FashionRNA: Interactive Reimagination of Fashion Heritage

Haneol Kim Jonghwan Bae Seonwoo Shin Sanghun Park*
Graduate School of Metaverse, Sogang University
Seoul, South Korea
{skykim, bjh0309, shine, mshpark}@sogang.ac.kr

Abstract

In this paper, we propose a novel framework that integrates physical fashion heritage with interactive digital creation, facilitating the reinterpretation of historical garments via an intuitive multimodal interface. This interface allows users to manipulate digital representations of historical artifacts, specifying stylistic modifications through a combination of visual and vocal inputs. These requests are processed by a computational pipeline comprising state-of-the-art artificial intelligence (AI) models. The pipeline's workflow is as follows: it first retrieves a target style based on user input, then generates a corresponding 3D garment model, and concurrently synthesizes Physically Based Rendering (PBR) materials derived from the source pattern. The resulting 3D asset is optimized for real-time rendering, enabling immediate visualization. The results demonstrate that this approach generates high-fidelity and stylistically coherent outcomes, effectively transforming static cultural artifacts into dynamic, interactive digital assets.

1. Introduction

Fashion transcends its utilitarian function as mere clothing, serving as a vital vessel for culture and heritage. Prestigious events such as the Met Gala demonstrate how fashion operates beyond commercial realms as a form of ceremonial, intellectual, emotional, and political living performance. These events emphasize the power of fashion to commemorate history and identity, elevating garments into narratives that embody social values and historical continuity. The case of Korea's legendary designer André Kim illustrates the significance of such heritage; by fusing traditional Korean elements with futuristic silhouettes, his works became a global symbol of Korean culture.

However, in the digital age, simply preserving these garments in museums is not enough. The challenge lies in transforming this static heritage into a living, interactive experience that can inspire future generations. In response, we propose fashion reimagined through neural augmentation (FashionRNA), a framework that bridges the gap between physical fashion heritage and the boundless potential of digital creation. Our system fundamentally rethinks the design workflow, moving from a complex, manual process to a seamless, AI-driven interaction, as shown in Figure 1.

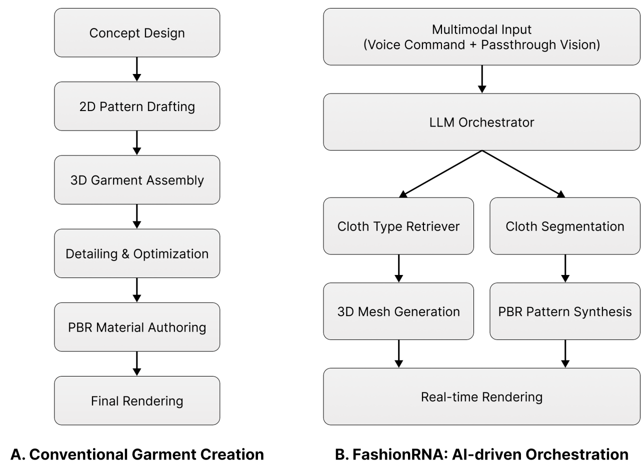


Figure 1. From manual to AI-automated garment production. (A) The conventional workflow is manual, slow, and requires specialized tools. (B) FashionRNA automates and unifies the process, generating 3D assets directly from intuitive user commands.

Our work makes several contributions. We first design a holistic workflow that integrates state-of-the-art AI models into a cohesive system, creating a novel user experience. We propose a novel role for the large language model (LLM) as an orchestrator that interprets complex user intent from multimodal inputs to automate the entire creative pipeline. This architecture establishes a paradigm for human-AI interaction, where users can intuitively reinterpret cultural heritage in real-time using voice and vision, transforming them from passive observers into active co-creators. Ultimately, FashionRNA is more than a digital archive. It serves as a bridge between past and future, turning cultural heritage into a source of endless inspiration across generations.

*Corresponding Author.

2. Related Work

2.1. Virtual Try-On and Fashion Generation

Virtual try-on (VTO) synthesizes realistic images of people wearing specific garments, revolutionizing online shopping by reducing return rates. The core challenge lies in realistically deforming non-rigid garments while preserving intricate details, a domain where 3D design software like CLO 3D has long excelled. Early GAN-based approaches like VITON [7] and CP-VTON [20] used two-stage pipelines: geometric warping followed by synthesis. However, these methods suffered from artifacts due to misalignment [1]. The field has recently shifted toward diffusion models, offering superior generation quality. TryOnDiffusion [21] introduced Parallel-UNet, integrating warping and synthesis through cross-attention. Recent works like StableVITON [9] and IDM-VTON [2] leverage pretrained text-to-image models. Interactive systems like PromptDresser [10] now enable text-based style editing, redefining VTO as a conditional generation task.

2.2. Egocentric Vision for Human-AI Interaction

Egocentric vision captures first-person perspectives through body-worn cameras, providing rich context about user attention and intent. This modality is crucial for developing embodied AI systems in AR/VR applications, though it presents challenges including camera motion, occlusion, and viewpoint variations. Large-scale datasets like Ego4D [5] and Ego-Exo4D [6] have accelerated research in multimodal understanding. These advances enable applications like xR-EgoPose [18] for full-body pose estimation in HMDs, demonstrating the trajectory from passive observation to active collaboration—essential for interactive heritage preservation systems.

2.3. Immersive and Generative Approaches in Digital Heritage

Digital technologies like 3D restoration and XR are converting static cultural heritage into interactive experiences, enabling immersive narratives that traditional exhibits cannot match [16]. VR significantly enhances user engagement by turning passive observers into active participants, thereby deepening their connection to cultural heritage [22]. More recently, generative AI facilitates the creation of culturally sensitive content by analyzing and reinterpreting historical archives. A notable example uses machine learning to generate novel designs based on Miao ethnic apparel [3]. This convergence of immersive interaction and generative AI now forms the foundation for systems that respect original designs while enabling creative reinterpretations for future generations [12].

3. Proposed Framework

Figure 2 illustrates the proposed architecture: a modular pipeline that reinterprets a physical garment based on user instructions. The system processes a multimodal input, comprising a voice command and a visual capture of the source garment. An LLM orchestrates the workflow by parsing the voice command to initiate two parallel processes. The first process, 3D Model Generation, generates a new geometric mesh corresponding to the user's desired clothing style. The second, Material Synthesis, extracts and synthesizes a texture from the source garment image captured by the user. Finally, the pipeline converges at the rendering stage, where the generated 3D mesh and PBR material are combined in Unity [19] to produce a real-time, high-fidelity visualization of the redesigned garment.

3.1. Multimodal Interaction

The process begins with a multimodal input from the user, which consists of a voice command and a visual capture of a garment. For instance, when a user says, "I want to redesign this dress as a cocktail dress", the voice command is directed to an LLM for intent parsing. The LLM extracts the target garment style (i.e., "cocktail dress") to be used in the 3D Model Generation process. Concurrently, the visual input is channeled directly to the Material Synthesis pipeline. The LLM supports this process by providing a textual prompt (i.e., "dress") needed for the initial segmentation step within that pipeline. Thus, the parsed voice command and the visual data act as coordinated inputs that initiate the two parallel workflows.

3.2. 3D Model Generation

This process generates a 3D model of the target garment style through a hybrid pipeline that combines a precomputed asset library with on-demand generation. Our system uses a database of 90 dress types with pre-generated 3D models. These types are categorized by attributes such as length, sleeve style, and silhouette, drawing inspiration from industry taxonomies like "The Definitive Guide to Dress Design" [17], which illustrates various dress styles with both images and textual descriptions. For user prompts corresponding to these established styles, the system efficiently retrieves the pre-built 3D model. To accommodate designs outside the established categories, users can augment the search database with new data and generate the corresponding 3D meshes. To do this, we first employ the ColPali [4] system, a multimodal retriever that connects text and images, to find the top-1 dress image from a comprehensive image database that best matches the user's prompt. This retrieved 2D image is then transformed into a 3D model using Meshy.ai, an AI-based service that generates a 3D mesh from a single input image. This model serves as the base geometry for material mapping.

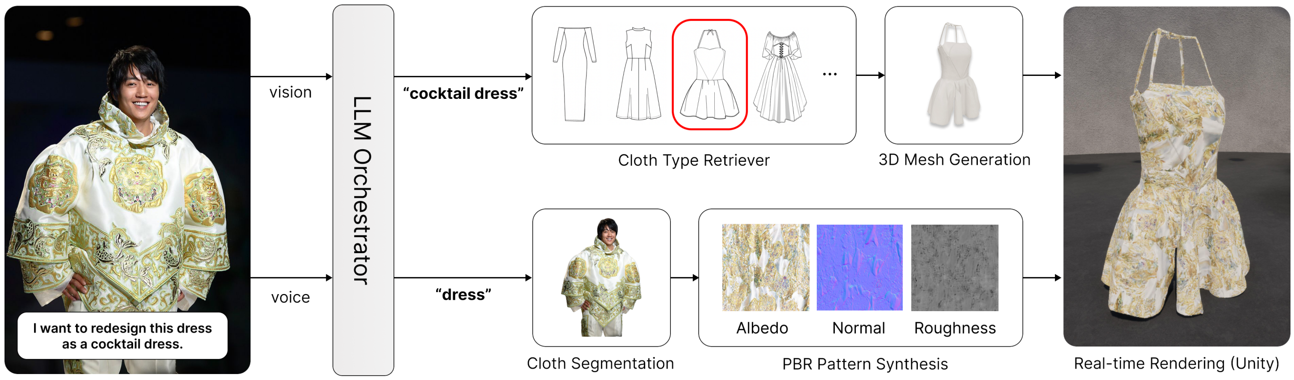


Figure 2. An example of reimagining an André Kim dress using FashionRNA. The user's request to "I want to redesign this dress as a cocktail dress" is processed by the LLM, which separates style and texture tasks. The system generates a new cocktail dress mesh and applies the PBR materials extracted from the original gown, resulting in a high-fidelity visualization of the redesigned garment.

3.3. Material Synthesis

In parallel with the 3D model generation, a three-stage pipeline is used to create a new PBR material from the garment the user is currently viewing. First, the system isolates the source garment using the Grounded Segment Anything Model (Grounded SAM) [13]. This process is driven by an LLM orchestrator that generates a segmentation prompt (e.g., "dress"), which Grounded SAM then uses to perform precise segmentation on the input image, creating a mask that isolates the clothing item. Next, the Material Palette [11] framework synthesizes a new pattern by fine-tuning a diffusion model to learn a material concept from the segmented region, generating a clean, tileable texture that removes scene-specific geometry and lighting. Finally, a domain-adapted network decomposes this generated texture into its intrinsic Spatially Varying Bidirectional Reflectance Distribution Function (SVBRDF) maps, specifically albedo, normal, and roughness. This process utilizes Unsupervised Domain Adaptation (UDA) to bridge the gap between synthetic training data and diffusion-generated textures. The final output is a palette of ready-to-use PBR materials corresponding to the selected garment regions in the original photograph. Despite the high-quality output, the entire process currently takes 3-4 minutes on an NVIDIA RTX 4090, presenting a significant limitation for real-time processing. This latency stems mainly from the computationally intensive fine-tuning process for 3D modeling and texturing, where methods like LoRA [8] and Dreambooth [15] are used to train a Stable Diffusion [14] model to learn new material concepts.

3.4. PBR Rendering in Unity

The final stage synthesizes the generated assets to produce the augmented visualization. Once both parallel processes are complete, the PBR texture maps are imported into the Unity (6000.0.50f1) High Definition Render Pipeline (HDRP). Within this environment, an Autodesk Interactive Shader is used to define a PBR material by correctly applying the corresponding texture maps. This newly defined material is then applied to the generated 3D mesh. The result is a high-fidelity rendering of the dress with the user-specified pattern, seamlessly integrated into the user's view, providing an immediate and compelling visualization of their requested design modification.

4. Results

To demonstrate FashionRNA's ability to bridge fashion heritage with contemporary digital creation, we conducted a qualitative experiment by reinterpreting the iconic works of the renowned late Korean designer, André Kim, a figure celebrated for his intricate and artistic designs. We tested the system's capacity to translate his classic, elaborate designs into modern garment styles using four distinct prompts: "I want to redesign this dress as a cocktail dress", "I want to redesign this dress as a bandage dress", "I want to redesign this dress as a bodycon dress", and "I want to redesign this dress as a harem dress".

As shown in Table 1, the multimodal retrieval system successfully identified a corresponding 2D reference image for all four prompts, demonstrating its strong grasp of complex user intent. The subsequent asset generation stage consistently produced accurate segmentation masks and high-fidelity PBR material maps. The resulting albedo, normal, and roughness maps effectively captured the intricate patterns and colors from the source garment. The final rendered outputs, presented in Table 2, confirm that the generated PBR materials were seamlessly applied to the new 3D models. Each result is a high-fidelity visualization of the redesigned garment, realistically reflecting the user's requested style while faithfully preserving the original pattern from the André Kim piece, thereby proving our method's success in fusing design history with new aesthetics.

Table 1. Intermediate Results of the Generation Process

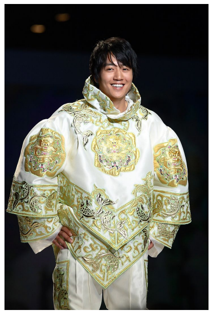
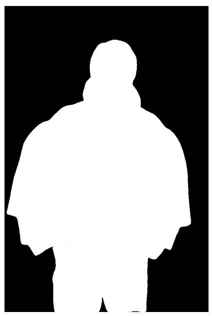
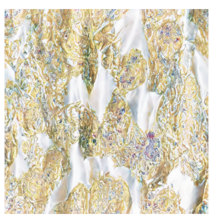
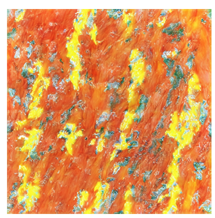
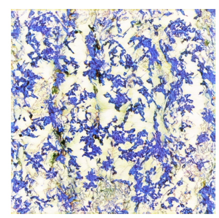
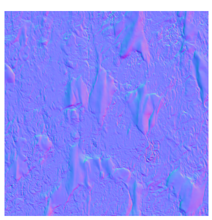
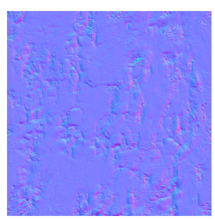
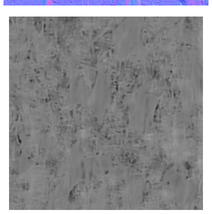
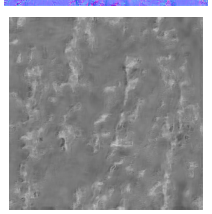
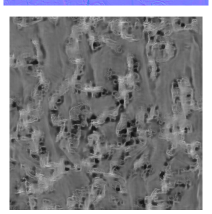
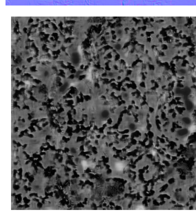
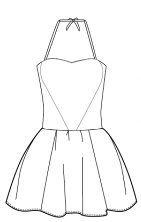
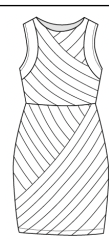
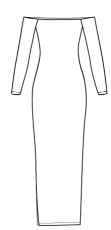
Prompt	"...cocktail dress"	"...bandage dress"	"...bodycon dress"	"...harem dress"
Image				
Mask				
Albedo				
Normal				
Roughness				
Retrieval				

Table 2. Synthesized Results of Redesigned Garments

Prompt	"I want to redesign this dress as a cocktail dress"	"I want to redesign this dress as a bandage dress"
		
Prompt	"I want to redesign this dress as a bodycon dress"	"I want to redesign this dress as a harem dress"
		

5. Discussion

In this work, we presented FashionRNA, a novel framework that orchestrates state-of-the-art AI models to bridge the gap between physical fashion heritage and interactive digital creation. Our primary contribution is not a single algorithmic improvement but rather the holistic design and integration of a workflow that streamlines digital content generation. By positioning the LLM as a central orchestrator, our system pioneers a new, intuitive creation paradigm that empowers users to transform cultural artifacts in immersive environments.

A significant technical limitation is the computational latency of our generative pipeline, which currently impedes a truly real-time user experience. To address this bottleneck, our future work will concentrate on system-wide optimization by employing techniques such as model distillation and asset caching. Furthermore, we plan to migrate to a more robust, open-source pipeline to enhance output fidelity and ensure academic reproducibility. The development of a comprehensive, open-access garment dataset will also be critical for rigorous quantitative benchmarking and user experience (UX) evaluation.

FashionRNA has the potential to become an impactful platform with applications ranging from reinterpreting cultural attire to powering personalized e-commerce VTO and fueling metaverse creator economy. Our goal is to evolve this concept into a powerful tool that bridges historical inspiration with future creation.

Acknowledgments

This work was supported by the NRF (National Research Foundation of Korea) and IITP (Institute for Information & Communications Technology Planning & Evaluation) grant funded by the Korea government (MSIT) (RS-2023-00251681, RS-2023-00259099, and RS-2022-00156318).

References

[1] Seunghwan Choi, Sunghyun Park, Minsoo Lee, and Jaegul Choo. Viton-hd: High-resolution virtual try-on via misalignment-aware normalization, 2021. 2

[2] Yisol Choi, Sangkyung Kwak, Kyungmin Lee, Hyungwon Choi, and Jinwoo Shin. Improving diffusion models for authentic virtual try-on in the wild, 2024. 2

[3] Meizhen Deng, Yimeng Liu, and Ling Chen. Ai-driven innovation in ethnic clothing design: an intersection of machine learning and cultural heritage. *Electronic Research Archive*, 31(9):5793–5814, 2023. 2

[4] Manuel Faysse, Hugues Sibille, Tony Wu, Bilel Omrani, Gautier Viaud, Céline Hudelot, and Pierre Colombo. Colpali: Efficient document retrieval with vision language models, 2025. 2

[5] Kristen Grauman, Andrew Westbury, et al. Ego4d: Around the world in 3,000 hours of egocentric video, 2022. 2

[6] Kristen Grauman, Andrew Westbury, et al. Ego-exo4d: Understanding skilled human activity from first- and third-person perspectives, 2024. 2

[7] Xintong Han, Zuxuan Wu, Zhe Wu, Ruichi Yu, and Larry S. Davis. Viton: An image-based virtual try-on network, 2018. 2

[8] Edward J Hu, yelong shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. LoRA: Low-rank adaptation of large language models. In *International Conference on Learning Representations*, 2022. 3

[9] Jeongho Kim, Gyojung Gu, Minho Park, Sunghyun Park, and Jaegul Choo. Stableviton: Learning semantic correspondence with latent diffusion model for virtual try-on, 2023. 2

[10] Jeongho Kim, Hoiyeong Jin, Sunghyun Park, and Jaegul Choo. Promptdresser: Improving the quality and controllability of virtual try-on via generative textual prompt and prompt-aware mask, 2024. 2

[11] Ivan Lopes, Fabio Pizzati, and Raoul de Charette. Material palette: Extraction of materials from a single image. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4379–4388, 2024. 3

[12] Federico Martusciello, Henry Muccini, and Antonio Bucchiarone. A reference architecture for gamified cultural heritage applications leveraging generative ai and augmented reality, 2025. 2

[13] Tianhe Ren, Shilong Liu, Ailing Zeng, et al. Grounded sam: Assembling open-world models for diverse visual tasks, 2024. 3

[14] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 10684–10695, 2022. 3

[15] Nataniel Ruiz, Yuanzhen Li, Varun Jampani, Yael Pritch, Michael Rubinstein, and Kfir Aberman. Dreambooth: Fine tuning text-to-image diffusion models for subject-driven generation, 2023. 3

[16] Mingxin Sun and Ahmad Nizam Othman. The impact of virtual reality technology on museum audience engagement. *International Journal of Academic Research in Progressive Education and Development*, 14, 2025. 2

[17] Techpacker. The definitive guide to dress design. Technical report, Techpacker, 2024. 2

[18] Denis Tome, Patrick Peluse, Lourdes Agapito, and Hernan Badino. xr-egopose: Egocentric 3d human pose from an hmd camera. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2019. 2

[19] Unity Technologies. Unity, 2025. Game development platform. 2

[20] Bochao Wang, Huabin Zheng, Xiaodan Liang, Yimin Chen, Liang Lin, and Meng Yang. Toward characteristic-preserving image-based virtual try-on network, 2018. 2

[21] Luyang Zhu, Dawei Yang, Tyler Zhu, Fitsum Reda, William Chan, Chitwan Saharia, Mohammad Norouzi, and Ira Kemelmacher-Shlizerman. Tryondiffusion: A tale of two unets, 2023. 2

[22] Filip Škola, Selma Rizvić, Marco Cozza, Loris Barbieri, Fabio Bruno, Dimitrios Skarlatos, and Fotis Liarokapis. Virtual reality with 360-video storytelling in cultural heritage: Study of presence, engagement, and immersion. *Sensors*, 20 (20), 2020. 2