A Review of Modern Fashion Recommender Systems

YASHAR DELDJOO and FATEMEH NAZARY, Polytechnic University of Bari, Italy
ARNAU RAMISA, Amazon, USA
JULIAN MCAULEY, UC San Diego, USA
GIOVANNI PELLEGRINI, Polytechnic University of Bari, Italy
ALEJANDRO BELLOGIN, Autonomous University of Madrid, Spain
TOMMASO DI NOIA, Polytechnic University of Bari, Italy

The textile and apparel industries have grown tremendously over the past few years. Customers no longer have to visit many stores, stand in long queues, or try on garments in dressing rooms, as millions of products are now available in online catalogs. However, given the plethora of options available, an effective recommendation system is necessary to properly sort, order, and communicate relevant product material or information to users. Effective fashion recommender systems (RSs) can have a noticeable impact on billions of customers' shopping experiences and increase sales and revenues on the provider side.

The goal of this survey is to provide a review of RSs that operate in the specific vertical domain of garment and fashion products. We have identified the most pressing challenges in fashion RS research and created a taxonomy that categorizes the literature according to the objective they are trying to accomplish (e.g., item or outfit recommendation, size recommendation, and explainability, among others) and type of side information (users, items, context). We have also identified the most important evaluation goals and perspectives (outfit generation, outfit recommendation, pairing recommendation, and fill-in-the-blank outfit compatibility prediction) and the most commonly used datasets and evaluation metrics.

CCS Concepts: • **Information systems** → **Recommender systems**; **Multimedia and multimodal retrieval**; • **Human-centered computing** → **User models;**

Additional Key Words and Phrases: Recommender systems, information retrieval, fashion retail, machine learning, artificial intelligence, computer vision, text mining, e-commerce

ACM Reference format:
Yashar Deldjoo, Fatemeh Nazary, Arnau Ramisa, Julian McAuley, Giovanni Pellegrini, Alejandro Bellogin, and Tommaso Di Noia. 2023. A Review of Modern Fashion Recommender Systems. *ACM Comput. Surv.* 56, 4, Article 87 (October 2023), 37 pages.
https://doi.org/10.1145/3624733

Authors' addresses: Y. Deldjoo (corresponding author), F. Nazary, G. Pellegrini, and T. Di Noia, Polytechnic University of Bari, Bari, Italy; e-mails: deldjooy@acm.org, {fatemeh.nazary, tommaso.dinoia}@poliba.it, g.pellegrini4@studenti.poliba.it; A. Ramisa, Amazon; e-mail: aramisay@amazon.com; J. McAuley, UC San Diego; e-mail: jmcauley@eng.ucsd.edu; A. Bellogin, Autonomous University of Madrid, Madrid, Spain; e-mail: alejandro.bellogin@uam.es.

0360-0300/2023/10-ART87 $15.00
https://doi.org/10.1145/3624733

1 INTRODUCTION

Recommender Systems (RSs) help users navigate large collections of products to find items relevant to their interests, leveraging large amounts of product information and user signals like product views followed or ignored items, purchases, or webpage visits to determine how, when, and what to recommend to their customers. RSs have grown to be an essential part of all large Internet retailers, driving up to 35% of Amazon sales [117] or more than 80% of the content watched on Netflix [32].

In this work, we are interested in RSs that operate in one particular vertical market: garments and fashion products. This setting introduces a particular set of challenges and subproblems that are relevant for developing effective RSs.

Due to market dynamics and customer preferences, there is a large vocabulary of distinct fashion products, as well as high turnover. This leads to sparse purchase data, which challenges the usage of traditional RSs [95]. Furthermore, precise and detailed product information is often not available, making it difficult to establish similarities between them.

To deal with the aforementioned problems, and given the visual and aesthetic nature of fashion products, there is a growing body of computer vision research addressing tasks like localizing fashion items [61, 179], determining their category and attributes [25, 54, 169], or establishing the degree of similarity to other products [6, 57, 111, 113], to name only a few. Although works in the computer vision literature often do not consider personalization (or recommendation), their predictions and embeddings can be leveraged by RSs and combined with past user preferences, thus mitigating sparsity and cold start problems. Another relevant fashion problem that has attracted the attention of computer vision research is that of combining garments into complete outfits [18, 27, 59, 107]. Several works have studied how to learn the compatibility between fashion items using both professional photos of models wearing designer-created outfits and social media pictures from 'influencers' and normal people [74, 148, 163]. In addition to allowing recommendations tailored to match the existing shopping basket or wardrobe of customers, these datasets help uncover other insights useful for RSs, such as the structure of fashion styles [73], social group preferences [93], or the evolution of trends across time and location [118, 119].

In addition to product-to-product relationships, fashion RSs also face particular product-to-user uncertainties, like *fit*, that can hurt the quality of recommendations if not taken into account. Fit prediction is a pain point for online fashion shopping according to customers, as well as the primary reason for product returns faced by online retailers [35]. The many sizing systems in use throughout the world, as well as their interpretation by different clothing manufacturers, make it very difficult to predict whether a particular product will properly fit a customer. Therefore, research on how to estimate a personalized garment fit has leveraged sources of information like co-purchase data [145], customer-reported measurements [128], or advanced imaging devices such as 3D scanners [12, 13, 37]. Needless to say, this information can be very valuable for recommendation.

Several online retailers, like Stitch-Fix and Amazon Prime Wardrobe, recreate the *try-out* experience of brick-and-mortar stores by shipping a highly tailored assortment of garments to a customer, who can proceed to try them on in their home and return those that they do not like. RSs can make this approach more sustainable by maximizing the number of items that customers keep [197], but the cold start problem and the commitment required from customers to stay in the membership plan make this approach hard to scale. Consequently, alternatives that would allow a customer to visualize the appearance and fit of clothes in an augmented reality or virtual environment are being researched [125].

The reverse approach, namely designing or automatically generating a computer model of a garment with the right appearance and fit for a customer and subsequently manufacturing it, is

another route being explored by, for example, Amazon with the *Made for You Custom T-shirt*, with the aim of bringing the made-to-measure fashion to the masses.[1]

1.1 Major Challenges

In this section, we will describe the major challenges faced by RSs in the fashion domain:

***Fashion item representation*:** Traditional RSs such as **Collaborative Filtering (CF)** and **Content-Based Filtering (CBF)** have difficulties in the fashion domain due to the sparsity of purchase data or the insufficient detail about the visual appearance of the product in category names [144]. Instead, more recent literature has leveraged models that capture a rich representation of fashion items through product images [62, 81], text descriptions or customer reviews [29, 192], or videos [181], which are often learned through surrogate tasks like classification or product retrieval. However, learning product representations from such input data requires large datasets to generalize well across different image (or text) styles, attribute variations, and so forth. Furthermore, constructing a representation that learns which product features customers take most into account when evaluating fashion products is still an open research problem.

***Fashion item compatibility*:** Training a model that is able to predict if two fashion items 'go together' or directly combine several products into an outfit is a challenging task. Different item compatibility signals studied in recent literature include co-purchase data [120, 162], outfits composed by professional fashion designers [59], or combinations found by analyzing what people wear in social media pictures [83, 153]. From this compatibility information, associated image and text data are then used to learn to generalize to stylistically similar products. Some works explicitly model the latent style types [20]. An additional under-explored difficulty for compatibility prediction is the dependency on trends, seasonality, location, or social group. Current approaches usually leverage image and text information.

***Personalization and fit*:** The best fashion product to recommend depends on factors such as the location where the outfit will be used [7, 24, 92, 166], the season or occasion [112, 118, 190], or the cultural and social background of the customer [93, 153, 193]. A challenging task in fashion recommendation systems is how to discover and integrate these disparate factors [148, 173]. Current research often tackles these tasks by utilizing large-scale social media data.

As discussed earlier, a personalization dimension very particular to the fashion domain is that of fit. In addition to predicting what size of a product will be more comfortable to wear, body shape can influence stylistic choices [69, 72, 143].

***Interpretability and explanation*:** Most of the existing fashion RSs in the literature focus on improving predictive performance, treating the model as a black box. However, deploying accountable and interpretable systems able to explain their recommendations can foster user loyalty in the long term and improve the shopping experience. Current models generally offer explanations through highlighted image regions and attributes or keywords [58, 70, 103, 126, 176, 183].

***Discovering trends*:** Being able to forecast consumer preferences is valuable for fashion designers and retailers to optimize product-to-market fit, logistics, and advertising. Many factors are confounded in what features are considered 'fashionable' or 'trendy,' like seasonality [118], geographical influence [7], historical events [75], or style dynamics [8, 55, 114]. Again, social media is a useful resource leveraged by researchers [51, 77].

[1] https://www.amazon.com/Made-for-You-Custom-T-shirt/dp/B08N6J8G5M

The preceding challenges, as well as many other issues, have been discussed in the studied research works, and they are reviewed in the following sections. For instance, visual modeling of fashion items (cf. Section 2.2.2), modeling of fashion outfits (cf. Section 2.1.2), geo-temporal localization (cf. Section 2.2.3), attribute prediction for fashion and leveraging multi-modal data (cf. Section 2.2.2), explainability in fashion recommendation (cf. Section 2.1.4), and learning style (cf. Section 2.2.2).

1.2 Search Strategy for Relevant Papers

We primarily relied on papers indexed in DBLP,[2] a prominent computer science bibliographic database, to identify publications that comprise the state-of-the-art in fashion item and outfit recommendation. Our search approach was divided into two stages: finding relevant publication collections, and post-processing and filtering the final list. The total number of articles analyzed exceeds 50, with the majority of recognized research papers published between 2014 and 2023, an indication of the topic's originality and freshness. We do not claim that this study is exhaustive in terms of the collected papers; however, we believe it gives a comprehensive overview of current achievements, trends, and what we deem to be the most significant challenges and tasks in fashion RSs. Additionally, it gives valuable information to both industry practitioners and academics.[3]

1.3 Related Surveys

To put this survey in context, we identified and present related review and survey articles to explain in which ways our article differs from and extends earlier work. In a recent work, Deldjoo et al. [45] presented a survey of RS leveraging multimedia content (i.e., visual, audio, and/or textual features). The domains studied in this survey include various ones such as media streaming for audio and video recommendation, e-commerce for recommending different products including fashion items, news, and information recommendation, social media, and so forth. Although fashion RS were also discussed, the authors only included a small portion of the topics and papers in this domain. Here, we discuss and present a comprehensive survey of significant tasks, challenges, and types of content used in the fashion RS field.

We have also identified surveys [30, 198] where the authors present a literature review of techniques at the intersection of fashion and computer vision and/or natural language processing. Although we find these works relevant to this article, they remain largely different from the review presented here, as those systems are not focused on RS but on other aspects of the fashion domain, such as text generation from images or pose estimation.

Perhaps the most relevant work to our current survey is a recent book chapter by Jaradat et al. [86] on fashion RS. This chapter focuses on discussing the state of the art of fashion recommendation systems; in particular, the authors affirm that deep learning represented a turning point with respect to the canonical approaches and therefore they examined four different tasks that use this new approach. Additionally, they provided examples and possible problems and their evaluation. In particular, the authors focused their review on tasks related to social media and the size recommendation problem (see Section 2.1.3, where we introduce this task in detail). In our survey, in addition to analyzing the state of the art of the most commonly used algorithms in a wide range of tasks, we went in depth to understand which are the main features used by the more modern fashion RSs. In fact, an extensive discussion is held on how both the user and the items, with their characteristics, can be a source for the definition of models with accurate recommendations.

[2] https://dblp.uni-trier.de/

[3] This link contains a comprehensive list of papers, which will be periodically updated: https://github.com/yasdel/FashionXrecsys

Table 1. List of Abbreviations Used throughout This Article

Abbreviation	Term
AI	Artificial Intelligence
BPR	Bayesian Personalized Ranking
CA	Context Aware
CBF	Content-Based Filtering
CF	Collaborative Filtering
CNN	Convolutional Neural Network
GAN	Generative Adversarial Network
KG	Knowledge Graph
LSTM	Long Short-Term Memory
MF	Matrix Factorization
RNN	Recurrent Neural Network
RS	Recommender System

Furthermore, since the fashion domain focuses on visual aspects, a concise reference is made regarding the utilization of computer vision techniques in the realm of fashion recommendation systems.

The Outline of the Survey. The structure of the article is organized as follows:

- Section 2 is dedicated to the categorization of fashion RSs, which we organize according to the *task* (cf. Section 2.1) such as fashion item, pair, outfit, and sizer recommendation, among others, based on the *input* (cf. Section 2.2) (by relying on collaborative information, textual and visual features, etc.).
- Section 3 is dedicated to the algorithms for fashion RSs, which we perform based on *visually aware CF* (cf. Section 3.1), *generative models* (cf. Section 3.2), and other models (cf. Section 3.3), where the preceding two sections also touch on topics related to *computer vision* and fashion recommendation systems in a brief manner.
- Section 4 is dedicated to evaluation and datasets, which we discuss based on *evaluation goal* (cf. Section 4.1); *evaluation perspective* (cf. Section 4.2) discussing themes such as whether an evaluation is intended to evaluate recommendation, explanation, and image quality; and finally *available datasets* (cf. Section 4.3).
- Finally, Section 5 presents our conclusion, and Section 6 reveals our future challenges.

In Table 1, we mention the list of abbreviations used throughout this article.

2 CATEGORIZATION OF FASHION RSS

We categorize fashion RSs according to their task (cf. Section 2.1), and the input data they use to perform that task (cf. Section 2.2). We will discuss these two categorizations in more detail in the next sections.

2.1 Categorization Based on Task

By *task*, we refer to the internal goal the fashion RS aims to achieve. This affects, in particular, the expected output of the algorithms. We identified five main tasks in the academic literature that fashion RS aim to achieve: (i) fashion item recommendation, (ii) fashion pair and outfit recommendation, (iii) size recommendation, (iv) explanation for fashion recommendation, and (v) other fashion prediction tasks. There are a few other tasks that so far have not gathered much attention but are growing in popularity; they will be summarized later under the general subsection (v).

We discuss these five categories by presenting a broad definition of each task formally while leaving a more detailed discussion of several prominent research works for Section 3.

2.1.1 Fashion Item Recommendation. The fashion item recommendation task, similar to the classical recommendation problem, focuses on suggesting individual fashion items (clothing) that match users' preferences.

Definition 1 (Fashion Item Recommendation). Let $\mathcal{U}$ and $\mathcal{I}$ denote a set of users and fashion items in a system, respectively. Each user $u \in \mathcal{U}$ is related to $\mathcal{I}_u^+$, the set of items she has consumed. Given a utility function $g : \mathcal{U} \times \mathcal{I} \rightarrow \mathbb{R}$, the *item recommendation task* is defined as

$$\forall u \in \mathcal{U},\ i_u^* = \underset{i \in \mathcal{I} \setminus \mathcal{I}_u^+}{\operatorname{argmax}}\, g(u, i), \tag{1}$$

where i_u^* denotes the best matching item not consumed by the user u before. The preference of user u on item i could be encoded as $s_{ui} \in \mathcal{S}$, a continuous-value score (e.g., 1–5 Likert scale), or implicit feedback in which we assume the user likes the item if she has interacted with (i.e., reviewed, purchased, clicked) the item (i.e., $s_{ui} = 1$). $\mathcal{I}_u^+$ represents the set of (u, i) pairs for which s_{ui} is known. The task of personalized top-K fashion item recommendation problem is formally defined as identifying, for user u, a set of ranked list of items $X_u = \{i_1, i_2, \ldots, i_K\}$ that match user preference.

A simple yet effective pure CF model that serves as a foundation for many model-based visual RS is **Bayesian Personalized Ranking—Matrix Factorization (BPR-MF)** [137]. Given a user u, and an item i, the core predictor in this model is given according to

$$\hat{s}_{u,i} = p_u^T q_i, \tag{2}$$

where $p_u, q_i \in \mathbb{R}^F$ are the embedding vectors for user u and item i, respectively, and F is the size of the embedding vector. This predictor is known as **Matrix Factorization (MF)**, and the parameters of the model can be learned using **Bayesian Personalized Ranking (BPR)**, a pairwise ranking optimization framework. We will show later in Section 3.1 how different authors have used this model as a starting point in the fashion domain so that it is extended to consider other types of inputs and learning scenarios.

2.1.2 Fashion Pair and Outfit Recommendation. Fashion outfits are sets of N items worn together (for an outdoor wedding, graduation party, baby shower, etc.). The simplest form of a fashion outfit is when $N = 2$—that is, two different items that look good when paired together, such as an orange shirt and a blue pair of jeans. However, in general, outfits in online fashion stores can be composed of items in different categories (e.g., show, bottom, top, hat, bag) that share some stylistic relationship.

Definition 2 (Fashion Outfit Composition). Let $O = \{i_1, i_2, \ldots, i_N\} \in \mathcal{O}$ denote a fashion outfit composed of a set of compatible fashion items, in which $i_n \in \mathcal{I}$ is the n-th fashion item in the outfit O, $\mathcal{O}$ is the set of all possible outfits, and N is the length of the outfit, whose value is not fixed and can change with each outfit. *Fashion outfit composition* is formulated as follows: "given a scoring function $s(O)$ that indicates how well the outfit $O \in \mathcal{O}$ is composed, find an outfit O that maximizes this utility":

$$O^* = \underset{O_j \in O}{\operatorname{argmax}}\, s(O_j), \tag{3}$$

where s is the outfit utility function, a scoring function that takes into account different types of relationships between fashion items to generate an outfit compatibility score. It is worth noting that, given this general task definition, the *evaluation* of fashion outfit composition is typically performed via **Fill-in-the-Blank (FITB)** or outfit compatibility score prediction [129] described in Section 4.

Moreover, it is possible to encode several objectives relevant to the fashion domain in the definition of the outfit composition scoring function by incorporating domain knowledge. For instance, McAuley et al. [120] defined compatibility according to *complementary* relationships (e.g., how much a white shirt complements blue pants) and *similarity* (how much one item in the outfit

is visually similar to another item). Hsiao and Grauman [74] model the outfit scoring function $s(O_j)$ as the superposition of two objectives:

$$O_j^* = \underset{O_j \in O}{\operatorname{argmax}}\, c(O_j) + v(O_j), \tag{4}$$

where $c(.)$ and $v(.)$ denote the *compatibility* score (how much pairs complement each other) and *versatility* score (defined as coverage of all styles), respectively. Parameters of the outfit scoring functions may be personalized to each style/user and learned from user interaction data. As an example, Lin et al. [107] propose a system where compatibility is computed according to the image and category of every item in the outfit.

Recommending a pair of fashion items or an outfit evidences several challenges, which have been addressed as follows:

- *Personalizing to a target customer*: Compatible fashion outfit recommendations can be studied in a personalized fashion with respect to the target user u to whom recommendations are computed:

$$O_u^* = \underset{O_j \in O}{\operatorname{argmax}}\, s(u, O_j). \tag{5}$$

  A common problem here is the availability of sufficient interaction data to learn personalized outfit models. Different approaches have been proposed to address this issue in the literature, notably by using fashion generation [27], graph-learning approaches [101], binary codes [116], and finally a self-attention mechanism to model the higher-order interactions between fashion items [115].
- *Modeling outfits as a sequence*: To take advantage of the representation of order-aware models such as **Long Short-Term Memory (LSTM)**, some works [59] model an outfit as an ordered sequence of items, where each item belongs to step t. Then sequence-aware methods model the visual compatibility relationships of outfits as a next item recommendation task:

$$\forall O_t = \{x_1, x_2, \ldots, x_t\},\ x_{t+1} = \underset{x \in \mathcal{Y}}{\operatorname{argmax}}\, s(O_t, x), \tag{6}$$

  where $\mathcal{Y}$ contains the set of allowed items. For instance, a user may want to find the best "shirt" that matches his/her current "shoes" and "pants." In this example, all the "shirts" in the database could be a candidate item set, thus filling the cycle: "shoes → pants → shirts," and this cycle can go on to add more items such as "shoes → pants → shirts → hat → bag" (bottom-up outfit representation), depending on the size of outfit considered.

  This problem might be generalized by allowing an input query x_q, either as a textual query (e.g., what outfit goes well with this mini skirt?) or a visual query (photo of a skirt). Moreover, when x_q is a visual query, the task is recognized as a *visual retrieval* task in the information retrieval community or *contextual recommendation* in the RS community.
- *CBF versus CF*: Although fashion item recommendation tasks have been dominated by CF-based approaches, which may or may not include side visual or textual information, the problem of outfit recommendation has been predominantly approached via content-based approaches, such as via visual metric learning problems. This can be explained due to scarcity of outfit interaction data. Nevertheless, there have been few works that approach the outfit problem in a CBF-driven approach. Consider the work by Hu et al. [76], in which the authors use multi-modal features of items and tensor factorization to model the interactions between users and fashion items.

Here, we review some notable research works that study fashion outfit composition tasks. Two crucial issues arise here [186]:

- How can we learn domain knowledge about fashion compatibility relationships between fashion items in the presence of subtleties and subjectivity that can exist across customers?
- How can we incorporate the learned domain knowledge into the design space of fashion recommendation models?

A number of works have tried to model visual/style compatibility as a *metric learning* problem, for instance, for clothes [65, 120, 129] and for furniture [5, 17, 127]. Yin et al. [186] propose a fashion recommendation system that, in the first phase, learns and incorporates visual compatibility relationships of items at the pixel level. In the second phase, a domain adaptation strategy is used to allow the use of the knowledge extracted from the source domain that can be consistent with a different domain—that is, how to make sure that the knowledge learned in the database used in their experiments can also be compatible with others.

Polanía and Gupte [129] proposed a neural model composed of two subnetworks, first a Siamese subnetwork that extracts visual features from a pair of images and then a metric learning approach that maps the pair of features and auxiliary cues (color information) to a fashion compatibility score. Li et al. [102] proposed learning a representation of an outfit by exploiting the information from the image, title, and category and a multi-modal fusion model. Prediction is made by computing a compatibility score for the outfit representation. The scoring component is learned in an end-to-end fashion. Han et al. [59] model the outfit generation process as a sequential process and employ a bidirectional LSTM to predict the next item from the current one. However, these works build on two key assumptions: (i) a fixed number of items in an outfit or (ii) a fixed order of items in terms of their categories, which may not exist in reality. To address this issue, other approaches have been proposed that do not build on these assumptions. To address the issue of outfit recommendation, Lin et al. [104] employ a multiple instance learning approach. This assumes that labels for a group of items are known, but not for each individual item. They propose a two-stage framework called *OutfitNet* for a fashion outfit selection. In the first stage, a Fashion Item Relevancy network (FIR) is used to learn the compatibility of fashion items and garment item relevance embeddings. In the second stage, an Outfit Preference network (OP) uses visual input to learn consumers' preferences. OutfitNet takes in various fashion products in an outfit, and learns the compatibility of the fashion items, the users' preferences for each item, and the users' attention on different pieces in the outfit.

Recent works in the fashion domain propose scalable approaches for personalized outfit recommendations. Chen et al. [27] present an industrial-scale fashion RS, POG, that uses a multi-modal neural model and transformer architecture to better capture dependencies among items. Sarkar et al. [142] propose OutfitTransformer, a scalable framework that uses self-attention to learn relationships between items in an outfit modeled as an unordered set. The system is trained using a novel set-wise outfit ranking loss, which generates a target item embedding and specification based on an outfit.

2.1.3 Size Recommendation. As the fashion industry continues to evolve, it has become increasingly important to consider more subjective factors such as fit, fabric softness, and material smell in fashion recommendation systems. According to Jaradat et al. [84], the ability to take measurements with mobile phones and apply algorithms will become crucial in this domain. However, the issue of size-fit remains a key concern for customers, leading to significant product returns and financial losses for companies. To address this challenge, size recommendation systems have been developed to help users select the best-fitting items without physically trying them on. Despite these efforts, determining the appropriate size and fit for customers still poses various challenges, as highlighted by Jaradat et al. [86]:

***Lack of consistency between brands*:** There is a large number of approved sizing systems around the globe for various clothes, such as dresses, tops, skirts, and pants, and brands. Moreover, there are different size systems such as numeric (38-39-40), standard (S,M,L), fractions (41 1/3, 42.5), convention sizes (36–38, 40–42), and country conventions (EU, FR, IT, UK), where inconsistencies and different ways of converting a local size system to another (as brands do not always comply with the same conversion logic) make the task challenging.

***Subjectivity*:** The exact size is a very subjective feature; users who have purchased items with the same style and shape may make future purchases with different sizes; and how an item fits on your body depends on or can be influenced by several factors, making an objective recommendation difficult. Moreover, customers may be driven by emotional aspects; even a piece of accurate size advice can come with a high emotional cost when the advised size differs from the customer's expectation.

***Data sparsity*:** On one hand, users are able to buy only a small part of the items of an e-commerce website, and on the other hand, articles have a limited stock, which can in turn hinder the task of RSs working with **User-Item (U-I)** fit feedback.

***Noise*:** The underlying interaction data where fashion recommenders are trained can be very noisy since users may make purchases not only for themselves but also share information about clothing with their friends and family, hindering the task of size recommendation.

Given a fashion garment and the shopping history of a user, size recommendation methods predict whether a given size will be too large, small, or correct. Approaches for size recommendation can be classified according to which type of input data they rely on:

- *Physical body related features*: The easiest way to make effective sizing recommendations is to use data from certain parts of the body [69, 72], such as bust, waist, and hip. Sometimes it may be useful to involve the user directly, such as by providing questionnaires to obtain other information, such as age, sex, height, and weight, or in more recent systems to provide their own photos and through vision algorithms to extract 3D avatars that reflect the physical characteristics of the users [94].
- *U-I fit feedback*: To provide personalized size recommendations, the interaction between the user and the item is essential; in this sense, using purchase data [1] or user returns and features of items [56, 69, 72] are used for this purpose while allowing the user not to have to provide data relating to their physical appearance.

A few research works on size recommendation are reviewed here. Abdulla et al. [1] propose a size RS that treats the recommendation problem as a classification problem. It learns a joint size-fit shared latent space to project users and fashion items into, using the skip gram based Word2Vec model. A gradient-boosting classifier is then used to predict the fit likelihood. The authors validate the usefulness of their system with both offline and online evaluation.

The implementation of Bayesian models is described in other works [56, 122] to address size and fit-related returns, which involve keeping, returning, or exchanging an article due to sizing issues. These models utilize prior knowledge and expertise to enhance their accuracy and have been tested on a massive scale of data from millions of customers. In summary, these models exhibit significant potential in reducing returns and enhancing customer satisfaction in the fashion e-commerce industry.

In the fashion industry, choosing clothing that fits well and flatters one's *body shape* is crucial. To address this issue, two approaches are discussed in the following. The first, ViBE (short for *Vi*sual *B*ody-aware *E*mbedding) proposed by Hsiao and Grauman [72], is a framework that captures clothing's association with various body shapes. ViBE uses an embedding learned from

images of fashion models of different shapes and sizes wearing various garments to recommend clothing that will suit the user's body shape. This approach allows for the creation of a 3D model of the human body that can be used for personalized recommendations. The second approach, proposed by Hidayati et al. [69], is a clustering-based body shape assignment approach that compares the body measurements of celebrities to assign them to body shape categories such as hourglass, rectangle, round, and inverted triangle. The system generates networks of images based on the visual appearance of clothing styles and body shape information, which are then used to recommend clothing for each body shape. However, one issue with this approach is that the body shapes of celebrities may not correspond to those of typical users. Overall, both approaches offer promising methods for personalized clothing recommendations that take body shape into account.

2.1.4 Learning Explanations for Fashion Recommendation. An explanation is a piece of information that is shown to users that describes why a particular item is recommended. In recent years, there has been an increase in the number of studies on *explainable recommendation* in various areas, especially the fashion domain, which can be categorized along the following dimensions:

- *Explanation goal*: Explanations have a significant impact on how individuals react to recommendations. An RS's aims in creating explanations for a particular recommendation, however, might range from enhancing the transparency to facilitating a speedier decision to convincing the user [14, 158].
- *Information source for explanation*: Recommendation explanations can be derived from many information sources and be presented in a variety of display styles. Examples of such styles in the fashion domain include textual explanations [29, 31], or visual explanations to specific regions in the image [29], feature-level explanations based on fashion items' content features [40, 155], and by comparison with the user purchase history [70].
- *Explanatory approach/focus*: Explainable recommendation research can focus on either developing interpretable *models* for increased transparency or developing separate post-hoc/model-agnostic approaches to explain recommendation *results*. Thus, the contrast lies in whether the explainability is focused on the method itself or the outcome.

In other words, the explanation goal asks, "What is the explanation's ultimate objective?" whereas the explanatory approach/focus seeks to comprehend "What will the explanation focus on?" In Figure 1, different types of visual explanation used for fashion recommendation are illustrated. Chen et al. [29] present a method that provides visual explanations to the user through certain regions of the items, where an LSTM combines visual information and user reviews. With the same goal, Tangseng and Okatani [155] propose a system that provides outfit recommendations and provides scores for each item or for each item feature to understand how much these influence the outfit composition. To explain whether an item/feature is compatible with the outfit and its influence on the score, three attributes have been extracted from the images: shape, texture, and colors. K-means clustering is applied for colors, and a **Convolutional Neural Network (CNN)** is used for shape and texture representation.

2.1.5 Other Fashion-Related Prediction Tasks. Besides fashion items and outfit recommendations, a few similar tasks are gaining attention in the fashion and retail industry. One of these areas includes *image-based fashion generation*, where methods have been proposed to generate outfits so that the user could explore them and judge their compatibility [91, 105, 184]. One of the key techniques used in these works is **Generative Adversarial Networks (GANs)**, which will be described in detail in Section 3.2. Other relevant prediction tasks concern the fashion **Artificial Intelligence (AI)** such as design [91], wardrobe creation [74], fashion trend forecasting [8], and societal/marketing biases such as socio-demographic inequality structures through recommendation

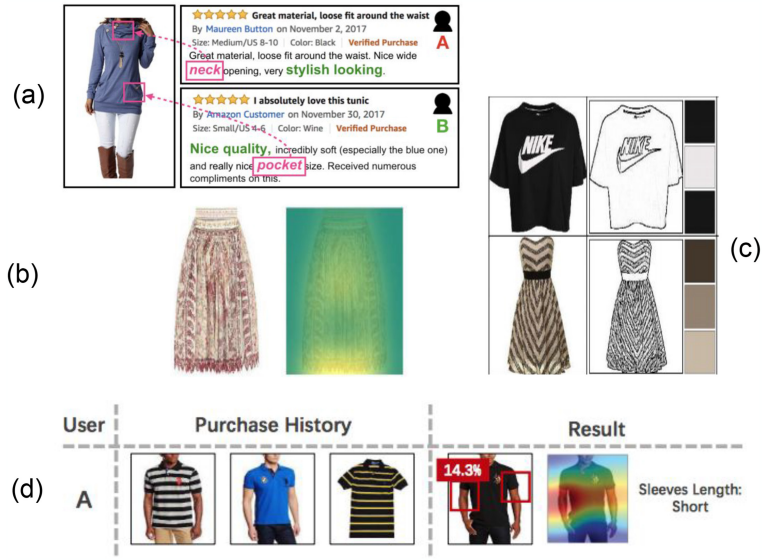

Fig. 1. (a) Personalized visual explanation by Chen et al. [29] related to the region of the fashion image based on user reviews. (b) Explainable outfit recommendation by Lin et al. [106] using a neural attentive model. (c) Outfit recommendation by Tangseng and Okatani [155] (d) Fashion item recommendation proposed by Hou et al. [70] comparing recommendation results with the target user's purchase history.

systems in fashion [21], or user shopping behavior online and in brick-and-mortar stores [174], or capturing users' seasonal and temporal preference changes [178].

2.2 Categorization Based on Input

By *input*, we refer to the data fashion RS used to train their models. They could be based on a *U-I matrix* together with *side information beyond the U-I matrix* (related to users and item), and contextual information. This information determines the type of RS, according to CF, hybrid (CF+CBF) systems, and **Context-Aware (CA)** systems. In Table 2, we provide a classification of data/computational features used for the design of fashion RS by authors over the years.

2.2.1 Fashion RS Relying on U-I Interaction Data. Recommendation models that are trained purely from U-I interaction data are known as CF. Neighborhood models are among traditional CF classes that predict the unknown U-I preferences based on the neighborhoods defined over user-user or item-item similarities (user-based CF and item-based CF, respectively), derived from the preferences or interactions of used with products. Machine-learned recommendation models are powered by model-based CF, a parameterized model whose parameters are learned in the context of an optimization framework, such as BPR [138]. In the work of Agarwal et al. [3], BPR is used to address the typically vast size of the catalog in e-commerce platforms with a two-stage recommendation formulation. The first stage generates non-personalized similar candidate products, and the second stage uses a model-based CF trained with BPR to provide user-level personalization (top 50 styles).

MF, or BPR-MF, is one of the most promising approaches to building CF. MF essentially encodes the complex relations between users and items into a lower-dimensional (latent) representation of users and items, whose dot product explains the unknown predictions [95]. Hwangbo et al. [79] propose a fashion RS that simultaneously proposes substitutes and complements based on

Table 2. Classification of Different Research Works Incorporating Various Side Information Kinds in Their Proposed Approach

Authors	Year	Data/Computational Features																		
		Side Info User		Side Information of Items													Other Information			
				Image Representation							Text Representation									
		Social network	Body features	Color	Texture	CNN	R-CNN	Siamese-CNN	RNN	Others	Pure	CNN	RNN	Word2Vec	Transformer	Graph embedding	Others	Celebrity/Social infl.	Fit feedback	Pair feedback	Others
Jing et al. [89]	2023												✓								
Ye et al. [185]	2023									✓											
Sarkar et al. [142]	2023									✓				✓							
De Divitiis et al. [38]	2023			✓						✓											
Jandial et al. [82]	2022					✓							✓								
Delmas et al. [49]	2022					✓							✓								
Chia et al. [33]	2022																				
Wu et al. [177]	2022					✓															✓
Sá et al. [140]	2022																				
Ravi et al. [135]	2021					✓				✓											
Hidayati et al. [68]	2021		✓	✓		✓															
Yu et al. [188]	2021			✓	✓	✓														✓	
Zhan et al. [189]	2021					✓								✓		✓					✓
Wen et al. [171]	2021					✓							✓								
Lee et al. [98]	2021					✓							✓								
Nestler et al. [122]	2021					✓													✓		
Xing et al. [178]	2020					✓			✓												✓
Wölbitsch et al. [174]	2020									✓											✓
Li et al. [99]	2020					✓					✓										
Banerjee et al. [16]	2020					✓					✓										
Dong et al. [52]	2020		✓								✓										✓
Tangseng and Okatani [155]	2020			✓	✓					✓											
Sridevi et al. [152]	2020				✓	✓															
Lin et al. [106]	2020					✓			✓	✓	✓									✓	
Chen et al. [28]	2019					✓		✓				✓			✓			✓			
Polanía and Gupte [129]	2019			✓				✓												✓	
Hou et al. [70]	2019					✓				✓											
Kumar and Gupta [96]	2019						✓			✓										✓	
Abdulla et al. [1]	2019		✓											✓					✓		
Lin et al. [105]	2019					✓									✓						
Pan et al. [127]	2019						✓	✓													
Yin et al. [186]	2019					✓															
Cardoso et al. [22]	2018					✓				✓		✓									
Tuinhof et al. [160]	2018					✓															
Guigourès et al. [56]	2018																		✓		
Hidayati et al. [69]	2018		✓	✓		✓										✓		✓			
Agarwal et al. [3]	2018																				✓
Hwangbo et al. [79]	2018										✓										✓
Zhou et al. [195]	2018			✓	✓	✓	✓				✓										
Liu et al. [108]	2018					✓											✓				
Sun et al. [153]	2018	✓				✓															
Jaradat [83]	2017			✓			✓										✓				
Zhang et al. [191]	2017		✓																		✓
Qian et al. [131]	2017			✓	✓		✓														
Zhao et al. [192]	2017							✓				✓									✓
Heinz et al. [67]	2017					✓			✓								✓				
Kang et al. [91]	2017							✓													
Bracher et al. [20]	2016					✓											✓				
McAuley et al. [120]	2015					✓				✓										✓	
Hu et al. [76]	2015			✓	✓												✓				
Bhardwaj et al. [19]	2014			✓																	
Jagadeesh et al. [81]	2014			✓	✓																

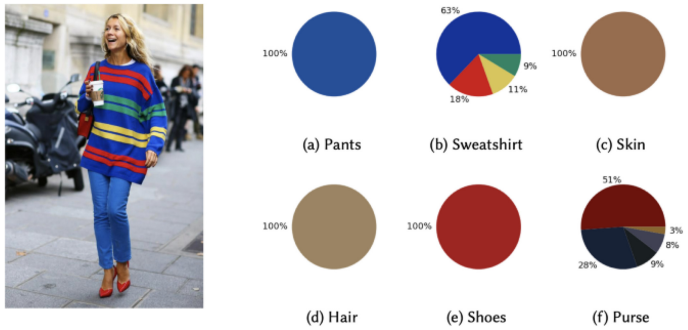

Fig. 2. Color extraction example from probabilistic color modeling of clothing items. Courtesy of Al-Rawi and Beel [9].

previously purchased items, with the substitutes driving the majority of purchases according to evaluation data. The authors further consider a preference decay factor to account for the life span of a product.

Thanks to deep neural models' ability to handle non-linear data through non-linear activation functions, recently *deep neural CF models* have been adopted by the research community to address the linearity issue of MF-based approaches and thereby uncover a more complex relationship between users and items. Given the challenges and nuances in the fashion domain, most of the works incorporate additional or side information to the problem and hence will be discussed in the following subsections.

2.2.2 Fashion RS Relying on U-I Interaction and Side Information. The range of information sources that RSs adopt can stretch beyond the U-I matrix, typically provided by attributes or information related to items such as the category of the user such as user gender, age, and demographics [146]. To surface valuable recommendations, it is essential to learn from available user interactions, understand, and uncover the underlying decision factors. We categorize this side information according to items and users.

Fashion RS Relying on U-I data and Side Information of Items. Understanding clothing provides a good platform for making recommendations [36]. State-of-the-art fashion RSs employ a variety of types of content features, such as visual and textual features, as side information to U-I interaction data. In the following, we review the most prominent features and attributes used for item representation in the fashion recommendation literature [36, 45]:

(1) *Color*: The most common means to identify how one looks is achieved via colors, materials, and silhouettes on the body. Color and color consistency are among the most recognizable fashion clothing features that consumers often consider to decide what they want to wear. Color can be represented in several forms. For instance, as shown in Figure 2, Al-Rawi and Beel [9] describe a technique to segment and extract the color of a person's picture through probabilistic color modeling for clothing items. This technique can either increase or replace the data entry procedure needed to add a fashion item to electronic commerce catalogs. Jagadeesh et al. [81] evaluate two image representations, HSV Histograms and color bag-of-words (BoW). Their results showed that color is a better descriptor than textures. In the work of Qian et al. [131], a recommendation model is proposed that uses a color map through k-means based on the value of the segmented pixel of the clothing items with an aim toward extracting a dominant color. Starting from an attribute query, Chae et al. [23] convert this

into a color vector space query; in particular, the Lab color space is used to simulate human visual perception through a palette of 269 colors.

(2) *Brand*: Product brands are a critical feature users consider when deciding among items. Wakita et al. [165] define a fashion-brand recommendation system, which through a deep (feed-forward) neural network has the primary purpose of predicting the user's favorite brand starting from user information, such as gender and height. Jaradat [83] proposes a clothing recommendation system that uses brands as the main feature, where the model suggests to users clothes of similar brands that they have already purchased and that are on trend.

(3) *Deep visual features*: Fashion recommendations can be improved using image-level features extracted through a deep network, such as a CNN [178, 195]. For instance, Zhou et al. [195] propose a CNN implementation to address the issue of two-piece apparel matching that is suitable with current fashion trends. They merged perception and reasoning models and constructed two parallel CNNs to enable the system to recognize garment features, one for upper-body clothing and another for lower-body apparel. A hierarchical topic model incorporates the resulting information into style topics with better semantic understanding to interpret the collocation patterns. The authors present a novel learning model based on Siamese CNNs for learning a feature transformation from clothing photos to a latent feature space expressing fashion style consistency.

(4) *Texture*: The texture describes the body and surface of a garment. It has a direct impact on users and is often used in recommendation systems. For example, Tangseng and Okatani [155] propose an outfit recommendation system based on attributes that are "human-interpretable," including texture, and highlight how useful it is to use them. Zou et al. [199] use eight different local binary patterns to extract texture features from pixels of an input image, after which a local vector is calculated for each region and encoded to obtain the final vector. Qian et al. [131] provide a system that uses a CNN to locate textures in an image dataset to create matches and provide recommendations. However, Jagadeesh et al. [81] affirm that the color of the clothing is more important than the type of texture because the former is a stronger descriptor than the latter.

(5) *Style*: Style is associated with how people intend to express themselves through visual elements such as accessories, clothing, hair, and other aesthetic features. Measuring style computationally and offering personalized style-based fashion recommendations is important for retail companies and the central focus of many research works. A wide range of methods has been proposed to model style in fashion recommendation, ranging from modeling the body characteristics of people with the features of clothes presented in Hidayati et al. [68], modeling fashion substitution and complementarity with other products by learning a parametric transform of distances proposed by McAuley et al. [120], and multi-modal representations to build trend-sensitive models within the fashion field as proposed by Zhou et al. [195].

(6) *Knowledge graph*: A **Knowledge Graph (KG)** is a heterogeneous graph where nodes serve as entities and edges serve as relationships between entities. The idea of making recommendations using side information from a KG has received considerable interest in the RecSys community. Such an approach can address the issues with RSs and provide explanations for recommended items. For instance, Li et al. [101] propose a model called a *Hierarchical Fashion Graph Network*, which uses a graph network to model the relationship between users and items for outfit recommendations through the propagation of information based on three levels, which are items, outfit level, and user level. Another possible use of KG is related to the solution of the cold start problem of many recommendation systems—for example, Yan et al. [180] use a model in which they build a U-K KG to discover the relationships that

exist between them and solve the cold start problem. Zhan et al. [189] address the problem of personalized outfit recommendation by proposing an Attentive Attribute-Aware Fashion Knowledge Graph (A3-FKG) that is used to associate various outfits with both outfit- and item-level features. Dong et al. [52] use a knowledge base to model the relationships between human bodies, fashion themes, and design factors using fuzzy techniques. The resulting ontology-based design knowledge base is used in an interactive, personalized design RS.

(7) *Textual features*: Textual features in the fashion domain can have several forms, such as reviews [29], article titles [192], article descriptions [105], and tags [20]. Zhao et al. [192] propose a *sentence model* that combines pairs of article titles to determine any compatibility of their styles. Textual features are integrated/modeled in the system for different purposes—for instance, for fashion generation [105]. In such a scenario, given an image of a top and a query vector description of the bottom, the fashion generator needs to generate a bottom image that matches the top image and the description as much as possible, process reviews [29], or attribute prediction [22]. A shown in Table 2, textual information has been used to represent item content in different forms, including term frequency–inverse document frequency [195], CNN [27, 192], Word2Vec [22, 76], transformer architecture [28], and graph embedding [56, 69], among others.

Some works have tried to evaluate the quality of the features mentioned previously for fashion recommendation. For example, Jagadeesh et al. [81] compare textual and visual features, and Deldjoo et al. [41] compare different CNN types for visual modeling of fashion items. In particular, the aim of Jagadeesh et al. [81] is to make efficient use of a large amount of fashion street images. To this end, two data-driven models are proposed, a deterministic fashion recommender (DFR) and a stochastic fashion recommender (SFR). The first is useful for identifying recommendations that are as objective as possible, and the latter is used to model the bias of users. An important aspect highlighted is that color is more important than the type of texture, which underlines how a simple descriptor such as color can be important. Deldjoo et al. [41] evaluate the quality of several visual fashion RSs, including **Visual Bayesian Personalized Ranking (VBPR)**, DeepStyle, ACF (attentive CF), and VNPR (introduced in Section 3) empowered by pre-trained CNN types such as AlexNet, VGG-19, and ResNet-50. Evaluation is performed on accuracy and beyond-accuracy objective and qualitative assessment of the visual similarities between pairs of images.

Fashion RS Relying on Interactions and Side Information of Users. These approaches in the most canonical form use data from social networks and fashion magazines and websites, in which user-user and body features (e.g., face, makeup, and size) information are extracted.

The main source of side information on users stretching beyond the U-I matrix is *social networks*, which has impacted a wide range of research disciplines in RS [146]. Mining social trust, influential figures, celebrities, and bloggers are among the influential factors impacting users' fashion choices. Different types of social network based relationships have been used in the RS domain to enhance the quality of recommendations, including membership, friendship, following, and so forth.

Different types of online platforms have been used for the purpose of advertising, marketing, social, and similar activities involving fashion and garments:

- *Social media*: With the advent of social media platforms like Instagram, the way fashion is advertised, and even the way it is designed, is being re-imagined. Hundreds of thousands of people share images of their "outfit of the day,"which prompts a flood of comments and queries from other users. Fashion brands can now use social media platforms such as Instagram to grow their businesses. Today, a single post by a fashion influencer receives much attention, allowing many followers to know more about the brands and styles of the clothes

that users mostly follow to understand which brands and models users follow the most, or to catch the trends of the moment or seasonality [83, 85].

- *Fashion magazines and beauty websites*: It is especially common for celebrities' styles to be viewed as fashion inspirations, as they pay fashion stylists to assist them with their wardrobe choices, which allows them to visually alter their figure. Hidayati et al. [69] extract a list of essential female celebrities from prominent fashion magazines such as *Vogue*, *Glamour*, and *Marie Claire*. In addition, a suite of body parameters is extracted from a body measurement website associated with celebrities. The dataset was completed using female celebrity photos taken from Google search engines. Polyvore[4] was another good example of a fashion website, where users could build and post information about fashion outfits reflecting their taste. These fashion outfits feature a wealth of multi-modal data, including photographs and descriptions of fashion products, the number of times the outfit has been liked, and the outfit's hash tags. Researchers have applied this data to a variety of fashion-related tasks (cf. Section 4).

Sun et al. [153] propose a personalized clothing recommendation system that takes into account both users' social circles and fashion style consistency of clothing products. Fashion style consistency refers to the fact that two clothing items, such as tops and skirts, can be visually different, but as long as they belong to the same style (e.g., sport, street, casual), the user may be likely buy them; hence, fashion style consistency is an essential element for the design of clothing RSs. To this end, the proposed approach considers three factors in the proposed clothing RS, namely interpersonal influence, personal interest, and interpersonal interest similarity; this was motivated by previous research works (e.g., [130]). In particular, five types of matrices are built by mining the social data available and other sources of information: representing user-user social influence, representing the user-user similarity of interests, representing user-clothing similarity, representing clothing-clothing fashion style similarity, and representing user-clothing ratings. Afterward, a probabilistic matrix factorization (PMF) framework is used to integrate the preceding observed matrices and recommend suitable clothing products to users by casting the problem in an optimization setting. Evaluation is carried out on real-world datasets collected from Moguije,[5] a Chinese website for social fashion blending a social network with online shopping options.

2.2.3 CA Fashion RSs. CA models improve CF by including contextual information in the model to account for the current information need of the user [4, 45]. Context awareness (i.e., understating the user's situational or contextual aspects) is essential in improving the user experience. Several taxonomies have been identified to classify context (e.g., based on the computing environment, user environment, and physical environment or based on time, location, and social information (user's friends, social circles)); see the work of Deldjoo et al. [47] for more information. For the scope of this survey, we choose a pragmatic approach and organize CA fashion RSs according to the types of context frequently used in the fashion literature, namely *spatio-temporal context* (e.g., GPS coordinates, location, place of interest, time, season), *multimedia and physiological context* (e.g., free text, image or user body/face features), and *affective context* (e.g., mood or emotion).

Spatio-Temporal Context. The *temporal* and *geographic* characteristics of fashion data influence human preferences. As for temporal aspects, first, fashion garments are sensitive to *seasonal* changes. Retail markets may face a decrease in preference over time after fashion products are released. Second, fashion *style trends* change as time progresses. For instance, trends can appear cyclically, with styles re-emerging after decades of absence, whereas others emerge and swiftly disappear.

[4] http://www.polyvore.com
[5] http://mogu-inc.com/

Understanding how style choices *vary with time*, and thereby what constitutes fashionable trends, is of crucial importance to the fashion retail business.

He and McAuley [63] propose an approach that involves learning the time evolution with which users decide to buy items through their implicit feedback, such as purchase history, clicks, and bookmarks. These features are used to define a predictor, helpful in approximating how much users interact with articles in a particular epoch. Dynamic FDNA proposed in the work of Heinz et al. [67] extends the static FDNA model with the inclusion of time-of-sale information. FDNA combines attributes and visual items using a neural architecture. Dynamic FDNA takes as input the timestamp of the purchase associated with a customer, and using existing information about the customer (purchase history so far), determines their current style. The dynamic model has the main advantage of providing recommendations for short-term customer intent. Adewumi et al. [2] propose a framework that provides (via a weather API) recommendations for the exact day the user needs them.

Multimedia and Body/Facial-Related Context. The term *multimedia context* refers to situations in which the system requires a user to provide a multimedia item (e.g., an image of a fashion item or an image plus textual query) to initiate the recommendation process (e.g., recommending a garment that complements an item the user is wearing):

- *Context = image*: Images are an important visual tool for users to communicate with a fashion RS. They could be used as input to the system to retrieve complementary or matching fashion products. For instance, given an image of a fashion item (e.g., "jeans"), one can identify matching fashion products (e.g., "tops") that complement an item the user is considering [81, 195]. An application in the context of fashion is "upselling in e-commerce" when an online shopper with an item in her shopping basket is prompted to buy more things that match the one they already have in their shopping carts.
- *Context = image + text*: In the fashion domain, multi-modal dialogue systems that can accurately recommend or generate matching fashion items based on a user's query image and textual description can potentially revolutionize how people shop for clothes. These systems can help users easily find the perfect outfit by providing personalized and accurate recommendations based on their preferences [49, 82, 98, 105, 171]. The problem addressed in these works can typically be formally started as given a query image and a textual description of modifications; the objective is to train a model that can learn an image-text representation that closely aligns with the visual representation of the target ground-truth image. Wen et al. [171] introduce a Comprehensive Linguistic-Visual Composition Network (CLVC-Net) that utilizes fine-grained local-wise and global-wise composition modules, along with a mutual enhancement module to retrieve the target image. The system can handle simple text modifications such as "change it to long sleeves" and abstract visual property adjustments like "change the style to professional." Lee et al. [98] propose a CoSMo (*Co*ntent-Style *Mo*dulation) algorithm that introduces two modules—the content and style modulators—to achieve desired modifications to the reference image feature for image retrieval with text feedback. Delmas et al. [49] propose an efficient model that consists of two modules (cross-modal and visual retrieval) with text-guided attention layers, each handling one modality of the query. These modules are trained jointly. These systems hold great potential for developing *multi-modal dialogue systems* and other tasks such as *virtual try-on* to effectively understand and respond to human queries involving text and images.
- *Context = body/facial-related attributes*: The user's physical appearance, including their face or body, can provide valuable contextual information to a recommendation system, which can then suggest cosmetic products [34, 191], clothing [68, 69], and hairstyle

Table 3. Classification of Fashion RS Based on the Input and Output of the System

Class	Input	Output	Research Works
Classical fashion RS	–	item	[56, 67, 70, 79, 83, 153, 191]
	–	pair/outfit	[1, 28, 76, 79, 153, 155]
Contextual fashion RS	image	item	[3, 19, 22, 81, 91, 131, 160, 186, 194]
	image	pair/outfit	[19, 96, 106, 108, 120, 129, 153, 186, 192, 194]
	image+text	item	[98, 171, 195]
	image+text	compl. pair/outfit	[105, 195]
	body/face	item	[34, 69]
	body/face	compl. pair/outfit	[191]
GAN-based fashion image gen.	–	item	[91, 147, 149, 184]
	–	compl. pair/outfit	[78, 96]

Note: Compl. pair/outfit refers to settings where ***complementary*** relationships between fashion items are considered, either for pairwise matches or outfit recommendation.

recommendations [109, 182]. For instance, Hidayati et al. [68] propose a framework that uses both body characteristics and clothing style to provide size recommendations. Meanwhile, in a different work by Hidayati et al. [69], a system requires users to provide information about their body, such as their shape, weight, height, bust, waist, and hip measurements, which is then correlated with photos of celebrities to suggest a style that matches the user's preferences.

In Table 3, we provide reference points to research works that explicitly use specific multimedia contexts (e.g., text and images) for fashion item information access (recommendation or search).

3 ALGORITHMS FOR FASHION RSS

For the purpose of this survey, we focus on the two most prominent approaches for fashion RS: (i) visually aware model-based CF (cf. Section 3.1) and (ii) generative fashion recommendation models (cf. Section 3.2). Finally, we briefly discuss other fashion RS approaches (cf. Section 3.3).

3.1 Visually Aware Model-Based CF

The fashion industry has a wide catalog of diverse items and a high rate of change or return as a result of market dynamics and customer preferences. This results in a lack of purchase data, which complicates the use of standard RSs. In addition, it is difficult to compute similarities across products when exact and extensive product information is not provided. Computer vision research is increasingly being used to address the issues outlined earlier. Given the visual and aesthetic nature of fashion products, a growing body of research addresses tasks like localizing fashion items, determining their category and attributes, or establishing the degree of similarity to other products. For instance, Qian et al. [131] use a CNN to segment and locate elements from the complicated backgrounds of street-style images. Hou et al. [70] use a semantic segmentation approach to derive region-specific attribute representations.

***VBPR (2016)* [64].** The VBPR method is built upon BPR and extends it by incorporating a latent content-based preference factor. The core predictor in VBPR is given by

$$\hat{x}_{u,i} = p_u^T q_i + \theta_u^T \theta_i, \tag{7}$$

where $\theta_i \in \mathbb{R}^K$ is typically designed to represent the visual signal of an item and $\theta_u \in \mathbb{R}^K$ represents the user preference on the visual dimension. He and McAuley [64] built the VBPR predictor

according to

$$\hat{x}_{u,i} = p_u^T q_i + \theta_u^T \, E \, \underbrace{\Phi_f(\mathbf{Img}_i)}_{f_i}, \tag{8}$$

in which $f_i = \Phi_f(\mathbf{Img}_i) \in \mathbb{R}^{F\times 1}$ represents the latent feature of item i (typically extracted from a pre-trained CNN) and Φ_f denotes the feature extractor. $E \in \mathbb{R}^{K\times D}$ linearly transforms the high-dimensional feature F into a lower-dimensional (e.g., $D = 20$) named the *visual space*. He and McAuley [64] show that VBPR improves the performance of BPR-MF and popularity-based recommenders for datasets chosen from Amazon fashion and Amazon phones.

***DeepStyle (2017)* [110].** Although VBPR has demonstrated effectiveness, it tends to learn the visual characteristics of items based on their category rather than their specific style (e.g., casual, aesthetic, or formal). Creating a style-CNN requires labeled data, which can be a time-consuming task that requires expert knowledge in the field of fashion. However, DeepStyle presents an innovative solution to this problem by using a U-I matrix to learn the style of items. The approach involves modeling the item as a combination of its style and category, where removing the category information is achieved by modeling the style as the difference between the item and the category (`style = item - category`). This transformation effectively converts the VBPR predictor into a dedicated style-CNN:

$$\hat{x}_{u,i} = p_u^T q_i + p_u^T(\mathbf{E} f_i - l_i), \tag{9}$$

where l_i is a latent factor that embodies the categorical information of item i.

***ACF (2017)* [26]** Attentive CF integrates an attention network to consider the varying levels of attention that users give to different components within multimedia content, such as the regions in an image or frames in a video when making recommendations. This approach is more effective than traditional visual recommendation systems that do not account for such nuances in user behavior:

$$\hat{x}_{u,i} = \left(p_u + \sum_{k\in I_u^+} a(u,k) v_k\right)^T q_i, \tag{10}$$

in which $a(u,k)$ denotes user u's preference degree toward item k, v_k is the attentive latent factor for item k, and q_i is the basic item vector in the MF model.

***DVBPR (2017)* [91]** VBPR utilizes pre-trained CNNs, such as ResNet-50, to extract visual information. However, these networks are trained for image classification tasks rather than recommendation tasks, and they come from different domains. In contrast, **Deep Visual Bayesian Personalized Ranking (DVBPR)** is an end-to-end model that trains the visual feature extractor and the preference prediction module simultaneously in a pairwise manner:

$$\hat{x}_{u,i} = \theta_u^T \Phi_e(\mathbf{Img}_i), \tag{11}$$

where the product i is associated with an image denoted as $\mathbf{Img}_i$ and its item content embedding is represented by $\Phi_e(\mathbf{Img}_i)$. The CNN model used to generate this embedding is trained directly in a pairwise manner.

Other relevant techniques in this area are TVBPR by He and McAuley [63], which incorporates seasonality and temporal changes into the VBPR model, and CO-BPR by Yin et al. [186], which recommends compatible outfits based on BPR. These methods draw inspiration from traditional content-based recommendation techniques, with the primary challenge being the higher dimensionality and computational costs associated with visual data. To address this, the image representations are projected into lower-dimensional spaces (as with VBPR) or handled through end-to-end (pixel-level) representations.

A general classification of computer vision in fashion RS is presented by Cheng et al. [30], according to (i) fashion detection (landmark detection, fashion parsing), (ii) fashion analysis (attribute prediction, style learning, popularity prediction), and (iii) fashion synthesis (style transfer, pose estimation).

3.2 Generative Fashion Recommendation Models

Most conventional RSs are not suitable for application in the fashion domain due to unique characteristics hidden in this domain. Recently, GAN-based models have been used for fashion generation and fashion recommendation with promising performance. For example, Kumar and Gupta [96] use an enhanced conditional GAN to generate bottom items that can fit with the top items received in input by their framework. GANs gain their power exploiting their *generative power*, allowing them to synthesize realistic-looking fashion items [43]. This aspect can inspire customers' and designers' aesthetic appeal/curiosity and motivate them to explore the space of potential fashion styles.

CRAFT. Huynh et al. [78] address the problem of recommending complementary fashion items based on visual features by using an adversarial process that resembles a GAN and uses a conditional feature transformer as $\mathcal{G}$ and a discriminator $\mathcal{D}$. One main distinction between this work and the prior literature is that the ⟨input, output⟩ pair for $\mathcal{G}$ are both features (here features are extracted using pre-trained CNNs [154]) instead of ⟨image, image⟩ or hybrid types such as ⟨image, features⟩ explored in numerous previous works [164, 196]. This would allow the network to learn the relationship between items directly on the feature space, spanned by the features extracted. The proposed system is named CRAFT (complementary recommendation using adversarial feature transform) since in the model, $\mathcal{G}$ acts like a feature transformer that—for a given query product image q—maps the source feature s_q into a complementary target feature $\hat{t}_q$ by playing a min-max game with $\mathcal{D}$ with the aim to classify fake/real features. For training, the system relies on learning the co-occurrence of item pairs in real images. In summary, the proposed method does not generate new images; instead, it learns how to generate features of the complementary items conditioned on the query item.

DVBPR. DVBPR [91] is one of the first works to exploit the *visual generative power of GANs* in the fashion recommendation domain. It aims to generate clothing images based on user preferences. Given a user and a fashion item category (e.g., tops, t-shirts, and shoes), the proposed system generates new images (i.e., clothing items) that are consistent with the user's preferences. The contributions of this work are twofold. First, it builds an end-to-end learning framework based on the Siamese CNN framework. Instead of using the features extracted in advance, it constructs an end-to-end system that turns out to improve the visual representation of images. Second, it uses a GAN-based framework to generate images that are consistent with the user's taste. Iteratively, $\mathcal{G}$ learns to generate a product image integrating a *user preference maximization objective*, whereas $\mathcal{D}$ tries to distinguish generated images from real ones. Generated images are quantitatively compared with real images using the preference score (mean objective value), inception score [141], and opposite SSIM [124]. This comparison shows an improvement in preference prediction in comparison with non-GAN-based images. At the same time, the qualitative comparison demonstrates that the generated images are realistic and plausible, yet they are quite different from any images in the original dataset—they have standard shape and color profiles but quite different styles.

MrCGAN. Shih et al. [147] propose a compatibility learning framework that allows the user to visually explore candidate *compatible prototypes* (e.g., a white t-shirt and a pair of jeans). The system uses a **Metric-regularized Conditional Generative Adversarial Network (MrCGAN)** to pursue the item generation task. It takes as input a projected prototype (i.e., the transformation of a query image in the latent "compatibility space"). It produces as the output a synthesized image

of a compatible item (the authors consider a compatibility notion based on the complementary of the query item across different catalog categories). Similar to the evaluation protocol in the work of Huynh et al. [78], the authors conduct online user surveys to evaluate whether their model could produce images that are perceived as compatible. The results show that MrCGAN can generate compatible and realistic images under compatibility learning settings compared to baselines.

Yang et al. *and* c^+***GAN***. Yang et al. [184] address the same problem settings of MrCGAN [147] by proposing a fashion clothing framework composed of two parts: a clothing recommendation model based on BPR combined with visual features and a clothing complementary item generation based GAN. Notably, the generation component takes as input a piece of clothing recommended in the recommendation model and generates clothing images of other categories (i.e., tops, bottoms, or shoes) to build a set of complementary items. The authors follow a similar qualitative and quantitative evaluation procedure as DVBPR [91] and further propose a *compatibility index* to measure the compatibility of the generated set of complementary items. A similar approach has also been proposed in c^+GAN [96] to generate a bottom fashion item paired with a given top.

3.3 Other Fashion RS Algorithms

Depending on the task, fashion item and outfit recommendation, and size recommendation as we identified in Section 2.1, a variety of other models have been used in the studied research work, such as **Recurrent Neural Networks (RNNs)** [22, 106], graph modeling [28, 101, 120], two-input (Siamese) CNNs [91, 129, 186, 192, 195], and attention mechanisms [106], among others. A detailed discussion on these approaches is left as a future direction.

4 EVALUATION AND DATASETS

Evaluation of fashion recommendations is a very complex task. It has to be adapted to the specific recommendation task (see Section 2.1) and consider the inputs of the algorithms (Section 2.2). Hence, in the next section, we present several goals that aim to be achieved when deciding for a specific evaluation setting (Section 4.1), whereas Section 4.2 presents the most important perspectives analyzed in the literature; finally, Section 4.3 shows and discusses currently available datasets.

4.1 Evaluation Goal

The goal of the evaluation is usually linked with the task the recommendation is intended to satisfy. However, we may find proposals where the algorithm is evaluated under different goals. For example, an outfit recommender can be evaluated only for that goal by ranking lists of outfits, but a more simple setting could be devised by using the recommender to score different outfits, to validate whether the algorithm is capable of discriminating those outfits that occur in the ground truth.

In the following, we present the most important goals we have identified in the literature:

- *Goal 1—Outfit generation*: The goal of this task is, given a fashion item x_q (e.g., skirt) representing the user's current interest, to find the best item $x \in \mathcal{I}$ (e.g., shirt) or the fashion outfit $F \in \mathcal{I}$ (e.g., shirt, pants, hat) that goes well with the input query. The input x_q can be specified as a textual query (e.g., what outfit goes well with this mini skirt?) or a visual query (photo of a skirt).

  When x_q is a visual query, the task is recognized as a *visual retrieval* task in the community of information retrieval or *contextual recommendation* in the community of RS (x_q is a multimedia context representing a user's interest—e.g., see the work of Kaminskas and Ricci [90] for a definition of multimedia context (cf. Section 2.2.3)).

- *Goal 2—Outfit recommendation*: This goal is related to the fashion and outfit recommendation task, where a set of objects is recommended to the user at once, by maximizing a utility function that measures the suitability of recommending a fashion outfit to a specific user.
- *Goal 3—Pair recommendation*: This goal is a simplification of the outfit recommendation goal when $N = 2$. This is typically performed as *top-bottom* or *bottom-up* recommendation. In this task, given clothes related to the upper part of the body, the aim is to predict the possible lower part and vice versa. For instance, in a mobile setting, the user may take a photograph of a single piece of clothing of interest (e.g., tops) related to the upper part of the body and look for the bottom (e.g., trousers or skirts) from a large collection that best match the tops [81]. This typically requires a collection of pairs of top and bottom images for compatibility modeling. With advancements in computer vision, in some works, automated top-bottom region detection from photograph images has been proposed [80].
- *Goal 4—FITB*: This goal is used in a setting where we are given an incomplete outfit F^- (e.g., shirt, pants, accessories) with a missing item (e.g., shoes), and the method must find the best missing fashion item $x \in \{x_1, x_2, x_3, x_4\}$ from multiple choices where $F = \{F^-, x\}$ has fashion items, which are compatible visually [60]. This is a convenient scenario in real life—for example, a user wants to choose a pair of shoes to match his pants and jacket.
- *Goal 5—Outfit compatibility prediction*: This goal is focused on, given a complete outfit $F \in \mathcal{F}$, predicting a compatibility score that best describes the composition of an outfit. All items in an outfit are compatible when all fashion items have similar style and go well together [60, 156]. This task is helpful since users may create their own outfits and wish to decide if they are compatible and trendy.

4.2 Evaluation Perspective

Besides the goal that is considered when evaluating an RS, we may find authors who take different perspectives when considering what is a good recommendation. Although most of the literature in the fashion domain has been focused on building more accurate prediction and recommendation models, as in other recommendation settings, recent research has gone beyond this perspective and analyzed the explanations that should be presented to the user, the images that go with the recommendations, and even social aspects that may reinforce biases in the system. We describe these perspectives next:

- *Evaluate the recommendation*: Most papers perform offline evaluation, where classification or ranking metrics are very popular (precision, NDCG, AUC); in this sense, business metrics like CTR or purchase percentage are less common. Later, Table 4 presents a summary. It should be noted that, depending on the recommendation task, other concepts, such as compatibility, might be measured to assess how well the user recognizes the recommendation. Recent approaches (e.g., [33]) propose comparative recommendations focused on product discovery tasks. Other works explicitly address the context of the recommendation—in particular, the *scene* in which the outfit is expected to be considered [185], natural language feedback [177], or maximum shopping budget [16].
- *Evaluate an explanation*: Understanding recommendation explanations is generally a difficult task, as theoretically it can be evaluated only by real users. However, these approaches can be costly and produce subjective judgments, but recent techniques aim to produce interpretable recommendations instead of providing an explicit explanation in natural language. Examples of research works where this perspective is considered for the fashion domain are presented elsewhere [176, 183].
- *Evaluate generated images*: When generating images, researchers typically measure three complementary metrics [91]: preference (how much each user would be interested in the rec-

ommended items), image quality using the inception score as a proxy, and diversity (where the visual similarity between every pair of returned images is computed). Sometimes qualitative evaluation is also included. For example, Lin et al. [105] show several examples of generated items and discuss their validity. A similar qualitative evaluation is observed in the work of Yang et al. [184].

- *Evaluate social perspectives*: Brand and Gross [21] highlight a very important aspect of today's society—gender equality—in detail, the attention is placed on the price differences that emerge between men's and women's products. Other works (e.g., [134]) argue whether users perceive in the same way recommendations generated by humans or by services exploiting AI, in terms of trustworthiness and acceptance of the recommendations. In a similar line, the diversity of the recommendations (where the product exposure is higher) has been linked to a higher purchase rate and amount, although it may depend on the type of customer [97]. In any case, special care must be taken by automatic fashion RSs to not reinforce social biases and perpetuate long-observed objectifications [157, 161], hence more research on these lines should be derived from the machine learning and RecSys communities.

From the summary presented later in Table 4, we conclude that the most prominent perspective remains the first one (*evaluate the recommendation*). Nonetheless, some recent papers pay attention to other perspectives, such as image generation and social perspectives. This can be observed in the table by checking how many papers report image quality/diversity metrics, or metrics beyond accuracy, such as diversity or some kind of explanation measurements. The general trend, hence, is that offline evaluation with classical metrics (error, classification, and ranking) are prominent, but the community is working toward introducing others, more focused on alternative dimensions, such as business metrics or those related to social and explanation perspectives.

4.3 Available Datasets

Table 5 summarizes a list of the most frequently used datasets, together with information on their availability, size, and the area to which the data pertain. They are discussed in more detail next:

- *Amazon Reviews dataset*: Amazon is the world's largest marketplace for selling several categories of products, including fashion products. The collection consists of about 140M records, including product reviews, ratings, and product information.
- *DeepFashion*: DeepFashion contains more than 800K photos annotated with 50 categories, 1K attributes, and clothing landmarks (each image contains four to eight landmarks), as well as more than 300K image pairs. They are classified according to several contexts, such as the store, street photo, and consumer.
- *Exact Street2Shop*: The images in this collection fall into two categories: (i) street images, which are photographs of individuals wearing various garments and taken in spontaneous, unplanned moments, and (ii) shop photos, which are photographs of clothing products from online clothing businesses
- *ExpFashion*: This dataset was gathered from a collection of Polyvore images. They created new outfits given an item from current outfits and placed them in the dataset, starting with 1K seed outfits and expanding to a total of 100K outfits.
- *Fashion IQ*: This dataset is commonly used as a benchmark for image retrieval with natural language feedback. It contains 77.6K images of fashion products over three categories: Dress, Toptee, and Shirt, with 46.6K images in the training and around 15.5K images for both the validation and test sets.
- *FashionVC*: This is a dataset consisting of the outfits of Polyvore's best fashion experts. Using a seed set of Polyvore popular outfits, the researchers identified 256 fashion experts, then retrieved historical outfits from them.

Table 4. Common Datasets and Evaluation Metrics Used in the Fashion Recommender Literature

Authors	Year	Evaluation										Clothing Type	Datasets
		Type	Metric										
			Offline Metrics				Online/Qual./Business						
			Error	Clf.	Rank	Beyond	Expl.	Survey	Business	Image quality	Image diversity		
Tangseng and Okatani [155]	2020	offline		✓			✓					generic	PO
Lin et al. [106]	2020	offline		✓	✓		✓					top-bottom	FashionVC
Chen et al. [28]	2019	off+online		✓					✓			generic	POG
Polanía and Gupte [129]	2019	offline			✓				✓			generic	PO
Hou et al. [70]	2019	offline		✓	✓							generic+shoe	Amazon
Kumar and Gupta [96]	2019	offline		✓		✓				✓	✓	top-bottom	Bing
Zhou et al. [194]	2019	online										generic	NA
Pan et al. [127]	2019	offline		✓				✓				generic	NA
Lin et al. [105]	2019	offline		✓	✓			✓				top-bottom	ExpFashion
Yin et al. [186]	2019	offline		✓						✓		generic	Amazon
Liu et al. [108]	2019	offline		✓								top-bottom	FarFetch
Abdulla et al. [1]	2019	off+online										generic	NA
Hwangbo et al. [79]	2018	online							✓			generic	NA
Cardoso et al. [22]	2018	offline		✓								generic	ASOS
He and Hu [66]	2018	online									✓	generic	PO
Sun et al. [153]	2018	offline	✓	✓								top-bottom	MouJIE
Cardoso et al. [22]	2018	offline		✓								generic	ASOS
Tuinhof et al. [160]	2018	offline		✓								top-bottom	CrowdFlower
Hidayati et al. [69]	2018	off+online			✓			✓				woman dress	Style4Body
Zhou et al. [195]	2018				✓							top-bottom	DeepFashion
Guigourès et al. [56]	2018	offline		✓		✓						shoes	NA
Agarwal et al. [3]	2018	offline		✓	✓							men t-shirt	Myntra
Heinz et al. [67]	2017	offline		✓								generic	Zalando
Kang et al. [91]	2017	offline		✓				✓		✓	✓	shoe+top-bot.	Amazon
Zhang et al. [191]	2017	offline		✓								–	NA
Jaradat [83]	2017	offline										–	Zalando+Instagram
Qian et al. [131]	2017	offline		✓	✓							top-bottom	–
Zhao et al. [192]	2017	offline		✓								generic	Amazon+Tao
Bracher et al. [20]	2016	online										generic	Zalando
McAuley et al. [120]	2015	offline		✓								–	Amazon
Hu et al. [76]	2015	offline			✓							generic	PO
Wang et al. [167]	2015	online						✓				–	–
Bhardwaj et al. [19]	2014	online						✓				top-bottom	Fashion-136K
Jagadeesh et al. [81]	2014	online						✓				top-bottom	Fashion-136K

Challenges IE: Interpretability/Explanation.
Metrics Err.: Error-based accuracy (RMSE, MAE); Clf.: Classification metrics (Precision, Recall, F1, Accuracy, AUC); RA: Rank-aware accuracy (MRR, MAP, NDCG); Beyond: Beyond accuracy metric (Novelty, Coverage, Diversity); Expl.: Explanation-related evaluation.
Datasets PO: Polyvore; NA: Anonymous.

- *Fashion-136K*: For this dataset, photographs of fashion models were gathered from the Web. All photographs will have both top and bottom garments in the same image.
- *IQON3000*: This dataset is composed of garment images and metadata. Garments are grouped by outfit and are associated to different users. These data were collected from IQON, a Japanese fashion community website, in which members could mix fashion items to create new outfits.
- *Polyvore*: This is a social commerce platform that allows users to exchange items and utilize them in image collages called *Sets*. As a dataset for many fashion recommendation systems (e.g., Hu et al. [76] have defined 150 sets of users), Sets are employed because they include

Table 5. List of the Most Commonly Used Datasets in the Fashion RS Literature

Dataset	Reference	Availability	Size	Area
Amazon Reviews	Link	Yes	140M interactions	Generic
DeepFashion	[113]	No	800K images	Generic
Exact Street2Shop	Link	Yes	-	Street and shop photos
ExpFashion	[106]	No	200K images and 1M comments	Item
Fashion IQ	[175]	Yes	77.6K images and textual descriptions	Dresses, shirts, tops & tees
FashionVC	[150]	No	20K images	Outfits by fashion experts
Fashion-136K	[81]	No	–	Fashion models
IQON3000	[151]	Yes	300K outfits by 3.5K users	Outfits
Polyvore	[76]	No	–	Item photos
Pog	Link	Yes	4M images and descriptions	Item and outfit
StyleReference	[68]	No	2K pictures	Body features
Style4BodyShape	[69]	No	3,150 celebrity records	Female celebrities

only the products and a clean background, which substantially facilitates extraction of object properties.

- *POG*: The POG dataset contains data retrieved from the Taobao website; specifically, the clicks on the most popular items and outfits were extracted, and each one was paired with a record including the image's background, title, and category.
- *StyleReference*: There are 2,160 photos of apparel in this collection. The photographs were sourced from a number of major fashion publications style magazines (e.g., *Elle* and *Vogue*).
- *Style4BodyShape*: In this dataset, it is possible to find three different kinds of information: (i) a list of the most stylish 3,150 female celebrities, who are well known for their refined sense of fashion; (ii) female celebrity body measurements, including dress, bust, waist, and hip measurements, derived from a website that compiles this data; and (iii) images of 270 fashionable celebrities crawled via the Google search engine.

In summary, we note that most of the reported datasets are not publicly available. This is a big hurdle to promote comparable and reproducible research in this domain. Moreover, we observe that the area to which the data belongs to is heterogeneous, as no two datasets share the same area except the three datasets classified as *Generic*. Finally, even though most of these datasets are quite large (reaching millions of interactions), some of them are very small, including few thousands of images, which might be insufficient information for some methods to learn interesting patterns from them.

5 CONCLUSION

In this survey, we examined RS specific to the clothing and fashion sector, introducing a taxonomy that categorizes them by task (e.g., item or outfit recommendations, size, explainability) and side information (users, items, context). We outlined key evaluation goals and metrics, such as outfit generation and compatibility prediction, along with community perspectives. This field offers distinct challenges vital to the success of RSs.

Datasets. We may recall how data collections began with simple "harvested" datasets and progressed to more "curated" datasets in subsequent years (e.g., images from fashion models rather than, e.g., co-purchase data). It is interesting to consider which of these methods is preferable and where the field is headed in terms of datasets. Even though harvested datasets are easy to collect and correspond well to real prediction tasks (e.g., purchase estimate), collected datasets are noisy; they may also not reflect the real semantics of visual preferences, compatibility, or other aspects of a user's experience. In contrast, curated datasets may not match the distribution of real data, or "models" may not represent the preference dimensions of regular users, or the datasets may

actually be contrary to one's goals. For instance, a marketing image may try to promote a pair of shoes by pairing it with a "boring" (non-distracting) outfit, thus curated data may not be any more "real" than harvested data (cf. Section 4.3).

Tasks. Whereas the traditional task in fashion RS research involved purchase or co-purchase prediction tasks, recent systems focus on combinatorial outputs (e.g., outfits and wardrobes) or even generative tasks (cf. Section 2.1). One needs to think whether this complexity is necessary. Do complicated non-pairwise functions actually exist in fashion semantics, or are pairwise functions adequate to represent the underlying decision factors? Is the increased complexity of combinatorial models worth the investment? This may require us to examine the long-term viability of simpler models in comparison to more complicated ones (cf. Section 2.1).

Interpretation. What exactly do visually aware models "learn"? Are they truly capturing fine-grained fashion semantics, or are they simply learning trivial factors (e.g., categories) from fashion data? (cf. Section 2.1.4) Visual RSs are mostly used to cope with noisy, sparse, cold start datasets including visual data, but it is unclear that we have made much progress toward acquiring "real" fashion semantics in the manner of a designer. If our objective is just to predict purchases, this may be irrelevant; if our objective is to fulfill the function of fashion designers, how can we develop more representative datasets or tasks (cf. Section 2.1.4)?

6 FUTURE OUTLOOK

We anticipate that the following areas will receive future attention and be the subject of future surveys.

Forecasting. The majority of fashion recommendations are focused on predicting the "present" based on previous interactions and trends—that is, what the user will do *right now* based on their history. How might such models be utilized to make predictions regarding the fashion trends of the future? This aspect can be linked to popularity forecasting in the fashion domain since trendy items will likely be popular [15, 87].

Ethics. Fashion recommendation raises a number of ethical considerations—for instance, how certain factors like gender [21], race, body shape, or religion, to name a few, result in stereotypical recommendations. A fashion recommendation system, for instance, may be biased toward recommending clothing items that are traditionally worn by men or women, thereby ignoring the fashion tastes of people who do not conform to these traditional gender roles. Besides the ethical considerations, such bias and unfairness can have other severe consequences, including decreased retail sales, dissatisfied customers, and degradation of the brand image (e.g., see [10, 42]). With the goal of rare product discovery, Sá et al. [140] explore how to diversify the recommendations in the context of luxury fashion, whereas Nestler et al. [122] use a Bayesian model to reduce the size- and fit-related product returns, hence reducing the risk of personal issues such as "it's not me" or other body image issues.

The ethics of "fast fashion" is also worth considering. *Fast fashion* is a term that refers to the manufacturing process, which means that the consumer may have to choose between purchasing something cheaper but manufactured under dubious conditions or paying more (or even having a very limited selection) for *sustainable* and *ethically* superior clothing.

Adversarial Robustness and Privacy. This topic revolves around protecting the security of machine learning/RS models and user data. In the fashion recommendation domain, adversarial attacks are a significant concern, as attackers can manipulate the system's output and deceive users [44]. One common form of such an attack is to manipulate the images of clothing items, making them appear different from their actual appearance. Another form of attack involves adding adversarial perturbations (noise) to the images that are invisible to the naked eye and promoting

or demoting certain items or item categories in the recommendation list for purposes such as marketing presentations. Consequently, users may be directed to clothing items they may not have otherwise considered, which could result in dissatisfaction and a reduction in retail sales [11, 159]. An example of an attack is presented by Treu et al. [159], which uses a fashion image to generate an adversarial texture applied to the clothing regions of the target image, preventing person segmentation from working correctly.

Deployability. What are the practical considerations in terms of deploying models from academia? Perhaps complexity and being naturally black box models (relating to the preceding discussion about interpretation) are important aspects for consideration given that the majority of the techniques described are based on neural networks. Furthermore, good performance on a small academic dataset may fail to reproduce at an Internet-size scenario, which compounded with the lack of interpretability may result in unforeseen problems when a new model is released to the public [123, 136]. Properly testing systematically and at scale, and how to better explain and understand model predictions, is thus a relevant area of research.

Multi-Modal and Cross-Modal Representations. A multi-modal fashion RS can be trained to recognize the semantic aspects of clothing items, such as their color, style, and material, and combine this information with visual data, such as images of the clothing items, to improve the product representations and ultimately provide more accurate recommendations [46]. However, it is costly to scale the training data required for all the fashion concepts with industry-level accuracy. However, recent advances in multi-modal text-and-image models such as CLIP [132] or ALBEF [100] have shown that it is possible to achieve a remarkable level of semantic understanding using only weakly related Internet text and image data. Future research will probably target how to use similar approaches for fashion RSs.

Similarly, cross-modal text-to-image generative models, such as DALL-E [133] and Stable Diffusion [139], could be incorporated into fashion AI systems to improve the accuracy of recommendations. These models can be trained to generate images of clothing items based on textual descriptions or retrieve textual descriptions of clothing items based on visual data, which can provide more accurate and diverse recommendations for users.

Iterative and Multi-Modal Recommendation. As conversational systems progress and gain wide adoption, fashion RSs will be able to utilize multi-turn interactions [48, 187], as well as multiple forms of input, to refine and improve the recommendations to the customer. Furthermore, different pieces of information provided by the customer may be contradictory or modify each other, and future fashion RSs will need to exhibit robustness to more informal forms of communication. Several works [50, 172, 177], among others, start to tap into this area, but more research will be needed to make such systems practical and train them at scale.

Virtual Try-on. Augmented reality is becoming commonplace in mobile camera apps, and in consequence it is very likely that customers will start expecting it in e-commerce settings. Some product verticals like beauty or eyewear, where the requirements are more aligned with current commercial augmented reality technology, are early adopters [170]. Clothing virtual try-on will require more research before it becomes useful, but it will create opportunities to make fashion RSs more personalized [88].

Causal Understanding. Another challenge for fashion AI and recommendation is to take a causal approach to fashion recommendation, which entails understanding the causal relationships between data and the effect of features on informed decision making. For instance, a fashion AI model can be trained to understand the causal relationships between a user's clothing preferences and their body type, age, or lifestyle, then use this knowledge to provide personalized recommendations that are more likely to suit the user's needs and preferences.

Integration with Generative AI and Large Language Models. Finally, as generative AI and large language models continue to advance, their roles in RSs—particularly in the field of fashion—are becoming increasingly significant [53, 71, 168]. These models can operate in both uni-modal and multi-modal capacities. For example, in the fashion industry, large language models can generate outfit recommendations by understanding the compatibility between various fashion items in a nuanced manner. In multi-modal scenarios, these models can combine both text and visual data to offer more personalized choices. For instance, a user could interactively decide what they would like to wear by inputting both textual and visual preferences into the system. Moreover, these models can also tackle common challenges in RSs such as sparsity and the cold start problem through prompt engineering techniques such as few-shot and zero-shot learning [39, 121].

REFERENCES

[1] G. Mohammed Abdulla, Shreya Singh, and Sumit Borar. 2019. Shop your right size: A system for recommending sizes for fashion products. In *Companion of the 2019 World Wide Web Conference (WWW'19)*. ACM, New York, NY, 327–334. https://doi.org/10.1145/3308560.3316599

[2] Adewole Adewumi, Adebola Taiwo, Sanjay Misra, Rytis Maskeliunas, Robertas Damasevicius, Ravin Ahuja, and Foluso Ayeni. 2019. A unified framework for outfit design and advice. In *Data Management, Analytics and Innovation*. Advances in Intelligent Systems and Computing, Vol. 1016. Springer, 31–41. https://doi.org/10.1007/978-981-13-9364-8_3

[3] Pankaj Agarwal, Sreekanth Vempati, and Sumit Borar. 2018. Personalizing similar product recommendations in fashion E-commerce. *CoRR abs/1806.11371* (2018). http://arxiv.org/abs/1806.11371

[4] Charu C. Aggarwal. 2016. Ensemble-based and hybrid recommender systems. In *Recommender Systems*. Springer, 199–224.

[5] Divyansh Aggarwal, Elchin Valiyev, Fadime Sener, and Angela Yao. 2018. Learning style compatibility for furniture. In *Pattern Recognition*. Lecture Notes in Computer Science, Vol. 11269. Springer, 552–566. https://doi.org/10.1007/978-3-030-12939-2_38

[6] Kenan E. Ak, Ashraf A. Kassim, Joo Hwee Lim, and Jo Yew Tham. 2018. Learning attribute representations with localization for flexible fashion search. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 7708–7717.

[7] Ziad Al-Halah and Kristen Grauman. 2020. From Paris to Berlin: Discovering fashion style influences around the world. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 10136–10145.

[8] Ziad Al-Halah, Rainer Stiefelhagen, and Kristen Grauman. 2017. Fashion forward: Forecasting visual style in fashion. In *Proceedings of the IEEE International Conference on Computer Vision*. 388–397.

[9] Mohammed Al-Rawi and Joeran Beel. 2021. Probabilistic color modelling of clothing items. *Recommender Systems in Fashion and Retail* 734 (2021), 21.

[10] Enrique Amigó, Yashar Deldjoo, Stefano Mizzaro, and Alejandro Bellogín. 2023. A unifying and general account of fairness measurement in recommender systems. *Information Processing & Management* 60, 1 (2023), 103115.

[11] Vito Walter Anelli, Yashar Deldjoo, Tommaso Di Noia, Daniele Malitesta, and Felice Antonio Merra. 2021. A study of defensive methods to protect visual recommendation against adversarial manipulation of images. In *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 1094–1103.

[12] Phoebe R. Apeagyei. 2010. Application of 3D body scanning technology to human measurement for clothing fit. *International Journal of Digital Content Technology and Its Applications* 4, 7 (2010), 58–68.

[13] Susan P. Ashdown, Suzanne Loker, Katherine Schoenfelder, and Lindsay Lyman-Clarke. 2004. Using 3D scans for fit analysis. *Journal of Textile and Apparel, Technology and Management* 4, 1 (2004), 1–12.

[14] Krisztian Balog and Filip Radlinski. 2020. Measuring recommendation explanation quality: The conflicting goals of explanations. In *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*. 329–338.

[15] Roja Bandari, Sitaram Asur, and Bernardo Huberman. 2021. The pulse of news in social media: Forecasting popularity. *Proceedings of the International AAAI Conference on Web and Social Media* 6, 1 (Aug. 2021), 26–33. https://ojs.aaai.org/index.php/ICWSM/article/view/14261

[16] Debopriyo Banerjee, Krothapalli Sreenivasa Rao, Shamik Sural, and Niloy Ganguly. 2020. BOXREC: Recommending a box of preferred outfits in online shopping. *ACM Transactions on Intelligent Systems and Technology* 11, 6 (2020), Article 69, 28 pages. https://doi.org/10.1145/3408890

[17] Sean Bell and Kavita Bala. 2015. Learning visual similarity for product design with convolutional neural networks. *ACM Transactions on Graphics* 34, 4 (2015), Article 98, 10 pages. https://doi.org/10.1145/2766959

[18] Elaine M. Bettaney, Stephen R. Hardwick, Odysseas Zisimopoulos, and Benjamin Paul Chamberlain. 2019. Fashion outfit generation for E-commerce. *arXiv preprint arXiv:1904.00741* (2019).
[19] Anurag Bhardwaj, Vignesh Jagadeesh, Wei Di, Robinson Piramuthu, and Elizabeth F. Churchill. 2014. Enhancing visual fashion recommendations with users in the loop. *CoRR abs/1405.4013* (2014). http://arxiv.org/abs/1405.4013
[20] Christian Bracher, Sebastian Heinz, and Roland Vollgraf. 2016. Fashion DNA: Merging content and sales data for recommendation and article mapping. *CoRR abs/1609.02489* (2016).
[21] Alexander Brand and Tom Gross. 2020. Paying the pink tax on a blue dress—Exploring gender-based price-premiums in fashion recommendations. In *Human-Centered Software Engineering*. Lecture Notes in Computer Science, Vol. 12481. Springer, 190–198. https://doi.org/10.1007/978-3-030-64266-2_12
[22] Ângelo Cardoso, Fabio Daolio, and Saúl Vargas. 2018. Product characterisation towards personalisation: Learning attributes from unstructured data to recommend fashion products. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'18)*. ACM, New York, NY,, 80–89. https://doi.org/10.1145/3219819.3219888
[23] Yeongnam Chae, Jiu Xu, Björn Stenger, and Soh Masuko. 2018. Color navigation by qualitative attributes for fashion recommendation. In *Proceedings of the IEEE International Conference on Consumer Electronics (ICCE'18)*. IEEE, Los Alamitos, CA, 1–3. https://doi.org/10.1109/ICCE.2018.8326138
[24] Yu-Ting Chang, Wen-Haung Cheng, Bo Wu, and Kai-Lung Hua. 2017. Fashion world map: Understanding cities through streetwear fashion. In *Proceedings of the 25th ACM International Conference on Multimedia*. 91–99.
[25] Huizhong Chen, Andrew Gallagher, and Bernd Girod. 2012. Describing clothing by semantic attributes. In *Proceedings of the European Conference on Computer Vision*. 609–623.
[26] Jingyuan Chen, Hanwang Zhang, Xiangnan He, Liqiang Nie, Wei Liu, and Tat-Seng Chua. 2017. Attentive collaborative filtering: Multimedia recommendation with item- and component-level attention. In *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR'17)*. ACM, New York, NY, 335–344.
[27] Wen Chen, Pipei Huang, Jiaming Xu, Xin Guo, Cheng Guo, Fei Sun, Chao Li, Andreas Pfadler, Huan Zhao, and Binqiang Zhao. 2019. POG: Personalized outfit generation for fashion recommendation at Alibaba iFashion. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'19)*. 2662–2670.
[28] Wen Chen, Pipei Huang, Jiaming Xu, Xin Guo, Cheng Guo, Fei Sun, Chao Li, Andreas Pfadler, Huan Zhao, and Binqiang Zhao. 2019. POG: Personalized outfit generation for fashion recommendation at Alibaba iFashion. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'19)*. 2662–2670. https://doi.org/10.1145/3292500.3330652
[29] Xu Chen, Hanxiong Chen, Hongteng Xu, Yongfeng Zhang, Yixin Cao, Zheng Qin, and Hongyuan Zha. 2019. Personalized fashion recommendation with visual explanations based on multimodal attention network: Towards visually explainable recommendation. In *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR'19)*. ACM, New York, NY, 765–774. https://doi.org/10.1145/3331184.3331254
[30] Wen-Huang Cheng, Sijie Song, Chieh-Yun Chen, Shintami Chusnul Hidayati, and Jiaying Liu. 2021. Fashion meets computer vision: A survey. *ACM Computing Surveys* 54, 4 (2021), 1–41.
[31] Zhiyong Cheng, Xiaojun Chang, Lei Zhu, Rose C. Kanjirathinkal, and Mohan Kankanhalli. 2019. MMALFM: Explainable recommendation by leveraging reviews and images. *ACM Transactions on Information Systems* 37, 2 (2019), 1–28.
[32] Sameer Chhabra. 2017. Netflix says 80 percent of watched content is based on algorithmic recommendations. *MobileSyrup*. Retrieved March 13, 2021 from https://mobilesyrup.com/2017/08/22/80-percent-netflix-shows-discovered-recommendation/
[33] Patrick John Chia, Jacopo Tagliabue, Federico Bianchi, Ciro Greco, and Diogo Gonçalves. 2022. "Does it come in black?" CLIP-like models are zero-shot recommenders. *CoRR abs/2204.02473* (2022). https://doi.org/10.48550/arXiv.2204.02473
[34] Kyung-Yong Chung. 2014. Effect of facial makeup style recommendation on visual sensibility. *Multimedia Tools and Applications* 71, 2 (2014), 843–853. https://doi.org/10.1007/s11042-013-1355-6
[35] Jon Cilley. 2016. *Apparel & Footwear Retail Survey Report*. Body Labs Inc., New York, NY.
[36] Humberto Corona. 2020. The State of Recommender Systems for Fashion in 2020. Retrieved January 19, 2021 from https://towardsdatascience.com/the-state-of-recommender-systems-for-fashion-in-2020-180b3ddb392f/
[37] Hein A. M. Daanen and Agnes Psikuta. 2018. 3D body scanning. In *Automation in Garment Manufacturing*. Elsevier, 237–252.
[38] Lavinia De Divitiis, Federico Becattini, Claudio Baecchi, and Alberto Del Bimbo. 2023. Disentangling features for fashion recommendation. *ACM Transactions on Multimedia Computing, Communications and Applications* 19, 1s (2023), 1–21.

[39] Yashar Deldjoo. 2023. Fairness of ChatGPT and the role of explainable-guided prompts. *CoRR abs/2307.11761* (2023). https://doi.org/10.48550/arXiv.2307.11761

[40] Yashar Deldjoo, Tommaso Di Noia, Danielle Malitesta, and Felice Merra, Antonio. 2022. Leveraging content-style item representation for visual recommendation. In *Advances in Information Retrieval*. Lecture Notes in Computer Science, Vol. 13186. Springer, 84–92.

[41] Yashar Deldjoo, Tommaso Di Noia, Daniele Malitesta, and Felice Antonio Merra. 2021. A study on the relative importance of convolutional neural networks in visually-aware recommender systems. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 3961–3967.

[42] Yashar Deldjoo, Dietmar Jannach, Alejandro Bellogin, Alessandro Diffonzo, and Dario Zanzonelli. 2023. Fairness in recommender systems: Research landscape and future directions. *User Modeling and User-Adapted Interaction*. Open access, April 24, 2023.

[43] Yashar Deldjoo, Tommaso Di Noia, and Felice Antonio Merra. 2021. A survey on adversarial recommender systems: From attack/defense strategies to generative adversarial networks. *ACM Computing Surveys* 54, 2 (2021), 1–38. https://doi.org/10.1145/3439729

[44] Yashar Deldjoo, Tommaso Di Noia, and Felice Antonio Merra. 2021. A survey on adversarial recommender systems: From attack/defense strategies to generative adversarial networks. *ACM Computing Surveys* 54, 2 (2021), 1–38.

[45] Yashar Deldjoo, Markus Schedl, Paolo Cremonesi, and Gabriella Pasi. 2020. Recommender systems leveraging multimedia content. *ACM Computing Surveys* 53, 5 (2020), Article 106, 38 pages. https://doi.org/10.1145/3407190

[46] Yashar Deldjoo, Markus Schedl, Balázs Hidasi, Yinwei Wei, and Xiangnan He. 2021. Multimedia recommender systems: Algorithms and challenges. In *Recommender Systems Handbook*. Springer, 973–1014.

[47] Yashar Deldjoo, Markus Schedl, and Peter Knees. 2021. Content-driven music recommendation: Evolution, state of the art, and challenges. *arXiv preprint arXiv:2107.11803* (2021).

[48] Yashar Deldjoo, Johanne R. Trippas, and Hamed Zamani. 2021. Towards multi-modal conversational information seeking. In *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 1577–1587.

[49] Ginger Delmas, Rafael Sampaio de Rezende, Gabriela Csurka, and Diane Larlus. 2022. ARTEMIS: Attention-based retrieval with text-explicit matching and implicit similarity. In *Proceedings of the 10th International Conference on Learning Representations (ICLR'22)*. https://openreview.net/forum?id=CVfLvQq9gLo

[50] Ginger Delmas, Rafael S. Rezende, Gabriela Csurka, and Diane Larlus. 2022. ARTEMIS: Attention-based retrieval with text-explicit matching and implicit similarity. In *Proceedings of the International Conference on Learning Representations*.

[51] Yujuan Ding, Yunshan Ma, Lizi Liao, Waikeung Wong, and Tat-Seng Chua. 2021. Leveraging multiple relations for fashion trend forecasting based on social media. *IEEE Transactions on Multimedia* 24 (2021), 2287–2299.

[52] Min Dong, Xianyi Zeng, Ludovic Koehl, and Junjie Zhang. 2020. An interactive knowledge-based recommender system for fashion product design in the big data environment. *Information Sciences* 540 (2020), 469–488. https://doi.org/10.1016/j.ins.2020.05.094

[53] Wenqi Fan, Zihuai Zhao, Jiatong Li, Yunqing Liu, Xiaowei Mei, Yiqi Wang, Jiliang Tang, and Qing Li. 2023. Recommender systems in the era of large language models (llms). *arXiv preprint arXiv:2307.02046* (2023).

[54] Beatriz Quintino Ferreira, Joao P. Costeira, Ricardo G. Sousa, Liang-Yan Gui, and Joao P. Gomes. 2019. Pose guided attention for multi-label fashion image classification. In *Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW'19)*. IEEE, Los Alamitos, CA, 3125–3128.

[55] Takao Furukawa, Chikako Miura, Kaoru Mori, Sou Uchida, and Makoto Hasegawa. 2019. Visualisation for analysing evolutionary dynamics of fashion trends. *International Journal of Fashion Design, Technology and Education* 12, 2 (2019), 247–259.

[56] Romain Guigourès, Yuen King Ho, Evgenii Koriagin, Abdul-Saboor Sheikh, Urs Bergmann, and Reza Shirvany. 2018. A hierarchical Bayesian model for size recommendation in fashion. In *Proceedings of the 12th ACM Conference on Recommender Systems (RecSys'18)*. ACM, New York, NY, 392–396.

[57] M. Hadi Kiapour, Xufeng Han, Svetlana Lazebnik, Alexander C. Berg, and Tamara L. Berg. 2015. Where to buy it: Matching street clothing photos in online shops. In *Proceedings of the IEEE International Conference on Computer Vision*. 3343–3351.

[58] Xianjing Han, Xuemeng Song, Jianhua Yin, Yinglong Wang, and Liqiang Nie. 2019. Prototype-guided attribute-wise interpretable scheme for clothing matching. In *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval*. 785–794.

[59] Xintong Han, Zuxuan Wu, Yu-Gang Jiang, and Larry S. Davis. 2017. Learning fashion compatibility with bidirectional LSTMs. In *Proceedings of the 25th ACM International Conference on Multimedia (MM'17)*. ACM, New York, NY, 1078–1086. https://doi.org/10.1145/3123266.3123394

[60] Xintong Han, Zuxuan Wu, Yu-Gang Jiang, and Larry S. Davis. 2017. Learning fashion compatibility with bidirectional lstms. In *Proceedings of the 25th ACM International Conference on Multimedia*. ACM, New York, NY, 1078–1086.

[61] Kota Hara, Vignesh Jagadeesh, and Robinson Piramuthu. 2016. Fashion apparel detection: The role of deep convolutional neural network and pose-dependent priors. In *Proceedings of the 2016 IEEE Winter Conference on Applications of Computer Vision (WACV'16)*. IEEE, Los Alamitos, CA, 1–9.

[62] Ruining He and Julian McAuley. 2016. VBPR: Visual Bayesian personalized ranking from implicit feedback. In *Proceedings of the 30th AAAI Conference on Artificial Intelligence*.

[63] Ruining He and Julian J. McAuley. 2016. Ups and downs: Modeling the visual evolution of fashion trends with one-class collaborative filtering. In *Proceedings of the 25th International Conference on World Wide Web (WWW'16)*. ACM, New York, NY, 507–517. https://doi.org/10.1145/2872427.2883037

[64] Ruining He and Julian J. McAuley. 2016. VBPR: Visual Bayesian personalized ranking from implicit feedback. In *Proceedings of the 30th AAAI Conference on Artificial Intelligence*. 144–150. http://www.aaai.org/ocs/index.php/AAAI/AAAI16/paper/view/11914

[65] Ruining He, Charles Packer, and Julian J. McAuley. 2016. Learning compatibility across categories for heterogeneous item recommendation. In *Proceedings of the IEEE 16th International Conference on Data Mining (IDCM'16)*. IEEE, Los Alamitos, CA, 937–942. https://doi.org/10.1109/ICDM.2016.0116

[66] Tong He and Yang Hu. 2018. FashionNet: Personalized outfit recommendation with deep neural network. *CoRR abs/1810.02443* (2018). http://arxiv.org/abs/1810.02443

[67] Sebastian Heinz, Christian Bracher, and Roland Vollgraf. 2017. An LSTM-based dynamic customer model for fashion recommendation. In *RecTemp@RecSys*. CEUR Workshop Proceedings, Vol. 1922. CEUR, 45–49.

[68] Shintami Chusnul Hidayati, Ting Wei Goh, Ji-Sheng Gary Chan, Cheng-Chun Hsu, John See, Lai-Kuan Wong, Kai-Lung Hua, Yu Tsao, and Wen-Huang Cheng. 2021. Dress with style: Learning style from joint deep embedding of clothing styles and body shapes. *IEEE Transactions on Multimedia* 23 (2021), 365–377. https://doi.org/10.1109/TMM.2020.2980195

[69] Shintami Chusnul Hidayati, Cheng-Chun Hsu, Yu-Ting Chang, Kai-Lung Hua, Jianlong Fu, and Wen-Huang Cheng. 2018. What dress fits me best?: Fashion recommendation on the clothing style for personal body shape. In *Proceedings of the 26th ACM International Conference on Multimedia (MM'18)*. ACM, New York, NY, 438–446.

[70] Min Hou, Le Wu, Enhong Chen, Zhi Li, Vincent W. Zheng, and Qi Liu. 2019. Explainable fashion recommendation: A semantic attribute region guided approach. In *Proceedings of the 28th International Joint Conference on Artificial Intelligence (IJCAI'19)*. 4681–4688. https://doi.org/10.24963/ijcai.2019/650

[71] Yupeng Hou, Junjie Zhang, Zihan Lin, Hongyu Lu, Ruobing Xie, Julian McAuley, and Wayne Xin Zhao. 2023. Large language models are zero-shot rankers for recommender systems. *arXiv preprint arXiv:2305.08845* (2023).

[72] Wei-Lin Hsiao and Kristen Grauman. 2020. ViBE: Dressing for diverse body shapes. In *Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR'20)*. IEEE, Los Alamitos, CA, 11056–11066. https://doi.org/10.1109/CVPR42600.2020.01107

[73] Wei-Lin Hsiao and Kristen Grauman. 2017. Learning the "latent look": Unsupervised discovery of a style-coherent embedding from fashion images. In *Proceedings of the IEEE International Conference on Computer Vision*. 4203–4212.

[74] Wei-Lin Hsiao and Kristen Grauman. 2018. Creating capsule wardrobes from fashion images. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 7161–7170.

[75] Wei-Lin Hsiao and Kristen Grauman. 2021. From culture to clothing: Discovering the world events behind a century of fashion images. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 1066–1075.

[76] Yang Hu, Xi Yi, and Larry S. Davis. 2015. Collaborative fashion recommendation: A functional tensor factorization approach. In *Proceedings of the 23rd ACM International Conference on Multimedia (MM'15)*. ACM, New York, NY, 129–138.

[77] Fu-Hsien Huang, Hsin-Min Lu, and Yao-Wen Hsu. 2021. From street photos to fashion trends: Leveraging user-provided noisy labels for fashion understanding. *IEEE Access* 9 (2021), 49189–49205.

[78] Cong Phuoc Huynh, Arridhana Ciptadi, Ambrish Tyagi, and Amit Agrawal. 2018. CRAFT: Complementary recommendation by adversarial feature transform. In *Computer Vision—ECCV 2018 Workshops*. Lecture Notes in Computer Science, Vol. 11131. Springer, 54–66.

[79] Hyunwoo Hwangbo, Yang Sok Kim, and Kyung Jin Cha. 2018. Recommendation system development for fashion retail e-commerce. *Electronic Commerce Research and Applications* 28 (2018), 94–101. https://doi.org/10.1016/j.elerap.2018.01.012

[80] Tomoharu Iwata, Shinji Watanabe, and Hiroshi Sawada. 2011. Fashion coordinates recommender system using photographs from fashion magazines. In *Proceedings of the 22nd International Joint Conference on Artificial Intelligence*.

[81] Vignesh Jagadeesh, Robinson Piramuthu, Anurag Bhardwaj, Wei Di, and Neel Sundaresan. 2014. Large scale visual recommendations from street fashion images. In *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'14)*. ACM, New York, NY, 1925–1934. https://doi.org/10.1145/2623330.2623332

[82] Surgan Jandial, Pinkesh Badjatiya, Pranit Chawla, Ayush Chopra, Mausoom Sarkar, and Balaji Krishnamurthy. 2022. SAC: Semantic attention composition for text-conditioned image retrieval. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV'22)*. IEEE, Los Alamitos, CA, 597–606. https://doi.org/10.1109/WACV51458.2022.00067

[83] Shatha Jaradat. 2017. Deep cross-domain fashion recommendation. In *Proceedings of the 11th ACM Conference on Recommender Systems (RecSys'17)*. ACM, New York, NY, 407–410. https://doi.org/10.1145/3109859.3109861

[84] Shatha Jaradat, Nima Dokoohaki, Humberto Jesús Corona Pampín, and Reza Shirvany. 2020. Second workshop on recommender systems in Fashion—FashionXrecsys2020. In *Proceedings of the 14th ACM Conference on Recommender Systems*. 632–634.

[85] Shatha Jaradat, Nima Dokoohaki, Kim Hammar, Ummul Wara, and Mihhail Matskin. 2018. Dynamic CNN models for fashion recommendation in Instagram. In *Proceedings of the 2018 IEEE International Conference on Parallel and Distributed Processing with Applications, Ubiquitous Computing and Communications, Big Data and Cloud Computing, Social Computing and Networking, and Sustainable Computing and Communications (ISPA/IUCC/BDCloud/SocialCom/SustainCom'18)*. IEEE, Los Alamitos, CA, 1144–1151.

[86] Shatha Jaradat, Guy Shani, Humberto Jesus, Corona Pampın, and Reza Shirvany. 2022. Fashion recommender systems. In *Recommender Systems Handbook*. Springer, 1015–1055.

[87] Amin Javari and Mahdi Jalili. 2014. Accurate and novel recommendations: An algorithm based on popularity forecasting. *ACM Transactions on Intelligent Systems and Technology* 5, 4 (Dec. 2014), Article 56, 20 pages. https://doi.org/10.1145/2668107

[88] Chamodi Jayamini, Asangika Sandamini, Thisara Pannala, Prabhash Kumarasinghe, Dushani Perera, and Kasun Karunanayaka. 2021. The use of augmented reality to deliver enhanced user experiences in fashion industry. In *Human-Computer Interaction—INTERACT 2021*. Lecture Notes in Computer Science, Vol. 12936. Springer, 1–11.

[89] Peiguang Jing, Kai Cui, Weili Guan, Liqiang Nie, and Yuting Su. 2023. Category-aware multimodal attention network for fashion compatibility modeling. *IEEE Transactions on Multimedia*. Early access, February 20, 2023.

[90] Marius Kaminskas and Francesco Ricci. 2012. Contextual music information retrieval and recommendation: State of the art and challenges. *Computer Science Review* 6, 2–3 (2012), 89–119. https://doi.org/10.1016/j.cosrev.2012.04.002

[91] Wang-Cheng Kang, Chen Fang, Zhaowen Wang, and Julian J. McAuley. 2017. Visually-aware fashion recommendation and design with generative image models. In *Proceedings of the 2017 IEEE International Conference on Data Mining (ICDM'17)*. 207–216. https://doi.org/10.1109/ICDM.2017.30

[92] Hirokatsu Kataoka, Yutaka Satoh, Kaori Abe, Munetaka Minoguchi, and Akio Nakamura. 2019. Ten-million-order human database for world-wide fashion culture analysis. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*.

[93] M. Hadi Kiapour, Kota Yamaguchi, Alexander C. Berg, and Tamara L. Berg. 2014. Hipster wars: Discovering elements of fashion styles. In *Proceedings of the European Conference on Computer Vision*. 472–488.

[94] Daejin Kim, Wooyeol Ryu, Seungho Lee, Byungsu Lim, and Sangjun Lee. 2016. A Unity3D-based mobile fashion coordination system. *International Journal of Advanced Media and Communication* 6, 1 (2016), 86–92.

[95] Yehuda Koren, Robert M. Bell, and Chris Volinsky. 2009. Matrix factorization techniques for recommender systems. *IEEE Computer* 42, 8 (2009), 30–37. https://doi.org/10.1109/MC.2009.263

[96] Sudhir Kumar and Mithun Das Gupta. 2019. c^+GAN: Complementary fashion item recommendation. In *Proceedings of the Workshop on AI for Fashion (KDD'19)*.

[97] Hyokmin Kwon, Jaeho Han, and Kyungsik Han. 2020. ART (Attractive Recommendation Tailor): How the diversity of product recommendations affects customer purchase preference in fashion industry? In *Proceedings of the 29th ACM International Conference on Information and Knowledge Management (CIKM'20)*. ACM, New York, NY, 2573–2580. https://doi.org/10.1145/3340531.3412687

[98] Seungmin Lee, Dongwan Kim, and Bohyung Han. 2021. CoSMo: Content-style modulation for image retrieval with text feedback. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR'21)*. IEEE, Los Alamitos, CA, 802–812. https://doi.org/10.1109/CVPR46437.2021.00086

[99] Eileen Li, Eric Kim, Andrew Zhai, Josh Beal, and Kunlong Gu. 2020. Bootstrapping complete the look at Pinterest. In *Proceedings of the 26th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD'20)*. ACM, New York, NY, 3299–3307. https://doi.org/10.1145/3394486.3403382

[100] Junnan Li, Ramprasaath Selvaraju, Akhilesh Gotmare, Shafiq Joty, Caiming Xiong, and Steven Chu Hong Hoi. 2021. Align before fuse: Vision and language representation learning with momentum distillation. *Advances in Neural Information Processing Systems* 34 (2021), 9694–9705.

[101] Xingchen Li, Xiang Wang, Xiangnan He, Long Chen, Jun Xiao, and Tat-Seng Chua. 2020. Hierarchical fashion graph network for personalized outfit recommendation. In *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR'20)*. ACM, New York, NY, 159–168. https://doi.org/10.1145/3397271.3401080

[102] Yuncheng Li, Liangliang Cao, Jiang Zhu, and Jiebo Luo. 2017. Mining fashion outfit composition using an end-to-end deep learning approach on set data. *IEEE Transactions on Multimedia* 19, 8 (2017), 1946–1955. https://doi.org/10.1109/TMM.2017.2690144

[103] Lizi Liao, Xiangnan He, Bo Zhao, Chong-Wah Ngo, and Tat-Seng Chua. 2018. Interpretable multimodal retrieval for fashion products. In *Proceedings of the 26th ACM International Conference on Multimedia*. 1571–1579.

[104] Yusan Lin, Maryam Moosaei, and Hao Yang. 2020. OutfitNet: Fashion outfit recommendation with attention-based multiple instance learning. In *Proceedings of The Web Conference 2020 (WWW'20)*. ACM, New York, NY, 77–87. https://doi.org/10.1145/3366423.3380096

[105] Yujie Lin, Pengjie Ren, Zhumin Chen, Zhaochun Ren, Jun Ma, and Maarten de Rijke. 2019. Improving outfit recommendation with co-supervision of fashion generation. In *Proceedings of the World Wide Web Conference (WWW'19)*. 1095–1105. https://doi.org/10.1145/3308558.3313614

[106] Yujie Lin, Pengjie Ren, Zhumin Chen, Zhaochun Ren, Jun Ma, and Maarten de Rijke. 2020. Explainable outfit recommendation with joint outfit matching and comment generation. *IEEE Transactions on Knowledge and Data Engineering* 32, 8 (2020), 1502–1516. https://doi.org/10.1109/TKDE.2019.2906190

[107] Yen-Liang Lin, Son Tran, and Larry S. Davis. 2020. Fashion outfit complementary item retrieval. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 3311–3319.

[108] Luyao Liu, Xingzhong Du, Lei Zhu, Fumin Shen, and Zi Huang. 2018. Learning discrete hashing towards efficient fashion recommendation. *Data Science and Engineering* 3, 4 (2018), 307–322. https://doi.org/10.1007/s41019-018-0079-z

[109] Luoqi Liu, Hui Xu, Si Liu, Junliang Xing, Xi Zhou, and Shuicheng Yan. 2013. "Wow! You are so beautiful today!"In *Proceedings of the 21st ACM International Conference on Multimedia (MM'13)*. 3–12.

[110] Qiang Liu, Shu Wu, and Liang Wang. 2017. DeepStyle: Learning user preferences for visual recommendation. In *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR'17)*. ACM, New York, NY, 841–844.

[111] Si Liu, Zheng Song, Guangcan Liu, Changsheng Xu, Hanqing Lu, and Shuicheng Yan. 2012. Street-to-shop: Cross-scenario clothing retrieval via parts alignment and auxiliary set. In *Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, Los Alamitos, CA, 3330–3337.

[112] Yujie Liu, Yongbiao Gao, Shihe Feng, and Zongmin Li. 2017. Weather-to-garment: Weather-oriented clothing recommendation. In *Proceedings of the 2017 IEEE International Conference on Multimedia and Expo (ICME'17)*. IEEE, Los Alamitos, CA, 181–186.

[113] Ziwei Liu, Ping Luo, Shi Qiu, Xiaogang Wang, and Xiaoou Tang. 2016. DeepFashion: Powering robust clothes recognition and retrieval with rich annotations. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 1096–1104.

[114] Ling Lo, Chia-Lin Liu, Rong-An Lin, Bo Wu, Hong-Han Shuai, and Wen-Huang Cheng. 2019. Dressing for attention: Outfit based fashion popularity prediction. In *Proceedings of the 2019 IEEE International Conference on Image Processing (ICIP'19)*. IEEE, Los Alamitos, CA, 3222–3226.

[115] Zhi Lu, Yang Hu, Yan Chen, and Bing Zeng. 2021. Personalized outfit recommendation with learnable anchors. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 12722–12731.

[116] Zhi Lu, Yang Hu, Yunchao Jiang, Yan Chen, and Bing Zeng. 2019. Learning binary code for personalized fashion recommendation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR'19)*. IEEE, Los Alamitos, CA, 10562–10570. https://doi.org/10.1109/CVPR.2019.01081

[117] Ian MacKenzie, Chris Meyer, and Steve Noble. 2013. How retailers can keep up with consumers. *McKinsey & Company*. Retrieved September 28, 2023 from https://www.mckinsey.com/industries/retail/our-insights/how-retailers-can-keep-up-with-consumers

[118] Utkarsh Mall, Kevin Matzen, Bharath Hariharan, Noah Snavely, and Kavita Bala. 2019. GeoStyle: Discovering fashion trends and events. In *Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision (ICCV'19)*. IEEE, Los Alamitos, CA, 411–420. https://doi.org/10.1109/ICCV.2019.00050

[119] Kevin Matzen, Kavita Bala, and Noah Snavely. 2017. StreetStyle: Exploring world-wide clothing styles from millions of photos. *arXiv preprint arXiv:1706.01869* (2017).

[120] Julian J. McAuley, Christopher Targett, Qinfeng Shi, and Anton van den Hengel. 2015. Image-based recommendations on styles and substitutes. In *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 43–52. https://doi.org/10.1145/2766462.2767755

[121] Fatemeh Nazary, Yashar Deldjoo, and Tommaso Di Noia. 2023. ChatGPT-HealthPrompt: Harnessing the power of XAI in prompt-based healthcare decision support using ChatGPT. *CoRR abs/2308.09731* (2023). https://doi.org/10.48550/arXiv.2308.09731

[122] Andrea Nestler, Nour Karessli, Karl Hajjar, Rodrigo Weffer, and Reza Shirvany. 2021. SizeFlags: Reducing size and fit related returns in fashion E-commerce. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD'21)*. ACM, New York, NY, 3432–3440. https://doi.org/10.1145/3447548.3467160

[123] Richard Nieva. 2015. Google apologizes for algorithm mistakenly calling black people 'gorillas.' *CNET*. Retrieved September 28, 2023 from https://www.cnet.com/tech/services-and-software/google-apologizes-for-algorithm-mistakenly-calling-black-people-gorillas/?lang=he

[124] Augustus Odena, Christopher Olah, and Jonathon Shlens. 2017. Conditional image synthesis with auxiliary classifier GANs. *Proceedings of Machine Learning Research* 70 (2017), 2642–2651.

[125] Ioannis Pachoulakis and Kostas Kapetanakis. 2012. Augmented reality platforms for virtual fitting rooms. *International Journal of Multimedia & Its Applications* 4, 4 (2012), 35.

[126] Charles Packer, Julian McAuley, and Arnau Ramisa. 2018. Visually-aware personalized recommendation using interpretable image representations. In *Proceedings of the 3rd International Workshop on Fashion and KDD (Held in Conjunction with KDD'18)*.

[127] Tse-Yu Pan, Yi-Zhu Dai, Min-Chun Hu, and Wen-Huang Cheng. 2019. Furniture style compatibility recommendation with cross-class triplet loss. *Multimedia Tools and Applications* 78, 3 (2019), 2645–2665. https://doi.org/10.1007/s11042-018-5747-5

[128] Jacqueline Parr and Sanjukta Pookulangara. 2017. The impact of true fit technology on consumer confidence in their online clothing purchase. In *Proceedings of the International Textile and Apparel Association Annual Conference*, Vol. 74.

[129] Luisa F. Polanía and Satyajit Gupte. 2019. Learning fashion compatibility across apparel categories for outfit recommendation. In *Proceedings of the 2019 IEEE International Conference on Image Processing (ICIP'19)*. 4489–4493. https://doi.org/10.1109/ICIP.2019.8803587

[130] Xueming Qian, He Feng, Guoshuai Zhao, and Tao Mei. 2014. Personalized recommendation combining user interest and social circle. *IEEE Transactions on Knowledge and Data Engineering* 26, 7 (2014), 1763–1777. https://doi.org/10.1109/TKDE.2013.168

[131] Yu Qian, P. Giaccone, Michele Sasdelli, Eduard Vasquez, and Biswa Sengupta. 2017. Algorithmic clothing: Hybrid recommendation, from street-style-to-shop. *CoRR abs/1705.09451* (2017). http://arxiv.org/abs/1705.09451

[132] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen Krueger, and Ilya Sutskever. 2021. Learning transferable visual models from natural language supervision. In *Proceedings of the International Conference on Machine Learning*. 8748–8763.

[133] Aditya Ramesh, Prafulla Dhariwal, Alex Nichol, Casey Chu, and Mark Chen. 2022. Hierarchical text-conditional image generation with clip latents. *arXiv preprint arXiv:2204.06125* (2022).

[134] Pei-Luen Patrick Rau, Ziyang Li, and Dinglong Huang. 2020. Who should provide clothing recommendation services: Artificial intelligence or human experts? *Journal of Information Technology Research* 13, 3 (July 2020), 113–125. https://doi.org/10.4018/JITR.2020070107

[135] Abhinav Ravi, Sandeep Repakula, Ujjal Kr Dutta, and Maulik Parmar. 2021. Buy me that look: An approach for recommending similar fashion products. In *Proceedings of the 2021 IEEE 4th International Conference on Multimedia Information Processing and Retrieval (MIPR'21)*. 97–103. https://doi.org/10.1109/MIPR51284.2021.00022

[136] Hope Reese. 2016. Why Microsoft's TayAI bot went wrong. *TechRepublic*. Retrieved September 28, 2023 from https://www.techrepublic.com/article/why-microsofts-tay-ai-bot-went-wrong/

[137] Steffen Rendle, Christoph Freudenthaler, Zeno Gantner, and Lars Schmidt-Thieme. 2009. BPR: Bayesian personalized ranking from implicit feedback. In *Proceedings of the 25th Conference on Uncertainty in Artificial Intelligence (UAI'09)*. 452–461.

[138] Steffen Rendle, Christoph Freudenthaler, Zeno Gantner, and Lars Schmidt-Thieme. 2009. BPR: Bayesian personalized ranking from implicit feedback. In *Proceedings of the 25th Conference on Uncertainty in Artificial Intelligence (UAI'09)*. 452–461.

[139] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. 2021. High-resolution image synthesis with latent diffusion models. *arXiv:2112.10752* (2021). https://doi.org/10.48550/ARXIV.2112.10752

[140] João Sá, Vanessa Queiroz Marinho, Ana Rita Magalhães, Tiago Lacerda, and Diogo Gonçalves. 2022. Diversity vs relevance: A practical multi-objective study in luxury fashion recommendations. In *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR'22)*. ACM, New York, NY, 2405–2409. https://doi.org/10.1145/3477495.3531866

[141] Tim Salimans, Ian J. Goodfellow, Wojciech Zaremba, Vicki Cheung, Alec Radford, and Xi Chen. 2016. Improved techniques for training GANs. In *Proceedings of the 30th International Conference on Neural Information Processing Systems (NIPS'16)*. 2226–2234.

[142] Rohan Sarkar, Navaneeth Bodla, Mariya I. Vasileva, Yen-Liang Lin, Anurag Beniwal, Alan Lu, and Gerard Medioni. 2023. OutfitTransformer: Learning outfit representations for fashion recommendation. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*. 3601–3609.

[143] Hosnieh Sattar, Gerard Pons-Moll, and Mario Fritz. 2019. Fashion is taking shape: Understanding clothing preference based on body shape from online sources. In *Proceedings of the 2019 IEEE Winter Conference on Applications of Computer Vision (WACV'19)*. IEEE, Los Alamitos, CA, 968–977.

[144] Dandan Sha, Daling Wang, Xiangmin Zhou, Shi Feng, Yifei Zhang, and Ge Yu. 2016. An approach for clothing recommendation based on multiple image attributes. In *Web-Age Information Management*. Lecture Notes in Computer Science, Vol. 9658. Springer, 272–285. https://doi.org/10.1007/978-3-319-39937-9_21

[145] Abdul-Saboor Sheikh, Romain Guigourès, Evgenii Koriagin, Yuen King Ho, Reza Shirvany, Roland Vollgraf, and Urs Bergmann. 2019. A deep learning system for predicting size and fit in fashion e-commerce. In *Proceedings of the 13th ACM Conference on Recommender Systems*. 110–118.

[146] Yue Shi, Martha Larson, and Alan Hanjalic. 2014. Collaborative filtering beyond the user-item matrix: A survey of the state of the art and future challenges. *ACM Computing Surveys* 47, 1 (2014), 3.

[147] Yong-Siang Shih, Kai-Yueh Chang, Hsuan-Tien Lin, and Min Sun. 2018. Compatibility family learning for item recommendation and generation. In *Proceedings of the 32nd AAAI Conference on Artificial Intelligence, the 30th Innovative Applications of Artificial Intelligence Conference, and the 8th AAAI Symposium on Educational Advances in Artificial Intelligence (AAAI/IAAI/EAAI'18)*. 2403–2410.

[148] Edgar Simo-Serra, Sanja Fidler, Francesc Moreno-Noguer, and Raquel Urtasun. 2015. Neuroaesthetics in fashion: Modeling the perception of fashionability. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 869–877.

[149] Montek Singh, Utkarsh Bajpai, Vijayarajan V., and Surya Prasath. 2020. Generation of fashionable clothes using generative adversarial networks: A preliminary feasibility study. *International Journal of Clothing Science and Technology* 32 (2020), 177–187. https://doi.org/10.1108/IJCST-12-2018-0148

[150] Xuemeng Song, Fuli Feng, Jinhuan Liu, Zekun Li, Liqiang Nie, and Jun Ma. 2017. NeuroStylist: Neural compatibility modeling for clothing matching. In *Proceedings of the 25th ACM International Conference on Multimedia (MM'17)*. ACM, New York, NY, 753–761. https://doi.org/10.1145/3123266.3123314

[151] Xuemeng Song, Xianjing Han, Yunkai Li, Jingyuan Chen, Xin-Shun Xu, and Liqiang Nie. 2019. GP-BPR: Personalized compatibility modeling for clothing matching. In *Proceedings of the 27th ACM International Conference on Multimedia (MM'19)*. ACM, New York, NY, 320–328. https://doi.org/10.1145/3343031.3350956

[152] M. Sridevi, N. ManikyaArun, M. Sheshikala, and E. Sudarshan. 2020. Personalized fashion recommender system with image based neural networks. *IOP Conference Series: Materials Science and Engineering* 981 (2020), 022073. https://doi.org/10.1088/1757-899X/981/2/022073

[153] Guang-Lu Sun, Zhi-Qi Cheng, Xiao Wu, and Qiang Peng. 2018. Personalized clothing recommendation combining user social circle and fashion style consistency. *Multimedia Tools and Applications* 77, 14 (2018), 17731–17754. https://doi.org/10.1007/s11042-017-5245-1

[154] Christian Szegedy, Sergey Ioffe, Vincent Vanhoucke, and Alexander A. Alemi. 2017. Inception-v4, Inception-ResNet and the impact of residual connections on learning. In *Proceedings of the 31st AAAI Conference on Artificial Intelligence*. 4278–4284. http://aaai.org/ocs/index.php/AAAI/AAAI17/paper/view/14806

[155] Pongsate Tangseng and Takayuki Okatani. 2020. Toward explainable fashion recommendation. In *Proceedings of the IEEE Winter Conference on Applications of Computer Vision (WACV'20)*. IEEE, Los Alamitos, CA, 2142–2151. https://doi.org/10.1109/WACV45572.2020.9093367

[156] Pongsate Tangseng and Takayuki Okatani. 2020. Toward explainable fashion recommendation. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*. 2153–2162.

[157] Marika Tiggemann, Amy Slater, Belinda Bury, Kimberley Hawkins, and Bonny Firth. 2013. Disclaimer labels on fashion magazine advertisements: Effects on social comparison and body dissatisfaction. *Body Image* 10, 1 (2013), 45–53. https://doi.org/10.1016/j.bodyim.2012.08.001

[158] Nava Tintarev and Judith Masthoff. 2015. Explaining recommendations: Design and evaluation. In *Recommender Systems Handbook*. Springer, 353–382.

[159] Marc Treu, Trung-Nghia Le, Huy H. Nguyen, Junichi Yamagishi, and Isao Echizen. 2021. Fashion-guided adversarial attack on person segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 943–952.

[160] Hessel Tuinhof, Clemens Pirker, and Markus Haltmeier. 2018. Image-based fashion product recommendation with deep learning. In *Machine Learning, Optimization, and Data Science*. Lecture Notes in Computer Science, Vol. 11331. Springer, 472–281. https://doi.org/10.1007/978-3-030-13709-0_40

[161] Laura Vandenbosch and Steven Eggermont. 2012. Understanding sexual objectification: A comprehensive approach toward media exposure and girls' internalization of beauty ideals, self-objectification, and body surveillance. *Journal of Communication* 62, 5 (2012), 869–887. https://doi.org/10.1111/j.1460-2466.2012.01667.x

[162] Andreas Veit, Balazs Kovacs, Sean Bell, Julian J. McAuley, Kavita Bala, and Serge J. Belongie. 2015. Learning visual clothing style with heterogeneous dyadic co-occurrences. In *Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV'15)*. IEEE, Los Alamitos, CA, 4642–4650. https://doi.org/10.1109/ICCV.2015.527

[163] Sirion Vittayakorn, Kota Yamaguchi, Alexander C. Berg, and Tamara L. Berg. 2015. Runway to realway: Visual analysis of fashion. In *Proceedings of the 2015 IEEE Winter Conference on Applications of Computer Vision*. IEEE, Los Alamitos, CA, 951–958.

[164] Riccardo Volpi, Pietro Morerio, Silvio Savarese, and Vittorio Murino. 2018. Adversarial feature augmentation for unsupervised domain adaptation. In *Proceedings of the 2018 IEEE Conference on Computer Vision and Pattern Recognition (CVPR'18)*. IEEE, Los Alamitos, CA, 5495–5504. https://doi.org/10.1109/CVPR.2018.00576

[165] Y. Wakita, Kenta Oku, and K. Kawagoe. 2016. Toward fashion-brand recommendation systems using deep-learning: Preliminary analysis. *International Journal of Knowledge Engineering* 2, 3 (2016), 128–131.

[166] Kaili Wang, Yu-Hui Huang, Jose Oramas, Luc Van Gool, and Tinne Tuytelaars. 2018. An analysis of human-centered geolocation. In *Proceedings of the 2018 IEEE Winter Conference on Applications of Computer Vision (WACV'18)*. IEEE, Los Alamitos, CA, 2058–2066.

[167] Lichuan Wang, Xianyi Zeng, Ludovic Koehl, and Yan Chen. 2015. Intelligent fashion recommender system: Fuzzy logic in personalized garment design. *IEEE Transactions on Human-Machine Systems* 45, 1 (2015), 95–109.

[168] Wenjie Wang, Xinyu Lin, Fuli Feng, Xiangnan He, and Tat-Seng Chua. 2023. Generative recommendation: Towards next-generation recommender paradigm. *arXiv preprint arXiv:2304.03516* (2023).

[169] Wenguan Wang, Yuanlu Xu, Jianbing Shen, and Song-Chun Zhu. 2018. Attentive fashion grammar network for fashion landmark detection and clothing category classification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 4271–4280.

[170] Yining Wang, Eunju Ko, and Huanzhang Wang. 2022. Augmented reality (AR) app use in the beauty product industry and consumer purchase intention. *Asia Pacific Journal of Marketing and Logistics* 34, 1 (2022), 110–131.

[171] Haokun Wen, Xuemeng Song, Xin Yang, Yibing Zhan, and Liqiang Nie. 2021. Comprehensive linguistic-visual composition network for image retrieval. In *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR'21)*. ACM, New York, NY, 1369–1378. https://doi.org/10.1145/3404835.3462967

[172] Haokun Wen, Xuemeng Song, Xin Yang, Yibing Zhan, and Liqiang Nie. 2021. Comprehensive linguistic-visual composition network for image retrieval. In *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 1369–1378.

[173] Yufan Wen, Xiaoqiang Liu, and Bo Xu. 2018. Personalized clothing recommendation based on knowledge graph. In *Proceedings of the 2018 International Conference on Audio, Language, and Image Processing (ICALIP'18)*. 1–5. https://doi.org/10.1109/ICALIP.2018.8455311

[174] Matthias Wölbitsch, Thomas Hasler, Simon Walk, and Denis Helic. 2020. Mind the gap: Exploring shopping preferences across fashion retail channels. In *Proceedings of the 28th ACM Conference on User Modeling, Adaptation, and Personalization (UMAP'20)*. ACM, New York, NY, 257–265. https://doi.org/10.1145/3340631.3394866

[175] Hui Wu, Yupeng Gao, Xiaoxiao Guo, Ziad Al-Halah, Steven Rennie, Kristen Grauman, and Rogério Feris. 2021. Fashion IQ: A new dataset towards retrieving images by natural language feedback. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR'21)*. IEEE, Los Alamitos, CA, 11307–11317. https://doi.org/10.1109/CVPR46437.2021.01115

[176] Qianqian Wu, Pengpeng Zhao, and Zhiming Cui. 2020. Visual and textual jointly enhanced interpretable fashion recommendation. *IEEE Access* 8 (2020), 68736–68746.

[177] Yaxiong Wu, Craig Macdonald, and Iadh Ounis. 2022. Multi-modal dialog state tracking for interactive fashion recommendation. In *Proceedings of the 16th ACM Conference on Recommender Systems (RecSys'22)*. ACM, New York, NY, 124–133. https://doi.org/10.1145/3523227.3546774

[178] Hao Xing, Zhike Han, and Yichen Shen. 2020. ClothNet: A neural network based recommender system. In *Fuzzy Systems and Data Mining VI—Proceedings of FSDM 2020*. Frontiers in Artificial Intelligence and Applications, Vol. 331. IOS Press, 261–272. https://doi.org/10.3233/FAIA200706

[179] Kota Yamaguchi, M. Hadi Kiapour, Luis E. Ortiz, and Tamara L. Berg. 2012. Parsing clothing in fashion photographs. In *Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, Los Alamitos, CA, 3570–3577.

[180] Cairong Yan, Yizhou Chen, and Lingjie Zhou. 2019. Differentiated fashion recommendation using knowledge graph and data augmentation. *IEEE Access* 7 (2019), 102239–102248. https://doi.org/10.1109/ACCESS.2019.2928848

[181] Ming Yang and Kai Yu. 2011. Real-time clothing recognition in surveillance videos. In *Proceedings of the 2011 18th IEEE International Conference on Image Processing*. IEEE, Los Alamitos, CA, 2937–2940.

[182] Wei Yang, Masahiro Toyoura, and Xiaoyang Mao. 2012. Hairstyle suggestion using statistical learning. In *Advances in Multimedia Modeling*. Lecture Notes in Computer Science, Vol. 7131. Springer, 277–287. https://doi.org/10.1007/978-3-642-27355-1_27

[183] Xun Yang, Xiangnan He, Xiang Wang, Yunshan Ma, Fuli Feng, Meng Wang, and Tat-Seng Chua. 2019. Interpretable fashion matching with rich attributes. In *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval*. 775–784.

[184] Zilin Yang, Zhuo Su, Yang Yang, and Ge Lin. 2018. From recommendation to generation: A novel fashion clothing advising framework. In *Proceedings of the 2018 7th International Conference on Digital Home (ICDH'18)*. IEEE, Los Alamitos, CA, 180–186.

[185] Tangwei Ye, Liang Hu, Qi Zhang, Zhong Yuan Lai, Usman Naseem, and Dora D. Liu. 2023. Show me the best outfit for a certain scene: A scene-aware fashion recommender system. In *Proceedings of the ACM Web Conference 2023 (WWW'23)*. 1172–1180.

[186] Ruiping Yin, Kan Li, Jie Lu, and Guangquan Zhang. 2019. Enhancing fashion recommendation with visual compatibility relationship. In *Proceedings of the World Wide Web Conference (WWW'19)*. ACM, New York, NY, 3434–3440. https://doi.org/10.1145/3308558.3313739

[187] Tong Yu, Yilin Shen, and Hongxia Jin. 2019. A visual dialog augmented interactive recommender system. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 157–165.

[188] Wenhui Yu, Xiangnan He, Jian Pei, Xu Chen, Li Xiong, Jinfei Liu, and Zheng Qin. 2021. Visually aware recommendation with aesthetic features. *VLDB Journal* 30, 4 (2021), 495–513. https://doi.org/10.1007/s00778-021-00651-y

[189] Huijing Zhan, Jie Lin, Kenan Emir Ak, Boxin Shi, Ling-Yu Duan, and Alex C. Kot. 2021. A3-FKG: Attentive attribute-aware fashion knowledge graph for outfit preference prediction. *IEEE Transactions on Multimedia* 24 (2021), 819–831. https://doi.org/10.1109/TMM.2021.3059514

[190] Xishan Zhang, Jia Jia, Ke Gao, Yongdong Zhang, Dongming Zhang, Jintao Li, and Qi Tian. 2017. Trip outfits advisor: Location-oriented clothing recommendation. *IEEE Transactions on Multimedia* 19, 11 (2017), 2533–2544. https://doi.org/10.1109/TMM.2017.2696825

[191] Yan Zhang, Xiang Liu, Yunyu Shi, Yunqi Guo, Chaoqun Xu, Erwen Zhang, Jiaxun Tang, and Zhijun Fang. 2017. Fashion evaluation method for clothing recommendation based on weak appearance feature. *Scientific Programming* 2017 (2017), Article 8093057, 12 pages.

[192] Kui Zhao, Xia Hu, Jiajun Bu, and Can Wang. 2017. Deep style match for complementary recommendation. In *Proceedings of the Workshops of the 31st AAAI Conference on Artificial Intelligence (AAAI Workshops'17)*, Vol. WS-17. http://aaai.org/ocs/index.php/WS/AAAIW17/paper/view/15069

[193] Jiachen Zhong, Rita Tse, Gustavo Marfia, and Giovanni Pau. 2019. Fashion popularity analysis based on online social network via deep learning. In *Proceedings of the 11th International Conference on Digital Image Processing (ICDIP'19)*, Vol. 11179. 1117938.

[194] Wei Zhou, P. Y. Mok, Yanghong Zhou, Yangping Zhou, Jialie Shen, Qiang Qu, and K. P. Chau. 2019. Fashion recommendations through cross-media information retrieval. *Journal of Visual Communication and Image Representation* 61 (2019), 112–120.

[195] Zhengzhong Zhou, Xiu Di, Wei Zhou, and Liqing Zhang. 2018. Fashion sensitive clothing recommendation using hierarchical collocation model. In *Proceedings of the 26th ACM International Conference on Multimedia (MM'18)*. ACM, New York, NY, 1119–1127.

[196] Shizhan Zhu, Sanja Fidler, Raquel Urtasun, Dahua Lin, and Chen Change Loy. 2017. Be your own prada: Fashion synthesis with structural coherence. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV'17)*. IEEE, Los Alamitos, CA, 1689–1697. https://doi.org/10.1109/ICCV.2017.186

[197] Kevin Zielnicki. 2019. Simulacra and selection: Clothing set recommendation at stitch fix. In *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval*. 1379–1380.

[198] Susana Zoghbi, Geert Heyman, Juan Carlos Gomez, and Marie-Francine Moens. 2016. Fashion meets computer vision and NLP at e-commerce search. *International Journal of Computer and Electrical Engineering* 8, 1 (2016), 31–43.

[199] Qin Zou, Zheng Zhang, Qian Wang, Qingquan Li, Long Chen, and Song Wang. 2016. Who leads the clothing fashion: Style, color, or texture? A computational study. *CoRR abs/1608.07444* (2016). http://arxiv.org/abs/1608.07444

Received 28 January 2022; revised 29 March 2023; accepted 21 August 2023