Directing User Attention via Visual Flow on Web Designs

Xufang Pang[*]   Ying Cao[*]   Rynson W.H. Lau   Antoni B. Chan

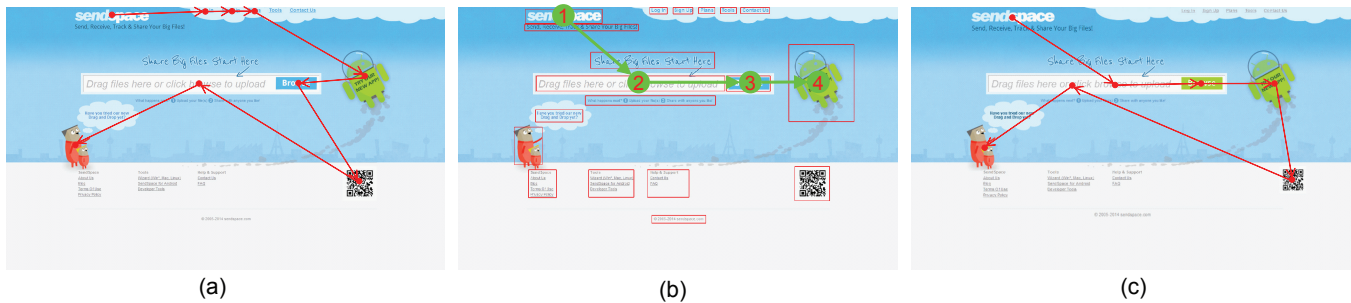Department of Computer Science, City University of Kong Kong

Figure 1: *Given an input web design in (a) and a designer-specified path (the green trajectory) over a subset of page components (red boxes) in (b), our method automatically generates a novel web design in (c) by modifying the properties of the components, such as position (e.g., the group of text links at the top), size (the QR Code at the bottom) and color (the button near the input field), so that the actual user attention paths match with the input path (i.e., the users' eye gazes when browsing the novel design should move across the subset of components successively along the input path). This results in a novel web design interaction that allows designers to easily create a visual flow to guide users (i.e., direct users'eyes to move along a specific path) through a web design. The red trajectories in (a) and (c) are the eye gaze paths of a test user on the input and novel designs.*

Abstract

We present a novel approach that allows web designers to easily direct user attention via visual flow on web designs. By collecting and analyzing users' eye gaze data on real-world webpages under the task-driven condition, we build two user attention models that characterize user attention patterns between a pair of page components. These models enable a novel web design interaction for designers to easily create a visual flow to guide users' eyes (i.e., direct user attention along a given path) through a web design with minimal effort. In particular, given an existing web design as well as a designer-specified path over a subset of page components, our approach automatically optimizes the web design so that the resulting design can direct users' attention to move along the input path. We have tested our approach on various web designs of different categories. Results show that our approach can effectively guide user attention through the web design according to the designer's high-level specification.

Keywords: web design, visual attention, layout

Concepts: •**Computing methodologies** → **Computer graphics; Graphics systems and interfaces;** *Perception;*

---

[*]Joint first authors

1 Introduction

Webpages have been ubiquitous medium for information communication since the birth of the Internet. They may be used in various fashions (e.g., personal, commercial and government webpages) to provide different functions. Webpages are a task-oriented medium, created with some objectives in mind, e.g., more sales and more clicks. Therefore, how well a webpage communicates information to users in a way intended by the designer to achieve certain objective is fundamental to web design [Jones 2011].

A web design essentially comprises a group of discrete, but semantically relevant elements (e.g., text, pictures, buttons), arranged in a 2D space. When browsing a webpage, users mentally assemble the elements as a whole to interpret the meaning behind it. Hence, how users are informed during the exploration heavily depends on them being presented with the right information at the right time. Therefore, a good web design is expected to guide users'eyes from one element to another, helping them determine what to look at next in order to make their final decisions efficiently [Guy 2011; Bradley 2013]. For example, a web designer may want potential customers of a commercial website to see something that convinces them to buy (e.g., "SALE" sign) *before* seeing the "Buy Now" button, in order to increase the chance of the customers taking a purchasing action [Bradley 2015]. Directing user attention along a specific path on a webpage via sophisticated design composition is one of the most common, yet important, strategies already adopted by web designers to convey information clearly and to make users quickly reach the goals set by the designers [Guy 2011; Bradley 2013]. While some high-level visual cues (e.g., arrows and images of people looking in one direction) have been proven to be effective in directing user attention, the frequent use of these cues in one web design is undesirable in practice, since it degrades the look and feel of the design [Bradley 2010]. Therefore, it is important for web designers to be able to guide users by manipulating the low-level visual cues of page components (e.g., color, size, position and space), to balance between the perceptual quality and business goal of a webpage. Unfortunately, while a skilled designer can wield some rules of thumb, such as alignment, symmetry and color compatibility, to achieve visually appealing designs, directing user attention

effectively on web designs by adjusting low-level design properties is non-trivial for both novice and professional designers — it requires a lot of expertise to reason about complex user attention behaviors under different configurations and a painstaking process to explore a large design space before getting a satisfactory result.

In this paper, we propose a novel web design interaction that allows designers to direct user attention using visual flow on web designs easily. Starting from an initial web design, a designer sketches a trajectory across a subset of page components to indicate the attention path that they expect users to follow, which we denote as the *designer-specified path* or *input path*. Then, the web design is optimized automatically so that the actual user attention behavior on it is as consistent with the designer's intent as possible. Our approach allows designers to easily manipulate how users move their eyes across multiple regions on web designs, thereby controlling how and in what order information is being perceived. To enable our novel design interaction, we use a data-driven approach to quantitatively model users' viewing behaviors on web designs, by gathering and analyzing eye tracking data of different users on various web designs under the task-driven condition. Unlike prior work on modeling visual saliency on webpages in the spatial domain [Shen and Zhao 2014], we focus on the temporal aspect of user attention, i.e., the transition of attention between page components, and the order that page components are visited. After acquiring eye movement data from different users on a collection of webpages, we first perform statistical analysis to study the characteristics of users' eye movements when browsing webpages. Then, we build two predictive models that describe the temporal behaviors of user attention between page components. These *user attention models* can then be used to produce optimized web designs by adjusting the attributes of the page components to minimize an objective function comprising three terms: *attention term*, *prior term*, and *regularization term*. The attention term encourages users' actual attention behaviors to match with designers' intended visual flow. The regularization and prior terms guarantee the validity of the final web design by imposing a set of visual design principles while keeping it close to the original design.

We show results of applying our approach to a variety of web designs, and conduct several user studies to validate the effectiveness of our method in directing user attention according to the designer's high-level intent. Furthermore, we also compare our approach with professional web designers in a task of directing user attention along the specified paths. The results show that the attention-directing task is quite challenging in practice and that the advantage of our approach becomes significant when the length of the specified paths increases. Note that our major goal is neither to generate design layouts of diverse styles nor to improve visual aesthetics of an input design. Instead, our approach is designed to introduce subtle variations to an input web design to effectively guide user attention behaviors, while still preserving the original layout style of the input design as much as possible.

Contributions. In summary, our main contributions include:

- A novel approach that allows web designers to easily direct user attention along a specified path through a web design.

- Two user attention models, which are learned from actual eye gaze data, to predict the temporal behaviors of user attention between two page components on a web design.

- An optimization framework, which automatically adjusts an existing web design to match the designer's intent in directing user attention.

2 Background

Layout and composition. Our approach optimizes geometric and visual properties of the components on a web design for a specific purpose. In this respect, it bears some similarity to recent works addressing layout and composition problems in various domains, such as floor plans [Merrell et al. 2010], furniture layout [Yu et al. 2011] and comic layout [Cao et al. 2012]. However, all these techniques are especially designed for their respective domains, and thus cannot be easily adopted to address our problem where understanding and modeling of user attention behaviors on web designs is required. A work more relevant to ours is [O'Donovan et al. 2014], which aims to predict the perceptual importance of components on single-page graphic designs for automtic layout generation. However, unlike their work that learns a model to predict per-pixel spatial saliency from human-labeled importance maps, we focus on modeling the temporal behaviors of user attention between page components using users' eye gaze data. In addition, our end goal is to enable intuitive manipulation of user attention behaviors on web designs, rather than automating layout generation based on a set of guidelines as in their work. Our work is also related to [Cao et al. 2014], which composes comic elements to direct reader attention over a comic page. Focusing on comics, which is a type of sequential art, they assume that there is a global reading path over a comic page and learns how the panel elements are composed around this global path. Hence, it is not amenable to webpages, where there are no pre-assumed global reading orders and viewers can start from any place and move their attention arbitrarily.

Assisting web design. Recent progress in data-driven web design has demonstrated how a large-scale corpus of webpages can be exploited to help web designers. Kumar et al. [2013] developed a scalable infrastructure for acquiring and managing a large repository of webpages crawled from the web. Such repository enables several novel design mechanisms, such as example-based webpage retargeting that allows any webpage to be used as a design template [Kumar et al. 2011], design-based search that allows designers to easily find relevant examples for inspiration [Kumar et al. 2013]. Our work shares the similar high-level goal of aiding web design process, but focuses on faciliating designers to control user attention behaviors on web designs, which has not been explored yet by the prior works.

Visual attention on webpages. Understanding where users look and how their visual attention moves on webpages are helpful for web designers to create more effective designs. Studies have found several reading patterns that users tend to adopt when navigating through webpages, e.g., Z-pattern and F-pattern [Nielsen and Pernice 2009]. These patterns have been formalized as rules that web designers mainly rely upon to encourage users to look through important regions. However, these patterns can only find limited use in modern webpages, as they are mainly applicable to text-heavy webpages [Bradley 2015]. Further, since user attention behaviors are influenced by complex factors and vary considerably under different contexts, these simple rules can only provide limited knowledge of how user attention changes with different design configurations. In contrast to using a few fixed paths to describe general user attention transition over the entire page (e.g., from left to right), our approach provides web designers with more fine-grained control over user attention, by manipulating how users' eyes move between page components to guide them along any reasonably-specified visual paths.

A recent work proposed a model to predict webpage saliency [Shen and Zhao 2014]. The major distinction between this work and ours is that, rather than predicting the perceptual importance of each location on a webpage (i.e., saliency map), we focus our prediction on temporal attention patterns between page components (i.e., how the gaze moves from one component to another) to facilitate our novel
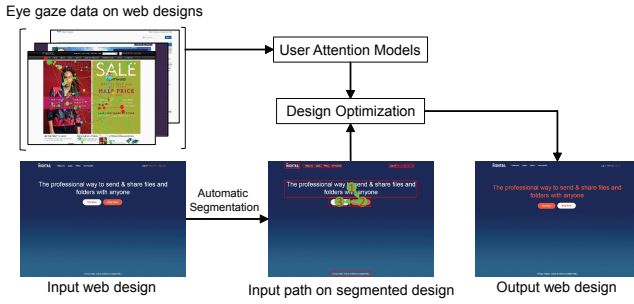
Figure 2: *Workflow of our method.*

design interaction. Some online services, such as EyeQuant [Eyequant 2016] and Feng-GUI [Feng-GUI 2016], analyze a snapshot of a given webpage and predict where viewers will look. The analysis results can then be used by the designer to improve the design iteratively. Unfortunately, they only offers analysis services. Low-level design refinement (e.g., adjusting positions, sizes and colors of components) has to be done manually by designers via a tedious trial-and-error process. In contrast, we automate the design refinement process, requiring only high-level input from web designers.

Predicting temporal behaviors of visual attention. We are not the first to investigate how human attention *transitions* on visual stimuli over time. There are many studies on analyzing the temporal aspects of eye movements on natural images [Schtz et al. 2011] or webpages [Eraslan et al. 2015] in psychology and cognitive science. Several saccadic models have also been proposed for scanpath estimation on natural images, e.g., a first-order Markov process [Ellis and Smith 1985], winner-take-all scheme [Itti and Koch 2000], and constrained random walk [Boccignone and Ferraro 2004]. Recent works incorporate more related factors such as fovea-periphery resolution discrepancy, visual working memory and viewing biases into the models to better simulate the human visual system [Wang et al. 2011; Liu et al. 2013; LeMeur and Coutrot 2016]. However, we are not aware of any prior methods especially designed for predicting the temporal behaviors of user attention on graphic designs (e.g., webpages and posters), which are created by designers via sophisticated choice of page layout, colors and typography. Our experimental results shows that our dedicated attention models outperform the general saccadic models in predicting eye movements on web designs in our problem setting. There are also some works on modeling scanpaths in visual search tasks [Najemnik and Geisler 2009]. However, their approaches are specific to the scenarios where observers search for a known target in a visual scene, and thus cannot be adapted to address our needs.

3 Overview

Given an input web design, we first segment it into a set of page components, and extract the component attributes from its HTML source file and a visual hierarchy tree from its DOM tree [Kumar et al. 2013]. Since there may exist regions that draw users' attention but are difficult to segment automatically, e.g., faces in a background image, designers are also allowed to manually segment some regions as page componets. Then, the designer draws a path across a subset of components, indicating the intended attention path of the components, Finally, the input design is automatically refined according to the designer's high-level intent. Note that as the objective of this work is not to improve visual aesthetics of a web design, we assume that the input web design already has a reasonably good quality in terms of some visual criteria (e.g., balance and alignment).

The key to our approach is a pair of user attention models that describe the temporal behaviors of user attention between a pair of page components. To understand and model user behaviors when viewing web designs, we collect and analyze eye-tracking data of different users on a web design corpus consisting of different genres (e.g., shopping, game and social media) under both task-driven and free-viewing conditions. Based on the collected data, we build an attention *transition* model, which can predict the likelihood of user attention moving from one page component to another, and an attention *order* model, which can predict the likelihood of one page component being visited before another. The two user attention models are then used in an optimization framework to generate refined web designs. Figure 2 shows the workflow of our method.

4 Data Collection

To understand the user behaviors when browsing web designs, we collected a corpus of 254 webpages from 6 different popular categories, each of which has a single and specific purpose. The categories are: email, file sharing, job searching, product promotion, shopping and social networking. All the webpages were crawled from the web and their snapshots are stored as web designs. We selected the webpages with varying design complexity in the number of components and the proportions of image and text components, so that our dataset, although not exhaustive, can be representative of mainstream web designs found online. Refer to Section S1 of the supplemental for the summary statistics of our dataset.

We used an eye tracker to record eye movements of human subjects under two different viewing conditions: task-driven and free-viewing. In each viewing session under the task-driven condition, each participant was asked to complete a sequence of tasks, each containing a task instruction relevant to the webpage category, followed by one or two web designs to view. Our tasks fall into two categories: *comparison* task and *shopping* task. In comparison tasks, the participants were told to compare two web designs offering similar services (e.g., social networking), and then select the one they preferred to use. In shopping tasks, participants were given certain amount of money and asked to view a single shopping web design to buy a favorite item. The presentation order of the tasks was randomized for each participant. In the free-viewing condition, we showed each participant a sequence of web designs randomly selected from our dataset, and asked them to view the pages casually without any constraint. For both conditions, participants were allowed to manually advance their own progress via mouse-clicks, without any time constraint. Participants did not view the same web design more than once. Refer to Section S2 of the supplemental for more details on our eye tracking experiment.

We captured the eye gaze data of each participant as a temporal series of fixation points. Each fixation point contains a 2D position $\mathbf{x} = (x, y)$ along with its duration d. For each web design under each mode, we collected eye movement data from 10 participants. In a post-processing stage, we used the Bento algorithm to segment each page in our dataset into a visual hierarchy tree of components as in [Kumar et al. 2013].

Attention transition graph. For each web design under each viewing condition, we construct an *attention transition graph* as in [Cao et al. 2014], which summarizes how users move their attention between page components. The nodes of the attention transition graph are the page components corresponding to the leaf nodes of the visual hierarchy tree. We add a directed edge from one component c_i to another c_j if there exists any user eye movement from c_i to c_j. We encode edge weights of the graph in an *attention transition matrix* $\mathbf{T}$, where each entry $\mathbf{T}(i, j)$ denotes the percentage of total users that move their eyes from c_i to c_j. Figure 3 shows
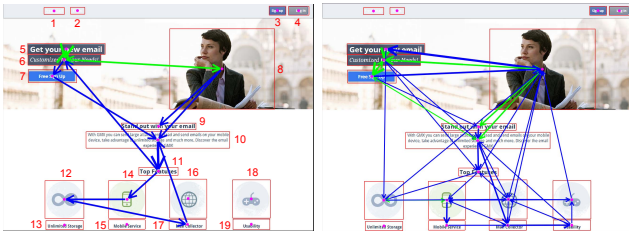
Figure 3: *Attention transition graphs for the task-driven condition (left) and the free-viewing condition (right) on a web design. Blue and green arrows represent single-direction and double-direction attention transitions, respectively. The arrow thickness is proportional to the fraction of users moving attention along the arrow.*

two example attention transition graphs for task-driven and free-viewing conditions on a web design.

Partial order matrix. To study the temporal order that users visit the page components, we also build a partial order matrix $\mathbf{O}$ for each web design, where each entry $\mathbf{O}(i,j)$ indicates the percentage of total users visited component i before component j. The probability that a page component is visited *before* any other components (i.e., first) is obtained from the partial order matrix by normalizing the matrix to sum to one and summing along each row, $\mathbf{h}(i) = \sum_k \hat{\mathbf{O}}(i,k)$, where $\hat{\mathbf{O}}$ is the normalized partial order matrix. We denote $\mathbf{h}$ as the *order probability*, which summarizes the temporal order that users look at each component on the page.

Using the collected data, we have performed a statistical analysis to verify if there exists discrepancies in user behaviors between different contexts. Our results show that user eye movements on web designs under the task-driven condition are different from those under the free-viewing condition (see details in Section S3 of the supplemental). With a task in hand, users tend to directly gravitate to the components that are likely to be useful for completing their goals, rather than looking at all information as in the free-viewing mode. Due to the task-oriented nature of webpages, it is more useful to model user attention behaviors under the task-driven condition. Hence, we focus our study in later sections on eye gaze data collected under the task-driven condition, and build task-driven user attention models. As such, for the rest of this paper, all eye gaze data refers to those from task-driven users, unless otherwise stated.

5   User Attention Models

In this section, our goal is to construct models that capture the correlation between the properties of two page components and user attention movement between them. For this purpose, we propose two independent user attention models, a user attention *transition* model and a user attention *order* model.

User attention transition model. Let $c_s \rightarrow c_d$ be a pair of ordered page components, where c_s and c_d are source and destination components, respectively. Our model f_t maps the features $\mathbf{v}$ of c_s and c_d to a probability of user attention shifting from c_s to c_d as: $p_t(c_s \rightarrow c_d) = f_t(\mathbf{v})$. It is learned using the collected eye gaze data, described in Section 4. For each web design, we enumerate all pairs of ordered components. For each component pair, we extract a set of features $\mathbf{v}_i$ and the corresponding user attention transition probability p_i from the attention transition matrix $\mathbf{T}$. We ignore pairs that start with a component whose corresponding row in $\mathbf{T}$ only has zeros. This is because users never looked at these page components and there would not be user attention transitions between the corresponding pairs. We also ignore pairs whose page components were visited by less than 10% of the users, as the

computation of their attention transition probabilities based on only a few samples is not statistically reliable. This results in a set of training examples $\mathcal{D} = \{(\mathbf{v}_i, p_i)\}$.

When selecting features for the attention models, computational efficiency is a main concern since the models are used in the cost function of our iterative optimization. Hence, we prefer simple and effective low-level features (e.g., position, orientation, intensity, colors, etc.), which have been shown to influence human visual attention [Itti and Koch 2000] and viewing behaviors on webpages [LeMeur and Coutrot 2016]. The feature vector $\mathbf{v}$ for a pair of ordered components $c_s \rightarrow c_d$ comprises two types of features: pairwise features and context features. The pairwise features capture the spatial and appearance properties of c_s and c_d that may affect user attention transition. They include:

- **Scale ratio** - the ratio between areas of c_s and c_d.
- **Normalized distance** - the minimum Euclidean distance between the bounding boxes of c_s and c_d, normalized with respect to the diagonal length of the web design. The distance is set to 0 if the two bounding boxes intersect.
- **Normalized position** - the center of the line that joins the center of c_s's bounding box and that of c_d's bounding box, relative to the center of the web design. It is normalized with respect to the dimensions of the web design.
- **Relative orientation** - the angle between the vector from the center of c_s to that of c_d and the horizontal vector. It is mapped to $[0^o, 360^o]$ and then normalized to $[0, 1]$.
- **Intensity difference** - the difference between the average intensities of c_s and c_d.
- **Color contrast** - χ^2 distance between the Lab color histograms of c_s and c_d.

The context features capture how the other components in the neighborhood of c_s and c_d affect attention transition between c_s and c_d. For example, if a component is closer to c_s than c_d, users may move their eyes directly to this component rather than to c_d. We define the neighbors of c_s as any components (except c_d) whose normalized distances (with respect to the diagonal length of the web design) to c_s are smaller than r. By inspecting our eye tracking data, we empirically found that the component pairs with normalized distances larger than 0.4 had almost zero attention transition probabilities. Thus, we refer to r as the *effective normalized distance* and set it to 0.4 in our implementation. Our context features include:

- **Normalized distance to the nearest neighbor** - the normalized distance from c_s to its nearest neighbor, calculated in a similar way as the pairwise feature. It is set to 1 if no nearest neighbors within r are found.
- **Relative orientation to the nearest neighbor** - the relative orientation feature between c_s and its nearest neighbor, calculated in the same way as the pairwise feature. It is set to 0 if no nearest neighbors within r are found.
- **Size ratio to the largest neighbor** - the ratio between the area of c_s and that of its largest neighbor. It is set to 1 if no nearest neighbors within r are found.
- **Brightness ratio to the brightest neighbor** - the ratio between the luminance of c_s and that of its brightest neighbor. It is set to 1 if no nearest neighbors within r are found.

To learn our models from the training data, we use random forest regression [Breiman 2001] with 100 trees, as it is fast in training and prediction.

User attention order model. Besides the attention transition probability, we are also concerned with the attention order between two components, i.e., if one component is visited before or after another. Therefore, we construct an attention order model f_o,

which takes as input the features $\mathbf{v}$ of a pair of ordered components $c_s < c_d$ and predicts the probability of c_s being visited before c_d, $p_o(c_s < c_d) = f_o(\mathbf{v})$. The attention order model is adapted from the attention transition model by changing the output of the attention transition model to be the probability of one component being visited before another. Consequently, during the training phase, p_i of each training example is computed from partial order matrix $\mathbf{O}$, instead of attention transition matrix $\mathbf{T}$, described in Section 4.

6 Design Optimization

Given an existing web design and a designer-specified path in the form of an ordered sequence of components $\mathcal{S} = \{c_1, c_2, \ldots, c_k\}$, our goal is to automatically adjust the web design, such that the actual user attention path matches with $\mathcal{S}$. We represent each web design as a union of page components. We parameterize the i-th component using its center position (x_i, y_i), size (width w_i and height h_i), and color r_i. Thus, a web design can be represented as $C = \{(x_i, y_i, w_i, h_i, r_i)\}$. We formulate the design adjustment as an optimization problem, and minimize an objective function comprising three terms: attention, prior, and regularization, i.e., $\mathbf{E}(C) = \alpha_1 \mathbf{E}_{att}(C) + \alpha_2 \mathbf{E}_{prior}(C) + \alpha_3 \mathbf{E}_{reg}(C)$, where $\alpha_1 = 0.5, \alpha_2 = 0.3, \alpha_3 = 0.2$ in our implementation. We optimize over the components corresponding to the leaf nodes in the visual hierarchy tree, and do not alter the components whose properties cannot be changed, e.g., an object that is part of a large background image (such as component 8 in Figure 3). In this section, we first introduce the formulation of each term in details, and then describe how to optimize our objective function.

6.1 Energy Terms

Attention term. The attention term $\mathbf{E}_{att}$ is used to encourage the actual attention path to match with the input path $\mathcal{S}$. In other words, it prefers the web design configuration where users will follow the input path $\mathcal{S}$, by successively looking at the components in $\mathcal{S}$. It is defined as: $\mathbf{E}_{att} = \mathbf{E}_{att}^O + \mathbf{E}_{att}^T$, where $\mathbf{E}_{att}^O$ and $\mathbf{E}_{att}^T$ are the order term and transition term, respectively.

The order term $\mathbf{E}_{att}^O$ constrains each component c_i in $\mathcal{S}$ to be read before its successors $\mathcal{S}_i = \{c_j | j > i, c_j \in \mathcal{S}\}$, and is defined as:

$$\mathbf{E}_{att}^O = -\frac{1}{|\mathcal{S}| - 1} \sum_{c_i \in \mathcal{S}} \frac{1}{|\mathcal{S}_i|} \sum_{s \in \mathcal{S}_i} [p_o(c_i < s) - p_o(s < c_i)], \quad (1)$$

where $p_o(s < c_i)$ is the probability of s being visited before c_i as predicted by our attention order model. $\mathbf{E}_{att}^O$ favors the user attention from c_i to all its successors and penalizes the user attention movement in the opposite direction. However, this term alone does not suffice to satisfy our requirement since it only enforces the components in $\mathcal{S}$ to be read in a given order, but not necessarily in *successive* manner. For example, after viewing c_i, users may move their eyes to other components that are not included in $\mathcal{S}$ before looking at c_{i+1}. The transition term $\mathbf{E}_{att}^T$ is introduced to ensure that user attention successively transitions across the components in $\mathcal{S}$. However, user attention transition is less likely to occur between two components with a distance larger than the effective normalized distance r, as discussed in Section 5. Thus, when computing the transition term, for each component c_i, we only consider its neighbors $\mathcal{N}_i$ within r. $\mathbf{E}_{att}^T$ is:

$$\mathbf{E}_{att}^T = -\frac{1}{|\mathcal{S}| - 1} \sum_{c_i \in \mathcal{S} \setminus c_k} [p_t(c_{i+1}|c_i) - \max_{e \in \mathcal{N}_i \setminus c_j} p_t(e|c_i)], \quad (2)$$

where c_k is the last component in $\mathcal{S}$. $p_t(c_j|c_i) = \frac{p_t(c_i \to c_j)}{\sum_{e \in \Omega} p_t(c_i \to e)}$, where Ω is the set of all components in the input design and

$p_t(c_i \to c_j)$ is the probability of user attention transition from c_i to c_j by our attention transition model. This term will encourage user attention transition from one component $c_i \in \mathcal{S}$ to its immediate successor, while suppressing the transition from c_i to other neighboring components.

In some scenarios, designers may want the starting component of the input path $\mathcal{S}$ to be the first component that users will look at on the web design. We can support this design intent by adding a first-fixation term $\mathbf{E}_{att}^F$ to the attention term defined above. Let $\mathbf{h} = \{h_i\}$ be the order probability calculated from the predicted partial order matrix. We then define the first-fixation term as:

$$\mathbf{E}_{att}^F = -(h_{c_1} - \max_{e \in \Omega \setminus c_1} h_e), \quad (3)$$

where c_1 is the first component in $\mathcal{S}$. Note that this term is optional and disabled by default. Hence, all the results shown in the paper are generated without this term, unless stated otherwise.

Prior term. An effective web design should closely adhere to several visual design principles, such as alignment and symmetry. To account for these design principles, we introduce a prior term that measures how well a given design conforms to the most widely used graphic design guidelines. The prior term consists of the energy terms for alignment $\mathbf{E}_a$, balance $\mathbf{E}_b$, white space $\mathbf{E}_w$ and overlap $\mathbf{E}_o$ that are implemented in [O'Donovan et al. 2014] to model single-page graphic designs. Moreover, since we modify the colors of page components during the optimization, we further introduce a color term $\mathbf{E}_c$ to enforce color compatibility of the optimized web design. We define $\mathbf{E}_c = 1 - \frac{s}{5}$, where $s \in [0, 5]$ is the score returned by the color compatibility model learned from a large dataset of color themes [O'Donovan et al. 2011]. Finally, the prior term is formulated as: $\mathbf{E}_{prior} = w_1 \mathbf{E}_a + w_2 \mathbf{E}_b + w_3 \mathbf{E}_w + w_4 \mathbf{E}_o + w_5 \mathbf{E}_c$. In our implementation, we apply equal weights, i.e., $w_i = 0.2$.

Regularization term. The regularization term is used to prevent the optimized web design from deviating too much from its original configuration as: $\mathbf{E}_{reg} = ||C - C_0||^2$, where C_0 is the configuration of the input web design.

6.2 Optimizer

Since our objective function is highly multimodal, we employ Markov chain Monte Carlo sampling techniques to efficiently explore its solution space. In particular, we use Metropolis-Hastings algorithm [Metropolis et al. 1953; Hastings 1970] to iteratively sample the design until a maximum number of iterations (i.e., 1000) is reached or the change in $\mathbf{E}(C)$ between two consecutive iterations is smaller than 0.001. At each iteration, we maintain a page configuration C, and generate a new configuration C' by randomly drawing among several proposal moves:

- **Update component position** - perturb the position of a random component by adding Gaussian noise.
- **Update component size** - perturb the size of a random component by scaling it with Gaussian noise.
- **Update text color**: replace the text color of a random component with a random color from a page-specific color palette.[1]
- **Update button color**: replace the color of a random button component with a random color selected from the design-specific color palette.

---

[1] The color palette is constructed by applying K-means clustering to extract 10 representative colors in Lab space from the input web design. To make sure that the text is visible, we exclude colors that are similar to the average color of the text's neighborhood. To guarantee visual coherence, we also force the text components belonging to the same group (i.e., the nodes sharing the same parent in the visual hierarchy tree) to take the same color, unless the selected text component is on the input path.

- **Align components**: use a random component as an anchor, randomly select an alignment type (e.g., left alignment), and move the other components that are near to the anchor (e.g., in terms of left boundary distance) to align with the anchor.

To maintain the visual structure of the input design, we identify the child-parent and sibling relations among the components using the extracted visual hierarchy tree, and neglect invalid moves based on these relations. In particular, we ignore the moves that result in a component falling outside the boundaries of the input web design or its parent container. We also discard the moves that cause the sibling components (i.e., the nodes belonging to the same parent node in the visual hierarchy tree) to overlap.

7 Results and Evaluation

In this section, we first present qualitative results on a variety of web designs from different categories and evaluate the effectiveness of our method in directing user attention and preserving the visual quality of original web designs via an eye-tracking experiment and a visual perception study. Then, we compare our results against those by professional web designers, and perform a preliminary A/B test to evaluate how guiding users along specific paths on web designs could affect user behaviors. Finally, we evaluate the performance of our user attention models and the importance of the various energy terms in our optimization.

We have run our method on a PC with an i7 3GHz CPU and 22GB RAM. We tested it on 30 web designs from 6 categories (5 web designs per category). For each web design, we specify three different input paths of lengths $\{3, 4, 5\}$ to generate three novel web designs, which results in a total of 90 optimized designs. The input paths are set to be as diverse in orientation and position as possible and have reasonable coverage of the entire page space. See the distributions of all the input paths in S4 of the supplemental. Note that all results shown in this paper and the supplemental are from web designs that are not part of the training dataset.

7.1 Qualitative Analysis

Figure 4 shows four example web designs along with their respective optimized versions. Note that directing user attention along a specified path entails complex and subtle changes to the input web design. For example, consider the middle result ($1 \rightarrow 6 \rightarrow 5 \rightarrow 7$) of the first row of Figure 4. To increase the probability of the user's eyes transiting from component 1 (the logo) to component 6 (the computer picture), components 3 and 4 are made smaller, and the group of text links above component 6 are pushed away from component 1. In addition, component 5 (the red button) and component 7 (the text "Learn more about ...") are set to the same color, so that the user can perceptually group them together and naturally transit from components 5 to 7. Furthermore, the three icons below component 7 are made smaller to reduce the probability of transiting from component 5 to the icons. See more results in Section S5 of the supplemental.

7.2 Eye tracking experiment

We next perform an eye-tracking experiment to evaluate how well our method aids designers in controlling viewer attention behaviors on web designs. In particular, we collect eye gaze data from users on the 90 optimized web designs generated by our method under the task-driven condition, and measure the consistency between the designer-specified path and actual users' eye gaze paths. In addition, to test the first-fixation term in our optimizer, for each of the 30 web designs, we select one of its input paths at random and run

our optimization with term $\mathbf{E}_{att}^{F}$ enabled, resulting in an additional 30 optimized web designs.

The experiment involves 40 participants, all of whom had no prior knowledge about the purpose of the study, and were different from the participants used for data collection in Section 4. The experiment setup is the same as in Section 4, where the participants viewed the web designs under the task-driven condition. Each comparison task comprises the optimized versions of two different input web designs from the same category. For each design, we collected eye gaze data from 10 different participants, and calculated the percentage of viewers whose eye gaze paths over the components in the input path exactly matched the input path (i.e., among all the components in the input path, viewers visit the starting component first, and then visit the succeeding components consecutively), which is referred to as *matching rate*. Higher matching rate indicates better ability to direct user attention. For comparison purpose, we also calculated the matching rate of the original input web designs, using eye tracking data from 10 different users.

Figure 5 (left) compares the average matching rates for all the optimized web designs and their corresponding input web designs. The optimized web designs achieve significantly higher matching rates, as compared with the input web designs. This implies that our method is effective in improving the input web designs to direct user attention behaviors in a way specified by the input paths.

We further study the impact of the input path length upon the performance of our method, by analyzing how the matching rate changes as the length increases from 3 to 5. We group the optimized web designs based on the input path length, and compute the average matching rate of each group, as shown in Figure 5 (right). We can see that the ability of our method in directing user attention decreases as the input path length increases. (Pearsons correlation coefficient between matching rate and input path length is -0.99.)

Finally, we evaluate the effect of the first-fixation term $\mathbf{E}_{att}^{F}$ by comparing pairs of output web designs optimized using the same input path with/without $\mathbf{E}_{att}^{F}$. With $\mathbf{E}_{att}^{F}$ enabled, the first component on an input path should be visited as early as possible. Hence, our evaluation is based on the time spent by users on a web design before visiting the specified first component. On average, it takes 1.9 ± 1.0 seconds for the users to visit the first component of an input path with $\mathbf{E}_{att}^{F}$, and 2.3 ± 1.1 seconds without $\mathbf{E}_{att}^{F}$. Such time reduction is statistically significant (independent t-test, $t(598) = 4.9$, $p < 0.001$). Comparison of results with/without $\mathbf{E}_{att}^{F}$ can be found in Section S6 of the supplemental.

7.3 Visual Quality Evaluation

To test whether our method can preserve the visual quality of the original web design, we perform a perceptual study where participants visually evaluate our results against original web designs via pairwise comparisons. We recruited 15 participants with no prior experience or training in web design. The participants were shown a pair of web designs and then chose which one was more visually attractive. "Tie" can be given if two web designs are regarded as similar visual quality. Each comparison pair includes an original web design and its optimized version by our method, which were displayed side by side in random order. For each of the 30 web designs used in Section 7.1, we paired it with each of its 3 optimized versions from 3 input paths, forming a total of 90 pairs for comparison. We distributed all the pairs among the participants, so that each participant compared 30 pairs and each pair was evaluated by 5 different participants. In total, 450 pairwise comparisons were generated. As shown in Figure 6, our results are perceived to have similar visual quality to the original webpages.

Figure 4: *Examples of web designs generated by our method from four input web designs. For each of the input designs (1st column), three optimized web designs (2nd, 3rd, and 4th columns) are generated using three different input paths shown on top of the optimized designs. This figure is best viewed on a computer screen.*

7.4 Comparison to Professional Web Designers

To assess how well professional web designers can complete the task of directing user attention behaviors, we perform a study to compare our results to those by professional web designers. We recruited 4 web designers, all of whom have more than 2 years of experience in web design. Given a set of existing web designs, each with an input path, the designers were asked to edit the web designs to accomplish two objectives: 1) to direct users' attention along the given paths; 2) to preserve the layout style and visual quality of original web designs as well as possible. For the task, the designers could choose their favorite webpage and image editing tools to edit the HTML source files (e.g., using Adobe Dreamweaver) or screenshots of web designs (e.g., using Adobe Photoshop). For fair comparison, the editing operations were restricted to be the same as what our optimizer does (i.e., changing the position and size of a page component, and changing the color of text and button). The

study was performed on the 30 web designs used in Section 7.1. For each design, we randomly selected an input path among the three paths of different lengths, in such a way that we end up with 10 designs for each of the three path lengths $\{3, 4, 5\}$. The 30 designs were distributed among the 4 designers, so that 2 designers edited 8 pages each while the other two edited 7 pages each. At the end of the study, the designers were asked to fill a questionnaire regarding the task. The designers spent an average of 15 ± 7 minutes on editing one web design. This suggests that editing web designs to direct user attention behaviors is a time-consuming process, requiring a lot of manual effort even for professionals. In contrast, our method only takes 5 ± 3 minutes to optimize one web design, which is almost three times faster than the manual process.

With the results by the designers, we performed an eye-tracking experiment using the same setup as Section 7.2. For each web design, we collected the eye gaze data from 10 different users. Then, we compared the eye gaze path matching rates of the designer-created
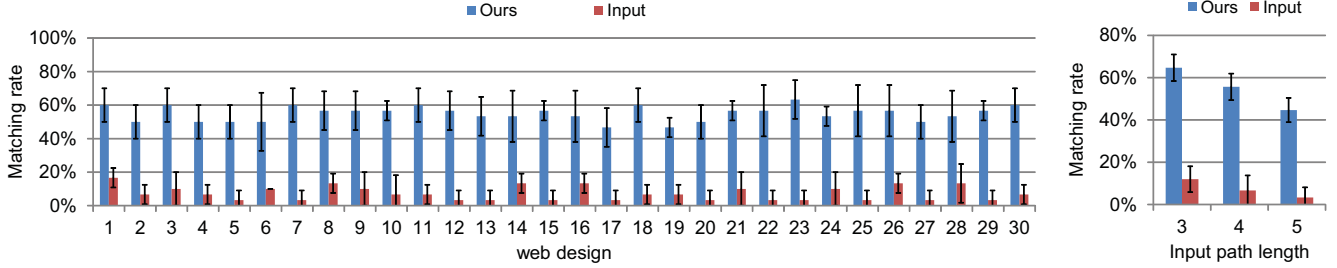
Figure 5: *Left: average matching rates for eye gaze paths on web designs generated by our method (Ours), compared with their respective input web designs (Input). Overall, Ours has an average matching rate of $55\% \pm 10\%$, which is significantly higher than $7\% \pm 7\%$ of Input (independent t-test, $t(178) = 36.6$, $p < 0.001$). Right: average matching rates versus input path length.*
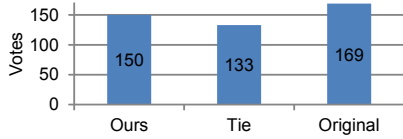

Figure 6: *Visual quality evaluation results. Although the original web designs are slightly preferred to ours, this preference is not statistically significant (Chi-squared test, $\chi^2(2) = 2.5$, $p = 0.11$).*

web designs with those of our results generated from the same set of inputs. Our method is able to achieve a higher average matching rate ($55 \pm 10\%$) than the designers ($49 \pm 17\%$), according to a Wilcoxon signed-rank test with $p = 0.009$. The matching rate of each web design can be found in Section S7 of the supplemental. Figure 8 shows that, similar to our method, the performance of the designers decreases as the input path length increases. For simple paths with 3 components, the results by the designers are slightly better than ours, whereas the performance of the designers drops dramatically for more complex paths with 4 or 5 components and becomes significantly worse than that of our method. One possible explanation is that, with more components in an input path, there is an exponential increase in the number of pairwise interactions between components and its neighbors that might affect user attention. This makes it difficult for the designers to balance these factors optimally when reasoning about user attention behaviors. Figure 7 shows some results generated by the designers and our method. More comparisons can be found in Section S6 of the supplemental. Note that, in this study, we only demonstrate the advantage of our method over professional web designers in a restricted setting, where the designers are only allowed to perform a limited set of editing operations (i.e., the proposal moves in our optimizer). In practice, to direct user attention, web designers may have more degrees of freedom to control, e.g., adding an additional visual cues (e.g., arrows and lines) to the input web design, which could possibly lead to better results. While we have shown that using our proposed proposal moves can already improve the attention-directing ability of web designs significantly, an interesting direction for future work is to extend our method to include more sophisticated operations and compare it with web designers with less restriction.

In the post-study questionnaires, when asked to rate the difficulty level of the task, two designers rated it as "difficult" and "very difficult", while the other two regarded it as "normal". When asked about where they learned their strategies for directing user attention, all designers reported that they relied on "experience", rather than learning from courses or books. This further confirms that directing user attention on web designs is a non-trivial and subjective task even for experienced web designers. Furthermore, we also asked the designers to rate their confidence level of achieving the

task objectives using the allowed operations on a scale of 1 to 5, with 1 least confident and 5 most confident. The designers gave an average rating of 3.75 ± 0.5, which suggests that the proposed operations in our optimizer, while limited, are reasonably sufficient for the designers to direct user attention on web designs.

7.5 A/B Test

We have performed a preliminary A/B test, with 160 participants, to study whether guiding users through web designs using our method can increase the chance of users taking actions expected by web designers. Our test is performed on randomly selected web designs from each of the 6 categories in our dataset. The results show that our optimized web designs achieve significantly higher action rates (47.9% on average) over the original web designs (35.8% on average), according to an independent t-test with $p < 0.001$. However, we have found that our method cannot produce statistically significant gains for the "social" category. This is perhaps because some original web designs have already been optimized for similar purposes. Furthermore, the frequent use of fancy images and personal photos could predominantly attract user attention, making it difficult to guide user attention along the given paths. Refer to Section S8 of the supplemental for more details. More thorough evaluations on how guiding users along visual flows may affect user behaviors are left as future works.

7.6 Evaluation of the User Attention Models

To evaluate the effectiveness of our user attention models, we apply a cross-validation approach on our training dataset. We randomly split the training data $\mathcal{D}$ into 10 folds. For each fold, we train our models on 9 folds, and then calculate the root mean squared error (RMSE) of model predictions on the remaining fold. The average RMSE over all 10 splits provides an estimate of the generalization performance of our models. Since there may exist great variation in attention behaviors across different users and our models are used to make predictions for unseen users, it is important to test how well our models can perform in such across-user prediction. Therefore, our folds are separated in such a way that the eye gaze data for the same user does *not* appear in both the training and test folds.

The RMSE only shows the model accuracy for predicting each individual probability in the attention transition or attention order matrix. It is also interesting to study how well the models can predict the relative ranking of components based on these matrices. Relative ranking is important in our optimization, where the models are used to encourage the most likely path to be consistent with the designer-specified path. To this end, we further evaluate our models via cross-validation using three correlation-based metrics:

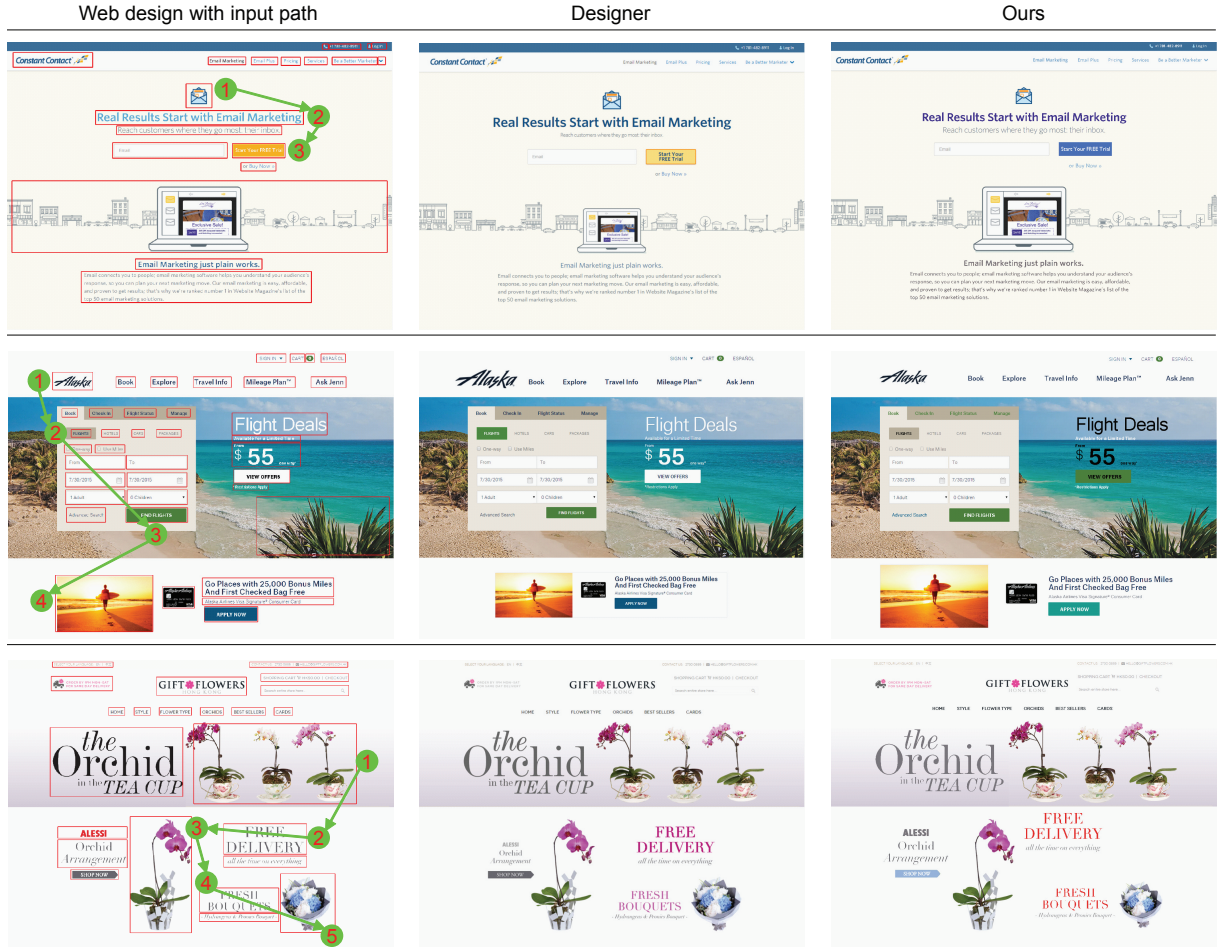- *Row rank similarity (RRS)* is used to evaluate the attention

Web design with input path	Designer	Ours

Figure 7: *Comparison of the web designs by professional designers and our method from the same input web designs and input paths.*
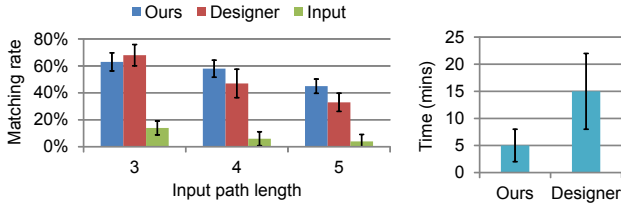

Figure 8: *Left: average matching rates of our method, professional designers and input web designs w.r.t. input path length. For input path length of 3, designers perform slightly better than our method (Wilcoxon signed-rank test, $p = 0.23$), while our method performs significantly better than designers for input path lengths of 4 and 5 (Wilcoxon signed-rank tests, $p = 0.008$ and $p = 0.004$, respectively). Right: average times of optimizing one web design by our method and professional web designer.*

transition model. Let $\mathbf{T}_p$ and $\mathbf{T}_g$ be the predicted and ground truth attention matrices for one web design. For the i-th row of $\mathbf{T}_g$ that does not have all zero values, we first obtain a ranking $\mathbf{g}_i$ for the components with non-zero probabilities. If multiple components have the same value, then we determine their order randomly. Next, for the corresponding row of $\mathbf{T}_p$, we obtain a rank $\mathbf{p}_i$ by sorting the components in $\mathbf{g}_i$ based on their corresponding values in $\mathbf{T}_p$. Finally, RRS is computed as Spearman's rank correlation coefficient between $\mathbf{g}_i$ and $\mathbf{p}_i$,

which is normalized into the range $[0, 1]$.

- *Pairwise order matching rate (POMR)* is used to evaluate the attention order model. Given a ground-truth attention order matrix $\mathbf{O}_g$, we first determine the visiting order between each pair of components (i, j), by calculating $\delta_{i,j}^g = \mathbf{O}_g(i, j) - \mathbf{O}_g(j, i)$. Component i is considered to be visited before component j if $\delta_{i,j}^g > 0$, and vice versa. Then, POMR is defined as the percentage of component pairs whose visiting orders are correctly predicted by our attention order model. We exclude the component pairs whose $\delta_{i,j}^g = 0$ since it is impossible to determine their visiting orders.

- *First component matching rate (FCMR)* is also used to evaluate the attention order model. It measures the accuracy in predicting the first component that users will fixate on a web design. Given a web design, the first-fixated component is found by computing the order probability $\mathbf{h}$ from the attention order matrix $\mathbf{O}$ as in Section 4, and identifying the component with the highest probability. FCMR is then defined as the percentage of web designs whose first visiting components are correctly predicted by our attention order model.

Effects of features. We first test the importance of each feature in our model, by comparing our models with weaker models where one of the features is excluded. The results, as shown in Table 1, suggest that our models with all the features outperform other variants with one feature excluded. More over, without the context fea-

	RMSE		Correlation-based metrics		
	transition	order	RRS	POMR	FCMR
all features	**0.066**	**0.072**	**0.82**	**83%**	**78%**
w/o scale ratio	0.083	0.085	0.75	79%	73%
w/o normalized distance	0.080	0.083	0.69	72%	71%
w/o normalized position	0.079	0.080	0.76	81%	72%
w/o relative orientation	0.078	0.086	0.78	75%	73%
w/o intensity difference	0.085	0.080	0.76	73%	70%
w/o color contrast	0.082	0.080	0.68	73%	69%
w/o normalized distance to nearest neighbor	0.078	0.082	0.69	72%	68%
w/o relative orientation to nearest neighbor	0.078	0.081	0.70	72%	70%
w/o size ratio to its largest neighbor	0.079	0.085	0.71	70%	68%
w/o brightness ratio to its brightest neighbor	0.078	0.082	0.70	73%	69%
w/o context features	0.109	0.112	0.63	65%	63%
[Boccignone and Ferraro 2004]	0.146	0.119	0.66	14%	45%
[LeMeur and Coutrot 2016]	0.129	0.102	0.70	21%	48%

Table 1: *Comparison of our model with full features against the variants of our model with a subsets of features and previous models that estimate eye movements from saliency maps.*

tures, the performance of our models dramatically decreases, which implies their importance.

Comparison to previous models. We next compare our models with two previous models that can be used to predict the temporal behaviors of visual attention: 1) the model in [Boccignone and Ferraro 2004], which formulates gaze shift as a Levy flight process, constrained by a saliency map; 2) the state-of-the-art saccadic model in [LeMeur and Coutrot 2016], which determines the next fixation point by modulating a saliency map using the viewing biases that are modeled using prior distributions of saccade amplitudes and orientations. We test both models on our dataset in Section 4. For the model in [LeMeur and Coutrot 2016], we learn the probability distributions of saccade amplitudes and orientations from our dataset. Since both methods require a saliency map as input for each web design, we use the *ground-truth* saliency map computed from the user eye gaze data of the web design. The saliency maps are constructed by convolving the users' fixation locations with a Gaussian kernel as in [Judd et al. 2009]. Moreover, for both methods, the first fixation points are selected randomly, according to a distribution of first fixation points learned from our eye gaze data for each web design.

Since both methods predict scanpaths instead of attention transition and order between page components as in our model, to evaluate their performance using the above metrics, we run each of the models 20 times on each web design to simulate the scanpaths of 20 different viewers. With the simulated eye gaze data for each web design, we compute the attention transition matrix and partial order matrix (as in Section 4) as predictions, which are compared against the ground-truth matrices of the webpage. Table 1 shows results of these methods. Our model outperforms both models on all the metrics. Moreover, as compared with our model, both models have difficulty in estimating the global visiting order of page components (very low POMR values). This is because both these methods only focus on modeling attention transition, whereas we train an independent model especially for predicting attention order, which leads to superior performance in the order prediction task.

7.7 Effects of the Energy Terms

We test the effect of each energy term in our optimization by generating a web design with each term omitted in turn. As shown in Figure 9, each term contributes to the final result.

8 Discussion

In this paper, we have proposed a novel technique for web designers to easily manipulate user attention behaviors on web designs. To this end, we have gathered and analyzed eye-tracking data from real-world users on webpages of various types, and proposed novel task-driven visual attention models, which predict user attention transition and order between page components from their low-level properties. Given an input web design and a designer-specified attention path over a set of page components, the input design is automatically adjusted via an optimization to match the specified and actual user attention behaviors. We envision that our method can be integrated into commercial web design editors to help designers create more effective web designs, and have potential to inspire future research for aiding in web design process. To encourage future works, we release our dataset and code at our website[2].

Limitations and future works. Our method is subject to several limitations and thus leave much room for further improvement. First, our method may fail to give satisfactory results for some extreme input paths. This is because in order to preserve the original layout style of the input web design our optimizer penalizes major modifications to the input web design (e.g., swapping the positions of two components), which are sometimes required to match an input path well. For example, as shown in Figure 10, if an input path starts with a component near the bottom of a page and then points towards a component at the top of the page, it will be difficult to generate a reasonable result without significantly changing the positions of the relevant components. To address this issue, we may consider exposing the weight of the regularization term to web designers, so that they could control the tradeoff between preserving their original layout and matching their specified paths. Second, our attention models only consider low-level features. However, web designers sometimes use high-level semantic cues in images (e.g., gaze direction and body poses) to guide user attention. Without access to the semantic information in the images, our model cannot properly deal with the influence of these semantic cues upon user attention. Incorporating image semantics into our models would further enhance the applicability of our method in practice, which is left as future work. Third, since the focus of this work is on directing user attention on a web design instead of improving its visual quality, we assume that input designs already have reasonably good layout. When input designs are not good to start with, our method might fail to generate visually satisfying results due to the use of limited design guidelines in the prior term of our optimization. A promising solution is to learn a visual quality model from examples, which can complement our user attention models towards a more sophisticated system to assist the web design process.

Acknowledgements

References

BOCCIGNONE, G., AND FERRARO, M. 2004. Modelling gaze shift

_____

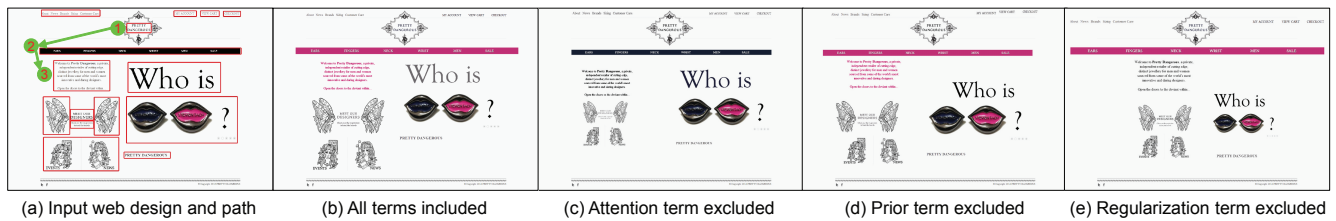[2]http://mmgroup.cs.cityu.edu.hk/projects/visualflow/

Figure 9: *The effect of each energy term on the optimized result. With the attention term excluded (c), the resulting design fails to guide user attention along the input path. With the prior term excluded (d), some visual design principles, e.g., alignment and balance, are violated. With the regularization term excluded (e), the layout style of the input design changes significantly.*
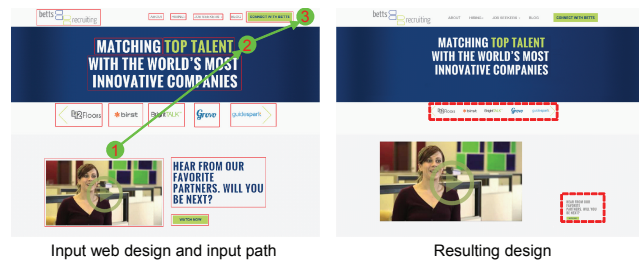

Figure 10: *Failure case: given a more extreme input path (left), our method may fail to generate a satisfactory result (right) by overly shrinking some of the components (red dash rectangles).*

as a constrained random walk. *Physica A: Statistical Mechanics and its Applications 331*, 12, 207–218.

BRADLEY, S., 2010. http://speckyboy.com/how-to-control-flow-within-your-web-designs/.

BRADLEY, S., 2013. http://vanseodesign.com/downloads/learn-design-fundamentals/.

BRADLEY, S., 2015. http://www.smashingmagazine.com/2015/04/design-principles-compositional-flow/.

BREIMAN, L. 2001. Random forests. *Machine Learning 45*, 1.

CAO, Y., CHAN, A., AND LAU, R. 2012. Automatic stylistic manga layout. In *ACM SIGGRAPH Asia 2012*.

CAO, Y., LAU, R., AND CHAN, A. 2014. Look over here: Attention-directing composition of manga elements. In *ACM SIGGRAPH 2014*.

ELLIS, S., AND SMITH, J. 1985. Patterns of statistical dependency in visual scanning. In *Eye Movements and Human Information Processing*, 221–238.

ERASLAN, S., YESILADA, Y., AND HARPER, S. 2015. Eye tracking scanpath analysis techniques on web pages: A survey, evaluation and comparison. *Journal of Eye Movement Research 9*, 1–19.

EYEQUANT, 2016. http://www.eyequant.com/.

FENG-GUI, 2016. https://feng-gui.com/.

GUY, T., 2011. http://uxmag.com/articles/visual-design-and-usability-yellow-brick-road.

HASTINGS, W. 1970. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika 57*, 97–109.

ITTI, L., AND KOCH, C. 2000. A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research 40*, 1489–1506.

JONES, B., 2011. http://webdesign.tutsplus.com/articles/understanding-visual-hierarchy-in-web-design.

JUDD, T., EHINGER, K., DURAND, F., AND TORRALBA, A. 2009. learning to predict where humans look. In *ICCV 2009*.

KUMAR, R., TALTON, J., AHMAD, S., AND KLEMMER, S. 2011. Bricolage: Example-based retargeting for web design. In *ACM SIGCHI 2011*.

KUMAR, R., SATYANARAYAN, A., TORRES, C., LIM, M., AHMAD, S., KLEMMER, S., AND TALTON, J. 2013. Webzeitgeist: Design mining the web. In *ACM SIGCHI 2013*.

LEMEUR, O., AND COUTROT, A. 2016. Introducing context-dependent and spatially-variant viewing biases in saccadic models. *Vision Research 121*, 7284.

LIU, H., XU, D., HUANG, Q., LI, W., XU, M., AND LIN, S. 2013. Semantically-based human scanpath estimation with hmms. In *CVPR*, 3232–3239.

MERRELL, P., SCHKUFZA, E., AND KOLTUN, V. 2010. Computer-generated residential building layouts. In *ACM SIGGRAPH Asia 2010*.

METROPOLIS, N., ROSENBLUTH, A., ROSENBLUTH, M., TELLER, A., AND TELLER, E. 1953. Equation of state calculations by fast computing machines. *Journal of Chemical Physics 21*, 1087–1092.

NAJEMNIK, J., AND GEISLER, W. 2009. Simple summation rule for optimal fixation selection in visual search. *Vision Research 49*, 10, 1286–1294.

NIELSEN, J., AND PERNICE, K. 2009. *Eyetracking Web Usability*. New Riders Publishing.

O'DONOVAN, P., AGARWALA, A., AND HERTZMANN, A. 2011. Color compatibility from large datasets. In *ACM SIGGRAPH 2011*.

O'DONOVAN, P., AGARWALA, A., AND HERTZMANN, A. 2014. Learning Layouts for Single-Page Graphic Designs. *IEEE TVCG 20*, 8, 1200–1213.

SCHTZ, A., BRAUN, D., AND GEGENFURTNER, K. 2011. Eye movements and perception: A selective review. *Journal of Vision 11*, 1–30.

SHEN, C., AND ZHAO, Q. 2014. Webpage saliency. In *ECCV2014*.

WANG, W., CHEN, C., WANG, Y., JIANG, T., FANG, F., AND YAO, Y. 2011. Simulating human saccadic scanpaths on natural images. In *CVPR*, 441–448.

YU, L., YEUNG, S., TANG, C., TERZOPOULOS, D., CHAN, T., AND OSHER, S. 2011. Make it home: automatic optimization of furniture arrangement. In *ACM SIGGRAPH 2011*.