GenWardrobe: A Fully Generative System for Travel Fashion Wardrobe Construction

Peng Jin
Donghua University
nlxxnlgz1@gmail.com

Yilin Wen
Singapore Management University
yilin.wen.2023@mitb.smu.edu.sg

Mingzhe Yu
Singapore Management University
mzyu1220@gmail.com

Yunshan Ma*
Singapore Management University
ysma@smu.edu.sg

Rong Zheng
Donghua University
rzheng@dhu.edu.cn

Jintu Fan
The Hong Kong Polytechnic University
jin-tu.fan@polyu.edu.hk

Chong Wah NGO
Singapore Management University
cwngo@smu.edu.sg

Abstract

With the increasing demand for outfit planning in real-world travel scenarios, the need for constructing a travel fashion wardrobe, a series of outfits tailored to a user's personalization and destination-specific context over a short travel period, has grown significantly. However, existing systems or works often focus on isolated factors and rely on retrieval-based methods, with insufficient utilization of generative models, limiting their adaptability to real-world travel scenarios. To address this issue, this study introduces GenWardrobe, a fully generative system for travel fashion wardrobe construction. GenWardrobe consists of three key modules: user query analysis, fashion knowledge retrieval via retrieval-augmented generation and wardrobe image generation. To facilitate users' usage, we encapsulate the solution into an interactive web application. Expert-level evaluation shows that GenWardrobe significantly outperforms traditional systems in both personalization and visual appeal. PowerPoint file and more materials of Genwordrobe can be found on our Github repository: https://github.com/ShanFengShanFeng/GenWardrobe.

CCS Concepts

• **Information systems** → **Multimedia and multimodal retrieval**; • **Computing methodologies** → **Image generation**.

Keywords

Fashion Wardrobe Construction; Image Generation; Multimodality

1 Introduction

In daily life, people often travel for various purposes such as vacations, business trips, or attending social events. During these trips, they need to pack a coherent and suitable set of fashion items, collectively referred to as a travel fashion wardrobe. This travel fashion wardrobe comprises the outfits worn across different travel days and occasions, typically under constraints like limited luggage space, diverse activities, and unfamiliar climates. Constructing such a wardrobe is a non-trivial task that requires simultaneously accounting for three key characteristics: **human** (body shape, skin tone, gender, age, demographic characteristics, etc), **complex context constrain** (destination, weather, time of day, and purpose, etc), and **fashion knowledge** (fashionability, popularity, mix-and-match, etc). Given the increasing consumer demand in both the travel and fashion industries—each valued at hundreds of billions of dollars globally—providing intelligent solutions for travel wardrobe construction carries substantial practical value [1].

However, existing systems or works only focus on one specific or a part of the requirements. Prior works typically focuses on isolated aspects such as personalized outfit generation or occasion-aware recommendation, without offering a comprehensive framework that holistically integrates human, complex context constrain, and fashion knowledge[3]. Moreover, most systems rely heavily on conventional retrieval-based pipelines or heuristic rules[2], which limit the ability to generate diverse and novel outfits tailored to real-world scenarios. Meanwhile, recent advances in generative AI—including large language models (LLMs), retrieval-augmented generation (RAG), and AI-generated content (AIGC)—have shown promise in producing high-quality and creative outputs, remaining underutilized in the fashion wardrobe construction.

To bridge the research gaps, we propose a fully generative system for travel fashion wardrobe construction, named as GenWardrobe. Specifically, this system consists of three key modules: (1) **user query analysis**, which parses both visual (user full-body photo) and textual (travel plan description) raw inputs and extract the structured representations; (2) **fashion knowledge retrieval via RAG**, which constructs a structured fashion knowledge base via LLM-based image filtering and extraction, and performs context-aware fashion knowledge retrieval through an LLM-based retrieval framework, i.e., Llama-Index; and (3) **wardrobe image generation**, which generates the fashion wardrobe images and the visualized fashion wardrobe with realistic background. The system takes a user's full-body photo and textual travel plan as input, and outputs both a pure fashion wardrobe and visualized fashion wardrobe aligned with the user input, offering richer visual incentives, enhanced user experience, and increased user satisfaction.

2 System Design

The system includes three key modules, as shown in Figure 1.

*Corresponding author.

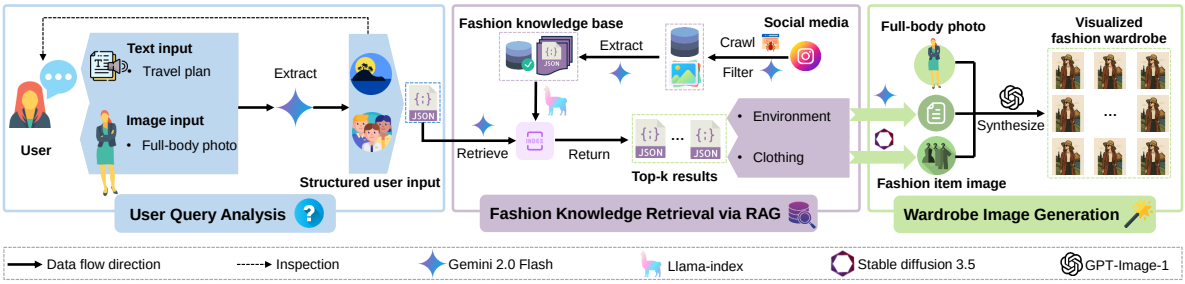

Figure 1: Illustration of the overall system design.

User Query Analysis. The initial stage of this system is from raw user input, which includes an image input of a full-body photo of the user and a text input which describes the user's travel plans. Then, we leverage Gemini to process the raw user input (both text and image input) to obtain the structured user input, which covers essential user information (gender, body shape, skin tone, etc) and travel plan (travel destination, time, weather, and occasion, etc).

Fashion Knowledge Retrieval via RAG. Given the structured user query, we leverage an RAG framework to obtain a set of fashion knowledge from a well-curated large-scale fashion knowledge base. Specifically, we first crawl fashion images from social media platforms, then filter out low-quality images using Gemini, and extract structured fashion knowledge from the high-quality images, finally, we obtain a large fashion knowledge base with approximately one million pieces of fashion knowledge, each associated with one high-quality social media image. For detailed fashion knowledge definitions, please refer to the GitHub README file. Next, we employ the popular Llama-Index framework to query the fashion knowledge base with the structured user query, resulting in top-k fashion knowledge that include both environment and clothing information.

Wardrobe Image Generation. Guided by the retrieved fashion knowledge, we employ the cutting edge text-to-image generation models to generate the wardrobe images. Concretely, we use the Stable Diffusion model to generate the fashion item images, tailored to the clothing information of the retrieved top-k knowledge. The fashion item images offer a preliminary results of the fashion wardrobe. To further enhance user's experience, we put on the fashion items to the given full-body photo of the user, as well as generating a photorealistic background based on the environment information of the retrieved knowledge. These two processes are implemented using the GPT-image-1 model. Ultimately, we provide both pure fashion wardrobe and visualized fashion wardrobe.

3 System Implementation and Demonstration

Following the aforementioned design, we implement the the system with a web application with the frontend-backend framework. Specifically, we use Vue.js to develop the frontend which takes user input and displays the generated results. For the backend implementation, we use Flask and Python, and we run the RAG module in our own server, while all the LLM calls use proprietary APIs. Details such as processing time can be found in the Github repository.

To showcase the end-to-end workflow of the system, we built an interactive demonstration, where the raw user input consists of a full-body photo and a travel plan text. The demonstration outputs

Figure 2: Illustration of the interface of the demo system.

both a pure fashion wardrobe and visualized fashion wardrobe. An example of a user query and its corresponding generated fashion wardrobe is illustrated in Figure 2. We also conduct an expert-level evaluation. We recruit six professional fashion reviewers to perform Academicism Aesthetic Test, and the evaluation results justify that our system significantly outperforms existing systems. More details of the evaluation are available in the GitHub README file.

4 Conclusion and Future Work

This study proposes GenWardrobe, a fully generative system for travel fashion wardrobe construction that seamlessly integrates user intent analysis, a RAG framework, and image synthesis. In the future, this system can be extended to a broader range of application scenarios, by integrating with AR/VR technologies, it holds the potential to further enhance user experience and unlock greater commercial value.

References

[1] Yujuan Ding, Yunshan Ma, Wenqi Fan, Yige Yao, Tat-Seng Chua, and Qing Li. 2024. Fashionregen: Llm-empowered fashion report generation. In *Companion Proceedings of the ACM Web Conference 2024*. 991–994.

[2] Yujuan Ding, PY Mok, Yunshan Ma, and Yi Bin. 2023. Personalized fashion outfit generation with user coordination preference learning. *Information Processing & Management* 60, 5 (2023), 103434.

[3] Yiyan Xu, Wenjie Wang, Fuli Feng, Yunshan Ma, Jizhi Zhang, and Xiangnan He. 2024. Diffusion models for generative outfit recommendation. In *Proceedings of the 47th international ACM SIGIR conference on research and development in information retrieval*. 1350–1359.