Mining Visual Evolution in 21 Years of Web Design

Ali Jahanian
CSAIL MIT
Cambridge, MA 02139, USA
ali-design@csail.mit.edu

Phillip Isola
UC Berkeley
Berkeley, CA 94720, USA
phillipi@berkeley.edu

Donglai Wei
CSAIL MIT
Cambridge, MA 02139, USA
donglai@csail.mit.edu

Abstract

The web contains a treasure trove of design data, with
many web pages being the product of careful thought about
layout, font, and color scheme. Not only does the current
web document current design trends, historical snapshots
of the web are a lens into past fashions. The Internet Archive
[2] has captured snapshots of the public Internet each year
going back to 1996. In this paper, we present a curated
dataset of 21 years of web design, scraped from the Inter-
net Archive. We report initial analysis of design trends ap-
parent in this data, and we demonstrate how the data can
be modeled with deep neural networks to enable novel de-
sign applications, such as predicting the apparent year of a
web design. The novelty of our work is two-fold: (1) mining
the long-term temporal evolution of designs on the Internet,
and (2) using deep neural networks as a tool for discovering
design elements, which can complement the hand-curated
features so far used in data-driven design mining.

Author Keywords

Design mining; data-driven design; web design; design evo-
lution; deep learning.

ACM Classification Keywords

H.5.2 [Information Interfaces and Presentation]: UI

(a) 1996   (b) 2016

Figure 1: These two web pages might seem generic, but each reflects the design trend of its own era. Can you tell what makes the 2016 web page look like 2016? Given a large dataset of web pages over 21 years (1996-2016), our algorithm can harvest the temporally-indicative visual patterns (e.g. patches in each row) to reveal the evolution of a design trend. Over each example web page above, we show regions our algorithm thinks are diagnostic of that web page's year. On each row, one patch from the web page is shown, followed by four patches, from other pages, that the algorithm considers to have a similar visual pattern. Noticeably, the old design (1996) mainly uses simple text and textual links for communicating the content, and a modern design (2016) applies more diverse graphical elements such as special fonts for text, graphic logos, and natural images.

Introduction

The design of web pages has changed dramatically over the last 21 years. From the animated gifs and marquees of the mid-1990s, to the embossed corners and drop shadows of the mid-2000s, to the flat minimalist pages of the 2010s, the web documents many important design trends of the last two decades. In this paper, we describe our work-in-progress on collecting and analyzing a large-scale dataset of historical web page snapshots. Our source of historical data is the Internet Archive [2], which has recorded snapshots of the public Internet at regular intervals since 1996 – often containing multiple snapshots of a page per month. We are gathering and organizing this data into a form amenable to data analysis. A primary contribution of this work will be making our dataset public as a resource to study temporal trends in web design.

In this extended abstract, we demonstrate several ways the data can be mined to reveal trends, and we propose several design applications that become possible when modeling data at this scale. First, we study how colors have changed in popularity over time. Second, we train deep neural network to analyze more complex design patterns. We train the network to predict the year of a web page from it's screenshot, demonstrating that this can be accomplished with reasonable accuracy. We then analyze some of the design features this network has learned to become sensitive to.

Related Work

Recent data-driven approaches have contributed to design evaluation and creation applications. Kumar et al. [10] suggest a design mining platform with the goal of querying web design assets (e.g., banners) based on the html struc-
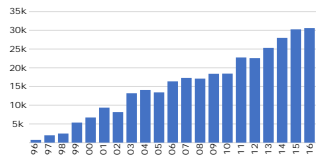
ture and graphical elements from more than 100,000 web pages. Reineke et al. model aesthetics at first impression from a dataset of web pages, using two measures of visual complexity and colorfulness [12]. Miniukovich and De Angeli further model the aesthetics of graphical user interfaces based on two datasets of screenshots from web pages and mobile apps [11]. Jahanian et al. model association of colors and linguistic concepts (color semantics) by design mining from a dataset of 2,654 magazine covers from 1998 to 2013, in 71 titles and 12 genres using topic modeling [8]. However, understanding the temporal aspect of data has not been the goal of these studies.

Considering temporal aspects of design, Ivory and Rodrick [7] evaluate the design evolution of more than 1,500 sites and 22,000 pages over the years of 2000, 2002, and 2003. They first define more than 150 quantitative measures over basic html elements (e.g., amount of text, links, images), page formatting (e.g., use of interactive elements), performance (e.g., download speed), and site architecture (e.g., consistency of the pages) [5]. Then, they use these measures to classify the pages based on quality ratings (e.g., visual design and ease of navigation). They further select a subset of highly-rated pages and compare the scores of features learned by their linear regression classifiers, and report several consistent design patterns for the three years. They further apply these analyses to confirm good design guidelines. Later Ivory manifested these evaluations in a book [6].

More recently, Alharbi and Yeh suggest a design mining framework for analyzing design patterns in Android apps [1]. They created a dataset by tracking 24,436 apps over 18 months. Then, they define a list of features including app details (e.g., title and description of the app), visual appearance, and behavioral features, and use these features

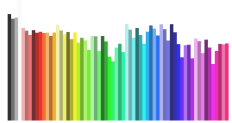
Figure 2: Distribution of data over years.

Figure 3: Distribution of colors over years. Magnitudes are in a logarithmic scale. The color space is a discretized version of sRGB with 4 bins per channel, and sorted based on hue.

to classify their data. Through this analysis, they delineate design patterns for eight aspects of user interface design (e.g., home screen widgets and navigation strategies).

We define the problem visual design mining as a computer vision task, and analyze on the perceptual aspects of screenshots of web pages. Our modeling framework is based on automatically learning perceptual features that can complement the so far hand-crafted measures. We also consider a wider range of years (and more frequent monthly screenshots) than prior work.

Dataset
In order to create a temporal dataset for the web, we retrieve web pages from the Internet Archive [2]. Previous scraped datasets from this archive are either *headless*, i.e. the Document Object Model (DOM) without the rendered page, or nonpublic screenshot datasets scraped for different goals than ours (for details, refer to the datasets and services available on the Internet Archive offered by individuals and private parties). Because we are studying the design aspect of the web pages, it is important to collect screenshots with the quality that modern web browsers provide. For this reason, we developed a scraper using the Selenium WebDriver [1] and Java with the Firefox driver. The scraper takes a screenshot of the entire page (scrolling to load full pages) and downloads the page DOM. The scraper also captures some details about how graphical elements are laid out, e.g., locations of graphical elements (i.e. <img> tags) on the page while it is rendered.

To generate a list of web pages to scrape, we use the list of high-traffic urls according to the Alexa [2] service. We then filter out the pornography urls using Web of Trust [3]. This

---

[1] http://www.seleniumhq.org/projects/webdriver
[2] http://www.alexa.com
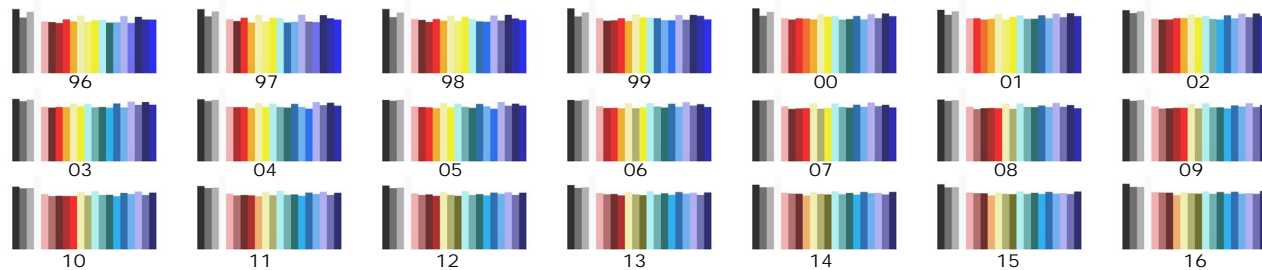[3] https://www.mywot.com/

Figure 4: Top 20 most frequent colors in each year. Magnitudes are in logarithmic scale. The color space is a discretized version of sRGB with 4 bins per channel, sorted based on hue. Through this analysis, we observe that the Internet is getting duller (less saturated) and a wider, more evenly spread, range of colors is being used as time progresses (distribution over colors is becoming higher entropy).

Figure 5: Confusion matrix for the test dataset.

current report is on 323,483 screenshots from 6,557 unique urls. For each year, we capture one screenshot per month (whenever a snapshot for that month is available) from 1996 to September 2016. Figure 2 illustrates the distribution of screenshots over the 21 years. Figure 3 illustrates the distribution of colors over all years, indicating a prevalence of blue hues in web design. Note these colors are presented in a discretized sRGB space with 4 bins per channels, and the histograms plot log frequencies.

Modeling

As a preliminary series of studies, we investigate temporal trends in the usage of various design elements. First, we analyze the evolution of color usage. Second, we mine for more complex visual patterns by training an AlexNet CNN model [9] to predict the year of a design, and then probing which patterns the network picked up on.

Statistical Approach: trends in color usage
In Figure 4, we analyze the evolution of color usage over time. For each year, we plot the frequency of each of the top 20 most common colors used on webpages snapshot-

ted from that year. Two trends are apparent: 1) web colors are becoming duller and less saturated over time, and 2) whereas in the 1990s a few specific colors were used very often (spikes in the plots, which show log frequencies), by the 2010s, a broader and more evenly spread range of colors was being used (indicated by flatter distributions in the plots).

Learning Approach: trends mining
Beyond color usage, we aim to find discriminative design elements that are indicative of time, which can be low-level features (e.g., usage of blue buttons) or high-level patterns (e.g., usage of face photos). However, the design elements are often outnumbered by other trend-irrelevant parts of the web pages. One common way to find needles in the haystack, e.g., [3], is to use handcrafted features and sophisticated clustering techniques. On the other hand, recent works [14] show that CNN models can automatically learn hierarchical features to use for pattern discovery through label prediction tasks. Below, we adopt the latter approach and train an AlexNet CNN model, *WebDesign-CNN*, on the task of year prediction. We use this network to find salient
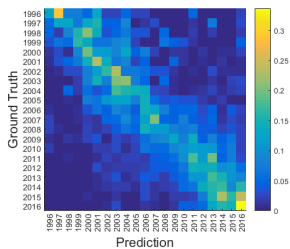
conv1

conv3

pool5

WebDesign-CNN

Figure 7: Visualization of WebDesign-CNN neurons at different layers. We show all 64 filters for conv1 units and four sample filters for conv3 and pool5. Noticeably, the WebDeisgn-CNN learns mostly horizontal and vertical edges (conv1), simple shapes (conv3), and more complex visual patterns like logos and faces (pool5).

1996   1998   2000   2002   2004

2006   2008   2010   2012   2014   2016

time

Figure 6: Here we show, for each predicted year, the top four web pages our model is most confident belong to that year, revealing designs that are highly characteristic for each year, such as simple textual web pages in the 1990s, and more image-heavy designs of the 2010s.

web pages that reveal temporal trends (Figure 6). Then, to further dissect the trends into visual patterns, we visualize the learned neurons in the model (Figure 7). Lastly, we interpret individual web page prediction results by analyzing the contributions of these visual patterns (Figure 1).

We trained the WebDesign-CNN to classify each 256×256 screenshot of a web page into one of the 21 years from 1996 to 2016. The network was trained on 310,770 images and we tested on an independent set of 11,713 images, ensuring that the test set contains no urls in common with the training set. This network achieves 18.0% top-1 classification accuracy (chance = 1/21 or 0.047%). In Figure 5, we show the confusion matrix of classifications on the test set. The diagonal structure of these confusions indicates that the network is successful at predicting the approximate year

for most test web pages. This degree of accuracy opens up the possibility of applications where a designer is given feedback about how new or old fashioned their web page looks. We discuss these applications more in the Future Work section.

With WebDesign-CNN, we now can reveal the design trends over the years through its prediction score. In Figure 6, we show the four test images that the CNN is most confident belong to each year. These are web pages that the model thinks are highly characteristic of each year. By examining these representative images, we can qualitatively identify design trends, such as the increasing use of photographic imagery over time.

Further, the WebDesign-CNN model learns a hierarchical

visual representation of design elements that are indicative of temporal trends. Specifically, we visualize the discriminative patterns stored inside neurons in the first, middle, and last layers of the network (conv1, conv3 and pool5) (Figure 7). For the conv1 layer, the neurons can be directly visualized as RGB images, where they capture horizontal and vertical patterns, perhaps picking up on the axis-aligned nature of web designs. For the conv3 and pool5 layers, we adopt the data-driven methods of [14] to find patches from the test data with the top activation for each neuron (we visualize 4 selected neurons for each of these layers in Figure 7). With increasing layer depth, neurons begin to capture more semantically interpretable patterns: e.g., boxes and text characters in the conv3 layer; logos and human faces in the pool5 layer.

After visualizing the patterns captured by pool5 neurons, we attempt to answer the question "what make a 2016 web page look like 2016". Given an input image, we can approximate the contribution of each pool5 neuron by its gradient with respect to the correct year classification [13]. In this way, we can select the top pool5 neurons for an image, and show which image patches caused these neurons to fire. In Figure 1, apply this method to two seemingly generic web pages, one from 1996 and the other 2016. Next to each detected patch, we show other patches in the test data that also highly activate these neurons, revealing the visual patterns each neuron is sensitive to (as described in the paragraph above). This analysis suggests that the two web pages in Figure 1 have distinct design elements, evolving from text and simple graphic elements to natural images and pretty icons.

Conclusions and Future Work

Being able to predict the apparent date of a web page opens up a number of design applications. First, we may be able to use this tool to identify web pages that are "trend setters", by finding pages whose predicted date is much later than the actual year they were created. A current difficulty in this analysis is that many older designs linger on the Internet, and therefore, training a model to simply predict the year of web page snapshot may not accurately reflect the year in which the design was conceived.

Another potential usage would be pointing out to a user if their web design looks out of date. We further envisage probing the model to identify which specific regions are causing it to look out of date. Such a tool could also guide a user to design a vintage look, by telling the user when their design takes on the appearance of, say, the year 2000.

We further plan to utilize generative neural networks (e.g., [4]) to create designs that look like a specific time. Such models can be utilized in the context of creativity support tools. For instance, there has been extensive study in HCI community about how design by example opens up new ways of creating alternatives of a design.

Our broader goal, however, is contributing to theories of design by understanding how people in different times and locations have created these artificial displays. Because our dataset contain metadata about geography of the web pages, in addition to date, we plan to incorporate the location aspect into the model. In this way, we can model both time and location of design, and perhaps answer questions such as: what does this design tell us about the culture (location and time) it comes from.

References

[1] Khalid Alharbi and Tom Yeh. 2015. Collect, decompile, extract, stats, and diff: Mining design pattern changes in Android apps. In *Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services*. ACM, 515–524. DOI:http://dx.doi.org/10.1145/2785830.2785892

[2] The Internet Archive. 1996. The Wayback Machine. (1996). Retrieved Jan 10, 2017 from https://archive.org

[3] Carl Doersch, Saurabh Singh, Abhinav Gupta, Josef Sivic, and Alexei Efros. 2012. What makes paris look like paris? *ACM Transactions on Graphics* 31, 4 (2012). DOI:http://dx.doi.org/10.1145/2185520.2185597

[4] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. 2014. Generative adversarial nets. In *Advances in Neural Information Processing Systems*. 2672–2680.

[5] Melody Yvette Ivory. 2001. *An empirical foundation for automated web interface evaluation*. Ph.D. Dissertation. University of California, Berkeley.

[6] Melody Y Ivory. 2013. *Automated Web Site Evaluation: Researchers' and Practitioners' Perspectives*. Vol. 4. Springer Science & Business Media. DOI:http://dx.doi.org/10.1016/j.ipm.2006.05.002

[7] Melody Y Ivory and Rodrick Megraw. 2005. Evolution of web site design patterns. *ACM Transactions on Information Systems (TOIS)* 23, 4 (2005), 463–497. DOI:http://dx.doi.org/10.1145/1095872.1095876

[8] Ali Jahanian, Shaiyan Keshvari, S. V. N. Vishwanathan, and Jan P. Allebach. 2017. Colors – Messengers of Concepts: Visual Design Mining for Learning Color Semantics. *ACM Trans. Comput.-Hum. Interact.* 24, 1, Article 2 (Jan. 2017), 39 pages. DOI:

http://dx.doi.org/10.1145/3009924

[9] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. 2012. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*. 1097–1105.

[10] Ranjitha Kumar, Arvind Satyanarayan, Cesar Torres, Maxine Lim, Salman Ahmad, Scott R Klemmer, and Jerry O Talton. 2013. Webzeitgeist: design mining the web. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 3083–3092. DOI:http://dx.doi.org/10.1145/2470654.2466420

[11] Aliaksei Miniukovich and Antonella De Angeli. 2015. Computation of interface aesthetics. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. ACM, 1163–1172. DOI:http://dx.doi.org/10.1145/2702123.2702575

[12] Katharina Reinecke, Tom Yeh, Luke Miratrix, Rahmatri Mardiko, Yuechen Zhao, Jenny Liu, and Krzysztof Z Gajos. 2013. Predicting users' first impressions of website aesthetics with a quantification of perceived visual complexity and colorfulness. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2049–2058. DOI:http://dx.doi.org/10.1145/2470654.2481281

[13] Karen Simonyan, Andrea Vedaldi, and Andrew Zisserman. 2014. Deep inside convolutional networks: Visualising image classification models and saliency maps. In *International Conference on Learning Representations*.

[14] Bolei Zhou, Aditya Khosla, Agata Lapedriza, Aude Oliva, and Antonio Torralba. 2015. Object detectors emerge in deep scene cnns. In *International Conference on Learning Representations*.