Frontiers of Information Technology & Electronic Engineering
www.jzus.zju.edu.cn; engineering.cae.cn; www.springerlink.com
ISSN 2095-9184 (print); ISSN 2095-9230 (online)
E-mail: jzus@zju.edu.cn

Automatic synthesis of advertising images according to a specified style*

Wei-tao YOU†1, Hao JIANG†‡2, Zhi-yuan YANG†1, Chang-yuan YANG†3, Ling-yun SUN1

[1]*Key Laboratory of Design Intelligence and Digital Creativity of Zhejiang Province, Hangzhou 310027, China*

[2]*International Design Institute, Zhejiang University, Hangzhou 310058, China*

[3]*International User Experience Business Unit, Alibaba Group, Hangzhou 311121, China*

†E-mail: weitao_you@zju.edu.cn; jiang_hao@zju.edu.cn; youngs@zju.edu.cn; changyuan.yangcy@alibaba-inc.com

Received July 22, 2019; Revision accepted Dec. 8, 2019; Crosschecked June 19, 2020; Published online Aug. 5, 2020

Abstract: Images are widely used by companies to advertise their products and promote awareness of their brands. The automatic synthesis of advertising images is challenging because the advertising message must be clearly conveyed while complying with the style required for the product, brand, or target audience. In this study, we proposed a data-driven method to capture individual design attributes and the relationships between elements in advertising images with the aim of automatically synthesizing the input of elements into an advertising image according to a specified style. To achieve this multi-format advertisement design, we created a dataset containing 13 280 advertising images with rich annotations that encompassed the outlines and colors of the elements, in addition to the classes and goals of the advertisements. Using our probabilistic models, users guided the style of synthesized advertisements via additional constraints (e.g., context-based keywords). We applied our method to a variety of design tasks, and the results were evaluated in several perceptual studies, which showed that our method improved users' satisfaction by 7.1% compared to designs generated by nonprofessional students, and that more users preferred the coloring results of our designs to those generated by the color harmony model and Colormind.

Key words: Image dataset; Data-driven method; Automatic advertisement synthesis
https://doi.org/10.1631/FITEE.1900367 **CLC number:** TP391

1 Introduction

Images in advertising play an important role in helping companies spread awareness of their products or services. A considerable amount of time and money is spent on the creation of advertisements (ads) every year. Although generating a creative ad with visual rhetoric requires professional design skills, most of these ads are simply synthesized using a set of graphical and text elements that have relatively structured visual appearances and comply with some measurable rules in the effort to clearly convey information. To reduce the cost of labor for the creation of these ads, technologies that support automatic synthesis of ads have received considerable attention, including automatic assembly of graphical elements using esthetic principles (Yang et al., 2016) and simultaneously creating a series of banners for different display sizes (Zhang et al., 2017).

Advertising images are normally composed of design elements such as products, texts, and logos instead of pixels laid on a regular lattice. The

‡ Corresponding author

* Project supported by the National Science and Technology Innovation 2030 Major Project of the Ministry of Science and Technology of China (No. 2018AAA0100700), the National Natural Science Foundation of China (No. 61672451), the Provincial Key Research and Development Plan of Zhejiang Province, China (No. 2019C03137), the China Postdoctoral Science Foundation (No. 2018M630658), and the Alibaba-Zhejiang University Joint Institute of Frontier Technologies

ORCID: Wei-tao YOU, https://orcid.org/0000-0002-9625-5547; Hao JIANG, https://orcid.org/0000-0002-3530-5133

© Zhejiang University and Springer-Verlag GmbH Germany, part of Springer Nature 2020

quality of an ad depends on the presence of these elements, and their attributes to and relationships with the other elements. Designers often create a series of advertising images by adjusting the placement and color of the elements, which is called the multi-format design approach (Zhang et al., 2017), commonly used when designing ads for a wide variety of display sizes, promotional activities, or target audiences. To automate the process of designing multi-format ads, most researchers have used data-driven methods to learn the design attributes in advertising images. However, acquiring an adequate number of ads with element-level labels is challenging and time-consuming. Most previous studies have focused on a particular design problem related to ad synthesis, such as the generation of pleasing colors (Lin et al., 2013) or understanding the content of an ad (Hussain et al., 2017).

Our previous work created a dataset of over 13 000 product ads and built two probabilistic models to automatically recolor input elements for different coloring tasks (You et al., 2019). In this study, we focused on the automatic synthesis of advertising images. In addition to recoloring design elements in an image, we need to solve the layout problems like arranging input elements and texts. Thus, we proposed methods to represent the features of characters and encode various types of layouts for layout synthesis. To ensure the quality of the synthesized images, we tested the reliability of our dataset and extended the existing data annotation. We evaluated the capability of our method by implementing two baseline methods, automatically synthesizing the input of elements into an ad according to user preference and recoloring each element according to a user-specified style. Independent perception studies verified that the layout synthesized by our method was preferable to those designed by nonprofessionals. Users preferred the esthetic appearance of our model's color suggestions to the colors generated from other models. The main contributions of this work are as follows:

1. We develop and introduce probabilistic models that capture the stylistic design attributes of training images, which incorporate the features of the graphic element's role to predict the design performance in the target context.

2. We propose functions and algorithms that are flexible enough to represent the unique features of Chinese characters, and encode various types of advertising image layouts to efficiently produce a layout from the input elements and texts.

3. We demonstrate how our dataset can be applied in the automatic synthesis of advertising images according to user-preference constraints to produce high-quality layout and color designs with a specified style.

2 Related work

2.1 Graphic design dataset

Graphic design generally consists of symbols, images, and texts to form visual representations of ideas and messages. Because the visual perception of a design depends largely on the arrangement and color combination of these elements, considerable effort has been made to build a dataset that contains labels for the design attributes to make it easy to learn the visual appearance.

The placement of elements in a layout has been extensively studied for several related topics, such as photo collage (Geigel and Loui, 2003), magazine cover design (Jahanian et al., 2012), and furniture object arrangement (Yu et al., 2011). Training images are often segmented into different panels and structured with layout attributes (e.g., position and size). Antonacopoulos et al. (2009) introduced a dataset with 1240 document images for the evaluation of layout analysis methods, containing the metadata for multiple types of elements, such as title, author, image, and font. Cao et al. (2012) built a dataset that contained approximately 4000 scanned manga pages to generate a stylistic manga layout, where the panel vertices on each page were manually labeled.

Color is one of the important design attributes and it has a significant impact on graphic design. Lin et al. (2013) collected 100 colored patterns for each of 82 artists to validate their model for automatically coloring two-dimensional patterns. To study theories on color compatibility, O'Donovan et al. (2011) developed a color selection tool by scoring the quality of a five-color set called a color theme. To enable a tool to learn a quantitative model, a dataset that contained a collection of 10 743 color themes and their ratings was created using Amazon's Mechanical Turk (MTurk). A large number of color problems have

been studied using this color compatibility model, including color theme extraction (Lin and Hanrahan, 2013) and web page coloring (Gu and Lou, 2016).

Such structured data have been effective in solving design problems aimed at a specific type of graphic design, such as a document image or a web page. However, the design of advertising images is challenging because it must clearly convey information while satisfying esthetic goals. The style of an ad must comply with its inherent content, product brand, and target audience. Although previous researchers have built several datasets that contain advertising images, the structured labels were often on a small scale (Zhang et al., 2017) and there was no target for automatic image synthesis (Yang et al., 2016).

2.2 Automated graphic design

Traditional methods for automated graphic design are generally driven by design rules or structured data. When creating a graphic design, designers often comply with design rules, such as the esthetic principles of a layout (Tuch et al., 2010) and harmonious color models (Tokumaru et al., 2002). A rule-based method uses these design rules to assist in the particular design tasks that vary from layout synthesis (O'Donovan et al., 2014) to color generation (Yang et al., 2016) and image thumbnailing (Choi and Kim, 2016). O'Donovan et al. (2014) proposed an energy function by assembling various heuristic visual cues and design principles to optimize single-page layouts, which was extended to an interactive tool (O'Donovan et al., 2015). Yang et al. (2016) presented a system to generate visual-textual presentation layouts, in which colors were determined automatically with the aid of a color harmony model and a color tone model, whereas topic colors were defined by the designers.

Instead of design rules, data-driven methods obtain particular design attributes from training images. Qiang et al. (2019) proposed a Bayesian network to characterize the relationships among design elements in a layout, which were learned from a small number of paper-poster examples. Charpiat et al. (2008) estimated multimodal distributions of local texture features for gray-scale image colorization, which was further developed and applied to automated pattern coloring (Lin et al., 2013) and automated web page coloring (Gu and Lou, 2016). Besides, Liu et al. (2019) introduced an intelligent banner release tool, Luban, which could automatically synthesize banners with different commodities. Recently, generative adversarial networks (GANs) have been successful for image synthesis such as with image-to-image translation (Zhu et al., 2017), text-to-image synthesis (Xu et al., 2018), and high resolution image generation (Wang et al., 2018). Tang et al. (2019) summarized the current emerging frameworks of design intelligence, including the methods for content generation using GANs. Although these GANs have demonstrated great power in generating realistic and natural-looking images, there are few models for graphic design because of the difficulty in finding data representations that are suitable for learning.

To summarize, most rule-based methods are generally context-free, which means that they struggle to predict the design performance in a real context. However, data-driven approaches are considered to be more practical because they depend on real images created by designers. Existing generative networks focus on the generation of pixel-based images, and no previous data-driven models could learn to create multi-format ads from large datasets.

3 Advertising images dataset

In this section, we introduce the methods used to construct a dataset of advertising images, describe strategies used to ensure consistent annotation, and demonstrate how to extend the dataset by extracting a panel layout and principal colors. Fig. 1 shows the workflow for constructing our dataset, with an example of the results at the bottom of each step.

3.1 Image acquisition and ad selection

Collecting a large number of advertising images with structured data for the elements is challenging because few source files (e.g., in PSD or SVG format) are shared on websites. To obtain sufficient design attributes for the elements, we first collect pixel-based images of ads and then invite annotators to structure each ad by labeling the elements in the images.

We search for ads from Huaban (https://huaban.com/), which is an online community that focuses on creating and sharing images. These ads cover diverse product types such as

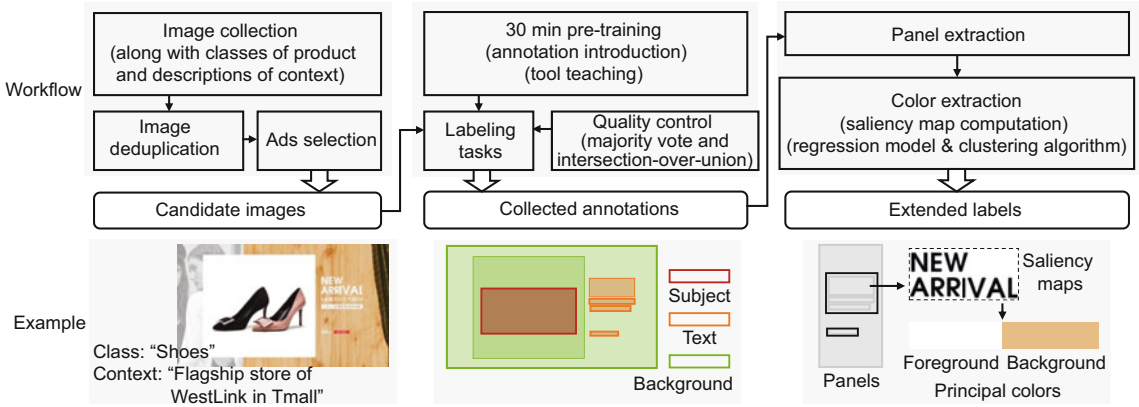

Fig. 1 Flow diagram for the construction of our dataset

clothing, electronics, food, and luggage. All of these images also include descriptions that contain keywords related to their usage context (e.g., "double 11 is coming," "the female apparel in 2017"). After removing all images smaller than 256 × 256 pixels, we obtain an initial pool of 20 302 noisy ads.

To remove duplicate images from this noisy set, we compute a fingerprint for each image using a perceptual Hashing algorithm (Buldas et al., 2013) and mark all pairs of images whose similarities are greater than the given threshold as the duplicates. Following the removal of duplicates, we show the annotators a large number of examples that we consider to be ads and those that we do not. The annotators are invited to filter images by answering the question—Could this image appear as an ad on an online shopping website? Finally, we collect the images that at least three-quarters of the annotators have labeled as an ad; thereby, we obtain 13 280 ads and 7022 non-ads.

3.2 Annotation collection

With the aid of our online annotation tool developed from LabelMe (Russell et al., 2008), each element is labeled by a bounding box. Also, we invite annotators to add a tag that describes its role. These tags are categorized into three groups, "Subject" (e.g., product), "Text" (e.g., description and logo), and "Background." We encourage the annotators to consider the adjacent elements when drawing a bounding box (Fig. 2). For example, multiple objects are allowed to appear in a single bounding box when they cluster together. A bounding box with the description tag could contain several lines of texts if they are adjacent and have similar design attributes (e.g., font, size, and color). Additionally, we ask the annotators to label the number of text lines for elements in the "Text" group.

To develop a consistent understanding of an advertising image, we require all annotators participate in a 30-min pre-training. This session allows them to study what an element is and how to effectively determine its role. To identify high-quality annotators, we assign a final labeling task that requires participants list the design elements from several advertising images. Approximately two out of three participants pass the labeling task and are invited to be the annotators.

During each of the labeling tasks, we ask one annotator to label 300 candidate images individually. While the annotator is completing the task, we randomly select 50 annotations from the results and ask three annotators to assess them. The primary considerations are whether important elements are omitted and whether the tag matches the labeled element well. We discard annotations of insufficient quality and add the corresponding ads back to the pool of candidate images.

We submit all collected annotations to our annotation tool and invite the annotators to check

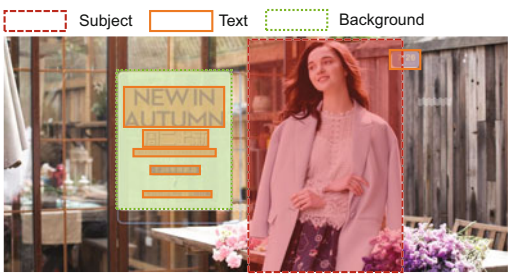

Fig. 2 Manually listed outlines of the design elements

the accuracy of the labeled bounding boxes. The annotators could modify suspicious bounding boxes that mistake the element or are not exactly aligned around the element. We select the relabeled images whose intersection-over-union (IoU) values are smaller than the pre-determined threshold. Then we invite the annotators to vote on the newly collected labels. The labels that obtain the majority of votes are finally accepted.

3.3 Dataset extension

For the collected annotations, we extend the dataset by extracting the panel layout and principal colors. Advertising images in our dataset contain a large number of text elements, which make it difficult to compute the relationships among the elements. Most previous approaches have learned to generate a panel-based layout for arranging the text and graphical elements (Cao et al., 2012; Qiang et al., 2019). To facilitate the learning of layout attributes, we merge the text elements into panels according to their adjacent distance. The pseudocode for computing the panel layout is given in Algorithm 1.

To obtain the principal color as a color attribute, we first use labeled bounding boxes and tags to refine the boundaries. For the elements in the "Subject" group, we use fully convolutional instance-aware semantic (FCIS) (Li Y et al., 2017) to recognize the instances. For instances that cannot be recognized, we invite the annotators to label their masks manually. For the elements in the "Text" group, considering that designers prefer to use a clean color as background to clearly convey message, we detect the saliency of texts and images by the Markov chain (Jiang et al., 2013), which performs substantially well when salient objects exist against a relatively clean background (Borji et al., 2015).

For the masks of products, we use the regression model in Lin and Hanrahan (2013) to obtain the first visually sensitive color. Using k-means clustering, we use the color of pixels in the most significant and non-significant regions of the saliency map to represent the colors of text and the corresponding background, respectively.

Algorithm 1 Panel layout computation

Input: set of labeled text elements T_n.
Output: set of computed panels P_m.

```
θ ← pre-determined threshold of the distance;
Initialize P_m;
Add first text element t_1 to P_m as p_1;
for each t_i (2 ≤ i ≤ n) in T_n do
    Compute distance d between t_i and each panel p in P_m;
    if d < θ then
        Add t_i to the corresponding p;
        Update [x, y, w, h] of p;
        Merge the overlapping panels in P_m;
    else
        Add t_i to P_m as a new panel;
    end if
end for
return P_m.
```

3.4 Dataset analysis

We collect annotations of 97 667 elements from 13 280 advertising images (Table 1) with their computed panel layouts and principal colors. To analyze the reliability of these annotations, we compute the precision and recall of the labels in the sampled images by comparing them with the ground truth obtained from the assessments by experts (co-workers on this project). These images are randomly selected from the entire dataset. As the role of an element is often ambiguous, even the experts tend to disagree on the element labeling results. We ask each of them to perform the labeling task separately, and determine the ground truth by majority vote. Although the labels from the experts are not completely correct, we use them to reflect the quality of annotations to a certain extent. We measure the recall of the annotations using

$$R = N_{\text{sampled}}(I_i)/N_{\text{groundtruth}}(I_i), \tag{1}$$

where $N_{\text{sampled}}(I_i)$ is the number of positive elements in sampled image I_i and $N_{\text{groundtruth}}(I_i)$ is the total number of elements in the corresponding ground truth image. Positive elements refer to those bounding boxes that contain the graphical elements or texts in the ground truth. Following the computation, we obtain an average recall of over 96%, which means that most of the elements are correctly labeled by

Table 1 Collected labels in our dataset

Element	Tag	Count
Elements in "Subject"	Product	16 057
Elements in "Text"	Description	44 835
	Logo	4955
	Promotion	20 168
Elements in "Background"	Background	13 280
	Sub-background	2625

the annotators. We then measure the precision of the tags:

$$P_{\mathrm{t}}=\frac{N_{\mathrm{tag}}^{\mathrm{c}}(I_i)}{N_{\mathrm{tag}}^{\mathrm{c}}(I_i)+N_{\mathrm{tag}}^{\mathrm{w}}(I_i)+N^{\mathrm{w}}(I_i)}, \tag{2}$$

where $N^{\mathrm{w}}(I_i)$ is the number of negative elements in sampled image I_i, and $N_{\mathrm{tag}}^{\mathrm{c}}(I_i)$ and $N_{\mathrm{tag}}^{\mathrm{w}}(I_i)$ are the numbers of positive elements with correct or incorrect tags, respectively. To measure the precision of the labeled bounding boxes, we compute IoU between the annotation and ground truth:

$$P_{\mathrm{b}}=\frac{S_{\mathrm{sampled}}(I_i)\cap S_{\mathrm{groundtruth}}(I_i)}{S_{\mathrm{sampled}}(I_i)\cup S_{\mathrm{groundtruth}}(I_i)}, \tag{3}$$

where $S(I_i)$ is the area of labeled pixels in the sampled image or ground truth image. The average scores for tags and bounding boxes are 89.5% and 83.2%, respectively. Because of the inherent ambiguity of graphic design, multiple tags are appropriate for a particular element in some cases (e.g., "description" and "logo" are both appropriate when the semantics of the text are related to the description of the given product, but its layout is similar to a logo). Similarly, the ambiguity of an element influences the precision of the bounding boxes (Fig. 3).

4 Probabilistic graphical model

To learn the design attributes in the training data, we propose two probabilistic graphical models to capture the stylistic properties for the automatic synthesis of advertising images.

4.1 Local graphical model

The local graphical model is used to estimate the distribution of design attributes in those elements

Fig. 3 Examples of outline ambiguity in the dataset
The green line represents the ground truth and the red line is the annotation from our dataset. It is difficult to decide a certain outline for some ads because of the ambiguity existing between elements. References to color refer to the online version of this figure

that have independent performance in color and layout. For a particular element e, we first obtain properties $\boldsymbol{x}_e^i$ of certain design attributes, and then convert the discrete distribution of these properties into a continuous distribution using kernel density estimation (KDE):

$$\hat{f}(\boldsymbol{x})=\frac{1}{n}\sum_{i=1}^{n}K_{\boldsymbol{H}}(\boldsymbol{x}-\boldsymbol{x}_e^i)=\frac{1}{n\,|\boldsymbol{H}|^{1/2}}\sum_{i=1}^{n}K\left(\frac{\boldsymbol{x}-\boldsymbol{x}_e^i}{\boldsymbol{H}^{1/2}}\right), \tag{4}$$

where $\boldsymbol{x}=(x_1,x_2,...,x_d)^{\mathrm{T}}$ is a d-variate vector, $\{\boldsymbol{x}_e^1,\boldsymbol{x}_e^2,...,\boldsymbol{x}_e^n\}$ is an independent sample of focal design attributes, and $\boldsymbol{H}$ is the bandwidth of the function and is a $d\times d$ matrix that is symmetric and positive definite. Because we do not confirm the underlying distribution of these properties, we cross-validate empirically to determine the best bandwidth. K is an evaluation kernel that is a symmetric multivariate density (Scott and Sain, 2005):

$$K\left(\frac{\boldsymbol{x}-\boldsymbol{x}_e^i}{\boldsymbol{H}^{1/2}}\right)=\frac{\exp\left(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{x}_e^i)'\boldsymbol{H}^{-1}(\boldsymbol{x}-\boldsymbol{x}_e^i)\right)}{(2\pi)^{d/2}}. \tag{5}$$

4.2 Conditional graphical model

In graphic design, the design attributes of an element are likely to be affected by known features. For example, the perceived color can be affected by background color, appearing more or less saturated. Thus, we propose a conditional graphical model.

To estimate design properties $\boldsymbol{x}_e$ (e.g., the lightness of the text) with known features $\boldsymbol{C}_e$ (e.g., the color of the background), we first use k-means clustering (k=10) to discretize all $\boldsymbol{x}_e$ into a finite number of clusters n_e. Then, we use conditional feature $\boldsymbol{C}_e$ and cluster label $\boldsymbol{n}_e$ to train a multinomial logistic regression (MNR) classifier, based on the list $\{\boldsymbol{x}_e^i,\boldsymbol{C}_e^i,\boldsymbol{n}_e^i\}$ for focal element e_i. This classifier could predict the probability of a cluster when giving a never-before-seen conditional feature. We use the corresponding probability as its density and place a Gaussian kernel at each cluster's center, the widths of which are the cluster's standard deviations σ. We finally obtain a continuous probability distribution:

$$P(\boldsymbol{x}|\boldsymbol{C}_e)=\sum_{i=1}^{10}\exp\left(\frac{-\left\|\boldsymbol{x}-\boldsymbol{n}_e^i\right\|^2}{2\sigma^2}\right)\cdot p(\boldsymbol{n}_e^i|\boldsymbol{C}_e). \tag{6}$$

This method was first introduced by Charpiat et al. (2008) and applied later in Lin et al. (2013) and

Gu and Lou (2016). In our previous work, we showed the capability of two graphical models by estimating the distributions of lightness and saturation, incorporating the features of the graphic element's role to predict the design performance in the target context (You et al., 2019). Using the local graphical model and conditional graphical model, we further solve the problem of layout synthesis (described in Section 5).

4.3 Sampling

To sample high-probability properties from graphical models, design properties $\boldsymbol{x}_e$ are first clustered into a finite group using mean shift clustering (bandwidth = 0.1). Based on the clusters selected according to their probabilities, we place a Gaussian distribution $v \sim \mathcal{N}(0, \sigma)$ at the center of the selected cluster for sampling, where σ is its standard deviation of the properties. We obtain N samples for each graphical model ($N = 500$), and the samples that result in the highest score for the corresponding models are accepted.

5 Methods

The automatic ad synthesis process is divided into layout synthesis and image recoloring. We first synthesize the product and text input into a layout according to user preference. Then, a variety of color suggestions are generated based on the context-related keywords or color of the product. We use these suggestions to guide element recoloring to output the final results.

5.1 Layout synthesis

Fig. 4 shows the layout synthesis process. With the extracted attributes of the panel layout in the dataset, we filter out the product ads that contain more than one product and divide them into different clusters using mean shift clustering (Algorithm 2). We discard the clusters with few ads and subjectively assign the remaining clusters to different types of layout (e.g., left-right and top-down). The advertising images in the user-desired type are used as training data for probabilistic estimation. Using the local graphical model, we estimate the layout attributes $(x_\mathrm{p}, y_\mathrm{p}, s_\mathrm{p}, r_\mathrm{p})$ for the image panel, where $(x_\mathrm{p}, y_\mathrm{p})$ is the center position, and $(s_\mathrm{p}, r_\mathrm{p})$ denotes the size and ratio of the panel. To reduce the computational complexity, we estimate $(s_\mathrm{p}, r_\mathrm{p})$ and $(x_\mathrm{p}, y_\mathrm{p})$, separately. In particular, considering the conditional dependence between two paired attributes, we filter out the training data that are different from the computed $(s_\mathrm{p}, r_\mathrm{p})$ for the estimation of $(x_\mathrm{p}, y_\mathrm{p})$. Then, we use the conditional graphical model to estimate the layout attributes of the ith text panel (in order of importance) by $P(\boldsymbol{x}_i|\boldsymbol{C}_i)$, where $\boldsymbol{C}_i = (x_\mathrm{p}, y_\mathrm{p}, s_\mathrm{p}, r_\mathrm{p}, i)$ consists of the computed $(x_\mathrm{p}, y_\mathrm{p}, s_\mathrm{p}, r_\mathrm{p})$ of the image panel and label i of the target text panel. With the estimated (x_i, y_i, s_i, r_i) of each panel P_i for texts, we infer the initial layout of the panels and then consider the composition of the raw content within each panel.

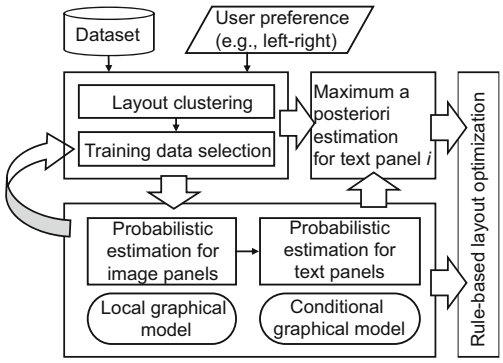

Fig. 4 Framework of the layout synthesis method

Algorithm 2 Layout clustering

Input: set of ads A_n with panel p_1 for product and largest panel p_2 for text.
Output: set of clustered results C_m and c_center.

```
Initialize C_m, c_center, D;
for each a_i in A_n do
    Compute distance d_i and angle θ_i between the centers of p_1 and p_2 in a_i;
    if d_i < 0 then
        Add a_i to c_center;
    else
        Add (cos θ_i, sin θ_i) to D;
    end if
end for
C_m ← mean shift clustering result of data D;
return c_center, C_m.
```

We place the input product at the center of the image panel as the product element. Considering that most of the ads in our dataset are China-specific, we assume that all the input texts consist of Chinese characters and that each character has the same width and height. Formally, let T_i be the input text for text panel P_i sorted by user-specified importance and $L_i = \{(h_1, d_1), (h_2, d_2), \ldots, (h_n, d_n)\}$ be the layout attributes that consist of n text elements.

For each text element, h is the height of the Chinese characters and d is the number of lines. We use the maximum a posteriori (MAP) method to determine the best-fitting text layout:

$$\begin{aligned} L_i^* &= \arg\max_{L_i} p(L_i|T_i) \\ &= \arg\max_{L_i} \left(\log p(T_i|L_i) + \log p(L_i)\right). \end{aligned} \tag{7}$$

Using the local graphical model, we compute the prior term $p(L_i)$:

$$p(L_i) \propto \prod_{k=1}^{n} \hat{f}(h^k, d^k|n, s_i, r_i). \tag{8}$$

We collect the panel attributes with similar size s_i and ratio r_i of P_i that consist of n text elements to train the local graphical model. Then, we define the likelihood term

$$p(L_i) \propto \prod_{k=1}^{n} \exp\left(-\frac{1}{2\sigma^2}\left|m_i^k - h_i^k/d_i^k\right|^2\right), \tag{9}$$

where m_i^k is the number of Chinese characters in the kth text element, the layout attributes of which are h_i^k and d_i^k. Following the method in Cao et al. (2012), we solve Eq. (9) by sampling a set of layouts from the local graphical model $L_i \sim \hat{f}(L_i)$ and selecting the top five that maximize the MAP estimation term as results. After placing the input product and text in each panel, we further refine the layout using several esthetic rules (O'Donovan et al., 2014). For example, the alignment rule is used to align text, the white space rule is used to adopt intervals between text lines, and the importance rule is considered when placing text at the bottom of the product. Fig. 5 shows an example of layout synthesis.

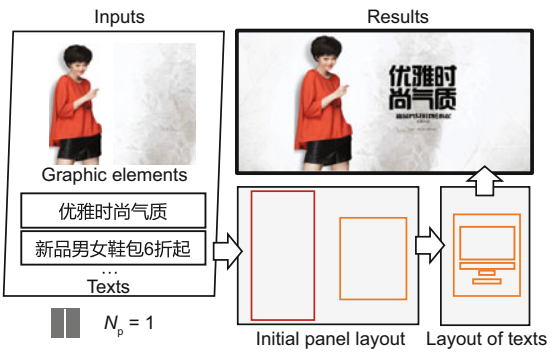

Fig. 5 Workflow of automatic layout synthesis

5.2 Image recoloring

The color of an ad often has a significant impact on marketing; designers use color strategically to influence the brand messaging of an ad and the purchase intent of the audience (Labrecque and Milne, 2012). To achieve a good color combination among the input text and graphic elements (i.e., product and background), we recolor each element automatically according to a coloring suggestion generated from the graphical models. For each advertising image in the training data, we translate the principal colors in the background into the hue, saturation, and value (HSV) color space, and use these discrete color attributes to train a local graphical model. We obtain the top five colors from the local graphical model, each of which is used as a feature to predict the text color using a conditional graphical model. We simplify the coloring problem by assuming that all the texts in the synthesized ad have the same color. For each labeled text element in the training data, we extract color properties $\boldsymbol{x}_i$ of the text and corresponding background colors $\boldsymbol{C}_i$, which result in a set of samples $\{\boldsymbol{x}_i, \boldsymbol{C}_i\}$ for training a conditional graphical model. With the computed color of the background as conditional feature $\boldsymbol{C}_{\text{b}}$, we obtain the text color using $P(\boldsymbol{x}|\boldsymbol{C}_{\text{b}})$.

To apply the generated color suggestions, we recolor the input graphic elements by changing the chromatic properties and lightness value. For chromatic properties, we use the recoloring method of linear template mapping in Seo et al. (2013), which manipulates global chromatic properties according to multiple reference colors. Because only one color is provided for each element in the coloring suggestions, we modify the original algorithm and map the center of the largest color cluster to the reference color. We adopt a strategy that the pixels whose lightness is close to the reference have larger adjustments than the pixels on the sides. Details were described in You et al. (2019). Fig. 6 shows an example of image recoloring.

6 Results and evaluation

In the implementation, we use 8021 clothing ads as the training data. For layout synthesis, we infer the design attributes for the layouts using a minimum number of text panels ($N_{\text{p}} \leq 2$) and split the

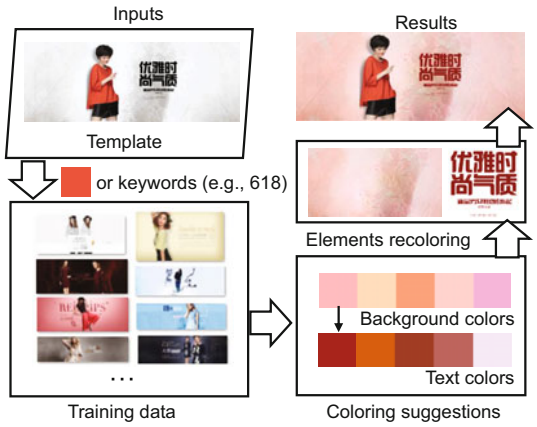

Fig. 6 Workflow for automatic image recoloring

input text into different panels randomly for MAP estimation. Users can guide the style of the results by selecting a specified type of layout. Fig. 7 shows the results with different types of layouts and numbers of text panels. Because most of our target audience has no experience in graphic design, we recruit 20 graduate students to participate in a perceptual study to understand how the synthesized ads compared with ads made by people. For each task, the participants are shown nine ads with a corresponding type of layout, of which three ads are created by students inexperienced in design, three by professional students, and three using our method. We prepare 36 ads as test material and ask the participants to rate the images using a five-point Likert scale.

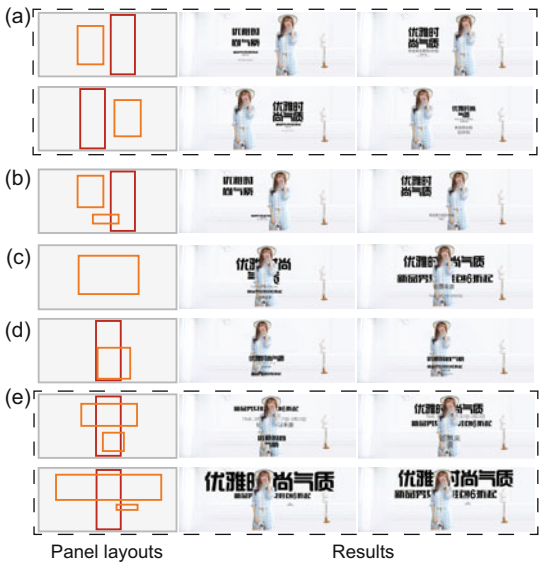

Fig. 7 Results with different types of layout and numbers of text panels (N_p): (a) left-right and $N_\mathrm{p} = 1$; (b) left-right and $N_\mathrm{p} = 2$; (c) center and $N_\mathrm{p} = 1$; (d) top-down and $N_\mathrm{p} = 1$; (e) top-down and $N_\mathrm{p} = 2$

We run a multivariate analysis of variance test using the creator of the ads and type of layout as dependent variables. The analysis reveals a significant difference among the professional, nonprofessional, and synthesized ads ($F = 4.874$, $p = 0.017$), but the type of layout shows no significant difference ($F = 0.214$, $p = 0.886$). The average score of our results is 3.208, which is higher than 2.996 for the nonprofessional students with $p = 0.069$; as such, our average score has an improvement of 7.1%. However, our results do not achieve a score commensurate with the score of 3.531 achieved by the ads created by professional students. For image recoloring, we apply our method to different sets of training data to generate colors with a specified style. To produce esthetically satisfying colors, we collect ads in which the Euclidean distance of the product color is smaller than the threshold as training data for probabilistic estimation. Fig. 8 shows the top four results for images of products with different colors. Because the generated colors need to accommodate various target contexts, we collect images of ads whose brief descriptions match the input keywords and use these images to train the models. Fig. 9 shows the results

Fig. 8 Advertising images generated with different product colors

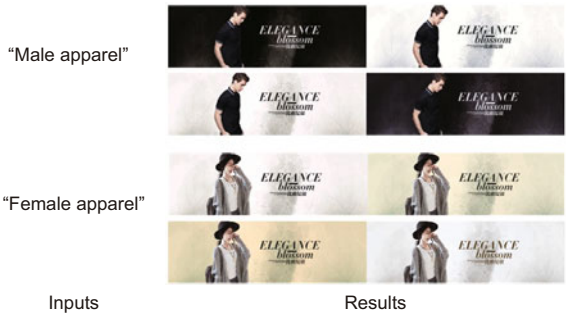

Fig. 9 Colors of advertisements generated with different input keywords for the target context

with keyword "male apparel" or "female apparel." To compare our coloring results with those of the other models, we conduct another user perception study. We obtain coloring suggestions from three resources, our method, the color harmony model (Yang et al., 2016), and Colormind (http://colormind.io/), which is a color scheme generator based on GAN. We prepare six templates with different colors of products and obtain coloring suggestions from different resources. When using the color harmony model, we apply the harmonic template including types "i," "L," "I," and "V," separately (Li X et al., 2015), and adopt an extended tone template (Tokumaru et al., 2002) to ensure sufficient visual contrast. When obtaining a five-color palette from Colormind, we lock the principal color of product to generate four other colors. We use each generated color in the scheme for the background, and randomly choose one other color for the text (some examples are presented in Fig. 10).

We invite 27 participates to attend another perceptual study. A random selection of 12 results generated from different sources is shown each time. We require participants choose four color schemes they like and four color schemes they dislike. Fig. 11 shows the percentage of images chosen as "like," "others," and "dislike" from different sources. A chi-squared test shows that the source of a coloring suggestion significantly affects its ranking in terms of preference by the subjects ($\chi^2 = 154.18, p < 0.01$). The coloring results learned from our dataset contain more "like" and fewer "dislike" images. The percentage of "top" images is 51%, which is improved by 22% and 31% compared with those obtained by the color harmony model and Colormind, respectively.

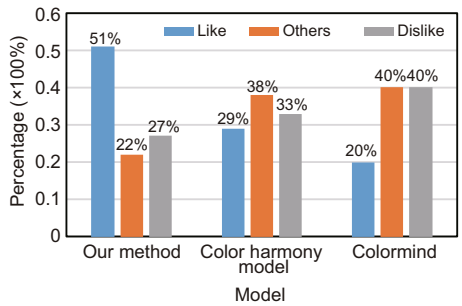

Fig. 11 Percentage of images selected from different models

7 Conclusions and future work

In this paper, we created a dataset for automatic synthesis of advertising images. This dataset contained basic information pertaining to images used for advertising, manually labeled bounding boxes, tags for their elements, and computed panels and colors. Using probabilistic graphical models, we tested the capability of our dataset by performing the tasks

Fig. 10 Comparison of different methods in color generation. Reprinted from You et al. (2019), Copyright 2019, with permission from Elsevier Ltd.

of layout synthesis and image recoloring with input graphic elements and texts to automatically generate ads. The styles of the synthesized ads were determined by different sets of training data selected according to user preference. Independent user perception experiments suggested that our method is effective. The results showed that our method improved users' satisfaction by 7.1% compared with designs from nonprofessional students, and more users preferred the coloring results of our designs to those generated by the color harmony model and Colormind.

A limitation of our method is that it considers only the general type of advertising images with basic groups of elements. There is still much work to synthesize ads with more complicated layouts and colors. To output ads that are more similar to those created by professional designers, we will need to improve the quality and quantity of the annotated ads, in addition to applying more powerful methods to capture the detailed distribution of different design attributes. In future work, we plan to follow the work of Li J et al. (2019) and use GANs to capture more discriminative features in advertising images according to the visual and textual semantics of the user inputs (Zheng et al., 2019). We are also considering to elaborate on the tags and labels with more detailed elements, such as graphic patterns, to support the synthesis of ads with more requisite design attributes.

Contributors

Wei-tao YOU and Zhi-yuan YANG designed the research. Wei-tao YOU and Hao JIANG processed the data. Wei-tao YOU drafted the manuscript. Hao JIANG and Ling-yun SUN helped organize the manuscript. Hao JIANG, Chang-yuan YANG, and Ling-yun SUN revised and finalized the paper.

Compliance with ethics guidelines

Wei-tao YOU, Hao JIANG, Zhi-yuan YANG, Chang-yuan YANG, and Ling-yun SUN declare that they have no conflict of interest.

References

Antonacopoulos A, Bridson D, Papadopoulos C, et al., 2009. A realistic dataset for performance evaluation of document layout analysis. 10th Int Conf on Document Analysis and Recognition, p.296-300. https://doi.org/10.1109/ICDAR.2009.271

Borji A, Cheng MM, Jiang H, et al., 2015. Salient object detection: a benchmark. *IEEE Trans Image Process*, 24(12):5706-5722. https://doi.org/10.1109/TIP.2015.2487833

Buldas A, Kroonmaa A, Laanoja R, 2013. Keyless signatures' infrastructure: how to build global distributed hash-trees. Nordic Conf on Secure IT Systems, p.313-320. https://doi.org/10.1007/978-3-642-41488-6_21

Cao Y, Chan AB, Lau RW, 2012. Automatic stylistic manga layout. *ACM Trans Graph*, 31(6):141-151. https://doi.org/10.1145/2366145.2366160

Charpiat G, Hofmann M, Schölkopf B, 2008. Automatic image colorization via multimodal predictions. European Conf on Computer Vision, p.126-139. https://doi.org/10.1007/978-3-540-88690-7_10

Choi J, Kim C, 2016. Object-aware image thumbnailing using image classification and enhanced detection of ROI. *Multim Tools Appl*, 75(23):16191-16207. https://doi.org/10.1007/s11042-015-2926-5

Geigel J, Loui A, 2003. Using genetic algorithms for album page layouts. *IEEE Multim*, 10(4):16-27. https://doi.org/10.1109/MMUL.2003.1237547

Gu Z, Lou J, 2016. Data driven webpage color design. *Comput Aid Des*, 77:46-59. https://doi.org/10.1016/j.cad.2016.03.001

Hussain Z, Zhang M, Zhang X, et al., 2017. Automatic understanding of image and video advertisements. IEEE Conf on Computer Vision and Pattern Recognition, p.1705-1715. https://doi.org/10.1109/CVPR.2017.123

Jahanian A, Liu J, Tretter DR, et al., 2012. Automatic design of magazine covers. IS&T/SPIE Electronic Imaging, Article 83020N. https://doi.org/10.1117/12.914596

Jiang B, Zhang L, Lu H, et al., 2013. Saliency detection via absorbing Markov chain. IEEE Int Conf on Computer Vision, p.1665-1672. https://doi.org/10.1109/TIP.2017.2766787

Labrecque LI, Milne GR, 2012. Exciting red and competent blue: the importance of color in marketing. *J Acad Mark Sci*, 40(5):711-727. https://doi.org/10.1007/s11747-010-0245-y

Li J, Xu T, Zhang J, et al., 2019. LayoutGAN: generating graphic layouts with wireframe discriminator. https://arxiv.org/abs/1901.06767

Li X, Zhao H, Nie G, et al., 2015. Image recoloring using geodesic distance based color harmonization. *Comput Vis Media*, 1(2):143-155. https://doi.org/10.1007/s41095-015-0013-5

Li Y, Qi H, Dai J, et al., 2017. Fully convolutional instance-aware semantic segmentation. IEEE Conf on Computer Vision and Pattern Recognition, p.2359-2367. https://doi.org/10.1109/CVPR.2017.472

Lin S, Hanrahan P, 2013. Modeling how people extract color themes from images. SIGCHI Conf on Human Factors in Computing Systems, p.3101-3110. https://doi.org/10.1145/2470654.2466424

Lin S, Ritchie D, Fisher M, et al., 2013. Probabilistic color-by-numbers: suggesting pattern colorizations using factor graphs. *ACM Trans Graph*, 32(4):37. https://doi.org/10.1145/2461912.2461988

Liu KL, Li W, Yang CY, et al., 2019. Intelligent design of multimedia content in Alibaba. *Front Inform Technol Electron Eng*, 20(12):1657-1664. https://doi.org/10.1631/FITEE.1900580

O'Donovan P, Agarwala A, Hertzmann A, 2011. Color compatibility from large datasets. *ACM Trans Graph*, 30(4):63-75. https://doi.org/10.1145/2010324.1964958

O'Donovan P, Agarwala A, Hertzmann A, 2014. Learning layouts for single-pagegraphic designs. *IEEE Trans Vis Comput Graph*, 20(8):1200-1213. https://doi.org/10.1109/TVCG.2014.48

O'Donovan P, Agarwala A, Hertzmann A, 2015. DesignScape: design with interactive layout suggestions. 33rd Annual ACM Conf on Human Factors in Computing Systems, p.1221-1224. https://doi.org/10.1145/2702123.2702149

Qiang YT, Fu YW, Yu X, et al., 2019. Learning to generate posters of scientific papers by probabilistic graphical models. *J Comput Sci Techn*, 34(1):155-169. https://doi.org/10.1007/s11390-019-1904-1

Russell BC, Torralba A, Murphy KP, et al., 2008. Labelme: a database and web-based tool for image annotation. *Int J Comput Vis*, 77(1-3):157-173. https://doi.org/10.1007/s11263-007-0090-8

Scott DW, Sain SR, 2005. 9 – multidimensional density estimation. *Handbook Stat*, 24:229-261. https://doi.org/10.1016/S0169-7161(04)24009-3

Seo S, Park Y, Ostromoukhov V, 2013. Image recoloring using linear template mapping. *Multim Tools Appl*, 64(2):293-308. https://doi.org/10.1007/s11042-012-1024-1

Tang YC, Huang JJ, Yao MT, et al., 2019. A review of design intelligence: progress, problems, and challenges. *Front Inform Technol Electron Eng*, 20(12):1595-1617. https://doi.org/10.1631/FITEE.1900398

Tokumaru M, Muranaka N, Imanishi S, 2002. Color design support system considering color harmony. IEEE World Congress on Computational Intelligence, p.378-383. https://doi.org/10.1109/FUZZ.2002.1005020

Tuch AN, Bargas-Avila JA, Opwis K, 2010. Symmetry and aesthetics in website design: it's a man's business. *Comput Human Behav*, 26(6):1831-1837. https://doi.org/10.1016/j.chb.2010.07.016

Wang TC, Liu MY, Zhu JY, et al., 2018. High-resolution image synthesis and semantic manipulation with conditional GANs. IEEE Conf on Computer Vision and Pattern Recognition, p.8798-8807. https://doi.org/10.1109/CVPR.2018.00917

Xu T, Zhang P, Huang Q, et al., 2018. AttnGAN: fine-grained text to image generation with attentional generative adversarial networks. IEEE Conf on Computer Vision and Pattern Recognition, p.1316-1324. https://doi.org/10.1109/CVPR.2018.00143

Yang X, Mei T, Xu YQ, et al., 2016. Automatic generation of visual-textual presentation layout. *ACM Trans Multim Comput Commun Appl*, 12(2):33-55. https://doi.org/10.1145/2818709

You WT, Sun LY, Yang ZY, et al., 2019. Automatic advertising image color design incorporating a visual color analyzer. *J Comput Lang*, 55:100910. https://doi.org/10.1016/j.cola.2019.100910

Yu LF, Yeung SK, Tang CK, et al., 2011. Make it home: automatic optimization of furniture arrangement. *ACM Trans Graph*, 30(4):86:1-86:12. https://doi.org/10.1145/2010324.1964981

Zhang Y, Hu K, Ren P, et al., 2017. Layout style modeling for automating banner design. Thematic Workshops of ACM Multimedia, p.451-459. https://doi.org/10.1145/3126686.3126718

Zheng X, Qiao X, Cao Y, et al., 2019. Content-aware generative modeling of graphic design layouts. *ACM Trans Graph*, 38(4):133. https://doi.org/10.1145/3306346.3322971

Zhu JY, Park T, Isola P, et al., 2017. Unpaired image-to-image translation using cycle-consistent adversarial networks. IEEE Int Conf on Computer Vision, p.2223-2232. https://doi.org/10.1109/ICCV.2017.244