SN Computer Science (2023) 4:514
https://doi.org/10.1007/s42979-023-01932-9

SURVEY ARTICLE

Study of AI-Driven Fashion Recommender Systems

Shaghayegh Shirkhani[1] · Hamam Mokayed[1] · Rajkumar Saini[1] · Hum Yan Chai[2]

Received: 30 March 2022 / Accepted: 18 May 2023

Abstract

The rising diversity, volume, and pace of fashion manufacturing pose a considerable challenge in the fashion industry, making it difficult for customers to pick which product to purchase. In addition, fashion is an inherently subjective, cultural notion and an ensemble of clothing items that maintains a coherent style. In most of the domains in which Recommender Systems are developed (e.g., movies, e-commerce, etc.), the similarity evaluation is considered for recommendation. Instead, in the Fashion domain, compatibility is a critical factor. In addition, raw visual features belonging to product representations that contribute to most of the algorithm's performances in the Fashion domain are distinguishable from the metadata of the products in other domains. This literature review summarizes various Artificial Intelligence (AI) techniques that have lately been used in recommender systems for the fashion industry. AI enables higher-quality recommendations than earlier approaches. This has ushered in a new age for recommender systems, allowing for deeper insights into user-item relationships and representations and the discovery patterns in demographical, textual, virtual, and contextual data. This work seeks to give a deeper understanding of the fashion recommender system domain by performing a comprehensive literature study of research on this topic in the past 10 years, focusing on image-based fashion recommender systems taking AI improvements into account. The nuanced conceptions of this domain and their relevance have been developed to justify fashion domain-specific characteristics.

Keywords Fashion recommender system · Artificial Intelligence (AI) · Computer vision · Deep learning · Content-based image retrieval (CBIR)

Introduction

Fashionable goods are in great demand, and as a result, fashion is viewed as a desired and prosperous industry. The fashion industry plays an important role in the global economy, with a large industrial value chain that includes garment design, manufacturing, and sales [4]. Indeed, there has been an increase in global demand for apparel in recent years. The fashion segment revenue is projected to reach 878,334 m US dollars in 2021, with an annual growth rate (CAGR 2021–2025) of 7.31%. The Value of the global fashion industry is alone 3 trillion US dollars today. It accounts for 2% of global GDP, implying that global apparel demand will grow. Fashion recommender systems make it easier for customers to find what they are looking for, but new advancements try to provide more personalized customized recommendations. Complimentary item recommendation is used as a cross-selling strategy in many marketing industries. A major challenge in the fashion domain is the increasing Variety, Volume, and Velocity of fashion production, which makes it difficult for consumers to choose which product to purchase, bringing choice overload to the customers [25].

General recommendation technology is now frequently used on e-commerce websites [17]. The concept of recommendation technology was first proposed in the mid-1990s [33, 44]; early work produced various heuristics for content-based and collaborative filtering (CF). A generic

✉ Shaghayegh Shirkhani
shaghayegh.shirkhani@ltu.se

Hamam Mokayed
hamam.mokayed@ltu.se

Rajkumar Saini
rajkumar.saini@ltu.se

Hum Yan Chai
humyc@utar.edu.my

[1] Department of Computer Science, Electrical and Space Engineering, Luleå Tekniska Universitet, Luleå, Sweden

[2] Department of Mechatronics and Biomedical Engineering, Universiti Tunku Abdul Rahman, Selangor, Malaysia

Published online: 05 July 2023

recommendation system's principal job is to anticipate things that future consumers would prefer and buy based on purchases of others with similar tastes or demographics [1]. Matrix Factorization (MF), which was popularized by the Netflix challenge, eventually became the standard recommender model for an extended period, from 2008 to 2016 [61, 100]. However, later, in the mid-2010s, the advent of deep neural networks in machine learning (also known as Deep Learning) transformed various fields, including speech recognition, computer vision, and natural language processing [35], also transformed the fashion domain.

In most of the domains in which Recommender Systems are developed (e.g., movies, e-commerce, etc.), the similarity evaluation is considered for recommendation [17, 93]. Instead, in the Fashion domain, compatibility is a critical factor. In addition, raw visual features belonging to product representations that contribute to most of the algorithm's performances in the Fashion domain are distinguishable from the metadata of the products in other domains [117].

A few studies [17, 37, 85, 107] have focused on recommender systems in general. However, this paper focuses on an image-based fashion recommender systems survey using deep learning.

Fashion recommendation systems in the apparel industry [37] classified the fashion recommender systems into four categories: style searching/retrieval, fashion coordination, wardrobe recommendation, and intelligent expert systems. Also, "Fashion Analysis" [80] specified two main streams of research in fashion analysis: clothing analysis and facial beauty (including makeup and hairstyle) analysis. They stressed clothing analysis tasks considering clothing recommendation, retrieval, and parsing. [107] described the advancements in fashion research using multimedia, categorizing fashion tasks into three categories: low-level pixel computation, mid-level fashion comprehension, and high-level fashion analysis. Human segmentation, landmark identification, and human posture estimation are examples of low-level pixel processing. The goal of mid-level fashion comprehension is to recognize fashion pictures, such as fashion goods and fashion styles. High-level fashion analysis encompasses fashion trend forecasts, fashion synthesis, and fashion recommendations. [17] recently explored the function of computer vision in fashion and classified fashion research subjects into four broad categories: detection, analysis, synthesis, and recommendation.

This paper presents the literature suitable for both novice and expert readers.

Complexity of the Fashion Domain

A fashion outfit is an ensemble of clothing items that maintains a cohesive style that may be deemed attractive in its overall composition and is in keeping with the general fashion taste of its current time. Stylish outfit construction is challenging since most of the traits that make an outfit fashionable are subjective and impossible to assess adequately [93]. While [6] tries to derive information and actionable insights from fashion data remains difficult due to the inherent subjectivity required to represent the domain adequately. Deriving information and data-driven insights remains challenging due to the inherent subjectivity required to represent the domain successfully. Creating harmonious fashion matching is difficult [17]. Because of the nuanced and subjective nature of the fashion concept, a large number of attributes for describing fashion and the compatibility of the fashion item extend themes in general and involves complex relationships.

Table 1 depicts the most related meanings linked with each idea based on the major phrases and conceptions regularly used in fashion recommendation literature. Some of these names have been used as umbrella terms, implying other meanings, while others have been used interchangeably.

Although compatibility has been used interchangeably with coherency, coordination, and complementary terms in different research papers in the FR domain, conceptually conveying the meaning, the compatibility notion has been perceived differently from a complementary concept from a task definition perspective in FRS. There are two key functions in the FRS domain: compatibility estimation and outfit completion. The former instructs the algorithm to differentiate between clothing that fits together and those that do not. The latter is commonly known as a Fill-In-The-Blank exercise (FITB). Finding the missing item that best completes an outfit from a subset of feasible selections is the goal of FITB. While these notions are closely related and interlinked, a conceptual framework has been taken from them as a coherent depiction of the fundamental conceptual aspects of subjectivity in FRS, as seen in Fig. 1 [93].

Application and Design Balance

Style is characterized by how well the application and design components of clothes work together. Domain compatibility refers to coherency in visual (appearance) and functional characteristics [17]. All fashion recommender systems are built around personalization. [111] noted, "Form follows function," which is regarded as a design principle. It should be noted that the Design and Application are guided by user attributes and on cloth and contextual factors based

Table 1 The frequently used notions in FR literature and their explanation as perceived by the researchers

Notion	Explanation
Compatibility and style	Style may be thought of as an aspect of the overall outfit [90]. Fashion trends evolve spontaneously from how individuals put together clothing combinations [46] Fashion recommendation is based on fashion compatibility, which measures how well different things may work together to generate fashionable ensembles [17]. Unlike style, which relates to how individuals dress [46], compatibility refers to how well-coordinated particular clothing is [50, 79, 118]
Compatibility and similarity	Visual similarity asks, "What looks like this?" On the other hand, compatibility asks, "What complements this?" It necessitates understanding how many visual things interact, frequently based on subtle visual features [46]. Incorporating the concept of compatibility into a more extensive definition of resemblance [117]
Compatibility and complementarity (visually and functionally)	Compatibility is determined by assessing how well-coordinated or complementing a particular pair of clothes is [46]. Describes how detecting links between goods is a critical challenge for an online fashion recommender system to assist consumers in discovering functionally complementary or visually comparable things [42]. Compatibility refers to coherency in both visual (appearance) and functional aspects [17]
Fashionability and compatibility	As measured by the number of "like" votes on a photograph posted online, the popularity of clothing items is referred to as fashionability [106]. Fashion compatibility was emphasized as a vital notion that is the foundation of any FRS to manufacture trendy clothing [17]. To effectively design trendy attire, the system must first and foremost have an innate grasp of product aspects such as color, form, style, fit, and so on [71]
Aesthetic perspective and design	It is essential for individuals to dress attractively; aesthetic adjectives used to describe clothes are connected to visual characteristics (e.g., "formal" or "casual") [54]. Wearing it correctly and attractively [79]. The style can also be viewed aesthetically [6]; each style can thus be defined in the consciousness of an observer as a unified aesthetic entity. Determining the corresponding rules from color combinations to generate impressions [67]. Visual information is crucial in the human decision-making process [139]. Fashion design activity serves as a foundation for dressmaking or pattern-making [62]. To optimize user preferences, fashionable clothes should be designed with a person's taste in mind [59]
Personalization	A well-defined user profile might help differentiate a more personalized recommendation system from existing systems [37]. In online services, recommender systems have been frequently utilized to forecast users' preferences based on their interaction histories [137]. The aesthetic component is critical in modeling and forecasting customer preferences, particularly in fashion-related domains such as apparel and jewelry [138]. The significance of personal preferences in style formation [17]
Style	Style is a consideration while picking each fashion choice for an ensemble. Style may be thought of as an aspect of the whole clothing [90]. The style can also be viewed aesthetically [6]; each style can thus be defined in the consciousness of an observer as a unified aesthetic entity. Choosing the style of a garment is influenced not only by the physical characteristics of the garment's components but also by the context [73]. What is a visual style? Fashion trends arise spontaneously from how individuals put together clothing items, making them challenging to predict with a computer model [46]. Outfits in online fashion data are made up of several distinct sorts of things (for example, tops, bottoms, and shoes) that share some style connection [117]. Style coherence is not the same as traditional conceptions of visual similarity [46]. Style coherency refers to constant fine-grained patterns reflected by varied combinations of clothes, and coherent styles reflect some latent appearance [46]

on clothing ontology. Fashion may be understood through domain ontology. Clothing ontology is primarily concerned with modeling the structure of physical feature values (e.g., sleeve length, colors, fabric). These elements have varying degrees of comprehension, and taking them into account allows FRSs to prescribe varying degrees of personalization [125].

Fashion Ontology

Outfits in online fashion data are formed of pieces of several sorts (e.g., top, bottom, shoes) that have some stylistic link [117]. An approach for learning both ideas of similarity (for example, when two tops are interchangeable) and compatibility is required for a representation for creating ensembles (items of a possibly different type that can go together in an outfit). To effectively design trendy ensembles, the system must first and foremost have an innate grasp of product aspects such as color, form, style, fit, and so on [71]. To represent nuanced concepts such as 'compatibility' on raw visual characteristics, we require expressive transformations capable of linking feature dimensions to describe the connections between pairs of items [42]. These products' characteristics may be conveyed in various ways, including photos, text, video, and audio. Identifying and comprehending complex and diverse links between product items is critical for any current recommender system [42]. We require expressive transformations capable of linking feature dimensions to describe the relationships between pairs of things to model sophisticated ideas like 'compatibility' upon raw visual characteristics [42]. Most clothing recommender systems use collaborative filtering and content-based

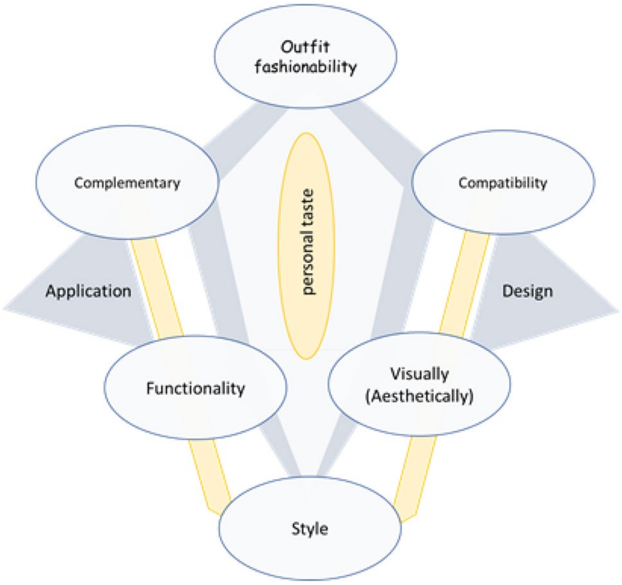


Fig. 1 Conceptual characteristics which convey fashion recommender system subjectivity

approaches, focusing on forecasting the user's item preferences based on enormous historical data and ignoring user contexts, such as weather, occasion, requirements, emotions, and relationships between clothing items [125]. The user entity description includes user characteristics such as height, weight, skin color, and so on. Cloth qualities such as color, texture, pattern, fabric, and so on are used to describe the clothing [4, 125]. The context entity contains information about the current weather, occasion, and so on [37]; apparel features are described from a formulation of colors, lines, shapes, patterns/prints, and textures that were studied through feature extraction, recognition, and encoding. Several studies have proven it feasible to reliably extract or detect visual information (such as sleeve length, color distribution, and garment pattern) from clothing photographs [133–135]. User features, on the other hand, are recognized as face characteristics, body traits, personal choice (taste), and wearing occasions. The authors of [125] developed three types of essential fashion ontology, as shown in Fig. 2, representing the user's linked elements, clothes, and context for clothing recommendation. A well-defined user profile might help differentiate a more personalized recommendation system from existing ones [37].

Fashion Recommendation System Tasks

When it comes to suggesting specific architecture and the unique scenario to get the best results, it is highly reliant on the task, domains, and recommender scenarios [142].

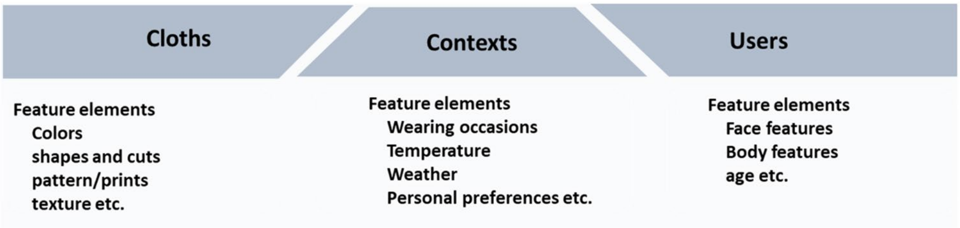
The FRS literature primarily classified recommendation into four tasks: item retrieval, complimentary item recommendation, whole outfit recommendation, and capsule wardrobe recommendation. The retrieval systems were seen in many ways. The researchers in [17] noted that the item-retrieval learns visual similarity between the same clothing type, but the FRS learns both visual similarity and visual compatibility between different types of clothing, differentiating the similarity task from compatibility estimation. Clothing-retrieval-based studies are included in this literature review since they have received much attention in clothing recommendation and play a significant role in introducing and evolving image-based fashion recommender systems. Table 2 shows the main tasks of FRSs discovered by this literature study and the connected ideas and attributes.

Similar Item Recommendation (Item Retrieval)

Researchers and scholars have introduced many image retrieval methods; [97] highlighted some of the most important and widely used techniques, including Text-based, Content-based, Multimodal Fusion, Semantic-based, and Relevance Feedback-based Image Retrieval. The work [53] proposed content-based recommender systems, which have sparked much interest in the field of fashion recommendation due to recent advances in deep learning methodologies, particularly in image processing, interpretation, and segmentation, as their amplified applicability in fashion.

Because of the necessity to appropriately handle the rapidly increasing volume of multimedia data, content-based image retrieval (CBIR) has gained much attention in the recent decade [10]. The content-Based Image Retrieval (CBIR) approach can retrieve relevant pictures from a database using an input image with the desired content [75]. This method is widely used in computer vision and artificial

Fig. 2 Fashion feature elements based on cloth ontology [125]



Cloths	Contexts	Users
Feature elements Colors shapes and cuts pattern/prints texture etc.	Feature elements Wearing occasions Temperature Weather Personal preferences etc.	Feature elements Face features Body features age etc.

Table 2 The main tasks of FRS with associated significant concepts and features

Recommender system	Key features and concepts
Imaged retrieval	Similar or identical item recommendation. Content-based image retrieval (CBIR) has received much interest among different image retrieval methods commonly used in CV and AI applications. Fashion instance-level image retrieval (FIR) as a sub-category of (CBIR), primarily concerned with cross-domain fashion image retrieval tasks
Complementary item recommendation	A relaxed version of the Outfit recommendations. Typically, only one item. Similar to the techniques used in similarity-based item recommendation. Most approaches rely on hybrid models. Usually including product-based, scene-based, and occasion-based Complementary Recommendation. Considered as FITB task. Often a model is given an incomplete outfit and then is asked to predict the missing items, given their categories
Outfit recommendation	The Complete Fashion Coordinators. Retrieving matching items. Formulated as three main stages: Learning Outfit Representation, Learning Compatibility, and personalization. Creation of an outfit from the scratch point or an incomplete one. Outfits' Compatibility Scoring using uni/multi-modal neural architectures. Sequential Outfits Representations and Predictors
Capsule wardrobes	Outfit subset selection problem. A minimal set of items that provides maximal mix-and-match outfits

intelligence applications. The CBIR system is technologically based on image representation and database search [75]. To identify pictures, feature vectors or image representations are supposed to be discriminative. It is also said to be immune to certain transformations. Based on visual representation, the similarity score between two images should communicate the semantic link. These interconnected characteristics are crucial to retrieval performance, and CBIR algorithms may be classified depending on how they impact these two variables. The fundamental stage for CBIR is image representation, which extracts the relevant characteristics from a given picture and then turns them into a fixed-sized vector (so-called feature vector) [75]. The extracted features are classified into three types: conventional features, classification CNN features, and retrieval CNN features.

Due to the rapid expansion of clothing e-commerce and the increase in the number of clothing photographs on the Internet, fashion instance-level image retrieval (FIR) as a sub-category (CBIR) is regarded as a hot topic in computer vision. (FIR) is required to address the growing demands for online purchases, fashion identification, and web-based recommendation. FIR focuses on cross-domain fashion image retrieval, which involves matching two photographs taken informally by users and professionally by vendors [75]. While there is a large domain variance between daily human photographs collected in a typical environment and clothing images taken under ideal settings, significant research efforts have been directed toward addressing the challenge of cross-scenario clothing retrieval (i.e., edited photos used in online clothing shops) [17].

The research community has focused heavily on creating cross-scenario image-based fashion retrieval tasks that find virtually similar goods from an inventory based on a fashion image query [17]. The seminal early work on automated image-based garment retrieval was presented by [120]. An unsupervised transfer learning approach [79] based on part-based alignment and sparse reconstruction attributes was proposed. The model integrates four potentials, including visual feature vs. attribute, visual feature vs. occasion, occasion vs. attribute, and attribute vs. attribute, in this Occasion-Based fashion advice system in the "magic closet." As a cross-scenario retrieval, [57] proposed a scalable solution to automatically recommend appropriate apparel goods given a single image without metadata.

It should be noted that the approaches used by [57, 79, 120] are based on hand-crafted characteristics. Building deep neural network architectures to address the apparel retrieval task has become popular as deep learning develops [17]. As large-scale fashion datasets were made available, many FIR techniques based on deep learning were created and successfully tested [10].

The authors of [49] designed a Dual Attribute aware Ranking Network (DARN) to represent in-depth features via attribute-guided learning. DARN modeled domain disparity while including semantic characteristics and visual similarity constraints in the feature learning stage. The street-to-shop retrieval challenge was initially attempted by [64], who built three separate algorithms for recovering the same fashion item in a real-world photograph from an online shop. Two deep learning baseline techniques were included in the three approaches, and one method learned the similarity between two street and store domains. In another work [55], researchers introduced a deep bi-directional cross-triplet embedding approach to describe the similarity of cross-domain pictures, which enhanced the one-way problem, street-to-shop retrieval job. They also extended this technology to fetch several related accessories to go along with the cloth item shop.

A Graph Reasoning Network (GRN) [70] was introduced to construct the similarity pyramid to improve on existing retrieval task methods that only considered global feature vectors and represented the similarity between a query and a clothing inventory by considering global and local representations. Other researchers incorporate text descriptions in addition to visual cues in garment retrieval tasks in some cases [68, 143]. A visual representation of the searched item was constructed by combining the visual representation of the query picture with the textual attributes in the query text; however, [71] employed a common multimodal embedding space to derive the semantic relationship between visual and textual elements. Because of recent advances in deep learning techniques, Content-Based Similar Fashion Items Recommendation approaches have received attention, mainly through computer vision and natural language processing. An early work [38] was introduced on the computer-generated design of fashion garments as a part of applications of Generative Adversarial Networks. The work [59] proposed a visually aware strategy as a solution for both the design and recommendation demands of the fashion industry. Another work [62] developed a method for generating clothing images.

Using metric learning advancements, several FIR systems employ various attention mechanisms. The Visual Attention Model (VAM) [121] developed an end-to-end network architecture by training a two-stream network with an attention branch and a global convolutional branch, then concatenating the resulting vectors to enhance a standard triplet objective function. FashionNet [81] trained a network using a triplet loss as well. Hard-aware Deeply Cascaded embedding (HDC) [140] combined a succession of models with varying complexity to my hard instances at several layers via a cascaded mechanism. The featured vectors from each subnetwork were weighted and merged to generate retrieval representations. The researchers [27] concentrated on improving the training and inference processes. They underlined the need for proper basic architecture training, trained the network with the triplet loss, and customized generic models to the specific function. In another work [143], authors created an adversarial network for Hard Triplet Generation (HTG) to increase the network's ability to differentiate comparable instances of different categories while grouping particular examples of the same categories.

A competitive FIR performance was aimed in [105]. They presented a unique method for transforming a query into a representation with the desired properties, as well as a novel idea of feature-level attribute modification. Some deep learning algorithms combine many methodologies. The Grid Search Network (GSN) [18] framed the training function as a search task to locate matches for a given picture query in a grid that included both positive and negative images. Attribute modules are used in several FIR methods [24, 49, 86, 105]. Other researchers [91] examined training methodologies and deep neural networks (DNNs) to increase retrieval performance. It has been demonstrated that better training procedures, data augmentation, and structural refinement can result in superior FIR outcomes.

Conventional recommender systems make predictions based on similarity (between items in content-based techniques and between users in collaborative-filtering methods), whereas the outfit completion task incorporates the concept of compatibility between the items that comprise an outfit [17, 93]. Each outfit often consists of several complementary things that match aesthetically or visually when worn together, such as tops, bottoms, shoes, and accessories [20]. As previously discussed, creating harmonious fashion matching is difficult. The suggestion of complementary clothing products has been a prominent problem in the fashion recommendations research area in recent years, gaining a lot of interest and resulting in a long list of algorithms and approaches.

Complementary Item Recommendation

Complimentary item recommendations [53] can be considered a relaxed version of the Outfit recommendations problem. The authors of [53, 93] declared that most of the previous research in fashion recommendations has focused on individual items' recommendation as outfit completion (complementary item recommendation or Fill-In-The-Blank task) and a limited amount of work has been conducted on whole outfits' recommendations. FITB consists in finding the missing item that best completes an outfit from a subset of possible choices. The type of missing items, their number, and the sampling modality for the list of candidates vary from paper to paper [93]. Typically, only one item is removed from the outfit, while the subset of proposed items to choose from contains the missing item and other three clothes [39, 117]; the only exception is in [74], that uses subsets of five elements, one of which is the correct one. The authors of [53] explained that the approaches proposed to address this problem are usually similar to the ones used in similarity-based item recommendations. So this problem can also be modeled by adding constrain in the similar item recommendation problem; however, when it comes to complementary fashion item recommendations, most approaches rely on hybrid models that use both user-item interactions and content-based features to generate recommendations. Following this, [93] introduced an extended task as Unconstrained Outfit Completion, a generalization of the FITB task. The number of relevant recommendations is then evaluated using some typical evaluation metrics used in the information retrieval domain, i.e., precision, recall, mean average precision, accuracy, and reciprocal ranking. In this approach, a model is given an incomplete outfit and then is asked to predict the missing items, given their categories. The authors in [93] clarified that all of these tasks are evaluated on a series of questions. Each question is a test set outfit (incomplete in FITB and UOC), and the answer is a binary label for the classification task (Compatibility Estimation), the selection of an item from a limited set (FITB), or the selection of one or more items from a collection (UOC).

An outfit consists of multiple clothing items; from a cloth ontology perspective [125], each item consists of different features. The compatibility of the clothing items necessitates employing the styling experts' guidance and the most recent trends. In addition to personal preferences and user and context characteristics, all of these features must first be understood, recognized, and extracted; as a result, when combining objects, the system must learn the interaction between these features. Creating customized outfit recommendations for consumers can be based on past purchases, specific input on products they like, or a customer query (e.g., photo) that can provide insight into their preferences or style. The style relationship for complementary recommendations is exploited [143]. They deduce this relationship between fashion items based on the title description, assuming that the title contains the most relevant information about the item. They employed (SCNN) Siamese Convolutional Neural Network to find the compatible pairs of items in a words space, then mapped them into an embedded style space. Words are the only inputs, making computation lighter; it also needs a few preprocessing stages without any feature engineering. Item-to-item compatibility modeling was proposed as a metric-learning issue based on co-purchase behavior by [118]. They used Amazon.com co-purchase data to train a Siamese CNN to learn style compatibility across categories and a nearest neighbor retrieval to generate compatible goods. The researchers in [89] focused on understanding parallels or complementary links between objects in the same way that the human brain does, as well as the applicability of a complimentary item. They constructed a general-purpose approach using textual and visual features from the Amazon data set of people who purchased/viewed things. While most earlier research focused on top-bottom matching issues, [48] worked on the personalized issue by modeling user-item and item-item relations using a functional tensor factorization approach. Exploring Visually Comparable Items returns visually similar items to a specific user query and are also regarded as stylish [42]. An occasion-oriented clothing recommender system proposed in [141] considers both the concepts of suitable and attractively wearing apparel. They utilized a unified latent Support Vector Machine to learn the recommendation model that used apparel matching criteria among visual aspects, qualities, and contexts (SVM). The topic was approached as a classification challenge via an end-to-end framework [74]. An outfit composition set was evaluated as fashionable or not based on aesthetics and compatibility. “Complete the Look,” developed by [60], suggests fashion items that suit the setting. They employed Siamese networks and category-guided attention approaches to assess both local (compatibility between each scene patch and product image) and global (compatibility between the scene and product images).

A mixed-category metric learning approach [11] can take numerous inputs to improve the traditional triplet neural network, which generally accepts just three items. They also stimulate the intra-category and cross-category items of fashion collocation by feeding the deep neural network of both well-collocated and poorly-collocated apparel items. The authors of [46] employed natural language processing to classify an outfit as a “document,” a clothing characteristic as a “word,” and a clothing style as a “topic.” The topic model addressed clothing matching. A STAMP (Short-term attention/memory priority model [83] was proposed for the session-based recommendation) model to address the limitations of previous approaches by considering the user's current actions to generate future recommendations in the same session in real-time. All this is doable by utilizing an attention model that can model the long-term session properties in parallel with modeling the user's last clicks to save all short-term attention tendencies. This novel idea soon becomes popular among other researchers. Inspired by the STAMP model [128] introduced a session-based approach with some improvements in the STAMP model [83] to produce complementary personalized item recommendations. [136] integrating new items for a recommendation automatically. They proposed a personalized fashion design network based on a query item, which generated a fashion item for each user, considering user interests and fashion compatibility. [14] developed a large-scale Personalized Outfit Generation (POG) model. They created POG on Alibaba's Dida platform to propose trendy tailored clothing to clients. They used user clicks on favored clothing items in combination with their purchase history to learn about consumers' tastes. They used a transformer encoder-decoder approach to simulate compatibility between clothing items and user interests. Another research in [92] makes suggestions for complementary clothing items given a query item using a Siamese network extracting features and a fully connected network learning fashion compatibility metric.

The majority of personalized outfit recommendation systems are based on the idea of providing an entire outfit based on a single clothing item that the client is exploring [53] or by offering a complimentary item that completes the look [60]. Furthermore, another classification is based on personal wardrobe [79]. The authors of [53] declared that outfit recommendation could be formulated as three main stages: Learning Outfit Representation, Learning Compatibility, and personalization. Learning the visual representations of the clothing items and/or their textual metadata attributes. Transforming each clothing item into an embedding. This transformation might be considered a concatenated representation of the image and its relative metadata. The style representation can also be deduced and used as input in the subsequent phases [53]. The compatibility among different clothing items for an outfit can be obtained through experts' opinions like fashion designers and stylists [108] or personal preferences of customers [40]. The system can also learn based on positive/negative samples of compatible items or scoring similarities between latent representations of different items [53]. Understanding personal preferences can be specified through direct input from customers themselves or through deductions obtained from their input and past behaviors and considering the collected representation of the users' preferences in the model used for learning [53].

Whole Outfit Recommendation

Pioneering research on whole outfit recommendation (e.g., The Complete Fashion Coordinator [116], What Am I Gonna Wear [104], and Magic Closet [79]) was dependent on user input concerning the cloths they hold in their wardrobes and the occasions they wear them. Outfits suggestion is also made based on historical data, deduced preferred style, and matched with the occasion. For instance, in the Complete Fashion Coordinator proposed by [116], the user enters pictures of their clothes along with information such as the occasion that item was worn there. It is also possible that by using social networks, they get feedback on them. Magic Closet suggested a method for retrieving matching items that online retailers can use. It matches each cloth item from the user's wardrobe [79].

Two main sub-categories of research in outfit recommender systems [53] have been identified, including models that used Outfits' Compatibility Scoring and Sequential Outfits Representations and Predictors. while [53] conducted a vast majority of research, including [5, 14, 60, 74, 89, 105, 107, 117, 118] in outfits recommendation is based on outfit scoring using uni- or multi-modal neural architectures for extracting and learning the feature representations of outfits and then applying a classifier network to predict a score that describes the outfit's style compatibility and adherence to the user's personal style. The second highlighted group [39, 56, 90, 123] used a sequential method to model the fashion outfit. Each piece of clothing represents a time step, and consistent order of clothing categories is used to ensure that no single item in an outfit is duplicated or missing. Bi-directional Long short-term memory (LSTM) architectures [36] enable modeling the interactions among current, past, and future preferred items using forward and backward passes, and the global dependencies between the clothing items can be identified. Now, we will introduce some most recent and significant outfit recommender research.

In [71], the primary purpose was to create an outfit regardless of the scratch point or an incomplete one. The attention-based fusion enhances item comprehension by fusing the product picture and description information to capture the most significant, fine-grained product attributes into the item representation. While they indicate that Outfit recommendation deals with two main challenges, (i) item understanding that demands visual and textual feature extraction and combination to make a better understanding, and (ii) item matching concerning the complexity of the Item compatibility relation, they focused on item understanding.

High-level semantic compatibility (e.g., style, functionality) cannot be managed just based on fashion photographs, as opposed to low-level visual compatibility (e.g., color, texture). The authors of [113] created a cutting-edge multimodal framework for fashion compatibility learning that merged semantic and visual embeddings into a single deep learning model. For discriminative semantic representation learning tasks, a multilayered Long Short-Term Memory (LSTM) is utilized. A deep CNN is used for visual feature extraction. The semantic and visual information of fashion products is then concatenated with a fusion module, which turns them into a latent feature space. Furthermore, a novel triplet ranking loss with compatible weights is given for assessing fine-grained associations between fashion products, which is more in accordance with human emotions when grasping the concept of fashion compatibility. Several trials demonstrated the efficacy of the suggested strategy, which outperforms the innovative methods used on the Amazon fashion dataset. while [23] suggested a unique Attentional Content-level Translation-based Fashion Recommender (ACTR) in the realm of sequential fashion recommendation. They improve the sequential fashion suggestion model by modeling the user's immediate intent and including item-level features.

Using a product picture and historical reviews, [128] created a Visual and Textual Jointly Enhanced Interpretable (VTJEI) model for fashion recommendations. By combining textual and visual information enhancement, the model delivers more accurate suggestions and visual and textual explanations. In addition, they presented a bidirectional two-layer adaptive attention review model to accept both implicit and explicit preferences from the user. Furthermore, [128] used a review-driven visual attention model to develop higher degrees of tailored picture representation based on the user's preferences. [109] suggested a customized data-driven fashion recommender system that creates recommendations for the user based on a given input.

The researchers in [16] proposed a new task in the deep-learning fashion industry. When it comes to Attribute editing, creating a photorealistic image incorporates the texture from several image sources. As a result, the highly complex qualities and the absence of paired data are critical difficulties to these tasks. To overcome these restrictions, [16] offers a novel self-supervised model to combine garment images with separated features (e.g., vest and sleeves) without paired data. Model training consists of two steps: self-supervised reconstruction learning and generalized attribute modifications using adversarial learning. For the learning process of each picture, a fully supervised training procedure was used, and an encoder-decoder structure was used for the self-supervised reconstruction.

Detect, Pick, and Retrieval Network (DPRNet) analyzes celebrity's fashion products from their videos and viewers' interests in their (former) apparel items to address video-to-shop issues. To boost performance in the video-to-shop operation, they updated the typical object detector, which automatically selects the best object areas without duplication. On DeepFashion, a multitask loss network has been used for fashion retrieval [144].

The authors of [110] developed an aesthetic-aware clothing recommender system that proposed a collaborative fashion recommendation system (CFRS), introducing a novel trend score metric. The trend score can be viewed by users and other product details to convey more insight about the products and is also used to sort products from trendiest to classic ones in each category. The system administrator is in charge of product management and, as a result, trend and user management. The administrator has the authority to update the trends in three categories: colors, printing, and materials. Fashion experts include fashion magazine editors, fashion designers, fashion bloggers, and others who may grade fashion trends by liking or hating things. Visitors cannot view or rate fashion trends and can watch them; however, users may see current fashion trends and follow (like) experts.

To address the shortcomings of previous works that focused on the compatibility of two items or represented an outfit as a sequence and failed to utilize the complex relations among items in an outfit fully, [21] proposed representing an outfit as a graph with each node representing a category and each edge representing the interaction between two categories. As a result, each outfit will be treated as a subgraph. To determine outfit compatibility from such a network, [21] proposed Node-wise Graph Neural Networks (NGNN). In NGNN, the node interaction for each edge varies. An attention mechanism is employed via learning node representations for outfit compatibility estimation. NGNN can be employed for modeling outfit compatibility from multiple modalities.

The authors of [94] proposed a model as an event-based outfit recommender system. In her proposed model, the type of event has been identified using object detection. Then, the clothes worn at that event are identified. Next, the correlation between the event and the clothes worn there has been understood. In this way, the most recently used clothes are recognized, and consequently, similar clothes have been recommended to employ the nearest neighbor approach for event recognition tasks employing RCNN as the meta-architecture for object detection.

The authors of [52] introduced two unique ways for dynamically using social media textual input in addition to visual categorization. The first is adaptive neural pruning (Dynamic Pruning), which activates the clothing attribute classifier's probable range of connections based on the clothing item category recognized through text analysis in postings. The second technique (Dynamic Layers) is a dynamic structure that has numerous attribute classification layers and dynamically activates an appropriate attribute classifier layer based on the image's mined text.

The authors of [63] created a two-stage deep learning system for recommending fashion clothing based on the visual resemblance style of other photos in the database. Using images as input, try understanding features. By doing so, a neural network classifier has been employed as an image-based feature extractor. A similarity algorithm is then responsible for generating recommendations and ranking suggestions. As a commonly used technology in image recognition, a convolutional neural network addresses the major functionality of classification and recommendation.

Capsule Wardrobe Recommendations

The capsule closet is undoubtedly one of the most popular and pervasive minimalist notions, focused on people's closets [43]. Whether it is referred to as a capsule closet, a closet detox, an apparel diet, or a minimalist wardrobe, it can help to shape the future of fashion and textile industries by shifting consumer mindsets, demand, and ambition away from maximalism to minimalism, materialism to idealism, and inviting companies to adapt their value chains to meet new consumer demands [43]. Given a stock of candidate articles of clothing and adornments, the algorithm must collect a minimal set of things that gives maximal mix-and-match outfits [46]. A capsule wardrobe is a collection of clothing articles that may be combined to create several sets of clothes that are aesthetically compatible. A method for creating a capsule wardrobe [46] automatically by characterizing the challenge as a subset selection issue. They contributed new perspectives into effective optimizations for combinatorial mix-and-match outfit selection as well as generative learning of visual compatibility. There are similar examples, including [40].

Computer vision in Fashion Recommender Systems

Computer vision has spread into numerous sectors, from gathering raw data to extracting picture patterns and analyzing data [126]. It includes principles, techniques, and ideas from digital image processing, pattern recognition, artificial intelligence, and computer graphics [19, 126]. Depending on the application domain and the data being processed, the methodologies used to handle computer vision challenges vary [126]. There is some overlap with Image Processing in terms of basic approaches, and some authors use both terms interchangeably [126].

Some critical tasks of computer vision include Object Detection (the location of the object), Object Classification (the broad category that the object lies in), Object Recognition (the objects in the image and their positions), and Object Segmentation (The pixels belonging to that object) [95]. "Fashion Meets Computer Vision: A Survey" [17] classified computer vision tasks in the fashion area into four categories: detection, analysis, synthesis, and recommendation. The primary goal of computer vision was to extract image features [95]. To put it succinctly, the fashion industry is a magnet for computer vision [46]. Following the quick evolution of the fashion industry into an online, social, and highly individualized business domain, new vision challenges are arising. Style models [64, 72, 88, 106], forecasting trends [3], interactive search [68, 143], and recommendation [48, 79, 118], all demand visual understanding considering all details with delicacy.

"Fashion Meets Computer Vision: A Survey" [17] conducted current research on intelligent fashion (intelligent fashion refers to fashion technology that is empowered by computer vision), which covers the research topics not only to detect what fashion items are presented in an image but also to analyze the items, synthesize creative ones, and finally creating customized recommendations. In recent years, computer vision researchers have focused on fashion, largely represented visually. Intelligent fashion is a complicated process from a technological standpoint since, unlike generic products, fashion items have significant aesthetic and design variances. Above all, there is a significant semantic gap between computable low-level features and the high-level semantic concepts that they encode. The authors of [107] presented a study in summarizing advancements in multimedia fashion research and categorizing fashion tasks into three categories: low-level pixel computing, mid-level fashion comprehension, and high-level fashion analysis. Low-level pixel computing includes human segmentation, landmark identification, and human posture estimation. The goal of mid-level fashion comprehension is to recognize fashion pictures, such as fashion goods and styles. High-level fashion analysis encompasses fashion trend forecasts, fashion synthesis, and fashion suggestions. Convolutional Neural Networks have excelled in various computer vision applications, including object recognition, detection, picture segmentation, and texture generation [28]. In another work [87], the authors outlined various ways to upgrade fashion technology with deep learning in "Fashion Object Detection and Pixel-Wise Semantic Segmentation." One of the fundamental concepts is to use artificial intelligence to develop fashion designs and recommendations. Similarly, an essential element is acquiring credible information on fashion trends, which involves analyzing current fashion-related photos and data.

Computer Vision and Deep Learning in FRS

The researchers [129] proposed content-based and hybrid recommendation models based on image information modeling. In content-based models, visual signals are utilized to build item visual representations, and consumer preferences are represented in the visual space [2, 45, 76, 82, 89, 96, 124, 131]. On the other hand, hybrid recommendation models employ item visual modeling to overcome data sparsity issues in CF [9, 13, 42, 122]. The researchers in [42] use visual content to create Visual Bayesian Personalized Ranking, a unified hybrid recommendation system (VBPR). This technique displays each user (item) in two latent spaces: a visual space projected using CNN-based visual features and a collaborative latent space used to discover users' latent preferences. The projected preference is learned by merging users' preferences from two regions, given a user-item pair and a related picture. Following the basic principle of VBPR, matrix factorization-based models have incorporated the item's visual content as a regularization term, ensuring that the learned item latent vector is similar to the visual image representation acquired by CNN.

CBIR techniques have attracted scholars' interest in the fashion recommender system domain. CBIR has also been widely used and improved via different computer vision and artificial intelligence approaches [75]. Content-based image retrieval can be determined via three main eras that vary in how they export the different low-level features representing an image's visual content [32, 115]. Following the proposed evolutionary eras of Content-based image retrieval techniques, we mapped out the main milestones of proposing fashion recommender systems in Fig. 3.

Conventional CV approaches, including feature descriptors, were employed for image retrieval tasks in the first two eras, notably for object detection. Before the advent of deep learning, feature extraction was utilized for applications such as image classification, as will be explained in the following sections. The advent of deep neural networks in machine learning (also known as Deep Learning) in the mid-2010 s changed various disciplines, including computer vision, natural language processing, and speech recognition [35].

Since 2012, a growing body of evidence has shown the primary application of CNNs in the advancement of image retrieval systems in the fashion domain, termed Era 3.0. Intelligent clothing recommender systems based on fashion and aesthetic concepts are being researched [79] to fulfill the

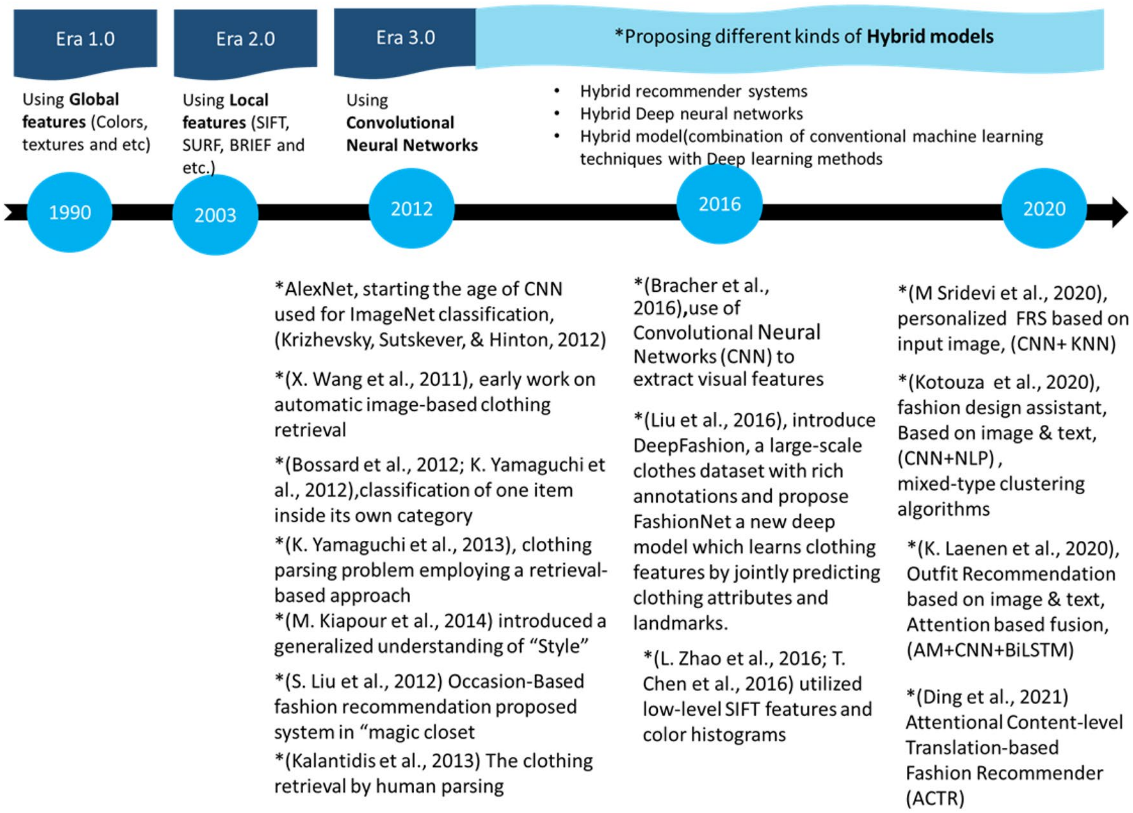

Era 1.0
Era 2.0
Era 3.0
*Proposing different kinds of Hybrid models
Using Global features (Colors, textures and etc)
Using Local features (SIFT, SURF, BRIEF and etc.)
Using Convolutional Neural Networks
• Hybrid recommender systems
• Hybrid Deep neural networks
• Hybrid model(combination of conventional machine learning techniques with Deep learning methods
1990
2003
2012
2016
2020
*AlexNet, starting the age of CNN used for ImageNet classification, (Krizhevsky, Sutskever, & Hinton, 2012)
*(X. Wang et al., 2011), early work on automatic image-based clothing retrieval
*(Bossard et al., 2012; K. Yamaguchi et al., 2012),classification of one item inside its own category
*(K. Yamaguchi et al., 2013), clothing parsing problem employing a retrieval-based approach
*(M. Kiapour et al., 2014) introduced a generalized understanding of "Style"
*(S. Liu et al., 2012) Occasion-Based fashion recommendation proposed system in "magic closet
*(Kalantidis et al., 2013) The clothing retrieval by human parsing
*(Bracher et al., 2016),use of Convolutional Neural Networks (CNN) to extract visual features
*(Liu et al., 2016), introduce DeepFashion, a large-scale clothes dataset with rich annotations and propose FashionNet a new deep model which learns clothing features by jointly predicting clothing attributes and landmarks.
*(L. Zhao et al., 2016; T. Chen et al., 2016) utilized low-level SIFT features and color histograms
*(M Sridevi et al., 2020), personalized FRS based on input image, (CNN+ KNN)
*(Kotouza et al., 2020), fashion design assistant, Based on image & text, (CNN+NLP) , mixed-type clustering algorithms
*(K. Laenen et al., 2020), Outfit Recommendation based on image & text, Attention based fusion, (AM+CNN+BiLSTM)
*(Ding et al., 2021) Attentional Content-level Translation-based Fashion Recommender (ACTR)

Fig. 3 The evolution of CV methods with DL advancements in FRS

varying demands of different users. The fundamental aim of [7, 132] was to categorize one item inside its category. Several researchers later proposed various methods for clothing parsing with various inputs [24]. The authors of [132] devised an efficient method for parsing apparel in fashion images, which is difficult due to the vast number of items and differences in setting and garment look, layering, and occlusion. Furthermore, they provided a large novel dataset and tools for labeling garment items for future research. The authors of [133] used a retrieval-based strategy to tackle the clothing parsing challenges. Following this road through the next milestone where the idea of "style" has received greater attention, notably by researchers such as [64], who tend to give a more generalized understanding of "Style" and build a useful dataset in this regard. It was used to generate garment suggestions, client ratings, and clothes [47]. To develop recommendations, [77] evaluated the history of clothing and accessories, as well as weather circumstances. There have recently been significant breakthroughs in computer vision tasks such as object recognition and detection and segmentation [12, 69, 98]. The revolution began when [69] used convolutional neural networks to significantly improve object recognition on the ImageNet challenge (CNN). This prompted to study and subsequent advancements in a variety of fashion-related tasks, including clothing categorization, predicting various types of properties of a given piece of clothing, and enhancing picture retrieval [7, 57, 64, 81, 130].

CNN extracting visual features from images [8] has given researchers a chance to possibly identify the complex and non-linear relationship between the visual characteristics of items. The work [103] also offered suggestions by extracting various aspects from photos to understand the contents, such as material, collar, sleeves, and so on. The work [81] constructs a substantial fashion dataset of around 800,000 photos with annotations for various types of clothing, their features, and the position of landmarks, as well as cross-domain pairings. In addition, they create a CNN to predict attributes and landmarks. The architecture is built on a 16-layer VGGNet with convolution and fully connected layers added to train a network to predict them. The works [29, 30], outlines the process of separating the content and style of photos and recombining them to produce new images utilizing image representations stored by many layers of VGGNet. They model an image's style by extracting the feature maps produced when the image is passed through a pre-trained CNN, in this case, a 19-layer VGGNet. The texture synthesis technique created by each convolutional layer provides the foundation for style extraction. When the pre-trained VGGNet analyses an image, the style is retrieved from convolutional layers. When the picture is processed, feature maps derived from higher network layers are used to generate the content. The mean squared error (MSE) between the input and output pictures is used to calculate the style and content losses (initiated from white noise). The researchers in [28] employed a 19-layer pre-trained VGG-Net in "Fashioning with Networks: Neural Style Transfer to Design Clothes." In another work [109], authors used neural networks to process images from the DeepFashion dataset and the nearest neighbor-backed recommender to generate final recommendations based on a given input image to find the most similar one in a Personalized fashion recommender system with image-based neural networks.

Deep Learning-Based FRS

DL-based fashion recommender systems have been categorized based on two main categories, including i) fashion recommender systems with single neural building blocks and ii) fashion recommender systems with deep Hybrid models. Each of this research has been introduced briefly as follows.

A data-driven innovative approach for FR was introduced by [109]. A Convolutional Neural Network and a Nearest Neighbor Based Recommender were used in this paper. After training the neural networks, an inventory is chosen for producing suggestions, and a database for the objects in the inventory is created. The nearest neighbor algorithm is used to determine the most relevant goods based on the input image, and suggestions are generated. The work in [66] suggested a semi-autonomous decision support system for assisting fashion designers by obtaining, synthesizing, and arranging data from various sources and taking the designer's preferences into account. The research in [23] proposed a unique Attentional Content-level Translation-based Recommender (ACTR) architecture that predicts both the immediate user intent and the likelihood of each intent-specific transition. Three categories of explicit instant intentions have been defined to model and forecast user intent: match, replace, and other. They improved the item-level transition modeling with many sub-transitions based on various content-level features in order to better use the peculiarities of the fashion domain and alleviate the item transition sparsity problem. The major goal of [71] was to design an outfit independent of the scratch point or an unfinished one. Fusing the product image and description information to capture the most significant, fine-grained product attributes into the item representation illustrates that attention-based fusion increases item comprehension. They indicate that Outfit recommendation faces two significant challenges: (i) item understanding, which necessitates the extraction and combination of visual and textual features to gain a better understanding, and (ii) item matching. Due to the complexity of the Item compatibility relation, they concentrated on item understanding. Because the relevance of different Item attributes in assessing compatibility may change depending on the sorts of items chosen to be matched, their suggested model was fed with picture embeddings and equivalent description embeddings as triplets. These triplets are sent to a semantic space since semantic spaces are better at capturing the ideas of picture similarity, text similarity, and image-text similarity. The research in [71] focused on relevant parts of the input using an attention mechanism. In order to bring fine-grained Item features to the surface, neural machine translation has been implemented in the attention process itself. It should be noted that this is the first time this approach has been applied to improve item comprehension in FRS. They examined numerous attention techniques to merge visual and textual information to discover superior performance while developing the proposed system, including Visual Dot Product Attention, Stacked Visual Attention, Visual L-scaled Dot Product Attention, and Co-attention. The models were tested on the fashion compatibility (FC) task and the fill-in-the-blank (FITB) task. The images were represented using the ResNet18 architecture [99], pre-trained on ImageNet. A bidirectional LSTM is used to represent the text descriptions. Both the ResNet18 architecture and the bidirectional LSTM have had their settings fine-tuned. The ADAM optimizer is used to train all models for ten epochs. All models are trained for five runs to reduce the impact of negative sampling. The performance is calculated by averaging the performance of the FC and FITB tasks throughout the five runs. This study found "the attention-based fusion mechanism is capable of integrating visual and textual information in a more purposeful manner than common space fusion [71]. attention-based fusion might also improve item comprehension by fusing information from the product image and description to capture the most significant, fine-grained product attributes into the item representation. The authors of [139] created aesthetic-aware apparel recommendation algorithms. Proposing a cutting-edge aesthetic deep model tensor factorization model that has been optimized with paired learning and negative sampling procedures. The work in [8] developed a DNN-based model that forecasts purchase likelihood for the customer–item pair, whereas the angle between vectors measures item similarity. Later, [42] expanded on the work of [89] by combining visual and historical user feedback data. Their suggested method included visual information in Bayesian Personalized Ranking with Factorization Matrix as the underlying predictor. Fashion similarity is computed as a K-nearest-neighbors issue. Interactions between users and items help to learn fashionability. [34] suggested a method for extracting style embeddings with a particular emphasis on style information such as textures, printing, material, and so on. The authors of [59] proposed a visually aware strategy as a solution to the fashion domain's design and recommendation demands. They enhanced a Bayesian personalized ranking

(BPR) formulation and used Siamese convolution neural networks (SCNN) to provide a fashion-aware visual representation of the items. [62] used Progressive Growing of GANs (P-GANs) to create designs and developed a clothing image generation method. In another work [51], researchers developed the GRU4REC approach with an alternative session-based nearest neighbor method, proposing the next item in an anonymous Session as session-based recommendations. Following the STAMP (Short-Term Attention/Memory Priority Model for Session-based Recommendation) model, while [127] proposed a session-based complementary FR strategy to tailor complementary item recommendations in the fashion domain. On the other hand, the work in [101] experimented with a method for re-ranking the most relevant items from the original recommendations to improve the similar-item recommendation by using an attention network to encode the user's session information in the session-based recommender field, employing the two-stage architecture of neural networks with a clear separation of candidate selection and ranking generation. The work presented in [89] used human preference modeling to uncover correlations between the appearances of pairs of items, which mapped compatible items to embeddings close in the latent space. In contrast, most approaches based on content extraction try to understand similarities or complementary relationships between items like a human brain does. The research in [89] designs graphs of images to address a network inference problem. Their proposed fashion item encoder employed both textual and visual attributes to understand the suitability of a complementary item. while [114] considered a sequential pattern of behavior (the most recent purchased items). They proposed a Convolutional Sequence Embedding Recommendation Model (Caser), Considering each user as a sequence of items that have interaction in the past and projection for the future of the most top-N potential interaction. The research [57] addressed a clothing retrieval problem by employing the human parsing method. Pose estimation generated an initial probability map of the human body for garment segmentation, and the segments were then categorized using locality-sensitive hashing. The visually equivalent objects were discovered by summing up the overlap similarities. Then, image retrieval algorithms with sub-linear complexity indexes were used to extract comparable objects from each of the discovered classes. FashionNet, as proposed in [41], comprises two components: a featured network for feature extraction and a matching network for compatibility computation. Users are recommended the outfits with the most outstanding ratings. In another work [63], researchers suggested a two-step deep learning framework for recommending fashion garments based on the visual resemblance style of another image. The neural style transfer technique is used to fashion in [28] to synthesize new personalized outfits. They devised a method for creating new personalized outfits based on a user's preferences and learning the user's fashion preferences from a small group of items in their wardrobe. The approach provided by [92] creates suggestions for complementary garment products given a query item. A Siamese network is utilized for feature extraction, followed by a fully connected network to learn a fashion compatibility measure. The research [118] learned feature transformation for compatibility measurements between pairs of objects using a Siamese CNN architecture. They modeled compatibility using co-occurrence data from large-scale user activity. A deep model that learns a unified feature representation for both users and pictures was given in [31]. This is accomplished by converting diverse user-image networks into homogenous low-dimensional representations. Another proposed approach is based on the idea that an item-level proxy can substitute for outfit compatibility [117]. A separate space was assigned to each item category pair to compute the compatibility between the items of each category. This approach increases the algorithm's time complexity and loses the relationships between the different pairs and the other subsets in an outfit. The approach uses a scoring system on a property (being close to that same space) to force embedding for items to be close in a generally shared space. Instead of simply learning the embedding of each item in the dataset in a common shared space, the authors create a first embedding space by using visual features extracted from a CNN and features representing the item's textual description via a visual-semantic loss; as a second, the authors use a learned projection that maps the general embedding to a secondary embedding space that scores compatibility between two different item types. The embeddings are then combined with a generalized distance measure to get object compatibility scores. The authors of [74] proposed an automated composition approach for fashion outfit candidates based on appearances and meta-data. The scoring module evaluates the aesthetics and set compatibility of instances. Their suggested model jointly learned modality embedding and fused modalities. As a similarity criterion, pictures are typically compared from only one unique perspective while learning similarity. While similarity cannot be represented in a single space, while [119] introduced Conditional Similarity Networks (CSNs), which train embeddings divided into semantically split subspaces that capture the various notions of similarities. Visual Bayesian Personalized Ranking (VBPR) was one of the earliest attempts to use visual material to develop a unified hybrid recommendation system [42]. Using heuristic graph convolution, GNNs have lately exhibited outstanding performance in graph data modeling [65, 129]. The researchers also proposed creating a heterogeneous graph of customers, clothes, and goods and employing hierarchical GNNs to encourage personalized outfits [25]. The researchers in [30] outlined the method of separating the content and style of

Table 3 Deep neural network-based fashion recommender systems

	Factor	Method	Literature
Input	Side information	Utilize (image/ Image and text)	[8, 28–31, 34, 41, 42, 42, 57, 59, 63, 65, 71, 74, 89, 92, 109, 117, 119, 129, 139]
	Behavior type	User clicking records/ interaction history	[23, 62, 101]
		User past feedback	[42]
		Sequential pattern of behavior (the most recent purchased items)	[114]
	Repeat consumption	User's purchased items, purchased/viewed items, user's co-purchase data	[23, 89, 118]
Model Structure	FRS with single Neural building blocks	P-GANs GNN STAMP NARM CNN SCNN AM+MTL CNN+KNN CNN+WNN CNN+SVM Deep CNN+KNN GRU4REC+KNN	[23, 28–31, 34, 42, 42, 51, 62, 63, 65, 89, 101, 101, 109, 118, 127, 129]

photos and recombining them to produce new images using image representations encoded by many layers of VGGNet. Table 3 also displays the input factor of each fashion recommender system.

Transformer-Based FRS

As textual data are widely used in recommendation systems, there is a growing interest in applying advanced sequential recommendation methods, which view customers as a sequence of interactions over time [26]. This approach helps to understand better customer preferences that change over time and simplifies feature engineering and modeling efforts. The attention mechanism was one of the models that are often used in sequential recommendation approaches [15, 58, 112]. Transformer-based models are particularly promising due to their ability to model long-range sequences and scalability. However, many previous studies have only evaluated models offline using open datasets, which may not be directly comparable to real-world datasets. Additionally, some studies do not consider side information like categorical inputs for items or customers, which is important for achieving the full potential of deep learning recommender systems [26, 58, 112]. More advanced studies focus on the application of Attention-based Fashion Recommendation Algorithms in different types of interactions with various fashion entities like items (such as shirts), outfits, and influencers, considering their diverse features [22]. Transform-based models are less frequently utilized for recommendation systems that involve visual data when compared to recommendation systems that deal with textual data. Rohan et al. [102] presented an Outfit-based Transformer model which can learn the compatibility of an entire outfit and enables large-scale complementary item retrieval. The model takes in an unordered set of items as input and utilizes a self-attention mechanism to learn the correlations between the items. Lorbert et al. [84] used a simpler single-layer self-attention-based framework for outfit generation, The study concentrated on providing clothing or accessory suggestions to customers based on their current outfits and the type of item they want to add to their outfits. Lin et al. [78] suggested utilizing the attention mechanism to combine the weights produced by the CNN, the category, and the target category. The resulting framework will aid in selecting appropriate subspace embeddings for the final embedding computation, which will enhance item retrieval for outfits.

Quantitative Measures of FRS

In general, different quantitative measures, such as FITB or Compatibility Estimation (CE), can be used to evaluate the accuracy of the outfit recommendation task. The outfit recommendation task can be evaluated using ranking accuracy measurements where the outfits offered to the user are necessary.

- *Compatibility estimation* Compatibility Estimation is a technique for evaluating a model's capacity to distinguish between compatible and incompatible outfits. Typically, for each compatible outfit, an incompatible instance may be generated by changing one item at a randomly selected location with another item from the clothing items dictionary [93]. This produced example is then tagged as incompatible. The objective is then transformed into a binary classification issue in which the model is trained with compatible and incompatible clothing. The area under the curve (AUC) of the receiver operating characteristic (ROC) is a commonly utilized assessment measurement in this method.

- *Fill in the Blanks* Fill in the Blanks is a standard fashion compatibility test [53]. The typical formulation is $= \frac{|guessed\ missing\ items|}{|questions|}$: One of the clothing items is randomly masked out of an ensemble consisting of n clothing items. The next step is to estimate which item in the whole outfit will go best with the others. The appropriately recommended items are the guessed missing items, and questions are the subsets of options for recommendation [93].
- *Unconstrained Outfit Completion (UOC)* All metrics (precision, recall, F1) utilized in the UOC task are averaged over all outfits in the test sets and across all removed pieces from each outfit. In many cases to evaluate, the missing item is only once per category, which restricts some of the metrics used for the UOC task, particularly in the Polyvore datasets [93].
- *Outfits ranking accuracy* Outfits ranking accuracy can be measured employing metrics such as Normalized Discounted Cumulative Gain (NDCG). Applying ranking metrics for ranking outfits is not considered an easy task, as it demands designing a representation of an outfit composed of various clothing pieces as an appropriate recommendation to the user [53]. This necessitates specifying guidelines for the user's positive or neutral choices. Some studies show that the outfit made or rated by the user (like/view) is a positive outfit, whereas the rest are neutral. The relevance of the top-n recommended outfits is then measured using NDCG, a commonly utilized criterion for comparing ranked lists [53].

Conclusion

Answering what distinguishes the fashion domain from that of other recommender systems leads to identifying fashion domain peculiarities. The main reasons why generic recommender systems cannot meet the needs in the FRS domain are: first, the subtle and subjective nature of fashion to be understood; second, while the fashion domain can be understood primarily through visual appearance properties and clothing ontology, the system should be able to handle a high number of items, a large number of attributes in each item, and a considerable number of interrelationships, as well as the high-dimensional and semantically complex features associated. We discuss all these concepts to demonstrate how closely they are interconnected. Toward answering how image-based fashion recommender systems have been affected by computer vision advancements, research provides a trajectory of the evolvement of computer vision techniques beginning from employing conventional image processing techniques such as SURF, SIFT, and so on to more recent ones utilizing Convolutional neural networks and vision transformers for image representations. It shows a big jump in proposing creative concepts like the style and then designs in this field, mainly since 2012 with CNNs. Observing the most recent advancements in using deep learning methods in the FRS domain indicates that the focus of studies has shifted from using just one neural building block to using deep hybrid models in FR's architectures. Due to the capability of Deep learning methods concerning the characteristics of the domain, including large dimensionality due to large amounts of attributes, the nonlinear nature of the relationship among features, and complex semantics, employing deep learning methods excels. In some cases, using these methods combined with conventional ones has been recommended.

Funding Information Open access funding provided by Lulea University of Technology. This study was not part of any funding.

Declarations

Conflict of interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

Compliance with ethical standards The authors of this article have not performed any studies with human subjects and animals.

References

1. Adomavicius G, Tuzhilin A. Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions. IEEE Trans Knowl Data Eng. 2005;17(6):734–49.
2. Alashkar T, Jiang S, Wang S, Fu Y. Examples-rules guided deep neural network for makeup recommendation. In: Proceedings of the AAAI conference on artificial intelligence, vol. 31. 2017.
3. Al-Halah Z, Stiefelhagen R, Grauman K. Fashion forward: forecasting visual style in fashion. In: Proceedings of the IEEE international conference on computer vision. 2017. p. 388–397.

4. Atharv Pandit Kunal Goel MJ, Katre N. A review on clothes matching and recommendation systems based on user attributes. Int J Eng Res Technol (IJERT). 2020;09(08).
5. Bettaney EM, Hardwick SR, Zisimopoulos O, Chamberlain BP. Fashion outfit generation for e-commerce. In: Joint European conference on machine learning and knowledge discovery in databases. Springer. 2020. p. 339–354.
6. Bollacker K, Díaz-Rodríguez N, Li X. Beyond clothing ontologies: modeling fashion with subjective influence networks. In: KDD workshop on machine learning meets fashion. 2016.
7. Bossard L, Dantone M, Leistner C, Wengert C, Quack T, Van Gool L. Apparel classification with style. In: Asian conference on computer vision. Springer. 2012. p. 321–335.
8. Bracher C, Heinz S, Vollgraf R. Fashion dna: merging content and sales data for recommendation and article mapping. 2016. arXiv preprint arXiv:1609.02489.
9. Cardoso Â, Daolio F, Vargas S. 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 & data mining. 2018. p. 80–89.
10. Chauhan N, Mahesh G. Comparison of different image retrieval techniques in cbir. In: National conference on computer science & security. 2013. p. 1–5. https://doi.org/10.13140/2.1.3033.6642.
11. Chen L, He Y. Dress fashionably: Learn fashion collocation with deep mixed-category metric learning. In: Proceedings of the AAAI conference on artificial intelligence, vol. 32. 2018.
12. Chen LC, Yang Y, Wang J, Xu W, Yuille AL. Attention to scale: Scale-aware semantic image segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2016. p. 3640–3649.
13. Chen J, Zhang H, He X, Nie L, Liu W, Chua TS. 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. 2017. p. 335–344.
14. Chen W, Huang P, Xu J, Guo X, Guo C, Sun F, Li C, Pfadler A, Zhao H, Zhao B. Pog: personalized outfit generation for fashion recommendation at alibaba ifashion. In: Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining. 2019. p. 2662–2670.
15. Chen Q, Zhao H, Li W, Huang P, Ou W. Behavior sequence transformer for e-commerce recommendation in alibaba. In: Proceedings of the 1st international workshop on deep learning practice for high-dimensional sparse data. 2019. p. 1–4.
16. Chen YC, Li L, Yu L, El Kholy A, Ahmed F, Gan Z, Cheng Y, Liu J. Uniter: Universal image-text representation learning. In: European conference on computer vision. Springer. 2020. p. 104–120.
17. Cheng W, Song S, Chen C, Hidayati SC, Liu J. Fashion meets computer vision: a survey. CoRR. 2020. arXiv:2003.13988.
18. Chopra A, Sinha A, Gupta H, Sarkar M, Ayush K, Krishnamurthy B. Powering robust fashion retrieval with information rich feature embeddings. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops. 2019.
19. Cosido O, Iglesias A, Galvez A, Catuogno R, Campi M, Terán L, Sainz E. Hybridization of convergent photogrammetry, computer vision, and artificial intelligence for digital documentation of cultural heritage-a case study: the magdalena palace. In: 2014 International conference on cyberworlds. IEEE; 2014. p. 369–376
20. Craik J. Fashion: the key concepts. Oxford: Berg Publishers; 2009.
21. Cui Z, Li Z, Wu S, Zhang XY, Wang L. Dressing as a whole: Outfit compatibility learning based on node-wise graph neural networks. In: The World Wide Web Conference. 2019. p. 307–317.
22. Deldjoo Y, Nazary F, Ramisa A, Mcauley J, Pellegrini G, Bellogin A, Di Noia T. A review of modern fashion recommender systems. 2022. arXiv preprint arXiv:2202.02757.
23. Ding Y, Ma Y, Wong W, Chua TS. Modeling instant user intent and content-level transition for sequential fashion recommendation. In: IEEE transactions on multimedia. 2021.
24. Dong Q, Gong S, Zhu X. Multi-task curriculum transfer deep learning of clothing attributes. In: 2017 IEEE winter conference on applications of computer vision (WACV). IEEE; 2017. p. 520–529.
25. Elahi M, Qi L. Fashion recommender systems in cold start. In: Fashion recommender systems. Springer; 2020. p. 3–21.
26. Fang H, Zhang D, Shu Y, Guo G. Deep learning for sequential recommendation: algorithms, influential factors, and evaluations. ACM Trans Inf Syst (TOIS). 2020;39(1):1–42.
27. Gajic B, Baldrich R. Cross-domain fashion image retrieval. In: Proceedings of the IEEE conference on computer vision and pattern recognition workshops. 2018. p. 1869–1871.
28. Ganesan A, Oates T, et al. Fashioning with networks: neural style transfer to design clothes. 2017. arXiv preprint arXiv:1707.09899.
29. Gatys LA, Ecker AS, Bethge M. A neural algorithm of artistic style. 2015. arXiv preprint arXiv:1508.06576.
30. Gatys LA, Ecker AS, Bethge M. Image style transfer using convolutional neural networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2016. p. 2414–2423.
31. Geng X, Zhang H, Bian J, Chua TS. Learning image and user features for recommendation in social networks. In: Proceedings of the IEEE international conference on computer vision. 2015. p. 4274–4282.
32. Gkelios S, Sophokleous A, Plakias S, Boutalis Y, Chatzichristofis SA. Deep convolutional features for image retrieval. Expert Syst Appl. 2021;177: 114940.
33. Goldberg D, Nichols D, Oki BM, Terry D. Using collaborative filtering to weave an information tapestry. Commun ACM. 1992;35(12):61–70.
34. Goncalves D, Liu L, Magalhães A. How big can style be? addressing high dimensionality for recommending with style. 2019. arXiv preprint arXiv:1908.10642.
35. Goodfellow I, Bengio Y, Courville A. Deep learning. Cambridge: MIT press; 2016.
36. Graves A. Generating sequences with recurrent neural networks. 2013. arXiv preprint arXiv:1308.0850.
37. Guan C, Qin S, Ling W, Ding G. Apparel recommendation system evolution: an empirical review. Int J Cloth Sci Technol. 2016.
38. Guo Y, Liu Y, Oerlemans A, Lao S, Wu S, Lew MS. Deep learning for visual understanding: a review. Neurocomputing. 2016;187:27–48.
39. Han X, Wu Z, Jiang YG, Davis LS. Learning fashion compatibility with bidirectional lstms. In: Proceedings of the 25th ACM international conference on Multimedia. 2017. p. 1078–1086.
40. Harada F, Shimakawa H. Outfit recommendation with consideration of user policy and preference on layered combination of garments. Int J Adv Comput Sci. 2012;2:49–55.
41. He T, Hu Y. Fashionnet: personalized outfit recommendation with deep neural network. 2018. arXiv preprint arXiv:1810.02443.
42. He R, McAuley J. 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. 2016. p. 507–517.

43. Heger G. The capsule closet phenomenon: a phenomenological study of lived experiences with capsule closets. 2016.
44. Hill W, Stead L, Rosenstein M, Furnas G. Recommending and evaluating choices in a virtual community of use. In: Proceedings of the SIGCHI conference on Human factors in computing systems. 1995. p. 194–201.
45. Hou M, Wu L, Chen E, Li Z, Zheng VW, Liu Q. Explainable fashion recommendation: a semantic attribute region guided approach. 2019. arXiv preprint arXiv:1905.12862.
46. Hsiao WL, Grauman K. Creating capsule wardrobes from fashion images. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2018. p. 7161–7170.
47. Hu X, Zhu W, Li Q. Hcrs: a hybrid clothes recommender system based on user ratings and product features. 2014. arXiv preprint arXiv:1411.6754.
48. Hu Y, Yi X, Davis LS. Collaborative fashion recommendation: A functional tensor factorization approach. In: Proceedings of the 23rd ACM international conference on Multimedia. 2015. p. 129–138.
49. Huang J, Feris RS, Chen Q, Yan S. Cross-domain image retrieval with a dual attribute-aware ranking network. In: Proceedings of the IEEE international conference on computer vision. 2015. p. 1062–1070.
50. Iwata T, Watanabe S, Sawada H. Fashion coordinates recommender system using photographs from fashion magazines. In: Twenty-second international joint conference on artificial intelligence. 2011.
51. Jannach D, Ludewig M. When recurrent neural networks meet the neighborhood for session-based recommendation. In: Proceedings of the eleventh ACM conference on recommender systems. 2017. p. 306–310.
52. Jaradat S, Dokoohaki N, Hammar K, Wara U, Matskin M. Dynamic cnn models for fashion recommendation in Instagram. In: 2018 IEEE Intl conf on parallel & distributed processing with applications, ubiquitous computing & communications, big data & cloud computing, social computing & networking, sustainable computing & communications (ISPA/IUCC/BDCloud/SocialCom/SustainCom). IEEE. 2018. p. 1144–1151.
53. Jaradat S, Dokoohaki N, Corona Pampin HJ, Shirvany R. Workshop on recommender systems in fashion and retail. In: Fifteenth ACM conference on recommender systems. 2021. p. 810–812.
54. Jia J, Huang J, Shen G, He T, Liu Z, Luan H, Yan C. Learning to appreciate the aesthetic effects of clothing. In: Proceedings of the AAAI conference on artificial intelligence, vol. 30. 2016.
55. Jiang S, Wu Y, Fu Y. Deep bi-directional cross-triplet embedding for cross-domain clothing retrieval. In: Proceedings of the 24th ACM international conference on Multimedia. 2016. p. 52–56.
56. Jiang Y, Qianqian X, Cao X. Outfit recommendation with deep sequence learning. In: 2018 IEEE fourth international conference on multimedia big data (BigMM). IEEE; 2018. p. 1–5.
57. Kalantidis Y, Kennedy L, Li LJ. Getting the look: Clothing recognition and segmentation for automatic product suggestions in everyday photos. In: Proceedings of the 3rd ACM conference on international conference on multimedia retrieval. Association for Computing Machinery; 2013. p. 105–112.
58. Kang WC, McAuley J. Self-attentive sequential recommendation. In: 2018 IEEE international conference on data mining (ICDM), IEEE. 2018. p. 197–206.
59. Kang WC, Fang C, Wang Z, McAuley J. Visually-aware fashion recommendation and design with generative image models. In: 2017 IEEE international conference on data mining (ICDM). IEEE; 2017. p. 207–216.
60. Kang WC, Kim E, Leskovec J, Rosenberg C, McAuley J. Complete the look: scene-based complementary product recommendation. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition; 2019. p. 10532–10541.
61. Karatzoglou A, Amatriain X, Baltrunas L, Oliver N. Multiverse recommendation: n-dimensional tensor factorization for context-aware collaborative filtering. In: Proceedings of the fourth ACM conference on Recommender systems. 2010. p. 79–86.
62. Kato N, Osone H, Oomori K, Ooi CW, Ochiai Y. Gans-based clothes design: Pattern maker is all you need to design clothing. In: Proceedings of the 10th augmented human international conference 2019. 2019. p. 1–7.
63. Keerthi Gorripati S, Angadi A. Visual based fashion clothes recommendation with convolutional neural networks. Int J Inf Syst Manag Sci. 2018;1(1).
64. Kiapour MH, Yamaguchi K, Berg AC, Berg TL. Hipster wars: Discovering elements of fashion styles. In: European conference on computer vision, Springer; 2014. p. 472–488.
65. Kipf T, Fetaya E, Wang KC, Welling M, Zemel R. Neural relational inference for interacting systems. In: International conference on machine learning, PMLR; 2018. p. 2688–2697.
66. Kotouza MT, Tsarouchis SF, Kyprianidis AC, Chrysopoulos AC, Mitkas PA. Towards fashion recommendation: an AI system for clothing data retrieval and analysis. In: IFIP international conference on artificial intelligence applications and innovations, Springer; 2020. p. 433–444.
67. Kouge Y, Murakami T, Kurosawa Y, Mera K, Takezawa T. Extraction of the combination rules of colors and derived fashion images using fashion styling data. In: Proceedings of the international multiconference of engineers and computer scientists, vol. 1. 2015.
68. Kovashka A, Parikh D, Grauman K. Whittlesearch: image search with relative attribute feedback. In: 2012 IEEE conference on computer vision and pattern recognition. IEEE; 2012. p. 2973–2980.
69. Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. Adv Neural Inf Process Syst. 2012;25:1097–105.
70. Kuang Z, Gao Y, Li G, Luo P, Chen Y, Lin L, Zhang W. Fashion retrieval via graph reasoning networks on a similarity pyramid. In: Proceedings of the IEEE/CVF international conference on computer vision. 2019. p. 3066–3075.
71. Laenen K, Moens MF. Attention-based fusion for outfit recommendation. In: Fashion recommender systems. Springer; 2020. p. 69–86.
72. Lee H, Seol J, Lee Sg. Style2vec: Representation learning for fashion items from style sets. 2017. arXiv preprint arXiv:1708.04014.
73. Li J, Zhong X, Li Y. A psychological decision making model based personal fashion style recommendation system. In: Proceedings of the international conference on human-centric computing 2011 and embedded and multimedia computing 2011. Springer; 2011. p. 57–64.
74. Li Y, Cao L, Zhu J, Luo J. Mining fashion outfit composition using an end-to-end deep learning approach on set data. IEEE Trans Multimedia. 2017;19(8):1946–55.
75. Li X, Yang J, Ma J. Recent developments of content-based image retrieval (cbir). Neurocomputing; 2021.
76. Lei C, Liu D, Li W, Zha ZJ, Li H. Comparative deep learning of hybrid representations for image recommendations. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2016. p. 2545–2553.
77. Limaksornkul C, Nakorn DN, Rakmanee O, Viriyasitavat W. Smart closet: statistical-based apparel recommendation system. In: 2014 Third ICT international student project conference (ICT-ISPC). IEEE; 2014. p. 155–158
78. Lin YL, Tran S, Davis LS. Fashion outfit complementary item retrieval. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020. p. 3311–3319.

79. Liu S, Feng J, Song Z, Zhang T, Lu H, Xu C, Yan S. Hi, magic closet, tell me what to wear! In: Proceedings of the 20th ACM international conference on Multimedia. 2012. p. 619–628.
80. Liu S, Liu L, Yan S. Fashion analysis: current techniques and future directions. IEEE Multimedia. 2014;21(2):72–9.
81. Liu Z, Luo P, Qiu S, Wang X, Tang X. Deepfashion: powering robust clothes recognition and retrieval with rich annotations. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2016. p. 1096–1104.
82. Liu Q, Wu S, Wang L. Deepstyle: learning user preferences for visual recommendation. In: Proceedings of the 40th international acm sigir conference on research and development in information retrieval. 2017. p. 841–844.
83. Liu Q, Zeng Y, Mokhosi R, Zhang H. Stamp: short-term attention/memory priority model for session-based recommendation. In: Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery & data mining. 2018. p. 1831–1839.
84. Lorbert A, Neiman D, Poznanski A, Oks E, Davis L. Scalable and explainable outfit generation. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2021. p. 3931–3934.
85. Lu J, Wu D, Mao M, Wang W, Zhang G. Recommender system application developments: a survey. Decis Support Syst. 2015;74:12–32.
86. Lu Y, Kumar A, Zhai S, Cheng Y, Javidi T, Feris R. Fully-adaptive feature sharing in multi-task networks with applications in person attribute classification. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2017. p. 5334–5343.
87. Mallu M. Fashion object detection and pixel-wise semantic segmentation: crowdsourcing framework for image bounding box detection & pixel-wise segmentation. 2018.
88. Matzen K, Bala K, Snavely N. Streetstyle: Exploring world-wide clothing styles from millions of photos. 2017. arXiv preprint arXiv:1706.01869.
89. McAuley J, Targett C, Shi Q, Van Den Hengel A. Image-based recommendations on styles and substitutes. In: Proceedings of the 38th international ACM SIGIR conference on research and development in information retrieval. 2015. p. 43–52.
90. Nakamura T, Goto R. Outfit generation and style extraction via bidirectional lstm and autoencoder. 2018. arXiv preprint arXiv:1807.03133.
91. Park S, Shin M, Ham S, Choe S, Kang Y: Study on fashion image retrieval methods for efficient fashion visual search. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops. 2019. p. 0–0.
92. Polanía LF, Gupte S. Learning fashion compatibility across apparel categories for outfit recommendation. In: 2019 IEEE international conference on image processing (ICIP). IEEE; 2019. p. 4489–4493.
93. Prato G. New methodologies for fashion recommender systems. 2019.
94. Ramesh N, Moh TS. Outfit recommender system. In: 2018 IEEE/ACM international conference on advances in social networks analysis and mining (ASONAM). IEEE; 2018. p. 903–910.
95. Ravula Samatha R, Pole Laxmi D. A literature survey on computer vision towards data science. Int J Creat Res Thoughts (IJCRT). 2020;08(06).
96. Rawat YS, Kankanhalli MS. Contagnet: Exploiting user context for image tag recommendation. In: Proceedings of the 24th ACM international conference on Multimedia. 2016. p. 1102–1106.
97. Reddy KS, Sreedhar K. Image retrieval techniques: a survey. Int J Electron Commun Eng. 2016;9(1):19–27.
98. Ren S, He K, Girshick R, Sun J. Faster r-cnn: towards real-time object detection with region proposal networks. Adv Neural Inf Process Syst. 2015;28:91–9.
99. Ren S, He K, Girshick R, Zhang X, Sun J. Object detection networks on convolutional feature maps. IEEE Trans Pattern Anal Mach Intell. 2016;39(7):1476–81.
100. Rendle S, Freudenthaler C, Schmidt-Thieme L. Factorizing personalized markov chains for next-basket recommendation. In: Proceedings of the 19th international conference on World wide web. 2010. p. 811–820.
101. Rodríguez JAS, Wu JC, Khandwawala M. Two-stage session-based recommendations with candidate rank embeddings. In: Fashion recommender systems. Springer; 2020. p. 49–66.
102. Sarkar R, Bodla N, Vasileva M, Lin YL, Beniwal A, Lu A, Medioni G. Outfittransformer: outfit representations for fashion recommendation. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2022. p. 2263–2267.
103. Sha D, Wang D, Zhou X, Feng S, Zhang Y, Yu G. An approach for clothing recommendation based on multiple image attributes. In: International conference on web-age information management. Springer; 2016. p. 272–285.
104. Shen E, Lieberman H, Lam F. What am i gonna wear? scenario-oriented recommendation. In: Proceedings of the 12th international conference on Intelligent user interfaces. 2007. p. 365–368.
105. Shin M, Park S, Kim T. Semi-supervised feature-level attribute manipulation for fashion image retrieval. 2019. arXiv preprint arXiv:1907.05007.
106. Simo-Serra E, Ishikawa H. Fashion style in 128 floats: Joint ranking and classification using weak data for feature extraction. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2016. p. 298–307.
107. Song S, Mei T. When multimedia meets fashion. IEEE Multimedia. 2018;25(3):102–8.
108. Song X, Feng F, Han X, Yang X, Liu W, Nie L. Neural compatibility modeling with attentive knowledge distillation. In: The 41st International ACM SIGIR conference on research & development in information retrieval. 2018. p. 5–14.
109. Sridevi M, ManikyaArun N, Sheshikala M, Sudarshan E. Personalized fashion recommender system with image based neural networks. In: IOP conference series: materials science and engineering, vol. 981, IOP Publishing; 2020. p. 022073.
110. Stefani MA, Stefanis V, Garofalakis J. Cfrs: A trends-driven collaborative fashion recommendation system. In: 2019 10th International conference on information, intelligence, systems and applications (IISA). IEEE; 2019. p. 1–4.
111. Sullivan L. Form follows function. De la tour de bureaux artistiquement.2010.
112. Sun F, Liu J, Wu J, Pei C, Lin X, Ou W, Jiang P. Bert4rec: Sequential recommendation with bidirectional encoder representations from transformer. In: Proceedings of the 28th ACM international conference on information and knowledge management. 2019. p. 1441–1450.
113. Sun GL, He JY, Wu X, Zhao B, Peng Q. Learning fashion compatibility across categories with deep multimodal neural networks. Neurocomputing. 2020;395:237–46.
114. Tang J, Wang K. Personalized top-n sequential recommendation via convolutional sequence embedding. In: Proceedings of the eleventh ACM international conference on web search and data mining. 2018. p. 565–573.
115. Tian X, Zheng Q, Xing J. Content-based image retrieval system via deep learning method. In: 2018 IEEE 3rd advanced information technology, electronic and automation control conference (IAEAC). IEEE; 2018. p. 1257–1261.
116. Tsujita H, Tsukada K, Kambara K, Siio I. Complete fashion coordinator: a support system for capturing and selecting daily

clothes with social networks. In: Proceedings of the international conference on advanced visual interfaces. 2010. p. 127–132.

117. Vasileva MI, Plummer BA, Dusad K, Rajpal S, Kumar R, Forsyth D. Learning type-aware embeddings for fashion compatibility. In: Proceedings of the European conference on computer vision (ECCV). 2018. p. 390–405.

118. Veit A, Kovacs B, Bell S, McAuley J, Bala K, Belongie S. Learning visual clothing style with heterogeneous dyadic co-occurrences. In: Proceedings of the IEEE international conference on computer vision. 2015. p. 4642–4650.

119. Veit A, Belongie S, Karaletsos T. Conditional similarity networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2017. p. 830–838.

120. Wang X, Zhang T. Clothes search in consumer photos via color matching and attribute learning. In: Proceedings of the 19th ACM international conference on Multimedia. 2011. p. 1353–1356.

121. Wang Z, Gu Y, Zhang Y, Zhou J, Gu X. Clothing retrieval with visual attention model. In: 2017 IEEE visual communications and image processing (VCIP). IEEE;2017. p. 1–4.

122. Wang S, Wang Y, Tang J, Shu K, Ranganath S, Liu H. What your images reveal: Exploiting visual contents for point-of-interest recommendation. In: Proceedings of the 26th international conference on world wide web. 2017. p. 391–400.

123. Wang J, Ma Y, Zhang L, Gao RX, Wu D. Deep learning for smart manufacturing: methods and applications. J Manuf Syst. 2018;48:144–56.

124. Wen J, Li X, She J, Park S, Cheung M. Visual background recommendation for dance performances using dancer-shared images. In: 2016 IEEE international conference on Internet of Things (iThings) and IEEE green computing and communications (GreenCom) and IEEE cyber, physical and social computing (CPSCom) and IEEE Smart Data (SmartData). IEEE; 2016. p. 521–527.

125. Wen Y, Liu X, Xu B. Personalized clothing recommendation based on knowledge graph. In: 2018 International conference on audio, language and image processing (ICALIP). IEEE; 2018. p. 1–5.

126. Wiley V, Lucas T. Computer vision and image processing: a paper review. Int J Artif Intell Res. 2018;2(1):29–36.

127. Wu JC, Rodríguez JAS, Pampín HJC. Session-based complementary fashion recommendations. 2019. arXiv preprint arXiv:1908.08327.

128. Wu Q, Zhao P, Cui Z. Visual and textual jointly enhanced interpretable fashion recommendation. IEEE Access. 2020;8:68736–46.

129. Wu L, He X, Wang X, Zhang K, Wang M. A survey on neural recommendation: from collaborative filtering to content and context enriched recommendation. 2021. arXiv preprint arXiv:2104.13030.

130. Xiao T, Xia T, Yang Y, Huang C, Wang X. Learning from massive noisy labeled data for image classification. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2015. p. 2691–2699.

131. Xu Q, Shen F, Liu L, Shen HT. Graphcar: Content-aware multimedia recommendation with graph autoencoder. In: The 41st International ACM SIGIR conference on research & development in information retrieval. 2018. p. 981–984.

132. Yamaguchi K, Kiapour MH, Ortiz LE, Berg TL. Parsing clothing in fashion photographs. In: 2012 IEEE conference on computer vision and pattern recognition. IEEE; 2012. p. 3570–3577.

133. Yamaguchi K, Hadi Kiapour M, Berg TL. Paper doll parsing: retrieving similar styles to parse clothing items. In: Proceedings of the IEEE international conference on computer vision. 2013. p. 3519–3526.

134. Yang Y, Ramanan D. Articulated pose estimation with flexible mixtures-of-parts. In: CVPR 2011. IEEE; 2011. p. 1385–1392.

135. Yang W, Luo P, Lin L. Clothing co-parsing by joint image segmentation and labeling. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2014. p. 3182–3189.

136. Yu Z, Lian J, Mahmoody A, Liu G, Xie X. Adaptive user modeling with long and short-term preferences for personalized recommendation. In: IJCAI; 2019. p. 4213–4219.

137. Yu W, Qin Z. Graph convolutional network for recommendation with low-pass collaborative filters. In: International conference on machine learning. PMLR; 2020. p. 10936–10945.

138. Yu W, He X, Pei J, Chen X, Xiong L, Liu J, Qin Z. Visually aware recommendation with aesthetic features. VLDB J. 2021;30(4):495–513.

139. Yu W, He X, Pei J, Chen X, Xiong L, Liu J, Qin Z. Visually aware recommendation with aesthetic features. VLDB J;2021:1–19.

140. Yuan Y, Yang K, Zhang C. Hard-aware deeply cascaded embedding. In: Proceedings of the IEEE international conference on computer vision. 2017. p. 814–823.

141. Zhang X, Jia J, Gao K, Zhang Y, Zhang D, Li J, Tian Q. Trip outfits advisor: location-oriented clothing recommendation. IEEE Trans Multimedia. 2017;19(11):2533–44.

142. Zhang S, Yao L, Sun A, Tay Y. Deep learning based recommender system: a survey and new perspectives. ACM Comput Surv (CSUR). 2019;52(1):1–38.

143. Zhao B, Feng J, Wu X, Yan S. Memory-augmented attribute manipulation networks for interactive fashion search. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2017. p. 1520–1528.

144. Zhao H, Yu J, Li Y, Wang D, Liu J, Yang H, Wu F. Dress like an internet celebrity: Fashion retrieval in videos. In: Proceedings of the twenty-ninth international conference on international joint conferences on artificial intelligence. 2021. p. 1054–1060.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.