FashionReGen: LLM-Empowered Fashion Report Generation

Yujuan Ding
dingyujuan385@gmail.com
The Hong Kong Polytechnic
University, HK SAR

Yunshan Ma
yunshan.ma@u.nus.edu
National University of Singapore,
Singapore

Wenqi Fan*
wenqifan03@gmail.com
The Hong Kong Polytechnic
University, HK SAR

Yige Yao
yi-ge.yao@connect.polyu.hk
The Hong Kong Polytechnic
University, HK SAR

Tat-Seng Chua
dcscts@nus.edu.sg
National University of Singapore,
Singapore

Qing Li
csqli@comp.polyu.edu.hk
The Hong Kong Polytechnic
University, HK SAR

ABSTRACT

Fashion analysis refers to the process of examining and evaluating trends, styles, and elements within the fashion industry to understand and interpret its current state, generating fashion reports. It is traditionally performed by fashion professionals based on their expertise and experience, which requires high labour cost and may also produce biased results for relying heavily on a small group of people. In this paper, to tackle the **Fashion Report Generation** (**FashionReGen**) task, we propose an intelligent **F**ashion **A**nalyzing and **R**eporting system based the advanced Large Language Models (LLMs), debbed as **GPT-FAR**. Specifically, it tries to deliver FashionReGen based on effective catwalk analysis, which is equipped with several key procedures, namely, catwalk understanding, collective organization and analysis, and report generation. By posing and exploring such an open-ended, complex and domain-specific task of FashionReGen, it is able to test the general capability of LLMs in fashion domain. It also inspires the explorations of more high-level tasks with industrial significance in other domains. Video illustration and more materials of GPT-FAR can be found in https://github.com/CompFashion/FashionReGen.

CCS CONCEPTS

• **Information systems → Multimedia content creation**.

KEYWORDS

Fashion Report Generation, Large Language Model, Multimodal Understanding and Generation, GPT

*Corresponding author: Wenqi Fan, Department of Computing, and Department of Management and Marketing, The Hong Kong Polytechnic University.

1 INTRODUCTION

The fashion industry is a vital component of global economy, characterized by its constant pursuit of novelty and changes. It is essential for fashion practitioners, enthusiasts and consumers to capture these shifts to get ahead of the right trends and make confident decisions [2–5]. To this end, specialized consulting and forecasting companies conduct seasonal or annual analysis, producing insightful fashion reports[1]. However, such **manual** production usually relies on a small group of experts to review, comprehend, summarize, create visual aids such as charts, analyze and write the reports. The manual process involves high labor and intellectual cost.

Developing an advanced approach to automate parts of or even the entire fashion report generation process is of great value. We term the task as **Fashion Report Generation** (**FashionReGen**), which has been barely explored because of the challenges in different stages. First, it requires fine-grained recognition of categories, attribute details, and even fashion-specialized concepts for high-level comprehension of fashion content. Second, domain-specific knowledge and experiences are required to perform professional statistical analysis. Finally, proper selection and combination of multi-modal information are needed to integrate text, charts and images into a coherent, complete and insightful report. Even though previous deep learning techniques have achieved enormous progress in image recognition and text generation, they are mostly developed for the general domain while far from achieving expert-level performance in specific domains such as fashion. Fortunately, the breakthrough of Large Language Models (LLMs) [1] shed light on this challenging problem, owing to the superb capability in multi-modal content comprehension and generation.

To this end, we propose an intelligent **F**ashion **A**nalyzing and **R**eporting (**FAR**) system empowered by GPT models (named as **GPT-FAR**) for FashionReGen. It is an intelligent, also first-of-its-kind pipeline that enables automatic fashion analysis and report generation. Our current FashionReGen is based on catwalk analysis since catwalk is the place where high brands present their new designs. It tends to lead the trends and holds great impact on downstream fashion industry. The proposed GPT-FAR includes the

---

[1]www.wgsn.com; https://www.mckinsey.com/industries/retail/our-insights/state-of-fashion

1. Catwalk Understanding

Garment Tagging | Tag Cleaning

2. Collective Organization and Analysis

Collective Category- and Attribute-Level Organization | Statistical Metrics: Mix, YoY, Trend

3. Fashion Report Generation

Chart Generation | Textual Analysis Generation | Image Retrieval

Input: Catwalk Outfit Images

Output: Fashion Analysis Report

Figure 1: Diagram of our GPT-FAR system.

processes of catwalk understanding, collective organization and analysis, and multi-modal fashion report generation. To facilitate the end-to-end FashionReGen based on catwalk observation, we first design a GPT-4V-based tagger to interpret the garments in the catwalk images, followed by a two-stage tag cleaner. Then, we employ professional metrics and charts to perform collective analysis. Finally, we develop an algorithm based on GPT-4V to generate textual analysis for the report with in-context learning based on statistical charts produced previously. Our contributions are:

- We present FashionReGen, a high-level domain-specific task with significant research and application value.
- We propose GPT-FAR, a system for automatic fashion report generation based on effective catwalk observation and analysis.
- We develop a complete platform to enable users to generate their own fashion reports on specific collections.

2 GPT-FAR SYSTEM

Our GPT-FAR consists of three major stages: 1) catwalk understanding, 2) collective analysis, and 3) fashion report generation.

2.1 Catwalk Understanding

2.1.1 Garment tagging with GPT-4V. We device an advanced **garment tagger** based on GPT-4V given its superior capability in image understanding [6]. For each catwalk image, the garment tagger first classifies each garment into one of the following categories: *dresses, skirts, jackets, coats, trousers, shorts, knit and jersey, sweater, top, blouses and woven tops*. The follow-up attribute-level tagging is strictly based on the methodology employed by WGSN[2]. Specifically, the tagger is required to generate keywords to describe each garment for its *style, silhouette, neckline, length, print and pattern, detail, embellishment* and *fabric* if applicable. Such a tagging procedure allows to generate systematic and comprehensive descriptive tags for each garment, which is achieved with the following prompt:

PROMPT 2.1. *Can you tag the outfit in the image at garment level, which including two main steps. 1. recognize garments in the image and label them with categories from the category set of #CATEGORY LIST and 2. for each garment, tag it from the the following aspects: #ASPECT LIST. You have to list as MANY tags as possible suitable for each aspect, not only one. Report the tagging results for each garment with a single line pattern, following the corresponding category. Do not*

---

output anything other than the category and tags for the mentioned aspects. Here is an example for your tagging, <image: {Category: Top; Style: Layered, Modern, ... Silhouette: Relaxed, ...; ...}, {Category: Skirt; Style: Casual, Street, ...; ... }>.

2.1.2 GPT-produced tag cleaning. Based on our experimental observation, the garment tagger is able to generate accurate, rich, and professional tags for target aspects. However, since the task is open-ended, it is difficult to strictly restrict the format of the output, even when we have emphasized the output format or shown examples in the prompt. A more tricky challenge is that it may generate different words meaning almost exactly the same attribute since the corpus is not and cannot be pre-defined. This is also a general issue of GPT-generated content, which is not friendly for subsequent processing, particularly statistical analysis.

To address this issue, we design a two-stage tag cleaning strategy for our GPT-produced tags, including **manual format unification** and **automatic synonym emergence** (shown in Figure 2). For format unification, we set some rules to address the inconsistency issue in terms of capitalization, plurality, and other formats such as extra space. Synonym emergence is designed as an iterative grouping process on the whole tag corpus of specific aspect based on GPT-4. As shown in the bottom left in Figure 2, we apply a GPT agent to group synonym tags together with the following prompt:

PROMPT 2.2. *You are an assistant to help group tagging corpus, specifically for the corpus describing #attribute for #CATEGORY. You are required to group similar words together and output a dictionary that map each word to the group? The corpus contains words as follows #TAGS CORPUS. Make sure only words with almost the same meaning are grouped, NOT those describing the same aspect at a larger scale. Here is an example: 'draped': ['draped', 'draping', 'draped front', 'draped neckline', 'draped panel', 'draped shoulders', 'draped overlay', 'draped look']. Only output the word groups as a dictionary, DO NOT output other descriptive or explanatory text!!*

We would expect that the guided agent would produce a perfect tag grouping dictionary. However, testing results suggest that so far it is challenging to achieve a one-step completion, especially when the tag corpus is relevantly large (with hundreds of identical tags). Therefore, we further design an iterative grouping process, which repeats the grouping operation for several times to ensure all tags with synonyms get well grouped. At each iteration, the GPT agent is required to group non-grouped tags into existing groups or create new groups with the following prompt.
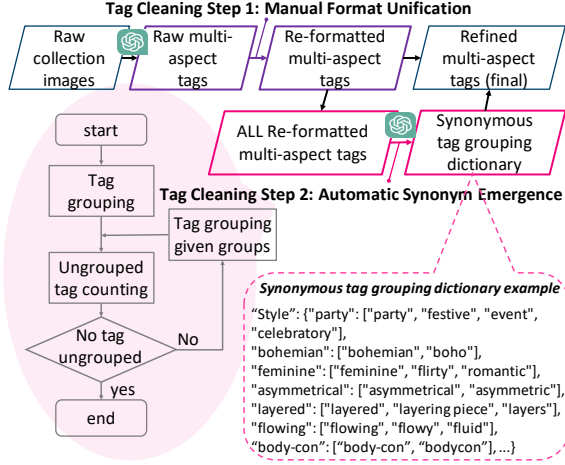
Figure 2: Data flow of the catwalk understanding process.

PROMPT 2.3. *... Candidate word groups include #TAG GROUPS. Check whether each word can be grouped into an existing group. If not, you can create new groups ...*

Note that this process groups only synonyms or different formats of words. We appreciate that GPT possesses expansive vocabulary to describe garment in detailed and professional words.

2.2 Collective Organization and Analysis

Our fashion analysis is at collection-level, which requires the organization of garment-level tagging results to obtain collective data. In specific, each garment i is tagged by its category $c^i \in C$ and the attribute details attached to the category $\{a^i\}_{a \in \mathcal{A}_a}$. The category set is $C = \{c_1, c_2, ..., c_{|C|}\}$. The attribute set for the specific category c is $\mathcal{A}^c = \{A_1^c, A_2^c, ..., A_{|\mathcal{A}^c|}^c\}$. Each attribute contains different values, for example, $A_1^c = \{a_{11}^c, a_{12}^c, ...\}$, where a_{11}^c denotes the specific attribute value of A_1^c. If the category *dresses*, one corresponding attribute could be *silhouette*, then under this attribute, the attribute value could be *fit-and-flare*, *layered*, etc.

Given a collection defined by its season (t) and brand (b) $\mathcal{S}_{t,b} = \{i_1, i_2, ..., i_{|\mathcal{S}_{t,b}|}\}$ (we may omit the subscripts for simplicity), the category-level index can be summarized as $\mathcal{D}_{\mathcal{S}}^C = \{d_{\mathcal{S}}^{c_1}, d_{\mathcal{S}}^{c_2}, ..., d_{\mathcal{S}}^{c_{|C|}}\}$, where $d_{\mathcal{S}}^{c_1}$ denotes the index of $\mathcal{S}$ of category c_1 and so on for others. Similarly, attribute-level index of collection $\mathcal{S}$ can be obtained. Taking the category c as an example, the attribute index set is $\mathcal{A}^c$ and its index set is $\mathcal{D}_{\mathcal{S}}^{\mathcal{A}^c} = \{\mathcal{D}_{\mathcal{S}}^{A_1^c}, \mathcal{D}_{\mathcal{S}}^{A_2^c}, ...\}$. Each attribute-level index set can be summarized as $\mathcal{D}^{A^c} = \{d^{a_1^c}, d^{a_2^c}, ...\}$.

We employ three key metrics for collective analysis, which are the mix (Mix), Year-on-Year index (YoY), and the list of evolving trend T. The category-level metrics (for category c) are as follows:

$$Mix_{\mathcal{S}_{t,b}}^c = \frac{d_{\mathcal{S}_{t,b}}^c}{\sum \mathcal{D}_{\mathcal{S}_{t,b}}^c} \times 100\%, \tag{1}$$

$$YoY_{\mathcal{S}_{t,b}}^c = \frac{Mix_{\mathcal{S}_{t,b}}^c - Mix_{\mathcal{S}_{t-1,b}}^c}{Mix_{\mathcal{S}_{t-1,b}}^c}. \tag{2}$$

Table 1: Catwalk Dataset

	Current Version	Expandable Version
Time Span	2019-2023 (5 years)	More years in the past or to come
Brand	Chanel, Saint Laurent, LV, Valentino, Dior, Givenchy	More high fashion brands leading trends
Seasons	Spring/Summer (SS)	Autume/Winter (AW), Pre-summer, Pre-fall

By replacing the category index $d_{\mathcal{S}}^{c,t}$ into attribute value index $d_{\mathcal{S}_{t,b}}^{a,t}$ and applying the corresponding attribute index set $\mathcal{D}_{\mathcal{S}_{t,b}}^A$, we can obtain the *Mix* of attribute value $Mix_{\mathcal{S}_{t,b}}^a$ as well as the Year-on-Year index $YoY_{\mathcal{S}_{t,b}}^a$. In addition, combining yearly data of *Mix* or *YoY* can obtain corresponding evolving trend list T.

2.3 Fashion Report Generation

In order to make our fashion report instructive and insightful, the report to generate will be covering three main parts: charts for statistical analysis, catwalk images for illustration, and textural demonstration for state summary and highlights. In the current version, we simply plot charts based on the collective analytic data from previous stage and retrieve images relevant to the analyzed collection. For the textural analysis generation, we survey professional fashion reports and propose the chart-based generation method. The idea is to uncover the notable changeability or trends presented by charts, hopefully to produce meaningful insights. For overall textual analysis, we first design a GPT-4V-based agent with the following prompt:

PROMPT 2.4. *You are given several charts describing the fashion status specifically for #REPORT TYPE of #COLLECTION and SEASON. Each chart is about one specific aspect, e.g., fabric, silhouette. Try to generate several paragraphs (less than FIVE) in the format of an article. You are also given several examples of textual analysis based on charts as follows: #EXAMPLEs. The length of the article should be around 250 characters. Do not use any key points or subtitles.*

We further present focused analysis in a short description for specific categories covered in the report with the following prompt:

PROMPT 2.5. *You are given several charts ... Try to generate a very short and neat piece of description (MUST less than two sentences) that can give an overview of the category or highlight the most significant trend. Please DO NOT make it too specific on specific aspects. You are also given several examples ... Please try to get the tone and style of the descriptions from the examples and apply then in your generation.*

3 RESULTS AND ANALYSIS: A CASE STUDY

To evaluate the effectiveness of the proposed GPT-FAR system to deliver automatic FashionReGen, we collect catwalk images and review the generated reports based on them. Considering the computational and economic cost, the preliminary data is small amount as shown in Table. 1. The system page of GPT-FAR, as shown in Figure 3, provides an input panel to select the analysis targets by year, season, category group and brand(s). Reports can be automatically generated by clicking the *generate* button.
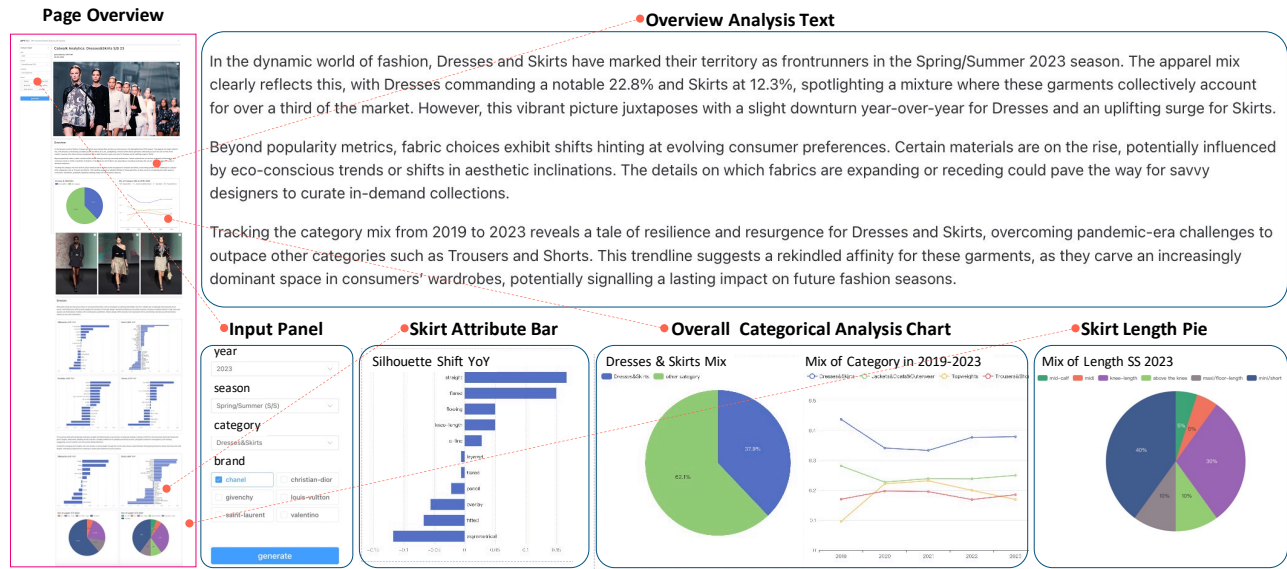
Figure 3: Illustration of the Fashion Analyzing and Reporting system and an example of the generated report.

3.1 Report Type and Structure

Our report performs catwalk analytics on specific group of categories, such as *Dresses & Skirts*, *Topweights*, *Trousers & Shorts*, *Jackets & Coats*. As a generated case shown in Figure 3, it is a multi-page report with a pre-defined layout. It starts with a cover page with the basic information, i.e., title, author and generating time of the report. The overall analysis page is next, which has a three-column layout with textual analysis in the left, charts lying vertically in the middle, and image illustrations in the right. The rest of the report is category-specific analysis.

3.2 Generation Results and Discussion

We have the following observation and discussion on the report generation and the middle-stage results of GPT-FAR: 1) GPT-4V enables to generate high-quality descriptive tags for garments owing to its visual understanding capacity. However, it also remains challenging to output only concise and clean tags, which makes the tag cleaning strategy critical for the further analytics on them; 2) The generated textual demonstration, especially for the overall analysis, is reasonable, fluent and professional; and 3) Overall, GPT-FAR generates high-quality fashion reports, which are comprehensive, illustrative, and in a hybrid modality of presentation. It offers a platform for automatic fashion analysis and report generation.

4 CONCLUSION

The paper posed a meaningful domain-specific task: fashion report generation (FashionReGen) relying on data observation and analysis. Empowered by LLM advances, a novel Fashion Analyzing and Reporting framework (GPT-FAR) was proposed, which enables automatic FashionReGen with the joint efforts of GPT-series models and domain-specific knowledge. A report generation system was further developed based on GPT-FAR, offering the platform for public users to generate their own fashion reports.

This work is still a preliminary demo, which has huge potential for further enhancement. In future, we will include more sources of data, such as social media, forum and business data, in order to offer more perspectives and opinions in the analysis. Meanwhile, the type of fashion reports can be expanded. We will also work on specific technical challenges to offer higher degree of automation and intelligence . The evaluation of FashionReGen is another important direction to study in the future.

ACKNOWLEDGMENTS

REFERENCES

[1] Liying Cheng, Xingxuan Li, and Lidong Bing. 2023. Is GPT-4 a Good Data Analyst?. In *EMNLP (Findings)*. ACL, 9496–9514.
[2] Yujuan Ding, Zhihui Lai, PY Mok, and Tat-Seng Chua. 2023. Computational Technologies for Fashion Recommendation: A Survey. *Comput. Surveys* 56, 5 (2023), 1–45.
[3] Yujuan Ding, Yunshan Ma, Lizi Liao, Wai Keung Wong, and Tat-Seng Chua. 2022. Leveraging Multiple Relations for Fashion Trend Forecasting Based on Social Media. *IEEE Trans. Multim.* 24 (2022), 2287–2299.
[4] Yujuan Ding, P. Y. Mok, Yunshan Ma, and Yi Bin. 2023. Personalized fashion outfit generation with user coordination preference learning. *Inf. Process. Manag.* 60, 5 (2023), 103434.
[5] Yunshan Ma, Yujuan Ding, Xun Yang, Lizi Liao, Wai Keung Wong, and Tat-Seng Chua. 2020. Knowledge Enhanced Neural Fashion Trend Forecasting. In *ICMR*. ACM, 82–90.
[6] Zhengyuan Yang, Linjie Li, Kevin Lin, Jianfeng Wang, Chung-Ching Lin, Zicheng Liu, and Lijuan Wang. 2023. The Dawn of LMMs: Preliminary Explorations with GPT-4V(ision). *CoRR* abs/2309.17421 (2023).