, *43*, 810–821. [CrossRef] [PubMed]
40. Dino, M.J.S.; Davidson, P.M.; Dion, K.W.; Szanton, S.L.; Ong, I.L. Nursing and human-computer interaction in healthcare robots for older people: An integrative review. *Int. J. Nurs. Stud. Adv.* **2022**, *4*, 100072. [CrossRef]
41. Tliba, M.; Kerkouri, M.A.; Ghariba, B.; Chetouani, A.; Çöltekin, A.; Shehata, M.S.; Bruno, A. SATSal: A multi-level self-attention based architecture for visual saliency prediction. *IEEE Access* **2022**, *10*, 20701–20713. [CrossRef]
42. Hendry, A.; Johnson, M.H.; Holmboe, K. Early development of visual attention: Change, stability, and longitudinal associations. *Annu. Rev. Dev. Psychol.* **2019**, *1*, 251–275. [CrossRef]
43. Zhang, R.; Chen, C.; Peng, J. Multi-scale graph feature extraction network for panoramic image saliency detection. *Vis. Comput.* **2024**, *40*, 953–970. [CrossRef]
44. Gao, S.; Zhou, C.; Ma, C.; Wang, X.; Yuan, J. AIATrack: Attention in attention for transformer visual tracking. In *Proceedings of the European Conference on Computer Vision (ECCV)*; Springer: Cham, Switzerland, 2022; pp. 146–164. [CrossRef]
45. Wang, B.; Hu, T.; Li, B.; Chen, X.; Zhang, Z. GaTector: A unified framework for gaze object prediction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*; IEEE: Piscataway, NJ, USA, 2022; pp. 19588–19597. [CrossRef]
46. Judd, T.; Ehinger, K.A.; Durand, F.; Torralba, A. Learning to predict where humans look. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*; IEEE: Piscataway, NJ, USA, 2009; pp. 2106–2113. [CrossRef]
47. Le Meur, O.; Coutrot, A. Introducing context-dependent and spatially-variant viewing biases in saccadic models. *Vis. Res.* **2016**, *121*, 72–84. [CrossRef]
48. Ziaei, M.; Salami, A.; Persson, J. Age-related alterations in functional connectivity patterns during working memory encoding of emotional items. *Neuropsychologia* **2017**, *94*, 1–12. [CrossRef]
49. Dowiasch, S.; Marx, S.; Einhäuser, W.; Bremmer, F. Effects of aging on eye movements in the real world. *Front. Hum. Neurosci.* **2015**, *9*, 46. [CrossRef]
50. Meng, Q.; Wang, B.; Cui, D.; Liu, N.; Huang, Y.; Chen, L.; Ma, Y. Age-related changes in local and global visual perception. *J. Vis.* **2019**, *19*, 10. [CrossRef]
51. Madden, D.J.; Siciliano, R.E.; Tallman, C.W.; Monge, Z.A.; Voss, A.; Cohen, J.R. Response-level processing during visual feature search: Effects of frontoparietal activation and adult age. *Atten. Percept. Psychophys.* **2020**, *82*, 330–349. [CrossRef] [PubMed]
52. Jaglarz, A. Color as a key factor in creating sustainable living spaces for seniors. *Sustainability* **2024**, *16*, 10251. [CrossRef]
53. Henderson, J.M.; Hayes, T.R. Meaning-based guidance of attention in scenes as revealed by meaning maps. *Nat. Hum. Behav.* **2017**, *1*, 743–747. [CrossRef] [PubMed]
54. Võ, M.L.-H.; Boettcher, S.E.P.; Draschkow, D. Reading scenes: How scene grammar guides attention and aids perception in real-world environments. *Curr. Opin. Psychol.* **2019**, *29*, 205–210. [CrossRef] [PubMed]
55. Li, J.; Zhang, J.; Tao, D. Deep automatic natural image matting. In Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence (IJCAI), Montreal, QC, Canada, 19–27 August 2021; pp. 800–806. [CrossRef]
56. Zhang, Q.; Qin, Q.; Yang, Y.; Jiao, Q.; Han, J. Feature calibrating and fusing network for RGB-D salient object detection. *IEEE Trans. Circuits Syst. Video Technol.* **2024**, *34*, 1493–1507. [CrossRef]
57. Bhat, S.F.; Alhashim, I.; Wonka, P. AdaBins: Depth estimation using adaptive bins. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*; IEEE: Piscataway, NJ, USA, 2021; pp. 4009–4018. [CrossRef]
58. Tseng, P.-H.; Cameron, I.G.M.; Pari, G.; Reynolds, J.N.; Munoz, D.P.; Itti, L. High-throughput classification of clinical populations from natural viewing eye movements. *J. Neurol.* **2013**, *260*, 275–284. [CrossRef]
59. Fernandes, E.G.; Tatler, B.W.; Slessor, G.; Phillips, L.H. Age differences in gaze following: Older adults follow gaze more than younger adults when free-viewing scenes. *Exp. Aging Res.* **2024**, *50*, 84–101. [CrossRef]
60. Guo, Y.; Pannasch, S.; Helmert, J.R.; Kaszowska, A. Ambient and focal attention during complex problem-solving: Preliminary evidence from real-world eye movement data. *Front. Psychol.* **2024**, *15*, 1217106. [CrossRef]
61. Pedziwiatr, M.A.; Heer, S.; Coutrot, A.; Bex, P.; Mareschal, I. Prior knowledge about events depicted in scenes decreases oculomotor exploration. *Cognition* **2023**, *238*, 105544. [CrossRef]
62. Açık, A.; Sarwary, A.; Schultze-Kraft, R.; Onat, S.; König, P. Developmental changes in natural viewing behavior: Bottom-up and top-down differences between children, young adults and older adults. *Front. Psychol.* **2010**, *1*, 207. [CrossRef]
63. Liang, Z.; Fu, H.; Chi, Z.; Feng, D.D. Refining a region based attention model using eye tracking data. In *Proceedings of the IEEE International Conference on Image Processing (ICIP)*; IEEE: Piscataway, NJ, USA, 2010; pp. 1105–1108. [CrossRef]
64. Sprangers, O.; Schelter, S.; de Rijke, M. Probabilistic gradient boosting machines for large-scale probabilistic regression. In *Proceedings of the ACM SIGKDD Conference on Knowledge Discovery & Data Mining*; Association for Computing Machinery: New York, NY, USA, 2021; pp. 1510–1520. [CrossRef]
65. Wilming, N.; Onat, S.; Ossandon, J.P.; Açık, A.; Kietzmann, T.C.; Kaspar, K.; Gameiro, R.R.; Vormberg, A.; König, P. An extensive dataset of eye movements during viewing of complex images. *Sci. Data* **2017**, *4*, 160126. [CrossRef] [PubMed]

66. Bylinskii, Z.; DeGennaro, E.M.; Rajalingham, R.; Ruda, H.; Zhang, J.; Tsotsos, J.K. Towards the quantitative evaluation of visual attention models. *Vis. Res.* **2015**, *116*, 258–268. [CrossRef] [PubMed]
67. Li, J.; Xia, C.; Song, Y.; Fang, S.; Chen, X. A data-driven metric for comprehensive evaluation of saliency models. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*; IEEE: Piscataway, NJ, USA, 2015; pp. 190–198. [CrossRef]
68. Ke, G.; Meng, Q.; Finley, T.; Wang, T.; Chen, W.; Ma, W.; Ye, Q.; Liu, T.-Y. LightGBM: A highly efficient gradient boosting decision tree. In *Advances in Neural Information Processing Systems 30 (NIPS 2017)*; Curran Associates Inc.: Red Hook, NY, USA, 2017; pp. 3146–3154. Available online: https://proceedings.neurips.cc/paper/2017/hash/6449f44a102fde848669bdd9eb6b76fa-Abstract.html (accessed on 19 March 2026).
69. Lundberg, S.M.; Lee, S.-I. A unified approach to interpreting model predictions. In *Advances in Neural Information Processing Systems 30 (NIPS 2017)*; Curran Associates Inc.: Red Hook, NY, USA, 2017; pp. 4765–4774. Available online: https://papers.nips.cc/paper_files/paper/2017/hash/8a20a8621978632d76c43dfd28b67767-Abstract.html.
70. Erdem, E.; Erdem, A. Visual saliency estimation by nonlinearly integrating features using region covariances. *J. Vis.* **2013**, *13*, 11. [CrossRef]
71. Spreng, R.N.; Turner, G.R. The shifting architecture of cognition and brain function in older adulthood. *Perspect. Psychol. Sci.* **2019**, *14*, 523–542. [CrossRef] [PubMed]
72. Hooker, S.; Erhan, D.; Kindermans, P.-J.; Kim, B. A benchmark for interpretability methods in deep neural networks. In *Advances in Neural Information Processing Systems 32 (NeurIPS 2019)*; Curran Associates Inc.: Red Hook, NY, USA, 2019; pp. 9734–9745. Available online: https://proceedings.neurips.cc/paper_files/paper/2019/hash/fe4b8556000d0f0cae99daa5c5c5a410-Abstract.html (accessed on 19 March 2026).
73. Molnar, C. Interpretable Machine Learning. 2020. Available online: https://christophm.github.io/interpretable-ml-book/ (accessed on 19 March 2026).
74. Chen, T.; Kornblith, S.; Norouzi, M.; Hinton, G. A simple framework for contrastive learning of visual representations. In *Proceedings of the 37th International Conference on Machine Learning (ICML)*; JMLR.org: New York, NY, USA, 2020; Volume 119, pp. 1597–1607. Available online: https://proceedings.mlr.press/v119/chen20j.html (accessed on 19 March 2026).
75. Fisher, A.; Rudin, C.; Dominici, F. All models are wrong, but many are useful: Learning a variable's importance by studying an entire class of prediction models simultaneously. *J. Mach. Learn. Res.* **2019**, *20*, 1–81. Available online: https://jmlr.org/papers/v20/18-760.html (accessed on 19 March 2026).