Multimedia Tools and Applications
https://doi.org/10.1007/s11042-020-09474-5

Computational attention model for children, adults and the elderly

Onkar Krishna[1,2] · Kiyoharu Aizawa[1] · Go Irie[2]

Received: 9 September 2019 / Revised: 23 July 2020 / Accepted: 28 July 2020 /
Published online: 06 September 2020

Abstract
Computational models of saliency estimation have been studied in a wide range of research fields, including visual perception, image processing, computer vision, multimedia, and their intersections. However, most of them seek to simulate scene viewing by adults only, and the impact of observer's age has rarely been considered. In this paper, we quantitatively analyze age-related differences in gaze landing positions during scene viewing. From the results, we draw the following three conclusions: child observers focus more on the foreground in a scene, i.e., locations that are near, while elderly observers tend to explore the background, i.e., locations farther in the scene; adult observers are more explorative than child and elder ones; and adult observers have significantly lower center bias compared to child and elderly observers. Based on these observations, we developed a novel computational model for age-dependent saliency estimation. The prediction accuracy suggests that our model better fits collected eye-gaze data of observers belonging to different age groups than several existing models do.

Keywords Saliency · Eye-tracking · Human visual system · Fixation dispersion · Depth bias

1 Introduction

Owing to the mechanism of selective attention in the human brain, the human visual system can pinpoint the most attractive region from large amounts of visual input. This visual

✉ Onkar Krishna
onkarkris@gmail.com

Kiyoharu Aizawa
aizawa@hal.t.u-tokyo.ac.jp

Go Irie
goirie@ieee.org

[1] Graduate School of Information Science and Technology, The University of Tokyo, Tokyo, Japan

[2] NTT Communication Science Laboratories, NTT Corporation, Kanagawa, Japan

perception phenomenon has inspired efforts to develop computational models for identifying the most attractive regions within images and videos, which is typically called saliency estimation, and these models have been widely used in many image processing applications, including image retrieval [42], object segmentation [16], and target tracking [62].

Significant progress has been made in developing effective saliency estimation methods. In the late '90s, Itti et al. [21] proposed one of the first computational models of perceptual saliency. They based their model on the human bottom-up attention mechanism to detect salient regions in an image by integrating its low-level features, including orientation, intensity, and color information. Ever since this seminal work, many saliency models based on the idea of integrating low-level features have been proposed (e.g., [13, 23, 45]). Subsequently, several studies [1, 40] formulated the saliency estimation problem as object detection/segmentation, and a number of approaches have been proposed along this line (e.g., [7, 51]) More recently, following the success of deep neural networks such as convolutional neural networks (CNNs) in image recognition, CNN-based saliency estimation models have been explored [32, 39]. Because they are capable of end-to-end learning, CNN-based models often outperform those with handcrafted features in terms of estimation accuracy.

Despite these remarkable advances, almost all previous approaches ignore differences between observers in attempting to predict a saliency map common to different observers. Specifically, several studies have suggested that the age of observers has a great impact on their viewing behaviors [10, 18]. As shown in Fig. 1, children and elderly observers have a dramatically different experience of visual saliency than adults while exploring the same scene. Since previous models have been developed and evaluated only with adult gaze data, it is unclear whether they can be applied directly to predictions for other age groups such as children and the elderly.

In this paper, motivated by the above observations, we consider the problem of predicting visual attention for observers in three age groups: children, adults, and the elderly. To this end, we first analyzed age-related differences over the groups in gaze landing positions while the observers viewed various types of images. Based on our observations in the analysis, we then developed a novel age-dependent saliency model. Our model is designed to adapt the scales and weights for the features according to the scene and the observer's age. Experimental results suggest that our model can better fit the eye-gaze data of observers

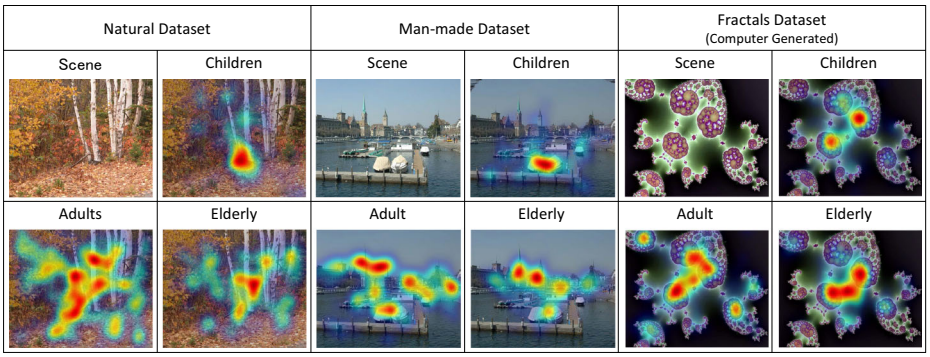

Fig. 1 Sample scenes from nature, man-made, and fractals datasets and their heat map generated from the recorded eye-gaze data of children, adults, and the elderly

belonging to different age groups than several existing models can. This work is a significant extension of the study [27] conducted for only one class of nature images.

2 Related work

We first review the available literature on age-related changes in scene-viewing behavior and then briefly cover existing computational models for saliency estimation.

2.1 Age-related changes in scene viewing

Many studies in psychology and neuroscience have examined the effects of aging on scene-viewing behavior. While earlier studies focused on analyzing basic developmental changes in eye-movement control using artificial stimuli [43, 48], more recent ones have examined the behavior in more realistic settings using natural stimuli [2, 6, 26, 61]. Some of the main findings can be summarized as follows. Ygge et al. [61] found that the fixation system to stably control gaze matures with increasing age between 4 to 15 years. Binda et al. [6] and Kirkorian et al. [26] clarified that the saccade amplitude increases with age, with the peak value reached between 10–15 years. Açık et al. [2] found that low-level (bottom-up) features of a scene such as color, luminance, and contrast often guide viewing behavior during the early stage of life (7–9 years), whereas it is dominated by contextual (top-down) information in the later stage (more than 72 years old). A series of more recent studies [18, 28, 29, 37] examined the age-impact on eye-gaze behavior of children from four age groups: 8–10 years old, 6–8 years old, 4–6 years old, and 2 years old. They analyzed the age-impact on saccade amplitude, fixation duration and the relation between fixation and saccade (ambient and focal) for artificial stimuli. The results in [18] revealed that fixation duration decreases while saccade length increases with age.

It has been suggested that changes in viewing behavior with aging may be influenced not only by low-level characteristics of a scene but also by its high-level characteristics and changes in the visual function itself. For example, young adults tend to focus on areas closer to them first and then scan a wider area [15], whereas older adults tend to see more distant parts of a scene due to presbyopia [53]. The size of the field of view and the control rate of eye movements usually decreases with age [5, 49], and the width of the area over which one can turn one's eyes varies with age. Furthermore, observers in different age groups may have different prior biases, such as central bias [27].

In this work, we experimentally analyzed several high-level attributes of a scene, including depth bias, center bias, and "fixation dispersion" of fixation landing positions, to derive useful observations for model development.

2.2 Saliency models

Tremendous efforts have been made to develop effective computational models of saliency estimation. The majority of the existing models typically rely on the psycho-physical theories of human visual attention, such as Treisman's feature integration theory (FIT) [54] and Wolfe's guided search model [60], and can be classified as either bottom-up [3, 13, 21] or top-down models [19, 23, 30, 32, 45, 47, 56]. Bottom-up models are fully scene-driven and focus on modeling the correlations between low-level scene features such as color, gradient, and contrast and the attractiveness of regions in an image. In addition, they often use

prior information to improve the saliency prediction accuracy. Typical examples are contrast prior [8], background prior [62], and compactness prior [63]. Top-down models, on the other hand, take into account gazing tendencies driven by human intentions and cognitive states of mind, and many of the models in this category [19, 30, 32, 47, 56] combine low-level features with human intentions through statistical modeling and machine learning approaches. For both, the common mechanism is to first compute various feature maps and then fuse the maps into a single final saliency map.

Other than these two types of models, biologically inspired models have been investigated. For example, Wang et al. [58] have proposed a model to simulate human saccadic paths on nature images based on an information maximization criterion. Along the same lines, Le Meur et al. [35] have attempted to generate sequences of saccade motions. Berga et al. [4] have proposed a neurodynamic model of saliency prediction in the primary visual cortex (area V1 or striate cortex).

In more recent years, many deep learning models have also been proposed [9, 19, 32, 33, 56]. Kümmerer et al. proposed transferring an AlexNet [31] pre-trained for object recognition to saliency estimation [33]. DeepFix [32] uses location biased convolution to capture location-dependent factors (e.g., center bias) for improving saliency estimation. ML-Net [9] introduces a multi-scale architecture to fuse different levels of the convolution feature map to predict the final saliency map. Rather than learning deep neural networks end-to-end from scratch, in this paper, we first experimentally uncover the relationship between fixation tendencies and scene characteristics at different ages and build a model based on the findings. Although various approaches have been investigated, most of them ignore age-dependent factors. As we have outlined in the previous subsection, despite several examples of cognitive and neurological studies suggesting that age influences gazing tendencies there are, to the best of our knowledge, only a few age-dependent models for computational saliency estimation [29, 37]. Even in those studies, however, the datasets were limited to observers who viewed very specific types of images collected from children's books or movies such as cartoon images. Moreover, the focus was only on children and adults, and the behavior of the elderly was ignored. Unlike these previous arts, in this study, we developed a simple but effective computational model for predicting saliency maps for age groups covering children, adults, and the elderly.

3 Dataset

The dataset used in this study was collected by Açik et al. [2]. It is publicly available and can be retrieved from the shared link[1] [59].

3.1 Participants, stimuli, and apparatus

Fifty-eight participants took part in the study: 18 children (age range of 7 to 9 years, mean age 7.6), 23 adults (age range of 19 to 27 years, mean age 22.1), and 17 elderly people (age range of 72 to 88 years, mean age 80.6). All participants reported normal or corrected-to-normal vision. All participants or their parents gave written consent to participate in the experiment.

[1] https://doi.org/10.1038/sdata.2016.126

The study used 192 color images belonging to three different categories "nature", "man-made", and "fractals" (64 images in each category). Nature represents scenes having trees, flowers, and bushes, without any artificial objects. The man-made category includes urban scenes such as streets, roads, buildings and construction sites, and fractals are computer-generated shapes taken from different web databases such as Elena's Fractal Gallery. All the image stimuli were randomly ordered to be balanced. All images had a resolution of 1280×960. EyeLink-1000 was used to record the gaze in its remote and hand-free mode.

3.2 Eye movement recording

Eye gaze was recorded while the observers viewed the images displayed for five seconds. The scene was subsequently replaced by a circular patch, and participants had to determine if the patch was part of the previous scene or not. This task was included only to maintain the motivation of participants; thus, the recognition results were not used in the subsequent analyses. The head position was tracked by using a target sticker placed on each observer's head. Observers viewed the stimuli from a distance of 65 cm on a 20-inch LCD monitor display (width: 40 cm). Fixation and saccades were identified via a fixation detection algorithm supplied by EyeLink. For more details about data acquisition and pre-processing, see [59].

3.3 Data representation

We generated human fixation and saliency maps from the collected eye-gaze data. For each age group the human fixation map for an image stimulus was generated by combining the fixation landing positions of all the observers of the group. Following [55], the ground truth saliency maps were generated from fixation maps by convolving a Gaussian over fixation locations by all the observers.

4 Analysis I: necessity of age-dependent saliency model

As reviewed in the previous section, previous studies have found that an observer's age has a significant impact on the fixation landing positions. However, it is still unclear whether we really need age-dependent saliency models or not. One straightforward approach would be to examine the correlation of fixation landing positions between observers belonging to different age groups. If the correlations were not high enough, it would confirm the need to build different models for each age group. We adopt yet another, more direct method in terms of prediction accuracy. Our hypothesis is that, if the fixation landing positions of an observer in a certain age group are "predictable" by using those of another observer in a different age group, there will be no significant advantage of preparing age-dependent saliency models, and vice versa. Therefore, we first analyze the correlations of fixation landing positions and predictability of the different age-groups.

Method Pearson's correlation coefficient (CC) is one of the popular measures to evaluate saliency estimation models [36]. Here, for the correlation analysis, we calculated Pearson's CC between fixation landing positions of different age groups. More specifically, for each given image and collection of fixation landing positions of observers within the same age group, we first generated one ground truth saliency map for each age group. Then, we calculated the CC value between the two ground truth saliency maps from two different age

groups. Finally, we calculated the average CC value for each pair of age groups over all the images in each image category.

For predictability, we first generated a "source saliency map" and "target fixation set" for each image. The source saliency map was generated from a single observer's fixation landing positions by convolving the Gaussian distribution at every fixation point. It was used to predict a target fixation set, which was made by collecting the fixation landing positions of all the (other) observers in an age group (Fig. 2). One target fixation set was generated for each age group so that we could compare the intra- and inter-group predictability. As a metric to measure the predictability of the target fixation points of an image with the corresponding source saliency map, we used area under the curve (AUC), which is frequently used as an accuracy measure of eye-gaze prediction [34]. Let us denote the source saliency map on the ith image of the jth observer by S^i_j and the corresponding target fixation set for the xth age group by T^i_{jx}. We compute the AUC score A^i_{jx} for every pair of S^i_j and T^i_{jx}. The AUC score is computed as done in [34]. Specifically, the source saliency map is thresholded to keep the top $\alpha\%$ of salient parts. Then, the target fixation set is mapped onto the thresholded map, where a fixation point falling into the salient parts is considered as a true positive (TP), and otherwise a false positive (FP). A TP vs. FP curve is drawn by repeating the above two steps for various α from 5% to 100%, and then the AUC score is computed based on the curve. The (observer-level) predictability P_{jx} for a pair of the jth observer as a source and the xth group as a target is specifically formulated as

$$P_{jx} = \frac{1}{N}\sum_{i=1}^{N} A^i_{jx}. \tag{1}$$

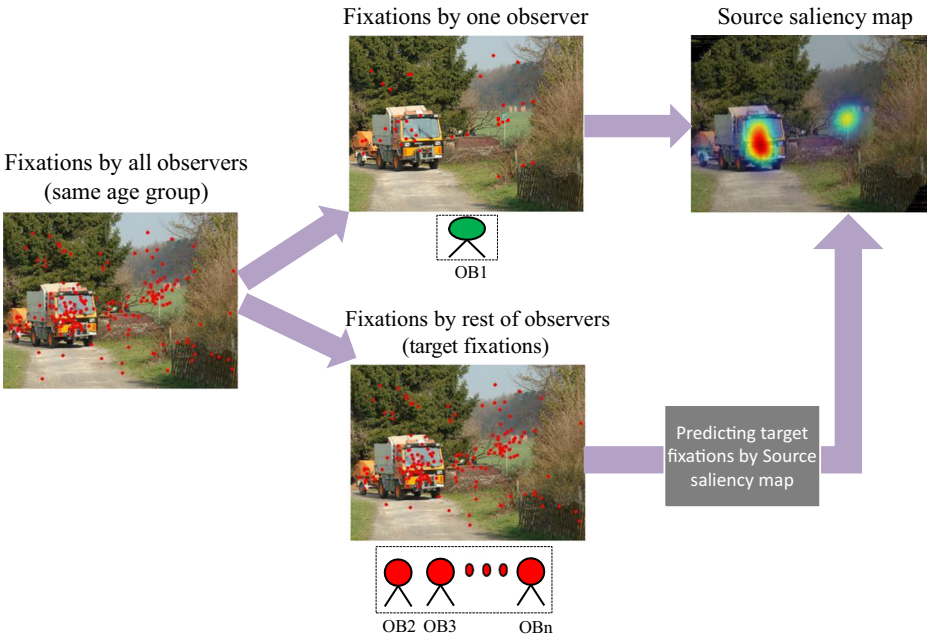

Fig. 2 Procedure to measure inter and intragroup predictability

If the jth observer is in the xth group, then we call it intragroup predictability, and otherwise we call it inter-group predictability. We also compute group-level predictability GP_{xy} between the xth group as a source and the yth group as a target by

$$\mathrm{GP}_{xy} = \frac{1}{|\mathcal{X}|} \sum_{j \in \mathcal{X}} P_{jy} \tag{2}$$

where $\mathcal{X}$ is the set of observers belonging to the xth group. The observer-level and group-level predictability scores are computed for each of the three image categories.

Results Group-level CC values are shown in Table 1, and the observer-level and group-level predictability scores are shown in Fig. 3 and Table 1, respectively.

From these results, we can make the following key observations.

- In Table 1, most CC values for different age groups are less than 0.70, which implies that there is a reasonable correlation between the fixation behaviors of different age groups, but it is not strong.
- In Fig. 3, the children are in better agreement with each other than the adults and elderly observers ($p < 0.001$ by one-way ANOVA).
- The diagonal entries in Table 2 are higher than the other entries in the same row, which suggests that the target fixation set can be best predicted by the source saliency map of an observer from the same age group.
- The influence of age in Table 2 is independent of the image category ($F(2, 192) = 41.39$, $p < 0.001$).

These results suggest that there is a significant difference between fixation patterns among observers in different age groups. Consequently, we can say that a model developed for a particular age group might not work well for other age groups, which suggests that we need an age-dependent saliency model to accurately predict fixations of each age group.

5 Analysis II: Age-dependent features for saliency prediction

The previous section proves the necessity of an age-dependent saliency model. However, what types of features could be used to capture age-dependent differences in fixation landing positions? In order to identify such features, we consider the center bias, depth bias, and fixation dispersion and analyze whether these factors actually differ across age groups.

Table 1 Average Pearson's CC values between ground truth saliency maps of different age groups

Category	Nature			Man-made			Fractals		
	Children	Adults	Elderly	Children	Adults	Elderly	Children	Adults	Elderly
Children	1.0	0.69	0.66	1.0	0.64	0.60	1.0	0.71	0.69
Adults	0.69	1.0	0.63	0.64	1.0	0.63	0.71	1.0	0.66
Elderly	0.66	0.63	1.0	0.60	0.63	1.0	0.69	0.66	1.0

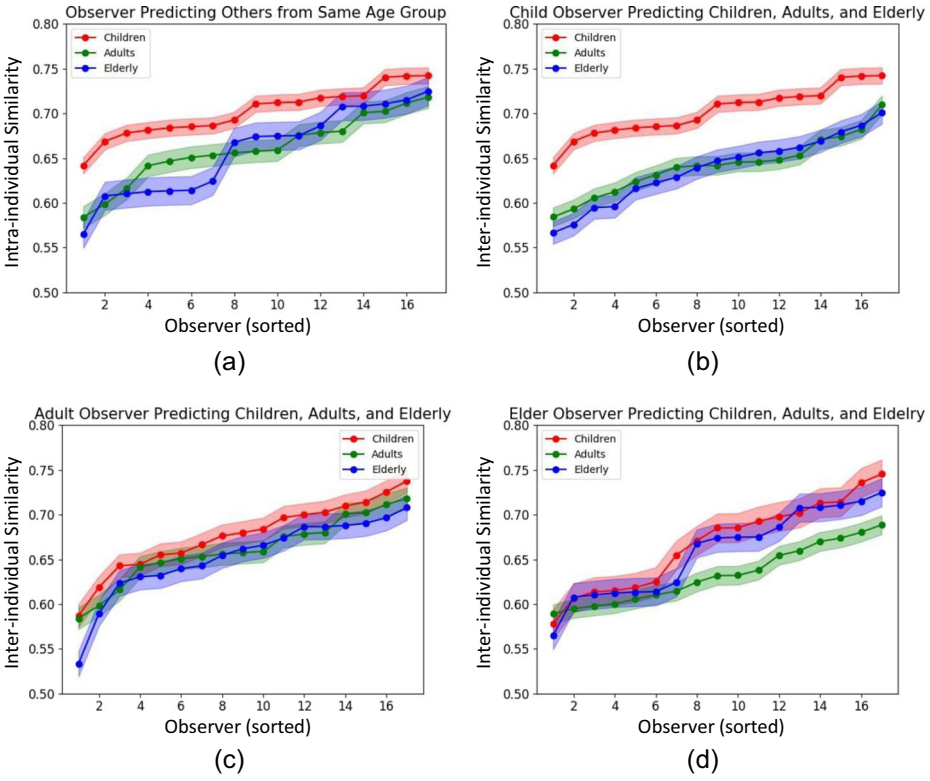

Fig. 3 Inter-individual predictability for man-made dataset: **a** AUC score of a participant predicting others from the same age group. **b**, **c**, **d** AUC score of a participant predicting others from different age group

5.1 Center bias

Center bias has been considered in many eye tracking studies [2, 18, 28], and it improves the prediction accuracy of computational saliency estimation [24, 29, 37].

However, it is unclear whether there is any age-dependent impact on its tendency, which we investigate below.

Table 2 Group-level predictability score for children, adults, and elderly in predicting observers of the same and different age groups for different image categories

Category	Nature			Man-made			Fractals		
Source \ Target	Children	Adults	Elderly	Children	Adults	Elderly	Children	Adults	Elderly
Children	0.8390	0.6937	0.7253	0.8483	0.7093	0.7172	0.8579	0.7257	0.7559
Adults	0.7548	0.7874	0.7193	0.7693	0.8085	0.7310	0.7870	0.8037	0.7507
Elderly	0.7432	0.6860	0.8063	0.7410	0.7094	0.8093	0.7693	0.7173	0.8326

Method Measuring the center bias of an age group is very straight forward. Here, we first generated an average map for each age group by taking an average of the ground truth saliency maps over all the images (for the definition of the ground truth map, see Section 3.3). The resulting maps are shown in Fig. 4a, b, c. We estimated the centriod for each average map, and, finally, the Euclidean distances between the centroid and center of an image were measured to reveal the bias of each age group towards the center of the image. The shorter the Euclidean distance, the higher the center bias.

Another interesting question related to the center bias would be: By using the average map for a certain age group, is it possible to predict the fixation locations of each observer in the same age group? If the answer is yes, at least for some age groups, it may be possible to use the average map as an "age-dependent center map", instead of the standard age-independent center map, i.e., a Gaussian distribution with the center of the image as its mean. Thus, we analyzed how accurately the average map can estimate each observer's fixation locations. To this end, we checked whether the fixation landing positions of one observer in a certain age group could be predicted by the average map of the corresponding age group in terms of the AUC score used in Section 4 to measure predictability.

Results The center bias showed the following tendencies:

– The children group had the highest center bias with the shortest Euclidean distance among age groups (150, 203, and 189 for the nature, man-made, and fractals datasets,

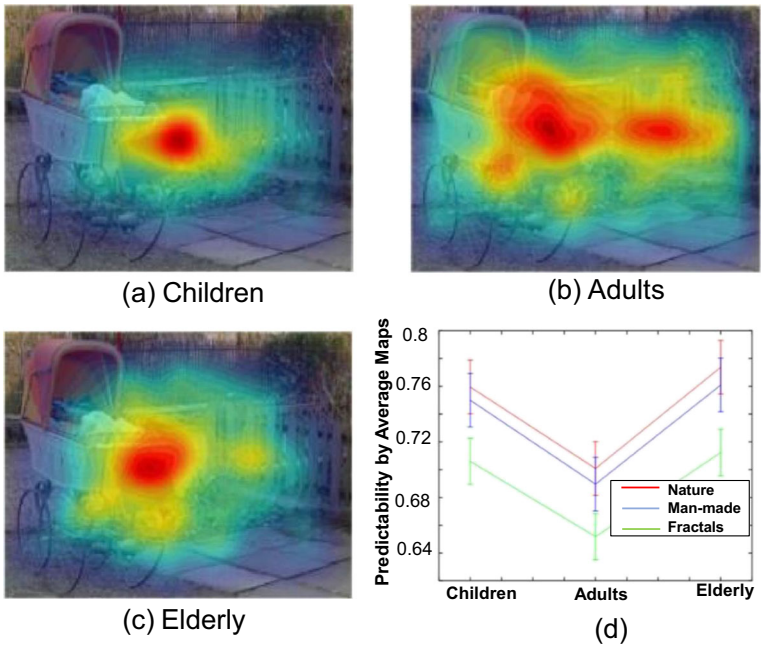

Fig. 4 Center bias: **a**, **b**, **c** Average (saliency) maps for children, adults and elderly are overlaid on a sample image. These maps were generated by averaging the ground truth saliency maps of all the images for each age groups. **d** Plot showing the predictability of fixation landing positions of an observer in a certain age group by using the average map computed for the corresponding age group. The y-axis is the average of the AUC score over all the observers in each age group indicated on the x-axis. Higher AUC means better predictability

respectively), whereas the adults had the lowest center bias with the longest Euclidean distance (210, 267, and 214 for the nature, man-made and fractals datasets, respectively). The result showed that the center bias tendency for elder participants was in between that for children and adults (184, 244, and 202 for the nature, man-made and fractals datasets, respectively).

– The results suggest that the age-impact on center bias tendency is independent of the type of images viewed.
– The results on the predictability are shown in Fig. 4d. They suggest that the average map for a certain age group could accurately predict the fixation landing positions by a single observer in the same age-group ($F(2, 63) = 34.10$, $p < 0.001$), especially for the children and elderly groups. The results are consistent across the three image categories. We can infer from these results that the average map provides an age-dependent alternative to the standard center map.

5.2 Depth bias

The role of the depth information of a visual stimuli for attention deployment has been investigated in many studies [15, 22, 38, 44]. Their results showed that the human visual system tends to focus earlier on objects placed in the foreground than on those in the background [22]. One question may arise here: Does this tendency hold true for all the age groups? In order to answer this question, we quantify the depth bias across different age groups.

Method To measure the depth bias, we labeled the most salient foreground and background regions in each image stimulus and calculated the percentage of fixations that landed on the foreground and background regions across all the images.

First, we obtained the most salient foreground and background regions by the following the steps illustrated in Fig. 5. Specifically, for each image stimulus, we computed an average map for each age group by collecting the fixation points of all the observers within the same group (examples of the obtained average maps are shown in Fig. 6) and generated an age-independent saliency map by linearly integrating the saliency maps of the three groups. We then applied a threshold to the age-independent saliency map to identify the most salient

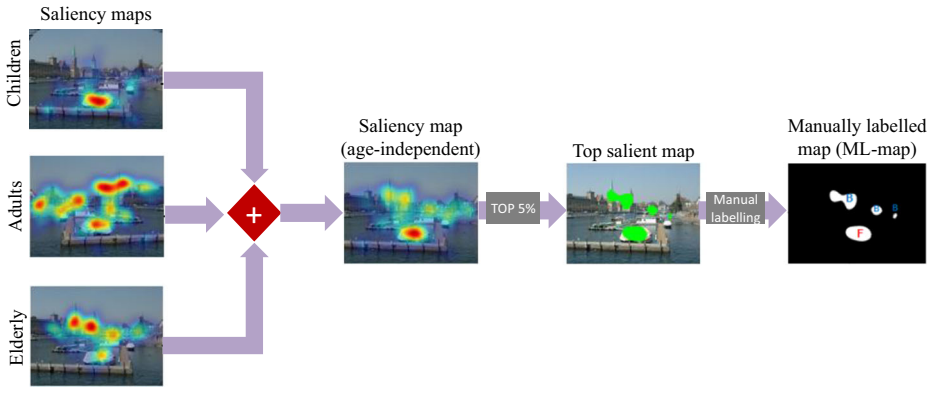

Fig. 5 Depth bias analysis framework. Depth bias was measured by evaluating the percentage of fixations landed on foreground and background as labelled in the ML-map

Fig. 6 Depth bias. Heat maps representing salient locations attended by children, adults, and elderly show clear impact of age on foreground and background viewing tendency

regions. The result can be seen as the "top-salient map" in Fig. 5. We empirically set the threshold to five and ten. Finally, we manually labeled the most salient regions as foreground ("F") or background ("B"). The labeling was performed by five different annotators, and the regions that reached consensus across the annotators were retained.

Based on the foreground and background regions, the percentage of the fixations that landed on either of these locations was computed over all the images.

Results The resulting percentages are shown in Fig. 7, which suggests the following:

- As can be seen in Fig. 7a, the percentage of the foreground regions was significantly higher for the child observers than for the other two groups. A one-way ANOVA suggested the significance of the differences ($F(2, 19) = 2.99$, $p < 0.05$), and the post-hoc analysis confirmed that the tendency in children is significantly higher than in adults and the elderly ($p < 0.05$).
- As can be seen in Fig. 7b, the percentage of the background regions is significantly higher for the elderly observers than for children and adults.

  A one-way ANOVA confirmed the significance of the differences ($F(2, 19) = 3.56$, $p < 0.03$), and the post-hoc analysis revealed that the tendency in the elderly is significantly higher than in adults ($p < 0.03$) and children ($p < 0.03$).

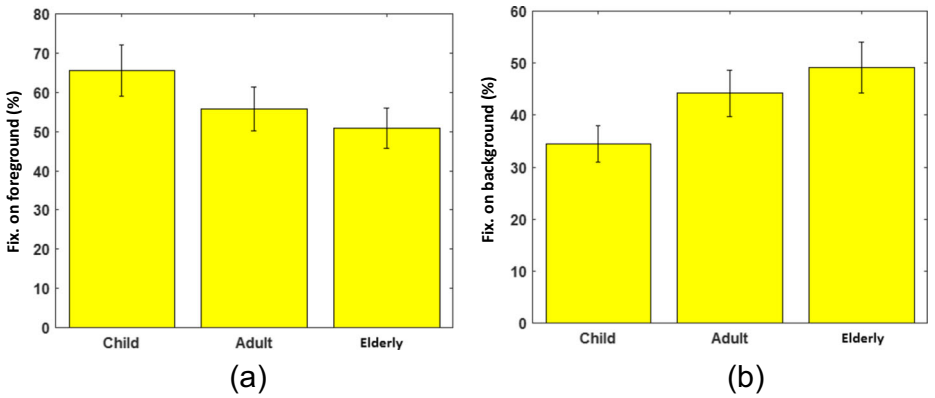

Fig. 7 Percentage of the gaze that landed on **a** the foreground and **b** background. Note that the summation of the foreground and background values is less than 100%, since a small number of fixation points do not fall into the selected foreground and background regions

5.3 Fixation dispersion

Previous studies have shown that useful field of view (UFOV), i.e., the area in a scene from which useful visual information can be extracted, shrinks in older people [5, 12, 50].

Inspired by these results, we measured the spatial scattering of the fixation locations of each age group. We call this measure "fixation dispersion".

Method To measure the scatter of fixation locations, we modeled the fixation dispersion as the first-order entropy of the saliency map, as was done in [28]. Specifically,

$$H(S_x^i) = \sum_l h_x^i(l) \log \frac{|h_x^i(l)|}{h_x^i(l)} \tag{3}$$

where S_x^i is the saliency map of the ith image in the group x, and $h_x^i(l)$ is the histogram entry of the intensity value l in S_x^i.

Examples of the histograms are shown in Fig. 8a. The higher the entropy, the more widely scattered the fixation locations, i.e., the behavior can be regarded as more explorative.

Results The results are shown in Fig. 8b–d, which suggests the following tendencies for the observers:

- There is a significant impact of the observers' age on the tendency of fixation dispersion independent from the image category ($F(2, 191) = 179.45$, $p < 0.001$).

  The post-hoc results showed that the children and elderly observers were less explorative than the adults ($p < 0.001$), whereas the elderly observers were more explorative than the children ($p < 0.001$).
- The age-related differences in fixation dispersion in the three different image classes suggest that this behavior is independent of the type of stimuli being observed.

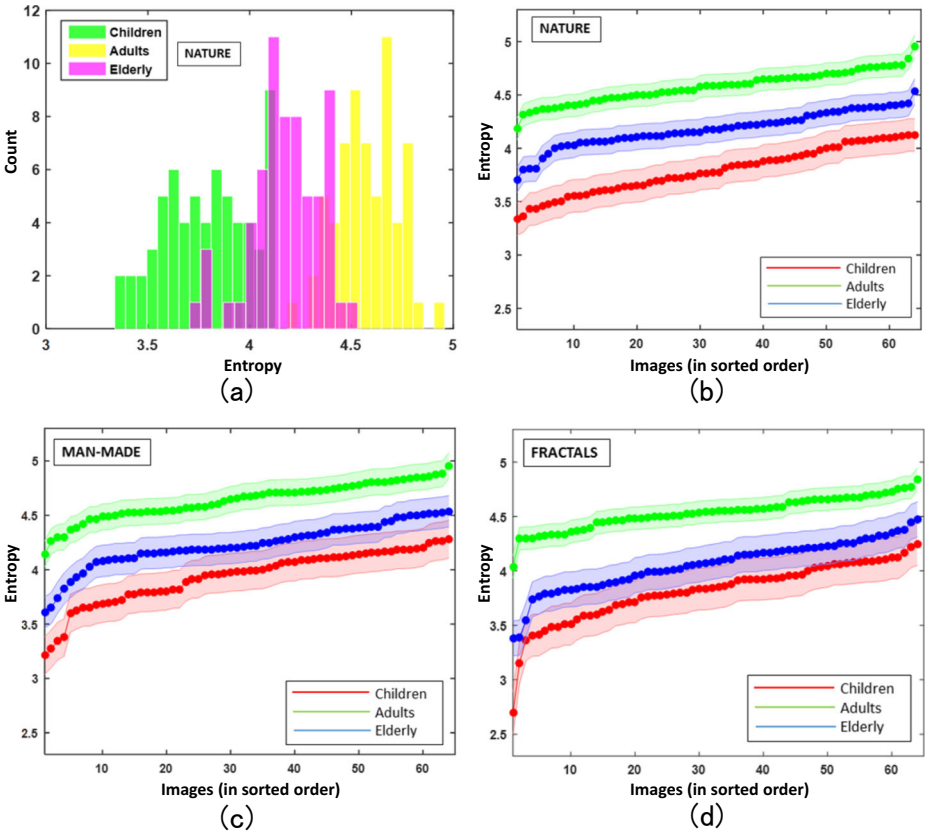

Fig. 8 Age impact on the fixation dispersion tendency. **a** Histograms of saliency intensity values show a shift of fixation dispersion from left to right with increasing age. **b**, **c**, **d** Plots of the entropy of fixation landing positions of all the observers for each image. Higher entropy means more dispersion

5.4 Recommendations for age-dependent saliency model

From the results of the analyses above, we make the following recommendations for designing an age-dependent saliency model.

- First, as observed in our study, the center bias can differ over age groups. One way to reflect these changes in the age-dependent saliency model would be to use the average map computed for each age group.
- Second, we saw from the depth bias analysis that the child observers focus more on the foreground regions, whereas the elderly observers focused more on the background.

  This can be incorporated into the age-dependent saliency model by including depth-dependent feature maps for children and elderly age groups.
- Finally, the analysis of fixation dispersion has shown that children are the least explorative among the three age groups, whereas adults are the most explorative.

  This can be included by choosing a suitable scale of feature maps from finer to coarser in accordance with the level of the fixation dispersion for the target age group.

A study has reported a relationship between image resolution (i.e., scale) and fixation distribution [25], which could be considered as supporting this recommendation from a different viewpoint.

6 Developing age-dependent saliency model

Based on the above recommendations, we developed an age-dependent saliency model for better prediction of saliency maps depending on the target observer's age.

6.1 Modifications for age-dependent saliency estimation

Our model was obtained by modifying several points of the feature-based model proposed in [24] based on the recommendations above.

As we mentioned in the related work section, models based on deep neural networks have attracted attention recently [32, 39, 47, 56], which have mostly been trained with adult observers' data. It would be possible to train such a deep model on a dataset of each age group to obtain a high-quality age-dependent saliency model. One advantage of relying on a feature-based model like [24] would be that we can avoid having to collect a huge amount of qualified training data for children and the elderly, which is challenging due to their physical or health conditions.

Hereafter, we first provide a brief description of the base model [24] and then describe our modifications.

6.2 Base model

The model [24] is designed to predict the saliency map by the weighted sum of multiple feature maps extracted from the input image or based on top-down knowledge.

The specific feature maps are described as follows.

- **Local Energy of Steerable Pyramids**. The pyramid sub band was generated in four orientations and three scales in a way similar to [52]. This is an effective feature for multi-scale analysis of scenes, which has been widely used in many image processing studies, ranging from object classification to saliency detection. In total, 13 bands of scales were used for the steerable pyramids.
- **Torralba's Saliency Map**. The output saliency map from Torralba's saliency model [46] was used as one of the features.
- **Itti-Koch Saliency Map**. The three features of the Itti-Koch saliency map [20] – intensity, orientation, and color contrast – were included.
- **Color Information**. Red, green, and blue and their probabilities were used as features. In addition, the probability of each color computed from the color histograms of the image filtered with a median filter at different scales was used as a feature.
- **Horizon Features**. Horizon detected by the algorithm reported in [46].
- **Objects**. The result of automatic object detection has been used in [24] as a high-level feature. Although "person" and "face" were detected in the original paper [24], we instead detect "person" and "car", which are frequently present in our dataset. As in [24], we use [14, 57] for person and object detection.
- **Center Map**. A feature map which indicates the distance to the center for each pixel is used.

Finally, the feature maps are linearly integrated to compute a single prediction result. The weights of the feature maps are learned by using linear SVM based on the gaze data of adults.

6.3 Age-dependent model

We modify the feature maps based on our recommendations.

In short, we introduce a new depth map and center map and adapt all the feature weights and the scale factors to the tendencies of children, adults, and the elderly.

- **Age-dependent Center Map**. Instead of using a common center map for all the age groups as was done in the base model [24], we prepare an age-dependent center map for each of the three age groups.

  The age-dependent center map is computed as described in Section 5.1.
- **Depth Map**. Unlike the base model, we newly add a depth feature map which is computed as follows. First, we estimate a depth map from the input RGB image using the algorithm proposed by Liu et al. [41]. An example of the estimated depth maps can be seen in Fig. 9. Our fixation dispersion analysis in Section 5.3 suggested that the scales of the feature maps can be tuned for the target age group. Hence, we apply the steerable pyramid filters to the depth map to obtain the depth feature map at different scales. Noteworthy is that this is the first work that brings a depth feature into the context of age-dependent saliency analysis and prediction.
- **Fixation-dispersion-oriented Scale Selection**. Age-related differences in the fixation dispersion tendency are incorporated by selecting different scales of steerable pyramids for different age groups. More specifically, we empirically split the 13 scale bands of the steerable pyramids into three groups: band 1 to 4 as Group 1 (finest), band 5 to 8 as Group 2, and band 9 to 13 as Group 3 (coarsest). A preliminary experiment was conducted to examine all possible combinations of the three groups. The optimal subsets for each age group are Groups 1, 2, and 3 for adults, Groups 2 and 3 for the elderly, and Group 3 for children. They are consistent with the recommendation made on the basis of the analysis results, which states that only a few scales are required to estimate the gaze landing positions of the less explorative children and elderly, and conversely, for adults, all the scales should be used.

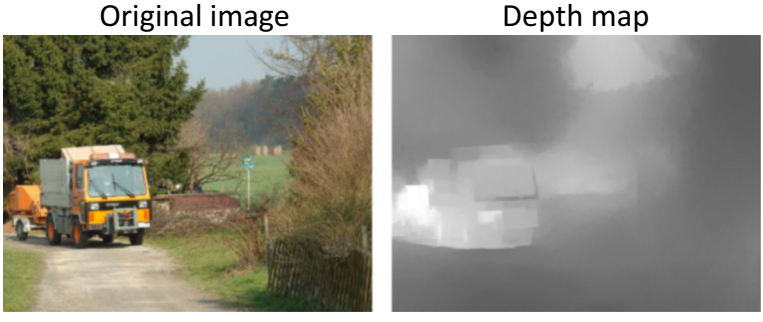

Fig. 9 Input image and its disparity map generated by the algorithm proposed by Liu et al. [41] (black and white areas represent the nearest and farthest locations in the scene, respectively)

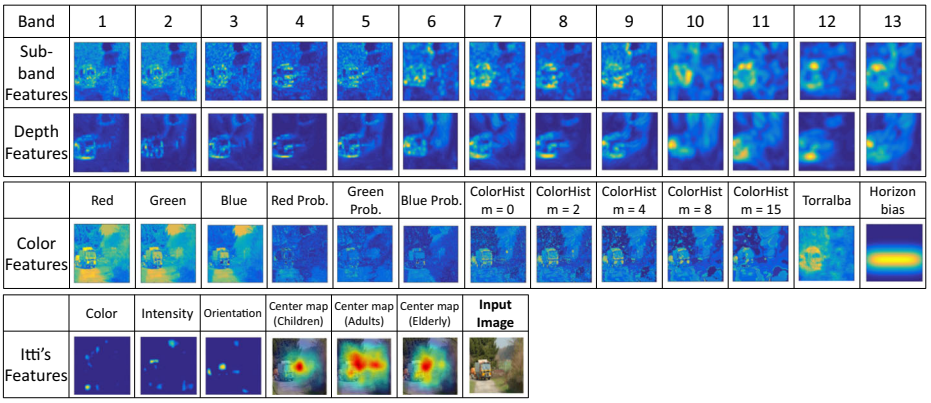

Fig. 10 Feature maps used in our model (the center-map feature is shown by overlaying it on top of an input image)

Figure 10 shows an example of feature maps extracted from an input image at different bands.

Finally, we train the weights of a linear SVM by using the dataset specific to each of the three age groups.

6.4 Evaluation

We conducted an experiment to analyze the effectiveness of our age-dependent saliency model.

Evaluation protocol We basically followed the protocol used in [24]. Specifically, for each image the 10 most salient pixels were randomly chosen from the top 10% salient locations of the ground truth saliency map as positive samples, and the 10 least salient pixels from the bottom 70% salient locations were chosen as negative samples. Then, a support vector machine (SVM) with linear kernels was used to train a classifier on the positive and negative samples over the set of precomputed image features. In our experiment setup, we split the dataset with 192 images into 120 training (40 images per image category) and 72 test sets (24 images per image category). After training the model on the training set, the evaluation was performed by using the test set. Specifically, given an input RGB image, the predicted saliency map was generated by the model. As in [24], we used the AUC scores to evaluate how well the saliency map estimated by our model can predict the ground truth map. We compared our model with several existing methods [11, 17, 21, 24]. For a fair comparison, we created two variants of the base model [24]. In the first variant (Judd-S), we trained the base model for each age group separately on our training dataset. The second variant (Judd-A) was created by training the base model on the combined fixations of all age groups. Finally, the performance of these two variants was evaluated on our test set for each age group. We also created two variants of our proposed model to understand the level of accuracy achievable by simply training the base models on age-specific data and combined data. Performance was evaluated using the AUC metric as was done in [24].

Results The results are shown in Table 3. The proposed age-dependent model outperformed the existing saliency models for most of the age groups independent of the image categories.

Table 3 Comparison of our age-dependent model with existing saliency prediction models. The values in the table are the AUC scores between the predicted saliency maps and ground truth saliency maps. Taking values from 0 to 1, higher is better. The best result of each age-groups is reported in bold emphasis

Category	Age	Itti [21]	GBVS [17]	Patch [11]	Judd-S [24]	Judd-A [24]	Ours-A	Ours-S
Man-made	Children	0.72	0.73	0.66	0.73	0.73	**0.75**	0.74
	Adults	0.67	0.67	0.61	0.65	0.66	0.67	**0.68**
	Elderly	0.69	0.71	0.67	**0.72**	0.70	**0.72**	**0.72**
Nature	Children	0.62	0.73	0.72	0.76	0.76	0.76	**0.77**
	Adults	0.60	0.67	0.64	0.67	0.67	0.69	**0.70**
	Elderly	0.61	0.73	0.72	**0.75**	**0.75**	**0.75**	**0.75**
Fractals	Children	0.70	**0.76**	0.71	0.75	**0.76**	**0.76**	**0.76**
	Adults	0.66	**0.71**	0.65	0.69	0.69	0.69	**0.71**
	Elderly	0.68	0.75	0.72	0.75	0.75	**0.76**	**0.76**

The best result of each age-groups is reported in bold emphasis

Compared to [11, 17, 21], the performance gain with our method is very high, especially for the images in the man-made and nature categories. Our age-dependent saliency model also outperformed the base model [24] not only for the children and elderly groups but also for adults, which suggests that our modifications, especially including the depth map in the leaning-based model, works well for adult observes as well.

The quantitative results in Table 3 show that our model has a clear edge over the baselines for all age groups; however, the significance of this gain is unclear. To better understand the extent that this gain in performance is reflected in the predicted saliency map of each age group, we show some sample images with their saliency maps (Fig. 11) computed by our model and baselines. The results show the excellent ability of our age-dependent saliency model to predict saliency maps for all the age groups (children, adults, and elderly). For instance, as shown in the red circle over the second image (second row) in Fig. 11, only our method can accurately predict the most salient regions across age groups, whereas the baseline methods severely fails to predict them for not only the children and elderly (which was expected) but for the adult observers as well.

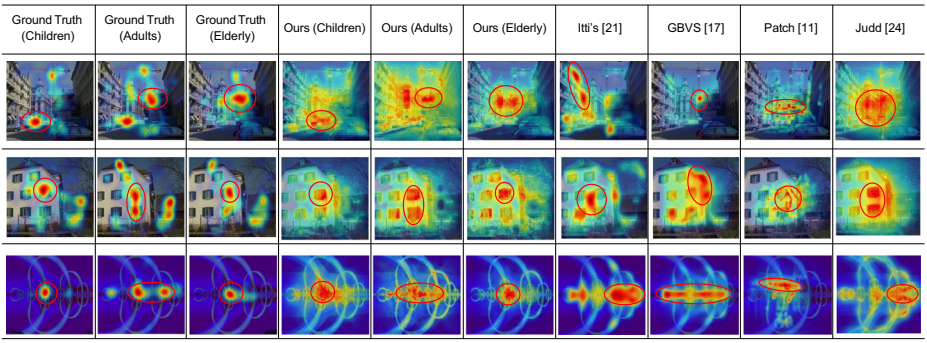

Fig. 11 Comparison of the saliency map generated from existing methods and our proposed age-dependent approach. The red circle overlaid onto the images shows the most salient image regions across observers in different age-groups

7 Conclusion

In this paper, we analyzed age-related differences in scene viewing behavior for children, adult, and elderly observers while they viewed scenes belonging to different categories. The analysis mainly focused on measuring age-related changes in depth bias, center bias, fixation dispersion, and inter-individual similarity. The results showed a significant impact of age on fixation landing positions independent of the category of the scene viewed. Further, they helped in feature scale selection, depth map insertions, and age-specific learning in our proposed age-dependent learning-based saliency model.

Our proposed model outperforms the existing saliency models in prediction accuracy for all the age groups. Predicting which part of a scene observers in different age groups would pay attention to will be useful in assisting people of different age groups in daily activities such as online shopping and driving by the elderly. In addition, the recommendations made from the analysis in the study will be useful in editing movies, books, and advertisement contents for the target-specific age groups.

Acknowledgements We thank Dr. Alpher Açık for providing the gaze data used in this study. This work is supported by JST CREST, JPMJCR1686.

References

1. Achanta R, Estrada F, Wils P, Süsstrunk S (2008) Salient region detection and segmentation. In: Proceedings of 2008 international conference on computer vision systems, pp 66–75
2. Açık A, Sarwary A, Schultze-Kraft R, Onat S, König P (2010) Developmental changes in natural viewing behavior: Bottom-up and top-down differences between children, young adults and older adults. Front Psychol 1:207
3. Bak C, Kocak A, Erdem E, Erdem A (2017) Spatio-temporal saliency networks for dynamic saliency prediction. IEEE Trans Multimed 20(7):1688–1698
4. Berga D, Otazu X (2018) A neurodynamic model of saliency prediction in V1. ArXiv Preprint arXiv:http://arxiv.org/abs/1811.06308
5. Beurskens R, Bock O (2012) Age-related decline of peripheral visual processing: the role of eye movements. Exp Brain Res 217(1):117–124
6. Binda P, Morrone MC (2018) Vision during saccadic eye movements. Annu Rev Vis Sci 4:193–213
7. Chang K-Y, Liu T-L, Chen H-T, Lai S-H (2011) Fusing generic objectness and visual saliency for salient object detection. In: Proceedings of 2011 international conference on computer vision, pp 914–921
8. Cheng M-M, Mitra NJ, Huang X, Torr PHS, Hu S-M (2014) Global contrast based salient region detection. IEEE Trans Pattern Anal Mach Intell 37(3):569–582
9. Cornia M, Baraldi L, Serra G, Cucchiara R (2016) A deep multi-level network for saliency prediction. In: Proceedings of 2016 IEEE international conference on pattern recognition, pp 3488–3493
10. Dowiasch S, Marx S, Einhäuser W, Bremmer F (2015) Effects of aging on eye movements in the real world. Front Hum Neurosci 9:46
11. Duan L, Wu C, Miao J, Qing L, Fu Y (2011) Visual saliency detection by spatially weighted dissimilarity. In: Proceedings of 2011 IEEE conference on computer vision and pattern recognition, pp 473–480
12. Edwards JD, Ross LA, Wadley VG, Clay OJ, Crowe M, Roenker DL, Ball KK (2006) The useful field of view test: normative data for older adults. Arch Clin Neuropsychol 21(4):275–286
13. Erdem E, Erdem A (2013) Visual saliency estimation by nonlinearly integrating features using region covariances. J Vis 13(4):11–11
14. Felzenszwalb P, McAllester D, Ramanan D (2008) A discriminatively trained, multiscale, deformable part model. In: Proceedings of 2008 IEEE Computer Society conference on computer vision and pattern recognition, pp 1–8

15. Gautier J, Le Meur O (2012) A time-dependent saliency model combining center and depth biases for 2D and 3D viewing conditions. Cogn Comput 4(2):141–156
16. Han J, Quan R, Zhang D, Nie F (2017) Robust object co-segmentation using background prior. IEEE Trans Image Process 27(4):1639–1651
17. Harel J, Koch C, Perona P (2007) Graph-based visual saliency. In: Proceedings of advances in neural information processing systems 19, pp 545–552
18. Helo A, Pannasch S, Sirri L, Rämä P (2014) The maturation of eye movement behavior: scene viewing characteristics in children and adults. Vis Res 103:83–91
19. Huang X, Shen C, Boix X, Zhao Q (2015) SALICON: reducing the semantic gap in saliency prediction by adapting deep neural networks. In: Proceedings of 2015 IEEE international conference on computer vision, pp 262–270
20. Itti L, Koch C (2000) A saliency-based search mechanism for overt and covert shifts of visual attention. Vis Res 40(10-12):1489–1506
21. Itti L, Koch C, Niebur E (1998) A model of saliency-based visual attention for rapid scene analysis. IEEE Trans Pattern Anal Mach Intell 20(11):1254–1259
22. Jansen L, Onat S, König P (2009) Influence of disparity on fixation and saccades in free viewing of natural scenes. J Vis 9(1):29
23. Jiang H, Wang J, Yuan Z, Wu Y, Zheng N, Li S (2013) Salient object detection: a discriminative regional feature integration approach. In: Proceedings of 2013 IEEE conference on computer vision and pattern recognition, pp 2083–2090
24. Judd T, Ehinger K, Durand F, Torralba A (2009) Learning to predict where humans look. In: Proceedings of 2009 IEEE international conference on computer vision, pp 2106–2113
25. Judd T, Durand F, Torralba A (2011) Fixations on low-resolution images. J Vis 11(4):14
26. Kirkorian HL, Anderson DR (2017) Anticipatory eye movements while watching continuous action across shots in video sequences: a developmental study. Child Dev 88(4):1284–1301
27. Krishna O, Aizawa K (2017) Age-adapted saliency model with depth bias. In: Proceedings of the 2017 ACM symposium on applied perception, pp 1–8
28. Krishna O, Yamasaki T, Helo A, Pia R, Aizawa K (2017) Developmental changes in ambient and focal visual processing strategies. Electron Imaging 14:224–229
29. Krishna O, Helo A, Rämä P, Aizawa K (2018) Gaze distribution analysis and saliency prediction across age groups, vol 13, p e0193149
30. Krishna O, Aizawa K, Reimerth S (2018) Signboard saliency detection in street videos. In: Proceedings of 2018 IEEE international conference on acoustics, speech and signal processing, pp 1917–1921
31. Krizhevsky A, Sutskever I, Hinton GE (2012) ImageNet classification with deep convolutional neural networks. In: Proceedings of advances in neural information processing systems 25, pp 1097–1105
32. Kruthiventi SSS, Ayush K, Babu RV (2017) DeepFix: a fully convolutional neural network for predicting human eye fixations. IEEE Trans Image Process 26(9):4446–4456
33. Kümmerer M, Theis L, Bethge M (2014) Deep gaze I: boosting saliency prediction with feature maps trained on ImageNet. ArXiv Preprint arXiv:http://arxiv.org/abs/1411.1045
34. Le Meur O, Baccino T (2013) Methods for comparing scanpaths and saliency maps: strengths and weaknesses. Behav Res Methods 45(1):251–266
35. Le Meur O, Liu Z (2015) Saccadic model of eye movements for free-viewing condition. Vis Res 116:152–164
36. Le Meur O, Le Callet P, Barba D (2007) Predicting visual fixations on video based on low-level visual features. Vis Res 47(19):2483–2498
37. Le Meur O, Coutrot A, Liu Z, Rämä P, Le Roch A, Helo A (2017) Visual attention saccadic models learn to emulate gaze patterns from childhood to adulthood. IEEE Trans Image Process 26(10):4777–4789
38. Leifman G, Rudoy D, Swedish T, Bayro-Corrochano E, Raskar R (2017) Learning gaze transitions from depth to improve video saliency estimation. In: Proceedings of 2017 IEEE international conference on computer vision, pp 1698–1707
39. Li X, Zhao L, Wei L, Yang M-H, Wu F, Zhuang Y, Ling H, Wang J (2016) DeepSaliency: multi-task deep neural network model for salient object detection. IEEE Trans Image Process 25(8):3919–3930
40. Liu T, Yuan Z, Sun J, Wang J, Zheng N, Tang X, Shum H-Y (2010) Learning to detect a salient object. IEEE Trans Pattern Anal Mach Intell 33(2):353–367
41. Liu F, Shen C, Lin G (2015) Deep convolutional neural fields for depth estimation from a single image. In: Proceedings of 2015 IEEE conference on computer vision and pattern recognition, pp 5162–5170
42. Lu X, Chen Y, Li X (2017) Hierarchical recurrent neural hashing for image retrieval with hierarchical convolutional features. IEEE Trans Image Process 27(1):106–120
43. Luna B, Garver KE, Urban TA, Lazar NA, Sweeney JA (2004) Maturation of cognitive processes from late childhood to adulthood. Child Dev 75(5):1357–1372

44. Ma C-Y, Hang H-M (2015) Learning-based saliency model with depth information. J Vis 15(6):19
45. Navalpakkam V, Itti L (2006) An integrated model of top-down and bottom-up attention for optimizing detection speed. In: Proceedings of 2006 IEEE Computer society conference on computer vision and pattern recognition, pp 2049–2056
46. Oliva A, Torralba A (2001) Modeling the shape of the scene: a holistic representation of the spatial envelope. Int J Comput Vis 42(3):145–175
47. Pan J, Sayrol E, Nieto XG, McGuinness K, O'Connor NE (2016) Shallow and deep convolutional networks for saliency prediction. In: Proceedings of 2016 IEEE conference on computer vision and pattern recognition, pp 598–606
48. Polat U, Schor C, Tong J-L, Zomet A, Lev M, Yehezkel O, Sterkin A, Levi DM (2012) Training the brain to overcome the effect of aging on the human eye. Sci Rep 2(1):1–6
49. Rogé J, Pébayle T, Campagne A, Muzet A (2005) Useful visual field reduction as a function of age and risk of accident in simulated car driving. Investig Ophthalmol Vis Sci 46(5):1774–1779
50. Sekuler AB, Bennett PJ, Mamelak M (2000) Effects of aging on the useful field of view. Exp Aging Res 26(2):103–120
51. Shen X, Wu Y (2012) A unified approach to salient object detection via low rank matrix recovery. In: Proceedings of 2012 IEEE conference on computer vision and pattern recognition, pp 853–860
52. Simoncelli EP, Freeman WT (1995) The steerable pyramid: a flexible architecture for multi-scale derivative computation. In: Proceedings of 1995 IEEE international conference on image processing, pp 444–447
53. Strenk SA, Strenk LM, Koretz JF (2005) The mechanism of presbyopia. Prog Retin Eye Res 24(3):379–393
54. Treisman AM, Gelade G (1980) A feature-integration theory of attention. Cogn Psychol 12(1):97–136
55. Velichkovsky B, Pomplun M, Rieser J (1996) Attention and communication: eye-movement-based research paradigms. Adv Psychol 116:125–154
56. Vig E, Dorr M, Cox D (2014) Large-scale optimization of hierarchical features for saliency prediction in natural images. In: Proceedings of 2014 IEEE conference on computer vision and pattern recognition, pp 2798–2805
57. Viola P, Jones M et al (2001) Robust real-time object detection. Int J Comput Vis 4(34-47):4
58. Wang W, Chen C, Wang Y, Jiang T, Fang F, Yao Y (2011) Simulating human saccadic scanpaths on natural images. In: Proceedings of 2011 IEEE conference on computer vision and pattern recognition, pp 441–448
59. Wilming N, Onat S, Ossandón JP, Açık A, Kietzmann TC, Kaspar K, Gameiro RR, Vormberg A, König P (2017) An extensive dataset of eye movements during viewing of complex images. Sci Data 4(1):1–11
60. Wolfe JM (1994) Guided search 2.0 a revised model of visual search. Psychon Bull Rev 1(2):202–238
61. Ygge J, Aring E, Han Y, Bolzani R, Hellström A (2005) Fixation stability in normal children. Ann New York Acad Sci 1039(1):480–483
62. Zhang L, Suganthan PN (2016) Visual tracking with convolutional random vector functional link network. IEEE Trans Cybern 47(10):3243–3253
63. Zhou L, Yang Z, Yuan Q, Zhou Z, Hu D (2015) Salient region detection via integrating diffusion-based compactness and local contrast. IEEE Trans Image Process 24(11):3308–3320