Webpage Saliency

Chengyao Shen[1,2] and Qi Zhao[2*]

[1]Graduate School for Integrated Science and Engineering,
[2]Department of Electrical and Computer Engineering,
National University of Singapore

Abstract. Webpage is becoming a more and more important visual input to us. While there are few studies on saliency in webpage, we in this work make a focused study on how humans deploy their attention when viewing webpages and for the first time propose a computational model that is designed to predict webpage saliency. A dataset is built with 149 webpages and eye tracking data from 11 subjects who free-view the webpages. Inspired by the viewing patterns on webpages, multi-scale feature maps that contain object blob representation and text representation are integrated with explicit face maps and positional bias. We propose to use multiple kernel learning (MKL) to achieve a robust integration of various feature maps. Experimental results show that the proposed model outperforms its counterparts in predicting webpage saliency.

Keywords: Web Viewing, Visual Attention, Multiple Kernel Learning

1 Introduction

With the wide spread of Internet in recent decades, webpages have become a more and more important source of information for an increasing population in the world. According to the *Internet World States*, the number of internet users has reached 2.4 billion in 2012. A recent survey conducted on US-based web users in May 2013 also showed that an average user spends 23 hours a week online [25]. This trend of increasing time spent on web has greatly reshaped people's life style and companies' marketing strategy. Thus, the study of how users' attention is deployed and directed on a webpage is of great research and commercial value.

The deployment of human attention is usually driven by two factors: a bottom-up factor that is memory-free and biased by the conspicuity of a stimuli, and a top-down factor which is memory-dependent and with variable selective criteria [16]. The saliency of stimuli is the distinct perceptual quality by which the stimuli stand out relative to their neighbors. It is typically computed based on low-level image statistics, namely luminance contrast, color contrast, edge density, and orientation (also motion and flickr in video) [16, 4]. Recent studies show that text, face, person, animals and other specific objects also contribute much to the saliency of a stimulus [18, 5, 9, 28, 29]

* Corresponding Author. Tel. (65) 6516-6658. E-mail: eleqiz@nus.edu.sg

Compared with natural images, webpage has its own characteristics that make a direct application of existing saliency models ineffective. For example, webpages are usually rich in visual media, such as text, pictures, logos and animations [21]. From the classical low-level feature based saliency point of view, a webpage is thus full of salient stimuli and competition arises everywhere, which makes an accurate prediction of eye fixations difficult. Besides, studies show that people's web-viewing behavior is different from that on natural images and reveals several distinct patterns such as F-bias to scan top-left region at the start [19, 4] and banner blindness to avoid banner-like advertisement [6, 12, 14]. These differentiating factors suggest new ways for webpage saliency.

1.1 Visual Attention Models on Webpages

Up to now, there is no report on webpage saliency models in the literature. There are, however, several conceptual models and computational models based on non-image information that investigate into the user viewing behaviors on different webpages.

For conceptual models, Faraday's visual scanning model [10] represents the first framework that gave a systematic evaluation of visual attention on webpages. This model identified six "salient visual elements"(SAE) in a hierarchy (motion, size, image, color, text-style, and position) that direct our attention in webpages and provided a description of how these elements are scanned by a user. A later research by Grier *et al.* [12] showed that Faraday's model is oversimplified for complex web-viewing behaviors (e.g., the salience order of SAE selected by the model might be inaccurate). Based on Faradays model, Grier *et al.* described three heuristics ("top left corner of the main content area is dominant, "overly salient items do not contain information, "information of similar type will be grouped together) from their observation and they further proposed a three stage EHS (Expected Location, Heuristic Search, Systematic Search) theory that explains the viewing behavior on webpages. These conceptual models give us a good foundation on developing a computational algorithm to predict webpage saliency.

For the computational models based on non-image information, the model from Buscher *et al.* [4] that utilized HTML-induced document object model (DOM) is among the most prominent. In [4], the authors first collected data when users were engaged in information foraging and page recognition tasks on 361 webpages from 4 categories (cars, diabetes, kite surfing, wind energy). They then performed a linear regression on features extracted from DOM and generated a model for predicting visual attention on webpages using decision trees. Their linear regression showed that size of the DOM is the most decisive factor and their decision tree get a precision of 75% and a recall of 53% in predicting the eye fixations on webpages. From their data, they also observed that the first few fixations (i.e., during the first second of each page view) are consistent in both tasks. Other models in this category either focus on a specific type of webpages [7] that does not generalize well, or based themselves on text semantics [22, 23] thus quite different from the goal in this work.

The only work we found sharing a similar goal with ours is from Still and Masciocchi [21]. The referred work, however, simply applied the classic Itti-Koch model [16] to predict the web-viewing entry points while we investigate features and mechanisms underlying webpage saliency and propose a dedicated model for saliency prediction in this context.

1.2 Motivations and Contributions

In this study, we aim to develop a saliency model that is purely based on visual features to predict the eye fixation deployment on webpages. To achieve this, we first collect eye tracking data from 11 subjects on 149 webpages and analyzed the data to get ground truth fixation maps for webpage saliency. We then extract various visual features on webpages with multi-scale filters and face detectors. After feature extraction, we integrate all these feature maps incorporating positional bias with multiple kernel learning (MKL) and use this integrated map to predict eye fixation on webpages. Comparative results demonstrate that our model outperforms existing saliency models. Besides the scientific question of how humans deploy attention when viewing webpage, computational models that mimic human behavior will have general applicability to a number of tasks like guiding webpage design, suggesting ad placement, and so on.

The main contributions of our research include:

1. We collect an eye fixation dataset from 11 subjects on 149 webpages which is the first dataset on webpage saliency according to our knowledge.
2. We propose a new computational model for webpage saliency by integrating multi-scale feature maps, face map and positional bias, in a MKL framework. The model is the first one for webpage saliency that is purely based on visual features.

2 Webpage Saliency Dataset

Since there is no publicly available eye tracking dataset on webpages, we create one dataset and plan to make it public to facilitate further research on webpage saliency. In the following, we describe the stimuli, data collection procedure, and data analysis for the dataset.

2.1 Stimuli

149 screenshots of webpages rendered in Chrome browser in full screen mode were collected from various sources on the Internet in the resolution of 1360 by 768 pixels. These webpages were categorized as pictorial, text and mixed according to the different composition of text and pictures and each category contains around 50 images. Examples of webpage in each category are shown in Figure 1 and the following criteria were used during the collection of webpage image samples:

Fig. 1. Examples of webpages in our dataset

- **Pictorial**: Webpages occupied by one dominant picture or several large thumbnail pictures and usually with less text. Examples in this category include photo sharing websites and company websites that put their products in the homepages.
- **Text**: Webpages containing informative text with high density Examples include wikipedia, news websites, and academic journal websites.
- **Mixed**: Webpages with a mix of thumbnail pictures and text in middle density. Examples are online shopping websites and social network sites.

The collected samples consisted of webpages from various domains. This was done to suppress the subjects's prior familiarity of the layout of the webpage as well as to prevent the subjects from developing familiarity during the experiment, so as to reduce personal bias or top-down factors.

2.2 Eye Tracking Data Collection

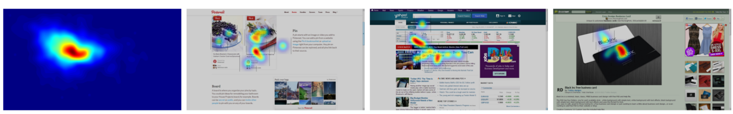
Subjects 11 students (4 males and 7 females) in the age range of 21 to 25 participated in data collection. All participants had normal vision or corrective visual apparatus during the experiment and all of them were experienced Internet users.

Apparatus and Eye Tracking Subjects were seated in a dark room with their head positioned on a chin and forehead rest, 60 cm from the computer screen. The resolution of the screen was 1360 × 768 pixels. Stimuli were placed across the entire screen and were presented using MATLAB (MathWorks, Natick, Massachusetts, USA) with the Psychtoolbox 3 [2] . Eye movement data were monocularly recorded using a noninvasive Eyelink 1000 system with a sampling rate of 1000 Hz. Calibration was done using the 9-point grid method.

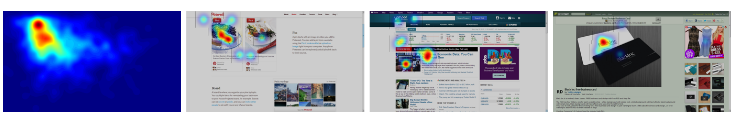
Procedure For each trial, an image was presented in random order for 5 seconds. Subjects were instructed to free-view the webpages and were informed that they had 5 seconds for each webpage. Each trial will follow by a drift correction where the subject would have to fixate at the center and initiate the next trial via a keyboard press.

2.3 Dataset Analysis

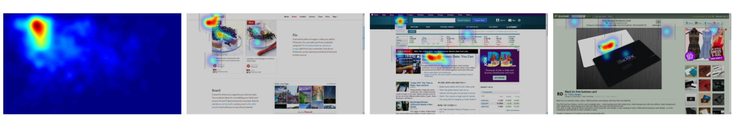
We analyze the eye fixation data collected from 11 subjects by visualizing their fixation heat maps. The fixation heat map was generated by convolving a 2D Gaussian filter on fixation points gathered from all the images in the dataset or in one particular category. In this work, a gaussian filter with a standard deviation of 25 pixels is used to smooth the fixation point and to generate a map. This size approximates the size of foveal region in human eye (1 visual degree approximates 50 pixels in our experimental setup).



(a) First Fixation



(b) Second Fixation

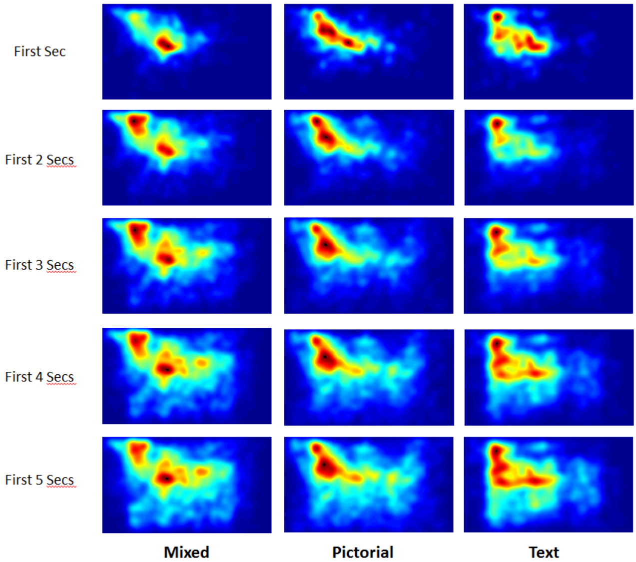


(c) Third Fixation

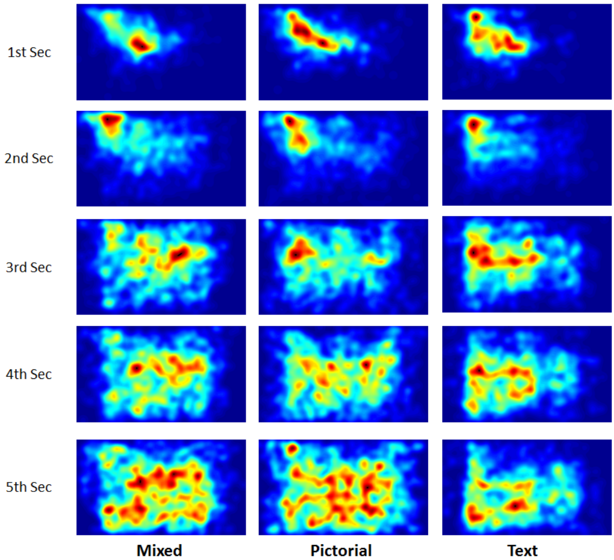
Fig. 2. Fixation heat maps of the first, second, and third fixations over all the webpage images (1st column) and the position distributions of the first, second and third fixations on three example images from the dataset.

Figure 2 visualizes the distributions of the first three fixations on all the webpages and on three individual webpages. Figure 3 illustrates category-wise fixation heat maps in first five seconds. From the figures we made the following observations: From the figures we made the following observations:

– **Positional Bias**: The positional bias on top-left region is evident in the visualization. From Figure 2, we observe that most of the first, second and third fixations fall in this region. More specifically, the first fixations tend to locate in the centre position that is slightly toward top-left corner and



(a) Accumulated fixation heat maps of fixations on three categories from the first second to the first five seconds



(b) Fixation heat maps on three categories with a second-by-second visualization

Fig. 3. Fixation heat maps on three categories of webpages.

the second and third fixations usually fall on the trajectory from center to top-left corner. From Figure 3, we can further observe that this top-left bias is common in all the three categories at first three seconds. These findings are in line with the F-bias described in [19, 12, 4]

- **Object and Text Preference**: By looking into the eye fixation distributions on each individual webpage, we found that the first several fixations usually fall on large texts, logos, faces and objects that near the center or the top-left regions (Figure 2, 2rd to 4th columns).
- **Category Difference**: From Figure 3, we observe that, in all categories, fixations tend to cluster at the center and top-left region in the first two seconds and start to diversify after the 3rd second. Webpages from the 'Text' category display a preference of the middle left and bottom left regions in 4th and 5th second while the fixations on the other two categories are more evenly distributed across all the locations.

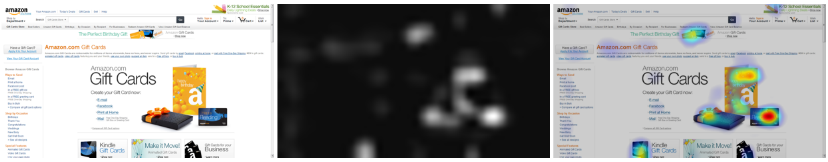
3 The Saliency Model

Data analysis from Section 2.3 suggest the following for computational modeling of web saliency: 1. positional bias is evident in the eye fixation distribution on webpages. 2. Face, object, text and website logo are important in predicting eye fixations on webpages. In this work, we propose a saliency model following the classic Itti-Koch saliency model [16, 17], which is one of the seminal works in the computational modeling of visual attention. We show below how in-depth analysis of the low-level feature cues and proper feature integration can effectively highlight important regions in webpages.

3.1 Multi-scale Feature Maps

The original Itti-Koch saliency model [16, 17] computes multi-scale intensity, color, and orientation feature maps from an image using pyramidal center-surround computation and gabor filters and then combines these maps into one saliency map after normalization. In our model, we further optimize this multi-scale representation by adding a thresholded center-surround filter to eliminate the edge artifacts in the representation. The edge artifacts are the scattered responses surrounding objects caused by center-surround/gabor filtering (especially for filters of low spatial frequency.). The additional thresholded center-surround filter is mostly to inhibit these false alarms and make the responses concentrated on the objects. The resulting model is able to capture higher-level concepts like object blobs and text, as illustrated in Figure 4.

Object representation Object blobs on webpages usually have large contrast in intensity and color to their backgrounds. In our experiments we found that object blobs could be well represented by the intensity and color maps in low spatial frequency, as shown in Figure 4(a).



(a) Object representation from intensity and color maps in low spatial frequency.

(b) Text representation from four orientation feature maps in high spatial frequency.

Fig. 4. Object blob and text representations from different feature maps. Left: input image, Middle: integrated feature map, Right: heat map overlay of input image and integrated feature map

Text representation Texts are usually in high spatial frequency and with large responses in all the orientation as they contain edges in each orientation. Based on this, the text representation can be derived from orientation feature maps. By integrating orientation feature maps in high spatial frequency, we found that almost all the texts can be encoded (Figure 4(b)).

In our implementation, we used Derrington-Krauskopf-Lennie (DKL) color space [8] to extract intensity, color and orientation features from webpage images. The DKL color space is defined physiologically using the relative excitations of the three types of retinal cones and its performance on saliency prediction is superior to RGB color space.

For the computation of multi-scale feature maps, we use six levels of pyramid images and we apply center surround filters and gabor filters with orientations of $0°$,$45°$,$90°$,$135°$ on the input image. In this way a total of 42 multi-scale feature maps are yielded on the seven channels (The DKL color space generates 3 feature maps: 1 intensity and 2 color maps, and center-surround filters are applied on these 3 channels. In addition, 4 orientation filters on the intensity map result 4 orientation maps, Thus 6 levels of image pyramid lead to a total of 42 feature maps). Based on the fact that different feature maps in different spatial frequency might encode different representations, we treat each feature map separately with MKL regression for feature integration thus each feature map would have a different contribution to the final saliency map.

3.2 Face Detection

From data analyses above, we found that in viewing webpages, human related features like eye, face and upper body also attract attention strongly, consistent with the literature [18, 5, 9, 28, 29]. We thus generate a separate face map by face and upper body detection. The upper body detector is used here to increase the robustness of face detection under different scenarios. The two detectors are based on cascaded object detector with Viola-Jones detection algorithm[26] implemented in Matlab Computer Vision System Toolbox. An scaling step of 1.02 on the input and a merging threshold of 20 and 30 are used for the face detector and the upper body detector to ensure correct detections and suppress false alarms. The face map is then generated by convolving the detection results with a Gaussian kernel.

3.3 Positional Bias

The positional bias in webpage saliency include center bias and top-left bias. In our implementation, the accumulated fixation map of all the webpages over 5 seconds is used as the positional bias and it is treated as one independent map in regression.

3.4 Feature Integration with Multiple Kernel Learning

We use multiple kernel learning (MKL) for regression and all the feature maps are integrated to predict eye fixations on webpage. MKL is a method that combines multiple kernels of support vector machines (SVMs) instead of one. Suppose we have N training pairs $\{(\mathbf{x}_i, y_i)\}_{i=1}^{N}$, where $\mathbf{x}_i$ denotes a feature vector that contains the values of each feature map on one particular position and $y_i \in \{-1, 1\}$ represents whether there is an eye fixation on the same position. An SVM model on them defines a discriminant function as:

$$f_m(\mathbf{x}) = \sum_{i=1}^{N} \alpha_{mi} y_i k_m(\mathbf{x}_i, \mathbf{x}) + b_m \tag{1}$$

where α represents dual variables, b is the bias and $k(\mathbf{x}_i, \mathbf{x})$ is the kernel. m is a subscript for each set of kernel in a standard SVM.

In its simplest form MKL considers a combination of M kernels as

$$\begin{gathered} k(\mathbf{x}_i, \mathbf{x}) = \sum_{m=1}^{M} \beta_m k_m(\mathbf{x}_i, \mathbf{x}) \\ \text{s.t. } \beta_m > 0, \sum_{m=1}^{M} \beta_m = 1 \end{gathered} \tag{2}$$

Then our final discriminant function on a vectorized input image $\mathbf{x}$ is

$$f(\mathbf{x}) = \sum_{m=1}^{M} \sum_{i=1}^{N} \alpha_{mi} y_i \beta_m k_m(\mathbf{x}_i, \mathbf{x}) + b_m \tag{3}$$

We utilized simpleMKL algorithm [20] in our model to solve this MKL problem and the probability of eye fixations on each position, or the final saliency map S, can then be obtained by

$$S = g \circ \max(f(\mathbf{x}), 0) \tag{4}$$

Where g is a gaussian mask that is used to smooth the saliency map.

4 Experiments

To verify our model in predicting eye fixations on webpage, we apply it to our webpage saliency dataset under different feature settings and then compare our results with the state-of-the-art saliency models. For a fair comparison and a comprehensive assessment, the fixation prediction results of all the models were measured by three similarity metrics and all the evaluation scores presented in this section are obtained as the highest score by varying the smooth parameter (standard deviation of a Gaussian mask) from $1\% - 5\%$ of the image width in a step of 0.05%.

4.1 Similarity Metrics

The similarity metrics we use include Linear Correlation Coefficient (CC), Normalized Scanpath Saliency (NSS) and shuffled Area Under Curve (sAUC), whose codes and descriptions are available online[1][1].

CC measures the linear correlations between the estimated saliency map and the ground truth fixation map. The more CC close to 1, the better the performance of the saliency algorithm.

AUC is the most widely used score for saliency model evaluation. In the computation of AUC, the estimated saliency map is used as a binary classifier to separate the positive samples (human fixations) from the negatives (uniform non-fixation region for classical AUC, and fixations from other images for sAUC). By varying the threshold on the saliency map, a Receiver Operating Characteristics (ROC) curve can then be plotted as the true positive rate vs. false negative rate. AUC is then calculated as the area under this curve. For the AUC score, 1 means perfect predict while 0.5 indicates chance level. Shuffled AUC (sAUC) could eliminate the influence of positional bias since the negatives are from fixations of other images and it generates a score of 0.5 on any positional bias.

NSS measures the average of the response values at fixation locations along the scanpath in the normalized saliency map. The larger the NSS score, the more corresponding between predictions and ground truths.

All these three metrics have their advantages and limitations and a model that performs well should have relatively high score in all these three metrics.

[1] https://sites.google.com/site/saliencyevaluation/evaluation-measures

4.2 Experimental Setup

To train the MKL, the image sample set was randomly divided into 119 training images and 30 testing images and the final results were tested iteratively with different training and testing sets separation. We collect positive samples and negative samples from all the webpage images in our dataset. For each image, positively labeled feature vectors from 10 eye fixation locations in eye fixation position map and 10 negatively labeled feature vectors from the image regions with saliency values below 50% of the max saliency value in the scene to yield a training set of 2380 training samples and 600 testing samples for training and validation. An MKL with a set of gaussian kernels and polynomial kernels is then trained as a binary-class regression problem based on these positive and negative samples.

4.3 Results and Performance

We first test our model in different feature settings including multi-scale feature maps with and without regression, face map, and positional bias. From Table 1, we could see that MKL regression and face map could greatly improve the model's performance on eye fixation prediction with all the three similarity metrics. However, the positional bias improve the performance in CC and NSS but does not improve sAUC, largely due to the design of sAUC that compensates positional bias itself [24, 27]

Feature Settings	CC	sAUC	NSS
Multiscale (no MKL)	0.2446	0.6616	0.8579
Multiscale	0.3815	0.7020	1.2000
Multiscale+Position	0.4433	0.6754	1.3895
Multiscale+Face	0.3977	**0.7206**	1.2475
Multiscale+Face+Position	**0.4491**	0.6824	**1.3982**

Table 1. The performance of our model under different feature settings.

Feature Settings	CC	sAUC	NSS
Pictorial	0.4047	0.7121	1.2923
Text	0.3746	0.6961	1.1824
Mixed	0.3851	0.7049	1.1928

Table 2. The performance of our model on three different categories in the webpage saliency dataset

We also test our model on each category in the webpage saliency data and train a MKL on the images inside the category with multi-scale feature maps and face map. From Table 2, we could see that the performance on all the three categories are close, however, the performance on Text is a bit smaller than that on Pictorial and the performance on Mixed is between them. This results indicate that text might be a difficult part to predict saliency. Besides, we also observe that almost all the scores in Table 2 is slightly smaller than the Multiscale+Face in Table 1, which may result from the fewer images got in each category for training.

Then, we compare our model (Multiscale+Face) with other saliency models on the webpage saliency dataset. These saliency models include GBVS [13], AIM [3], SUN [27], Image Signature [15], AWS[11] and Judd's saliency model with both a face detector and a learning mechanism [18]. All the evaluation scores presented here are also obtained as the highest score by varying the smooth parameter for a fair comparion. From the prediction results listed in Table 3, we could see that our model outperforms of all the other saliency models with all the three similarity metrics.

For a qualitative assessment, we also visualize human fixation maps and saliency maps generated from different saliency algorithms on 9 images randomly selected from our webpage saliency dataset in Figure 5. From the visualization, we could see that our model predicts important texts like title or logo to be more salient than other objects and the background. It highlights all the regions where evident texts and objects locate.

Measure	GBVS [13]	AIM [3]	AWS [11]	Signature [15]	SUN [27]	Judd *et. al.* [18]	Our Model
CC	0.1902	0.2625	0.2643	0.2388	0.3137	0.3543	**0.3977**
sAUC	0.5540	0.6566	0.6599	0.6230	0.7099	0.6890	**0.7206**
NSS	0.6620	0.9104	0.9216	0.8284	1.1020	1.0953	**1.2475**

Table 3. The performance of different saliency models on the webpage saliency dataset

5 Conclusion

Despite the abundant literature in saliency modeling that predicts where humans look at in a visual scene, there are few studies on saliency in webpages, and we in this work make a first step to the exploration on this topic. Particularly, we build a webpage saliency dataset with 149 webpages from a variety of web sources and collect eye tracking data from 11 subjects free-viewing the images. A saliency model is learned by integrating multi-scale low-level feature representations as well as priors observed web-viewing behavior. MKL is used as the computational technique for a robust integration of features, without assumption of kernel functions. Experiments demonstrate the increased performance of the proposed method, compared with the state-of-the-art in saliency prediction. As far as we know, this is the first computational visual saliency model to

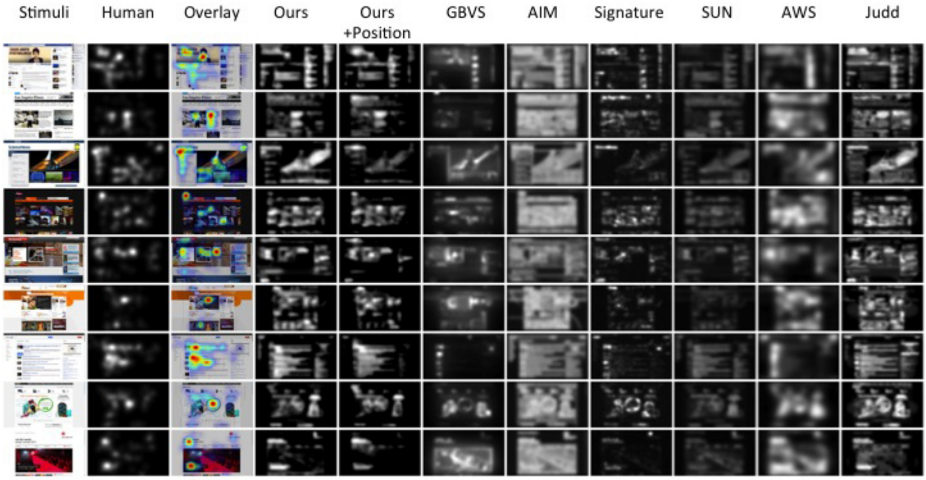


Fig. 5. Qualitative comparisons of the proposed models and the state-of-the-art on the webpage saliency dataset

predict human attention deployment on webpages, and we expect development along this line to have large commercial values in webpage design and marketing strategy.

6 Acknowledgement

The authors would like to thank Tong Jun Zhang for his help in data collection and model implementation. The work is supported by the Singapore Ministry of Education Academic Research Fund Tier 1 (No.R-263-000-A49-112).

References

1. Borji, A.: Boosting bottom-up and top-down visual features for saliency estimation. In: Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on. pp. 438–445. IEEE (2012)
2. Brainard, D.H.: The psychophysics toolbox. Spatial vision 10(4), 433–436 (1997)
3. Bruce, N., Tsotsos, J.: Saliency, attention, and visual search: An information theoretic approach. Journal of Vision 9(3) (2009)
4. Buscher, G., Cutrell, E., Morris, M.R.: What do you see when you're surfing?: using eye tracking to predict salient regions of web pages. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. pp. 21–30. ACM (2009)
5. Cerf, M., Harel, J., Einhäuser, W., Koch, C.: Predicting human gaze using low-level saliency combined with face detection. Advances in neural information processing systems 20 (2008)

6. Cho, C.H., Cheon, H.J.: Why do people avoid advertising on the internet? Journal of Advertising pp. 89–97 (2004)
7. Cutrell, E., Guan, Z.: What are you looking for?: an eye-tracking study of information usage in web search. In: Proceedings of the SIGCHI conference on Human factors in computing systems. pp. 407–416. ACM (2007)
8. Derrington, A.M., Krauskopf, J., Lennie, P.: Chromatic mechanisms in lateral geniculate nucleus of macaque. The Journal of Physiology 357(1), 241–265 (1984)
9. Einhäuser, W., Spain, M., Perona, P.: Objects predict fixations better than early saliency. Journal of Vision 8(14) (2008)
10. Faraday, P.: Visually critiquing web pages. In: Multimedia ?89, pp. 155–166. Springer (2000)
11. Garcia-Diaz, A., Leborán, V., Fdez-Vidal, X.R., Pardo, X.M.: On the relationship between optical variability, visual saliency, and eye fixations: A computational approach. Journal of vision 12(6), 17 (2012)
12. Grier, R., Kortum, P., Miller, J.: How users view web pages: An exploration of cognitive and perceptual mechanisms. Human computer interaction research in Web design and evaluation pp. 22–41 (2007)
13. Harel, J., Koch, C., Perona, P.: Graph-based visual saliency. Advances in neural information processing systems 19, 545 (2007)
14. Hervet, G., Guérard, K., Tremblay, S., Chtourou, M.S.: Is banner blindness genuine? eye tracking internet text advertising. Applied Cognitive Psychology 25(5), 708–716 (2011)
15. Hou, X., Harel, J., Koch, C.: Image signature: Highlighting sparse salient regions. Pattern Analysis and Machine Intelligence, IEEE Transactions on 34(1), 194–201 (2012)
16. Itti, L., Koch, C.: A saliency-based search mechanism for overt and covert shifts of visual attention. Vision research 40(10), 1489–1506 (2000)
17. Itti, L., Koch, C.: Computational modelling of visual attention. Nature reviews neuroscience 2(3), 194–203 (2001)
18. Judd, T., Ehinger, K., Durand, F., Torralba, A.: Learning to predict where humans look. In: Computer Vision, 2009 IEEE 12th International Conference on. pp. 2106–2113. IEEE (2009)
19. Nielsen, J.: F-shaped pattern for reading web content (2006)
20. Rakotomamonjy, A., Bach, F.R., Canu, S., Grandvalet, Y.: Simplemkl. Journal of Machine Learning Research 9(11) (2008)
21. Still, J.D., Masciocchi, C.M.: A saliency model predicts fixations in web interfaces. In: 5 th International Workshop on Model Driven Development of Advanced User Interfaces (MDDAUI 2010). p. 25 (2010)
22. Stone, B., Dennis, S.: Using lsa semantic fields to predict eye movement on web pages. In: Proc. 29th Cognitive Science Society Conference. pp. 665–670 (2007)
23. Stone, B., Dennis, S.: Semantic models and corpora choice when using semantic fields to predict eye movement on web pages. International Journal of Human-Computer Studies 69(11), 720–740 (2011)
24. Tatler, B.W.: The central fixation bias in scene viewing: Selecting an optimal viewing position independently of motor biases and image feature distributions. Journal of Vision 7(14), 4 (2007)
25. http://www.emarketer.com/Article/Social Usage-Involves-More-Platforms-More-Often/1010019: Social usage involves more platforms, more often
26. Viola, P., Jones, M.: Rapid object detection using a boosted cascade of simple features. In: Computer Vision and Pattern Recognition, 2001. CVPR 2001. Pro-

ceedings of the 2001 IEEE Computer Society Conference on. vol. 1, pp. I–511. IEEE (2001)
27. Zhang, L., Tong, M., Marks, T., Shan, H., Cottrell, G.: Sun: A bayesian framework for saliency using natural statistics. Journal of Vision 8(7) (2008)
28. Zhao, Q., Koch, C.: Learning a saliency map using fixated locations in natural scenes. Journal of Vision 11(3) (2011)
29. Zhao, Q., Koch, C.: Learning visual saliency. In: Information Sciences and Systems (CISS), 2011 45th Annual Conference on. pp. 1–6. IEEE (2011)