Vinci: An Intelligent Graphic Design System for Generating Advertising Posters

Shunan Guo
iDVX Lab, Tongji University
g.shunan@gmail.com

Zhuochen Jin
iDVX Lab, Tongji University
chjzcjames@gmail.com

Fuling Sun
iDVX Lab, Tongji University
fulingsun.idvx@gmail.com

Jingwen Li
iDVX Lab, Tongji University
jinwere.lee@gmail.com

Zhaorui Li
iDVX Lab, Tongji University
zhaorayli@tongji.edu.cn

Yang Shi
iDVX Lab, Tongji University
yangshi.idvx@tongji.edu.cn

Nan Cao
iDVX Lab, Tongji University
nan.cao@gmail.com

Figure 1: Vinci system employs Artificial Intelligence techniques to automatically generate advertising posters based on a product image and several taglines. This figure demonstrates a gallery that exhibits the posters generated by the system.

ABSTRACT

Advertising posters are a commonly used form of information presentation to promote a product. Producing advertising posters often takes much time and effort of designers when confronted with abundant choices of design elements and layouts. This paper presents Vinci, an intelligent system that supports the automatic generation of advertising posters. Given the user-specified product image and taglines, Vinci uses a deep generative model to match the product image with a set of design elements and layouts for generating an aesthetic poster. The system also integrates online editing-feedback that supports users in editing the posters and updating the generated results with their design preference. Through a series of user studies and a Turing test, we found that Vinci can generate posters as good as human designers and that the online editing-feedback improves the efficiency in poster modification.

CCS CONCEPTS

• **Computer systems organization** → **Embedded systems**; *Redundancy*; Robotics; • **Networks** → Network reliability.

KEYWORDS

Graphic design, deep generative networks, design tool

ACM ISBN 978-1-4503-8096-6/21/05…$15.00
https://doi.org/10.1145/3411764.3445117

ACM Reference Format:
Shunan Guo, Zhuochen Jin, Fuling Sun, Jingwen Li, Zhaorui Li, Yang Shi, and Nan Cao. 2021. Vinci: An Intelligent Graphic Design System for Generating Advertising Posters. In *CHI Conference on Human Factors in Computing Systems (CHI '21), May 8–13, 2021, Yokohama, Japan.* ACM, New York, NY, USA, 17 pages. https://doi.org/10.1145/3411764.3445117

1 INTRODUCTION

Advertising posters are a form of information presentation combing images and overlaid text [26]. It is a graphic design that is commonly used to promote a product by demonstrating its features and communicating intended benefits to the target audience [23]. Creating an effective poster can be difficult and time-consuming for designers when facing an overwhelming space of design choices [27, 30]. Given an image of a product and the text descriptions, designers often need to explore a large range of design elements (e.g., background images, embellishments, text styles) to find matches for the advertised product and arrange the design elements to satisfy aesthetic goals. Not to mention that designers are sometimes required to create posters in large quantities [6]. Despite that some techniques were proposed to automatically generate graphical designs, they mostly focus on arranging the layout of user-specified images and texts [25, 29, 39, 41]. The design process of advertising posters, however, is beyond deciding the layouts and styles of predefined design elements. Designers also need to select proper design elements in the context of a given product to enrich the content of the poster and ensure that all the elements are good matches to each other in terms of both semantics and styles. For example, typical considerations from designers when creating an advertising poster can be *"which background should I use to make the product more prominent"*, and *"how to coordinate embellishments to enhance the informativeness of my design?"*

Recent advances in deep learning enable Artificial Intelligence (AI) to assist designers in generating aesthetic and creative visual content. For example, some deep learning models have been successfully applied to image synthesis [18, 49], style transfer [5, 40], and assisted drawing [4, 33, 43]. However, to the best of our knowledge, no prior published study targets automatic creating visual content for advertising posters. To automatically generate advertising posters, three challenges need to be specifically addressed:

C1 *How to formulate the human design process in a way that machine can understand.* The design process of an advertising poster is complex and abstract, including a series of design decisions on a variety of design elements. It requires a technique that can transform this design process into a machine learning model that captures key design steps and common collocation of design elements.

C2 *How to select design elements that match in both semantics and visual style.* When selecting design elements, a designer need to ensure that the context of the selected elements is in accordance with the semantic given by the product image, and make sure the visual styles, such as colors, shapes and sizes of the design elements are in perfect harmony. For example, a poster promoting fruits can benefit from embellishments such as leaves and sunshine to convey fresh and organic impressions. Also, fruits, leaves, and sunshine all have the design style – "natural", thus they shall form a harmonious combination.

C3 *How to arrange design elements on the canvas.* The layout of design elements can affect how the product image is perceived [25]. An ideal layout should emphasize the product image, and meet certain aesthetic goals (e.g., clean, non-overlapping, symmetrical).

To address these challenges, we introduce Vinci, an intelligent design system for automatically generating advertising posters. Given a product image and the text descriptions specified by the user, Vinci selects design elements that match the product and arranges them on a canvas to generate an advertising poster. To inform the process of designing an advertising poster, we first formulate the design space of advertising posters through interviews with graphic designers. We identify four dimensions of the design space: object, background, embellishment, and text. Each poster in our dataset is then transformed into a sequence of design choices in the design space. A deep learning model based on Sequence-to-Sequence Variational Autoencoder (VAE) [21], is employed to learn the patterns of design choices in the design sequences of human-designed posters. Vinci incorporates a web-based user interface, from which a user can easily modify a generated poster to refine its design. The system can generalize the editing operations to all generated posters to meet with the user's design preference. The direct contributions and the novel aspects of this work include:

- **Algorithm.** We present an intelligent framework for learning the patterns of design choices in human-designed advertising posters and generating new poster designs based on the user-input product image and taglines. In particular, we propose a design space to characterize the design elements in advertising posters and introduce the design sequence to formalize the design decisions of human designers in creating the posters.
- **System.** We introduce an intelligent graphic design system, Vinci, to support the interaction process of automatic poster generation. The system displays posters generated based on the product image and taglines uploaded by the user and allows the user to edit the posters by preference. Vinci is also featured with an *online-editing feedback* mechanism that can automatically make tweaks on all generated posters according to the user's design preferences that are reflected in the edits.
- **Evaluation.** We report results from three forms of user studies: (1) a user study verifying the effectiveness of different components in the poster generation model, (2) a Turing test evaluating the quality of the generated posters, and (3) a within-subject controlled user study testing the capability of the *online-editing feedback*. The results show that Vinci can generate advertising posters as good as human designers, and the *online-editing feedback* can improve the efficiency in poster editing. We also report feedback from professional designers in an expert interview.

2 RELATED WORK

This section reviews existing studies that are relevant to our research, including (1) automatic generation of information presentations, and (2) deep learning models for generating graphic designs.

2.1 Automatic Generation of Information Presentations

Information representations, such as graphic user interface, posters, and magazine covers, are widely used in our daily life as an essential medium of visual communication. To generate aesthetic graphic designs, designers often need to put significant time and effort. Therefore, an extensive range of techniques has been proposed to aid the design of information presentations. Many studies in this area focused on developing optimal layout algorithms that can automatically place images and text in proper positions by following certain aesthetic criteria [26]. One representative example is DesignScape [28], which is an interactive system for generating layout suggestions on a set of user-input images and texts. To enhance the interactivity of the automatic layout generation process, Todi et al. proposed SketchPlore [38], which is a real-time layout optimizer that can automatically infer a designer's action (e.g., place and arrange elements on canvas) and use predictive models to suggest more feasible layouts for generating a user interface. Tabata et al. successively proposed two works [36, 37] for automatically creating diverse candidates of magazine layouts that retain the original design styles given the user input texts, images and page numbers. A more recent work proposed by Zheng et al. [47] is capable of laying out images on a text document according to the content of the paragraphs. While these techniques are powerful in generating aesthetic text-image layouts, the layout rules derived from these techniques may not be applicable to generating advertising posters. Moreover, the analytical tasks of generating poster designs and text-image layouts are essentially different. In previous works of generating text-image layouts, all design materials are provided by the user, and the machine is trained only to determine the positions of these elements. In contrast, designing an advertising poster usually starts with very limited user-specified design elements (i.e., one product image and a few taglines). The machine needs to further select design elements (e.g., backgrounds, embellishments) that match with the product image and arrange all elements properly on the canvas to complete the design.

Some more complicated systems were developed with the focuses on other graphic design issues. For example, to generate the content of the information representation, Yin et al. [42] introduced an automatic social media snippets generation system that extracts a major picture and textual descriptions from a media corpus to compose a text overlaid image. Qiang et al. [29] introduced a data-driven framework for automatically generating a scientific poster by extracting content from a research paper. Zhao et al. [45] built an automatic icon generation and ideation system that is capable of creating compound icons by linking and arranging individual icons mapped from the keywords input by the user. Besides static graphic representation, Chi et al. [7] also explored the possibility of generating video from web page contents. Compared with these studies that aim at extracting or arranging critical contents from a source, our work focuses more on the specific design task of generating advertising posters, which is analyzing the semantic of the input product image to match proper design materials and place all design elements in proper positions on canvas for forming an aesthetic poster.

To produce a better and styled graphic design, Jahanian [16] proposed graphic design guidelines for magazine covers and designed a framework composed of three main modules, including the layout of cover elements, coloring, and typography of cover lines to approach these guidelines. Yang et al. [41] designed a system that automatically generates digital magazine covers by summarizing a set of topic-dependent templates and introducing a computational framework that incorporates the key elements of layout design. Although having a similar goal to our work, the systems above are usually rule-based with less intelligence. Advanced machine learning techniques are utilized to support generating information presentations more efficiently and intelligently. For example, Zheng et al. [46] proposed a system based on a deep style-embedding model to recommend the styled design templates for generating a web-based graphic user interface. Lei et al. [25] introduced LayoutGAN for laying out design elements into proper positions to form a scene. However, none of these techniques focuses on generating advertising posters, which have entirely different patterns in the choice and arrangement of design elements. Luban system[1] from Alibaba Group is most relevant to our work, which also uses deep learning techniques for generating advertising posters. However, there is no published research work regarding this technique. We compare the quality of posters generated by Luban and Vinci in our user experiment, and the result suggests that Vinci can generate posters with higher qualities.

2.2 Deep Generative Models for Graphic Design

The recent advances in deep learning enable the application of AI to graphic design and thus change the nature of creative processes. Style transfer generates artistic imagery by disentangling and recombining the image content and style. Inspired by the potential of convolutional neural network (CNN) [24], Gatys et al. [10] first proposed a CNN-based neural algorithm of artistic style, which is capable of generating new images with the content of a photograph and the style of artworks. Wang et al. [40] introduced a hierarchical deep CNN architecture that learns artistic style cues at multiple scales (color, texture, and brushwork) and achieves style transfer in nearly real-time. These CNN-based techniques use images as input and reconstruct results from pixel values. Compared to these pixel image modeling approaches, our work models posters as a sequence of design choices and generates higher quality results.

To better assist in generating content in graphic design, two generative models are frequently used, including Variational Autoencoder (VAE) [21] and generative adversarial network (GAN) [11]. For example, Ha and Douglas proposed Sketch-RNN [12], a Sequence-to-Sequence VAE that generates a vector representation of an image as a short sequence of strokes. As a subsequent effort, AI-Skecther [4] uses a CNN-based autoencoder to capture the positional information of each stroke, an influence layer to guide the generation of each stroke, and a conditional vector to facilitate multi-class sketch generation, supporting generating high-quality sketches. The work mentioned above represents sketches as a sequence of actions controlling a pen while our work models posters as a sequence of design choices selecting and organizing visual elements. Also, to ensure that the generated sequence of design choices matches the product image specified by the user, we feed the information of the product image into the encoder.

As for GAN-based methods, Zhu et al. [49] proposed a cycle-consistent adversarial network (CycleGAN) and applied it to style transfer, object transfiguration, and season transfer. Zhu [48] introduced a GAN-based model to learn the manifold of natural images, allowing for modifying the color and the shape of the generated

[1] https://luban.aliyun.com/

image. Jin et al. developed MakeGirlsMoe [18] that allows users to specify the desired features of female animators such as hair color and eye color and then create customized portraits. When compared with GAN-based techniques, our work requires relatively small-sized training data to generate high quality advertising posters.

3 INITIAL INTERVIEW AND SYSTEM OVERVIEW

We aim to design Vinci as a virtual bot that can mimic the behaviors of human graphic designers to create an advertising poster. To better understand the designers' workflow and design thinking when creating advertising posters, we conducted an in-depth interview with four experienced graphic designers to investigate the general workflows of designing an advertising poster. Based on the feedback from interviewees, we develop a web-based system driven by a pre-trained deep generative model for automatically generating advertising posters. In this section, we describe the interview procedure, discuss the feedback from participants, and give an overview to the Vinci system.

3.1 The Interview

We conducted a two-hour interview with four graphic designers(all females, aged 22–24). All designers had more than five years of graphic design experience. We asked each designer to create an advertising poster from scratch with a given product image and some optional taglines before the interview. With a goal of understanding the human designers' design process and habits, we set no specific limitation to designers in creating the posters. During the interview, we guided them to explain their design decisions with questions under three topics: the general workflow of creating the poster, how they searched and selected design element, and difficulties they have when designing an advertising poster according to their past design experience. We summarize their feedback as follows:

Workflow of creating the poster. According to our interviewees, there were three common steps they followed when making the poster. Given a product image and taglines, they first chose an essential tone of the poster that matches with the color of the product. For example, a product with predominately bright colors shall fit better with the bright tone than the dark one. In particular, designers mentioned that the tone of the poster is mainly determined by the choice of background image. The second step is to search and choose the matching embellishments to emphasize the main characteristics and selling points of the product. Finally, the designers arrange all the design elements (i.e., background image, product image, embellishments, taglines) on the canvas by adjusting their sizes, positions, and styles to generate an aesthetic poster. From the above feedback, we inferred a sequential ordering in the design decisions of a poster, and formalized this sequential workflow into a design sequence (later introduced in Section 4.2).

Collecting and selecting design elements. When discussing how to collect and select design elements, designers mentioned that they first searched for semantically associated objects with the product. For example, one designer used sea waves as background to associate seafood with the ocean, while the other designer put seafood in a dining scene to associate seafood with restaurants. They then selected the matching design elements by considering the essential tone determined, the size and aspect ratio of the design elements, and its relevancy to the product. From the above feedback, we recognized a requirement of considering element relevancy and semantic consistency when selecting the design elements, which inspired our model design (later introduced in Section 5).

Design difficulties. All the designers agreed that collecting design elements is the most time-consuming and labor intensive process. Specifically, one designer commented that even if they had certain ideas of what to look for, she said that "it is not easy to find design elements of high-quality, harmonious color and suitable aspect ratios". Another designer added that "It is more difficult to find harmonious combination of design elements. Sometimes we need to overturn the entire design intention when proper design elements are not to be found." The designers also brought up another issue that advertising posters are often required to be designed in different sizes for different advertising scenarios. For example, one designer explained that "outdoor posters and banners have different aspect ratios, which can largely influence the choices and arrangements of design elements. It is tedious and time-consuming to adjust the design elements when the size of the poster changes."

3.2 System Overview

Based on the feedback gathered from the initial interview, we develop an interactive web application (as shown in Fig. 2), Vinci , for automatically generating advertising posters with user-specific product image and taglines. The system contains two major functionality: automatic poster generation and online editing-feedback. In the following, we first present our design goals extracted from our discussion with the designers. Then, we illustrate how we achieve our design goals in the introductions of each system functionality.

Design goals. Vinci is designed based on the following goals. First, Vinci is designed to relieve designers from the tedious routine work, such as repeatedly choosing, comparing, and arranging the design elements when creating a poster design from scratch (**G1**), and enable Vinci to generate large quantities of posters in a short time (**G2**). We also aim to boost creative inspirations with various design suggestions (**G3**), as designers are sometimes required to create multiple design alternatives for different use cases (e.g., websites, indoor and outdoor) and to ensure the diversity of the advertisements.

Automatic poster generation. After loading the system website, the user can upload a image of the promoted product(Fig. 2(a)), and enter the product taglines, promotion and product descriptions(Fig. 2(b)). We tended not to incorporate other users' specifications, such as the choice of design elements and layout templates, but leave these tedious work to the automatic generation model, so as to save the designer's effort on iterating the design elements and enable efficient poster generation (**G1, G2**). After clicking the "Generate" button, the design materials are sent to a pre-trained deep generative model for automatic poster generation. Fig. 3 gives an overview of the model architecture, which will be detailed in Section 4 and Section 5. It consists of four major modules: a preprocessing module, a *generator*, a *reconstructor*, and an *estimator*. The preprocessing module extracts reusable design elements from a collection of human-designed advertising posters, and abstracts the

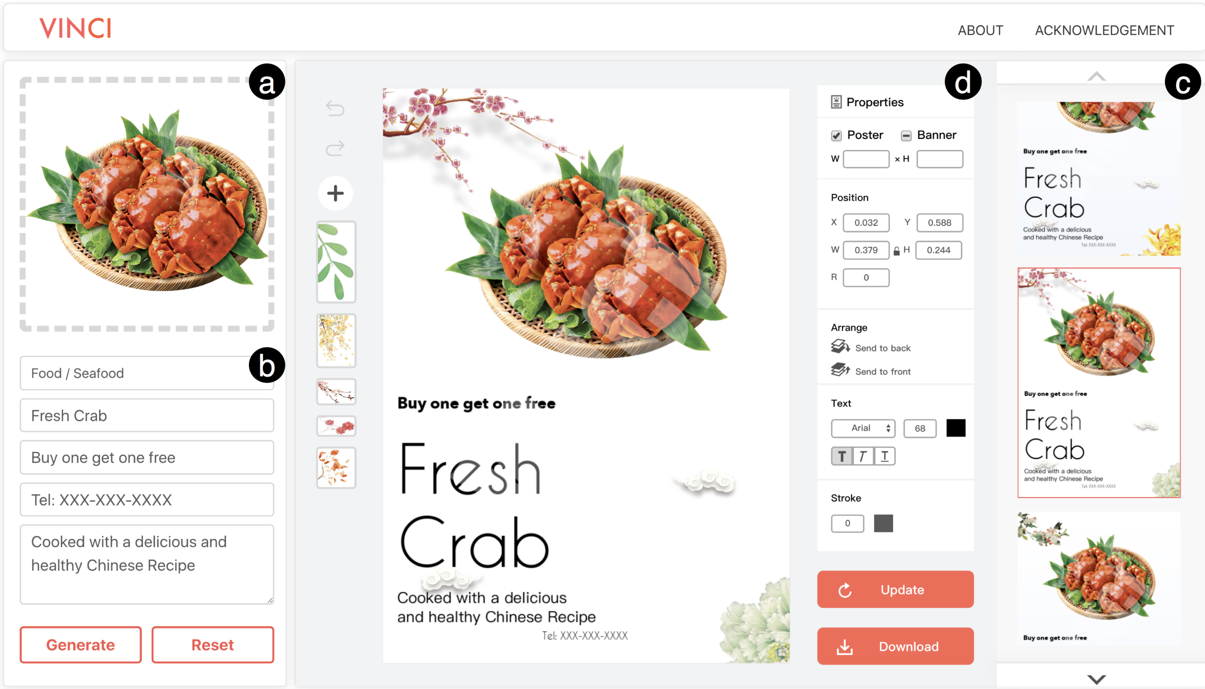

Figure 2: The online user interface of Vinci system consists of (a) an image input box for users to upload a product image; (b) the input box for text descriptions; (c) the list of generated posters; and (d) the edit panel, in which a selected poster can be further modified based on user's preference.

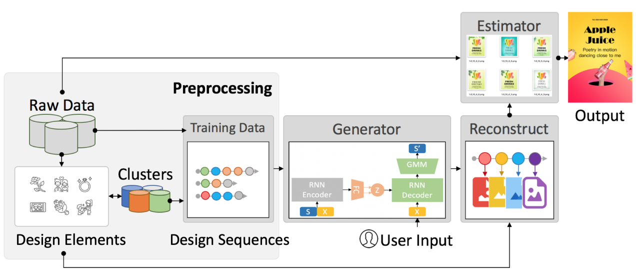

Figure 3: An overview of the running pipeline and the architecture of Vinci system.

design of each poster in a sequence of design choices (i.e., design sequence) to generate training data of the *generator*. The *generator* is the deep generative model that is trained to produce design sequences, and the *reconstructor* decodes each generated design sequence into a series of design elements in the material library and adjust their positions to form an advertising poster. The *estimator* is trained to rate the quality of the posters generated from the *reconstructor*. To inspire designers with various design options (**G3**), we preserve not only the poster with the highest quality, but outputs a list of high-quality ones to the user in a ranked list (Fig. 2(c)). Beyond the automated generation process,Vinci also provides free space for users to add their design specifications. The user can select one poster and make changes through the editing panel (Fig. 2(d)) and download the poster by clicking the "download" button.

Online editing-feedback. In addition to enabling users to make tweaks on a focal poster, Vinci is also designed to capture the intention of each modification (e.g., the change of an embellishment's size or position) and quickly broadcast the modifications to the rest of the generated posters. This function aims to help users modify the generated posters in batches, which is inspired by the second design difficulty – designers often need to change the sizes of the posters in their work. Specifically, when the user clicks "update", the system searches design elements similar to the modified ones in the rest of the posters and apply the corresponding modifications to them. The similarity is determined based on image features extracted based on a VGG-16 [31] pre-trained on the ImageNet [8] and the embellishment's original position in the poster. We evaluate online editing-feedback with a within-subject controlled user study in the usage scenario of resizing posters, so as to see whether this function can improve users' editing efficiency. The details of the study is introduced in Section 6.4.

4 PREPROCESSING

The preprocessing module parses human-designed posters stored in a layered format (i.e., the PSD format used in Adobe Photoshop) to extract and cluster reusable design elements (Fig. 3(1,2)). It then transforms the design of each poster into a design sequence that is composed of design elements in the order of background, embellishments, the object image, and text descriptions to train the *generator* (Fig. 3(3)). In this section, we first introduce a design space of the design elements in advertising posters, followed by the process of converting poster designs into design sequences.

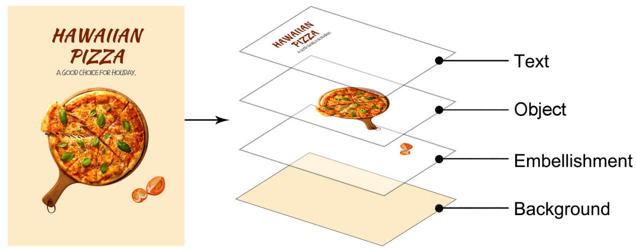

Figure 4: An advertising poster is composed of layers of design elements, including background, embellishment, object, and text.

4.1 Design Space

As suggested by the designers in our initial interview, the workflow of creating an advertising poster generally involves collecting and selecting design elements of four basic categories: the object image, the background, the embellishments, and the text descriptions. Therefore, we formalize the design space of an advertising poster into the following four dimensions:

- **D1 Background.** The background of a poster determines the primary color and is usually laid at the bottom. It can be filled with a solid color or an image with a certain texture that is in accord with the context of the product.
- **D2 Embellishment.** The embellishments, such as photos of product-related items, symbols, and borders, serve as complements to the product image. They are often used to enrich the content of the poster and bring out the distinctive features of the product. A poster can have multiple embellishments, and each embellishment is stored in an individual layer of a poster.
- **D3 Object.** The object refers to the product to be promoted, which is usually presented via a photo or an illustration in a poster. The overall design style of a poster and the choice of embellishment can be largely influenced by the category (e.g., food & drink, beauty & cleansing, electronics) and the appearance of the product. For example, food & drink advertisements tend to use "appetizing" colors [19] and avoid colors such as blue, black and purple.
- **D4 Text.** Adding text descriptions to the poster provides more information of the object such as the brand of the product, special promotions, and the contact information of the seller. The style (e.g., font, size, color) of texts usually hinges on the importance of the information.

Preprocessing training posters. As stated in Section 3.2, the *generator* of Vinci is trained to learn design patterns of human designers from a collection of human-designed advertising posters. For ease of identifying design elements, the collected advertising posters shall be stored in PSD format where design elements are clearly separated into different layers, and each graphical element is labeled as either background, embellishment, or product image. Each poster should also be labeled with a product category (e.g., electronics, groceries, etc.). Given that the design styles of advertising posters under different product categories may differ greatly, the model shall be trained with only one category of advertising posters at a time to ensure the homogeneity of the training data.

4.2 Design Sequence

To train a model that can automatically design advertising posters, we first abstract the design of human-created posters into a data format that is understandable to the machine. To this end, we leverage the structure of layers that formalize the design of posters (as shown in Fig. 4) and transform each preprocessed poster into a sequence of design elements derived from our design space. For example, the poster in Fig. 4 can be converted into a sequence of a yellow background, a tomato embellishment, a photo of the promoted product (i.e., a pizza), and text descriptions, from the bottom to the top. The order of elements is also matched with the design workflow mentioned in the initial interview. In this way, the design of an advertising poster can be formalized as a series of design elements arranged for the given product image.

However, this formalization can lead to high sparsity in data. Suppose that we have a library of 100 reusable design elements and 100 posters, then each poster can be converted to a unique design sequence that lies in a $100 \times 100 \times 4$ space. This cannot be easy for the model to extract common patterns of design elements. Not to mention that we also need to consider their positions in continuous space. To build design sequences that are more representative, we attempted to reduce the design space by categorizing design elements in each dimension. Specifically, we categorized the backgrounds and embellishments by extracting their visual features (e.g., color, shape, position, texture) using a VGG-16 network pre-trained on the ImageNet. This results in a group of clusters background clusters and embellishment clusters, each containing similar design elements. We also defined several layout templates for framing the layout of the product image and taglines by clustering the positions, size, and compositions of the product image and texts in the sample posters. In this way, the templates record the frequently used layout configurations and reduce the design choices in the layout dimensions. The layout of other design elements (backgrounds and embellishments) is not included in the template to allow for more diverse layouts of the entire poster.

After clustering the design elements and the layout for the product and taglines, a poster can be represented as a sequence of design choices on clusters in the background, and embellishment. Formally, given a poster of product image P_i, its corresponding design sequence is denoted as $\boldsymbol{S_i} = \{\boldsymbol{s}_1, \boldsymbol{s}_2, \cdots, \boldsymbol{s}_m\}$. In the sequence, each design step $\boldsymbol{s}_j = [\boldsymbol{c}_j, x_j, y_j]$ is a feature vector that captures a design choice on a specific dimension to generate a poster. Here, $\boldsymbol{c}_j \in \{0, 1\}^M$ is a one-hot vector that indicates the selection of element clusters. For example, $\boldsymbol{c}_{jk} = 1$ indicates that a design element in the k-th cluster is used at the j-th design step. Generalizing specific design elements to element clusters also prevent the model from learning the exact element combinations that exist in the original posters, so that the diversity of the generated designs can be guaranteed. The 2D-coordinates $[x_j, y_j]$ of this element in the poster are also recorded. Considering that each poster includes one background that fills the entire canvas, together with one or multiple embellishments to match with the product image, we set the first step in the sequence $\boldsymbol{s}_1$ to represent the design decision on the background dimension and the following steps in the sequence

$\{s_2, \cdots, s_{m-1}\}$ to represent the choices on the embellishment dimension. The design sequence ends with s_m that denotes the layout for text and object determined by the choice of layout template.

5 POSTER GENERATION

In this section, we introduce the model for generating advertising posters given a product image and the corresponding taglines. The model consists of three modules: (1) a *generator* trained to generate design sequences of the posters; (2) a *reconstructor* designed to select design elements based on the design sequences and arrange the positions of design elements to produce a poster; (3) a pre-trained *estimator* designed to rate the quality of the generated posters.

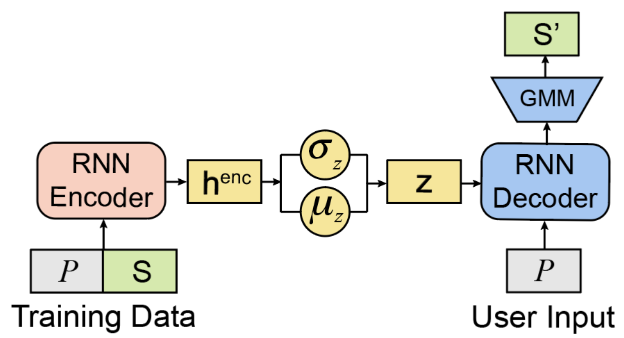

Figure 5: The *generator* of the Vinci system is a sequence-to-sequence Variational Autoencoder (VAE).

5.1 Generator

As shown in Fig. 5, the Generator is a sequence-to-sequence Variational Autoencoder (VAE) [3, 21], consisting of an encoder and a decoder that are based on Long Short-Term Memory recurrent neural network (LSTM-RNN) [14, 15]. The generator extracts the product image P and their corresponding design sequences S from each processed human-designed posters and feeds them to the encoder in batches for training the model. The encoder extracts a high-level latent vector $\boldsymbol{z}$ for representing the features of the training data. Based on the abstract information captured by the distribution of $\boldsymbol{z}$ and a user-specified product image P, the decoder generates a design sequence that constructs the design of the poster for the user-specific product image.

<u>VAE Encoder.</u> The encoder of the model is trained to capture the sequential distribution of design steps in the design sequences. In particular, a hidden vector h_e is trained to parameterize a multinomial distribution through the RNN:

$$\boldsymbol{h}_e = encoder(\boldsymbol{P}, \boldsymbol{S}) \tag{1}$$

where $\boldsymbol{P}$ is a collection of product images and $\boldsymbol{S}$ are their corresponding poster design sequences extracted from the training data. Here, $\boldsymbol{h}_e$ is the hidden state of the RNN, which is used to calculate the mean ($\boldsymbol{\mu}_z \in R^K$) and the standard deviation ($\boldsymbol{\sigma}_z \in R^K$) of the multinomial distribution as follows:

$$\boldsymbol{\mu}_z = W_\mu \cdot \boldsymbol{h}_e + \boldsymbol{b}_\mu,\ \boldsymbol{\sigma}_z = Exp(\frac{W_\sigma \cdot \boldsymbol{h}_e + \boldsymbol{b}_\sigma}{2}) \tag{2}$$

A latent vector $\boldsymbol{z} \in R^K$ is randomly sampled from the above distribution according to Eq. (3), which is later used in the decoder to construct the design sequence. The random sampling ensures that the decoded design sequence is non-deterministic (i.e., the model generates a new sequence instead of merely restoring the training design sequences).

$$\boldsymbol{z} = \boldsymbol{\mu}_z + \boldsymbol{\sigma}_z \cdot \boldsymbol{\epsilon} \quad where\ \boldsymbol{\epsilon} \sim N(0, 1) \tag{3}$$

<u>VAE Decoder.</u> In the decoder, we use an RNN together with a Gaussian Mixture Model (GMM) [2] to predict the next design step $\boldsymbol{s}_i = [c_i, x_i, y_i]$ in a design sequence $\boldsymbol{S}$ based on the previous design steps $\boldsymbol{s}_p = \{\boldsymbol{s}_0, \cdots, \boldsymbol{s}_{i-1}\}$ and the latent vector $\boldsymbol{z}$. Initially, $\boldsymbol{s}_0$ is set to be the feature vector of the input product image, which is calculated with image hashing using pre-trained VGG-16 network. Here, the RNN is used to restore the sequential information of the design steps from $\boldsymbol{z}$, and the GMM is used to construct the distribution of the design steps based on a set of n normal distributions with different weights for predicting the next design step. This procedure starts with the following decoding process:

$$\boldsymbol{h}_d = decoder(\boldsymbol{s}_p, \boldsymbol{z}) \tag{4}$$

where $\boldsymbol{h}_d$ is the output hidden state of each RNN unit that captures the latent design information stored in the design elements of the previous design sequence. It is further transformed into Y through a fully connected layer to restore the parameters of a Gaussian Mixture Model (GMM). Formally, Y is calculated as follows:

$$Y = W_y \cdot \boldsymbol{h}_d + \boldsymbol{b}_y \tag{5}$$

and it can be further decomposed in the form of

$$Y = [\boldsymbol{c}, (\omega_1, q_1), \cdots, (\omega_k, q_k)]$$

where $\boldsymbol{c}$ is a vector with each field in the vector indicates the probability that a certain cluster of design elements is chosen. (ω_i, q_i) represents the i-th normal distribution where the coordinates of the next design element can be drawn from. In particular, ω_i and $q_i = [\mu_{x_i}, \mu_{y_i}, \sigma_{x_i}, \sigma_{y_i}]$ are the weight and the parameters of the i-th normal distribution respectively. They are the learned parameters that control the distribution of the x and y coordinates of the next design element.

<u>Loss Function.</u> The VAE is trained with a goal of minimizing the following loss:

$$Loss = l_r + w_{kl} \cdot max(l_{kl}, \xi) \tag{6}$$

where l_r is the reconstruction loss that measures the differences between the training samples and the generated design sequences. l_{kl} is the Kullback-Leibler (KL) divergence loss [22] that estimates the divergence between the distribution of the latent vector $\boldsymbol{z}$ and the distribution of a multinomial distribution $N(0, I)$. The KL divergence ensures the estimated distribution of $\boldsymbol{z}$ follows a multinomial distribution that can be estimated by a mean $\boldsymbol{\mu}_z$ and a standard deviation $\boldsymbol{\sigma}_z$. ξ is the lower bound of l_{kl} that ensures the optimizer can turn to minimize l_r when l_{kl} reaches the minimum. w_{kl} is used to balance between l_r and l_{kl}, which is gradually increased during the training process to prioritize the minimization of l_r so as to guarantee the quality of the generated design sequences.

Specifically, l_r and l_{kl} are defined as follows:

$$l_c = -\frac{\sum_{i=1}^{n}\sum_{j=1}^{M} c_{ij} log(c'_{ij})}{n} \tag{7}$$

$$l_p = -\frac{\sum_{i=1}^{n} log(\sum_{j=1}^{k} \omega_j N(x_i, y_i | \mu_{x_j}, \mu_{y_j}, \sigma_{x_j}, \sigma_{y_j}))}{n} \tag{8}$$

$$l_r = l_c + w_p \cdot l_p \tag{9}$$

$$l_{kl} = -\frac{\sum_{i=1}^{K}(1 + log(\boldsymbol{\sigma}_{z_i}) - \boldsymbol{\mu}_{z_i}^2 - \boldsymbol{\sigma}_{z_i})}{K} \tag{10}$$

where n is the length of the design sequence. l_c is the cross entropy that estimates the difference between $\boldsymbol{c}'_i$, the probabilities of selecting design elements from the each cluster in the i-th step of the design sequence, and $\boldsymbol{c}_i$, the one-hot vector representing the cluster of the i-th design element in the training data. l_p estimates the likelihood of the position of each design element in the training sample under a GMM parameterized by $(\omega, \mu_{x_j}, \mu_{y_j}, \sigma_{x_j}, \sigma_{y_j})$.

In our implementation, the LSTM-RNNs in both the decoder and the encoder consist of 512 units. The dimension of the latent vector $\boldsymbol{z}$ (i.e., K) is set to 16, and the total number of the normal distributions in each GMM (i.e., k) is set to 10. In addition, we set $\xi = 0.10$ and both the upper bound of w_p and w_{kl} as 1.00. The model is optimized with Adam optimizer [20], where the gradient clipping value is set to 1.0 to avoid the exploding gradient. The batch size of the input data for each training step is set to 20.

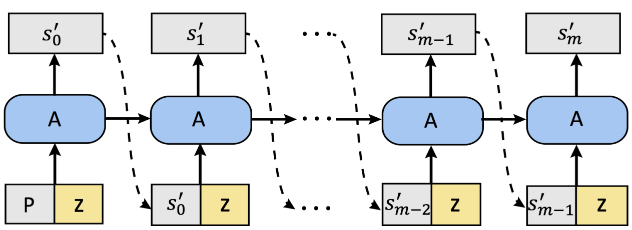

Figure 6: The generation of a design sequence is to use previous design steps to predict the next design step.

Design Sequence Generation. Once trained, the decoder can be used to generate new design sequences in real-time based on a sampled latent vector $\boldsymbol{z}$ and a product image uploaded by the user (as shown in Fig. 6). The generated sequence is initialized with the input product image, and the decoder iteratively predicts the next design steps based on all the previous steps. This process ends when the generated step indicates a cluster of layout templates for texts, given that texts are often arranged at the topmost graphic layer of the poster and the last step of the training design sequences. As a result, each generated sequence can be restored as a series of design choices on the clusters of design elements and their corresponding coordinates on the poster, which will be later used in the *reconstructor* to generate the poster.

5.2 Reconstructor

The *reconstructor* is responsible for transforming a generated design sequence into a final advertising poster following three steps (Fig. 7): (a) selecting specific design elements from each cluster in the design sequence; (b) arranging the design elements using their coordinates derived from the design sequence; and (c) refining the layouts and styles of design elements to finalize the generated poster.

Design Element Selection. To convert a design sequence into a poster, specific design elements need to be selected from the clusters in the design sequence. In order to find the best combination of design elements, the *reconstructor* first enumerates all possible element combinations by selecting one design element in each of the clusters in the design sequence at a time. Then, the *reconstructor* then calculates the similarity of design elements that are combined together by their visual features learned from the VGG-16 network, and removes element combinations with low visual similarity, so as to ensure consistency in design styles.

Layout. Layout is a process of arranging visual elements and taglines at proper positions on the canvas. In particular, for each design sequence, we first roughly arrange user inputs (i.e., the product image and taglines) on the canvas based on the last element of the sequence (i.e., the layout template). The size of the product image and the style of texts are also adjusted accordingly. Then, we add the embellishments to the canvas by iterating all previous design elements in the sequence and place them at the corresponding coordinates that are estimated by the *generator*.

Refinement. After arranging all design elements on the canvas, we further refine the generated poster by making minor adjustments to the position of embellishments and the color of texts. In particular, some embellishments are originally placed in the corner or the edge of the poster, which are partially clipped and are not aesthetically pleasing to be placed in the middle. To address this issue, we labeled each embellishment that is clipped with a position tag, such as "corner" and "edge", when collecting design elements in the preprocessing module, and tweak their positions to fit with the nearest corner or edge accordingly. We also adjust the color of the taglines based on the perceptive luminance value [1] of their background colors. Specifically, this value indicates the brightness of the background, and the text color is chosen oppositely by a proper contrast ratio to ensure the readability of the taglines.

5.3 Estimator

The *generator* of Vinci can quickly create dozens of posters in a few seconds following the process mentioned earlier. Intending to deliver only the best quality posters to the user, we further incorporate Vinci with an *estimator* that is responsible for evaluating

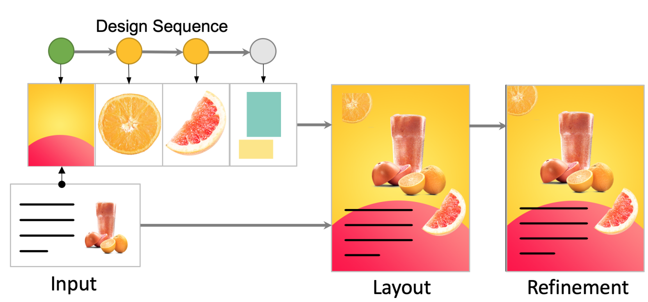

Figure 7: Vinci reconstructs a poster based on the generated design sequence and user input via three major steps: (a) design element selection, (b) layout, and (c) refinement.

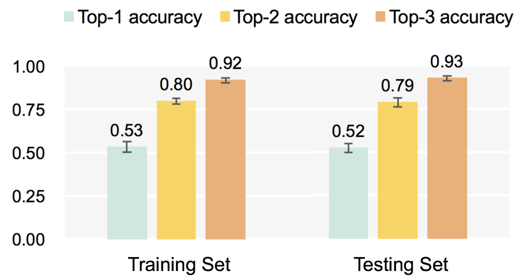

Figure 8: The accuracy of the *estimator* on the training set and the testing set.

and ranking the quality of the generated posters. The *estimator* is designed as a classifier that takes the image of the generated poster as input and classifies the image into one of the five categories that respectively indicates image quality from 5 (the best) to 1 (the worst). We implement the *estimator* with the pre-trained VGG-16 network for extracting the feature of the input image, and two additional fully connected layers for categorizing the posters.

To generate the training samples, we manually labeled 3227 posters that are automatically generated from Vinci . The labeling process was guided by consistent criteria made by a group of 4 professional designers based on common aesthetic standards in designing advertising posters. Concretely, the total score of each poster is 5 points, and we deduct points if the poster contains (1) mismatched elements (-1 point), (2) element overlaps (-1 point), (3) other inaesthetic layouts besides overlaps (-1 point), or (4) irrational style settings (e.g., size of embellishments, color of texts) (-1 point). Note that we deduct one point for each design issue regardless of how many times the issue occurs. For example, both the posters with one-element overlap and two-element overlaps will be deducted with one point. The final score is used to label the poster. We balanced the classes by preserving approximately 645 posters in each category. The *estimator* is trained using 5-fold cross-validation, and the performance of the trained *estimator* on the testing set is shown in Fig. 8. In particular, the top-1,2,3 accuracy are above 52%, 75% and 93% respectively. We also test the consistency between ratings from the *estimator* and human aesthetics via a user study, which is later discussed in Section 6.2. The trained *estimator* is directly incorporated into the pipeline introduced in Section 3 for filtering the model output and delivering the best quality posters to the user.

6 EVALUATION

We evaluated the capability of Vinci in automatically generating aesthetic advertising posters from three aspects. First, we made a user survey to examine the effectiveness of two key components in the poster generation model – the *estimator* and the VAE-based *generator*. Second, we evaluated Vinci 's functionality – poster generation and *online-editing feedback*– through two individual user studies: a Turing test that assessed the quality of the generated posters, and a controlled user study that shows the benefit of the *online-editing feedback* in accelerating poster modification. Last, we conducted an in-depth interview with two professional graphic designers discussing the usability of Vinci . In this section, we first introduce the data preparation process for training the model before conducting the experiments. Then, we introduce the study designs, report study results, and discuss subjective feedback from study participants regarding each of the experiments mentioned above.

6.1 Data Preparation

Prior to the experiments, we assembled a group of three graduate students from design school to generate a dataset of 173 advertising posters about Chinese seafood and soft drinks for training the model. All designers have 5–6 years of graphic design experience. The designers were allowed to either create the posters from scratch or collect the posters that they deem well-designed from online resources. Each designer was asked to create 20 - 30 advertising posters for each of the two product types. We set no constraints on the choice of product images, background, and embellishments. All posters in the training dataset are stored in the PSD format, with each design element separated and labeled.

We established two libraries to store reusable design elements for each product category, respectively. In particular, we extracted 26 backgrounds and 270 embellishments for Chinese seafood, which are further clustered into 3 and 42 element clusters, respectively. For the soft drinks, we extracted 6 clusters of 33 backgrounds and 44 clusters of 220 embellishments. We also clustered the position of product image and taglines in all posters and summarized a representative template for each of the four layout clusters by averaging the positions. Each template is integrated with ten copyrighted text styles that are identified from the posters for selection.

To ensure the performance of the deep generative model and avoid overfitting [35], we expand the sample set by exchanging similar product images under the same category in the posters. This gives us a set of synthetic posters that are similar to the original posters but are different. We asked two designers, listed as coauthors of this paper, to go through each synthetic posters and remove those with significant element overlaps and irrational element matchings. As a result, we obtained 2004 samples for generating the training data. For each sample, we extracted a design sequence following the method introduced in Section 4.2, and utilized all design sequences for training the *generator* for automatic poster generation.

6.2 Experiment 1: Performance of Model Components

We examined the performance of two model components – the *estimator* and the VAE-based *generate* – through a user study with two types of comparison tasks. Given that the data for training the *estimator* was labeled under the guidance from a small group of designers, it is unclear whether the taste of the *estimator* is aesthetically in accord with a broader range of users. Therefore, in the first type of comparison task, we compared posters that were rated low (P_{low}) and posters that were rated high (P_{high}) to see whether human judgments are consistent with the ratings given by the *estimator*. Our hypothesis was that the human designers preferred the design of P_{high} to P_{low} (**H1**). In the second type of comparison tasks, we compared the posters that are respectively generated based on the design sequences produced by the VAE (P_{vae}) and those derived from random design sequences (P_{rand}), so as to have a better understanding on the benefit of using the VAE-based *generator* to produce design sequences. Our hypothesis

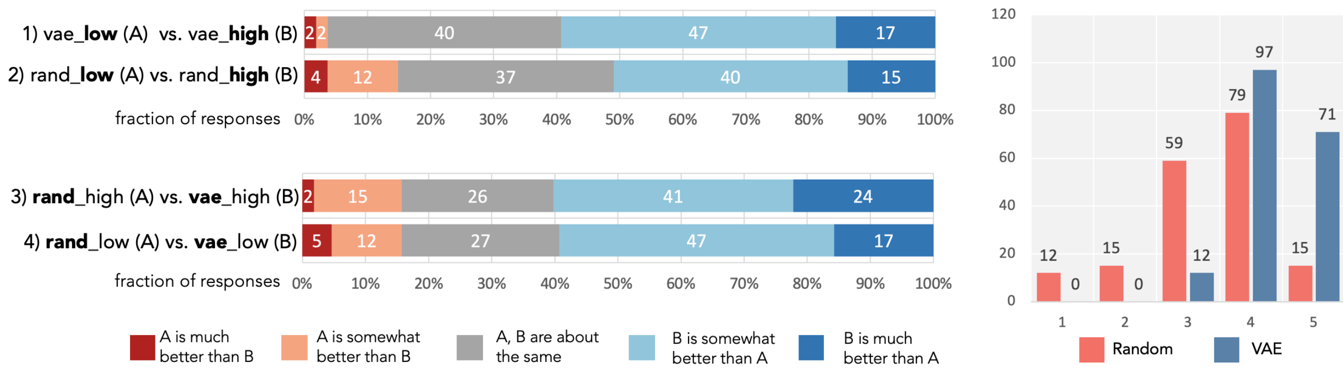

Figure 9: Left: the distribution of responses in experiment 1 under four comparison conditions. Right: the distribution of *estimator*'s ratings on 180 posters generated by random design sequences and VAE-generated design sequences, respectively.

was that the participants preferred the design of P_{vae} to P_{rand} (**H2**).

<u>*Study data.*</u> We collected 6 product images (3 from each product category) to generate the different groups of posters (i.e., P_{low}, P_{high}, P_{vae}, and P_{rand}) for the study tasks. For each product image, we first generate 30 posters using the generation model introduced in Section 5, which gave us a total number of 180 P_{vae}. To test the capability of VAE, we replaced the *generator* in our original model to a random design sequence generator, and further generated 180 P_{rand} (30 for each product image) for comparison. In particular, a random design sequence is generated based on the following steps: (1) determining the length of the sequence m by randomly sampling a number from the length distribution of the training data; (2) randomly select a background cluster in the first step; (3) randomly select an embellishment cluster for each of the steps of $\{s_2, \cdots, s_{m-1}\}$; and (4) randomly select a template from the four layouts of the product and the taglines at the last step of the sequence. The produced random design sequences are then sent to the *reconstructor* and the *estimator* to generate P_{rand}. To generate P_{low} and P_{high}, we analyzed the distribution of ratings for P_{vae} and P_{rand}, respectively. As shown at the right of Fig. 9, the ratings of P_{vae} range from 3 to 5, whereas P_{rand} have lower ratings. Considering that close ratings (i.e., 3 and 4) may have a fuzzy boundary and larger subjective uncertainties, we chose posters that were rated 3 as P_{low} and 5 as P_{high}.

<u>*User tasks.*</u> For each product image, we randomly sampled four posters with different conditions: a poster that was generated from the VAE and was rated low (P_{vae_low}), a poster that was generated using random design sequence and was rated low (P_{rand_low}), a poster that was generated from the VAE and was rated high (P_{vae_high}), and a poster that was generated using random design sequence and was rated high (P_{rand_high}). This gives us four valid comparison tasks with controlled variable for each product image: two for evaluating the *estimator* – (1) P_{vae_low} vs. P_{vae_high} and (2) P_{rand_low} vs. P_{rand_high}, and two for evaluating the VAE-based *generator* – (3) P_{rand_high} vs. P_{vae_high} and (4) P_{rand_low} vs. P_{vae_low}. Therefore, we have a total number of 24 comparison tasks for all product images. In each task, the participants were given two posters (A and B) and were asked to provide an answer using a 5-point Likert scale (A is much better than B, A is somewhat better than B, A and B are about the same, B is somewhat better than A, B is much better than A), reflecting their decision on which poster is better designed. We also asked the participants to explain the reason for their choice in each task briefly.

<u>*Procedure.*</u> We recruited 18 participants (8 males and 10 females, mean age 26.8) in this experiment. All participants are either students from design schools or professional designers in design companies. Each participant was asked to complete 24 comparison tasks. The order of the tasks and the order of the posters for comparison were randomized. As a result, we collected $24 \times 18 = 432$ responses (108 for each of the four comparison conditions) and their explanations.

<u>*Results.*</u> We categorized the responses from the participants by four comparison conditions (as shown in Fig. 9(1–4)), and tested the statistical significance in the participants' preferences using Chi-Square Goodness of Fit test. For the purpose of understanding the Likert scales for participants' preferences, we collapsed categories of "much better than" and "somewhat better than" to signify an explicit preference, and preserved the category of "about the same" to signify neutral.

We first analyzed the first two comparison conditions to investigate the designers' preference on P_{low} and P_{high}. As shown in Fig. 9 (1, 2), an average of 55% of the responses (119/216) exhibited a preference for P_{high} ($\chi^2(2, N = 216) = 68.58, p < .0001$). Only around 9% of the responses (20/216) indicated a preference on P_{low}. These findings supported **H1** that human designers preferred the design of P_{high} to P_{low}, which is in accordance with the ratings given by the *estimator*. Moreover, we found the participants' preferences are more significant when the posters are generated by VAE (59% of the responses, $\chi^2(2, N = 108) = 50.67, p < .0001$) compared to random design sequence (51% of the responses, $\chi^2(2, N = 108) = 21.17, p < .0001$), reflecting a better performance of the *estimator* in distinguishing good and bad the poster designs when the designs are generated by VAE. Around 35% of the responses indicate similar design qualities between P_{low} and P_{high}, which can be resulted from the small gap of low and high ratings (3 and 5).

We then investigated the designers' preference on P_{rand} and P_{vae} by analyzing the last two comparison conditions. As reported in Fig. 9 (3,4), more than 59% of the responses (129/216) showed a tendency towards P_{vae} ($\chi^2(2, N = 216) = 64.19, p < .0001$). Only

(a) P_{rand_low} (b) P_{rand_high} (c) P_{vae_low} (d) P_{vae_high}

Figure 10: An example of posters generated with an orange juice product image in experiment 1.

around 15% of the responses (34/216) exhibited a preference on P_{rand}. These results supported **H2** that the participants preferred the design of P_{vae} to P_{rand}.

Feedback. We analyze subjective feedback from participants to have a better understanding of the reasons for their choices. In the comparison tasks between P_{low} and P_{high}, the most commonly mentioned reasons for P_{high} being better than P_{low} include "no layout overlaps", "more compact layouts", "more prominent product image", "embellishments / colors are more harmonious", "embellishments are more relevant to the product". For example, as shown in Fig. 10, poster (a) and (c) that were rated low by the *estimator* have significant element overlaps comparing to (b) and (d) that were rated high. The feedback is consistent with our labeling guidance introduced in Section 5.3. Similarly, the reasons for P_{vae} being better than P_{rand} includes "no layout overlaps", "layouts are balanced", "the embellishments and backgrounds are more interconnected". For example, poster (a) and (b) in Fig. 10 that were generated by random design sequences contain irrelevant embellishments (e.g. tomato and flamingo). The colors of embellishments and the background image are also arbitrary and failed to match with the product image. In contrast, the semantic of the embellishment(i.e., a slice of orange) in (c) and (d) are highly related to the product image. The colors of embellishment and background also matched well with each other and with the product image. This result is expected because the VAE is designed to predict design elements based on the arrangement of previous design elements, which tend to result in more interrelated design elements in VAE-generated design sequences.

6.3 Experiment 2: Turing Test

The experiment for evaluating the quality of the generated posters consists of two phases: a Human-AI design challenge with 15 participants to collect posters for the experiment, followed by the Turing test with 100 participants.

Procedure. In the first phase, we organized a Human-AI design challenge to encourage creativity by inviting participants to design advertising posters. We recruited three teams of participants (each with 5 participants) through an online procedure with one team containing only professional designers (the *Designer* team). Each team was asked to design 10 posters based on a given product image and taglines. In particular, the *Designer* team was asked to use graphic design tools (e.g., Adobe Photoshop) and online design resources (e.g., stock photo, illustration) for designing posters. Each participant in the *Designer* team was asked to design two advertising posters, one for each product category (i.e., the Chinese seafood and the soft drinks). The other two teams were instructed to use AI designers, *Luban*[2] and *Vinci*, to generate posters respectively. As both systems can generate a batch of posters based on one product image, the participants in each team were asked to decide the best one for their final submission. Also, to limit the influence of confounding factors, the two teams were not allowed to refine the results by using editing functions integrated in the two AI designers. All teams should submit their poster designs within two hours. At the end of the challenge, 30 posters designed by the three teams (*Designer*, *Luban*, *Vinci*) were collected (as shown in Figure 11).

In the second phase, we recruited 100 participants, 50 designers (26 males and 24 females, mean age 23.08), and 50 non-designers (28 males and 22 females, mean age 24.82), to conduct a Turing test using an online questionnaire. Specifically, we randomized the order of the 30 posters and demonstrated one poster to a participant at a time. For each poster, the participant was asked to identify whether it was designed by a human designer or an AI designer, and rate the quality of the poster on a 5-point Likert scale from "very poor" to "very good". The participants were encouraged to leave the reasons for their choices. It took an average of 10 minutes for the participants to complete the questionnaire.

Hypotheses. Before the formal user study, we conducted a pilot study with 30 participants to compare the posters generated by *Luban* and *Vinci*. We also investigated literature regarding graphic design basics [23], layout instructions [32], and color design [34] to obtain insights on the potential differences between *Vinci* and human designers. Based on the results of the pilot study and literature survey, we form hypotheses as follows:

- **H1** The percentage of cases perceived as human (ρ_h) of *Vinci* and that of *Designer* has no significant difference (a). The ρ_h of *Vinci* is significantly higher than that of *Luban* (b).
- **H2** *Vinci* can generate posters as good as *Designer*. Users give significantly higher ratings to *Vinci* than *Luban*.

Results. To analyze the results of the Turing test, we define ρ_h, the percentage of the cases a poster is perceived as designed by a human-designer. A poster with a larger ρ_h value indicates that people think the poster is more likely to be designed by a human-designer. To estimate the design quality, we directly used participants' ratings on each poster (1 ="very poor", 5="very good"). We used repeated measures one-way ANOVA to examine if significant differences exist among different groups, while Bonferroni correction was used for pairwise comparisons. The normality of the data and the homogeneity of variances were respectively tested via the Shapiro-Wilk test and F-test, and the unsatisfied data were transformed using the Normal Inverse Cumulative Distribution Function.

We compare ρ_h for the posters designed by each team (*Designer*, *Luban*, and *Vinci*). In the case where the participants are designers (Fig. 12(a)), the difference in ρ_h is significant ($F_{2,98} = 61.69$, $p < .01$, $\eta^2 = .56$). The *Designer* team had the best performance (i.e., the largest ρ_h on average, $M = .55$, $SD = .03$), which was significantly better than both *Luban* ($M = .19$, $SD = .03$, $p < .01$) and *Vinci* ($M = .41$, $SD = .03$, $p < .01$)(**H1a** rejected). When comparing *Vinci* and *Luban*, the difference in ρ_h was also significant(**H1b** accepted). In the case where the participants are non-designers (Fig. 12(b)), the posters

[2]https://luban.aliyun.com

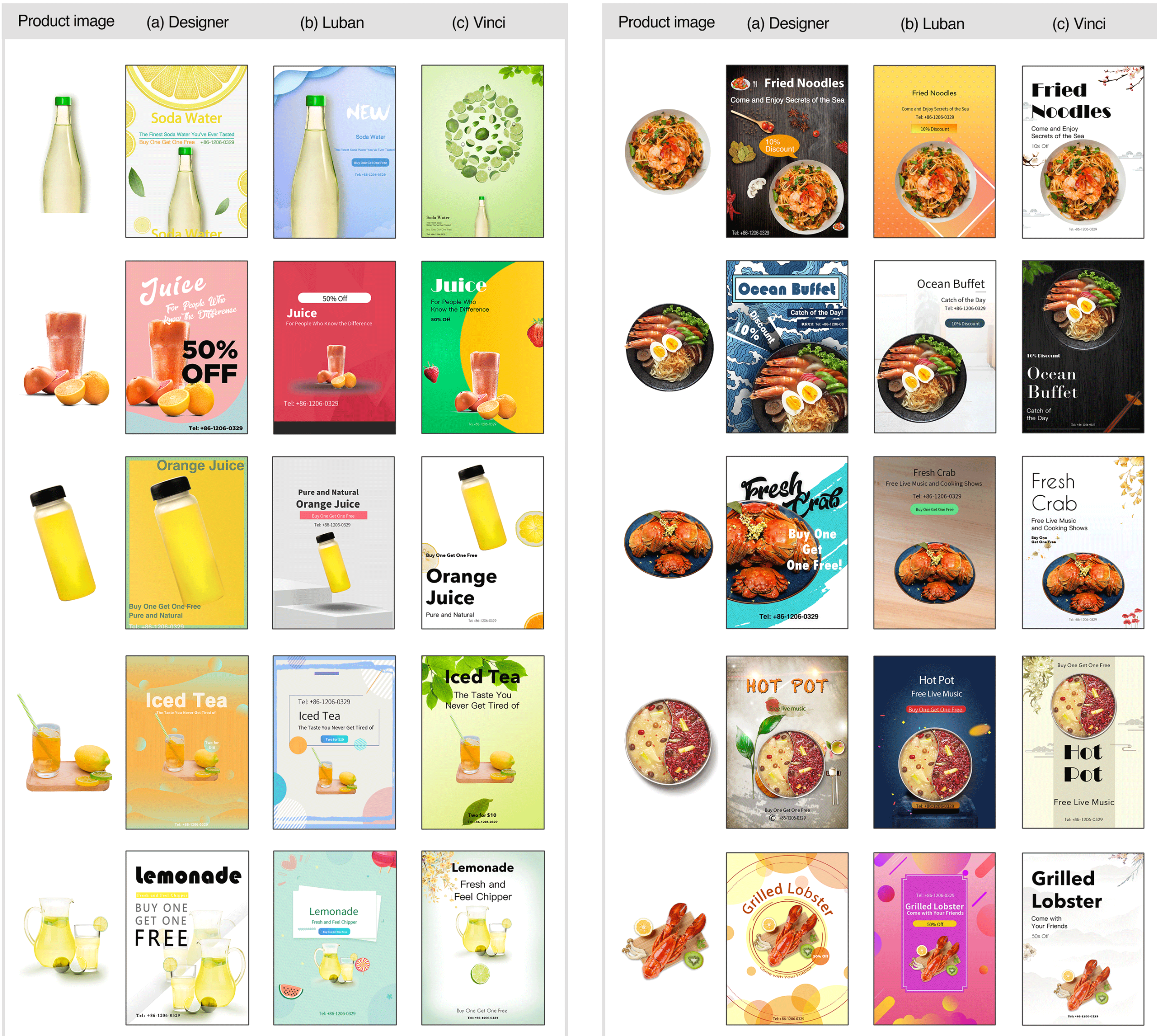

Figure 11: The results of the posters design challenge. The advertising posters shown in this figure are generated respectively by (a) designers, (b) Luban system, and (c) Vinci.

produced by the *Designer* team still had the best performance ($M = .55$, $SD = .03$) and was significantly better than *Luban* ($M = .26$, $SD = .03$; $p < .01$) and *Vinci* ($M = .46$, $SD = .03$, $.01 < p < .05$)(**H1a** rejected). The performance of *Vinci* was still significantly better than *Luban* ($p < .01$)(**H1b** accepted). This results showed that our technique outperformed the state-of-the-art AI designer in Turing test with $\rho_h > 0.4$ on average.

In terms of design quality, there was no significant difference between *Vinci* and *Designer*. At the same time, both *Vinci* and *Designer* can generate posters with significantly better qualities when compared to that of *Luban* in all cases. In particular, when all the participants were designers (Fig. 12(c)), *Vinci* ($M = 3.08$, $SD = .06$) and *Designer* ($M = 3.24$, $SD = .07$) were significantly better than *Luban* ($M = 2.33$, $SD = .09$, $p < .01$)(**H2** accepted). When all the participants were non-designers (Fig. 12(d)), *Vinci* ($M = 3.47$, $SD = .08$) and *Designer*($M = 3.53$, $SD = .09$) were still significantly better than *Luban* ($M = 2.89$, $SD = .11$, $p < .01$)(**H2** accepted).

6.4 Experiment 3: Evaluation on Online Editing-Feedback

We also conducted a controlled within-subject user study to evaluate the performance of Vinci 's online editing-feedback function by comparing user's editing efficiency respectively with and without the support of the *editing-feedback*. In particular, the study was designed according to a task that was frequently performed by designers in their work – changing the size of a poster. As discussed

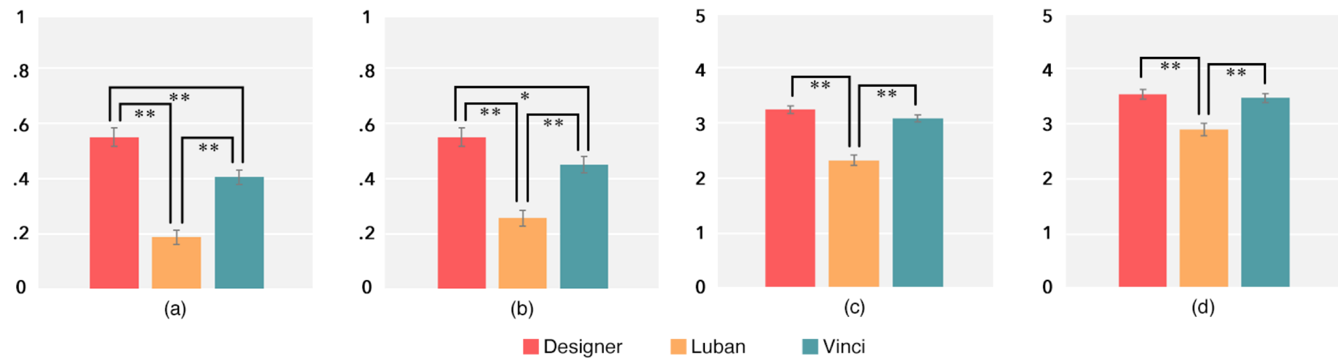

Figure 12: Means and standard errors of each item ($* : .01 < p < .05, ** : p < .01$). (a) Percentage of cases perceived as human by designers, (b) percentage of cases perceived as human by non-designers, (c) ratings on quality by designers using a 5-point Likert scale, (d) ratings on quality by non-designers using a 5-point Likert scale.

in Section 3.1, this task can be time-consuming when performed manually as all design elements need to be rearranged. 20 students (4 males and 16 females, mean age 24) from a design school were invited to participate in our experiment. Each of them was asked to modify a given set of posters in the size of 420 * 570 into the corresponding banners in the size of 420 * 250. The participants were asked to only change the positions and sizes of the design elements, while using the same set of design elements (e.g., background, embellishments, font-face/color) to maintain the design style. Each user was required to perform the task both with and without the support of online editing-feedback. To avoid learning effects, we prepared two different sets of posters. Each set consisted of 12 different posters. To ensure similar difficulty for all tasks, we generate studied posters using Vinci and set the total number of used embellishments in the generated posters strictly to 4.

Results. We recorded the task-completion time and the number of operations for each poster to measure users' performance. We employed the paired t-test with a significance level of .01 to compare the means of completion times and numbers of operations with and without the support of the *editing-feedback*, respectively. As reported in Fig. 13 and Table 1, there was no significant difference in both completion time and the number of operations when users modified the first poster. As the participants continued to perform

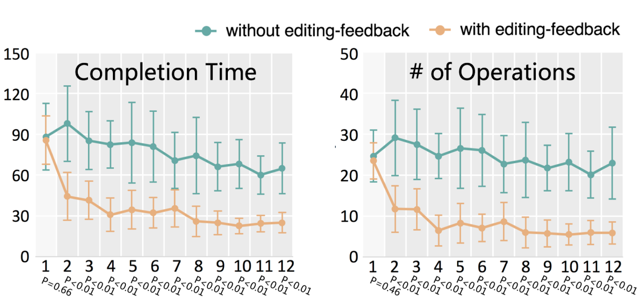

Figure 13: Users' task-completion time (left) and number of operations (right) for modifying 12 posters in a series with/without the support of online editing-feedback. Error bars indicate 99% confidence intervals.

more tasks, these two measurements decreased dramatically with the support of editing-feedback, while the measurements without the support decreased slightly. Significant differences remained in both the completion time and the number of operations from the second poster to the last one. Overall, these results indicated that the support of editing-feedback could significantly improve working efficiency when changing the size of the posters.

6.5 Expert Interview

To further evaluate our system, we conducted a semi-structured interview with two professional designers, **E1** and **E2**. Both of them had over 10 years of design experience. Before the interview, we briefly introduced the design of our technique and demonstrated the use of our system. The two experts were asked to use Vinci to design posters with a given set of product images and were encouraged them to think-aloud during the design process. After trying out the system for 20 minutes, we carried out the interview on main three main topics: the quality of the generated posters, the usefulness of the technique, and the usability of the system. The experts were also asked to provide some comments on the performance of *Luban* compared to Vinci . We reported some of the comments below and discuss others in the Section 7.

The quality of the generated posters. The two experts were impressed by the quality of the generated posters. **E1** mentioned: *"The posters generated by Vinci follow principles of composition and design guidelines in graphic design such as vertical composition and triangular composition. It is surprising that these principles are all presented in Vinci 's designs."* **E2** was impressed by Vinci 's ability to generate matching embellishments: *"The system is definitely very smart! It uses embellishments that match the product image, in both color and semantics. For example, it adds fresh strawberries and ice cubes to the product iced tea, even I did not think about it."* Compared with Vinci , *Luban* seldom adds embellishments to the generated posters. *"The product may be more notable, but the poster can be monotonous if the background image is plain, "* commented by **E1**. **E2** further added: *"Without the embellishments, the poster designs generated by Luban are more universal and not specific to the product. The designs are also more similar across different product categories compared to Vinci ."* Both **E1** and **E2** believed that Vinci has potential to

	Completion time					Number of Operation				
	Without Feedback		With Feedback		Sig	Without Feedback		With Feedback		Sig
ID	Mean	SD	Mean	SD	p-value	Mean	SD	Mean	SD	p-value
1	88.12	41.47	85.62	29.97	0.66	24.6	10.7	23.45	7.41	0.46
2	97.79	47	44.19	29.85	4.80E-05	29.05	15.49	11.7	9.55	6.40E-05
3	85.22	36.09	41.39	23.88	1.80E-05	27.45	14.48	11.6	8.48	6.40E-05
4	82.39	29.41	30.72	20.92	4.00E-06	24.6	9.33	6.45	6.4	5.10E-07
5	83.71	50.44	34.37	24.07	3.70E-04	26.5	16.47	8.25	8.23	9.70E-05
6	80.97	44.19	32.13	18.96	1.60E-06	26	14.85	7.1	5.65	6.60E-07
7	70.69	34.83	35.48	23.13	2.00E-04	22.65	11.75	8.65	7.85	3.20E-05
8	74.28	47.62	25.84	18.8	3.50E-05	23.65	15.49	6	6.32	1.60E-05
9	66.1	30.08	24.82	14.74	9.90E-06	21.65	9.38	5.75	5.49	9.50E-07
10	68.15	30.4	22.46	9.75	7.40E-06	23.1	11.75	5.5	4.32	3.70E-06
11	59.94	23.54	24.21	10.25	2.50E-06	20.05	9.69	5.95	4.9	3.90E-06
12	64.82	31.68	24.78	12.61	8.00E-06	22.9	14.76	5.85	4.56	2.50E-05

Table 1: Users' performance on editing 12 posters in serious respectively with/without the support of online editing-feedback. This table illustrates the underlying data of the diagram shown in Fig. 13.

be applied to real-world scenarios. **E2** noted: *"Vinci works like an experienced designer, who arrives at good design solutions quickly. If I'm tasked with designing a poster with the help of AI, Vinci would be my first choice."* However, **E1** raised questions about how creative and professional an "AI designer" could be. He said that *"although the poster generated by Vinci is aesthetically appealing, its diversity is quite limited. For example, the embellishments are always placed on the corners."* and suggested that Vinci should increase its variety and creativity in poster design in future development.

System. Both **E1** and **E2** believed that Vinci was an effective tool and its user interface satisfied the principles of "fast cognition" and "efficient design". **E2** mentioned that *"I enjoy using Vinci to generate posters. It automatically generates hundreds of results with different styles within a short time and I can browse to select the one I like."* **E1** thought that the function of *online editing-feedback* was useful, *"although AI generates a nice poster for me, I would still like to fine-tune it such as changing the color of the text. It's thoughtful of developers to integrate Vinci with an editing panel."* In contrast, *Luban* does not support making tweaks on the generated posters. **E2** also commented on the *online editing-feedback* and said: "we often need to generate posters with various sizes to accommodate to different scenarios such as advertising banners and leaflets. This function can be especially time-saving in this type of task." Overall, **E1** and **E2** felt that Vinci was a "powerful" and "effective" system, but still had some potential directions of improvement. **E2** suggested that it would be better if the system can take more forms of input, for example, the logo of the company. He also mentioned about allowing users to upload their own design elements. **E1** posed his concern about applying automatic design generation techniques to fulfill the work of designers, and said: *"AI is increasingly lowering the high barrier of becoming a professional graphic designer by actively guiding creative processes. I would perceive it as a challenge for those designers who might lack talents and skills."*

7 DISCUSSION

As suggested in our user experiments, Vinci is incorporated with effective model designs, and is capable of generating advertising posters as good as human designers in a short period of time. The editing-feedback function also allows for modifying posters in batch efficiently. In this section, we discuss the limitations and implications of our work.

7.1 Limitations

In the following, we discuss several limitations of our system that are identified from the experiments and user feedback.

Data preparation. As stated in Section 4.1, the collected advertising posters need to be preprocessed before used for generating the design sequences and training the *generator*. Despite that there are some online resources for advertising posters in the PSD format, it can take manual efforts to label the graphical design elements into the background, embellishments and product image. This is a major obstacle to the immediate generation of posters of a new product category. A potential solution is to train a classifier with the collected and labeled design elements for automatically labeling the new ones, which is feasible because each type of the design elements is discriminable with features of their sizes and positions. For example, the background of the poster is often laid at the bottom and fills the canvas. The embellishments are often small and scattered around the canvas, whereas the product image is usually placed at the center.

System efficiency. While the model can choose and arrange design elements for thousands of posters in only a few seconds, it may take longer to generate the PNG image for users to observe due to the time consumed in rendering process. We tested the efficiency of Vinci on a Linux server with an Intel Xeon CPU (GD6148 2.4GHz/20-cores), 192GB RAM, and a Tesla V100-PCIE-16GB graphics card, in terms of the key steps for producing a poster: generating design sequences from the *generator*, laying out design elements in the *reconstructor*, transforming design information into an image, and *evaluating the quality of the generated poster image.* As reported in Fig 14, the system has an approximately linear time complexity in terms of both the number of the generated posters and the length of the design sequence. The major bottleneck lies in the image rendering process, which can be addressed via GPU acceleration.

Diversity. Some users found our system tended to generate results with similar styles. This is mainly due to the limited number

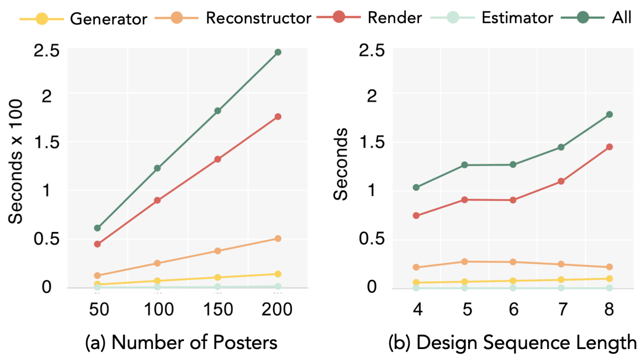

Figure 14: The performance of Vinci in terms of time consumption: (a) the performance with different numbers of posters being generated; (b) the performance of generating a poster with different length of design sequence.

of training posters and design elements we collected, which can greatly affect the diversity of the generated posters. For example, most participants that stayed neutral in *Experiment 1* explained their reasons as: "the design styles are similar", "the layout structures are similar". This issue could be solved as we expand our library of design elements and templates by collecting more posters from more diverse sources.

User inputs. Vinci currently takes user inputs with a fixed form: a product image, the product category, and the taglines for the product, and Vinci searches over a pre-established element library to select and arrange design elements for generating the poster. However, as suggested by a professional designer in our expert interview, designers may need to upload other images or their own elements. This could be a future direction of upgrading our system.

Font-face design. Text is an important part of the poster and can greatly affect the aesthetics, readability, and informativeness of a poster. Therefore, designers usually spend serious time designing fonts when creating a poster. However, our system currently doesn't support font-face design, which could be a future direction.

Copyright issue. Vinci system generates posters by selecting and packing the collected design elements together. This may lead to copyright issues of the collected design elements. One potential solution is to incorporate a sub-generative model (e.g., GAN [11]) in Vinci for generating original and reusable design elements.

7.2 Human Designer vs. AI Designer

Our evaluation showed that human designer outperformed Vinci in the Turing test while Vinci achieves comparable performance with a human designer in design quality. As suggested by the study participants, their criteria for whether a poster is designed by humans included creativity, uniqueness of concept, level of detail, visual appeal, and harmoniousness. For example, one participant commented that the poster designed by Vinci was *"too simple"* or *"too universal"* and could be equally applicable to other products. According to our expert interview, this issue also exists for *Luban*, and is more obvious under the absence of embellishments. In contrast, human designers usually fine-tune details (e.g., aligning text along a curved path), which made the results more elegant. In terms of design quality, participants paid more attention to criteria such as the use of a design element to enhance a theme, artistic merit, visual appeal, and the composition of the poster. Given that Vinci was designed under the consideration of these aesthetic criteria, it is not difficult to understand why Vinci was rated almost as good as a human designer. In particular, the participants emphasized Vinci 's ability to select backgrounds and the embellishments that fit well with the product image.

In addition, the participants were also impressed by Vinci 's power on generating a large number of posters within a short period, which "can never be achieved by a human designer". They believed this feature could be especially useful in the case where the quantity of the posters delivered is more important than the quality—for example, helping online stores to design and generate advertising posters before holiday seasons.

7.3 Generalizability

As we have seen in both the quantitative analysis and qualitative feedback, Vinci was favored by a majority of users in their design practice. They also expressed a preference for applying Vinci to other graphic design formats such as infographics, slides, and leaflets. In fact, our approach can be generalized to other design scenarios where design processes can be abstracted into sequences of inner-correlated design decisions. For example, infographics can be perceived as a sequence of overlaid graphic layers in a design space of background, visualization charts, icons, and annotations. Once the designs are transformed into design sequences, they can be utilized to train the *generator* and create new designs in a way similar to generating advertising posters. The only problem, however, is to identify the design space for a particular type of information representation. Currently, the design space of advertising posters is determined using a heuristic method, which may not be applicable to other types of information presentation. To better exploit design knowledge codified within existing graphic design, a potential solution is to incorporate machine learning techniques (e.g., [9]) to help derive design spaces from data automatically.

7.4 Application of Vinci in Graphic Design

We observed that our Vinci users are mostly positive toward the automatic generated posters. They suggest that Vinci's capability of generating matching embellishments that decorate the product provides a good starting point for poster design. It is feasible to adjust these elements and add new ones on canvas for refinement. Some participants expressed a desire for more intelligence in Vinci . For example, participants suggested that Vinci can learn their design styles and transfer them to newly generated posters. On the other hand, participants expressed their concerns with the application of Vinci in graphic design. First, although being perceived as aesthetic, the design of advertising posters generated by Vinci lack sufficient diversity in design style. For example, participants noted that the current personality of our poster design is limited to modern and suggested that more personalities should be involved, such as cute, futuristic, and vintage [44]. Second, the lack of controllability, comprehensibility, and predictability in the creative process might provide a negative user experience, which is also pointed out as limitations of intelligent user interfaces in prior research work [13, 17]. Third, most participants agreed that Vinci is creative and could generate some unexpected design element combinations, which poses a challenge for those designers who lack talents.

8 CONCLUSION AND FUTURE WORK

In this paper, we present an intelligent design system, Vinci, for automatic advertising poster design. To the best of our knowledge, it is one of the first research work towards applying deep learning-based techniques to generate poster designs. Given a product image and text specified by the user, Vinci is able to select design elements to match with the product image and arrange their positions on canvas to generate a poster with high quality. The user studies showed that the posters generated by our system were rated higher regarding quality compared with *Luban*, and achieved comparable performance compared with posters created by human-designers. The editing function of Vinci were also demonstrated to be effective in improving the efficiency of poster modification. Future work includes overcoming the current limitations on diversity and font-face design, and extending the system to generate more complicated graphic designs such as infographics.

ACKNOWLEDGMENTS

Nan Cao is the corresponding author. We would like to thank all the study participants and the anonymous reviewers for their valuable comments. This work was supported in part by NSFC (62072338) and NSF Shanghai (20ZR1461500).

REFERENCES

[1] Fernando Alonso, José Luis Fuertes, Ángel Lucas González, and Loïc Martínez. 2010. On the Testability of WCAG 2.0 for Beginners. In *Proceedings of the 2010 International Cross Disciplinary Conference on Web Accessibility (W4A)* (Raleigh, North Carolina) *(W4A '10)*. Association for Computing Machinery, New York, NY, USA, Article 9, 9 pages. https://doi.org/10.1145/1805986.1806000

[2] Christopher M Bishop. 1994. *Mixture density networks*. Technical Report. Citeseer.

[3] Samuel R. Bowman, Luke Vilnis, Oriol Vinyals, Andrew Dai, Rafal Jozefowicz, and Samy Bengio. 2016. Generating Sentences from a Continuous Space. In *Proceedings of The 20th SIGNLL Conference on Computational Natural Language Learning*. Association for Computational Linguistics, Berlin, Germany, 10–21. https://doi.org/10.18653/v1/K16-1002

[4] Nan Cao, Xin Yan, Yang Shi, and Chaoran Chen. 2019. AI-Sketcher : A Deep Generative Model for Producing High-Quality Sketches. *Proceedings of the AAAI Conference on Artificial Intelligence* 33, 01 (Jul. 2019), 2564–2571. https://doi.org/10.1609/aaai.v33i01.33012564

[5] Jie Chang, Yujun Gu, and Ya Zhang. 2017. Chinese Typeface Transformation with Hierarchical Adversarial Network. *CoRR* abs/1711.06448 (2017), 9. arXiv:1711.06448 http://arxiv.org/abs/1711.06448

[6] Gang Chen, Peihong Xie, Jing Dong, and Tianfu Wang. 2019. Understanding programmatic creative: The role of AI. *Journal of Advertising* 48, 4 (2019), 347–355.

[7] Peggy Chi, Zheng Sun, Katrina Panovich, and Irfan Essa. 2020. Automatic Video Creation From a Web Page. In *Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology* (Virtual Event, USA) *(UIST '20)*. Association for Computing Machinery, New York, NY, USA, 279–292. https://doi.org/10.1145/3379337.3415814

[8] J. Deng, W. Dong, R. Socher, L. Li, Kai Li, and Li Fei-Fei. 2009. ImageNet: A large-scale hierarchical image database. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, Miami, FL, USA, 248–255. https://doi.org/10.1109/CVPR.2009.5206848

[9] V. Dibia and Ç. Demiralp. 2019. Data2Vis: Automatic Generation of Data Visualizations Using Sequence-to-Sequence Recurrent Neural Networks. *IEEE Computer Graphics and Applications* 39, 5 (2019), 33–46. https://doi.org/10.1109/MCG.2019.2924636

[10] Leon A Gatys, Alexander S Ecker, and Matthias Bethge. 2016. Image style transfer using convolutional neural networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, Las Vegas, NV, USA, 2414–2423.

[11] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. 2014. Generative adversarial nets. In *Advances in neural information processing systems*. Curran Associates, Inc., Palais des Congrès de Montréal, Montréal CANADA, 2672–2680.

[12] David Ha and Douglas Eck. 2017. A Neural Representation of Sketch Drawings. *CoRR* abs/1704.03477 (2017), 15. arXiv:1704.03477 http://arxiv.org/abs/1704.03477

[13] Melanie Hartmann. 2009. Challenges in Developing User-Adaptive Intelligent User Interfaces. In *LWA 2009: Workshop-Woche: Lernen, Wissen, Adaptivität, Darmstadt, 21.-23. September 2009*, Melanie Hartmann and Frederik Janssen (Eds.), Vol. TUD-CS-2009-0157/TUD-KE-2009-04. FG Telekooperation/FG Knowledge Engineering, Technische Universität Darmstadt, Germany, Darmstadt, Germany, ABIS:6–10.

[14] Simon Haykin. 1994. *Neural Networks: A Comprehensive Foundation* (1st ed.). Prentice Hall PTR, USA.

[15] Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long short-term memory. *Neural Computation* 9, 8 (1997), 1735–1780.

[16] Ali Jahanian. 2016. *Automatic Design of Self-Published Media: A Case Study of Magazine Covers*. Springer International Publishing, Cham, 69–99. https://doi.org/10.1007/978-3-319-31486-0_5

[17] Anthony David Jameson. 2009. Understanding and dealing with usability side effects of intelligent processing. *AI Magazine* 30, 4 (2009), 23–23.

[18] Yanghua Jin, Jiakai Zhang, Minjun Li, Yingtao Tian, Huachun Zhu, and Zhihao Fang. 2018. MakeGirlsMoe - Create Anime Characters with A.I.! https://make.girls.moe/.

[19] David Johnson. 2005. Psychology of Colors.

[20] Diederik P. Kingma and Jimmy Ba. 2017. Adam: A Method for Stochastic Optimization. arXiv:1412.6980 [cs.LG]

[21] Diederik P Kingma and Max Welling. 2014. Auto-Encoding Variational Bayes. arXiv:1312.6114 [stat.ML]

[22] Solomon Kullback and Richard A Leibler. 1951. On information and sufficiency. *The Annals of Mathematical Statistics* 22, 1 (1951), 79–86.

[23] Robin Landa. 2010. Graphic design solutions.

[24] Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. 1998. Gradient-based learning applied to document recognition. *Proc. IEEE* 86, 11 (1998), 2278–2324.

[25] Jianan Li, Jimei Yang, Aaron Hertzmann, Jianming Zhang, and Tingfa Xu. 2019. LayoutGAN: Generating Graphic Layouts with Wireframe Discriminators. arXiv:1901.06767 [cs.CV]

[26] Simon Lok and Steven Feiner. 2001. A survey of automated layout techniques for information presentations. *Proceedings of SmartGraphics* 2001 (2001), 61–68.

[27] Jock Mackinlay. 1988. Applying a theory of graphical presentation to the graphic design of user interfaces. In *Proceedings of the ACM SIGGRAPH symposium on User Interface Software*. ACM, Alberta, Canada, 179–189.

[28] Peter O'Donovan, Aseem Agarwala, and Aaron Hertzmann. 2015. Designscape: Design with interactive layout suggestions. In *Proceedings of the 33rd annual ACM conference on human factors in computing systems*. ACM, Seoul, Republic of Korea, 1221–1224.

[29] Yu-ting Qiang, Yan-Wei Fu, Xiao Yu, Yan-Wen Guo, Zhi-Hua Zhou, and Leonid Sigal. 2019. Learning to Generate Posters of Scientific Papers by Probabilistic Graphical Models. *Journal of Computer Science and Technology* 34, 1 (2019), 155–169.

[30] Steven F. Roth, John Kolojejchick, Joe Mattis, and Jade Goldstein. 1994. Interactive graphic design using automatic presentation knowledge. In *Conference on Human Factors in Computing Systems, CHI 1994, Boston, Massachusetts, USA, April 24-28, 1994, Proceedings*, Beth Adelson, Susan T. Dumais, and Judith S. Olson (Eds.). ACM, Massachusetts, USA, 112–117. https://doi.org/10.1145/191666.191719

[31] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, et al. 2015. Imagenet large scale visual recognition challenge. *International Journal of Computer Vision* 115, 3 (2015), 211–252.

[32] Timothy Samara. 2017. Making and Breaking the Grid, Updated and Expanded: A Graphic Design Layout Workshop.

[33] Yang Shi, Nan Cao, Xiaojuan Ma, Siji Chen, and Pei Liu. 2020. EmoG: Supporting the Sketching of Emotional Expressions for Storyboarding. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. ACM, Oahu, Hawaii, USA, 1–12.

[34] Terry Lee Stone, Sean Adams, and Noreen Morioka. 2008. Color design workbook: A real world guide to using color in graphic design.

[35] Chen Sun, Abhinav Shrivastava, Saurabh Singh, and Abhinav Gupta. 2017. Revisiting unreasonable effectiveness of data in deep learning era. In *Proceedings of the IEEE international conference on computer vision*. IEEE, Venice, Italy, 843–852.

[36] Sou Tabata, Haruka Maeda, Keigo Hirokawa, and Kei Yokoyama. 2019. Diverse Layout Generation for Graphical Design Magazines. In *SIGGRAPH Asia 2019 Posters*. ACM, BCEC, Brisbane, Australia, 1–2.

[37] Sou Tabata, Hiroki Yoshihara, Haruka Maeda, and Kei Yokoyama. 2019. Automatic layout generation for graphical design magazines. In *ACM SIGGRAPH 2019 Posters*. ACM, Los Angeles, USA, 1–2.

[38] Kashyap Todi, Daryl Weir, and Antti Oulasvirta. 2016. Sketchplore: Sketch and explore with a layout optimiser. In *Proceedings of the ACM Conference on Designing Interactive Systems*. ACM, Brisbane QLD Australia, 543–555.

[39] Sreekanth Vempati, Korah T Malayil, Sruthi V, and Sandeep R. 2019. Enabling Hyper-Personalisation: Automated Ad Creative Generation and Ranking for

Fashion e-Commerce. arXiv:1908.10139 [cs.IR]
[40] Xin Wang, Geoffrey Oxholm, Da Zhang, and Yuan-Fang Wang. 2017. Multimodal transfer: A hierarchical deep convolutional neural network for fast artistic style transfer. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, Honolulu, HI, USA, 5239–5247.
[41] Xuyong Yang, Tao Mei, Ying-Qing Xu, Yong Rui, and Shipeng Li. 2016. Automatic generation of visual-textual presentation layout. *ACM Transactions on Multimedia Computing, Communications, and Applications* 12, 2 (2016), 33.
[42] Wenyuan Yin, Tao Mei, and Chang Wen Chen. 2013. Automatic generation of social media snippets for mobile browsing. In *Proceedings of the ACM International Conference on Multimedia*. ACM, Barcelona, Spain, 927–936.
[43] Lvmin Zhang, Chengze Li, Tien-Tsin Wong, Yi Ji, and Chunping Liu. 2018. Two-stage sketch colorization. In *SIGGRAPH Asia Technical Papers*. ACM, ACM, Tokyo, Japan, 261.
[44] Nanxuan Zhao, Ying Cao, and Rynson WH Lau. 2018. What characterizes personalities of graphic designs? *ACM Transactions on Graphics* 37, 4 (2018), 116.
[45] Nanxuan Zhao, Nam Wook Kim, Laura Mariah Herman, Hanspeter Pfister, Rynson WH Lau, Jose Echevarria, and Zoya Bylinskii. 2020. ICONATE: Automatic Compound Icon Generation and Ideation. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. ACM, Oahu, Hawaii, USA, 1–13.
[46] Shuyu Zheng, Ziniu Hu, and Yun Ma. 2019. FaceOff: Assisting the Manifestation Design of Web Graphical User Interface. In *Proceedings of the ACM International Conference on Web Search and Data Mining*. ACM, Melbourne, Australia, 774–777.
[47] Xinru Zheng, Xiaotian Qiao, Ying Cao, and Rynson WH Lau. 2019. Content-aware generative modeling of graphic design layouts. *ACM Transactions on Graphics (TOG)* 38, 4 (2019), 1–15.
[48] Jun-Yan Zhu, Philipp Krähenbühl, Eli Shechtman, and Alexei A Efros. 2016. Generative visual manipulation on the natural image manifold. In *Proceedings of the European Conference on Computer Vision*. Springer, Springer, Amsterdam, The Netherlands, 597–613.
[49] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A. Efros. 2017. Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks. In *IEEE International Conference on Computer Vision*. IEEE, Venice, Italy, 2242–2251.