A Deep Study into the History of Web Design

Bardia Doosti
School of Informatics and Computing
Indiana University Bloomington
bdoosti@indiana.edu

David J. Crandall
School of Informatics and Computing
Indiana University Bloomington
djcran@indiana.edu

Norman Makoto Su
School of Informatics and Computing
Indiana University Bloomington
normsu@indiana.edu

ABSTRACT

Since its ambitious beginnings to create a hyperlinked information system, the web has evolved over 25 years to become our primary means of expression and communication. No longer limited to text, the evolving visual features of websites are important signals of larger societal shifts in humanity's technologies, aesthetics, cultures, and industries. Just as paintings can be analyzed to study an era's social norms and culture, techniques for systematically analyzing large-scale archives of the web could help unpack global changes in the visual appearance of websites and of modern society itself. In this paper, we propose automated techniques for characterizing the visual "style" of websites and use this analysis to discover and visualize shifts over time and across website domains. In particular, we use deep Convolutional Neural Networks to classify websites into 26 subject areas (e.g., technology, news media websites) and 4 design eras. The features produced by this process then allow us to quantitatively characterize the appearance of any given website. We demonstrate how to track changes in these features over time and introduce a technique using Hidden Markov Models (HMMs) to discover sudden, significant changes in these appearances. Finally, we visualize the features learned by our network to help reveal the distinctive visual elements that were discovered by the network.

CCS CONCEPTS

• **Information systems** → **Surfacing**; • **Human-centered computing** → **Web-based interaction**; • **Computing methodologies** → **Interest point and salient region detections**; *Supervised learning by classification*; *Neural networks*; • **Mathematics of computing** → Kalman filters and hidden Markov models;

KEYWORDS

Web Design, Deep Learning, Convolutional Neural Networks, Cultural Analytics

1 INTRODUCTION

The advent of digital technologies has brought about a revolution in analyzing and unpacking "culture." Cultural Analytics is a relatively new field that aims to study the humanities and other disciplines through computational analysis of large-scale cultural data [16]. For example, work in cultural analytics has automatically analyzed patterns in large historical and contemporary samples of art [25], pop music [6], comic books [17], Vogue magazine covers [24], and architecture [14]. Ironically, however, perhaps the most important and best reflection of today's "new media" [16] – the world wide web itself – has had little examination through the lens of cultural analytics. There is growing recognition that such new media should be preserved. For instance, The Internet Archive [4] attempts to store a comprehensive history of the web, while the University of Michigan Library houses the Computer and Video Game Archive [2]. The web is now a first-class cultural artifact with at least one quarter of a trillion archived pages across nearly 30 years [4].

Recent work has argued that analyzing the *visual designs* of websites could provide a window into the evolution of the web, and specifically how visual design reflects changes in visual aesthetics, role of technology, cultural preferences, and technical innovations [8, 22]. Reinecke *et al.* [22, 23] defined specific low-level metrics for quantifying visual properties of websites, such as color distribution, amount of white space, and structure of page layout, and developed a model of perceived visual complexity based on these low-level measures. They used this technique to reveal cultural preferences for particular aesthetic styles, for example. Chen *et al.* [8] asked web designers, developers, and artists to view historical collections of web pages and reflect on the changes they observed across time, and specifically to speculate on the web's design "periods" including the key changes and causes of changes that have occurred over time. While limited in scope, these papers represent a first step in understanding cultural patterns via analysis of the visual designs of websites.

Although similar in many respects, the web has a number of key differences from more traditional cultural artifacts. Automated studies of art and music tend to be limited by the quantity of available data, while the number of web artifacts is essentially boundless, with millions of new pages coming online each day. On one hand, this may make it easier to find statistically-significant general patterns among all of the "noise" of individual web pages; on the other hand, it makes manual study and organization of web design impractical. Moreover, unlike art and music, the web lacks well-developed theories to compare and contrast visual design styles. We thus need to develop automated techniques that can be used to characterize, compare, and contrast visual design styles in a meaningful way.

In this paper, we take initial steps towards this goal and make four key contributions, using recent progress in computer vision and machine learning. We first ask how much "signal" can be derived from the visual designs of websites – how much information about cultural patterns are encoded solely in the visual appearance of sites? We describe an approach in which we train state-of-the-art classifiers borrowed from computer vision, specifically deep Convolutional Neural Networks (CNNs), to recognize the eras and

WebSci'17, June 25-28, 2017, Troy, NY, USA.

978-1-4503-4896-6/17/06…$15.00
DOI: http://dx.doi.org/10.1145/3091478.3091503

genres of web pages. Our results show that classifiers are able to classify websites by their genre by more than 4 times the baseline and recognize website design era by 2.5 times the random baseline, which indicates that modern computer vision can indeed discern particular patterns from large scale datasets of website designs. The results are accurate enough to suggest that visual appearance could be an important signal for societal change as reflected by the web.

Second, we show that the key features identified by the classifier during this supervised learning task can be used to characterize *new web pages* in an unsupervised way, without predefining handcrafted visual attributes. Intuitively, these features generate a new visual similarity space that is automatic, objective, and potentially more meaningful than metrics defined by human intuition. With this framework, we can, for instance:

(1) Measure the visual design similarity between two pages. How similar is a website to a prominent, trend-setting website – how "Apple-like" is `cnn.com`?
(2) Measure the similarity of a given page to a particular genre of website. How "news-site-like" or "entertainment-site-like" is the visual design of `cnn.com`?
(3) Measure the similarity between a given page and a particular website era. How "modern" is the design of `cnn.com`?

Third, we present techniques for using these measures to perform cultural analytics over time at a large scale. We take historical snapshots of a single website, as captured by the Internet Archive, and characterize each individual snapshot using our objective metrics to give a (noisy) time series quantifying how the design has changed over time. We then develop a Hidden Markov Model to robustly identify sudden changes in the time series corresponding to actual changes in visual design (and ignoring the day-to-day changes caused by updates to content). This historical perspective gives us a glimpse on how websites have evolved over time, which could lead to broader, more fundamental insights on the relation of visual design and society and culture. For instance, we could measure how a particular company's website has become influential on other websites' designs (e.g., to what extent have companies sought to emulate Apple's aesthetics or vice versa?).

Lastly, we show that with our trained CNN, we can randomly generate *novel* website designs by inverting the process, by using the CNN to sample a design for a given point in the similarity space. Such tools not only shed light on the most basic question of "what does it mean to be a website design?" but also could serve to inspire current website designers on the future of web design.

2 RELATED WORK

While we are aware of very little work that tries to characterize website design automatically across time and genres, the basic idea of applying machine learning and data mining algorithms to analyze websites is not new. Kumar *et al.* [12] introduced a data mining platform called Webzeitgeist which used WebKit [1] to generate features such as ratio, dominant colors, and number of words. Webzeitgeist can extract useful information about a website, but is based on HTML code and thus captures only a rough sense of a page's visual design. Many other papers classify the genre and other properties of web pages from text (HTML) analysis (see Qi *et al.* [20] for an overview). In contrast, our goal is to analyze websites based on the way that people experience them: viewing the visual appearance of a website and making inferences based only on it, without examining the underlying HTML code.

Perhaps the closest line of work to ours is that of Reinecke *et al.* [23] and Reinecke and Gajos [22], who design visual features to measure properties of websites like color distribution, page layout, amount of whitespace, etc. They proposed a model to predict higher-level perceived visual properties of websites such as visual complexity from these low-level features, and used them to characterize cross-cultural differences and similarities in web design. Similar to our study, their features are based on rendered images as opposed to HTML. However, they rely on hand-engineered features that may be subjective and not necessarily representative. In contrast, we learn visual features in a data-driven way, and use these to track changes in web design across genres and eras. The two approaches are complementary: our automatic approach might discover visual features that are hard to describe or would not otherwise be apparent, whereas their features are explicitly tied to a human perceptual model, which may make their results easier to describe and interpret.

We also draw inspiration from the work of Chen *et al.* [8], who interviewed experts and asked them to critique and group together the design of prominent websites over time. Their study identified certain design elements or "markers" (such as search and navigation bar placement, color scheme, relative proportion of text and imagery, etc.) that distinguished various "eras" of web design. The study also suggested that web design evolution is driven by multiple factors including technological developments (e.g., new HTML capabilities and features), changing roles and functions of the web over time, impression management of companies and individuals (e.g., companies wishing to project confidence, friendliness, etc.), as well as changing aesthetic preferences. We use this non-technical exploration of web design as the inspiration for our paper, which seeks to study visual design automatically and objectively, at a large scale.

Our work uses computer vision to define visual features and similarity metrics for comparing and characterizing web design. Quantifying the similarity of images is a classic and well-studied problem in computer vision. A traditional approach is for an expert to engineer visual features, by hand, that they think are relevant to the specific comparison task at hand, and then apply machine learning to calculate the similarity of two images. Often these features are specialized versions of more general features, such as Scale Invariant Feature Transforms (SIFT) [15] which tries to capture local image features in a way that is invariant to illumination, scale, and rotation. A disadvantage of these approaches is that they are fundamentally limited by the programmer's ability to identify salient visual properties and to design techniques for quantifying them.

More recently, deep learning using Convolutional Neural Networks (CNNs) has become the de facto standard technique for many problems in computer vision. An advantage of this approach is that it does not require features to be designed by hand; the CNNs instead learn their own optimal features for a particular task from raw image pixel data. These networks, introduced by LeCun *et al.* [13] in the 1990's for digit recognition, are very similar to traditional

feed-forward neural networks. However, in CNNs some layers have special structures to implement image convolution or sub-sampling operations. These deep networks with many alternating convolution and downsampling layers help make the classifiers insensitive to spatial structure of the image and recognize both global and detailed image features. In 2012 Krizhevsky *et al.* [11] succeeded in adapting LeCun *et al.*'s CNN idea for a more general class of images using many more layers, very large training datasets with millions of images [9], huge amounts of computational power with Graphics Processing Units (GPUs), and innovations in network architecture. Since then, CNNs have been successfully applied to a large range of applications.

Here we show how to apply CNNs to the specific domain of web design. Contemporaneous with our work, Jahanian *et al.* [10] have also applied deep learning to website design, although they focus on the classification task of identifying a website's era based on its visual appearance. We take this idea a step further: we create CNNs to classify particular genres and eras of websites, and then use the trained features as data mining tools to characterize the visual features of websites and how they change through time.

3 METHODOLOGY

Our overall goal is to analyze the archive of a website, consisting of a time series of HTML pages, and characterize the changes in its visual appearance over time, including finding key transition points where the design changed. While we could analyze the HTML source code directly to detect design changes, this is difficult in practice since the HTML itself may reveal little about the physical appearance of a page: a page may be rewritten using a new technology (e.g. CSS and JavaScript) such that the source code is completely different but the visual appearance is the same, or a small change in the HTML (e.g. new background image) may create a dramatically different appearance. We thus chose to analyze the visual characteristics of rendered pages, emulating how a human user would see the page. The main challenge is that many pages (e.g., news sites) are highly dynamic, so that nearly every piece of content and most pixels in the rendered image change on a day-to-day basis. We wish to ignore these minor variations and instead find the major changes that correspond to evolutions in web site design.

We use two techniques for addressing this challenge. First, we extract high-level visual features that have been shown to correspond with semantically-meaningful properties using deep learning with Convolutional Neural Networks [13]. These features abstract away the detailed appearance of a page, and instead cue on more general properties that may reflect design, such as text density, color distribution, symmetry, busyness or complexity, etc. However, even these abstract properties vary considerably on a highly dynamic website, where content like prominent photos might change on a daily basis. We thus also introduce an approach for smoothing out these variations over time, in effect looking for major changes in the "signal" as opposed to minor variations caused by "noise." We apply Hidden Markov Models, a well-principled statistical framework for analyzing temporal and sequential data in domains like natural language processing, audio analysis, etc.

3.1 Dataset

We began by assembling a suitably large-scale dataset of visual snapshots (images) of webpages. Our dataset consists of (1) a large number of *current* snapshots of a wide variety of websites organized by *genre* (news, sports, business, etc.), and (2) a longitudinal collection of a large number of *historical* snapshots of a handful of pages over time.

For the genre dataset, we used CrowdFlower's URL categorization dataset, which consists of more than 31,000 URL domains, each hand-labeled with one of 26 genres [3]. We downloaded the HTML code for each URL and then rendered the page into an image using PhantomJS [5], a headless Webkit API [1], at a resolution of 1200×1200 pixels. (We chose this resolution because it works well with both the wide-screen format that many websites today support and earlier, less technologically-advanced designs). For websites that must be rendered at greater than 1200 pixels on either dimension (as is frequently the case along the vertical dimension), we cropped the snapshot.

For our second, longitudinal dataset, we collected snapshots for a set of prominent websites with a long history (from the early 1990s through the present) from the Internet Archive [4]. Unfortunately, this set of websites is sparse since most well-known websites did not exist or were not well-known before the 1990s (and thus were not archived by the Internet Archive until more recently). Many websites from the 1990s also disappeared after 2000. We chose the same 9 websites studied by Chen *et al.* [8] as well as 26 additional websites which were present in the 1990s (covering most of the genres mentioned in the CrowdFlower dataset). In total, we captured 7,303 screenshots from our 35 chosen websites from *archive.org*, spanning 1996 through 2013. We used the same process described above for rendering these websites to images.

We acknowledge that our relatively small dataset introduces limitations: a small dataset makes it difficult to pinpoint the accuracy and generalizability of a classifier. Our intent, however, is not to provide a robust, production-ready classifier. Instead, our results provide evidence that such classifiers can indeed reflect how information valuable to cultural informatics are signaled purely by a website's aesthetics. We hope future work will build validated CNNs or other models from which we can glean cultural signals.

3.2 Visual Features

For automatically and objectively measuring visual properties of large-scale collections of web pages, we need to develop quantitative measures of the visual appearance of a page. To do this, we first develop a technique for measuring the visual similarity between two rendered web page images, but in a way that attempts to ignore minor differences between pages and instead focuses on overall design.

In the Computer Vision community, deep learning with Convolutional Neural Networks (CNNs) has recently emerged as the *de facto* standard image classification technique, yielding state-of-the-art results on almost all vision tasks [11]. The basic idea is that unlike traditional approaches which use hand-designed visual features and whose performance is thus limited by human skill and intuition, CNNs learn the (hopefully) optimal low-level visual features for a given classification task automatically, directly from

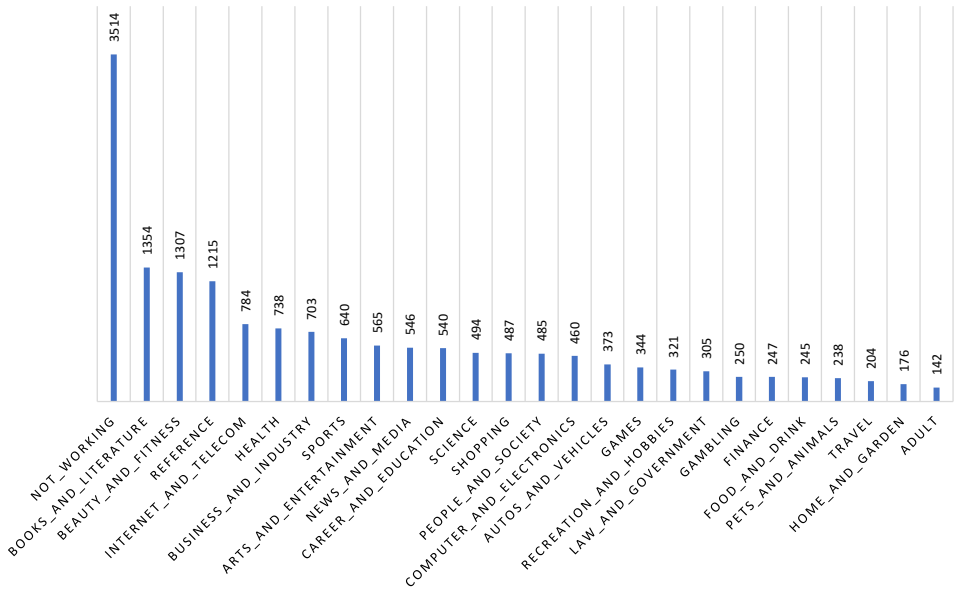


Figure 1: Frequency of website genres in our dataset

the raw image pixel data itself. Learning a CNN classifier can thus be thought of as simultaneously solving two tasks: (1) finding the right low-level features that are most distinctive and meaningful for a particular domain, and (2) learning a classifier to map these features to a high-level semantic class or category.

We use this property of CNNs in the following way. We train a classifier to estimate properties of web pages, including their genre and their "era" (when they were created), using visual features of the page itself. We can measure performance of the classifiers on these tasks, but producing accurate classifications is *not* our goal. (If it were our goal, we would just analyze the HTML source code itself instead of using the rendered image.) Instead, our goal is to train a classifier so that the CNN learns the important low-level features corresponding to visual style; we can then discard the classifier itself, and simply use these "optimal" features to compare websites directly.

3.2.1 Network Details. More specifically, we train CNNs for both genre and era classification tasks using the dataset described above. We use the popular AlexNet [11] CNN architecture, which consists of 5 convolutional layers that can be thought of as feature extractors, and 3 fully connected layers that classify the image based on those features. Each convolutional layer applies a series of filters (which are learned in the training phase) of different sizes to the input image and then finally pushes a 4096-d vector to the fully connected layers for classification. Since our dataset is not large enough to learn a deep network from scratch (which has tens of millions of parameters), we follow recent work [19] and initialize our network parameters with those pre-trained on the ImageNet dataset [9], and then "fine-tune" the parameters on our dataset.

3.2.2 Classification Results. For website genre categorization, the specific task was to classify each website into one of 26 different genre categories. We partitioned the dataset into training and test subsets, using half of the websites for training and half for testing. Since the frequency of classes was non-uniform (Figure 1), we also balanced the classes in both training and testing, so that the probability of randomly guessing the correct class is about 3.8%. Our classifier achieved about 16% correct classification rate on this task, i.e., four times the random baseline. These results may seem low, but we stress the difficulty of the task: the classifier only sees the visual rendering of the website (no textual or other features), and there is substantial noise in the dataset because many sites could be labeled in multiple ways (e.g. is Sports Illustrated a sports website or a news website?).

For website era categorization, we discretized time into four eras, 1996-2000, 2001-2004, 2005-2008, and 2009-2013, and again balanced classes. We split the test and training sets by *website*, i.e. all of the historical snapshots of `cnn.com` were in either the training set or test set, since the task could be very easy if very similar snapshots for the same site were in training and test sets. Here our CNN-based classifier achieved about 63% accuracy relative to a baseline of 25%, or about 2.5 times baseline. This is again a relatively difficult task because pages near the end of one era can be easily misclassified as belonging to the beginning of the next era, for example. Table 1 shows a confusion matrix on this task, which confirms that many misclassifications occur in adjacent bins.

These results show that while the visual classifiers are not perfect, the fact that the recognition rates are significantly higher than baseline shows that they are learning meaningful visual features. This suggests that our hypothesis of defining a similarity measure for website visual design using these features may succeed.

3.2.3 Visualization. The above results suggest that features extracted by deeply-trained networks carry meaningful information about visual style, but reveal little about what exactly this information is. Of course, this is one of the major disadvantages of deep

Table 1: Confusion matrix for recognizing website era.

	Predicted class			
	1996-2000	2001-2004	2005-2008	2009-2013
1996-2000	41	18	14	5
2001-2004	15	18	33	12
2005-2008	5	8	37	28
2009-2013	0	2	13	63

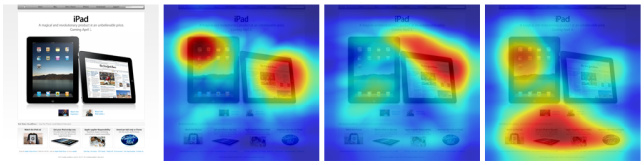


Figure 2: Heat maps showing which parts of a website support each of three different class hypotheses, according to the classifier.

machine learning: it is very difficult to interpret or "debug" the parameters of a learned classifier. One potential way of gaining some insight is to create visualizations that try to reveal which image features the network is actually cueing on.

We use a modification of the technique of Zhou *et al.* [26] to do this. Very briefly, this technique inserts a general average pooling layer into the network, which allows us to visualize the importance of each pixel as a heat map for each class. Although this method successfully generates an attention map of the network, it decreases the accuracy of the network by a few percentage points. Our compromise solution to preserve accurate results is to train both networks separately and substitute the convolutional layer weights learned by the classic AlexNet in the attention map network. This network has good results with both the classification and visualization of the attention maps. Figure 2 shows a sample input website snapshot and the attention map of three different classes for this image. In these visualizations, red regions are the most influential cues used by the network to conclude that the image belongs to a specific class, and blue corresponds to less important regions. These generated heat maps help us find the most important parts of each image for each class.

3.2.4 Website Generation. Although we trained our deep networks to extract features from websites and classify them into categories, an interesting property of these networks is that they can actually be run "in reverse" to generate novel exemplars. The intuitive idea here is that the networks learn a mapping from visual features into mathematical vectors in some high-dimensional space and it is possible to reverse the process by generating a random high-dimensional vector and then producing an image that has that feature representation.

To do this, we use Generative Adversarial Networks (GANs) [21]. Figure 3 shows some examples. These images are, at least in theory, novel website designs that do not appear in the training set but have similar features to websites that are in the training set. We believe that many of these designs seem quite plausibly "real," although it can be subjective. The network's architecture also limits the resolution of these generated images, so they appear blurry. Nevertheless, such a technique could provide a means of inspiring web developers, helping suggest new directions of visual design to pursue.

3.3 Temporal Smoothing

Given a time series of a visual feature over time, like that shown in Fig 8, our goal now is to segment into periods of generally homogeneous design. The challenge is to ignore spurious variations due to day-to-day changes in site content (e.g. different photos or text), and instead identify more major, longer-term changes that reflect shifts in the underlying visual design. This problem is reminiscent of filtering problems in signal processing, where the goal is to reconstruct an underlying low-frequency signal in the presence of added high-frequency noise.

Using this signal processing view, we initially tried applying a low-pass filter (e.g. a mean filter or a Gaussian filter) to the time series of visual feature values. While this succeeded in smoothing out the time series, it has the unfortunate effect of also smoothing out the sharp changes in the signal that are the transitions we are looking for. As shown, the problem is that a low-pass filter imposes the (implicit) assumption that the visual feature's value on one day should be almost the same as the value on the next, and does not permit the occasional discontinuities caused by major design changes.

We thus explored an alternative model that explicitly permits discontinuities. Suppose that we wish to analyze a website over a period of N days. For each day i, let f_i denote the value of the visual feature computed on the rendered image for that day, or the special value $\emptyset$ if the value is bad or missing. We assume that the value of this visual feature is a result of two different forces: the inherent design of the page, and the content on that day. We model this as an additive process, i.e. $f_i = d_i + c_i$, where d_i represents the value related to the design and c_i is the "noise" caused by changes in day-to-day content.

Our goal is to infer d_i, which we cannot observe, from the observations f_i which can be observed. From a probabilistic perspective, we want to find the values for $d_1, d_2, \ldots, d_N$ so as to maximize their probability given the observations,

$$\arg\max_{d_1,\ldots,d_N} P(d_1,\ldots,d_N|f_1,\ldots,f_N). \quad (1)$$

To do this, we build a model that makes several assumptions; although these assumptions are not likely to hold perfectly in practice, we have found that they work well enough for our purposes in analyzing the noisy visual feature time series. We first assume that the design in effect on any given day depends only on the design of the day before it, i.e. d_i is independent from earlier days, conditioned on d_{i-1}. This reflects the assumption that a site's design tends to stay consistent over time and not, for example, flip back and forth between two designs on alternating days of the week. Second, we assume that c_i is independent from c_j for any $j \neq i$, conditioned on d_i. This means that content changes from one day to the next are independent from one another. Taken together, and using Bayes'

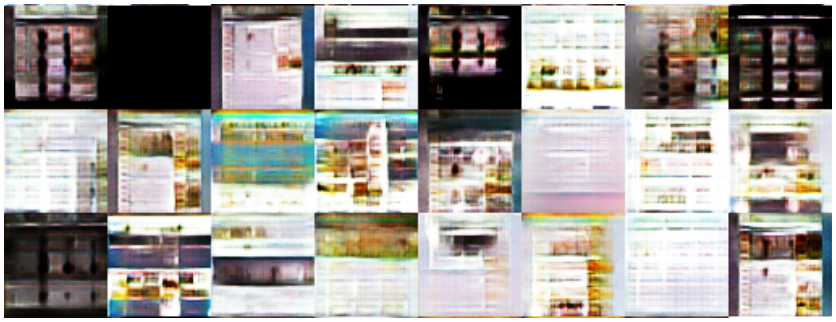


Figure 3: Novel website designs generated automatically by our network. (Images are blurry because the network architecture limits the resolution of generated images.)

law, these assumptions let us rewrite equation (1) as,

$$\arg\max_{d_1,\ldots,d_N} \frac{P(f_1,\ldots,f_N|d_1,\ldots,d_N)P(d_1,\ldots,d_N)}{P(f_1,\ldots,f_N)} \tag{2}$$

$$= \arg\max_{d_1,\ldots,d_N} P(f_1,\ldots,f_N|d_1,\ldots,d_N)P(d_1,\ldots,d_N) \tag{3}$$

$$= \arg\max_{d_1,\ldots,d_N} \prod_{i=1}^{N} P(f_i|d_i) \prod_{i=2}^{N} P(d_i|d_{i-1}) \tag{4}$$

where the denominator can be ignored since it depends only on observations, which do not depend on the values of d_i.

Equation (4) is a Hidden Markov Model, which are widely used to analyze sequences in areas such as natural language processing, speech recognition, robotics, computer vision, etc., and can be solved efficiently using the Viterbi algorithm. All that remains is for us to define the emission and transition probability distributions (i.e. first and second terms of equation (4), respectively). For the emission probability, we model c_i as a zero-mean Gaussian distribution, which assumes that day-to-day content changes modify the visual feature of the page in a similar way to white noise in signal processing (sometimes adding a bit, sometimes subtracting a bit),

$$P(f_i|d_i) = \mathcal{N}(f_i - d_i; \mu = 0, \sigma_1),$$

with some constant variance σ_1. For the transition probability distribution, we assume that on any given day, there is some small probability p_1 that the design changes, which causes an arbitrary change in the visual feature d_i, while with probability $1 - p_1$, the feature of the underlying design remains nearly the same (with the change modeled by a Gaussian distribution with a small sigma),

$$P(d_i|d_{i-1}) = p_1 + \mathcal{N}(d_i; \mu = d_{i-1}, \sigma_2).$$

We stress that our intention is not to create an accurate model of the process by which design changes happen, but instead to propose a reasonable enough model that it can be used to identify potential design changes automatically. The constants p_1, σ_1, and σ_2 are parameters of our analysis technique, and can be used to adjust the type of results that are found. For instance, increasing p_1 permits the technique to find more frequent design changes, but is also more likely to respond to random noise patterns. Increasing σ_1 or σ_2 allows greater flexibility in the observed visual feature relative to the underlying design feature, but again is likely to cause more noise. For our analyses in this paper, we set $p_1 = 0.1$, $\sigma_1 = 1.0$, and $\sigma_2 = 0.01$.

4 RESULTS

We now present results of our analysis techniques on our corpus of webpages. Since CNNs hide the "logic" behind their reasoning in a "black box," we first present visualizations that elucidate the features of the trained deep network itself. Next, we demonstrate application of our deep learning with website designs to cultural analytics. In this set of analyses, we examine the degree that visual designs of websites resemble other websites over time. We show how this can be applied to individual websites and a group of websites. Lastly, we propose that our deep network can also serve to inspire the future by generating new website designs that encapsulate the visual aesthetics of a specific era or category.

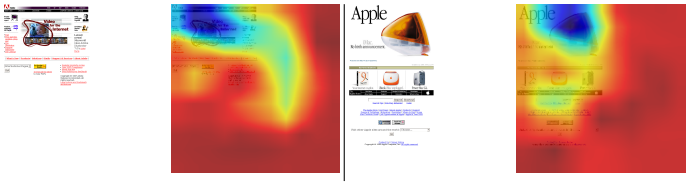
4.1 Visualizing Features of the Network

We visualized the output of the neurons to understand *how* the deep network recognizes the era and genre of the website. We found that in most cases, these features at least resemble what humans might look for when performing the same task.

4.1.1 Features for Era. Figure 4(a) depicts the attention map of the "1990s website" class. The visualizations suggest that the margins (the white areas) on the right side and the bottom of the image are the most important parts for this class. By "most important," we mean the part of the image that is most influential for the classifier in reaching a conclusion that the design is from the 1990s. Thus, one possible interpretation is that the network believes margins are a key marker for 1990's-era websites.

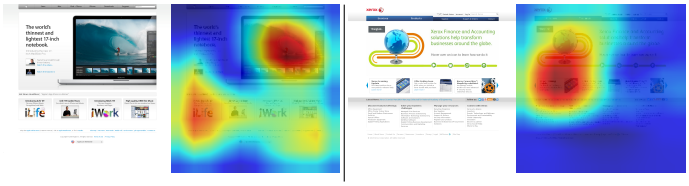
In contrast, Figure 4(b) shows that the attention map of the "2009-2013" class has multiple areas of visual importance. Many websites in this class encompass the entire browser frame; we surmise that an image that fills up the frame is important for the network. Interestingly, the visualization also points to a growing importance in designs that prominently feature computer technologies (e.g., laptops, tablets, and smartphones). For instance, in `apple.com` we see that their products increasingly take up a larger proportion of a website's space over time (Figure 5).

4.1.2 Features for Genre. In recognizing a website's genre, it appears that specific objects get the attention of the network. For



Adobe.com in 1997 Apple.com in 1999

(a) 1990's class



Apple.com in 2009 Xerox.com in 2013

(b) 2003-2013 class

Figure 4: Heat maps showing the importance of various image features in reaching the conclusion that the websites are from (a) the 1990's and (b) 2009-2013.

2001 2004 2008 2013

Figure 5: Changing product sizes in `Apple.com`'s design through time.

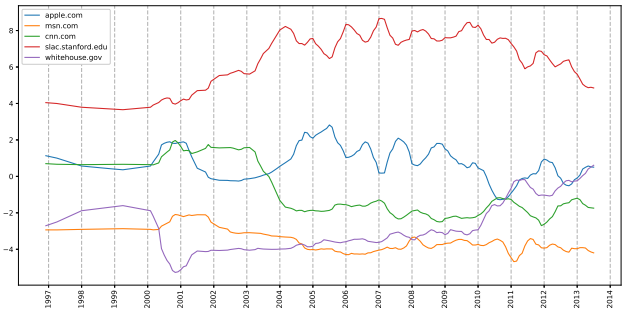
`bmw.com.tr` `browneyedbaker.com`

Figure 6: A website from `automotive` class (left) and a website from `food and drink` class (right), and the network's attention for recognizing the classes.

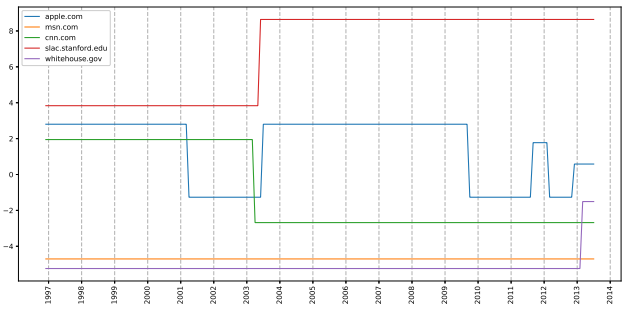
example, our sports website classifier seems to focus on sports objects such as basketballs, baseballs, bicycles, and so on. This same observation follows for the food and drink, automotive, home and garden, pets and animals, and adult genres. Figure 6 shows some examples.

4.2 Web Designs Across Eras

We characterize the changing appearance of a website over time by comparing its appearance to a small number of exemplar sites. We chose five websites as our exemplars: `apple.com`, `msn.com`, `cnn.com`, `slac.stanford.edu`, and `whitehouse.gov`. We chose these as "canonical" pages because they are regarded as leaders in the technology and design industry. For instance, `apple.com` and `msn.com` are regarded by some as interaction design pioneers. One hypothesis is that, just as their products have been influential, their website designs may have influenced other organizations' websites.



(a) Raw similarity scores

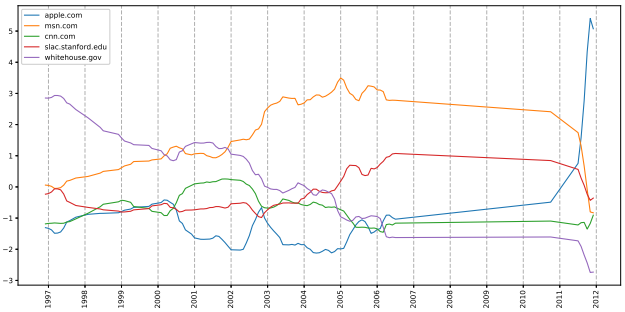


(b) Sudden moves of similarity, as identified by an HMM

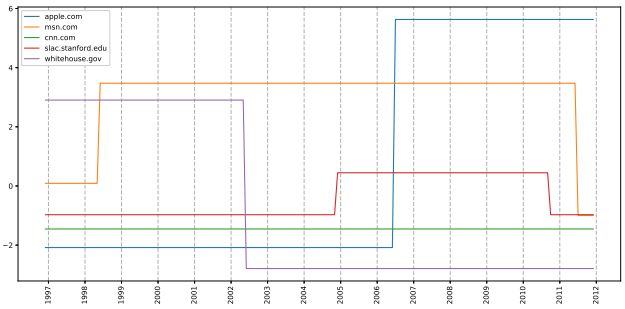
Figure 7: Similarity of `indiana.edu` to each canonical site.

For example, Figure 7 shows the similarity of one particular academic website, `indiana.edu`, to these fives sites over time. We see that `slac.stanford.edu` (another academic website) is the most similar website among our five exemplars. This is intuitive since a university's web design should most resemble another university's website versus websites of companies in the technology and design industry. (Note that in all of our plots, the *absolute* y-axis units are not really meaningful for either the raw similarity scores or the HMM outputs, so we focus on examining the relative values and trends over time.)

The web portal `yahoo.com`, founded in 1994, has a particularly long history. Since `yahoo.com` and `msn.com` are both portal websites, we might expect the two to be alike across eras, and Figure 8(a) verifies this similarity for nearly six years, until 2010 when it became more similar to `apple.com`. This behavior is even more apparent in the results smoothed by the Hidden Markov Model in Figure 8(b). Note that our dataset is missing snapshots from `yahoo.com` between 2006 and 2010. This explains why the lines have no fluctuation during this period on both graphs in Figure 8. Yahoo is also judged to be similar to `whitehouse.gov` initially, but

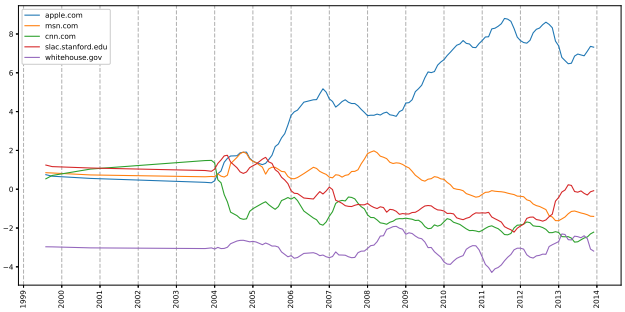


(a) Raw similarity scores

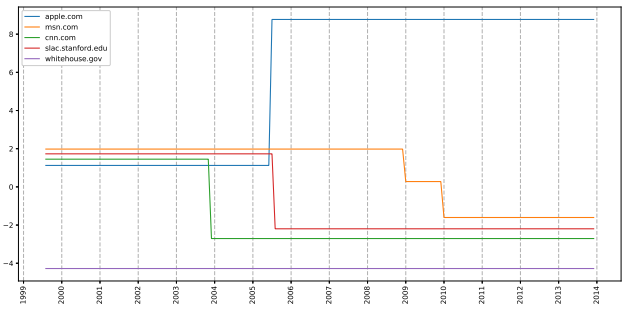


(b) Sudden moves of similarity, as identified by an HMM

Figure 8: Similarity of `yahoo.com` to each canonical site.



(a) Raw similarity scores



(b) Sudden moves of similarity, as identified by an HMM

Figure 9: Similarity of `amazon.com` to each canonical site.

loses that similarity over time. This is consistent with the observation that websites of some genres initially looked like each other but became more diverse as the web evolved.

Figure 9(a) repeats this analysis for `amazon.com`. We observe that similarity to `apple.com` has increased while it has become less similar to the other websites. Figure 9(b) starkly shows that this move began shortly before 2004. This phenomenon may be because both websites have followed the design trend of featuring large product images, which caused the network to recognize them as very similar to each other.

Lastly, to give a sense for how design on the web has changed in general, we took the "basket" of 35 pages in our dataset and compared all of them (aggregated together) to the five canonical pages. Figure 10 shows that websites were similar to Microsoft for nearly eight years but Apple gradually became dominant after 2005. We may speculate that the visual designs of Microsoft's website was globally representative or influential of the web, but after some time the designs of Apple better represented the state of visual design on the web.

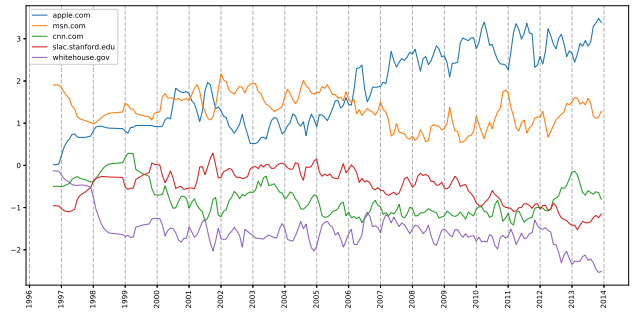
4.3 Multi-Genre Classification

Categorization of websites into particular genres is not an exact science. Many websites (especially contemporary ones) now serve a number of different purposes: `bloomberg.com` could be easily categorized as, for example, a financial, news, or media website. Thus instead of requiring the classifier to produce a single genre estimate for each page, we ask it for five hypotheses along with the confidence of each class. These confidences reflect the extent to which the network believes a given site belongs to a given genre. For example, confidences of the network for seven categories and seven classes are shown in Figure 11. If we were to take the highest confidence genre, our classifier yields some good results: shopping for `amazon.com`, computer for `apple.com`, adult for `pornhub.com`, and finance for `bloomberg.com`. More importantly, however, we also observe some intuitive confidence values for multiple genres: `espn.com` has high scores for both sports and news, for example.

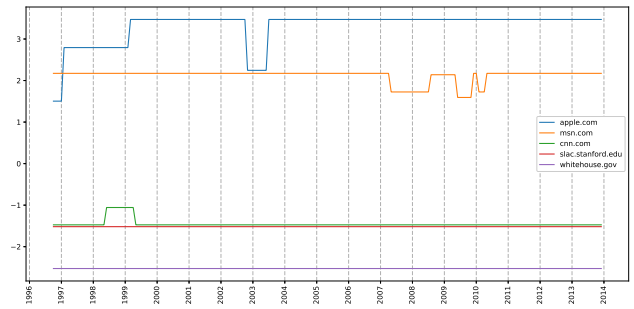
Thus, a strength of our techniques is in showing a more nuanced view on which *multiple* genres are signaled by the visual designs of websites. That is, we may speculate that there is probably something inherent in visual designs that suggests certain genres and that, more intriguingly, these designs can also suggest a mixture of genres.

4.4 Generating New Website Designs

The utility of a deep network lies in its ability to uncover patterns to classify images. Yet one innovative function of a trained network is its power to use its learned features to "hallucinate" an entirely new image based on the snapshots in the training data. Figure 12 shows images generated by the network trained on snapshots of about 17,000 contemporary websites. Each image is generated by a random point in the similarity space. We observe that the new images "look" a lot like real designs. This is a highly subjective observation; nevertheless, the method could potentially give designers not only a tool to analyze which features are important for each genre or each design era, but also a framework to find inspirations for new designs. For example, such a network can generate multiple images



(a) Raw similarity scores



(b) Sudden moves of similarity, as identified by an HMM

Figure 10: Similarity of all 35 websites in our corpus, aggregated together, to each of our 5 canonical sites.

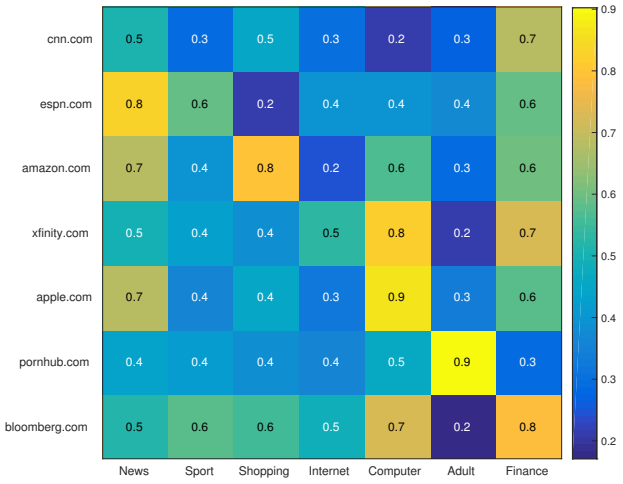


Figure 11: Confidence of the network for seven websites on seven genres.

as templates for designing a news or shopping website, which in turn could serve as a basis for designers to creatively build their own website.

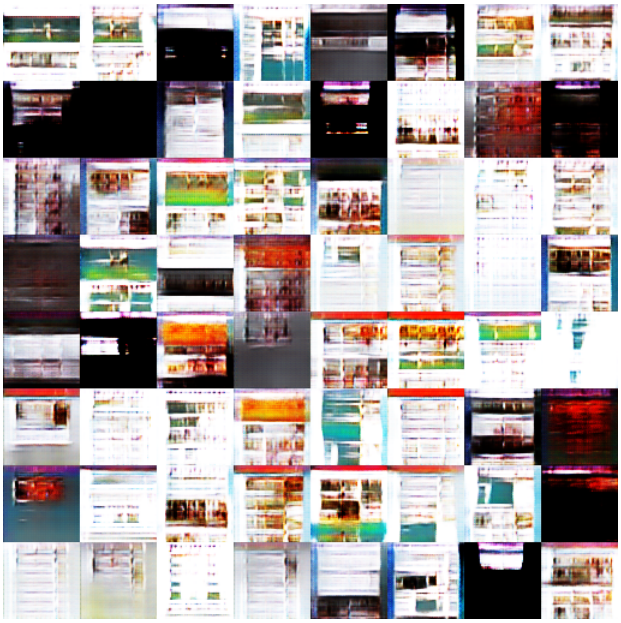


Figure 12: Websites generated by our GAN network.

5 CONCLUSION AND FUTURE WORK

In recent years, deep learning and large-scale datasets have helped artificial intelligence meet or surpass human abilities in some narrow tasks [18]. It remains a question to what degree artificial intelligence can support and complement the arts and humanities. In this paper, we take first steps towards answering, computationally, a question we often ask ourselves when viewing any visual object created by people: irrespective of its context (e.g., creator, history, circumstances) can we find real meaning in the image itself? In some respects, we might claim that our algorithm adheres to the philosophy behind formalism [7], that what matters in art is contained purely in the art piece itself. Thus, we believe our work represents an application of computer vision and machine learning that takes notions of art and design seriously. The techniques we introduce in this paper could complement and add to extant theories in the humanities.

We have begun to develop techniques that researchers can use to study large-scale collections of web pages, from a design perspective, without analyzing all pages manually. In particular, we have shown how CNNs could provide quantitative measures of website design styles, how to compare these designs, chart their evolution over time, and find transition points among these noisy signals. We have also showed how novel designs could be sampled from the CNN. Future work can include a more detailed analysis of the feature(s) being cued on by the CNN, as well as application to larger and broader datasets. We hope that our work can be a step towards developing practical tools for Cultural Analytics and some day lead to finding patterns and insights into the study of culture that would not be possible to find through human analysis alone.

ACKNOWLEDGMENTS

We thank Wen Chen and Mingze Xu for early development of our HMM for temporal smoothing. This work was supported in part by the National Science Foundation through CAREER grant IIS-1253549 and Nvidia, and used the compute facilities of the FutureSystems Romeo cluster which is supported by Indiana University and NSF RaPyDLI grant 1439007.

REFERENCES

[1] 2017. Apple Inc. WebKit. http://www.webkit.org. (2017).
[2] 2017. Computer & Video Game Archive | U-M Library. https://www.lib.umich.edu/computer-video-game-archive. (2017).
[3] 2017. Data For Everyone Library. https://www.crowdflower.com/wp-content/uploads/2016/03/URL-categorization-DFE.csv. (2017).
[4] 2017. The Internet Archive. http://www.archive.org. (2017).
[5] 2017. PhantomJS. http://www.phantomjs.org. (2017).
[6] Arram Bae, Doheum Park, Juyong Park, and Yong-Yeol Ahn. 2016. Network Landscape of Western Classical Music. *Leonardo* 49, 5 (March 2016), 448–448.
[7] Clive Bell. 1914. *Art.* Frederick A. Stokes Company, New York.
[8] Wen Chen, David J. Crandall, and Norman Makoto Su. 2017. Understanding the Aesthetic Evolution of Websites: Towards a Notion of Design Periods. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. ACM, New York, NY, USA, 5976–5987. DOI: http://dx.doi.org/10.1145/3025453.3025607
[9] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. 2009. Imagenet: A large-scale hierarchical image database. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 248–255.
[10] Ali Jahanian, Phillip Isola, and Donglai Wei. 2017. Mining Visual Evolution in 21 Years of Web Design. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems (CHI EA '17)*. ACM, New York, NY, USA, 2676–2682. DOI: http://dx.doi.org/10.1145/3027063.3053238
[11] 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.
[12] 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 (CHI '13)*. ACM, New York, NY, USA, 3083–3092. DOI: http://dx.doi.org/10.1145/2470654.2466420
[13] Yann LeCun, Leon Bottou, Yoshua Bengio, and Patrick Haffner. 1998. Gradient-based learning applied to document recognition. *Proc. IEEE* 86, 11 (Nov 1998), 2278–2324. DOI: http://dx.doi.org/10.1109/5.726791
[14] Stefan Lee, Nicolas Maisonneuve, David Crandall, Alexei Efros, and Josef Sivic. 2015. Linking past to present: Discovering style in two centuries of architecture. In *IEEE International Conference on Computational Photography (ICCP)*. 1–10.
[15] David G Lowe. 1999. Object recognition from local scale-invariant features. In *IEEE International Conference on Computer Vision (ICCV)*, Vol. 2. IEEE, 1150–1157.
[16] Lev Manovich. 2016. The science of culture? Social computing, digital humanities and cultural analytics. *The Datafied society: social research in the age of big data.* (2016).
[17] Lev Manovich, Jeremy Douglass, and William Huber. 2011. Understanding Scanlation: How to Read One Million Fan-Translated Manga Pages. *Image & Narrative* 12, 1 (2011), 206–228.
[18] John Markoff. 2015. A Learning Advance in Artificial Intelligence Rivals Human Abilities. https://www.nytimes.com/2015/12/11/science/an-advance-in-artificial-intelligence-rivals-human-vision-abilities.html. (Dec. 2015).
[19] Maxime Oquab, Leon Bottou, Ivan Laptev, and Josef Sivic. 2014. Learning and transferring mid-level image representations using convolutional neural networks. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 1717–1724.
[20] Xiaoguang Qi and Brian D. Davison. 2009. Web Page Classification: Features and Algorithms. *ACM Comput. Surv.* 41, 2, Article 12 (Feb. 2009), 31 pages. DOI: http://dx.doi.org/10.1145/1459352.1459357
[21] Alec Radford, Luke Metz, and Soumith Chintala. 2015. Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks. *CoRR* abs/1511.06434 (2015). http://arxiv.org/abs/1511.06434
[22] Katharina Reinecke and Krzysztof Z. Gajos. 2014. Quantifying Visual Preferences Around the World. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '14)*. ACM, New York, NY, USA, 11–20. DOI: http://dx.doi.org/10.1145/2556288.2557052
[23] 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 (CHI '13)*. ACM, New York, NY, USA, 2049–2058. DOI: http://dx.doi.org/10.1145/2470654.2481281
[24] Holly Rushmeier, Ruggero Pintus, Ying Yang, Christiana Wong, and David Li. 2015. Examples of challenges and opportunities in visual analysis in the digital humanities. In *SPIE/IS&T Electronic Imaging*. International Society for Optics and Photonics, 939414–939414.
[25] Maximilian Schich, Sune Lehmann, and Juyong Park. 2008. Dissecting the Canon: Visual subject co-popularity networks in art research. In *European Conference on Complex Systems*.
[26] Bolei Zhou, Aditya Khosla, Agata Lapedriza, Aude Oliva, and Antonio Torralba. 2016. Learning Deep Features for Discriminative Localization. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2921–2929.