A Trend-Driven Fashion Design System for Rapid Response Marketing in E-commerce

Lianghua Huang, Yu Liu, Bin Wang, Pan Pan and Rong Jin
Machine Intelligence Technology Lab, Alibaba Group
{ly103369, xuangen.hlh, ganfu.wb, panpan.pp, jinrong.jr}@alibaba-inc.com

Abstract

Fashion is the way we express ourselves and has grown into one of the largest industries in the world. Despite the significant evolvement of the fashion industry over the past decades, it is still a great challenge to respond to the diverse preferences of a large number of different consumers in time and accurately. To deal with the problem, we present an innovative demonstration of a trend-driven fashion design system using deep generative modeling, which enables automatic fashion design and editing based on trend reports. Our system consists of three components, including trend-driven fashion design, interactive fashion editing, and popularity estimation. The system offers a unified framework for mass-production of fashion designs that conform to the trend, which helps businesses better respond to market demands.

Introduction

Fashion is a form of self-expression and autonomy in a particular period and context (Kaiser and Green 2021), and it implies a look defined by the fashion industry as *trending*[1]. The global fashion market accounts for up to 2% of the world's Gross Domestic Product (GDP)[2], and the revenue is expected to grow at an annual rate of 7.31%[3]. Although the fashion industry has a huge market capacity and has evolved for decades, how to mass-produce trendy and novel designs quickly and accurately is still a major difficulty[4].

The revolution of artificial intelligence (AI) brings potential for automatic fashion analysis and design, and a wide range of innovative applications in this field come into being (Cheng et al. 2020). These applications ranging from fashion detection (Ge et al. 2019), fashion analysis (Caba Heilbron et al. 2019; Ma et al. 2019), fashion design (Pandey and Savakis 2020), to fashion recommendation (Dong et al. 2019). However, few work focuses on automatic trendy design in line with the market demand, where the *trend* usually represents a vague composition of styles, attributes, colourways, popular elements, and/or prices.

[1]https://en.wikipedia.org/wiki/Fashion

[2]https://fashionunited.com/global-fashion-industry-statistics/

[3]https://www.statista.com/outlook/dmo/ecommerce/fashion/worldwide

[4]https://en.wikipedia.org/wiki/Fast_fashion

Figure 1: An overview of our trend-driven fashion design system. (a) Trend-driven fashion design. (b) Interactive fashion editing. (c) Popularity estimation.

In this demo, we present a trend-driven fashion design system built upon deep generative modeling, which takes trend reports as input and is able to mass-produce trendy designs. Specifically, the system implements three essential functions: trend-driven fashion design, interactive fashion editing as well as popularity estimation. Unlike previous works, we mainly focus on rapid production of fashion designs that conform to the popularity trend, helping manufacturers respond to market demands quickly and accurately.

System Framework

The proposed demo system takes trend descriptions and/or user editing operations as input, and outputs the generated design images along with popularity estimation. Figure 1 shows the overall framework. The trend-driven fashion design module generates designs based on trend conditions; the interactive fashion editing module allows the user to smoothly modify the designs; while the popularity estimation module gives a prediction of the designs' popularity. We introduce the components respectively in the following.

Trend-Driven Fashion Design

Trend-driven fashion design system aims to generate novel design images based on structured trend descriptions[5]. As shown in Figure 1 (a), each trend description can be parsed and represented as a list of *key-value* tokens, such as *style-loafer, material-rubber, pattern-lattice, etc.* In learning phase, we build a vocabulary covering all possible tokens in the training set, including categories, properties, prices and words extracted from commodity titles and descriptions. Then we train a conditional StyleGAN2 (Karras et al. 2020) to learn the correspondence between these tokens and generated images, where the *condition* is a vector mapped through a multi-layer MLP from a multi-hot vector whose dimension is the vocabulary size. During test, we also parse each query trend description into *key-value* tokens and encode them to a condition vector. The condition vector and several random style vectors are then fed into the pretrained StyleGAN2 to obtain generated images. We choose conditional StyleGAN2 instead of other conditional generators (*e.g.,* DALL-E (Ramesh et al. 2021)) because StyleGAN2 runs much faster, produces higher fidelity images and has better controllability over generated images.

Interactive Fashion Editing

As shown in Figure 1 (b), the system implements three types of editing operations to control the image generation process: attribute editing, style transfer and sketch editing. *Attribute editing* allows a user to smoothly change the semantic properties of a design image. We train an attribute classifier in the latent space and learn disentangled walking directions to achieve this (Zhuang, Koyejo, and Schwing 2021). *Style transfer* enables sellers to slightly modify trendy designs to conform to their own brand styles. This capability is achieved by linear interpolation between images in the $W+$ space of StyleGAN2 (Karras et al. 2020). *Sketch editing* further allows the users to control the shape or details of a design. We achieve this capability by using image-to-image translation (Richardson et al. 2021). These editing tools provide different levels of controllability for fashion design.

Popularity Estimation

Several approaches have been proposed to estimate the popularity of commodities (Wu et al. 2019). However, most methods require inflexible retraining or finetuning of the model when trends are updated, while many others are based on multimodal or social cues, which are unavailable for AI-designed images. In contrast to them, as illustrated in Figure 1 (c), we adopt a simple but effective strategy: for each generated image, we search for the top-K nearest neighbors in the commodity database using an image search engine (*i.e.,* a pretrained feature extractor); then we compute a soft voting over these samples to obtain the estimated price, click-through rate, selling potential and trading volume. We find the simple strategy to work pretty well in practice.

[5]Note we do NOT predict trends in this work, but rather assume that trends are already available and we take them as input.

Metric	Autoregressive	cStyleGAN2
Resolution	256	**1024**
Speed	0.06 fps	**5 fps**
Accuracy	**0.82**	0.79
Interp.		✓
Attr. editing		✓

Table 1: Performance comparison between our trend-conditional StyleGAN2 (*abbr.* cStyleGAN2) and an autoregressive generation model (*i.e.,* GPT with VQ-VAE2).

Experiments

We investigate the contributions of our work and implement multiple architectures for trend-driven fashion design. The experiments are conducted on a *shoes* dataset containing 100K commodity images, each with descriptions consisting of different levels of categories, title, price and properties. We evaluate two models: the conditional StyleGAN2 and an autoregressive model, *i.e.,* GPT (Ramesh et al. 2021) with VQ-VAE2 (Razavi, van den Oord, and Vinyals 2019) on the dataset. 24 trend descriptions are taken as the condition. On user-annotated images that conform to these trend descriptions, we pretrain a trend classifier and use the classification accuracy as the metric for conditionally generated images. Table 1 compares the results. Our conditional StyleGAN2 achieves comparable trend accuracy with the autoregressive model, but it runs two magnitudes faster, has better controllability over generated images, and it is able to generate much higher resolution (1024×1024) designs.

Conclusion

In this work, we present a trend-driven fashion design system using generative modeling. The system consists of three different components for novel fashion design generation, editing and popularity estimation, respectively. We show that our system achieves good performance in terms of generation speed, resolution, controllability and trend accuracy. We believe that the system can help fashion businesses improve their ability in rapid response marketing.

References

Caba Heilbron, F.; Pepik, B.; Barzelay, Z.; and Donoser, M. 2019. Clothing Recognition in the Wild using the Amazon Catalog. In *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops*, 0–0.

Cheng, W.-H.; Song, S.; Chen, C.-Y.; Hidayati, S. C.; and Liu, J. 2020. Fashion meets computer vision: A survey. *arXiv preprint arXiv:2003.13988*.

Dong, X.; Song, X.; Feng, F.; Jing, P.; Xu, X.-S.; and Nie, L. 2019. Personalized capsule wardrobe creation with garment and user modeling. In *Proceedings of the 27th ACM International Conference on Multimedia*, 302–310.

Ge, Y.; Zhang, R.; Wang, X.; Tang, X.; and Luo, P. 2019. Deepfashion2: A versatile benchmark for detection, pose estimation, segmentation and re-identification of clothing images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 5337–5345.

Kaiser, S. B.; and Green, D. N. 2021. *Fashion and cultural studies*. Bloomsbury Publishing.

Karras, T.; Laine, S.; Aittala, M.; Hellsten, J.; Lehtinen, J.; and Aila, T. 2020. Analyzing and improving the image quality of stylegan. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 8110–8119.

Ma, Y.; Yang, X.; Liao, L.; Cao, Y.; and Chua, T.-S. 2019. Who, where, and what to wear? Extracting fashion knowledge from social media. In *Proceedings of the 27th ACM International Conference on Multimedia*, 257–265.

Pandey, N.; and Savakis, A. 2020. Poly-GAN: Multi-conditioned GAN for fashion synthesis. *Neurocomputing*, 414: 356–364.

Ramesh, A.; Pavlov, M.; Goh, G.; Gray, S.; Voss, C.; Radford, A.; Chen, M.; and Sutskever, I. 2021. Zero-shot text-to-image generation. *arXiv preprint arXiv:2102.12092*.

Razavi, A.; van den Oord, A.; and Vinyals, O. 2019. Generating diverse high-fidelity images with vq-vae-2. In *Advances in neural information processing systems*, 14866–14876.

Richardson, E.; Alaluf, Y.; Patashnik, O.; Nitzan, Y.; Azar, Y.; Shapiro, S.; and Cohen-Or, D. 2021. Encoding in style: a stylegan encoder for image-to-image translation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2287–2296.

Wu, B.; Cheng, W.-H.; Liu, P.; Liu, B.; Zeng, Z.; and Luo, J. 2019. Smp challenge: An overview of social media prediction challenge 2019. In *Proceedings of the 27th ACM International Conference on Multimedia*, 2667–2671.

Zhuang, P.; Koyejo, O. O.; and Schwing, A. 2021. Enjoy Your Editing: Controllable GANs for Image Editing via Latent Space Navigation. In *International Conference on Learning Representations*.