Mediating Color Filter Exploration with Color Theme Semantics Derived from Social Curation Data

ZIMING WU, Hong Kong University of Science and Technology, Hong Kong SAR
ZHIDA SUN, Hong Kong University of Science and Technology, Hong Kong SAR
TAEWOOK KIM, Hong Kong University of Science and Technology, Hong Kong SAR
MANUELE REANI, University of Manchester, United Kingdom
CAROLINE JAY, University of Manchester, United Kingdom
XIAOJUAN MA, Hong Kong University of Science and Technology, Hong Kong SAR

Despite the popularity of photo editors used to improve image attractiveness and expressiveness on social media, many users have trouble making sense of color filter effects and locating a preferred filter among a set of designer-crafted candidates. The problem gets worse when more computer-generated filters are introduced. To enhance filter findability, we semantically name and organize color effects leveraging data curated by creative communities online. We first model semantic mappings between color themes and keywords in everyday language. Next, we index and organize each filter by the derived semantic information. We conduct three separate studies to investigate the benefit of the semantic features on filter exploration. Our results indicate that color theme semantics constructed through social curation enhance filter findability, and provide important evidence with regard to using the wisdom of the crowd to improve user experience with image editors.

CCS Concepts: • **Human-centered computing** → **Collaborative and social computing design and evaluation methods**;

Additional Key Words and Phrases: Color filter; color theme semantics; data-driven design; social curation

1 INTRODUCTION

Digital filters which are used on photos to perfect their appeal have become increasingly popular among social media users. Instead of tweaking individual image parameters manually, users can now dramatically transform a photo, for instance making it appear richer, softer or dated, with a simple click. Interestingly, research has shown that filtered images can affect viewers' interpretations, judgments, attitudes, and memories of the visual content [30], and most importantly, can attract

Authors' addresses: Ziming Wu, Hong Kong University of Science and Technology, Hong Kong SAR, zwual@connect.ust.hk; Zhida Sun, Hong Kong University of Science and Technology, Hong Kong SAR, zhida.sun@connect.ust.hk; Taewook Kim, Hong Kong University of Science and Technology, Hong Kong SAR, tw.kim@connect.ust.hk; Manuele Reani, University of Manchester, United Kingdom, m.reani@manchester.ac.uk; Caroline Jay, University of Manchester, United Kingdom, Caroline.Jay@manchester.ac.uk; Xiaojuan Ma, Hong Kong University of Science and Technology, Hong Kong SAR, mxj@cse.ust.hk.

Input Image	Filtered Results

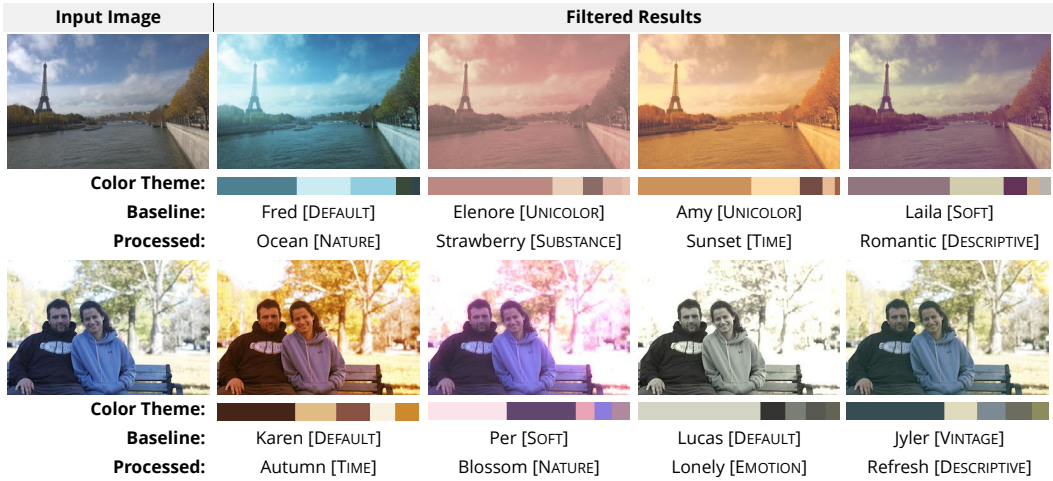|			
Color Theme:			
Baseline: Fred [DEFAULT]	Elenore [UNICOLOR]	Amy [UNICOLOR]	Laila [SOFT]
Processed: Ocean [NATURE]	Strawberry [SUBSTANCE]	Sunset [TIME]	Romantic [DESCRIPTIVE]
Color Theme:			
Baseline: Karen [DEFAULT]	Per [SOFT]	Lucas [DEFAULT]	Jyler [VINTAGE]
Processed: Autumn [TIME]	Blossom [NATURE]	Lonely [EMOTION]	Refresh [DESCRIPTIVE]

Fig. 1. Examples of the baseline filter names and CATEGORIES compared with the results generated by our framework. The first line below the image presents the color themes regarding each filter. The second line shows the baseline, *i.e.*, the filter names and CATEGORIES of a commercial application, *Pixlr* , while the third line introduces the names of filters and CATEGORIES generated by the framework we are proposing. The "input image" presents the original images, and here we take a landscape and a portrait as an example. The filter effects in "filtered results" are also borrowed from *Pixlr*.

more clicks, likes and comments online [6]. Many social media platforms, including *Instagram* and *Snapchat*, have integrated photo filter features into their services.

Most of the existing photo filters on the market are crafted by designers and named anecdotally [35]. For example, one of the most famous *Instagram* filters, "Juno", was named after one of the company's co-founder's dog, and "Kelvin" denotes Lord Kelvin, a scientist studying temperature [1]. Generally, the way in which designers organize and label color filters is not based on the preferences and experiences of the general public. As the stories behind the creation of these filters are largely unknown to users, they may find recognizing and memorizing the visual effects of filters by their names difficult and counterintuitive. It follows that when users need to locate a photo effect, they tend to click on the available filters one by one. This may be manageable when the size of the filter pool is relatively small. However, when interacting with more advanced photo editors that have a larger selection of filters, current popular filter naming and organization schemes may be detrimental to the user experience.

With the recent success of automatic image processing techniques [35, 54], an increasing number of computer-generated filters are being introduced into photo applications to transform ordinary shots into stunning yet authentic artworks (*e.g.,* the *Smart Looks* feature in *Photoshop*). While the aesthetic achievement may be enjoyable, this approach poses new challenges for the user, as the algorithmic filters come in significant quantities and are largely unnamed. Although some researchers have tried to alleviate the problem by recommending filters based on image content [31] or image aesthetics [53], how to better facilitate free exploration and targeted search in a vast filter repository has not yet been fully explored. In particular, previous research has failed to exploit the full potential of the wisdom of the crowd for color filter design.

In this paper, we investigate using color theme semantics to enhance the findability of color filters, supporting the exploration of a large filter space along the semantic dimensions. In this context, findability measures the efficiency with which users locate a target and is an important component

of the user experience with photo editors [15]. Previous research suggests that organizing data into conceptual dimensions can provide better findability, preventing users getting lost in complex data [55]. Inspired by these findings, we present a data-driven approach to semantically organizing color filters based on a knowledge extraction method applied to a large set of socially curated data. Leveraging social curation allows us to fully exploit the wisdom of the crowd [23, 45], and specifically, to model color filter semantics from a crowd perspective. In the current investigation, we first generate mappings between color themes (*i.e.,* a finite set of colors, usually three or five, in order) and verbal descriptors from an ordinary vocabulary using a probabilistic model constructed from online design material curation sites, including *COLOURlovers* [3] and *Adobe Kuler* [2]. We augment the model with social engagement information (*e.g.,* favorites) obtained from the curation sites of these user-named color themes, to reflect crowd consensus on their semantic relevance [43]. Next, we organize the verbal descriptors into a two-level structure based on semantic association, as a way to index and categorize the corresponding color filters. The results of three user studies show that color filters equipped with semantic information provide better findability and conceptual support for filter exploration. In particular, we demonstrate that the generated filter semantics allow for better filter effect interpretation, intention expression, and navigation management for both casual and expert users.

The main contributions of this paper are as follows:

(1) To enhance digital filter findability, we repurpose social curation data to model and create semantic indices of color filter effects that reflect crowd preferences and opinions. This enables users to navigate filters along semantic dimensions that are both familiar and intuitive.

(2) We conduct three user studies to demonstrate the usefulness and effectiveness of color filter semantics for mediating filter exploration. The results show that the semantic naming and organization generated by our approach enhance filter findability, and that the effect is stronger for casual users than for experts.

(3) We provide insights into the opportunities and challenges of repurposing socially curated data to facilitate data-driven design and describe their implications for future design.

The rest of the manuscript is organized as follows. We first introduce the related work in Section 2. Then, Section 3 describes a two-step preliminary study. This includes both the analyses of conventional applications and an online user survey to grasp users' needs for better filter findability. We propose our computational model based on large scale socially curated data to semantically index and organize filters in Section 4. The generated filter semantics are evaluated in two phases: a small-scale laboratory experiment with an existing application (Study II in Section 5) and a large-scale online study with a web-based mock-up (Study III in Section 6). Finally, we discuss design considerations and present the lessons learnt from using social curation data in Section 7. We end the manuscript with a brief conclusion presented in Section 8.

2 RELATED WORK

2.1 Photo Filtering

Digital filters achieve visual effects on photos by automatically adjusting image parameters, such as chrominance, contrast, and vignetting [6, 35]. Some applications like *Picasa* even provide filters that can render an image into different artistic styles [16]. Previous studies have shown that digital filters can improve the attractiveness, interestingness, and uniqueness of a photo [6]. Photo filtering can also influence viewers' perceptions by adding a new layer of context to pictures [11, 48]. Though it is argued that people prefer unfiltered photos in some cases [6, 49], the analysis of *Flickr* data indicates that, statistically, a filtered photo attracts more comments and viewers [6]. While these studies focus on users' perceptions of filter use, there is a lack of research on helping users to explore

a vast filter space, especially when provided with an increasing number of available algorithmic filters [31].

2.2 Color Semantics

Previous research shows that visual stimuli and attributes (*e.g.,* color) have an embodied meaning and depend on the aesthetic properties of the stimulus/attribute itself [19, 44]. However, visual stimuli and attributes can also have a referential meaning which is generated by a set of semantic associations learned through experience [28]. Ecological Valence Theory (EVT) states that individual preferences for colors are linked to people's affective responses to color-associated objects [46]. In other words, people prefer colors that are associated with the objects they like. As colors appear to carry contextual information through color-object association, diffusion of online shared images can be significantly impacted by color [5]. In addition, the experience that users have when using color filters to edit an image may be enhanced by the use of appropriated color tones based on emotional signals [6]. Moreover, research in the field of psychology has shown that people are generally faster at recognizing words describing objects when these are presented in congruent colors (*e.g.,* the color yellow with the word banana) [20]. The opposite effect is also true. As demonstrated by the Stroop effect, the identification of colors is greatly impaired (*e.g.,* longer reaction time) if the name in which the colored word is written is inconsistent with the actual color (*e.g.,* the word "red" written in green) [36]. Finally, there is also evidence that the effect of colors can be altered by attached textual information [5]. As the perception of colors is related to their semantics, the recognition and utilization of color filters may be enhanced by the use of an appropriate semantic structure; the potential for this is investigated here.

2.3 Modeling Color Semantics

Prior studies associate linguistic concepts to colors primarily by modeling their statistical distribution from a collection of data [8, 19]. Heer *et al.* [19] utilized crowd-sourced color name judgments to develop a probabilistic model for color naming. This was extended to find category-associated colors [51].

A color theme, which is the composition of multiple colors [54], may be considered complicated when analyzing its semantics for two main reasons. Firstly, a color theme can be annotated with different names based on a given context [24]. Secondly, different users can perceive a given color theme subjectively, *e.g.,* in line or not in line with their preferences [43]. Previous work aimed at associating names and color themes was performed either by controlled user experiments or by learning strategies automatically. In the former case, the pioneering work comes from Kobayashi's Color Image Scale [27]. He organized 1170 three-color combinations into 2D space and matched them to 180 predefined keywords through crowd-sourced experiments. While it has been further extended by Solli *et al.* [52] to fit any color combinations, it does not support the addition of further descriptive keywords. For the latter case, different data sources are leveraged for learning color theme semantics. For instance, some studies take advantage of annotated images by mapping their annotations to the color theme of the corresponding images [21, 34, 35]. However, the ambiguity of image annotations can lead to the content conflation problem. That is, an image annotation can be matched to the image content rather than the color theme, *e.g.,* an image annotated with "happy" probably has a happy face instead of a happy color tone. Other works like [14, 41] utilized the data shared in online communities, e.g. *COLOURlovers* , to extract the semantics by classifying them into several predefined keywords. We also harness social curation data, but use a probabilistic model to provide a scalable means of mapping color themes to any existing keywords. More recently, Jahanian *et al.* [24] studied color theme semantics by linking the topic of magazine covers to the topic of color themes, extracted from those covers. However, this method investigates the

semantic association solely from the perspective of the designers rather than that of the actual users. Uniquely, we construct the color theme semantics computationally from the social curation of color themes to reflect crowd consensus.

2.4 Leveraging Semantics

Semantic information has been successfully exploited to provide navigation to users exploring a large collection of data [7, 55]. For example, Yee *et al.* [55] proposed hierarchical faceted meta-data to allow image searching along conceptual dimensions, suggesting that proper semantic navigation can enhance feelings of discovery "without being lost". Using crowd-sourced data, O'Donovan *et al.* [44] designed a tool that allows users to select a font type by using high-level attributes, such as "dramatic" and "graceful". There are also studies on supporting image enhancement according to a linguistic concept. [29, 34, 35]. For example, Liu *et al.* [35] used a collection of images whose annotation matched an input keyword for color transfer. Similarly in [29], annotated images were utilized to learn a model that supports high-level image editing. However, all these studies examine only how to adjust an image according to a given keyword, without understanding how semantic information benefits user exploration. The study bearing the closest resemblance to ours is that of Lin [32], which suggests that chart reading speed is linked to semantically-resonant colors in visualization. In contrast to that, we investigate how color theme semantics can facilitate color filter findability.

2.5 Social Curation

Mapping color categories and filters to names that convey strong emotional signals and semantics is challenging, as the mapping must be meaningful to the majority of users. A potential way of overcoming such an issue is to construct a model by aggregating crowd preferences and opinions. For instance, in the case of mapping single colors to names, previous research has utilised crowd-sourcing methods [13, 19, 40, 42] and, more recently, text analysis techniques on large Google datasets [51]. Nevertheless, a crowd-sourcing approach could potentially be extended to the case of mapping more complex constructs, such as color filters, which categorize and organize a large set of colors using semantics. An important method for extracting knowledge from the crowd is social curation, a collaborative sharing of Web contents organized around one or more themes or topics. One of the aims of social curation is to aggregate user feedback and opinions into useful knowledge that can be used for a wide variety of objectives. Social curation is emerging as a new way to extract collective knowledge from social media platforms for several reasons [23]. An advantage of using social curation, compared with methods that extract knowledge from search engines, is that social curation is usually conducted on relatively smaller and more focused data sets, which are already pre-filtered. This provides superior search results in terms of relevancy [45]. Moreover, social curation can be successfully used for tasks as complex as user profiling and image context discovery [17, 26]. This stream of research indicates that curated content is potentially richer than content automatically generated by an algorithm [26]. However, there are some challenges linked to using social curation data, such as information completeness, context dependency and the nature of the data itself [9]. While these studies show the value of using knowledge from social curation, there is little understanding of how to repurpose curated online data, especially for facilitating data-driven design. Thus, the implications of this work extend beyond color filter design.

3 STUDY I: PRELIMINARY STUDY

The objective of the first study was to capture users' preferences, attitudes, and opinions about photo filtering to derive design needs. To identify these factors, we first scrutinized features such as filter

Applications	# filters	Increasable	Layout	Categories	Filter Names
Facebook	7	No	Juxtaposition	N/A	Vintage, Spring
Flickr	11	No	Juxtaposition	N/A	Tender, Pigeon
Snapseed	11	No	Juxtaposition	N/A	Portrait, Accentuate
VSCO	10	Yes	Juxtaposition	N/A	B1, B5
MOLDIV	31	Yes	Categorization	Basic, Natural	Happy, Lovely
Meitu	51	Yes	Categorization	Lomo, Artistic	Norway, Ruby
PhotoWonder	58	Yes	Categorization	Portrait, Classic	Natural, Elegant
Pixlr	112	No	Categorization	Too Old, Subtle	Agnes, Alex
B612	117	No	Customization	N/A	Clean, Loveletter
Instagram	40	No	Customization	N/A	Clarendon, Gingham
SNOW	41	No	Customization	N/A	Clean, Loveletter

Table 1. Each row shows the condition of each application. *Increasable* means that users can add more photo filters through in-app purchase. The *Categories* and *Filter Names* columns each present two examples for the given application.

names and filter organization of the existing popular photo filter applications. We then conducted a survey to learn how users interact with those features and their effect on user experience.

3.1 Analysis of Conventional Photo Filter Applications

We tabulated three social media applications (*i.e., Facebook, Flickr,* and *Instagram*), which are popular for their photo sharing features, and added eight further photo filter applications. All the chosen applications have been on the app store's top favorite lists in the "Camera & Video" category for more than three months. As shown in Table 1, the number of photo filters embedded in each application varies, ranging from 10 to over 100 photo filters. Furthermore, some of these applications, including *VSCO, Meitu, MOLDIV,* and *PhotoWonder* allow users to increase the number of photo filters through in-app purchases.

There are three different ways in which one can enumerate photo filters: juxtaposition, customization, and categorization. These methods enable users to select the best candidate from a large number of photo filters. Some applications, especially those which have around ten photo filters (*e.g., Facebook, Flickr, Snapseed,* and *VSCO*) merely juxtapose each of the filters one by one. More sophisticated applications with a larger number of filters (often more than 30) are divided into two main families; one family of applications has fixed categories, and the second family allows users to customize the layout of given filters.

Compared with *Facebook* and *Snapseed*, which randomly juxtapose photo filters, *Flickr* and *VSCO* have a color bar above each filter's thumbnail icon which shows the relevant effect. The color bar may alleviate users' misperception of filters, arising from unrepresentative filter names, by giving additional perceptual clues to users. *MOLDIV, Meitu, PhotoWonder,* and *Pixlr* all provide thumbnails presenting the filter effects. The thumbnails allow users to preview the results of the filter application. *MOLDIV* has representative images for categories, *e.g.,* a landscape picture for the "Basic" category and a portrait for the "Natural" category. Those images should (theoretically) help users choose the desired category based on the image content. *Pixlr* displays filter previews together with their labels. These labels are based on common names, *e.g.,* "Jessica", "John" and "Camilla" for the color category "Soft". Lastly, in *B612, Instagram,* and *SNOW,* users can customize the arrangement of photo filters according to their own preferences. They can select the desired
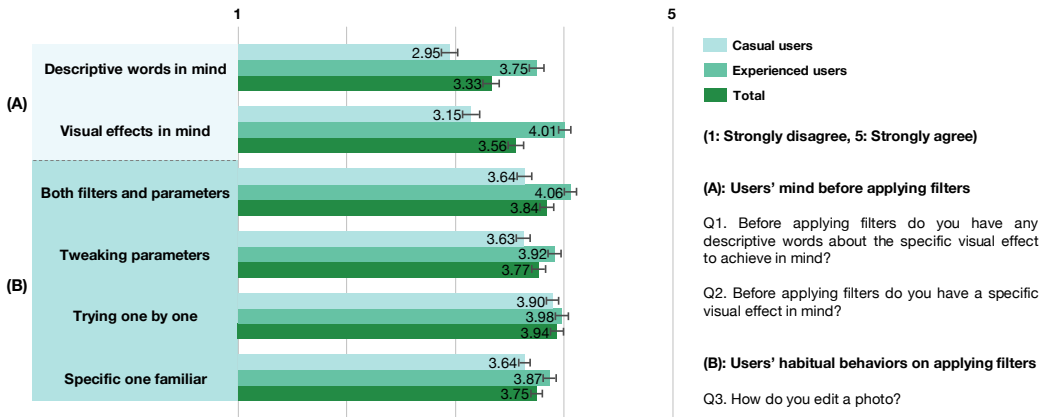
Fig. 2. The results of the preliminary user survey. (A) are the responses to Q1 and Q2, while (B) are the responses to Q3. The questionnaires are arranged on a 5-point Likert scale. The error bars show the standard errors.

filters from a comprehensive list, which is provided at the end of the filter array. The chosen ones will then appear on the filter array.

3.2 Online User Survey

3.2.1 Method. We conducted an online survey using Amazon Mechanical Turk (MTurk) [1]. We chose MTurk as it is widely used to study general users' behaviors, and there is evidence that data collected using this method is of a similar quality to data collected using other methods [37]. Without referring to any particular application, the survey asked questions, on a 5-point Likert scale, about how respondents use photo filtering. After filtering out 78 respondents who answered *"Never"* to the question *"How often do you apply photo filter?"*, those respondents who gave nonsense type-in answers, and those who completed the survey in less than 3 minutes (indicating poor engagement with the survey), we obtained 225 valid participants (114 males and two who did not disclose their gender). Then, we split the respondents into two groups: Casual users and Experienced users, by asking how familiar they were with photography and photo editing.

In total, there were five questions. We first asked whether users had a particular word in mind describing the filter they intended to apply. The second question asked whether users had a particular visual effect in mind representing the filter they intended to apply. These two questions revealed the users' mental process before he/she applied filters to an image, measured on two distinct dimensions: Verbal and Visual.

3.2.2 Results. The results from the survey (see Fig. 2. (A)), show that experienced users are more likely to think about expressive words ($Mean = 3.75, SD = 1.05$) and visual effects ($Mean = 4.01, SD = 0.75$) before applying color filters than casual users ($Mean = 2.94, SD = 1.05$ and $Mean = 3.15, SD = 1.2$ respectively). Overall, for both types of users, Verbal clues ($Mean = 3.33, SD = 1.05$) seem to evoke less than Visual clues ($Mean = 3.56, SD = 1.05$). This might be because language alone may not be able to express certain concepts and ideas which can be easily imagined using mental visualization.

---

[1]Amazon Mechanical Turk: https://www.mturk.com/

The third question asked about the way in which users actually edit images when using photo filtering applications. The responses were grouped into four different categories (see Fig. 2. (B)). The results indicate that experienced users are more inclined to edit images both by using filters and tweaking parameters ($Mean = 4.06, SD = 0.75$), and to choose specific filters that are familiar to them ($Mean = 3.87, SD = 0.75$), compared with casual users. Overall, the most common way to use photo filter applications for the respondents was to try each of the given filters one by one.

The fourth question asked participants to report how easy it is to find the desired filter with the photo filtering application they generally use. Only 35 percent of respondents (80 responses out of total 225) answered that it is easy to find a satisfactory color filter.

The fifth question asked participants what difficulties they encountered when using color filtering applications, followed by five different options that participants had to score from 1 to 5. Overall, the top two reasons, for all participants, were that there is a weak relationship between names and color effects in certain image editors ($Mean = 3.42, SD = 0.9$) and that there are too many filters in certain editors ($Mean = 3.50, SD = 1.05$). Interestingly, for experienced users only, the top two reasons were that there are a limited number of filters in certain image editors ($Mean = 3.62, SD = 1.2$) and that there are too many filters in certain editors ($Mean = 3.61, SD = 1.5$). In this question, there was an option where participants could provide additional reasons regarding why it can be difficult to use color filters. For this we received 41 meaningful responses, of which the most common related to the fact that it might be hard to compare the effects of filters (9 responses), the filters may not be well categorized (6 responses), or the filter may make the image worse (5 responses).

3.2.3 Discussion. According to the results, it appears that different users use different strategies for applying color filters to images and these may be dependent on some mental processes that occur prior to applying those filters. Filter findability might be the link to this phenomenon, though this was not tested in this study. Furthermore, a large proportion of users find it difficult to select the desired color filter. From the participants' responses, it appears that there are several issues with how color filters are categorized and named. The results from this study were used to design study II and study III.
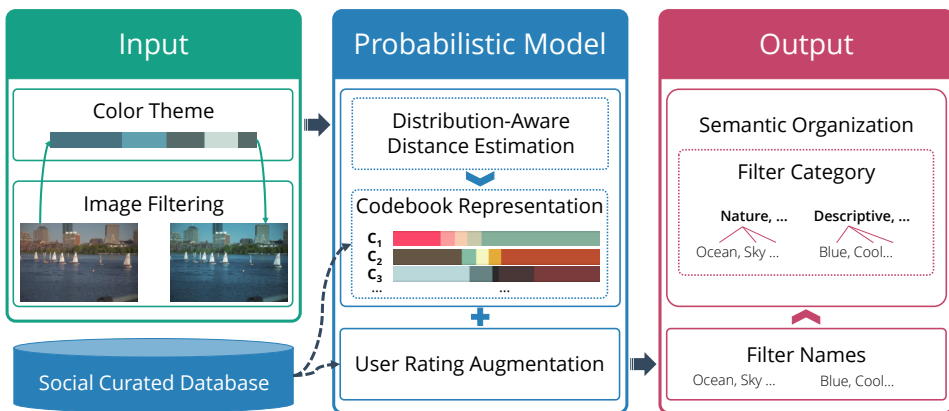
4 FRAMEWORK DESIGN

Fig. 3. The overview of the proposed pipeline for computationally indexing and organizing color effects.

The objective of our proposed framework is to efficiently support users in their search for target filters. Previous work has highlighted that organizing data into semantic dimensions assists data exploration [55], and that there are important links between color-name semantics and data reading speed [32]. Meanwhile, our intial study (Study I) reveals that filter naming and categorization in existing applications impairs filter findability. Inspired by these findings, the next part of our work aims to enhance the findability of color filters by manipulating their semantic information. This task is nontrivial as the index of color filters should be semantically relevant to their visual effects and intuitive to general users. To address the challenge, we exploit socially-curated data to extract the crowd consensus on the semantic mapping from keywords to color effects. In particular, we use socially curated data from the online community *COLOURlovers* because the user-curated names convey the perception, meanings and emotional responses of color themes [43]. Moreover, filter effects can be computationally expressed by color themes [54]. We first use a probabilistic model to construct color theme semantics by aggregating socially-curated information (as shown in Fig. 3). Then, we index each color filter by the semantics generated by the model and further organize the filters into a two-level structure via a computational scheme.

4.1 Model Color Theme Semantics

4.1.1 A Probabilistic Model for Color Theme Semantics. Leveraging the social curation data, we learn the perceptual semantics of color themes by modeling their probabilistic distribution. Heer *et al.* [19] utilized multinomial conditional probability to semantically name a color but not specifically for the composition of colors. We adapt Heer's model to extract the semantics of color themes based on the social curation data. Specifically, given a color theme c, the likelihood of c named by the keyword w can be defined as:

$$P(w|c) = \sum_{c_i \in \zeta} P(w|c_i)P(c_i|c) \tag{1}$$

where c_i is the ith color theme in the dataset ζ. $P(c_i|c)$ denotes the probability of c_i given c. When the distance between c_i and c is small enough, $P(c_i|c) \approx 1$. In this case, we can approximate $P(c_i|c)$ to $1 - \frac{dist(c_i, c)}{maxDist}$, where $dist(c_i, c)$ denotes the distance between c_i and c, and $maxDist$ is the maximum observed distance found in the dataset.

Distribution-Aware Distance Estimation. For estimating $P(c_i|c)$, we merge visually similar color themes. This requires measuring the distance between two color themes. Typically in *COLOURlovers*, users can assign different proportions for each color swatch within a theme to elicit a unique look and feeling, as shown in Fig.4). Thus, color themes composed of different-proportion colors might convey different semantics. Previous work has measured the similarity between color themes directly, by estimating the overall pairwise color distance using the L2 distance, thus failing to consider the relationship among the color swatches present within a color theme. To fully exploit the proportion information, we consider each color theme as a different distribution, and use the Earth Mover's Distance (EMD) [50] to measure the distribution difference between two color themes. EMD is a widely used metric for measuring the minimized transportation cost of transforming one distribution into another. Here, we use EMD to calculate the minimum color value to be changed for making one color theme the same as another. More specifically, each color theme c is represented as $\{(v_1, w_1), (v_2, w_2)...(v_5, w_5)\}$, where v_i is the ith color swatch of c and w_i is the proportion of v_i. Thus, given two color themes (c_x, c_y), their distance $dist(c_x, c_y)$ can be mathematically defined as:

$$dist(c_x, c_y) = EMD(c_x, c_y) = \frac{\sum_{i=1}^{5} \sum_{j=1}^{5} f_{ij}^* d_{ij}}{\sum_{i=1}^{5} \sum_{j=1}^{5} f_{ij}^*} \tag{2}$$

where f_{ij}^* is the optimized flow that minimizes the overall cost for making c_i the same as c_j. d_{ij} is the Euclidean distance between v_i of c_x and v_j of c_y. The color distance is measured in CIELAB space as it can decorrelate different color channels well [31]. The full details for calculating f_{ij}^* are omitted here as they are outside the scope of this paper, and can be found in [50].

Codebook Representation. After calculating all the pairwise distances of the collected color themes, we bag color themes using a clustering approach. We create a color theme codebook $O = \{o_1, o_2, ..., o_k\}$ by clustering all color themes based on the estimated distance using the K-centroid algorithm. Each cluster center o_i defines the center theme of the ith cluster, such that visually similar color themes are merged whereas different themes are separated. The selection of K is critical. With a small K, visually distinct color themes would be grouped into the same cluster while a large K entails a large dictionary, leading to an inefficiency of the representation. After setting $K = 4096$ (as discussed in section 4.1.2), Eq.1 becomes:

$$P(w|c) = \sum_{c_i \in \zeta \, and \, \{c_i, c\} \sim o_j} P(w|c_i)(1 - \frac{EMD(c_i, o_j)}{maxDist_{emd}}) \tag{3}$$

where $\{c_i, c\} \sim o_j$ denotes that c_i and c share the same cluster center o_j and $maxDist_{emd}$ is the maximum EMD distance observed in the dataset.

Model Augmentation by User Rating. O'Dononvan *et al.* [43] suggested that the evocative names of color themes have a significant impact on users' ratings, indicating the link between names and user ratings. Driven by this, we explore the incorporation of rating information into the probabilistic model to highlight impactful items (*i.e.,* highly rated color themes). We augment Eq. 3 with a rating term and emphasize users' preference for (c, w) with a larger λ. Then, from a probabilistic perspective, given a color theme, c can be named by:

$$\arg \max_{w} \; \lambda r(c, w) + P(w|c) \tag{4}$$

where λ is a weighting parameter and the rating term $r(c, w)$ corresponds to the intra-cluster averaged rating of the color theme c associated with w.
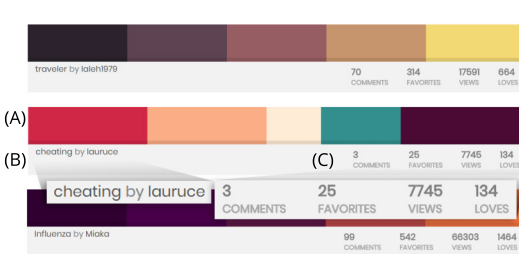

Fig. 4. Examples of color themes collected from *COLOURlovers* . (A) is the color scheme with different proportions; (B) is the color theme name provided by the author; (C) is the information gathered from the social platform.
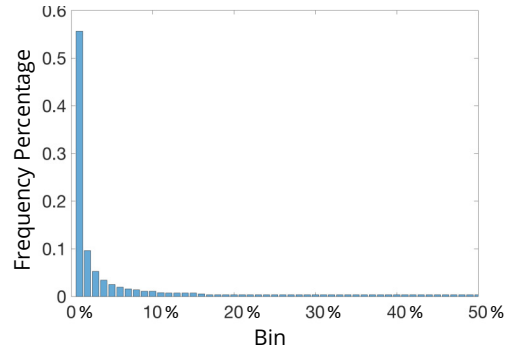
Fig. 5. The histogram of name distribution of the collected dataset.

4.1.2 Experiment. We detail the experiment as follow:

Dataset Collection. Five-color palettes are widely used to represent color themes in many online creative communities, *e.g., COLOURlovers, Adobe Kuler*[2], *Coolors*[3], *etc..* We therefore focus on 5-color themes in this work, but it should be noted that theoretically our framework is applicable to color themes with any number of colors. We collected a dataset of 5-color themes and their user-curated information from *COLOURlovers*. *COLOURlovers* is a popular social curating platform in which users create and share color themes. Each color theme (the combination of five colors, usually assigned with different proportions) is annotated with a name by the creator and rated with "LOVES" by viewers, as shown in Fig. 4 (C). We crawled all the 5-color themes together with their names, ratings and the number of views from the website. We only included the color themes named by single-word keywords but filtered out those named by numbers, stopwords, and special characters. We removed color themes with less than 5 *loves* and 10 *views* [43]. We ended up with 908,405 themes, from which we randomly chose 15% (*i.e.,* 136,261) themes as the validation set and 15% different themes as the test set for parameter selection and evaluation. Then, the remaining 635,883 themes were used for training our model. The name distribution of the collected data is shown in Fig. 5. This shows that the top 1% of the most frequent names spans 57% of the occurrences in the dataset.

Parameter Selection and Model Evaluation. The model has two hyper-parameters λ and K, the value of which can be influential on the model's performance. We chose an optimal combination by grid search, based on the Top-n metric on the validation set. Specifically, for each color theme c, we acquired n keywords with the highest possibility given by Eq.4 and checked any candidate names that match c. The percentage of the matched instances indicates the precision of the model. The results with different settings are reported in Fig.6. From these results, we set $K = 4096$ and $\lambda = 0.1$, which yield the optimal value for our measure of precision.
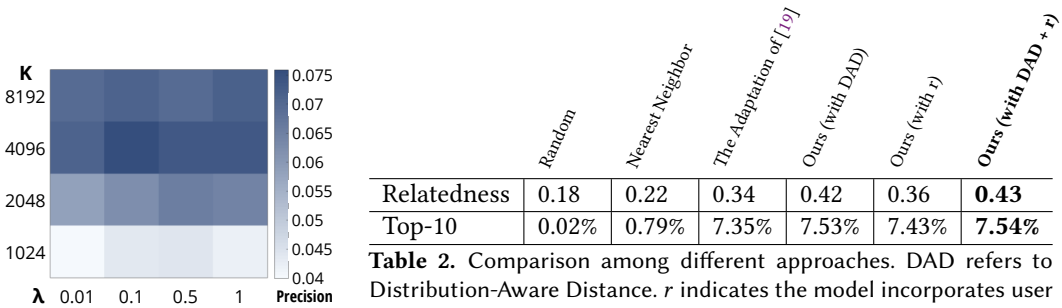

Fig. 6. The model performance with respect to different K and λ. Higher color intensity indicates better performance.

	Random	Nearest Neighbor	The Adaptation of [19]	Ours (with DAD)	Ours (with r)	Ours (with DAD + r)
Relatedness	0.18	0.22	0.34	0.42	0.36	**0.43**
Top-10	0.02%	0.79%	7.35%	7.53%	7.43%	**7.54%**

Table 2. Comparison among different approaches. DAD refers to Distribution-Aware Distance. *r* indicates the model incorporates user rating information. Relatedness is measured by the average Word2Vec similarity score between two terms. A higher value indicates better performance. Top-10 metric measures how often the correct name appears in the first ten most likely labels.

To validate the model, we tested the naming precision on the collected test set by using the Top-10 metric and the relatedness score. As described previously, the Top-10 metric gives a positive value to the results that match the ground-truth names and a false value to words with different canonical forms regardless of the similarity of their meaning, even if they are synonyms. For example, "Happy" and "Glad" are regarded as a mismatch by the top-10 metric even though they could both be valid names for a "Happy" color theme. Therefore, following a similar process to [22], we used the

---

[2] *Adobe Kuler*: https://color.adobe.com/explore/
[3] *Coolors*: https://coolors.co/

relatedness score to take the semantic similarity of words into consideration. More specifically, we used the word2vec model [38] to embed names into vector representations. Word2vec is a word embedding model that takes a text corpus as input and outputs the corresponding vector representation. It is able to preserve the fine-grained semantic and syntactic regularities of words. The relatedness between two names is computed as the cosine similarity between their vector representations. A higher relatedness score indicates higher semantic similarity. To report the result shown in Table.2, we first demonstrated the difficulty of the task by showing random guess results (*i.e.,* candidate names are randomly chosen) and the nearest neighbor's names (*i.e.,* acquired by finding the names of the closest neighbors). Moreover, we chose Heer's probability model [19] as the baseline for our evaluation. The distinction between the original model and our extension is that we exploit the socially curated information of color themes by incorporating a user-defined color distribution and users' ratings into the model. The results can be seen in Table.2. These results indicate that our model achieves superior performance, and verifies the effectiveness of aggregating curated information. The high relativeness score of our model suggests high relevance in meaning between the generated names and the ground truths. This relevance is also verified by users' feedback in Study II and Study III.
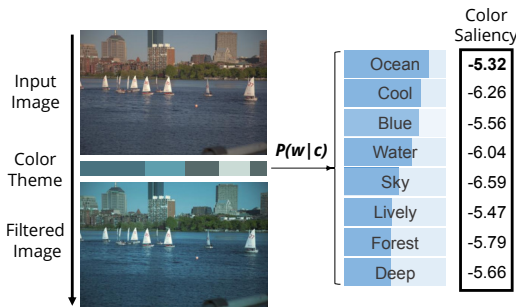

Fig. 7. An example of the semantic names generated from a color effect. The proportion of the dark blue color indicates the probability $P(w|c)$ and the corresponding saliency scores are shown on the right.
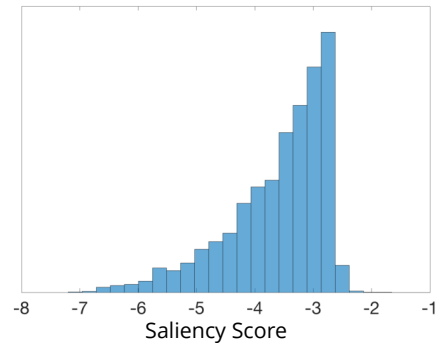

Fig. 8. The histogram of salience scores.

4.2 Organizing Color Filters by Semantics

4.2.1 Index color filters by semantic names. We name the color filters in accordance with their associated color themes on selected images. The color theme of an image with the filter f is found by its quantized 5-color composition c_f using the median cut algorithm [18]. Following this, we acquired ten candidate names with the highest semantic relevance in relation to the filter f based on Eq.4. While c_f is named according to those candidate names with the largest probability, it is possible that these names are also associated with other color themes. In these cases, the candidate names lack canonical color representations, and users may thereby find it difficult to recall the color effects represented by those words, *e.g.,* mapping the word "Deep" to certain colors. To address this issue, we choose the ones with the lowest naming ambiguity from the list of names as the final output of our model. As demostrated in [19], the color naming ambiguity of a keyword is negatively correlated with its saliency. Therefore, we choose names with high saliency. More specifically,

Saliency is used to measure the color ambiguity of keywords [19], defined as the negative entropy of $P(C|w)$:

$$Saliency(w) = -Entropy(P(C|w)) = \sum_c p(c|w) \log p(c|w) \tag{5}$$

A higher entropy (*i.e.,* lower saliency) of P(C|w) indicates that the keyword w has been used to name a larger number of color themes which are not visually similar to one another. In other words, w is a weaker canonical color representation and it would be more difficult for users to recall the color theme denoted by w. Fig. 7 showcases the generated candidate names of a blue-tone color effect with their related saliency scores listed on the right. For instance, the word "Ocean" has a higher saliency score than the word "Cool". The saliency distribution of all the collected names is shown in Fig.8. Overall, the above processes aim to ensure the generation of a reasonable and intuitive filter indexing scheme based on the social naming of color themes. In particular, we first extract color themes whose semantics are agreed upon by the majority of users (*i.e.,* that achieve the maximum likelihood in Eq.4) then, we choose names with a high saliency score to represent these color themes. This helps exclude uncommon and ambiguous names that are only meaningful to a small group of users.

4.2.2 Filter Categorization. To facilitate top-down navigation [55], we categorize the filters into semantic dimensions using a computational scheme. We first embed the single-word names of the collected color themes into vectors using Word2vec [38]. We utilize Google's word2vec model which is pre-trained on a Google News dataset[4]. The dataset consists of 100B corpus from Google News that covers broad and general themes such as location, emotion, art, *etc* [4]. The lexicon of these themes has a considerable overlap with the common vocabularies that people would use to label color themes [43]. Then with the 300-dimension vectorized features, we use k-mean clustering for categorizing names. This process ensures names with high semantic similarity, *e.g.,* "Happy" and "Glad", are merged into a cluster. We start by setting the seed $k = 5$ and increase it until the clusters are conceptually merged. For example, the word "Tree" and the word "Flower" are merged into the same cluster called "Nature". Empirically, we found that $k = 13$ yields the optimal clustering performance. Once the clustering process is complete, we remove those clusters with a relatively small number of elements in them, as well as those with semantically meaningless items, *e.g.,* "B12" and "aaa". Each cluster is named according to the higher-level category labels in WordNet [39]. Eventually, we attained eight clusters which formed our eight categories, namely "Emotion", "Human", "Nature", "Art", "Descriptive", "Substance", "Location" and "Time". If a new autumn-tone color theme comes, it would be matched to the closest center of the codebook C (see Section 4.1.1) and be labeled by the semantics (*e.g.,* 'Autumn') shared by the center according to Section 4.2.1. It would be assigned to the Nature category based on the clustering result.

5 STUDY II: IN-LAB USER STUDY

We conducted an in-lab user study to evaluate the effect of the constructed semantics on filter findability. For this, we designed a prototype application based on our previously described framework and compared it with an existing commercial application.

5.1 Method

5.1.1 Prototype Design. The prototype, named *IFil*, has 135 filters, of which the visual effects were generated by [31] and then labeled and organized by our proposed approach. The filters were organized into a two-level structure. The first level represents filter categories (*e.g.,* "Emotion"), the

---

[4]word2vec: https://code.google.com/archive/p/word2vec/

second level, nested within the first, represents specific filters that are described using semantically relevant names (*e.g.,* "Lonely") together with the filter preview (as shown in Fig.1). By default, the names are arranged in alphabetical order.

We compared this system to the commercial application, *Pixlr* [5], which is a popular photo editing product developed by *Autodesk*. We used this editor because it is one of the most highly rated (*i.e.,* 4 out of 5 stars) photo editing applications in the iOS Store [6] and it provides more than a hundred pre-defined photo filters, a number which poses a potential challenge for filter selection.Another important feature of *Pixlr* is that, in contrast with some other commercial applications, it uses filter categorization. This is directly relevant to our investigation since our model also uses categories to organize filters.

5.1.2 Task Design. It is common practice for studies examining the exploration of information to involve a browsing activity (*i.e.,* serendipitous discovery) and a searching activity (*i.e.,* look for answers to specific questions) [47]. As such, we proposed two tasks, *i.e.,* filter browsing, and filter searching:

Filter Browsing. In the browsing task, participants were asked to explore the prototype and choose their favorite filter for a given image. The image was randomly chosen from the genres of portrait, still life, and landscape respectively, collected from the FiveK dataset[7] [10].

Filter Searching. In the searching task, each participant was given an input image and a target image. Then, they were asked to modify the input image to make it as similar to the target image as possible in the shortest possible time. Again, we selected three genres (*i.e.,* portrait, still life, and landscape) of images for each category from the most popular images found in photo sharing copyright free websites[8]. Those images were filtered *a priori*.

5.1.3 Procedure. We conducted a within-subjects experiment with 24 participants (12 females, $Mean_{Age} = 20.75$, $SD_{Age} = 3.37$) recruited from a local university through poster advertisements. By asking about their experience on photo filter applications, we classified 11 participants as experts (they use such applications at least once a week); the remaining 13 participants were classified as casual users. After providing informed consent, the overall procedure of the experiment was described to the participants. The study was video-recorded and the post-task interviews were voice-recorded. After a brief training session, each participant performed the browsing and searching tasks. The order of the tasks was counterbalanced. Each task took on average around five minutes. At the end of the experiment, we conducted a post-task interview with each participant and analyzed the qualitative data emerging from the open questions to extract in-depth impressions regarding the two applications, *IFil* and *Pixlr*. On average, the overall duration of the experiment was 40 minutes. We analyzed the videos of the experiment at the end of the study to gain insight into how users behaved when interacting with both systems.

5.2 Results

5.2.1 Questionnaire Analysis. At the end of each task, participants were asked to fill in a 5-point Likert scale questionnaire (in which "1" indicates strongly disagree and "5" strongly agree) with four questions: the first asking how understandable the filter names are, the second asking how much the filter names reflect the filter effect they describe, the third asking whether the filter names are helpful in finding the desired effect and the fourth question asking whether the categorization of filters is helpful in finding the desired effect. The results of the interviews are shown in Fig. 9.

---

[5]*Pixlr*: https://pixlr.com/
[6]iOS Store: https://itunes.apple.com/us/app/pixlr-photo-collages-effect/id526783584?mt=8
[7]FiveK dataset: http://groups.csail.mit.edu/graphics/fivek_dataset/
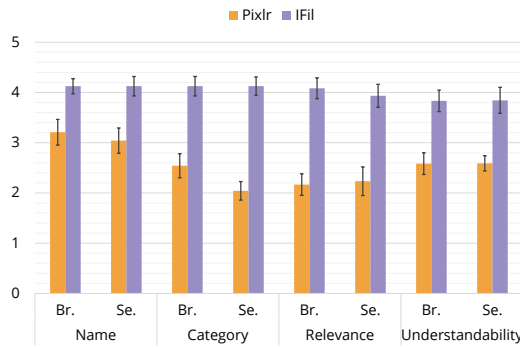[8]Pexels: https://www.pexels.com/

Fig. 9. Bar chart for the four questions in Study II. The y axis corresponds to the score obtained for each question. The bar height represents the mean value across the type of task and the two systems. Br = browsing; Se = searching. The error bar represents the standard error.

Overall, participants' responses support the superior efficacy of our framework compared to *Pixlr*. First, the filter names of *IFil* are, on average, more understandable (Q1, $Mean = 4.08, SD = 0.56$) compared with the ones used by *Pixlr* ($Mean = 2.17, SD = 0.64$). This may be linked to the use of everyday language from social curation data to construct the model. Second, the names of filters produced by our model appear to better reflect the effect they describe (Q2: $Mean = 3.83, SD = 0.82$) compared to the names used by *Pixlr* ($Mean = 2.85, SD = 1.05$). This may also be linked to the direct use of crowd opinions to extract filter names. Finally, participants perceived the filter names (Q3, $Mean = 4.13, SD = 0.91$) and category (Q4, $Mean = 4.12, SD = 0.95$) of *IFil* to be superior, in terms of how helpful they are for finding the desired color effects, compared with *Pixlr* (Name: $Mean = 2.54, SD = 0.97$, Category: $Mean = 3.04, SD = 1.08$). These results suggest that semantically indexing and organizing filters can have an impact on the ease of use and the findability of color filter effects. To validate these results, we conducted a further qualitative analysis of the open questions asked of the participants, which is reported in the next section.

5.2.2   Interview Analysis. Participants were asked what their impressions were of both systems, and if they had any comments regarding how they perceived the interaction with the two systems during the tasks. We analyzed the user feedback to gain insights into the opinions and perceptions that users had when interacting with *IFil* and *Pixlr*, especially in relation to the semantics of filters. The results for filter names and filter categories are reported separately.

Filter Naming. In the interviews, 21 participants expressed a preference for *IFil*'s filter names over those in *Pixlr* for a number of reasons. These were: 1) When users, especially non-experts, see a filter name in *IFil* , they can easily imagine how a picture will turn out after applying this filter. For instance, one user reported *"I really like the comprehensible names in IFil. As a beginner, I pretty much relied on the names when working with the application ... It becomes clear about what the filtered images are gonna to be with those names like 'Strawberry' makes pink ... Pixlr doesn't work that well"* (P13, male). In contrast, some participants reported that *"The names in Pixlr such as 'Ronny' and 'Sanna' don't mean anything to me"* (P1, female).

2) When users have a specific visual effect in mind, they can quickly translate it into a word which is often found in *IFil* prototype, for instance, a participant reported *"When I want to make it more red and yellow, I just go to the 'sunshine' or 'sunset' [filter]."* (P8, female).

3) The prototype semantic association makes a search path easier to recall, which is particularly useful when users want to go back to a filter previously visited, or compare two filters back and

forth. *"A couple of times I was inclined to grab a pencil and take notes, because there were so many filters I liked. But it turned out that I was able to remember how to find each of them as the names are really descriptive. Like, 'Carnival' was nice...but 'Twilight' looked good as well. In IFil I could effortlessly switch back between the two for comparison"* (P9, female) and *"I always forget the filters I used. But with IFil, I can easily remember the names to locate the one I want"* (P3, Female).

4) The diverse conceptual dimensions captured by *IFil* categories create a sense of serendipity during users' exploration of the filter space. *"When I chose 'Joy', it actually made the picture happier with a brighter and warmer tone. Then I tried 'Autumn', and the picture suddenly had a seasonal flavor ... It feels like I can add a layer ... any layer of context to my picture. It is very exciting"* (P16, Male). Overall, some participants indicated that speaking the same language achieved a state of mutual understanding between them and *IFil*. It was like having a conversation with the app. As reported by several participants: *"I can make perfect sense of what the names try to say"* (P4, Female), and *"I feel IFil can understand and reflect my intention better than Pixlr"* (P18, Female).

Filter Categorization. Compared to *Pixlr*, the majority of participants reported that the filters in *IFil* are better organized in term of intuitiveness and interpretability. Participants felt that *IFil* captures the *"relationships in the real-world"* (P15, male). As P14 said, *"I can easily guess what would be under each main category, like 'Happy' and 'Sad' belong to 'EMOTION' ... whereas the grouping of filters in Pixlr doesn't make sense to me at all."* Some participants felt that the hierarchical structure of *IFil* is at an appropriate level of complexity, making the size of the filter pool more manageable compared to that of *Pixlr*. *"Though there are plenty of filters in IFil, I don't feel them overwhelming because they are well organized"* (P6, female). In particular, such a structure seems to help participants to quickly narrow their choices down to a few conceptual dimensions that it is easier to relate to the image content. For example, P7 reported that he *"just focused on the categories of 'NATURE' and 'TIME'"* when looking for a filter to be applied to a scenery picture.

5.2.3 Video analysis. We analyzed the video recordings to determine whether there were consistent patterns of user behaviors. Different behaviors across the two systems were observed. In the baseline system, most participants (17 out of 21) chose the desired filters by repetitively trying out the whole filter space. More diverse searching behaviors appeared when participants interacted with *IFil* . Some (*e.g.,* P1, P3, and P14) switched frequently between categories and only experimented with a few color filters under each category; others (*e.g.,* P8 and P15) preferred to explore locally within only a few categories, where they experimented with several filters. Both strategies might be used by other users to narrow down the filter selection.

When using *Pixlr* , some users struggled to locate a filter they had previously tried. They had to go back to previously visited filters and sometimes mistakenly clicked on filters that were located close to the target (P3, P9, P15, P16). With *IFil* , there were fewer such errors. For example, P9 and P18 made fewer clicks before reaching the target with *IFil* (P9: Mean = 12.3; P18: Mean = 17.5) than with *Pixlr* (P9: Mean = 21.7; P18: Mean = 23.4). We observed that these participants both like to switch between two filter effects many times for comparison. It took them more time and a greater number of clicks to locate the filters they wished to compare when using *Pixlr* . One possible explanation is that the intuitive names given by *IFil* help users recall the filter effects and increase filters' memorability, thus providing navigational support.

5.3 Discussion

Compared with *Pixlr*, *IFil* has filter names and categories that appear to be more effective in supporting users to intuitively explore color filters. These results suggest a potential benefit of semantic features on filter findability. We discuss strategies that may be used by users during both tasks. These are our own interpretations of the data extracted from questionnaires, interviews,

and observations. One limitation of our approach is that we cannot prove that these strategies reflect the actual mental process of users. The comparison between *IFil* and *Pixlr* has a number of confounding factors, as the user interface (UI) design and the built-in filters between these two systems are slightly different. In the next study, we thus narrowed down the scope of our investigation to investigate how semantics in filter mapping, obtained through social curation, can affect filter findability and user experience. For this, we used only one UI and one filter space, in which we apply two different approaches to filter naming and categorization.

6 STUDY III: ONLINE USER STUDY

In this study, we evaluated the proposed semantic features of color filters applied to their names and categories and, specifically, how this can affect filter findability. We designed a comparative study using two versions of a filter application, one with and one without the assistance of filter semantics. We recruited a large number of MTurk crowd workers to perform two tasks, filter browsing and filter searching, using the two versions of the prototype.
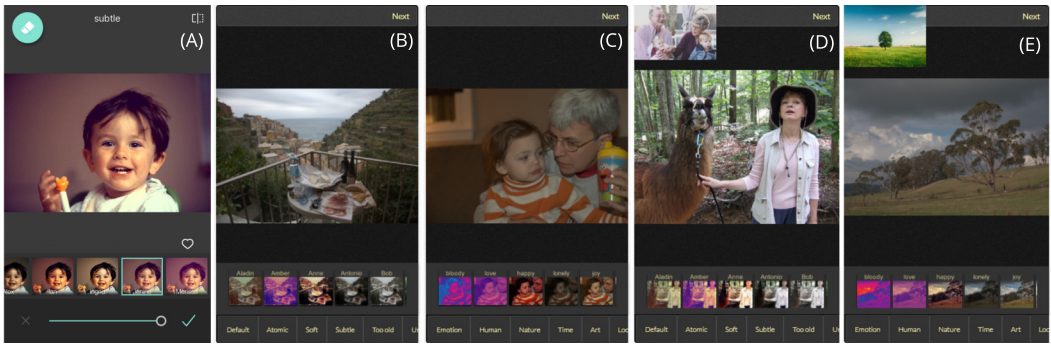

Fig. 10. Interfaces of the five different configurations designed for Study III. (A) is the interface of *Pixlr*. (B) to (E) show the interfaces of our prototype systems: (B) and (D) have the same filter configuration as *Pixlr* used in the browsing task and searching task respectively; (C) and (E) have the same filters but configured by the semantic names and categories in the browsing and searching tasks respectively.

6.1 Method

6.1.1 Prototype Design. We designed two prototypes in which the user interface is similar to that of conventional photo filtering applications. We developed a web-based baseline prototype equipped with *Pixlr*'s filters and configured those filters by the original names and categories proposed by the *Autodesk* system. The other prototype is exactly the same system in terms of UI and color filter space, but in this case, the filters are named and categorized by our framework. Examples of the processed names and categories are shown in Fig. 1. We call this the semantics prototype. Since we were unable to access the filter implementation of *Pixlr*, we pre-filtered all the images used in the experiment by all the built-in filters so that the prototypes have all the filter results. As mentioned in Study I (shown in Table 1), *Pixlr* has 112 filters with designer-defined names (*e.g.,* "Fred", "Carl", etc.), which are organized into eight categories (*e.g.,* "DEFAULT", "SUBTLE", etc.). We borrowed those filters and structures directly from *Pixlr* and removed the non-color effect filters. These excluded filters can change other features of the image, *e.g.,* its layout, which are out of the scope for the current study. There are a total of 13 filters under the "CREATIVE" category in *Pixlr*. We ultimately used 99 filters with their original names in the baseline prototype and with semantically mapped names in the semantics prototype.

6.1.2   Procedure. We conducted a $2 \times 2 \times 2$ split-plot design study to determine the effects of the within-subjects factor *Semantics* (two levels: *baseline* and *semantics* generated by social curation); the between-subjects factor *Task* (two levels: *browsing* and *searching* task); and the between-subjects factor user *Expertise* (two levels: *casual* users and *expert* users), on user experience and performance during filter exploration. Thus, in our study, participants were either required to perform a browsing task or a searching task. We took this approach to avoid fatigue and learning effects, which may occur when using a repeated measures approach. We recruited 110 MTurk crowd workers, and each of them performed one of the tasks using both of the prototypes. They had to state their level of expertise with color filters applications before the task by selecting one of the two options, *i.e.,* casual or expert user. During the experiment, we collected *Number of Clicks* for each task, and used this as a proxy for findability, with more clicks indicating lower findability. After finishing the task with one prototype, participants were asked to fill in a 5-point Likert scale questionnaire (where "1" is strongly disagree and "5" is strongly agree) for *Understandability* (*i.e.,* the filter names are understandable), *Relevance* (*i.e.,* the filter names are relevant to the filter effects), the *Usefulness of Filter Names* and *Categories* (*i.e.,* the filter names/categories help me to find the desired filter) and overall *Findability* (*i.e.,* The system helps me to find the desired filter).

6.2   Results

6.2.1   Interaction Analysis. The descriptive statistics for the number of clicks for the system with semantic filters and the baseline system are reported in Table 3. It can be noted that the *Number of Clicks* is overall lower for the *semantic* version of the interface. This appears to differ between tasks as, for *browsing*, there are on average 8.35 ($SD = 8.79$) clicks for the *semantic* version and 17.31 ($SD = 20.38$) for the *baseline* version of the interface. Whereas, for *searching*, there are 14.88 ($SD = 18.08$) for *semantic* and 15.330 ($SD = 22.70$) for *baseline* respectively. The dependent variable *Proportion of Clicks* represents the number of clicks in the filter area, as a proportion of the total number of clicks on the screen; thus this is a number between 0 and 1, and indicates whether participants clicked more on the filter area or on the category area. A number close to one indicates the former behavior, a number close to zero indicates the latter behavior. The descriptive statistics for this variable are also reported in Table 3. It is worth noting here that the proportion of clicks for the filter area was overall higher than 60% regardless of the condition, arriving at almost 80% for the semantic version of the system. This indicates that the participants clicked more on the filter area overall, compared with the category area.

Conditions		Clicks		Prop_Clicks	
		Mean	SD	Mean	SD
Browse	Semantics	8.35	8.79	0.79	0.28
	Baseline	17.31	20.38	0.62	0.27
Search	Semantics	14.88	18.08	0.78	0.33
	Baseline	15.33	22.7	0.61	0.3

Table 3. Descriptive statistics for the variables *Clicks* and *Proportion of Clicks*, across types of tasks and semantics level.

In this analysis, we excluded all those participants who had fewer than three clicks per task from the dataset, as these participants did not experiment enough with the tool to produce results that are useful in our interaction analysis. Thus, from an initial pool of 110 participants we ended up with 62, including 20 expert users. We conducted a $2 \times 2 \times 2$ factorial split-plot ANOVA for the dependent variable *Number of Clicks*. We tested the main effects of the three factors and their

interactions. The main effects of expertise and Task were not significant. There was a significant effect for Semantics, $F(1, 58) = 4.32, p = 0.04, \eta^2 = 0.168$. It shows that users click significantly less on the Semantics version than on the Baseline system. No significant interaction effect was found between the factors.

We also conducted a second $2 \times 2 \times 2$ factorial split-plot ANOVA for the dependent variable *Proportion of Clicks* (for filter names). We tested the main effects of the three factors and their interactions. The main effects of *Expertise*, *Task* and *Semantic* variables are not significant. There is only a close-to-significant interaction effect of Semantics with expertise, $F(1, 58) = 3.53, p = 0.06, \eta^2 = 0.111$. This effect is shown in Fig. 11. It can be noted that expert users (*mean = 0.73, SD = 0.27*) tend to click more on the filter area of the prototype (compared with the category area), than casual users (*mean = 0.65, SD = 0.26*) only in the semantic prototype. In the baseline prototype, in which the filters are not semantically mapped, an opposite effect is found (for *expert mean = 0.55, SD = 0.24*, for *casual mean = 0.67, SD = 0.25*).
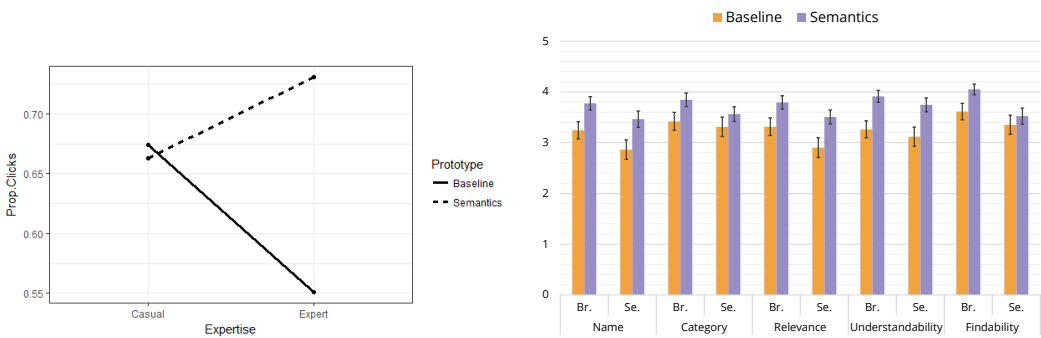

Fig. 11. Interaction effect between the factor Expertise and Semantics, for the variable *Proportion of Clicks*.

Fig. 12. Bar chart for the five questions in Study III. The y axes represent the score obtained for each question. The bar height represents the mean value across the type of task and level of semantics. The error bar represents the standard error.

6.2.2 System Evaluation. Fig. 12 shows the average values for the five questions in the after-task questionnaire, as a function of factor semantics and task. Overall, the semantic interface elicited higher scores for all the questions across tasks. The browsing task was associated with slightly higher scores regardless of the other variables. For instance, for the question measuring filter findability, for the browsing task, participants in the *Semantics* condition gave an average score of 4.05 compared with a score of 3.62 in the baseline condition. This indicates that findability was higher for the semantic version of the prototype.

To analyze participants' responses to the five different questions, we conducted 5 (Bonferroni adjusted) Wilcoxon tests for the browsing task and for the searching task, for both *expert* and *casual*. For the *Findability* question in the browsing task, there is a significant difference for the variable *Semantic*, $Z = 2.25, p = 0.024, r = 0.50$ (medium effect size). When splitting the data into *expert* and *casual*, there is a significant difference only for *casual*, $Z = 2.55, p = 0.021, r = 0.57$, indicating that *expert* users did not perceive higher filter findability for the semantic system compared to the baseline. A similar effect is found for the other 4 questions (*Usability of Filter Category*, *Usability of Filter Name*, *Relevance* and *Understandability*), in which only for casual users was a significant difference between the system mapped using the semantic and the baseline system observed. For

the search task, we found a similar effect for the variables *Relevance*, $Z = 2.432, p = 0.030, r = 0.504$, and *Understandability*, $Z = 2.287, p = 0.044, r = 0.51$, in which only casual users seems to perceive the difference brought about by the use of semantics mapping. In this task, there was no significant difference for the measure of *Findability*, *Usability of Filter Category*, and only a marginal difference for the question Filter, *i.e.,* only when the data was not split into *casual* and *expert*, $Z = 2.663, p = 0.008, r = 0.596$. To sum up, the factor *Semantic* seems to affect how casual users (only) perceive findability and other usability measures, and this effect is greater for the browsing task. In general, casual users rated the system with semantically mapped filters higher in almost all the five measures of experience. An interpretation of these findings is provided in the discussion. The full descriptive statistics are reported in the supplementary material.

6.3 Discussion

The results show that users find the desired filter with fewer clicks when using the semantics prototype compared to the baseline. This suggests that the derived semantics help users to locate the desired filter. This is in line with past research on color semantics, for instance, showing that chart reading speed is linked to semantically-resonant colors in visualization [32]. As shown in Table 3 this effect was stronger for the browsing task; this is potentially due to the different user behaviors in the two task settings. In the searching task, users are provided with a target image and they might simply look at the preview for a filter with a visual effect that closely matches the target. This is consistent with the findings from Study II in which some users mention they mostly rely on visual similarity between the target and the preview. However, in the browsing task, they might seek cues from filter names as they have no specific target in mind. In that case, they could feel confused when the filter name is inconsistent with the filter effect. This is further supported by the system evaluation, shown in Fig. 12 in which scores are higher for *Usability of Filter Name*, *Usability of Filter Category* and *Findability* in the browsing task than the searching task. When considering different levels of user expertise with color filter applications, we did not find a significant difference in the number of clicks between expert and casual users. Nevertheless, we found a relationship between the level of expertise users have and the semantic representation of color filters; specifically, the focus on the filter areas tends to increase for experts (higher click proportion) when the system includes semantic mappings. This might be explained by the fact that, for users who are very familiar with color filter applications, semantics may reduce the global search (search on categories) as it allows experts to find the desired category quickly so that they can experiment with the color filters within a specific category (local search). In the system evaluation, casual users found it easier to locate filters in the semantics version, compared to the baseline version of the prototype. Such an effect was not found in experts. It is possible that experts are already familiar with those technical names (*e.g.,* "Nightvision" and "Unicolor") used in commercial systems (*e.g., Pixlr*); this familiarity with technical terms may mean experts perceive the semantic prototype less differently to the baseline in terms of its findability and the other measures of usability. Nevertheless, experts only show a difference in the proportion of clicks when the semantics are manipulated. This indicates that people's perception (in this case, regarding the system characteristics) is not necessarily related to their actual behavior (in this case, related to how they perform in a task).

7 GENERAL DISCUSSION

7.1 Lessons Learned from Repurposing Social Curation Data

This paper explores a method which repurposes social curation data to improve the findability of photo filters. We show that a large set of accessible crowd data can serve as the basis for developing

computational models for data-driven design. However, such data repurposing is not a trivial process. It entails several significant technical challenges:

Data transferability. Repurposing social curation data requires an in-depth study of the underlying transferability of cross-domain knowledge. That is, the chosen source data should bring relevant value to the target domain. However, the criteria for assessing relevance is unclear and highly context-specific. Thus, it is hard to determine which data should be collected and analyzed [9]. To address this issue, one needs to look at the characteristics of the domain data in detail. In this work, we aimed to transfer the semantics of color themes curated by online communities on *COLOURlovers* to digital filters. This is because the user-curated names convey the perception, meanings and emotional responses of color themes [43], and color themes can naturally express filter effects [54]. These factors make color theme data highly relevant for modeling filter semantics. There are alternative online databases that can be used as a source of data, such as user-generated meta-data (*e.g.,* online images tags). However, these suffer from the content conflation problem [34]. That is, an image tag might describe the image content rather than the color distribution, meaning this data is less reliable for our task.

Data consistency and quality. Repurposing the data created under different contexts requires dealing with data inconsistency and bridging the gaps across domains [12]. More specifically, the source data might come in different formats, context, and levels of details. As such, it is impractical to directly deploy the information from one domain to another. In this work, in order to repurpose the semantics of color themes collected from *COLOURlovers* for digital filters, we managed to build a connection by matching color themes to color filters using colors as a mediator. It is also important to handle the uncontrolled quality of social curation data from open online platforms. Users can name a color theme using multilingual words, by using special characters, or by inventing terms which are meaningful for the creator only. Due to these shortcomings, cleaning the data requires significant effort.

The balance between crowd consensus and knowledge diversity. Curated data contains diverse opinions given by the content creators who often have different preferences and cultural backgrounds. In many situations, it is difficult to attain one consensus across the board. To be more specific, while we seek to model color theme semantics by mining the majority's patterns from curation data, the valid responses from a minority group would be overlooked. Therefore, the balance between mining the consensus from the majority and ensuring a certain degree of knowledge diversity should be considered. In this work, we aggregated user ratings to highlight the valuable items (*i.e.,* highly rated color themes) instead of only considering the most frequent patterns appearing in the data. Future research could extend this work by exploring user profiles, *e.g.,* by giving higher weights to the responses from those users who always contribute to the production of high-quality data.

7.2 Design Implications and Future Work

The results of our work show the potential value of social curation in computational design by repurposing data generated from online communities. It opens a discussion on how we could respond to the opportunities and challenges of the future paradigm of design, reshaped by the increasing availability of online data. However, our study has some limitations. First, we compared *IFil* with only one of the most popular applications, *i.e., Pixlr*. Other commercial applications categorize and name color filers using different approaches. Second, our study only focuses on color filters without considering those filters that can change other photo properties, *e.g.,* image texture, and layout. Although these choices are necessary for reducing the number of confounders in our evaluation, they may limit the generalisability of our results. Third, we model color theme semantics to reflect the user consensus of the general public. However, we do not incorporate

domain knowledge about color theory into the implementation of our current model. For example, it has been shown that color-dominance and color-saliency are important for perceiving the color distribution of an image [25, 33]. In future work, we could assign larger weights to dominant and salient colors when modeling color theme semantics. Fourth, the social curation data of color themes was not fully exploited. For instance, future work could utilize aesthetic attributes, tags, and comments in *COLOURlovers* to enhance the model. Finally, we computationally extracted names from curated color theme data for color filters but did not compare those names with the words users would use to describe filters. Such comparison can verify our generated names in terms of their similarity to user-provided groundtruths, and potentially give insights into refining model-generated filter names. This could be done by collecting names directly from crowd users in the future.

8 CONCLUSION

In this paper, we test the feasiblity of using a probabilistic model to construct color theme semantics from social curation data and use it to semantically index and organize filter effects. Three studies have been conducted to verify the framework, showing how the derived filter semantics can improve filter findability. We find that the filter names and categories generated by our model help users to locate the desired filter more efficiently when compared with the baseline. This effect is stronger for casual users than expert users. To our knowledge, this is the first study which fully evaluates the impact of color filter semantics on filter findability. Although the study focused on filter findability, it is worth noting that color theme semantics extracted from social curation data also carry the potential for application in a number of other areas, including but not limited to keyword-based photo retouching and color theme-based image retrieval.

REFERENCES

[1] 2015. The Easter Eggs Hidden In The Names Of Instagram Filters. http://www.refinery29.com/2015/12/99486/instagram-filters-names. (December 2015).

[2] 2017. Adobe Kuler: Color wheel | Color schemes. https://color.adobe.com/. (September 2017).

[3] 2017. Colourlovers. http://www.colourlovers.com/. (September 2017).

[4] 2018. Google News. https://news.google.com/. (July 2018).

[5] Saeideh Bakhshi and Eric Gilbert. 2015. Red, purple and pink: The colors of diffusion on Pinterest. *PloS one* 10, 2 (2015), e0117148.

[6] Saeideh Bakhshi, David Shamma, Lyndon Kennedy, and Eric Gilbert. 2015. Why We Filter Our Photos and How It Impacts Engagement. (2015). https://www.aaai.org/ocs/index.php/ICWSM/ICWSM15/paper/view/10573/10484

[7] Jörg Becker, Christian Brelage, Karsten Klose, and Michael Thygs. 2003. Conceptual Modeling of Semantic Navigation Structures: The MoSeNa-approach. In *Proceedings of the 5th ACM International Workshop on Web Information and Data Management (WIDM '03)*. ACM, New York, NY, USA, 118–125. https://doi.org/10.1145/956699.956726

[8] Robert Benavente, Francesc Tous, Ramon Baldrich, and Maria Vanrell. 2002. Statistical Modelling of a Colour Naming Space. In *European Conference on Colour in Graphics, Imaging and Vision (CGIV'02)*. Poitiers (France), 406–411. http://www.cat.uab.cat/Public/Publications/2002/BTB2002

[9] Alessandro Bertoni. 2018. Role and Challenges of Data-Driven Design in the Product Innovation Process. In *16th IFAC Symposium on Information Control Problems in Manufacturing-INCOM 18*.

[10] V. Bychkovsky, S. Paris, E. Chan, and F. Durand. 2011. Learning Photographic Global Tonal Adjustment with a Database of Input/Output Image Pairs. In *Proceedings of the 2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR '11)*. IEEE Computer Society, Washington, DC, USA, 97–104. https://doi.org/10.1109/CVPR.2011.5995413

[11] Elena Caoduro. 2014. Photo Filter Apps: Understanding Analogue Nostalgia in the New Media Ecology. *Networking Knowledge: Journal of the MeCCSA Postgraduate Network* 7, 2 (2014). http://ojs.meccsa.org.uk/index.php/netknow/article/view/338

[12] José M Cavanillas, Edward Curry, and Wolfgang Wahlster. 2016. *New horizons for a data-driven economy: a roadmap for usage and exploitation of big data in Europe.* Springer.

[13] Jason Chuang, Maureen Stone, and Pat Hanrahan. 2008. A probabilistic model of the categorical association between colors. In *Color and Imaging Conference*, Vol. 2008. Society for Imaging Science and Technology, 6–11.

[14] Gabriela Csurka, Sandra Skaff, Luca Marchesotti, and Craig Saunders. 2010. Learning moods and emotions from color combinations. In *Proceedings of the Seventh Indian Conference on Computer Vision, Graphics and Image Processing.* ACM, 298–305.

[15] Leah Findlater and Joanna McGrenere. 2007. Evaluating reduced-functionality interfaces according to feature findability and awareness. In *IFIP Conference on Human-Computer Interaction.* Springer, 592–605.

[16] L. A. Gatys, A. S. Ecker, and M. Bethge. 2016. Image Style Transfer Using Convolutional Neural Networks. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR).* 2414–2423. https://doi.org/10.1109/CVPR.2016.265

[17] Xue Geng, Hanwang Zhang, Zheng Song, Yang Yang, Huanbo Luan, and Tat-Seng Chua. 2014. One of a kind: User profiling by social curation. In *Proceedings of the 22nd ACM international conference on Multimedia.* ACM, 567–576.

[18] Paul Heckbert. 1982. Color Image Quantization for Frame Buffer Display. *SIGGRAPH Comput. Graph.* 16, 3 (July 1982), 297–307. https://doi.org/10.1145/965145.801294

[19] Jeffrey Heer and Maureen Stone. 2012. Color Naming Models for Color Selection, Image Editing and Palette Design. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '12).* ACM, New York, NY, USA, 1007–1016. https://doi.org/10.1145/2207676.2208547

[20] Loïc P Heurley, Thibaut Brouillet, Gabrielle Chesnoy, and Denis Brouillet. 2013. Color perception involves color representations firstly at a semantic level and then at a lexical level. *Cognitive processing* 14, 1 (2013), 19–29.

[21] Xiaodi Hou and Liqing Zhang. 2007. Color Conceptualization. In *Proceedings of the 15th ACM International Conference on Multimedia (MM '07).* ACM, New York, NY, USA, 265–268. https://doi.org/10.1145/1291233.1291288

[22] Bernd Huber, Daniel McDuff, Chris Brockett, Michel Galley, and Bill Dolan. 2018. Emotional Dialogue Generation Using Image-Grounded Language Models. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18).* ACM, New York, NY, USA, Article 277, 12 pages. https://doi.org/10.1145/3173574.3173851

[23] Katsuhiko Ishiguro, Akisato Kimura, and Koh Takeuchi. 2012. Towards automatic image understanding and mining via social curation. In *Data Mining (ICDM), 2012 IEEE 12th International Conference on.* IEEE, 906–911.

[24] 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. https://doi.org/10.1145/3009924

[25] EunJin Kim and Hyeon-Jeong Suk. 2017. Thoughts and Tools for Crafting Colors: Implications from Designers' Behavior. In *Proceedings of the 2017 Conference on Designing Interactive Systems (DIS '17).* ACM, New York, NY, USA, 321–331. https://doi.org/10.1145/3064663.3064769

[26] Akisato Kimura, Katsuhiko Ishiguro, Makoto Yamada, Alejandro Marcos Alvarez, Kaori Kataoka, and Kazuhiko Murasaki. 2013. Image context discovery from socially curated contents. In *Proceedings of the 21st ACM international conference on Multimedia.* ACM, 565–568.

[27] Shigenobu Kobayashi. 1981. The aim and method of the color image scale. *Color Research & Application* 6, 2 (1981), 93–107. https://doi.org/10.1002/col.5080060210

[28] Lauren I Labrecque, Vanessa M Patrick, and George R Milne. 2013. The marketersâĂŹ prismatic palette: A review of color research and future directions. *Psychology & Marketing* 30, 2 (2013), 187–202.

[29] Pierre-Yves Laffont, Zhile Ren, Xiaofeng Tao, Chao Qian, and James Hays. 2014. Transient Attributes for High-level Understanding and Editing of Outdoor Scenes. *ACM Trans. Graph.* 33, 4, Article 149 (July 2014), 11 pages. https://doi.org/10.1145/2601097.2601101

[30] Hyojin Lee, Xiaoyan Deng, H Rao Unnava, and Kentaro Fujita. 2014. Monochrome Forests and Colorful Trees: The Effect of Black-and-White versus Color Imagery on Construal Level. 41, 4 (2014), 1015–1032. http://www.jstor.org/stable/10.1086/678392

[31] J. Y. Lee, K. Sunkavalli, Z. Lin, X. Shen, and I. S. Kweon. 2016. Automatic Content-Aware Color and Tone Stylization. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR).* 2470–2478. https://doi.org/10.1109/CVPR.2016.271

[32] Sharon Lin, Julie Fortuna, Chinmay Kulkarni, Maureen Stone, and Jeffrey Heer. 2013. Selecting Semantically-Resonant Colors for Data Visualization. In *Computer Graphics Forum*, Vol. 32. Wiley Online Library, 401–410.

[33] Sharon Lin and Pat Hanrahan. 2013. Modeling How People Extract Color Themes from Images. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '13).* ACM, New York, NY, USA, 3101–3110. https://doi.org/10.1145/2470654.2466424

[34] A. Lindner and S. SÃijsstrunk. 2015. Semantic-Improved Color Imaging Applications: It Is All About Context. *IEEE Transactions on Multimedia* 17, 5 (May 2015), 700–710. https://doi.org/10.1109/TMM.2015.2410175

[35] Yiming Liu, Michael Cohen, Matt Uyttendaele, and Szymon Rusinkiewicz. 2014. AutoStyle: Automatic Style Transfer from Image Collections to Users' Images. In *Proceedings of the 25th Eurographics Symposium on Rendering (EGSR '14)*. Eurographics Association, Aire-la-Ville, Switzerland, Switzerland, 21–31. https://doi.org/10.1111/cgf.12409

[36] C. M. MacLeod. 1991. Half a century of research on the Stroop effect: an integrative review. *Psychological Bulletin* 109, 2 (March 1991), 163–203.

[37] Winter Mason and Siddharth Suri. 2012. Conducting behavioral research on AmazonâĂŹs Mechanical Turk. *Behavior research methods* 44, 1 (2012), 1–23.

[38] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Efficient Estimation of Word Representations in Vector Space. *CoRR* abs/1301.3781 (2013). http://arxiv.org/abs/1301.3781

[39] George A. Miller. 1995. WordNet: A Lexical Database for English. *Commun. ACM* 38, 11 (Nov. 1995), 39–41. https://doi.org/10.1145/219717.219748

[40] Nathan Moroney. 2003. Unconstrained web-based color naming experiment. In *Color imaging VIII: Processing, hardcopy, and applications*, Vol. 5008. International Society for Optics and Photonics, 36–47.

[41] Naila Murray, Sandra Skaff, Luca Marchesotti, and Florent Perronnin. 2012. Toward automatic and flexible concept transfer. *Computers & Graphics* 36, 6 (2012), 622–634.

[42] Dimitris Mylonas, Lindsay MacDonald, and Sophie Wuerger. 2010. Towards an online color naming model. In *Color and Imaging Conference*, Vol. 2010. Society for Imaging Science and Technology, 140–144.

[43] Peter O'Donovan, Aseem Agarwala, and Aaron Hertzmann. 2011. Color Compatibility from Large Datasets. *ACM Trans. Graph.* 30, 4, Article 63 (July 2011), 12 pages. https://doi.org/10.1145/2010324.1964958

[44] Peter O'Donovan, Jānis Lībeks, Aseem Agarwala, and Aaron Hertzmann. 2014. Exploratory Font Selection Using Crowdsourced Attributes. *ACM Trans. Graph.* 33, 4, Article 92 (July 2014), 9 pages. https://doi.org/10.1145/2601097.2601110

[45] Carlos Padoa, Daniel Schneider, Jano Moreira De Souza, and Sergio Palma J Medeiros. 2015. Investigating social curation websites: A crowd computing perspective. In *Computer Supported Cooperative Work in Design (CSCWD), 2015 IEEE 19th International Conference on*. IEEE, 253–258.

[46] Stephen E Palmer and Karen B Schloss. 2010. An ecological valence theory of human color preference. *Proceedings of the National Academy of Sciences* 107, 19 (2010), 8877–8882.

[47] Zoran Pečenović, Minh N. Do, Martin Vetterli, and Pearl Pu. 2000. *Integrated Browsing and Searching of Large Image Collections*. Springer Berlin Heidelberg, Berlin, Heidelberg, 279–289. https://doi.org/10.1007/3-540-40053-2_25

[48] Yilang Peng. 2017. Time Travel with One Click: Effects of Digital Filters on Perceptions of Photographs. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. ACM, New York, NY, USA, 6000–6011. https://doi.org/10.1145/3025453.3025810

[49] Judith Redi and Isabel Povoa. 2014. Crowdsourcing for Rating Image Aesthetic Appeal: Better a Paid or a Volunteer Crowd?. In *Proceedings of the 2014 International ACM Workshop on Crowdsourcing for Multimedia (CrowdMM '14)*. ACM, New York, NY, USA, 25–30. https://doi.org/10.1145/2660114.2660118

[50] Yossi Rubner, Carlo Tomasi, and Leonidas J Guibas. 2000. The earth mover's distance as a metric for image retrieval. *International journal of computer vision* 40, 2 (2000), 99–121.

[51] Vidya Setlur and Maureen C Stone. 2016. A linguistic approach to categorical color assignment for data visualization. *IEEE transactions on visualization and computer graphics* 22, 1 (2016), 698–707. https://doi.org/10.1109/TVCG.2015.2467471

[52] Martin Solli and Reiner Lenz. 2010. Color semantics for image indexing. In *Conference on Colour in Graphics, Imaging, and Vision*, Vol. 2010. Society for Imaging Science and Technology, 353–358.

[53] W. T. Sun, T. H. Chao, Y. H. Kuo, and W. H. Hsu. 2017. Photo Filter Recommendation by Category-Aware Aesthetic Learning. *IEEE Transactions on Multimedia* 19, 8 (Aug 2017), 1870–1880. https://doi.org/10.1109/TMM.2017.2688929

[54] Baoyuan Wang, Yizhou Yu, Tien-Tsin Wong, Chun Chen, and Ying-Qing Xu. 2010. Data-driven Image Color Theme Enhancement. *ACM Trans. Graph.* 29, 6, Article 146 (Dec. 2010), 10 pages. https://doi.org/10.1145/1882261.1866172

[55] Ka-Ping Yee, Kirsten Swearingen, Kevin Li, and Marti Hearst. 2003. Faceted Metadata for Image Search and Browsing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '03)*. ACM, New York, NY, USA, 401–408. https://doi.org/10.1145/642611.642681