

MYCloth: Towards Intelligent and Interactive Online T-Shirt Customization based on User's Preference

Yexin Liu

AI Thrust, HKUST(GZ)

Email: yliu292@connect.hkust-gz.edu.cn

Lin Wang[†]

AI/CMA Thrust, HKUST(GZ) and

Dept. of CSE, HKUST

Email: linwang@ust.hk

Abstract—In conventional online T-shirt customization, consumers, *i.e.*, users, can achieve the intended design only after repeated adjustments of the design prototypes presented by sellers in online dialogues. However, this process is prone to limited visual feedback and cumbersome communication, thus detracting from users' customization experience and time. This paper presents an intelligent and interactive online customization system, named MYCloth, aiming to enhance the T-shirt customization experience. Given the user's text input, our MYCloth employs ChatGPT to refine the text prompt and generate the intended paint of the cloth via the Stable Diffusion model. Our MYCloth also enables the user to preview the final outcome via a novel learning-based virtual try-on model. The whole system allows to iteratively adjust the cloth till optimal design is achieved. We verify the system's efficacy through a series of performance evaluations and user studies, highlighting its ability to streamline the online customization process and improve overall satisfaction.

I. INTRODUCTION

Traditionally, obtaining T-shirts involves physical store visits or more recently, online shopping. Consumers, *i.e.*, users, seek out pre-made clothes that match their personal style. The clothing industry responds by providing a wide range of fashion items for various tastes and occasions [1]–[3]. However, these options are limited by mass production and standard designs. A growing trend focuses on personalized T-shirts, where individuals can express their unique style, turning their attire into a form of self-expression. The essence of personalized T-shirts lies in their capacity to offer consumers an active role in the design and conceptualization of their clothing. It is a journey where consumers transit from passive wearers to active participants in its creation. Through a blend of innovative technology and collaborative design processes, individuals are empowered to make choices concerning fabric, color, patterns, and even the minutiae of stitching details.

Recently, online customization has gained great attention as it allows customers to design and order clothes from the comfort of their homes or anywhere with internet access [4]–[7]. However, it also comes with a set of challenges: **1) Communication Complexity.** Traditional systems for customized clothing often follow a linear path, with users initially conveying their specific design requirements to sellers. The burden of interpreting these requirements and crafting design prototypes rests primarily with the sellers. This approach, while common,

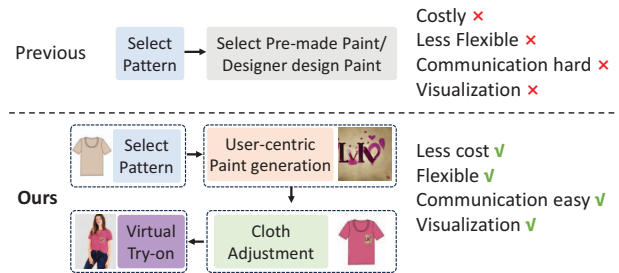


Fig. 1: Comparison between previous T-shirt customization and our proposed AI-assisted personalized customization.

can lead to communication complexities, misunderstandings, and protracted interactions. **2) Limited Visual Feedback.** In many personalized clothing processes, there's a lack of effective tools for users to gain an immediate and intuitive visual understanding of how their chosen designs look on them. This absence of direct visual feedback can lead to uncertainty and dissatisfaction among users.

To address these challenges, some systems have been developed for interactive T-shirt design [8], [9], enabling users to rearrange the clothing pattern elements, choose the available colors, and fabric. However, these systems 1) only allow users to select existing elements in the system during design, such as clothing texture, layout, *etc.*, and cannot generate some elements according to the user's own preferences, such as printing; 2) only allow users to view the garment itself without previewing how it looks when worn by a person.

To this end, this paper proposes an intelligent and interactive online customization system, named **MYCloth**, aiming to enhance the T-shirt customization experience. Through MYCloth, it is possible to empower users to seamlessly select T-shirt styles, customize colors, and explore diverse design aesthetics, while offering interactive tools for comprehensive T-shirt customization. Technically, given the user's text input about the design theme, our MYCloth system first refines the text prompt by employing large language models (LLMs), such as ChatGPT [10], to re-prompt the text to make it more consistent with the design theme. We then leverage the artificial intelligence (AI)-driven text-to-image generation model, *e.g.*, Stable Diffusion model [11], to generate paint

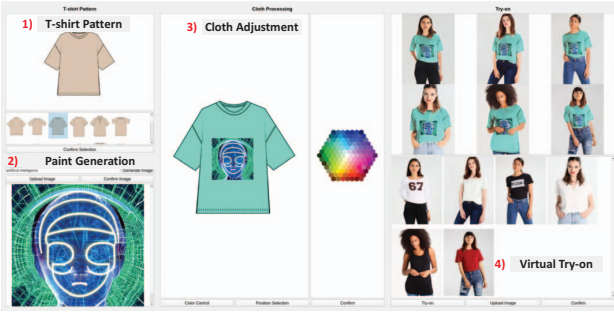


Fig. 2: Overview of the proposed system: 1) **Pattern selection** allowing users to choose from a variety of T-shirt patterns, 2) **Paint Generation** intelligently generating design concepts aligned with user-defined themes, 3) **Cloth Adjustment** empowering users to meticulously refine design elements like color and the position of the paints, bridging the conceptualization and tangible design, and 4) **Virtual Try-On** enabling users to visualize their creations on 2D avatars for design validation.

images that can be used in T-shirt design. Moreover, we propose a novel AI-based virtual try-on model that allows users to preview their customized T-shirts comprehensively. This enables individuals to gain valuable insights into the final appearance of their T-shirt and make necessary adjustments. In summary, the major contributions and novel aspects of our work are as follows:

- **Novel System.** We introduce the MYCloth system, an intelligent and interactive platform for online T-shirt customization. MYCloth empowers users to seamlessly design and personalize their T-shirts by offering intuitive tools for style selection, color choices, and design customization.
- **Novel Algorithm.** 1) **A LLM-assisted Paint Generation Model.** Our algorithm presents a novel approach by integrating ChatGPT to refine and specify user-provided textual descriptions. This enhancement results in more detailed and accurate inputs for our text-to-image generation model, ultimately leading to higher-quality, personalized clothing designs. 2) **A Novel Learning-based Virtual Try-On Model.** We propose an advanced virtual try-on model that leverages state-of-the-art computer vision and deep learning techniques. This model enables users to visualize and interact with their customized T-shirts on virtual avatars, providing a realistic and immersive online try-on experience.
- **Comprehensive Evaluation.** 1) Numerical evaluation. 2) Visual comparison; 3) User study. Through quantitative metrics and qualitative feedback, we demonstrate the effectiveness and efficiency of the MYCloth system.

II. RELATED WORK

A. Computer-aided Clothing Design

The intersection of artificial intelligence (AI) and fashion design, notable for its academic and industrial signifi-

cance [12]–[15], has led to significant advancements in apparel Computer-Aided Design (CAD) systems, fabric and color selection, fit optimization, image synthesis, and the evaluation of virtual 3D designs. Systems like StyleMe [15] exemplify AI's role in automating design generation. Despite these developments, traditional design tools still heavily rely on the expertise of skilled designers, limiting inclusivity for non-experts. Interactive design systems [16]–[18] allow for pattern and fabric customization, but their utility in broader aspects of fashion customization, such as T-shirt design where customers focus on colors and prints rather than patterns, remains limited.

B. Text-to-image Generation

Text-to-image generation models, like TediGAN [19] and DALL-E [20], use textual descriptions to create realistic images, bridging linguistic creativity and visual design [21]–[23]. These models, applicable in fashion design, fall into two categories: GAN-based methods (e.g., TediGAN using StyleGAN [24]) and two-stage methods. The latter involves transforming images into discrete tokens via variational autoencoders and then reconstructing them into images, as seen in DALL-E and Cogview [20], [22]. ARMANI enhances this process by integrating textual information early on for more expressive tokens [23]. While these models expand fashion visuals, they lack the ability for detailed garment customization, like specific print incorporation. This paper diverges from generating garments via text-to-image models, focusing instead on a re-prompting method based on large language models, such as ChatGPT, coupled with stable diffusion techniques for generating specific designs [10], [11].

C. AI-driven Virtual Try-on

Virtual try-on technologies, crucial in digital garment visualization, utilize a two-stage process involving garment deformation and image synthesis [25]–[33]. Previous methods focused on neural network-based control points for deformation [33], [34] or appearance flow maps for non-rigid deformation [34], [35]. This paper proposes a novel one-stage framework for more efficient virtual try-on generation.

III. SYSTEM DESIGN AND IMPLEMENTATION

In this section, we introduce the proposed intelligent and interactive online T-shirt customization system, named MYCloth, tailored to provide users (*i.e.*, consumers) with an engaging and efficient T-shirt customization experience. The overall system is shown in Fig. 3. The MYCloth system is composed of four key components: Pattern Selection, Paint Generation, Cloth Adjustment, and Virtual Try-On. Designed to simplify the creative customization process, users can seamlessly design T-shirts through intuitive interactions, such as clicking buttons and drag-and-drop actions. The sections are structured as follows: Section III-A illustrates the details of the pattern selection. Subsequently, Section III-B introduces a novel method for generating paints in accordance with user preferences. The subsequent Section III-C is dedicated to the image processing methods of cloth adjustment. Section III-D,

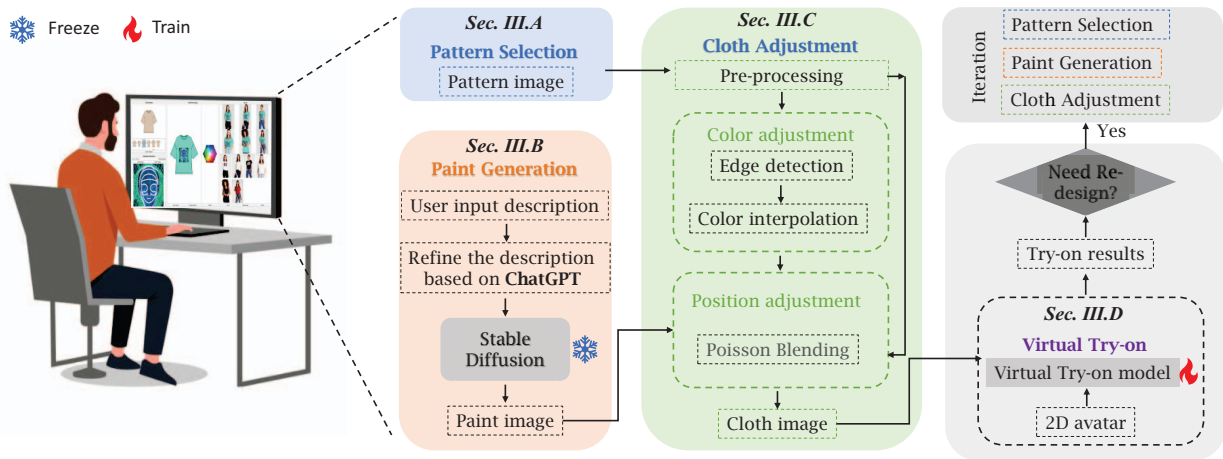


Fig. 3: The overall structure of our system.

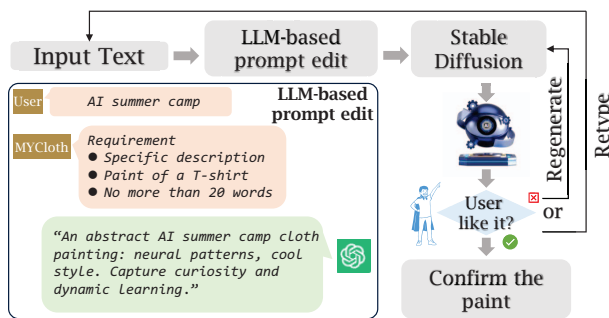


Fig. 4: The overview of the paint generation. Users input text, which is optimized using ChatGPT, and the stable diffusion model is employed to generate prints.

introduces a robust learning-based virtual try-on model that vividly visualizes the designs on virtual avatars.

A. Pattern Selection

The pattern selection phase serves as a fundamental starting point within the system. In this phase, customers are presented with a collection of T-shirt patterns. These patterns have been meticulously designed by experienced fashion designers. This collaboration ensures that those customers can access high-quality patterns that serve as a basis for their creative exploration. In essence, the pattern selection embodies MYCloth’s commitment to inclusivity and collaboration. Through this approach, we aim to provide users with a solid foundation for their design journey, regardless of their experience level.

B. LLM-assisted Paint Generation Model

Customizing prints currently involves three methods: 1) Consumers design the print themselves, 2) they select from existing designs offered by sellers, or 3) they propose a theme and have professional illustrators design it, which incurs additional costs and time. This paper explores using a text-to-image generation model to create paint based on textual

descriptions of design themes. However, slight variations in descriptions can lead to significant differences in outcomes. To address this, we propose a two-step solution: first, refining the textual description using ChatGPT, and then feeding these enhanced descriptions into a text-to-image model employing stable diffusion techniques for paint generation, as depicted in Fig 4. This process starts with the user’s theme description, which is refined through interaction with ChatGPT, and then used in the text-to-image process to create the paint. This method combines creative input with AI, producing paint that accurately reflects the intended theme.

C. Cloth Adjustment

The cloth adjustment component allows users to easily tailor and refine T-shirt designs to their liking. It offers an intuitive interface for adjusting elements like color, as well as the position and scale of the paint, thus enhancing the customization and personalization of T-shirts. This component comprises two main features: color adjustment and position adjustment.

1) Color adjustment. The cloth image processing involves several steps to maintain visual integrity and color consistency. Initially, pre-processing techniques denoise the image and analyze the pixel distribution. Edge detection algorithms are then applied to accurately identify and preserve the cloth’s boundary colors and positions. Following this, a pixel interpolation method is used to ascertain the color information for each pixel. This pre-processing phase gathers the cloth’s overall color distribution, recording the main image color as (R_m, G_m, B_m) . For each pixel, the process determines whether it lies on the cloth’s edge; if so, its color remains unchanged. Otherwise, the final pixel color is calculated through interpolation, considering its original color, the cloth’s main color (R_m, G_m, B_m) , and the target cloth color (R_t, G_t, B_t) .

2) Position adjustment. The position adjustment function enables users to precisely control the placement of the paint image within the interface. By clicking and dragging within

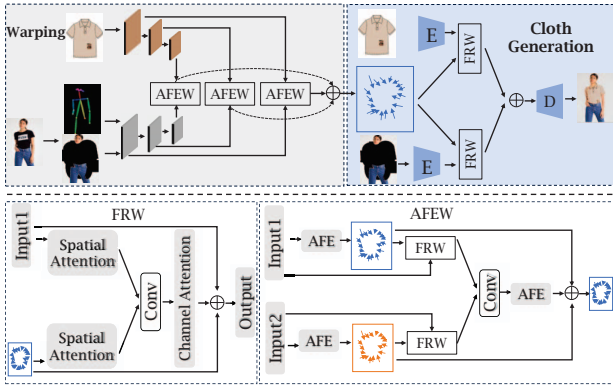


Fig. 5: The virtual try-on framework. **AFEW**: attention-based flow estimation and warping module. **FRW**: flow rectification and warping block. **AFE**: appearance flow estimator

the boundaries of the paint, users can freely relocate it to any desired position and scale within the cloth.

D. Learning-based Virtual Try-On Model

Our system integrates a virtual try-on, allowing users to project their T-shirt designs onto virtual avatars. This offers a seamless transition from digital design concepts to practical visuals, aiding users in making more informed design choices. As depicted in the figure, the system employs a single-stage framework for generating virtual try-on results. This framework comprises two key components: a warping module and a cloth generation module.

1) *Warping Module*: The inputs comprise the cloth image (x_c) and the 2D human image (x_h). The cloth image undergoes processing through the cloth branch, while the 2D human images are initially subjected to preprocessing to generate the human pose image and the human agnostic image. The latter masks the upper body of the depicted person. Subsequently, the human image branch receives the concatenated form of the human pose image and the human agnostic image. Both the cloth and the human branch have the same feature extraction structure, namely the Feature Pyramid Network (FPN) [36]. Notably, the two branches do not share the parameters. For each scale, individual feature maps (f_c^i and f_h^i) representing the cloth and human branches are obtained.

To facilitate the transformation of the cloth image to align with the contours of the human body, a methodology inspired by [32], [35], [37] is adopted to entail the estimation of appearance flow across scales. Specifically, predicated upon the extracted features f_c^i and f_h^i for each scale, an attention-based flow estimation and warping module (AFEW) is introduced, as depicted in Fig.6 (b). The appearance flow estimator (AFE) [35] is introduced to deduce flow vectors F_c^i and F_h^i . The formulation is as follows.

$$\begin{aligned} F_c^i &= AFE(f_c^i) \\ F_h^i &= AFE(f_h^i) \end{aligned} \quad (1)$$

To refine the acquired appearance flow, a flow rectification and warping block (FRW) is proposed. Given the output of the cloth feature and human features, the AFE estimates the fused appearance flow. Subsequently, a residual connection is implemented to further enhance the quality of the generated appearance flow.

The structure of the FRW block is as follows: the image feature (whether cloth or human image) and the predicted appearance flow are processed through a spatial attention mechanism. This is followed by the fusion of the image feature maps and the appearance flow. A convolutional layer and a channel attention mechanism are subsequently introduced to rectify channel-related information. The ultimate step involves the application of a residual process to yield conclusive warping results.

2) *Cloth Generation Module*: Upon acquiring the flow and warping results, the cloth generation module is introduced to generate the virtual try-on images. This process entails several key steps. A shared encoder is deployed to map both the cloth image and the agnostic images into the latent space. Subsequently, the generated flow, warping outcomes, and one of the image features (either the cloth image or the agnostic image) are fused within the AFEW module. The aggregated features are then directed fed into a shallow decoder, ultimately generating the final virtual try-on result (y_p). Notably, the employed shallow encoder and decoder are each composed of two convolution layers without any down sampling.

3) *Training Losses*: The loss function has two primary components: the similarity loss function and the perceptual loss function [38]. The similarity loss is rooted in L1 distance and quantifies the dissimilarity between the predicted image and the corresponding ground truth image.

$$\mathcal{L}_s = \|y_p - y_g\|_1, \quad (2)$$

Here, $\|\cdot\|_1$ denotes the L1 distance. The variables y_p and y_g respectively signify the predicted virtual try-on image and the ground truth image. The perceptual loss function is harnessed to ensure the consistency of feature maps, which are extracted through the VGG19 model [39], between the predicted virtual try-on image and the ground truth image. This is mathematically formulated as follows:

$$\mathcal{L}_{per} = \sum_{i=1}^5 \|\phi_i(y_p) - \phi_i(y_g)\|_1, \quad (3)$$

where $\phi_i()$ signifies the features of the i_{th} layer of the VGG-19 model. The comprehensive loss is presented as:

$$\mathcal{L} = \sum_{n=1}^N (n+1) * (\lambda_s \mathcal{L}_s^n + \lambda_{per} \mathcal{L}_{per}^n), \quad (4)$$

where, \mathcal{L}^n symbolizes the loss at the n_{th} scale.

IV. EVALUATION

This section presents an evaluation of the MYCloth system's capabilities in two aspects: 1) virtual try-on performance,

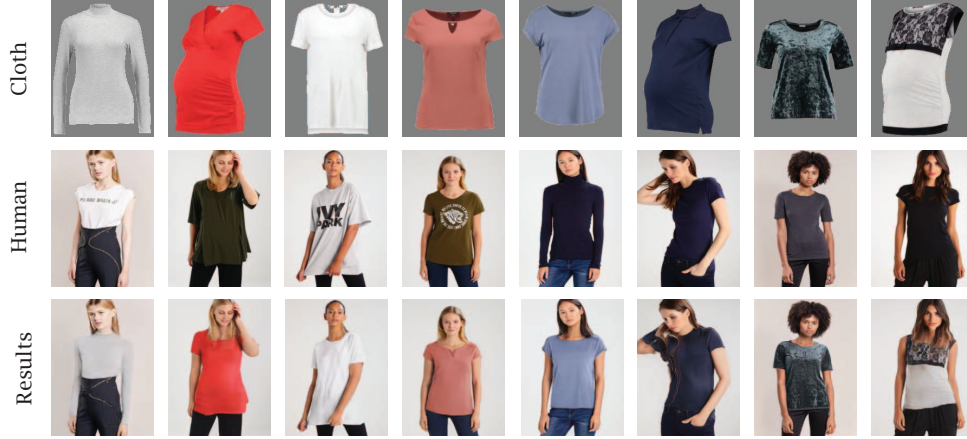


Fig. 6: The visual results of the virtual try-on method on the VITON dataset.

TABLE I: Quantitative comparisons with previous virtual try-on methods on the VITON dataset (paired setting).

Method	FID (\downarrow)	SSIM (\uparrow)	IS (\uparrow)	PSNR (\uparrow)
CP-VTON	30.50	0.784	2.757	21.01
ClothFlow	23.68	0.843	-	23.60
ZFlow	15.17	0.885	-	25.46
Ours	11.32	0.887	2.846	26.19

and 2) user experience. Firstly, we conduct a comparison experiment to validate the effectiveness of the virtual try-on model and an ablation study to validate the proposed modules. Then, we conducted a user study to determine if the MYCloth system can reduce discrepancies between the envisioned and final T-shirt. Finally, we conducted a user study to evaluate if the MYCloth system can improve the consumer’s interactive experience when customizing clothes online.

A. Evaluation of the virtual try-on

Implementation Details. The cloth and human encoders within the warping module are both structured according to the Feature Pyramid Network (FPN) [36], comprising five downsampling stages. Notably, these encoders are independently parameterized. The appearance flow estimator, crucial for accurate warping, is constructed using four convolutional layers. Their hidden dimensions follow a sequence of 512, 256, 128, and 64, respectively. In the cloth generation module, both the cloth and human encoders adopt an identical structure. This is reinforced by shared parameters. Each encoder and decoder is structured with three convolution layers, each with hidden dimensions of 32, 64, and 128. It’s important to highlight that feature maps are not subsampled in this process. For our experiments, we utilized the PyTorch framework on A6000 GPUs. Employing the Adam optimizer, we adopted a batch size of 16 for training. The training spanned 200 epochs, initialized with a learning rate of 5×10^{-5} . Subsequently, the learning rate underwent a reduction to 0.1 times the original

every 50 epochs. The loss function weight factors, λ_s and λ_{per} , were both set to 1, ensuring a balanced impact of the similarity and perceptual loss components.

Datasets. Our experimentation is grounded in the widely recognized VITON dataset [40], a standard resource for virtual try-on research. This dataset serves as a pivotal benchmark in the field and comprises a training set encompassing 14,221 image pairs, alongside a distinct testing set containing 2,032 image pairs. Notably, each pair is composed of a front-view photograph and an in-shop clothing image, all calibrated to a resolution of 256×192 pixels. This dataset’s established relevance and comprehensive coverage underscore its suitability for evaluating our proposed system.

Comparison Experiments. To evaluate the effectiveness of our method, we conducted a comparison experiment with previous virtual try-on methods, including Cloth-Flow [41], CP-VTON [42], and ZFlow [43]. In the comprehensive evaluation, we utilize several established metrics to assess the quality of our synthesized images compared to ground truth images. These metrics include the Structure Similarity Index Measure (SSIM) [44], the Peak Signal-to-Noise Ratio (PSNR) [45], the Fréchet Inception Distance (FID) [46], and the Inception Score (IS) [47]. These metrics collectively measure structural similarity, image fidelity, distributional similarity, and the realism of the generated images. Ground truth images of the same individuals wearing identical clothing are employed as references for these evaluations, ensuring a robust and quantitative assessment of our system’s performance.

In our comprehensive quantitative evaluation of virtual try-on methods using the VITON dataset, as shown in Tab. I, our proposed method demonstrates exceptional performance across multiple key metrics. With the lowest FID of 11.32, the SSIM of 0.887, a competitive IS of 2.846, and the highest PSNR of 26.19, our method consistently outperforms other previous approaches, such as CP-VTON, ClothFlow, and ZFlow. These results underscore the effectiveness of our method in producing virtual try-on images that are not only

highly realistic and structurally similar to real images but also diverse in content, making it a compelling choice for enhancing the online customized shopping experience.

To evaluate the efficacy of our algorithm, we conducted an empirical study employing the VITON dataset. The visual results of this study are elucidated in Fig. 6. In this figure, we observe three columns of images. The first column showcases the reference cloth images, while the second column exhibits the human images. The third column presents the outcomes achieved through the application of our method. Notably, our algorithm demonstrates remarkable proficiency in capturing intricate textures and faithfully reproducing poses.

Ablation Experiments. In Tab. II, we systematically analyzed the impact of different modules on our framework’s performance, using abbreviated notations for clarity. We started with the baseline configuration and progressively introduced components. The Attention-based Flow Estimation and Warping Module (AFEW) significantly improved both SSIM and PSNR, highlighting its value in enhancing image quality. The incorporation of the Flaw Rectification and Warping Block (FRW) within the Warping module and later within the cloth generation module further elevated these metrics, indicating their crucial roles in refining the warping and generation processes. Ultimately, the combined presence of all modules led to the highest SSIM (0.887) and PSNR (26.19) values, showcasing the effectiveness of these components and underscoring their collective contribution to superior performance.

B. Evaluation of the User Experience

To assess the practicality of our system, we conducted comprehensive evaluations involving 15 students and 3 T-shirt customization sellers. Data collection was accomplished through a combination of quantitative analysis, involving subjective ratings, and qualitative research via semi-structured interviews, with the overarching goal of thoroughly assessing the user experience.

1) *Study Design:* We structured our evaluation approach with tailored surveys for consumers and T-shirt customization sellers. Each participant was invited to use our system to create a customized T-shirt item based on a specific theme, followed by individual interviews to gather insights.

2) *Participants:* Among the participants, 8 participants were female and 7 participants were male, they were between 18 and 27 years old. They represented diverse academic backgrounds, including public policy (4 participants), computer artificial intelligence (3 participants), mechanical engineering (2 participants), applied mathematics (2 participants), bioengineering (2 participants), and finance (2 participants). They all had experience wearing group-customized T-shirts. For the T-shirt customization sellers, we engaged three online T-shirt customization sellers based on their established sales.

3) *Task and procedures:* Each participant was tasked with completing four specific assignments:

- **Style and Color Selection.** Participants are required to select a T-shirt style as the starting point for their

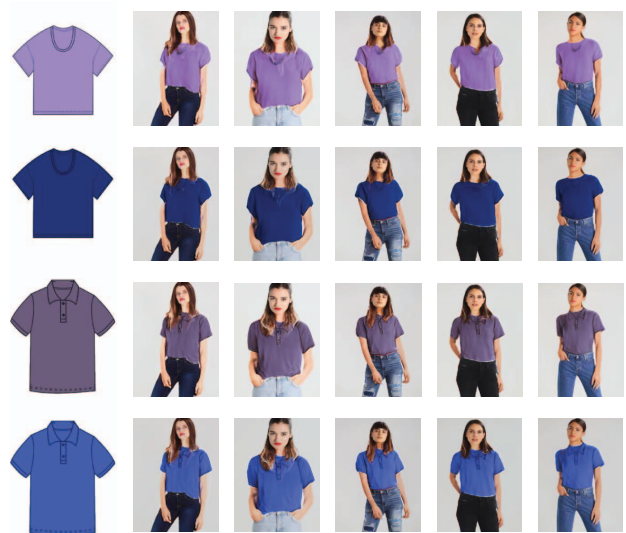


Fig. 7: Example of the user uses our system to change the color of a cloth in a chosen pattern and uses the virtual try-on to see how it looks on the body..



Fig. 8: Users use our system to generate prints based on a given theme, change the color of the T-shirt, change the size and position of the prints, and use the virtual try-on to see an example of how it looks on the body.

personalized design. The objective is to establish the foundational elements of the customized T-shirt design.

- **Paint Generation.** Participants input a theme or text that will guide their T-shirt design. They can enter slogans, quotes, or any text they wish to feature on the T-shirt. This text provides a personalized touch to the design, aligning it with their creative vision.
- **Cloth Editing.** This task allows participants to edit and fine-tune the pattern elements of their T-shirt design. They can change the T-shirt’s color, adjust print sizes and positions, and even add additional pattern elements like icons or shapes.
- **Virtual Try-On.** Participants virtually try on their per-

TABLE II: Quantitative comparisons with previous virtual try-on methods on the VITON dataset (paired setting).

Baseline	AFEW	FRW within the Warping module	FRW within the cloth generation module	SSIM (\uparrow)	PSNR (\uparrow)
✓				0.815	20.26
✓	✓			0.862	24.73
✓			✓	0.828	21.95
✓	✓		✓	0.871	25.95
✓	✓			0.887	26.19

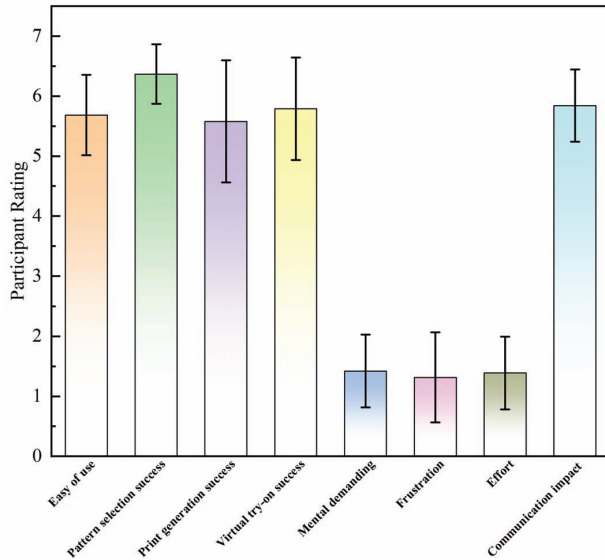


Fig. 9: The responses from participants were evaluated using a seven-point Likert scale for all questions.

sonalized T-shirt designs. This step provides insights into the final look and fit of the customized T-shirt.

4) *Measurement*: After participants completed their tasks, they were asked to provide assessments on various aspects of the system using a 7-point Likert scale.

5) *Results: overall assessment*: In this section, we provide an overview of the questionnaire responses collected from users who engaged with the MYCloth system, as depicted in Fig. 9 and Fig. 10. Participants are identified by labels ranging from P1 to P18, with P1-15 representing consumers and P16-18 representing T-shirt customization sellers. The ratings encompass various aspects, including system usability, pattern selection satisfaction, print generation success, virtual try-on effectiveness, mental demand, frustration, effort, and communication impact. Examples obtained by some users using our system are shown in Fig. 7 and Fig. 8.

V. CONCLUSION

In this paper, we introduced MYCloth, an innovative online T-shirt customization system designed to enhance personalized fashion experience, offering users the opportunity to become designers themselves and create a T-shirt that aligns with

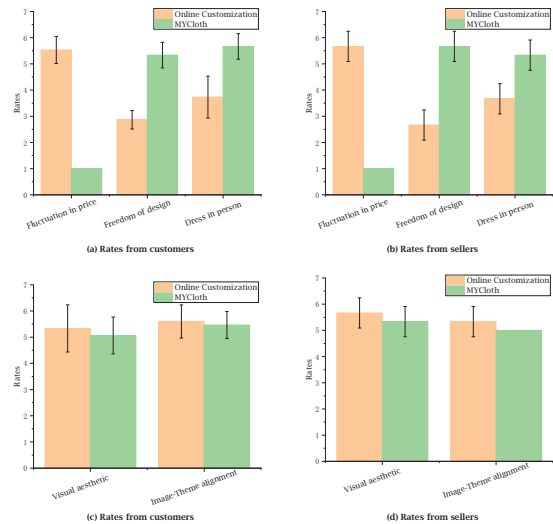


Fig. 10: The responses from customers and sellers were evaluated using a seven-point Likert scale.

their unique tastes. With an intuitive interface, advanced print generation based on a large language model and stable diffusion model, and a robust virtual try-on model, MYCloth redefines the customization process. Through extensive user surveys, interviews, and usability assessments involving both consumers and T-shirt customization sellers, this research has underscored the system’s efficacy. Users consistently expressed high satisfaction with MYCloth’s user-friendliness, the realism of virtual try-ons, and the quality of generated prints. These results emphasize how MYCloth enhances T-shirt customization, offering an enjoyable, efficient, and unique way to personalize fashion choices.

REFERENCES

- [1] Y. Zhao, “Manufacturing personalization models based on industrial big data,” *Journal of Discrete Mathematical Sciences and Cryptography*, vol. 21, no. 6, pp. 1287–1292, 2018.
- [2] D. Matthews, L. Rothenberg, and S. Gopalakrishnan, “The impact of mass customization on fashion-innovative students: an assessment of need for uniqueness, self-identity, and perceived performance risk,” *International Journal of Fashion Design, Technology and Education*, vol. 12, no. 3, pp. 293–300, 2019.
- [3] R. De Silva, T. D. Rupasinghe, and P. Apeageyi, “A collaborative apparel new product development process model using virtual reality and augmented reality technologies as enablers,” *International Journal*

- of Fashion Design, Technology and Education*, vol. 12, no. 1, pp. 1–11, 2019.
- [4] B. A. Manko, “The evolution of fashion: How digital technology makes a basic graphic t-shirt a marketing staple,” *Journal of Information Technology Teaching Cases*, p. 20438869231151971, 2023.
 - [5] T. H. Nobile, A. Noris, N. Kalbaska, and L. Cantoni, “A review of digital fashion research: before and beyond communication and marketing,” *International journal of fashion design, technology and education*, vol. 14, no. 3, pp. 293–301, 2021.
 - [6] B. Huang, C. Juaneda, S. Sénécal, and P.-M. Léger, ““now you see me”: the attention-grabbing effect of product similarity and proximity in online shopping,” *Journal of Interactive Marketing*, vol. 54, no. 1, pp. 1–10, 2021.
 - [7] B. Tudjarov, A. Bachvarov, and I. Boyadjiev, “Web virtual reality for product customization,” in *Proceedings of the 3rd international conference on mass customization and personalization in Central Europe (MCP-CE 2008)*, Novi Sad, Serbia, pp. 7–13, 2008.
 - [8] “Teespring is now & spring,” 2023.
 - [9] “Zazzle,” 2016.
 - [10] “Openai. chatgpt (feb 13 version) [large language model].,” 2023.
 - [11] R. Rombach, A. Blattmann, D. Lorenz, P. Esser, and B. Ommer, “High-resolution image synthesis with latent diffusion models,” in *CVPR*, pp. 10684–10695, 2022.
 - [12] Y. Zhu, W. Wang, and C. Jiang, “The application of clothing patterns based on computer-aided technology in clothing culture teaching,” *Computer-Aided Design and Applications*, pp. 145–155, 2022.
 - [13] J. Li, “The application of curve algorithm and computer aided design in garment design,” in *2021 International Conference on Big Data Analytics for Cyber-Physical System in Smart City: Volume 2*, pp. 565–570, Springer, 2022.
 - [14] N. Särnkäri and A. Vänskä, ““just hit a button!”—fashion 4.0 designers as cyborgs, experimenting and designing with generative algorithms,” *International Journal of Fashion Design, Technology and Education*, vol. 15, no. 2, pp. 211–220, 2022.
 - [15] D. Wu, Z. Yu, N. Ma, J. Jiang, Y. Wang, G. Zhou, H. Deng, and Y. Li, “Styleme: Towards intelligent fashion generation with designer style,” in *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, pp. 1–16, 2023.
 - [16] X.-j. Zhu, H. Lu, and M. Rättsch, “An interactive clothing design and personalized virtual display system,” *Multimedia tools and applications*, vol. 77, pp. 27163–27179, 2018.
 - [17] X. Zhu, X. Li, Y. Chen, J. Liu, X. Zhao, and X. Wu, “Interactive genetic algorithm based on typical style for clothing customization,” *Journal of Engineered Fibers and Fabrics*, vol. 15, p. 1558925020920035, 2020.
 - [18] Z. Wang, Y. Xing, J. Wang, X. Zeng, Y. Yang, and S. Xu, “A knowledge-supported approach for garment pattern design using fuzzy logic and artificial neural networks,” *Multimedia Tools and Applications*, pp. 1–21, 2022.
 - [19] W. Xia, Y. Yang, J.-H. Xue, and B. Wu, “Tedigan: Text-guided diverse face image generation and manipulation,” in *CVPR*, pp. 2256–2265, 2021.
 - [20] A. Ramesh, M. Pavlov, G. Goh, S. Gray, C. Voss, A. Radford, M. Chen, and I. Sutskever, “Zero-shot text-to-image generation,” in *ICML*, pp. 8821–8831, PMLR, 2021.
 - [21] P. Esser, R. Rombach, and B. Ommer, “Taming transformers for high-resolution image synthesis,” in *CVPR*, pp. 12873–12883, 2021.
 - [22] M. Ding, Z. Yang, W. Hong, W. Zheng, C. Zhou, D. Yin, J. Lin, X. Zou, Z. Shao, H. Yang, et al., “Cogview: Mastering text-to-image generation via transformers,” *Advances in Neural Information Processing Systems*, vol. 34, pp. 19822–19835, 2021.
 - [23] X. Zhang, Y. Sha, M. C. Kampffmeyer, Z. Xie, Z. Jie, C. Huang, J. Peng, and X. Liang, “Armani: Part-level garment-text alignment for unified cross-modal fashion design,” in *Proceedings of the 30th ACM International Conference on Multimedia*, pp. 4525–4535, 2022.
 - [24] T. Karras, S. Laine, and T. Aila, “A style-based generator architecture for generative adversarial networks,” in *CVPR*, pp. 4401–4410, 2019.
 - [25] A. Baldtrati, D. Morelli, G. Cartella, M. Cornia, M. Bertini, and R. Cucchiara, “Multimodal garment designer: Human-centric latent diffusion models for fashion image editing,” *arXiv preprint arXiv:2304.02051*, 2023.
 - [26] Z. Xie, Z. Huang, X. Dong, F. Zhao, H. Dong, X. Zhang, F. Zhu, and X. Liang, “Gp-vton: Towards general purpose virtual try-on via collaborative local-flow global-parsing learning,” in *CVPR*, pp. 23550–23559, 2023.
 - [27] K. Yan, T. Gao, H. Zhang, and C. Xie, “Linking garment with person via semantically associated landmarks for virtual try-on,” in *CVPR*, pp. 17194–17204, 2023.
 - [28] L. Zhu, D. Yang, T. Zhu, F. Reda, W. Chan, C. Saharia, M. Norouzi, and I. Kemelmacher-Shlizerman, “Tryondiffusion: A tale of two unets,” in *CVPR*, pp. 4606–4615, 2023.
 - [29] Z. Huang, H. Li, Z. Xie, M. Kampffmeyer, X. Liang, et al., “Towards hard-pose virtual try-on via 3d-aware global correspondence learning,” *Advances in Neural Information Processing Systems*, vol. 35, pp. 32736–32748, 2022.
 - [30] H. Yang, X. Yu, and Z. Liu, “Full-range virtual try-on with recurrent tri-level transform,” in *CVPR*, pp. 3460–3469, 2022.
 - [31] X. Dong, F. Zhao, Z. Xie, X. Zhang, D. K. Du, M. Zheng, X. Long, X. Liang, and J. Yang, “Dressing in the wild by watching dance videos,” in *CVPR*, pp. 3480–3489, 2022.
 - [32] Y. Ge, Y. Song, R. Zhang, C. Ge, W. Liu, and P. Luo, “Parser-free virtual try-on via distilling appearance flows,” in *CVPR*, pp. 8485–8493, 2021.
 - [33] S. Lee, G. Gu, S. Park, S. Choi, and J. Choo, “High-resolution virtual try-on with misalignment and occlusion-handled conditions,” in *ECCV*, pp. 204–219, Springer, 2022.
 - [34] T. Issenhuth, J. Mary, and C. Calauzenes, “Do not mask what you do not need to mask: a parser-free virtual try-on,” in *ECCV*, pp. 619–635, Springer, 2020.
 - [35] S. Bai, H. Zhou, Z. Li, C. Zhou, and H. Yang, “Single stage virtual try-on via deformable attention flows,” in *ECCV*, pp. 409–425, Springer, 2022.
 - [36] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, “Feature pyramid networks for object detection,” in *CVPR*, pp. 2117–2125, 2017.
 - [37] R. A. Güler, N. Neverova, and I. Kokkinos, “Densepose: Dense human pose estimation in the wild,” in *CVPR*, pp. 7297–7306, 2018.
 - [38] J. Johnson, A. Alahi, and L. Fei-Fei, “Perceptual losses for real-time style transfer and super-resolution,” in *ECCV*, pp. 694–711, Springer, 2016.
 - [39] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” *arXiv preprint arXiv:1409.1556*, 2014.
 - [40] X. Han, Z. Wu, Z. Wu, R. Yu, and L. S. Davis, “Viton: An image-based virtual try-on network,” in *CVPR*, pp. 7543–7552, 2018.
 - [41] X. Han, X. Hu, W. Huang, and M. R. Scott, “Clothflow: A flow-based model for clothed person generation,” in *CVPR*, pp. 10471–10480, 2019.
 - [42] B. Wang, H. Zheng, X. Liang, Y. Chen, L. Lin, and M. Yang, “Toward characteristic-preserving image-based virtual try-on network,” in *ECCV*, pp. 589–604, 2018.
 - [43] A. Chopra, R. Jain, M. Hemani, and B. Krishnamurthy, “Zflow: Gated appearance flow-based virtual try-on with 3d priors,” in *CVPR*, pp. 5433–5442, 2021.
 - [44] K. Seshadrinathan and A. C. Bovik, “Unifying analysis of full reference image quality assessment,” in *IEEE International Conference on Image Processing*, pp. 1200–1203, IEEE, 2008.
 - [45] A. Hore and D. Ziou, “Image quality metrics: Psnr vs. ssim,” in *International Conference on Pattern Recognition*, pp. 2366–2369, IEEE, 2010.
 - [46] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, and S. Hochreiter, “Gans trained by a two time-scale update rule converge to a local nash equilibrium,” *Advances in neural information processing systems*, vol. 30, 2017.
 - [47] T. Salimans, I. Goodfellow, W. Zaremba, V. Cheung, A. Radford, and X. Chen, “Improved techniques for training gans,” *Advances in neural information processing systems*, vol. 29, 2016.