# predicting webpage aesthetics with heatmap entropy

Abstracts

Today eye trackers are extensively used in user interface evaluations. However, it’s difficult to analyze and interpret eye tracking data from an aesthetic point of view. To find quantitative links between eye movements and aesthetic experience, we tracked 30 observers’ eye movements during initial landings of 40 webpages (each page for 3 seconds). The webpages were also rated based on the observers’ aesthetic judgments.

The Shannon entropy was used to analyze the eye tracking data. The result shows that the heat map entropy (visual attention entropy, VAE) is highly correlated with the observers’ aesthetic judgements of the webpages. VAE is interpreted as the fluency or chaos in observers’ attentional processes. Its improved version, relative VAE (rVAE), has more significant correlation with the perceived aesthetics. (r=-0,65; F= 26.84,P=0.000007). This single metric alone can distinguish between good and bad looking pages with about 87.5% correctness. Further investigation reveals that the performance of both VAE and rVAE became stable after 1 second. The curves indicate that their performances could be better if the tracking time were extended slightly beyond 3 seconds.

Key words: entropy , visual attention , eye tracking , aesthetics, web page

### Introduction

Aesthetics are important factors for engaging users in browsing a website. Big data reveal that most visitors will determine whether to stay on or leave a webpage in just a few seconds after landing [ ]. Appealing webpages feel more trustworthy to visitors [ ]. Beautiful webpages are seemingly more usable [ ].

Aesthetic is rooted in the Greek term for "to perceive." Our vision system, the eye and the connected cortex behind, enables us to perceive and cognize the beauty around us. An eye behaves like a spotlight [  ]: it has a narrow but very high-resolution foveal vision surrounded with a broader but much lower resolution peripheral area. visual attention keeps selecting where our spotlight moves, targeting things like faces, words, images on screens, and a variety of other meaningful objects, bringing salient details into the focus and filtering out background clutter. The spotlight keeps glancing from one spot to another and piecing all snapshots together to give an overall impression of the scene to our brain.

Eye tracking devices are designed for recording fixations (gaze) and saccades (eye movements). In usability and marketing research, eye tracking has been used in evaluating the accessibility of key information. Though useful, this is still a fairly limited measure of user experience [].

Is it possible to use eye trackers to anticipate observers’ aesthetic satisfaction with a webpage? This paper reports our efforts in answering this question. The results are positive with evidences that eye tracking devices do enable us some new sense of beauty in the eye of the beholder.

Thirty observers’ visual explorations of 40 webpages were recorded individually using an eye tracker in a normal lab setting. Each subject stared at each page for 3000ms. The webpages were later rated by the observers based on their personal aesthetic judgments. Shannon entropy was introduced to analyze the eye tracking data. The result shows that the entropy in a heat map (attention map), which visualizes gaze distribution, is highly correlated with the observers’ aesthetic judgements of the webpages (r=-0,65; F= 26.84,P=0.000007). This single metric alone can differentiate good or bad looking web pages to a certain degree (about 87.5%). The entropy can be interpreted as an objective and quantitative metric of chaos in our attentional processes. Therefore it could be the direct evidence to the fluency hypothesis (Reber et al), which proposes that beauty is grounded in the processing experience of the perceiver. The more fluently perceivers can process an object, the more positive their aesthetic response.

### Related work

1\ Eye tracking and experimental aesthetics

It is a natural thought that our eye’s attentional responses may somehow reflect our sensory contemplation or appreciation of an object.

To examine how eye movements relate to judging aesthetically pleasing works of art, Berlyne (1971) concluded that aesthetic evaluations are based on two types of visual exploration, one is global diverse exploration and the other is local specific information gathering. The diverse exploration is characterized by long saccades and short fixation (gaze) durations, while specific information gathering has longer fixation durations and shorter saccades. Berlyne proposed that the pattern of exploration, featured by the oscillation of the durations and saccade lengths, is crucial for judging images as aesthetically pleasing or not. Berlyne's concept of visual exploration was influential to the following studies. Locher (2006) demonstrated that simply altering the color-balance of the original abstract compositions will change the distribution of gaze and scan path. Franke et al. 2008 observed an increased number of eye fixations and longer durations on more preferable 3-D renderings. Plumhoff 2009 observed that with pleasant graphics, fixation durations increase and saccade lengths fluctuate more over the viewing period. Wallraven et al. 2009, analyzed eye tracking data from 20 participants who looked at 275 artworks from different periods of styles, and found a strong effect of art period (style) on both the number and duration of fixations. Massaro et al 2012 use categorized (color, grey, human and nature) art paintings as test material to investigate the contributions of bottom-up and top-down processes of visual attention to eye-scanning on the art paintings. Khalighy et.al 2015 developed an empirical aesthetic formula by conducting three eye-tracking experiments on simple geometric forms of stimuli. He induced that beauty is positively related to the product of the number of fixations and standard deviation of duration.

difference in the distribution of the gaze existed between the three versions of each composi-

tioempirical evidence that affected the way viewers attention and cognitive processes that occur during an aesthetic episode visually

The researches above revealed the potential of eye tracking for various objects of aesthetic studies. The results so far, however, do not go much beyond Berlyne’s idea. Some reported observations, such as the increased number of fixations, longer duration and oscillation of saccade lengths, etc, only confirm Berlyne’s idea that eye behaviors on aesthetically pleasing objects are more active and dynamic. It is indeed hard to interpret those results from an aesthetic point of view. Although eye trackers have been used as a new apparatus for experimental aesthetics, there is little research that gives us a convincing in-depth interpretation about how our eyes’ behavior is connected to our aesthetic responses.

2\Eye tracking and webpage aesthetics

Research on webpage aesthetics is mainly focusing on finding predictive features of webpages(e.g., the complexity[]) . Some progress has made in this direction. Ivory and Sinha [ ] Zheng et al. [ ] reneick etal [ ]use low level image features to predict perceived visual complexity and colorfulness, and they found the two models are largely correlated to people’s aesthetic preferences.

Eye trackers have been extensively used in webpage evaluations for visualizing accessibility or interest bias towards specific objects. The existing visualization and analysis tools have nothing to do with the user experiences of aesthetics. Proven quantitative links between eye movements and aesthetics are still lacking [ ]. Can eye tracking data provide a more general means of quantifying a viewer’s aesthetic experience? It may need more interpretable indexes formulated from an aesthetic perspective and some data mining skills.

# Hypothesis

Our idea is derived from the fluency hypothesis. Reber et al propose that beauty is grounded in perceivers’ fluency of processing an object.

Viewing a webpage is essentially a procedure of image processing in the human visual system.

How does one evaluate the fluency of visual processing？We could start from the entrance of the processing pipeline – attention, which serves as a precursor to all other neurological and cognitive functions.

At its most basic, attention is defined as the process by which we select a subset from all of the available information for further processing [ ]. Attention keeps alerting with peripheral vision and selecting targets in both bottom up (visual saliences) and top down (attractive contents) ways.

If the fluency hypothesis is true, the attentional process must be also influential to our aesthetic judgements. Then, how to measure the fluency of visual attention?

Reasonably a fluent attentional process may have:

1\ Less conflict and more easy in competitive selection, which is the process determining which visual object gains access to working memory, between multiple attention cues.

2\ Less distraction and more easy in orienting and localizing attention onto an interesting object.

This paper adopts entropy as the quantitative metric to the fluency/chaos in our visual attention behaviors and further inspects whether it correlated with our aesthetic judgements or not. Several existing indexes mentioned will also be investigated.

The Entropy measure will be applied on two aspects of the eye tracking data:

Entropy in gaze plot (*Saccade sequence*) represents how complex and uncertain the temporal order of attentional shifts among multi-objects.

Entropy in heat map *(Fixation distribution*) represents how noisy and deconcentrating the spatial allocation of attention on those objects.

## Experiment and data collection：

The experiment simulates natural and accidental landings of websites. The only unnatural part of the experiment is the form we called Time-limited Exploration. That means the experiment forces/encourages the observers to do quick general explorations rather than specific information gatherings on the pages. After three pages of adaptive testing in advance, the observers should aware that they only have a few seconds to view each of the pages displayed. This is reasonable as we already knew, users are used to making a quick decision whether stay or leave in first seconds after landing a new page, to avoid wasting more time than necessary on bad pages [ ].

The restricted exposure time could be enough to allow us to examine the early unintentional gazing behavior, meanwhile avoids the fatigue of the observer.

Equipment

The eye tracker we used is a Tobii T50 that works with Tobii Studio 9 software. The eye tracker resolution is 1280 \* 1024 pixel.

Materials

We selected snapshots of 40 websites’ main pages as the testing material. To make the webpages more representative of good and bad aesthetics, we sampled pages from（\*）and a site where users can vote for ugly web designs - [www.websitesfromhell.com](http://www.websitesfromhell.com/), and only selected the odd number pages in the collections.

Pages that contained non-Latin characters or well known logos and faces were removed, to avoided interference from non-aesthetics based social opinions. Finally, we narrowed our selection to 20 webpages from each site. All pages were snapshotted in landscape form with 1280\*800 pixels resolution to match the window of the eye tracker. The snapshot includes the frame of the web browser. Figure xx is all the pages used as the stimuli.

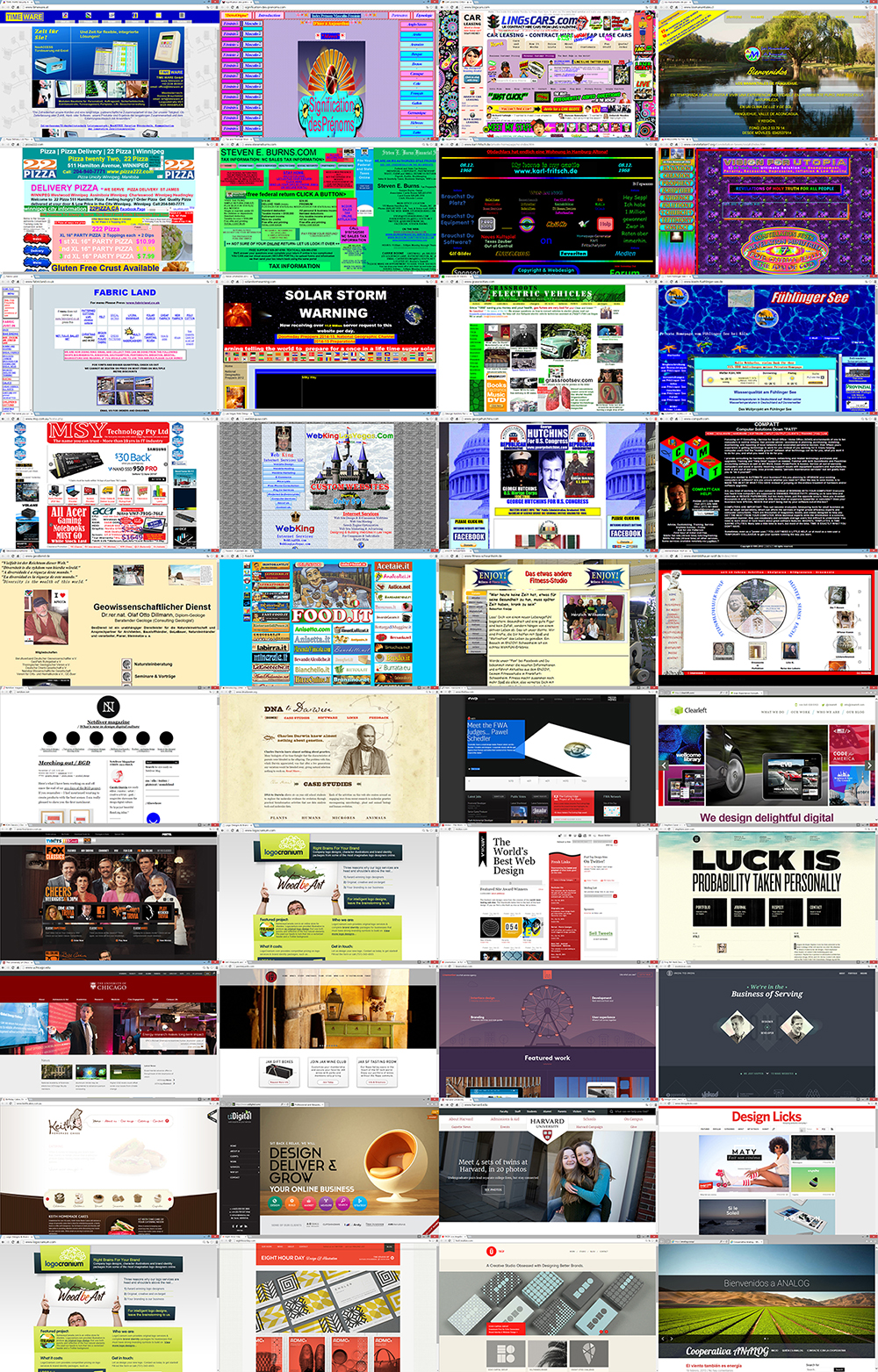


Figure 1 All the pages used in our eye tracking test are arranged from top to bottom according to their scores, from low to high.

Subjects

A total of 30 subjects (13 men, 17 women) participated in the eye tracking experiment. They are students or teachers from a number of departments in Shanghai Jiao Tong University. Some of them are foreign students from South Korea. Their ages are between 19 and 27 years old.

Site setting

The experiment was conducted in a quiet room. The curtains of the room were pulled up to avoid uncontrollable light and reflections on the screen. The eye tracker was placed in front of a pure white wall to avoid possible distractions.

Procedure

Each subject was instructed to browse these pages just as they would when mindlessly browsing the web. He/she sat in a working chair and was asked to lean forward to rest his/her chin comfortably on a soft high support to fix his/her head at a 60 inch distance from the eye tracker screen. The elbows were rested on the table with a mouse in one hand, even though clicking and scrolling did not produce any response.

In order to reduce fatigue, the test was divided into two phases with a minute of rest in between. In each phase, the screen of the tracker automatically displayed half of the total 40 experimental pages in a random order. Each page was only displayed for 3 seconds, followed with 1 second of a black screen. Three dummy pages were placed at the beginning of the first phase as the adaptive test to make them feel comfortable with the rhythm of the test. During the eye tracking, there was no interaction between the operator and the subjects.

The subjective ratings of the pages were arranged separately after the experiment. The subjects reviewed the 40 experimental pages once again with no time limit. The subjects were asked to judge the aesthetic quality of "good" or "bad" for each page. And all subjects confirmed the pages they saw in the test had never been seen before.

Data collection：

Each page in the test gets a final score based on how many “good” or “bad” ratings it received. The figure xx is the distribution of the 40 pages on the scores.

The distribution is apparently bimodal. The 20 pages with score less than 0.5 are exactly the same pages collected from the www. In other words, these 20 pages are recognized as "bad" pages without any doubt.

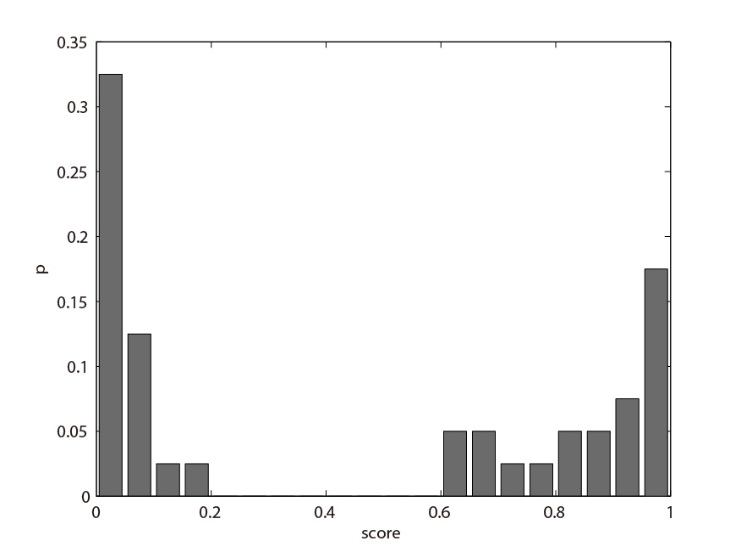


Figure 2. the distribution of page on scores. There are clearly two groups. The bad pages are generally more definite, while the good pages are more evenly distributed.

T50 collects the eye focus on the screen at a frequency of 50Hz, and through interpolation it is possible to obtain the speed of eye movement at all times. Then to estimate where the fixation gaze happens (if speed==0).

The raw data produced by the eye tracker consisted of a series of gazes and fixations. Each fixation contains four parameters: the start time, the duration, position X and Y on the screen. The following analysis is just based on this format of eye-tracking data.

## - Analysis and results

## 3.1 Traditional statistical indexes

Firstly, we tested some traditional descriptive statistical indexes on the collected eye tracking data. Most of them have been used in the aesthetics research mentioned in section 2.

The indexes we use include:

num of fixation, mean of duration, std of duration, mean of saccade length, std of saccade length, num of AOI, mean of AOI fixNum, std of AOI fixNum

Areas of Interests (AOIs) are spatial clusters of the fixations on the pages, we uses the default AOI clustering algorithm in the Tobii Studio package. Figure xx is an AOI example obtained by the clustering.

|  |  |  |
| --- | --- | --- |
| LINK Excel.SheetBinaryMacroEnabled.12 "C:\\Users\\Administrator\\AppData\\LocalLow\\WINZIP\_P1a19\\3.1\_传统指标与美感的相关性.csv" "3.1\_传统指标与美感的相关性!R1C1:R9C3" \a \f 5 \h \\* MERGEFORMATX | pearson correlation | ANOVA  p-value |
| number of fixations | 0.3322 | 0.0429 |
| mean of durations | -0.1526 | 0.327 |
| std of durations | 0.1165 | 0.6219 |
| mean of saccade length | -0.2203 | 0.0871 |
| std of saccade length | -0.1933 | 0.2732 |
| number of AOIs | -0.2607 | 0.1631 |
| mean of AOI fixNum | 0.3228 | 0.0782 |
| std of AOI fixNum | 0.2556 | 0.0991 |

Table 1 the performances of some traditional statistics of fixations, saccades and AOIs. Only the number of fixations has a weak correlation with the aesthetic scores r=0.33. and the ANOVA is significant. P=0.04

The results of the analysis are listed in table xxx. Each index has a linear correlation value with the aesthetic scores, and a p-value of anova test separating the good and bad two classes.

The results show that only number of fixations barely passed the anova test. P<0.05, all other indexes failed in the anova test. The number of fixations has a weak positive correlation with the page aesthetics. In other words, the nice pages generally get more gazes, implying the eyes are more active. Meanwhile, the number of AOI has a weak negative correlation with the page aesthetics, meaning the aesthetically pleasing page’s fixations are concentrated on fewer AOI. It may also imply subjects are more swiftly entering the stage of specific information gathering [ ]

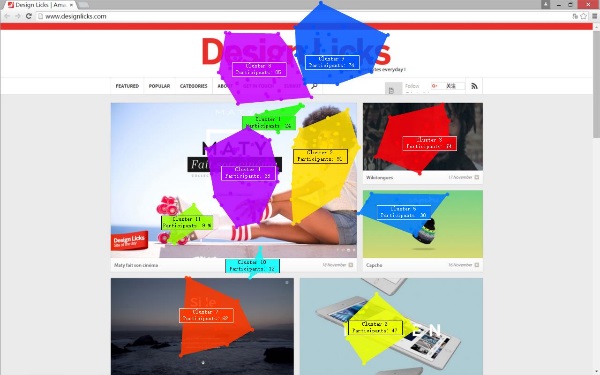
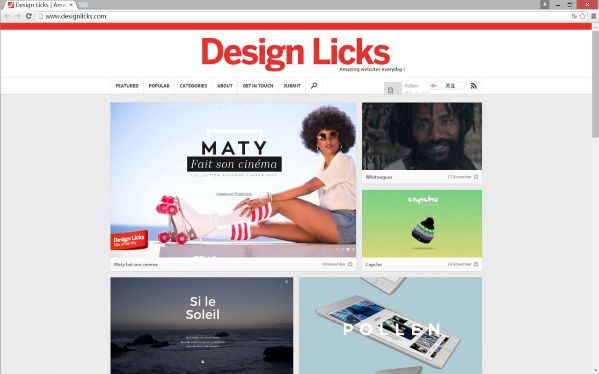


Figure 3. left is one of the testing images. The right is the image overlapped with AOIs, which are generated using default clustering tool in Tobii studio. There are total 11 AOIs.

## 3.2 Entropies in eye movements

The existing definition of entropy includes thermodynamics and information theory, conceptual, entropy represents a system’s chaos, degree of confusion, or vice versa, order and certainty.

The concept of visual entropy actually expresses the determinism of peoples’ eye movements in spatial and temporal dimensions. A smaller visual entropy may indicate a more fluent and concentrated visual process.

The visual entropies defined in this paper are based on the Shannon entropy of information. The entropy is a measure of the uncertainty based on the probabilities of all possible events.

For a discrete random variable X with probability spaces Ω = {x1, ..., xn} and probability distribution P (X). The information entropy H (the Greek letter eta) is defined as H (X) = -ΣP (xi) log2 P(xi). The higher the degree of H (X), the higher the degree of chaos on the distribution of X and the greater the amount of uncertainty, whereas the smaller the H (X), the more ordered the distribution of X and the smaller the amount of uncertainty.

Now we analyze the visual entropies from two aspects of the eye tracking data: large saccades (transitions) among AOIs and fixation distribution

## 3.2.2 Entropy in gaze transitions

Entropy of eye movements was first introduced in a paper by Tole et al., (1983) as a measure of cognitive loads, describing the changes of looking behaviors of eleven pilots in the conditions of different tasks. The model was also adopted in analysis of eye tracking data of car driving [Jess Gilland]

Hooge [ ] used a similar metric of “Scan path entropy” to quantify gaze guidance to a specified target.

This paper follows the Tole’s definition of visual entropy based on Markov chain models.

The entropy represents the uncertainty of the transition between the fixations. As we knew, the number of possible positions of fixations is 1024\*800, equal to the resolution of the images. That is a huge space of states. To estimate the probabilities of the transitions among those states need a very big sample set. That is impossible to get via eye tracking tests. Tole’s solution was to subdivide the view into a number of areas. In this research, we take the advantage of the system-generated AOI. Then only the inter AOI transitions are taken into account. Inter AOI transitions are relative large saccades. A transition is thought to be an attention driven mechanism that requires a slight but unconscious attentional shift from the current target to peripheral target acquisition (Henderson, 1993). That is just the meaning of global exploration suggested by Berlyne (1971). By AOI clustering, we can convert each observer’s eye movements on a page into a series of AOI jump sequences.(see table xx)

|  |  |
| --- | --- |
| LINK Excel.SheetBinaryMacroEnabled.12 "C:\\Users\\Administrator\\AppData\\LocalLow\\WINZIP\_Pae13\\pic17\_AOIseqs.csv" "pic17\_AOIseqs!R1C1:R31C1" \a \f 4 \h \\* MERGEFORMATX6 ->7 ->11 ->3 ->11 ->10 ->9 ->2 ->3->4 |  |
| 7 ->3 ->4 ->8 ->6 ->4 ->3 ->2 |  |
| 3 ->5 ->9 |  |
| 7 ->4 ->8 ->11 ->3 ->7 ->11 |  |
| 7 ->4 ->8 ->11 ->4 ->7 ->2 ->9 |  |
| 7 ->3 ->7 ->4 ->8 ->11 ->9 ->2 ->9 |  |
| 7 ->8 ->4 ->7 ->11 ->2 ->9 ->2 ->3 |  |
| 3 ->7 ->11 ->10 ->9 ->2 ->9 ->4 |  |
| 7 ->11 ->2 ->6 ->7 ->8 |  |
| 4 ->11 ->5 ->2 |  |
| 7 ->3 ->11 ->4 ->8 ->7 ->2 |  |
| 7 ->3 ->4 ->8 ->3 ->2 ->9 |  |
| 6 ->3 ->4 ->8 ->3 ->10 |  |
| 7 ->3 ->1 ->10 ->11 ->7 |  |
| 7 ->11 ->8 ->4 ->7 ->8 ->9 ->2 ->6 ->2 |  |
| 7 ->4 ->8 ->5 ->9 ->2 ->9 |  |
| 11 ->7 ->11 ->10 |  |
| 7 ->3 ->4 ->8 ->11 ->10 ->9 ->2 |  |
| 7 ->11 ->10 ->2 ->9 ->4 ->8 |  |
| 7 ->4 ->8 ->7 ->3 ->1 ->3 |  |
| 7 ->11 ->10 ->2 ->9 |  |
| 7 ->3 ->11 ->10 ->2 |  |
| 3 ->4 ->1 ->3 ->8 ->4 ->11 |  |
| 7 ->11 ->4 ->8 ->3 ->9 ->2 ->9 |  |
| 7 ->4 ->8 ->3 ->2 ->9 ->8 ->4 ->11 |  |
| 7 ->4 ->8 ->11 ->7 ->3 ->2 |  |
| 7 ->3 ->4 ->8 ->11 ->7 |  |
| 6 ->7 ->8 ->4 ->7 ->8 ->11 ->4 ->2 |  |
| 7 ->11 ->7 ->3 ->11 ->8 ->4 ->8 ->7 |  |
| 3 ->7 ->4 ->8 |  |
|  |  |

Table 2. 30 subjects’ gaze transitions among the 11 AOIs on the image (see figure 3)

The Markov nature of eye movements

If a discrete time series {x0, x1, ..., xt} satisfies P (xt + 1 | xt) = P (xt + 1 | xt, xt-1, xt-2, ..., x0), then the time series is called the Markov chain. The Markov chain of AOI sequences could be simply explained that which AOI we glance toward is related only to which AOI we have just gazed upon.

For the 1st order Markov chain, the total transitional probability information can be summarized by a one-step transition probability matrix.

In the experiment, the matrix of probability is estimated by counting all the occurrences of the transitions between the AOIs on the page. The figure xx is the probability matrix obtained from the eye movement data of the example image in fugure xx:



Table 3 Markov transition probability matrix of the 11 AOIs

In the Matrix, pij represents the probability of transition from AOI i to AOI j.

Then, a visual entropy based on the 1st order markov transition matrix are calculated by the formula:

【引文】

【Σ(Σp(Y|X))p(X)/Hmax】

Where p (Y | X) = pxy represents the conditional probability of transition from X to Y. p (X) represents the prior probability of X, that is, the probability of starting from X, which is obtained by counting the frequency of occurrence of X in all AOI sequences on the statistics page.

Hmax is the maximum entropy of the current AOI transitions, and the maximum entropy is obtained by supposing all transition probabilities are equal and all the prior probabilities are equal. By dividing the Hmax, the visual entropies of different pages with different numbers of AOIs are comparable to each other.

Then entropy could inform us the certainty of the attentional shifts between AOI. We expected that the entropy of a good page would have a smaller value. However, the result was disappointing. Based on the 3s eye tracking data, The correlation between the entropies and the aesthetic scores was 0.1585, and the two classes’ anova P-value is 0.4741.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Source | SS | df | MS | F | Prob>F |
| Groups | 0.00166 | 1 | 0.00166 | 0.52 | 0.4741 |
| Error | 0.12101 | 38 | 0.00318 |  |  |
| Total | 0.12268 | 39 |  |  |  |

table 4. the Anova analysis of Visual Entropy based on the gaze transitions

The failure of the entropy based on AOI markov chain might be caused by the following two reasons:

The 1st order markov chain might be an over-simplified model. However, increasing the order of the chain requires an exponentially bigger data set.

Relevant information was lost and some errors was introduced in the clustering and the default clustering algorithm in tobii studio was not smart enough to avoid some wrong clusters.

## 3.2.1 Entropy in heat map

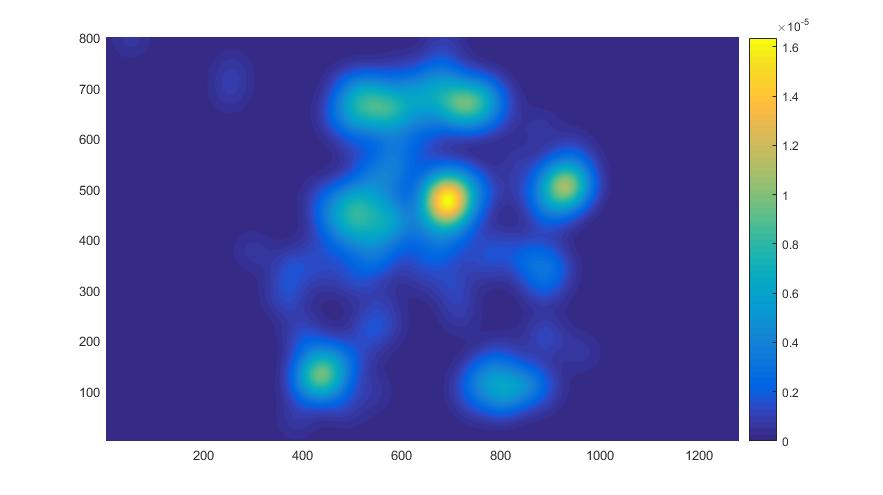
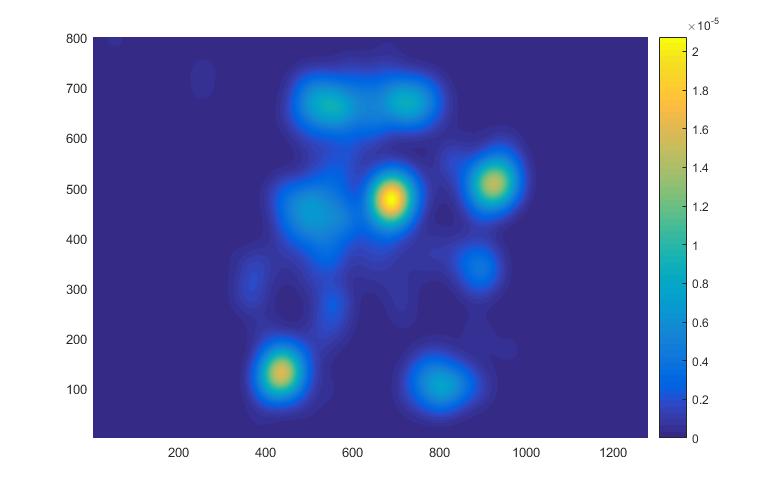


Figure 4 the heat map of the example pages in figure 3. The left is duration-weighted heat map. The right is weightless heat map. You can see the brightness of the dots are slightly different. The two maps use Gaussian interpolation (σ=30px). Weightless means that all the kernels have the same weight.

Entropy based on Markov chain regards eye movements as a serially-related events of eye fixations and saccades.

Here we introduce a new entropy metric based on the heat map, which is the best-known visualization technique for eye tracking studies.[[35]](https://en.wikipedia.org/wiki/Eye_tracking" \l "cite_note-35) In contrast to the gaze chain, heat map has no information about the order of fixations. That means the new entropy metric assumes all the gazing events are independent.

Heat map visualizes spatial probability distribution of fixations on a page.

Considering a two-dimensional random variable Z = (X, Y) which represents the position of a fixation, the resolution of the screen constitutes its probability space (ie, each pixel is a possible position for a fixation).

Based on the probability distribution P (X, Y), The visual entropy are defined as

H (Z) = Σ (-P (xi, yi) \* log (P (xi, yi)), for all (x, y) st P (x, y)! = 0 ,ΣP(Xi, Yi) = 1

The probability distribution of Z in the XY space, which is actually the heat map, is posterior and can be estimated by the eye tracking data.

The entropy of the gazemap reflects the consistency or agglomeration of the subjects’ eye movements in the space: when all the gazes concentrate on a single pixel, it obtains the smallest value, and when the gazes are evenly scattered on the screen, it gets the maximum value.

To estimate the P(X,Y) with a small sample of fixations, an important technique is Gaussian Mixture, which interpolates all the fixations by placing a Gaussian kernel at each observed fixations.

The expression of the mixture model is

Σ(-1/2/r/r\*e^(-((x-xi)^2+(y-yi)^2)/2/r/r)\* ti)/ Σti

Where σ is the standard deviation of the kernel，weight ti is the duration of the ith fixation xi yi，

the model will be normalized to make sure the total probability P(X,Y)=1.

Placing a proper size Gaussian at a fixation is also reasonable given the fact that the fixation is actually a round area corresponding to the fovea. In addition, the fixation in reality is not absolutely fixed due to the fixational movements []. The mixture model also tolerates the unavoidable measure error of the eye tracker. The figure xx is a typical heat map with σ =30px

The visual entropies of 40 pages based on their heat map were turned out to be negatively correlated with their aesthetic scores. Pearson correlation r= -0.54891. Anova f=xxx p = xxxx. The results are pretty significant.

It is interesting that the weightless heat map (see figure xx), a heat map without the duration information, has even better performance than, or at least as good as the duration based heat map. Pearson correlation is -xxxx. Anova F=16.6 with p = 0.0002. This fact may imply that the duration information has no contribution to predicting the aesthetic scores. It works like a kind of random perturbation. It is not so strange considering the fact that the duration can range greatly by individual (Green, 2002; Irwin, 1996; Zwahlen & Schnell, 1998) from less than 100ms to well over a second depending on the situation (Harris et al, 1988).

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Source | SS | df | MS | F | Prob>F |
| Groups | 1.15861 | 1 | 1.15861 | 15.79 | 0.0003 |
| Error | 2.78901 | 38 | 0.07339 |  |  |
| Total | 3.94762 | 39 |  |  |  |

Table xx the ANOVA analysis of the entropy of duration weighted heat map

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Source | SS | df | MS | F | Prob>F |
| Groups | 1.01285 | 1 | 1.01285 | 16.6 | 0.0002 |
| Error | 2.31802 | 38 | 0.061 |  |  |
| Total | 3.33087 | 39 |  |  |  |

Table xx the ANOVA analysis of the entropy of weightless heat map

Present results show that the visual entropy based on heat maps is the most promising index to predict a web page’s aesthetic score. Heat maps are sometimes also referred to as attention map. So we name this metric Visual Attention Entropy (VAE).

## 3.3.3

## Relative VAE

There is a defect with using VAE as metric. VAE is negatively correlated with the aesthetic judgements. However, an absolutely lower VAE does not necessarily mean a very high aesthetic quality. For instance, a web page with very little content may result in a very low VAE while a web page containing many objects may have a relatively high VAE. This certainly does not mean the former is better than the latter. Theoretically, the VAEs of pages with different contents are not comparable to each other.

To solve this problem, we introduced a concept called Relative VAE (R-VAE). To compare VAEs of different webpages, it is necessary to take into account their different Base VAE (B-VAE), which is supposed to be the noise free VAE of a page. It represents the required attention resources of the page . The B-VAE on the page is estimated by averaging the individual VAEs of all subjects who viewed the page.

B(H(Z)) =ΣH i (Z)/n, i= 1, 2,…n

Where H i (Z) is the individual VAE measured on ith subject.

With the Base VAE as a precondition, now we get the relative VAE:

R(H(Z)) = H(X)/ B(H(Z)).

公式检查一下

Compared to VAE, The performance of Relative VAE was significantly improved, the Pearson Correlation –are increased. from r= 0.55 to r= 0.65. ANOVA F=25.88 with p-value =0.00001

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Source | SS | df | MS | F | Prob>F |
| Groups | 0.00295 | 1 | 0.00295 | 25.88 | 1.01E-05 |
| Error | 0.00433 | 38 | 0.00011 |  |  |
| Total | 0.00727 | 39 |  |  |  |

Table xx Relative VAE ANOVA (weightless)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Source | SS | df | MS | F | Prob>F |
| Groups | 0.00336 | 1 | 0.00336 | 26.84 | 7.53E-06 |
| Error | 0.00475 | 38 | 0.00013 |  |  |
| Total | 0.00811 | 39 |  |  |  |

table xx Relative VAE ANOVA (duration-weighted)

The formula xx indicates that the higher the Base-VAE, the lower the Relative-VAE, therefore benefits the aesthetics. That’s not fully true, however, and in reality the higher Base-VAE may result in possibly a much higher VAE, since the two indexes are significantly correlated (r=0.7). see table xx

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| LINK Excel.SheetBinaryMacroEnabled.12 "C:\\Users\\Administrator\\AppData\\LocalLow\\WINZIP\_Pce65\\3.4\_重要指标相关系数表.csv" "3.4\_重要指标相关系数表!R1C1:R7C7" \a \f 5 \h \\* MERGEFORMATX | aesthetic score | number of fixations | VAE | Base-VAE | relative  -VAE |
| aesthetic score | 1 | 0.332174 | -0.54891 | -0.11156 | -0.64971 |
| number of fixations |  | 1 | 0.039015 | 0.591096 | -0.24901 |
| VAE |  |  | 1 | 0.702389 | 0.933015 |
| Base-VAE |  |  |  | 1 | 0.399264 |
| relative-VAE |  |  |  |  | 1 |

Table xx: the Pearson correlations in-between the main indexes. All VAEs are weightless.

On other hand, generally a lower Base-VAE corresponds to a lower VAE, but this does not mean a lower Relative-VAE either. The analysis shows that the aesthetic score is weakly correlated with the Number of Fixations (r=0.33). So the good pages generally have more fixations (interesting points). And Base-VAE is also positively correlated with Number of Fixations (r=0.59). This implies that the good web pages generally are not based on lower Base-VAEs. Those facts may also explain why Base VAE has ignorable linear correlation with the aesthetic score (r=-0.11). Figure xx shows Base-VAE has no difference between the good and bad pages.

It seems that a moderate Base-VAE is better. A moderate Base VAE means a proper amount of specific information for the viewer to perceive. It also represents the required attention resource needed for the visual exploration in the viewport.

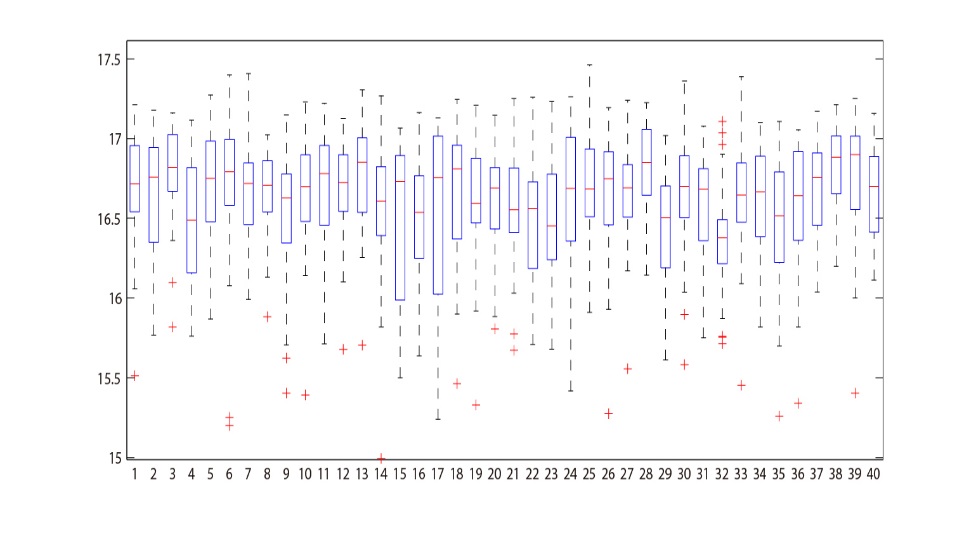


Figure xx the box diagram of 30 individual VAEs on the 40 webpages. The red lines in the boxes represent the Base-VAEs (means).

The introduction of Base-VAE enables us to compare VAEs of pages with different complexities of contents, and to get more precise predictions to their aesthetics.

【公式】

The figure xx is the ANOVA box diagrams of VAE and Relative-VAE. Figure xx is the scatter plots 40 webpages in the planes of score and two versions of VAE. Compared to the VAE, the Relative-VAE is more correlated with the aesthetic score.

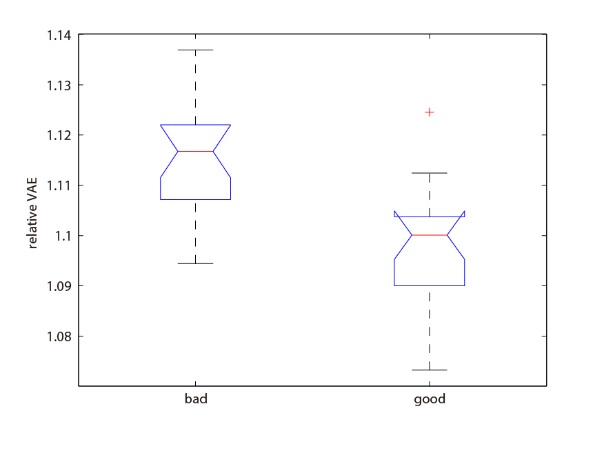
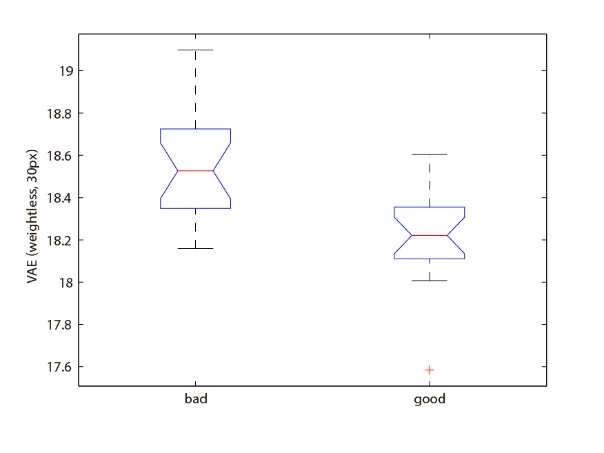


Figure xx. the box diagram of VAE (left) and R-VAE (right).both are weightless. R-VAE is able to better separate the two classes.

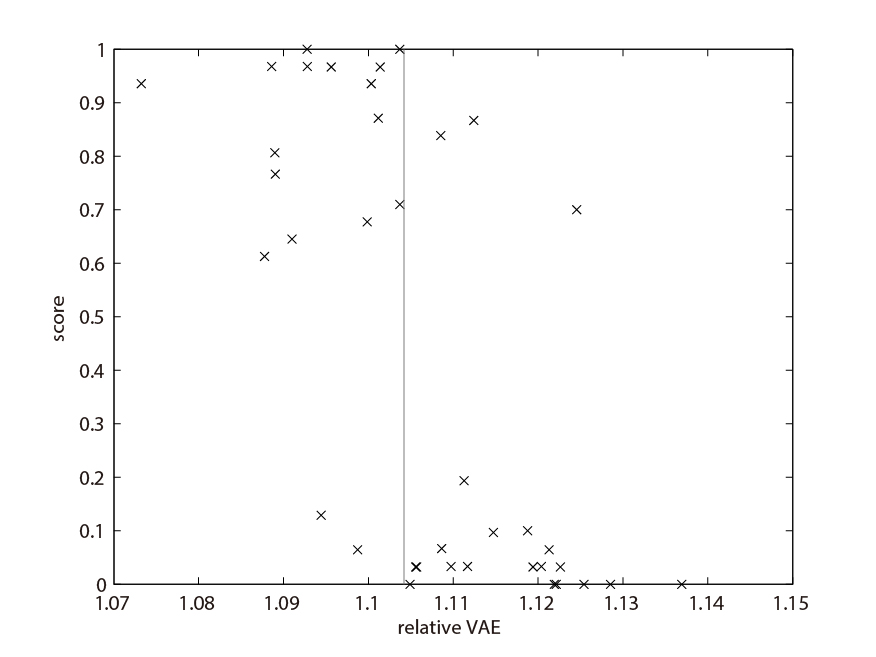
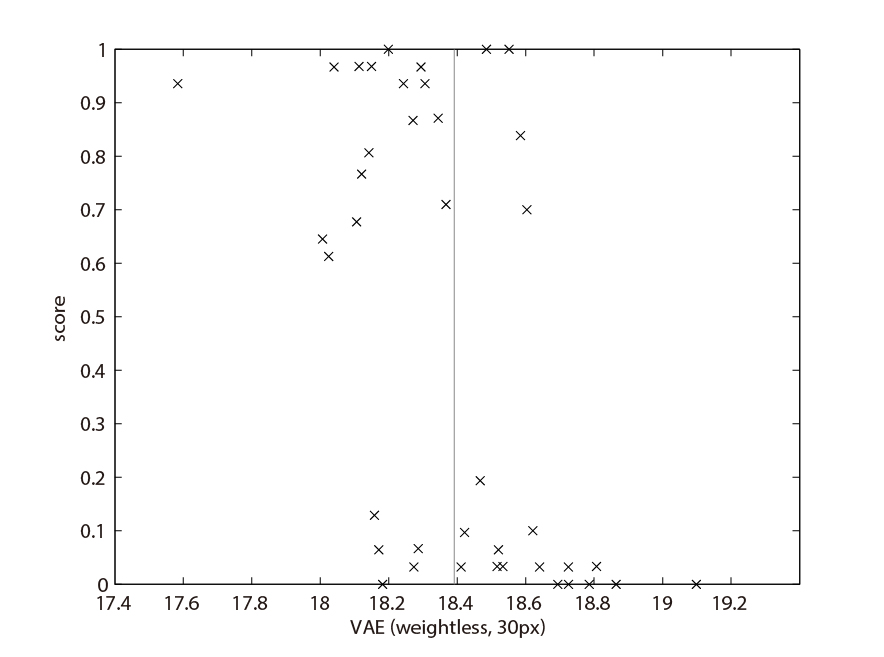


Figure xx. 40 pages scatter over VAE (top) and R-VAE (bottom) axis. Both VAEs are weightless. Apparently. The point cloud in the bottom one is more significantly skewed.

## Outlier

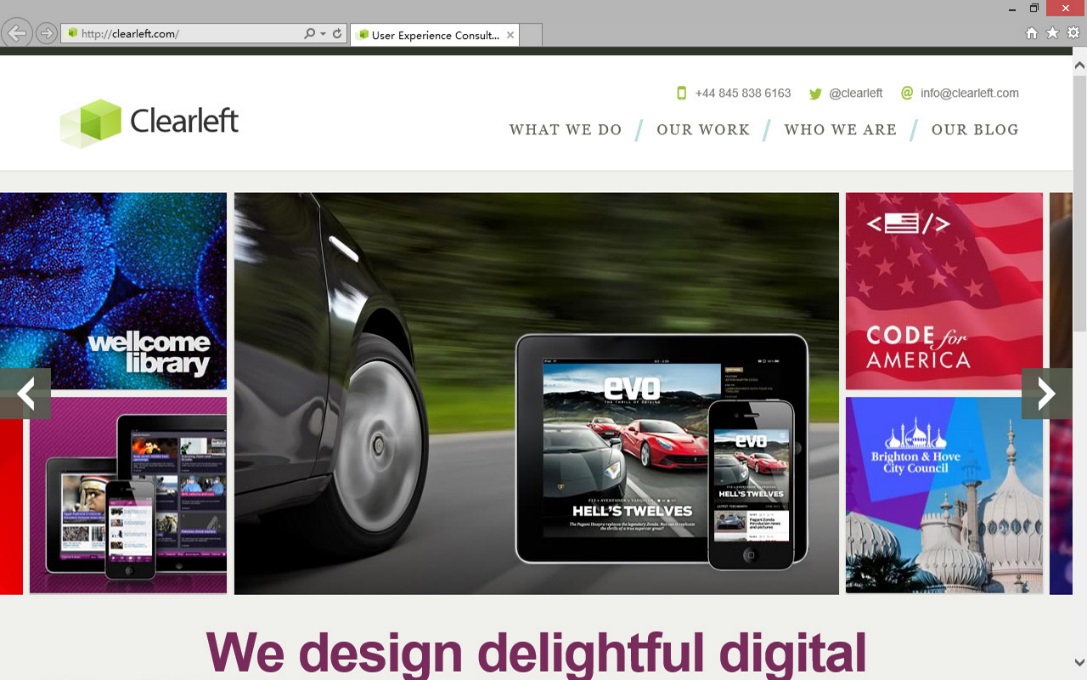


Figure xx. the outlier page.

Figure xx. the revised version of the outlier page.

图片需要处理 两侧白线加深，地下文字反白。颜色变浅 上面黑线变浅，上面图标加上阴影。左侧的红色变灰色，补上线条，右下图片变成灰色。

By carefully comparing the two scatter plots in figure xx, we find that generally the points with higher scores (above 0.5) are located towards the left, while the points with lower scores are skewed right. Make them more separable. however, there are still some misclassified cases. An extreme outlier (see the dot with red circle in figure xx) has a moderate aesthetic score of 0.7, but its VAE and R-VAE are quite high.

When we check the original image of the page (see figure xx) full screen, we can feel our eyes are somewhat strained with selection pressure. Thus, it is seemingly not an experimental error. The R-VRE really captures the unpleasant viewing experience for the observers’ eyes.

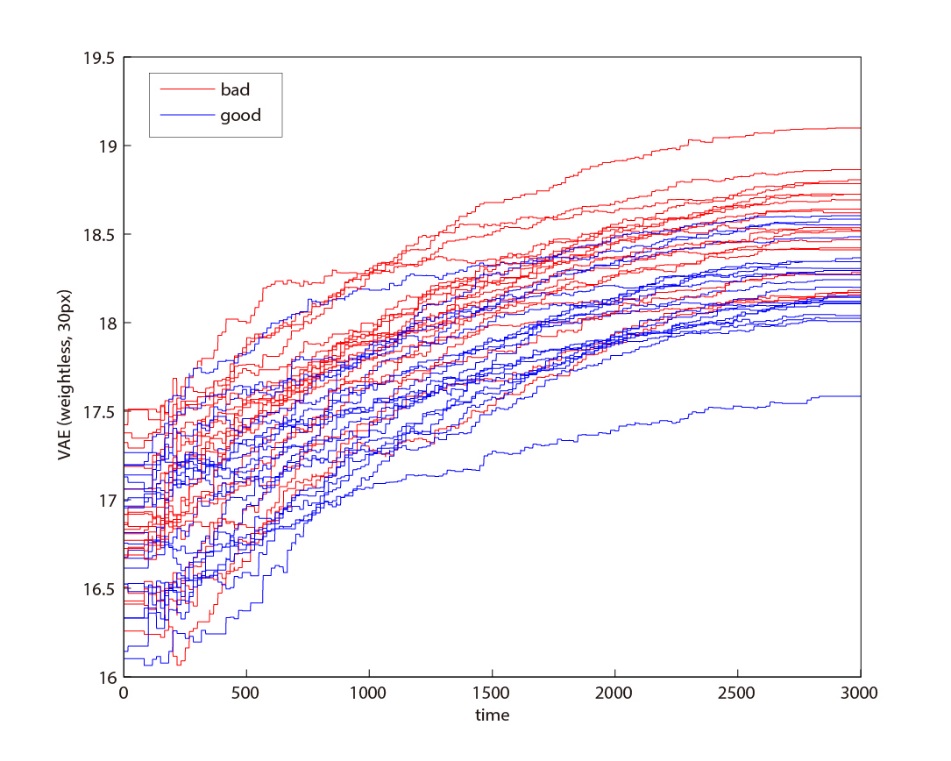
Why did this page get a fairly good aesthetic score? We think it might be attributed to the beautiful contents. Objectively speaking, this is not such a bad webpage, especially since the images are nice. However, from the designer’s point of view, it still has much room to improve. Figure xx bottom is a slightly revised version of the page.

It is no doubt that the contents of a webpage are an important factor to our attention and aesthetic cognition. VAE is not only about webpage aesthetics. VAE is actually more affected by bottom up low-level image features such as edge contrast or complexity.

## 4 The stability of VAE

Figure 4a visualizes how the VAEs of 40 pages increase over time. For every moment of time ti, we calculate the entropies based on the accumulated fixations during the period from t0 to ti, until i= 3000ms. Then for each of the 40 pages, we get a corresponding curve over time. All of the 40 curves fluctuate greatly at the beginning, and then gradually flatten and stabilize. In same way, Figure 4b visualizes how the R-VAE of each pages decreasing over time.

The blue curves are the aesthetically good pages while the red curves are the bad pages. Apparently, the good pages generally have lower VAE and R-VAE all the time. And the two groups are more clearly separated at the end.



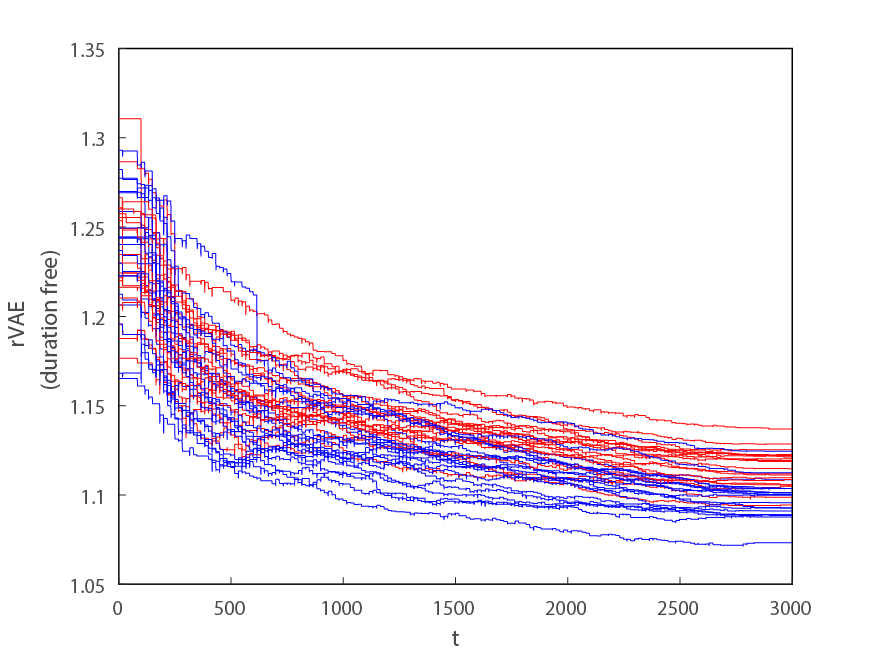


Figure xx. The VAEs (top) and the R-VAEs (bottom) of the 40 pages are calculated over time line. The blue curves are good pages. The red curves are bad pages.

In order to optimize the performance of VAE we traverse the correlations between the entropies and the aesthetic scores depending on the variables of the testing time, the number of subjects and the sigma.

Figure xx visualizes how the correlations between the entropies and the aesthetics developing over the testing time. The correlations become most significant at 3s. As we did not record the eye movements after 3s, we are not able to see the development afterwards. We suppose the correlation will continue to become stronger, judging from the tendency, which is still slightly decreasing.

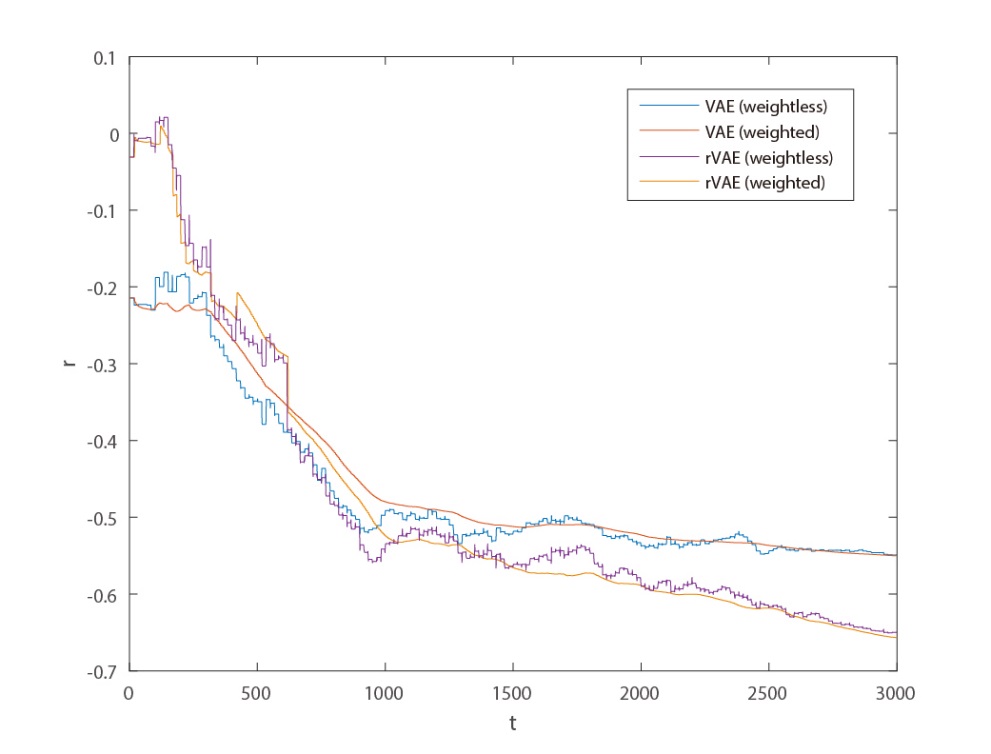
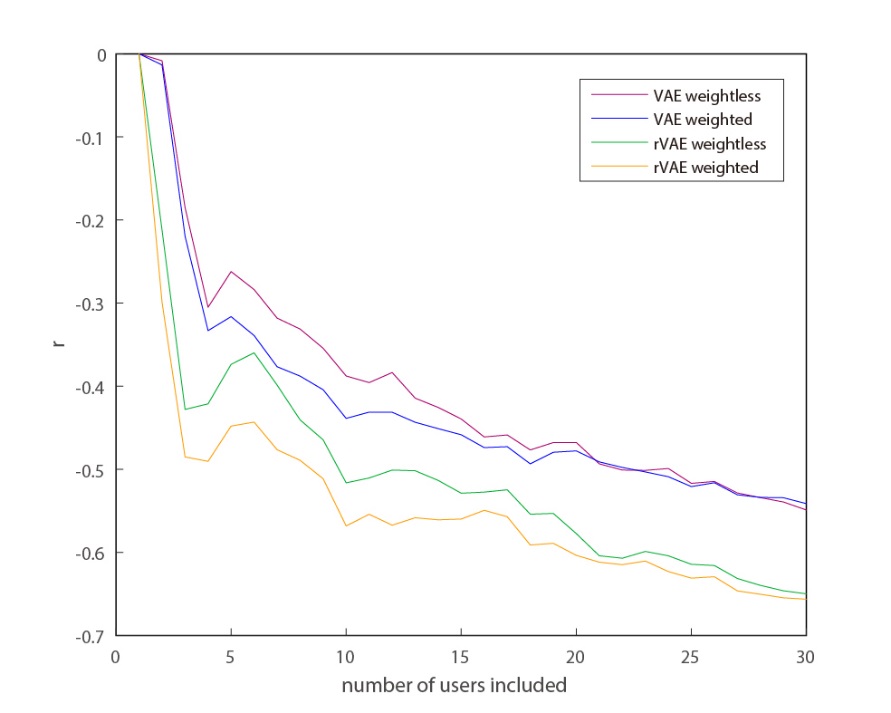


Figure xx. Four curves represent VAE and R-VAE (both with and without duration weights), their correlations to the aesthetic scores developing over the time line. It is clear that R-VAEs keeps below the VAEs after 1000ms. It seems that the R-VAEs could perform better even after the 3000ms.

The figure xx visualizes how increasing the number of subjects gradually stabilizes the performances of the visual entropies. They gradually become stable by enlarging the sample groups of subjects, sized 2, 3, 4… 29, which are generated by random from the total 30 subjects.

The coefficients are gradually stabilized with the increasing number of subjects and gradually converges between -0.5 and -0.6. The graph also indicates that 30 subjects might be a balance point between the precision and the costs.

The larger the number of subjects, the smaller the testing error. However, it also means more time and more money. Furthermore, the accuracy of the posteriori probability gradually ceases to increase after a certain number of samplings.



sigma = 30px。

Figure xx. Four curves represent VAE and R-VAE (both with and without duration weights), their correlations to the aesthetic scores gradually increase with the number of subjects.

Until now, all the analysis of the entropies is based on the heat map with sigma = 30px. How sensitive is the performance when the sigma’s value is changing?

The figure xx visualizes the Pearson correlations changing by the sigma. For a wide range of the sigma, from 13px-60px, the correlation coefficients stays below -0.5. And the curves have minima at about 20-30px. So the entropies are not so sensitive to the size of the Gaussian.

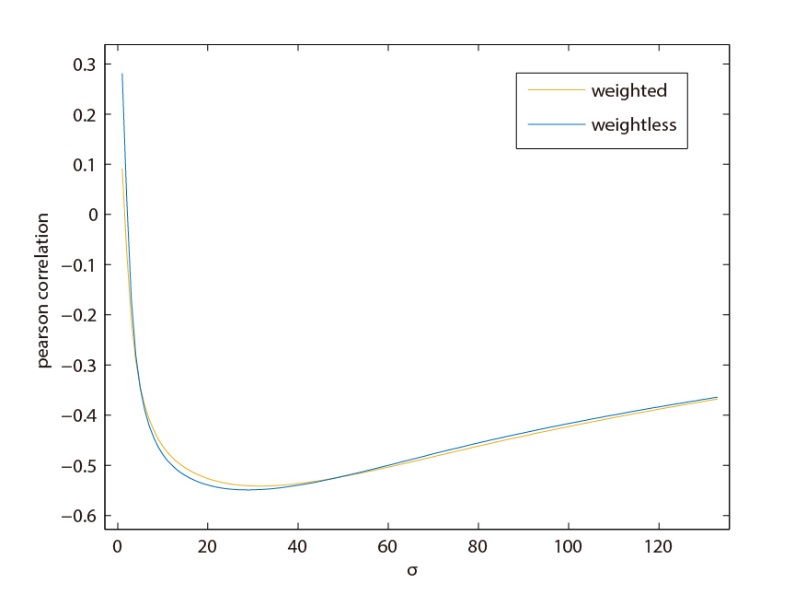


figure xx. Two curves are for weighted and weightless VAEs. Their Pearson correlations with aesthetics smoothly changing by gradually increased sigma.

The above analysis verified the stability of the visual attention entropies in predicting our aesthetic judgements.

## Discussion and Conclusion

the results of this experiment prove that our aesthetic judgments of webpages is related to our eye movement behavior. The Shannon Entropy measure was testified on two aspects of the eye tracking data: gaze sequence and heat map. The entropy in heat map, VAE, was proved predictive to our aesthetic judgements, while the entropy in gaze series (Markov chain) was not. VAE as the entropy of visual attention can be interpreted as the chaos in allocating of our limited attentional resource. Its improved version, R-VAE has significant correlation (r=-0.65) with our perceived aesthetics. This single metric alone can predict whether a webpage aesthetically pleasing or not. to a certain degree (87.5%).

The success of VAE is at least partially attributed to the interpolation with Gaussians in the belief that the fixation distribution is continuous. In addition, projecting the fixations during the entire 3s on a single plane makes the data points denser. Correspondingly, the failure of the entropy in gaze sequence is possibly due to the extra temporal dimension making the data points much sparser. It is hard to conclude so far that temporal information is irrelevant to our aesthetic judgements. Maybe in the future a more informative model could be found that is more effective in solving this problem.

A lower VAE can be interpreted as less effort expended in eye orienting, searching and less distraction in looking - in other words, lower perceptual or attentional load. The evidence that aesthetically pleasing pages have lower VAEs supports the fluency hypothesis [], and somehow explains why “what is beautiful is usable” [ ]. VAE is based on the special information of eye fixations. As a new objective and quantitative index of eye tracking data, it is surely not just a metric for aesthetics. VAE in nature is more related to our perceptual and motor control ability: eye gazing is like mouse pointing. That’s the reason why some visual design principles for eye guidance are quite similar to Fitts law [].

Relative-VAE can be interpreted as a balance between using as low of a VAE as possible to perceive as much as possible useful information in the constraint of the screen. This explanation is in line with the aesthetic principle: “maximum effect for minimum means”[], which is rooted in the Evolutionary Aesthetics [ ]. The theory argues that the basic aesthetic preferences of Homo sapiens have evolved in order to enhance survival and reproductive success.

Eye movements are the collaboration of overt and covert attention. Though we found that to predict a web page’s aesthetics needs at least 1000ms eye tracking data, it has been reported that our first impression of a webpage are formed as quickly as in 50-500ms [ ]. 50ms is barely enough for a snapshot and our eyes are still at their original positions. It implies that brain can make an aesthetic judgements as soon as a covert attention map is formed by the visual residual after mere 50ms exposure.

Acknowledgment