DEGREE PROJECT

Image-based fashion recommender systems

Considering Deep learning role in computer vision development

Shaghayegh Shirkhani

Master Programme in Data Science
2021

Luleå University of Technology
Department of Computer Science, Electrical and Space Engineering

LULEÅ UNIVERSITY OF TECHNOLOGY

Abstract

Fashion is perceived as a meaningful way of self-expressing that people use for different purposes. It seems to be an integral part of every person in modern societies, from everyday life to exceptional events and occasions. Fashionable products are highly demanded, and consequently, fashion is perceived as a desirable and profitable industry. Although this massive demand for fashion products provides an excellent opportunity for companies to invest in fashion-related sectors, it also faces different challenges in answering their customer needs. Fashion recommender systems have been introduced to address these needs. This thesis aims to provide deeper insight into the fashion recommender system domain by conducting a comprehensive literature review on more than 100 papers in this field focusing on image-based fashion recommender systems considering computer vision advancements. Justifying fashion domain-specific characteristics, the subtle notions of this domain and their relevancy have been conceptualized. Four main tasks in image-based fashion recommender systems have been recognized, including cloth-item retrievals, Complementary item recommendation, Outfit recommendation, and Capsule wardrobes. An evolvement trajectory of image-based fashion recommender systems concerning computer vision advancements has been illustrated consists of three main eras and the most recent developments. Finally, a comparison between traditional computer vision techniques and deep learning-based has been made. Although the main objective of this literature review was to perform a comprehensive, integrated overview of researches in this field, there is still a need for conducting further studies considering image-based fashion recommender systems from a more practical perspective.

Table of Contents

1. Introduction

In this chapter, we introduce the significance of the research and the gap, our objectives, and contributions to addressing these gaps, highlighting the main research questions.

1.1. Significance of research

There are three main reasons which motivate us to perform this research are outlined here as follows:

- The importance of the subject in the industry domain
- The gap exists in the research area
- The importance of doing this comprehensive literature review as the first stage of a 3-staged research design

Here we point out the gap which exists in researches and academic studies in the fashion recommender domain; the importance of the fashion domain and consequently fashion recommender systems have been discussed in section 2.1., the third significance of research as a cornerstone of three-staged designed research also has been explained in objective section 1.4.

Although there are some studies have been done reviewing recommender systems in general, during recent years, no research work, to the best of our knowledge, has been focused on reviewing particularly in-depth imaged-based fashion recommender systems considering deep learning advancements from a computer vision perspective, while the study of this domain is very significant for both researchers and real-world developers because of the great demand which exists in different layers of the apparel and fashion industry value chain concerning the rapid growth of e-commerce industry. Through this thesis, we are going to bridging this gap. In the following, we introduce a few studies which are rather close to the fashion domain, such as in (Cheng et al., 2021; S. Song et al., 2018; C. Guan et al., 2017; J. Lu et al., 2015).

(C. Guan et al., 2017) studied fashion recommendation systems in the apparel industry which classified the fashion recommender systems into four categories, including style searching/retrieval, fashion coordination, wardrobe recommendation, and intelligent expert systems, based on some matching criteria (or models), Also in " Fashion Analysis " (S. Liu et al., 2014) specified two main streams of researches in fashion analysis: clothing analysis and facial beauty (including makeup and hairstyle) analysis. S. Liu et al. believed that Clothing analysis tasks include clothing recommendation, retrieval, and parsing. In 2018, Song and Mei (S. Song et al., 2018) introduced the progress in fashion research with multimedia, which categorized the fashion tasks into three aspects: low-level pixel computation, mid-level fashion understanding, and high-level fashion analysis. Low-level pixel computation aims to generate pixel-level labels on the image, such as human segmentation, landmark detection, and human pose estimation. Mid-level fashion understanding aims to recognize fashion images, such as fashion items and fashion styles.

High-level fashion analysis includes recommendation, fashion synthesis, and fashion trend prediction. More recently, “Fashion Meets Computer Vision: A Survey” introduced intelligent fashion with respect to the role of computer vision in fashion. (Cheng et al., 2020) categorized fashion research topics into four main categories: detection, analysis, synthesis, and recommendation; as its name suggest, this review has been focused on the intersection of fashion and computer vision. Although this research is valuable in its case, it seeks a wide perspective on the generality of fashion and has not focused on image-based Fashion Recommender Systems (FRS).

1.2. Research Questions

The main questions we aimed to answer through this literature review can be outlined as follows:

(a)What makes the fashion domain distinctive from other recommender systems domains?

(b)What are the main tasks which have been defined for fashion recommender systems?

(c)How image-based fashion recommender systems have been affected by computer vision advancements?

1.3. Contribution

The contributions of our work in this thesis can be summarized as follows:

- We provide a comprehensive survey in the fashion recommender domain, including the most recent research progress.
- We point out the main areas of complexity in FRS
- We illustrate how conceptually are interconnected the main concepts of FRS in an integrated structure.
- For each aspect of the beforementioned areas, we introduce significant and recent researches.
- Main tasks in FRS have been identified and categorized.
- We introduce prominent studies in each category of tasks.
- We investigate the evolution of computer vision's most frequent technique in image-based fashion recommender systems.
- We investigate the role of deep learning in computer vision approaches used in fashion recommender systems.
- We categorized the most recent studies in fashion recommender systems based on the deep learning method they used

1.4. Research Objectives

This thesis has been designed as a part of three staged research on personalized fashion recommender systems, as illustrated in the Figure below. What has been done through this research is considered as stage one.

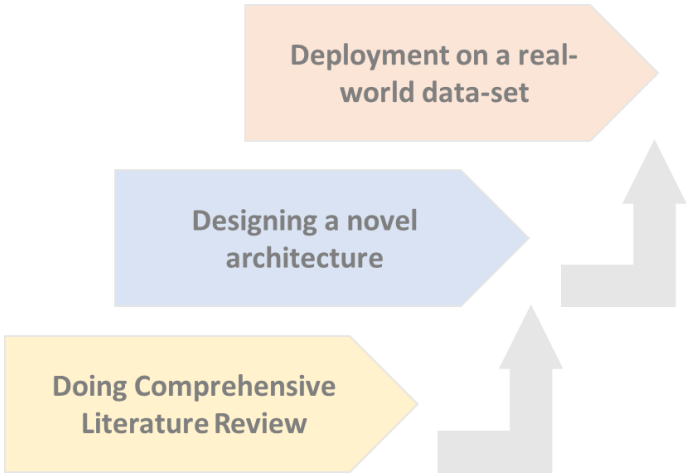

Figure 1: The three-staged research study on proposing a novel personalized fashion recommender system

With deep-diving into the subject, we aimed to perform comprehensive and fruitful research by taking the issue from determining aspect to account, particularly when we find out that there is a lack of conceptually and methodologically integrated multifaceted study. Nonetheless, there are many precious researches that have been done by different scholars from different perspectives, remarkably during recent years. Consequently, this gap motivates us to create a bigger picture in this context by putting different pieces of the puzzle along with each other. While literature review is known as a research method, it has also been considered as a cornerstone in doing any scientific research; by doing so, not only we will be equipped with getting more profound insight into our subject, but also this prepares us with a robust framework to choose/create the best fit method for what we supposed to do in subsequent steps by avoiding follow up one-sided approach. We hope this literature review will be illuminative for the next researchers who have the same interest in this field.

2. Methodology

The aforementioned research questions and objectives will be addressed through this literature review thesis, respectively. The literature review is perceived as a scientific method of performing research. There are many reasons behind this importance; there are some of the more chief reasons, including putting the most related previous studies together to get a better understanding of prior researchers' knowledge, ideas, and experiences to generate new knowledge with both understanding what has been done before, and what is remained as a gap to be filled, these will lead to improving the researcher mindset with comprehension of limitations and findings in the context of their research topic. Furthermore, this is important to locate the research stream with the determination of the scope of the research. Nonetheless, while past findings and experience can help support a proposition, keep in mind that they are not a substitute for logical reasoning; according to (Sutton et al., 1995), we put logical reasoning on top of all previous research through this thesis. Thereby, we are going to build a more clarified, integrated perspective here.

We follow the guidelines of (Brocke et al., 2009) to explain our methodology based on their proposed framework for the literature review, which contains five main steps, illustrated as Figure 2. This framework is concentrated on the process of searching the literature for performing IS literature reviews. In the first stage, the definition of review scope focuses on clarifying the scope and flavor of the study. The scope definition is perceived as a major challenge in reviewing the literature (Brocke et al., 2009). In the second phase, topic conceptualization emphasizes how a review must begin with the definition of the critical terms (Zorn et al., 2006), with "a conception of what is known about the topic and potential areas where knowledge may be needed" (Torraco, 2005). The search process has been remarked as the third step, including "database, keyword, backward, and forward search, as well as an ongoing evaluation of the sources" (Brocke et al., 2009). The previous step's outcome as collected literature on the research topic has to be analyzed and synthesized in the next stage. Concept- centric approach developed by (Salipante et al., 1982), and according to (Webster et al., 2002), has been suggested for review structuring. (Brocke et al., 2009) also mentioned that using concept-matrix is helpful in this phase since it "subdivides topic-related concepts into different units of analysis and allows for arranging, discussing, and synthesizing prior research." The outcomes of this stage consist of "sharper and more insightful questions for future research (Webster et al., 2002), "which is expected to result in a research agenda" (Brocke et al., 2009) as the final stage of this process.

Figure 2: The framework for literature review, Brocke et al., 2009

Here, we map out how this study has been done concerning the forementioned literature review framework proposed by (Brocke et al., 2009) as follows:

1. **Definition of review scope**:

we follow "Cooper´s taxonomy" framework for this purpose, as suggested by (Brocke et al., 2009). This framework is comprised of 6 main characteristics, each consisting of different classifications. According to (Brocke et al., 2009), some of these categories have to be determined exclusively, such as perspective and coverage. In contrast, the rest can be combined, including audience, organization, goal, and focus. Depicted as Table 1., The scope specifications of this thesis are remarked by grey light, corresponding with "Cooper´s taxonomy" framework.

Table 1. “Taxonomy of Literature Reviews” by Cooper, 1988

characteristic	categories			
Focus	Research outcomes	Research methods	theories	applications
Goal	integration	criticism	Central issues	
Organization	historical	conceptual	methodological	
Perspective	Neutral representation		Espousal of position	
Audience	Specialized scholars	General scholars	Practitioners/politicians	General public
coverage	Exhaustive	Exhaustive and selective	representative	Central/pivotal

The focus of the research has been identified through four different categories: research outcomes, research methods, theories, and applications. In this thesis, the focus is mostly on research methods considering the applications. The goal of the literature review may be integration, criticism, or central issues. The purpose of the present thesis is integration. While the organization of the literature review can be historical, conceptual, or methodological, the organization in this thesis is both conceptual and methodological. The perspective of the review indicates that “whether a certain position is espoused or not” (Brocke et al., 2009). It can be neutral representation or espousal position. In this thesis, the position is neutral. The audience of a literature review may be different among specialized scholars, general scholars, practitioners/ politicians, and the general public; however, according to (Brocke et al., 2009), it also can be combined, as it is within this thesis. According to Cooper, four levels of coverage can be recognized, including exhaustive, exhaustive and selective, representative, or central/pivotal. Based on the definition “exhaustive (including the entirety of literature on a topic or at least most of it)” (Brocke et al., 2009), the coverage of this thesis is considered as exhaustive and selective with reviewing most of the related literature with the topic and in some cases highlighting the most significant ones.

2. Conceptualization of the topic:

Regarding (Brocke et al., 2009), the key terms mostly related to the topic have to be identified in this step. Concept mapping has been suggested in this regard within the same reference. It has to be said that finding these key words are perceived as significant of importance in the literature review, mostly because searching these keywords will result in finding material that contains highly relative content which not only will be efficient in structuring our mindset but will lead to getting a better understanding of the researcher position with respect to the purpose and scope of the research via solving the puzzle with putting all pieces correctly together. For this purpose, we seek a multi-layer analysis to have a broad scene of all approaches towards

a personalized fashion recommendation system. To better understand where we are, we continue our research by taking one step back to find the mainstream of our exact topic by paying attention to its parent node, which is "recommendation systems" in general. After reviewing different related papers, we found out that research around Fashion recommender systems seeks different approaches. We try to map out the areas of interest in research that exist on FRS, as shown in Figure 3. It has to be said that within this thesis, the focus is on image-based fashion recommendation systems from a computer vision and deep learning perspective. In contrast, the rest categories have been touched.

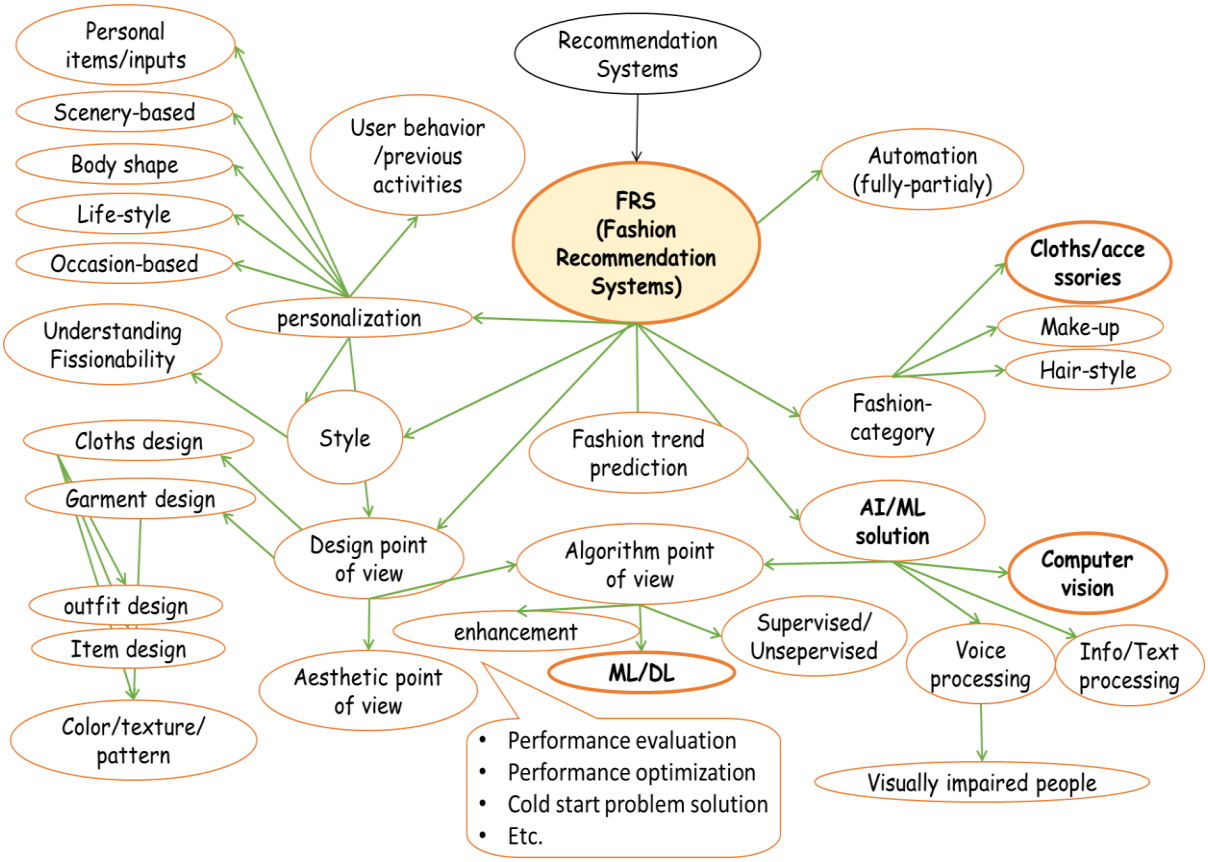

Figure 3: The different approaches towards FRS (fashion recommendation systems) studies

3. Literature search:

We were inspired by (Webster et al., 2002) structured approach to identify the sources of materials we need, including searching the primary academic databases to find the significant contributions which resulted in collecting desired papers. The databases and journals, including IEEE explorer, ProQuest, ScienceDirect,

Elsevier, Springer, Google Scholar, and selected conference papers, were used for keyword searches. There are also other databases like the Diva portal that have been used for collecting material.

We also follow (Malone et al., 1994) approach via going backward-forward fashion. In this technique, we first review the citations of those articles we find in step 1 to recognize the previous studies on the same field. Then we tried to find the essential papers, and which of them should be kept and which of them should be excluded; we also have reviewed the abstraction and introduction, and in most cases, we used the scan and skim technique and finally studied articles by details if it was required for deciding to keep or releasing reports because in some cases it was not the same material we were looking for. Still, it contains some fundamental knowledge it was necessary to complete or reconstructing our mindset about our approach towards the topic with an ongoing evaluation of materials. A systematic search should ensure that you accumulate a relatively complete census of relevant literature. Of course, you will miss some articles. However, if these are critical to the review, they are likely to be identified by colleagues who read your paper either before or after your submission (Webster et al., 2002).

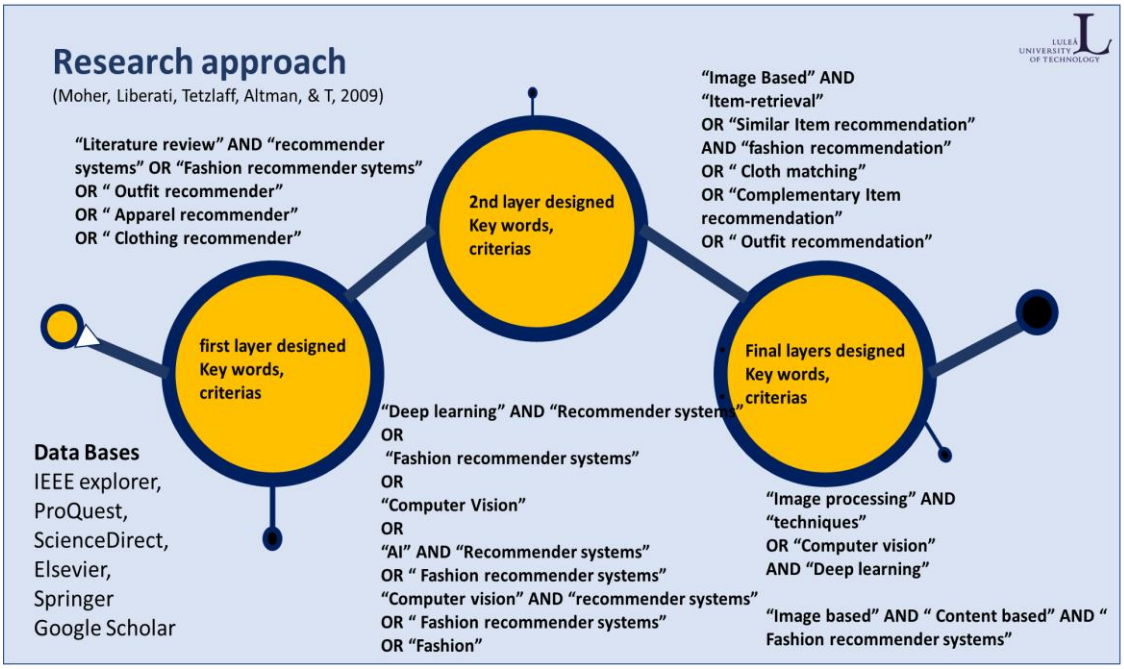

Figure 4: The three layers of researching the most significant keywords

Illustrated in Figure 4., in line with what has been mentioned above, three layers of research have been designed with highlighting the most significant keywords for search in data-based considering exclusive and inclusive criteria in terms of language, novelty, importance, and connection, and relevancy of the

material. Obviously, in addition to what has been designed separately, there was a need for a comprehensive deductive overview of material and contents to provide integration and links in needed aspects.

4. Literature analysis and synthesis

As mentioned before, the outcomes of the previous step as collected related literature on the topic have to be analyzed and synthesized through this step. Inspired by Concept- centric approach suggested by (Salipante et al., 1982) and (Webster et al., 2002), we have reconstructed the previous knowledge conceptually through this step which mainly has been done in paralleled with step 2. The results have been updated through step 3. (Brocke et al., 2009) also mentioned: "subdivides topic-related concepts allows for arranging, discussing, and synthesizing prior research." According to the importance of each context in our study, we deep-dive into it. We mainly contribute to developing the concepts via integrating and making conclusions through logical deduction. Besides, through this step, central pieces of interest from reviewed papers are summarized and highlighted.

5. Research agenda

The outcomes of the previous steps generate "sharper and more insightful questions for future research (Webster et al., 2002), which "is expected to result into a research agenda" (Brocke et al., 2009) in the final stage of this process. Furthermore, more than 200 abstracts papers were read to narrow the findings down more. More than 100 papers have been reviewed, summarized, categorized, and integrated for this thesis research agenda through a last additional backward and forward search.

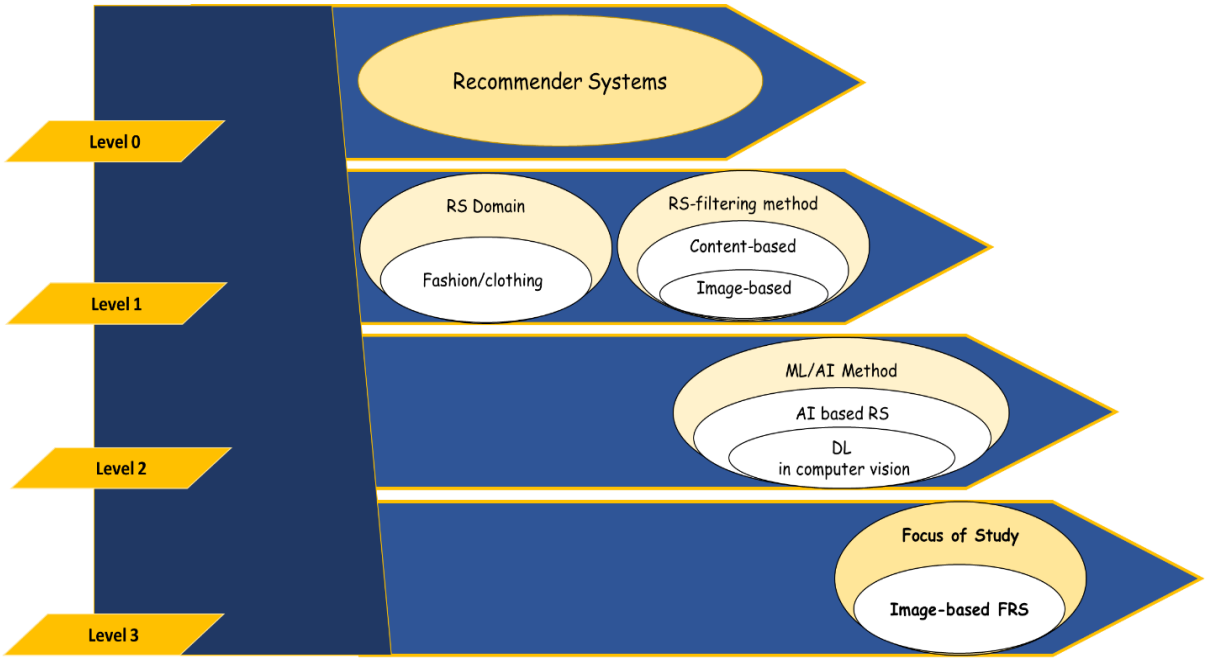

Figure 5: The scope and focus of the thesis

Overall, the main focus of interest in this thesis has been depicted as the Figure above. Performing this literature review required digging up the subject into four layers, each layer considering different aspects to finally integrate all of them in a way that makes sense to answer our research questions. In the very beginning stage of research, level-0, we try to introduce the general idea of a recommender system, then different categories and purposes for which those systems have been defined, however; while our purpose of conducting this research was not reviewing recommender systems in general, just suffice to provide main definitions and classifications. In level-0, we made an effort to locate our research domain, fashion, and notably clothing, which usually includes accessories and the recommender systems filtering methods focused on Image content-based. In level-2, we introduced two main areas of Artificial Intelligence, including Computer vision and Deep learning, looking at how the relative methods have been used and evolved in the fashion domain and particularly in image-based FRS as specified in Level-3.

3. Fashion Recommender Systems

3.1. Fashion Domain

Fashion is how we present ourselves to the world. The way we dress and makeup defines our unique style and distinguishes us from others (Cheng et al., 2020). Fashion in modern society has become an indispensable part of who I am. Dressing in a socially acceptable combination of clothes is considered vital in contemporary society, especially where professionalism is synonymous with clothing (A. Pandit et al., 2020). In everyday life, people need to find appropriate clothes to wear. Wearing clothes that match color, texture, style, skin tone, etc., is an essential aspect of fashion and personality. Outfit selection is a common problem that people face every day. This problem is extensive and involves many visual and social factors that can be implicit and abstract. As its definition in Cambridge dictionary, in its contemporary significance, is a style popular at a particular time, especially in clothes, hair, make-up, etc.

Fashionable products are highly demanded, and consequently, fashion is perceived as a desirable and profitable industry. The fashion industry occupies a significant position in the global economy and involves a sizeable industrial value chain, including garment design, production, and sales (A. Pandit et al., 2020). In fact, in recent years, there has been an expanding demand for clothing worldwide. The fashion segment revenue is projected to reach 878,334 m U.S. 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 percent of the world's Gross Domestic Product (GDP) [1], indicating that the demand for clothing will rise across the entire world.

What has been mentioned above were some main reasons that make the fashion industry highly attractive for investors. Other reasons and challenges in this industry make use of fashion recommender systems essential as an integral part of all today's businesses. 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. In many marketing industries, complementary item recommendation is used as a cross-selling strategy. As (M. Elahi 2021) indicated, "A major challenge in the fashion domain is the increasing Variety, Volume, and Velocity of fashion production which makes it difficult for the consumers to choose which product to purchase." (M. Elahi, 2021) believes that while having more choices available provide a better chance for consumers to choose appealing products, this phenomenon may result in the problem of choice overload, i.e., the issue of having an unlimited number of choices, especially when they do not differ significantly from each other, besides the fact that (A. Pandit et al., 2020) In today's fast-

[1] https://www.statista.com/outlook/244/100/fashion/worldwide

moving world people don't have enough time to dedicate to fashion and personality; as a result, they wear the same monotonous dress in their routine.

Furthermore, in (C. Guan et al., 2017; Liew et al., 2011), reported survey results indicated that 94% of the respondents admitted that their clothing purchasing decisions rely upon advice from others, such as friends and family. As we can see, "personal style advisor is much needed for ordinary people with less fashion knowledge and/or individual tastes in dressing." With tens of thousands of cloth styles in current online stores, it is challenging for a stylist to find appropriate clothes to match individual needs and occasional needs. In addition, the recommendation result highly depends on a stylist's knowledge and practical experiences (Cheng et al., 2020) "while not everyone is a natural-born fashion stylist. In support of this need, fashion recommendation has attracted increasing attention, given its ready applications to online shopping for fashion products".

Recommender systems can mitigate the above-mentioned problems by "suggesting a personalized selection of products (i.e., fashion items) that are predicted to be the most attractive for a target user (i.e., fashion consumer); this is fulfilled through filtering irrelevant things and recommending a shortlist of the most relevant items for the users. Analyzing and learning user preferences makes recommendations effective (M. Elahi, 2021).

3.2. Recommender Systems (fashion domain)

This literature review's technological solution field belongs to the Recommender Systems domain, mainly because the ambition of this thesis is to survey image-based personalized fashion recommender systems from a computer vision perspective for the fashion; thus, Recommender Systems are the closest available domain.

Employing general recommendation technology has been widely integrated into e-commerce websites (Cheng et al., 2020). The idea of Recommendation technology was initially introduced in the mid90s (D. Goldberg et al., 1992; W. Hill et al., 1995); the early work developed many heuristics for content-based and Collaborative Filtering (CF). (G. Adomavicius et al., 2005) the primary function of a general recommendation system is to predict products that potential consumers might want to buy based on their stated preferences, online shopping choices, and purchases of people with similar tastes or demographics. Popularized by the Netflix challenge, Matrix Factorization (MF) later became the mainstream recommender model for a long time from 2008 until 2016 (A. Karatzoglou et al., 2010; S. Rendle et al., 2010). However, the linear nature of factorization models makes them less effective when dealing with large and complex data, e.g., the complex user-item interactions, and the items may contain complex semantics (e.g., texts and images) that require a thorough understanding of them. Around the same time, in the mid-2010s, the rise

of deep neural networks in machine learning (aka., Deep Learning) has revolutionized several areas, including speech recognition, computer vision, and natural language processing (I. Goodfellow et al., 2016).

Recommender systems can be classified according to different principles depending on the task they are focused on –i.e., predicting item ratings and ranking item sets–, the approach to extract user preferences –i.e., implicit or explicit–, and the recommendation dynamics they follow -e.g., single-shot or unique answer and conversational or iterative approaches (L. Q. Sánchez et al., 2020). (Cheng et al., 2020) based on the literature, two main categories of recommender systems are usually considered (L. Q. Sánchez et al., 2020) based on the way recommendations are generated, (1) content-based (CB) systems, which recommend items similar to those liked in the past, (2) collaborative filtering (CF) systems, which suggest to user's items preferred by 'similar' people. (Cheng et al., 2020) Content-based systems examine properties of the recommended items by conducting a classification of users and products profile data according to the product features. Collaborative filtering systems recommend items based on similarity measures between users and/or items through clustering products bought from similar users. These systems recommend products on the basis of the prediction of users' preferences by analyzing an extensive scalable database from users' activities recorded through purchase or browsing history, click rate, products questionnaire, and user profiles. (L. Q. Sánchez et al., 2020) in general, the former makes use of item similarities based on textual representations, whereas the latter exploits are rating patterns. Also, recommender systems can be placed by the algorithmic approach they use for each of the above categories. In this sense, there are again two main types: (1) heuristic-based, which estimates the relevance of items through mathematical formulas, and (2) model-based, which predict the relevance of items through machine learning techniques. (M. Elahi, 2021), while has studied cold stat problem in Fashion Recommender Systems, highlighted four main categories of Techniques for Fashion Recommendation based on main classifications of techniques in recommender systems: Collaborative filtering base-models, Content-based models, Machine learning class models, and a Hybrid class of techniques.

3.3. The complexity of the fashion domain

As a broad definition, a Recommender System is a subclass of information filtering system that seeks to predict the "rating" or "preference" a user would give to an item (Ricci et al., 2011; G. Prato, 2019). This definition refers to generic Recommender Systems across all domains, "in some specific areas of application for that category of algorithms (such as the Fashion domain) the definition itself fails to incorporate the nuances and the peculiarities of solutions in the domain," therefore, it is perceived that General Purpose Recommender Systems (i.e., not specifically designed for a single application domain) are not able to address the specific needs of a domain with a particular characteristic like Fashion.

(Bollacker et al., 2016) stated fashion is an inherently subjective, cultural notion, and (G. Prato, 2019) indicated “a fashion outfit is an ensemble of clothing items that maintains a coherent style, that can be considered pleasing in its overall composition, and that is in line with the general taste of fashion of its present time frame,” obviously, we are dealing with a domain with some specific complexity in terms of concept understanding, tasks definition and consequently design and development of solutions, (G. Prato, 2019) “fashionable outfits composition is challenging both in terms of problem definition and in terms of the definition of the evaluation metrics because most of the characteristics that make an outfit fashionable are rather subjective and consequently difficult to be measured effectively.” (Bollacker et al., 2016) stated that extracting knowledge and actionable insights from fashion data still presents challenges due to the intrinsic subjectivity needed to model the domain effectively.

In most of the domains in which Recommender Systems are developed (e.g., movies, e-commerce, etc.), the similarity between items has been used as a proxy to evaluate which item to recommend (Cheng et al., 2021; G. Prato, 2019). Instead, in the Fashion domain, compatibility is a critical factor to be considered. In addition, raw visual features belonging to products representations that contribute for the most part of the algorithm's performances in the Fashion domain are distinguishable from the metadata of the products in other domains (M. Vasileva et al., 2018).

Considering what has been mentioned above, challenges in the fashion domain mostly stems from complexity in a context which makes it necessary that those key terms and notions which are frequently used in fashion recommendation literature be clarified conceptually in order to provide a unified framework not only to relax digesting definitions which are linked and interconnected but more importantly soften vague understanding, particularly among those researchers who are newly attracted to this field. These concepts are principals which particular tasks in FRS have been evolved mainly based on them and directly influence technical and methodological considerations in FRS architecture design; as (Bollacker et al., 2016) cited, “in order to understand fashion in any rigorous way, this subjectivity must be an intrinsic part of the model.”

Some of these concepts have been used as umbrella terms and implicitly convey the other ones’ definitions, and some of these terms have been used interchangeably. Here as part of our contribution in this literature review, we are going to show how correlated and interconnected these notions are in FRS by mapping Main areas of complexity in FRS as a conceptual framework illustrated in Figure 6.

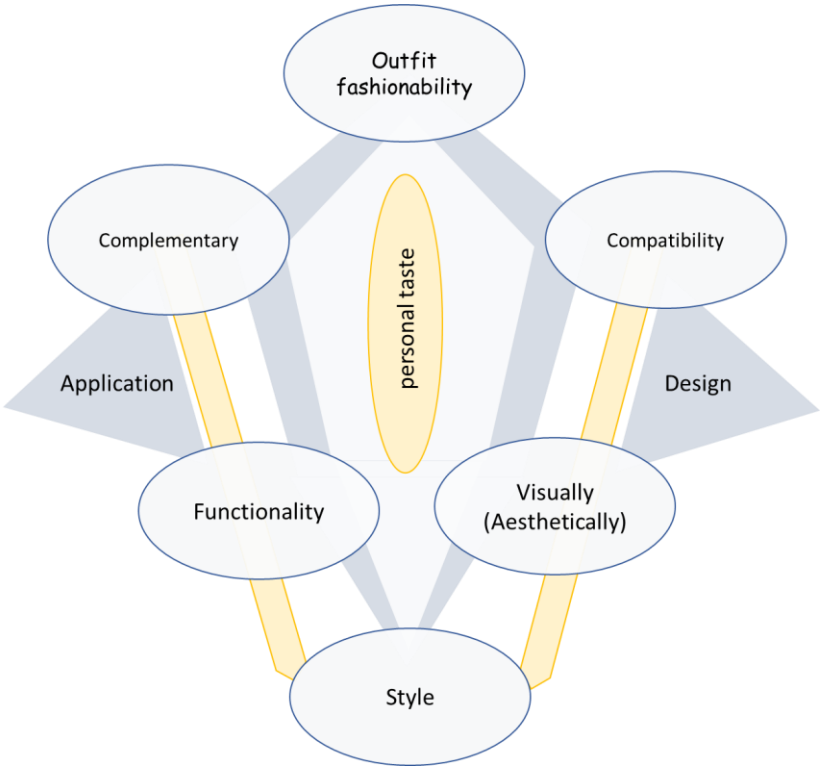

Figure 6: The main disciplines which any fashion recommender system can be understood through them

3.3.1 The balance between Application and Design

Here, we define the term Application to define the functionality, which cloth's items are usually created and used for, and Design reflects aesthetic aspects as a narrative of form. Application of the clothing item and its design are two main aspects in fashion, which people usually choose their cloths. Fashionable items are perceived through these two main disciplines. (R. He et al., 2016) explains that identifying relationships between clothing items is a key task of an online fashion recommender system to help users discover functionally complementary or visually compatible items. In FR's literature, Style is defined based on how these two aspects of clothing fit well together, also in (Cheng et al., 2020) stated, "in different research papers in FR's domain compatibility refers to coherency in both aspects of visually(appearance) and functionally." As fashion is considered a historical, geographical, and cultural notion, application and design vary over time and region, so the style does. Besides, personal preferences have a significant effect on forming styles. While (Y. Wen et al., 2018) stated fashion can be understood through its domain ontology, however; Clothing ontologies primarily model the structure of physical feature values (e.g., sleeve-length, colors, fabric), it is worth noting that the Design and Application follow user, cloth and contextual characteristics based on clothing ontology. Application is defined based on contextual

characteristics, for instance, temperature, location, occasion, etc., while most of the time, the design seeks both user characteristics (e.g., skin color, height, size, etc.), and item characteristics (e.g., texture, color, shape, etc.). Nonetheless, design follows the application, as (Sullivan 2010) stated, "Form follows function," which is perceived as a principle of design. This is what has to be understood via FRS to make recommendations with different degrees of personalization.

3.3.2 Personalization and Understanding notion of Style

Personalization is the heart of all fashion recommender systems; this represents the importance of personal preferences considerations in designing a system. Any fashion recommender system cannot be perceived without taking personal preferences into account for making the most satisfying recommendations. Nonetheless, the recommendations may be developed by offering different degrees of personalized recommendations. (C. Guan et al., 2017) A well-described user profile could distinguish a more customized recommendation system from available systems. In a very initial state, recommender systems can satisfy customers by suggesting similar or identical items they are looking for among clothes in a particular inventory. In the next steps, ranking systems have been added to these primitive RS's, based on understanding user preferences according to previous behavior or scoring systems. (Q. Zhang et al., 2020) Recommender systems provide personalized service support to users by learning their previous behaviors and predicting their current preferences for particular products. (W. Yu et al., 2020) Recommender systems have been widely used in online services to predict users' preferences based on their interaction histories. In more advanced levels, recommender systems employed more sophisticated architectures to suggest the similarity-based items and consider more complex aspects necessary for their customers to choose clothing items, including style coherency and outfit compatibility. (Cheng et al., 2020) highlighted that the main difference between fashion item-retrieval and fashion recommendation is that the former learns the visual similarity between the same clothing type.

In contrast, the latter learns both visual similarity and visual compatibility between clothing types distinguishing similarity task from compatibility estimation. (Hsiao et al., 2018) Whereas visual similarity asks "what looks like this?" and is relatively well understood, compatibility instead asks, "what complements this?" It requires capturing how multiple visual items interact, often according to subtle visual properties. According to (Cheng et al., 2020), "in apparel recommendations, there is a distinctive function which is not only recommending similar products to meet users' current dressing-style but providing personalized styling advice to develop a better understanding of personalized styling" highlights the importance of considering personal preferences in style creations.

Figure 7: A comparison among Instance match system, label-based system, and proposer Latent look system on how style is perceived and represented, source: (Hsiao et al., 2017), "Learning the Latent "Look": Unsupervised Discovery of a Style-Coherent Embedding from Fashion Images."

3.3.3 Style and compatibility

How has style been perceived through the literature? (Nakamura et al., 2018) indicated style is a criterion in selecting each fashion item for creating an outfit. It means that style can be regarded as a feature of the overall outfit. In another definition, (Hsiao et al., 2017) signify "What defines a visual style? Fashion styles emerge organically from how people assemble outfits of clothing, making them difficult to pin down with a computational model". (M. Vasileva et al. 2018) Putting the notion of compatibility in the broader definition of similarity, "Outfits in online fashion data are composed of items of many different types (e.g., top, bottom, shoes) that share some stylistic relationship. A representation for building outfits requires a method that can learn both notions of similarity (for example, when two tops are interchangeable) and compatibility (items of possibly different type that can go together in an outfit)". Outfit Compatibility refers to have a coherent style and coordinated clothing; these terms are mainly used interchangeably in literature. (Hsiao et al., 2017) believed "Style coherency differs from traditional notions of visual similarity". They explained " style coherency located between the two different perspectives in the literature: on one side of the spectrum are methods based on robust instance matching, e.g., (S. Liu et al., 2012; Kalantidis et al., 2013; M. Kiapour et al., 2015; S. Vittayakorn et al., 2015; Z. Liu et al., 2016) (see Figure 7(a)); on the other side, are methods based on coarse style classification, e.g., to label an outfit as one of a small number predefined categories like Hipster, Preppy, or Goth (M. Kiapour et al., 2014; E. Simo-Serra et al., 2016).

According to Hsiao et al., proposed model, coherent styles reflect some latent "look," and style coherency refers to consistent fine-grained trends represented by different combinations of garments. (Hsiao et al., 2018) indicates compatibility requires judging how well-coordinated or complementary a given set of garments is. (R. He et al., 2016) explains that identifying relationships between items is a key task of an online fashion recommender system to help users discover functionally complementary or visually compatible items. In domains like clothing recommendation, this task is particularly challenging since a successful system should be capable of handling a large corpus of items, a huge number of relationships among them, and the high-dimensional and semantically complicated features involved.

Furthermore, the human notion of "compatibility" goes beyond mere similarity: For two items to be compatible—whether jeans and a t-shirt or a laptop and a charger—they should be similar in some ways but systematically different in others. (Cheng et al., 2020) highlighted the fashion compatibility as a key concept which is a basis of any FRS to generate fashionable outfits, "Fashion recommendation works based on fashion compatibility, which performs how well items of different types can collaborate to form stylish outfits.

To successfully create fashionable outfits, first and foremost, the system requires an inherent understanding of product features related to color, shape, style, fit, etc. (K. Laenen et al., 2020). These product features are expressed in different modalities such as images, text, video, or audio. Identifying and understanding complicated and heterogeneous relationships between items in the product is a vital component of any modern recommender system (R. He et al., 2016). To model subtle notions like 'compatibility' upon the raw visual features, we need expressive transformations capable of relating feature dimensions to explain the relationships between pairs of items (R. He et al., 2016). Most clothing recommender systems employed the approach of collaborative filtering and content-based methods, which pay more attention to predicting the user's item preferences based on massive historical data, overlooked user's context, such as weather, occasion, requirements, emotions, and correlations between clothing items (Y. Wen et al., 2018). These systems ignored the user's context, such as weather, event, needs, feelings, and correlations between clothing items. (A. Pandit et al., 2020; Y. Wen et al., 2018) The user entity description comprises user features such as height, weight, skin, etc. The clothes are defined using cloth characteristics like color, texture, pattern, fabric, etc. The context entity consists of information about the current weather, occasion, etc. (C. Guan et al., 2017); apparel features are described from a formulation of colors, lines, and shapes, pattern/prints, and textures which were studied through the process of feature recognition, extraction, and encoding. On the other side, user features are recognized as facial features, body features, personal preference (taste), and wearing occasions. (Y. Wen et al., 2018) introduced three kinds of fundamental ontology, Figure 8., which describes the related entities of the user, clothing, and context for clothing

recommendation. A well-described user profile could distinguish a more personalized recommendation system from available systems (C. Guan et al., 2017).

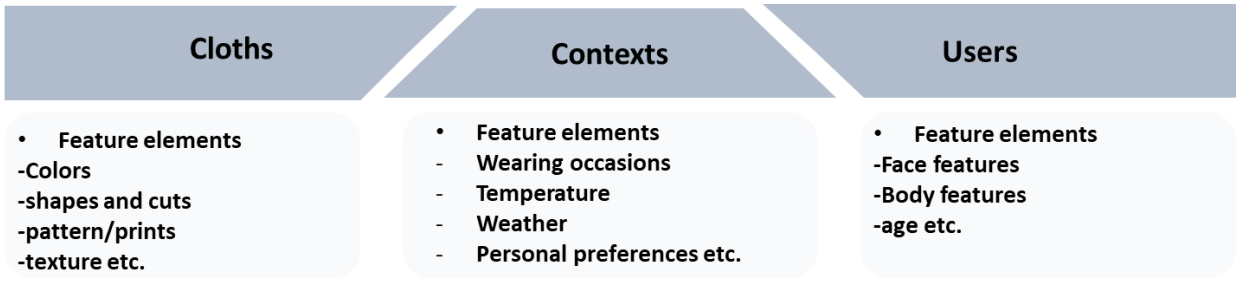

Figure 8: Fashion feature elements based on cloth ontology (Y. Wen et al., 2018)

It is worth mentioning that, although in different research papers in FR's domain, compatibility has been used interchangeably with coherency, coordination, and complementary terms, conceptually conveying the meaning, compatibility notion has been perceived differently from complementary concept based on the task's definition in FRS. In this meaning, compatibility has been defined in a broader definition among items of an outfit when the purpose of an FRS is outfit recommendation as a whole. In contrast, a complementary concept usually refers to the Fill-In-The-Blank task. According to (G. Prato 2019), two main functions in the FRS domain can be identified, including Compatibility estimation and outfit completion. The former asks the algorithm to distinguish between those clothes that fit together and those that do not. The latter is generally referred to as a Fill-In-The-Blank task. FITB consists in finding the missing item that best completes an outfit from a subset of possible choices". As (K. Laenen et al., 2020) explained in fashion e-commerce, a complementary item is an item that can be worn with the current item as a Cross-selling strategy to the customer to encourage the purchase of something else in conjunction with the current product. Outfit recommendation is an extension of complementary item recommendation where the system recommends not one but multiple items that create a fashionable outfit with the current item. Compatibility Estimation (CE) and Fill-In-The-Blank task (FITB) tasks are explained separately in the FRS task metric evaluation section. (Wong et al., 2009b) emphasized that providing mix-and-match recommendations is a 'must' strategy for fashion retailers to enhance customer service and improve sales. (Hsiao et al., 2017) whereas style refers to a characterization of whatever it is people wear, compatibility refers to how well-coordinated individual garments are (A. Veit et al., 2015; S. Liu et al., 2012; T. Iwata et al., 2011), and fashionability refers to the popularity of clothing items, e.g., as judged by the number of "like" votes on an image posted online (E. Simo-Serra et al., 2015). Recent work explores forecasting the popularity of styles (Z. Al-Halah et al., 2017).

For providing insight into the meaning of visual style, The Hipster Wars represented five style categories (Hipster, Goth, Preppy, Pinup, Bohemian). They recognized them based on key-point locations on bodily

parts (M. Kiapour et al., 2014). Another approach pre-trains a neural network for style using weak meta-data labels (E. Simo-Serra et al., 2016). These two retrievals work (M. Kiapour et al., 2014; E. Simo-Serra et al., 2016) and (Hsiao et al., 2017) aimed to capture a broader notion of style. Unlike those two (Hsiao et al., 2017) treat styles as discoverable latent factors rather than manually defined categories. (Y. Ma et al., 2017) looking at the most popular clothing fashion styles of this season Reported by Vogue1, including romantic, elegant, and classic, indicates" these styles depend strongly on some specific visual details, such as lapel collar nipped waist, matched with high-waistlines dress or pencil trousers. Since clothing fashion styles benefit a lot from graphic details", Many efforts have been made to address can bridge the gap between them automatically, namely, (Fan et al., 2014) proposed an approach to parse refined texture attribute of clothing. (Liang et al., 2016) also created an integrated system to parse a set of clothing images. In addition, some researchers try to analyze visual features by considering occasion and scenario considerations. (S. Liu et al. 2012) proposed a scenario-oriented clothing recommendation system. (Jia et al. 2016) suggested to taking the aesthetic effects of upper-body menswear into account (Y. Ma et al., 2017) believed that there is still a lack of unity, and it is overlooked that the pairing of top and bottom has a considerable influence on fashion designs. They proposed their approach to addressing these limitations from two perspectives, including suggesting a Fashion Semantic Space (FSS) based on the Image-Scale Space noticing the aesthetic domain (Yasuda et al., 1995), and developing a Bimodal Correlative Deep Autoencoder (BCDA), a fashion-oriented multimodal deep learning-based model, to extract the correlation between visual features and fashion styles by employing the inherent matching rules of tops and bottoms.

3.3.4 Aesthetic perspective

The aesthetic perspective is another concept that can play a key role in proposing FRS's is perceived as visual perception. (Jia et al., 2016) However, it is also quite significant for people to wear aesthetically. (Liu et al. 2012) takes two criteria, wearing properly and aesthetically, into consideration. (Kouge et al. 2015) obtains the associated rules from color combinations to derive impressions. The style also can be perceived aesthetically (Bollacker et al., 2016); each style can then be described as a coherent aesthetic entity in the mind of an observer. While traditionally, this style may be quantitatively described by its physical features, consider the alternative aspect of its subjective qualities shared with other cultural entities. In a few papers, Aesthetic concepts directly have been highlighted (W. Yu et al., 2018) indicated that Visual information plays a critical role in the human decision-making process. Recent developments on visually aware recommender systems have taken the product image into account. They argue that the aesthetic factor is essential in modeling and predicting users' preferences, especially fashion-related domains like clothing and jewelry. According to (Jia et al. 2016), It has been documented that people reinforce their mood and express their feelings through their clothing (Kang et al., 2013).

A variety of researches have made it possible to extract or recognize the visual features (e.g., sleeve length, color distribution, and clothing pattern) from clothing images accurately (Yang et al., 2011; Yamaguchi et al., 2013; W. Yang et al., 2014). (Jia et al. 2016) tried to answer how people describe clothing, they explained, "The aesthetic words like "formal" or "casual" are usually used rather than comments like the sleeves are long, or the collar is round. These aesthetic words are related to visual features. For example, suits with more than three buttons look formal, while tank tops seem casual". (J. Li et al., 2011) stated choosing the style of a garment is affected not only by the physical attributes of the components of the garment but also the context. The attraction, similarity, and compromise effect in multi-alternative decision-making have been widely reported and studied in psychology. A good and precise prediction of a customer's choice should account for not only the rational part but also the irrational part.

3.3.5 Design Perspective

While the AI approaches in the fashion recommendation domain mostly is concentrated on apparel recommendation (Zeiler et al., 2013), identifying apparel attributes (M. Abadi et al., 2015; Z. Liu et al., 2016), and segmenting major apparel components (Z. Liu et al., 2016). Some studies have followed a different perspective, mostly related to the most creative side of the fashion domain, design.

Apparel Design using AI has seen recent success with generative adversarial network (GAN) models and style transfer (Zeiler et al., 2013; O. Russakovsky et al., 2015) and availability of rich datasets like DeepFashion (Z. Liu et al., 2016). More recently (Kang et al., 2017) proposed a model to address the recommendation problems within the fashion domain and design fashion garments considering a person's taste. In "Visually-Aware Fashion Recommendation and Design with Generative Image Models" exploits Siamese convolutions neural networks (CNN) to create a fashion-aware visual representation of the items. Moreover, the system can create new fashion items by understanding the user's past interactions to understand the taste and constraints. Kato et al. (N. Kato et al., 2017) propose the use of Progressive Growing of GANs (P-GANs) in the task of fashion design to serve as a base for dress-making or pattern-making. However, while this approach shows some interesting and promising results, it was found through a user study that the technology does not yield the same results as an experienced pattern maker in the same tasks, highlighting the importance of the craft in the design of fashion garments. There has also been a growing interest in generating apparel designs using GANs, given their ability to generate appealing images. (Kang et al., 2017) uses Conditional GAN to generate novel fashion items that maximize user preferences. (Zhu et al., 2017) presented a method for creating new apparel for a wearer using textual descriptions. Employing generative adversarial training (Yu et al., 2017) introduced a personalized fashion design framework that can automatically model users' preferences and design a compatible cloth item based

on a given query item. StyleGAN is a shape-conditioned model suggested by (Sbai et al., 2018) for producing fashion design images (Banerjee et al., 2018).

Various GAN architectures for context-based fashion generation were investigated by (Rajdeep et al., 2018). (Dong et al., 2019) introduced the Fashion Editing Generative Adversarial Network (FEGAN), which allows users to edit a fashion image with only a sketch and a few sparse color strokes. GarmentGAN was proposed by (Raffiee et al., 2020) to synthesize high-quality photographs and reliably transfer photographic features of clothing. A shape transfer network and an appearance transfer network are the two GANs that make up the system. Unlike these approaches, (Bhardwaj et al., 2020) generated new designs by combining multiple apparels and then use style-transfer to add further variation.

3.4. Fashion recommendation system tasks

Emphasizing the fact that how task-oriented is proposing a recommendation system, in general, there is a growing body of knowledge that indicates it is not possible to be understood how a fashion recommendation system has been developed or should be developed unless foster a deeper primitive understanding of why it has been built. Therefore, obtaining a better understanding of various applications and needs of which any FRS has been made as an answer or solution will be illuminative for developing a more efficient fashion recommendation system, particularly in choosing technology. Consequently, scholars have mainly developed their ideas on the basis of different task categories. Even when the main purpose of some studies was mentioned as introducing or developing a novel architecture or providing some improvement in architecture, it still has been defined as a solution that has been evolved around a problem definition as a task in the FRS domain.

Reviewing the literature on fashion recommender systems indicates a few main tasks that FRSs usually have evolved around them from item-retrieval systems to recommending capsule wardrobes. Nonetheless, some scholars believe that the retrieval systems cannot be considered as a fashion recommender system; the main reason behind this belief goes back to the functionality differences between them that have been mentioned by (Cheng et al., 2020), who explained that the item-retrieval learns the visual similarity between the same clothing type, whereas the FRS learns both visual similarity and visual compatibility between different types of clothing, distinguishing similarity task from compatibility estimation. Clothing-retrieval-based researches are included in this literature review since this field has attracted the most attention in clothing recommendation systems. The main tasks which are frequently assigned to fashion recommender systems have identified through this literature review with a primary focus on image-based systems have represented as follows:

- Similar or identical item recommendation (item retrieval)
- Complementary Item recommendation
- Outfit recommendation
- Capsule Wardrobe

3.4.1 Similar or identical item recommendation (item retrieval)

Figure 9: Identical Item Recommendation
Source: (M. Kiapour et al., 2015) M. Hadi Kiapour, Xufeng Han, Svetlana Lazebnik, Alexander C. Berg, Tamara L. Berg, Where to Buy It: Matching Street Clothing Photos in Online Shops, 2015 IEEE International Conference on Computer Vision

There are lots of image retrieval methods that have been introduced by researchers and scholars, (Reddy et al., 2016) highlighted some of the most important and widely used of these techniques, including Text-Based Image Retrieval, Content-Based Image Retrieval, Multimodal Fusion Image Retrieval, Semantic-Based Image Retrieval, Relevance Feedback Image Retrieval. (Jaradat et al., 2021) claimed that because of recent advances in deep learning methodologies, particularly in image processing, interpretation, and segmentation, and because this effect can be amplified in fashion, content-based recommender systems have sparked a lot of interest in the field of fashion recommendation. With respect to the wide attraction and application of Content-based image retrieval (CBIR) in FRS researches and considering the scope and focus of interest in this thesis in following (CBIR), Content-based image retrieval has been the focal point of this review in this section.

Content-based image retrieval (CBIR) has received a lot of attention in the previous decade, owing to the need to properly handle the fast-rising volume of multimedia data (Nandish et al., 2013). The content-Based Image Retrieval (CBIR) technique can retrieve relative images from a database with an input image of the content we are looking for (X. Li et al., 2021). This approach is commonly utilized in computer vision and artificial intelligence applications. (X. Li et al., 2021) From a technological viewpoint, the CBIR system is based on image representation and database search. Feature vector or image representation is expected to be discriminative so as to distinguish images. Furthermore, it's supposed to be resistant to certain

transformations. The similarity score between two photos should express the semantic connection based on visual representation. These two related factors are critical to retrieval performance, and CBIR algorithms can be classified based on how they affect these two factors. (Nandish et al. 2013) almost all CBIR systems have been designed based on color, texture, and shape features of images. (X. Li et al., 2021) Image representation is perceived as the key step For CBIR that extracts the critical features from a given image and then transforms them into a fix-sized vector (so-called feature vector). The extracted features can be divided into three main categories: conventional features, classification CNN features, and retrieval CNN features.

Fashion instance-level image retrieval (FIR) as a sub-category (CBIR) is considered a hot topic in computer vision due to the rapid development of clothes e-commerce and the increase in the number of clothing images on the Internet. (FIR) is necessary for meeting the increasing needs of online purchasing, fashion detection, and web-based recommendation. (X. Li et al., 2021) FIR is primarily concerned with cross-domain fashion image retrieval, which entails matching two photos taken casually by users and the other professionally by vendors. According to (Cheng et al., 2020), While there is a considerable domain variation between daily human photos obtained in a typical setting and clothing images taken in perfect conditions, significant research investigations have been focused on tackling the problem of cross-scenario clothing retrieval (i.e., edited photos used in online clothing shops).

Many academic communities are paying significant attention to the development of cross-scenario image-based fashion retrieval tasks to retrieve nearly equivalent or the same products from the inventory according to a fashion image query (Cheng et al., 2020). The notable early work on automatic image-based clothing retrieval was presented by (X. Wang et al., 2011). (S. Liu et al., 2012) suggested using an unsupervised transfer learning method based on part-based alignment and sparse reconstruction attributes. This Occasion-Based fashion recommendation proposed system in “magic closet” employed latent SVM; the model incorporates four potentials, including visual feature vs. attribute, visual feature vs. occasion, occasion vs. attribute, and attribute vs. attribute. The clothing retrieval problem was addressed from human parsing (Kalantidis et al., 2013). For clothing segmentation, an initial probability map of the human body was produced by pose estimation, and the segments were subsequently classified using locality-sensitive hashing. By adding up the overlap similarities, the visually comparable items were found then. Then employed image retrieval techniques used indexes of sub-linear complexity to retrieve similar items from each of the detected classes. It has to be mentioned that (Kalantidis et al., 2013; S. Liu et al., 2012; X. Wang et al., 2011) are based on hand-crafted features. With the advances of deep learning, there has been a trend of building deep neural network architectures to solve the clothing retrieval task (Cheng et al., 2020). Many

FIR approaches based on deep learning were developed and worked successfully as large-scale fashion datasets were provided (Nandish et al., 2013).

Employing attribute-guided learning, (J. Huang et al., 2015) created a Dual Attribute aware Ranking Network (DARN) to represent in-depth features. DARN modeled the disparity between domains while embedding semantic attributes and visual similarity constraints into the feature learning step. The street-to-shop retrieval problem was originally attempted by (M. Kiapour et al., 2015) as the first attempt at same item retrieval, who produced three distinct algorithms for retrieving the same fashion item in a real-world image from an online shop. The three methods contained two deep learning baseline methods, and one method learned the similarity between two different street and shop domains.

(S. Jiang et al., 2016) proposed a deep bi-directional cross-triplet embedding algorithm to model the similarity between cross-domain photos, improved the one-way problem, street-to-shop retrieval task. They also expanded this method to retrieve a series of complementary accessories to pair with the cloth item shop.

(Kuang et al., 2019) introduced a Graph Reasoning Network to build the similarity pyramid to enhance the existing retrieval tasks methods, which only considered global feature vectors and represented the similarity between a query and a clothing inventory by considering global and local representation. There are also other scholars who consider text descriptions in addition to visual features in clothing retrieval tasks, in some (B. Zhao et al., 2017; Kovashka et al., 2012) By connecting the visual representation of the query image with the textual properties in the query text, a visual representation of the searched item was created, but (Laenen et al., 2018) used a shared multimodal embedding space to deduce the semantic linkage between visual and textual features. Thanks to recent advancements in deep learning techniques, Content-Based Similar Fashion Items Recommendation approaches have attracted more attention, particularly via computer vision and natural language processing (Y. Guo et al., 2016) introduced an early work on the computer-generated design of fashion garments as a part in applications of Generative Adversarial Networks. (Kang et al., 2017) introduced a fashion-aware visual representation that used Siamese convolutions neural networks (SCNN) to create a fashion-aware visual representation of the items. Besides, the system is able to suggest new fashion items by learning the user's past interactions to capture preferences and a sort of constraints. This system followed BPR and Siamese networks to create a visually aware personalized recommender system. Their model produces an image with maximized personalization through integration with Generative Adversarial Networks (GANs) to generate images by employing the activation maximization technique. (Kato et al.,2019) employed Progressive Growing of GANs (P-GANs) to make patterns and proposed an approach of clothing image generation. They contribute to experiment with how the process of creating patterns may be influenced by the quality of images generated by GANs.

Based on their findings, they indicate that “In contrast to design, making patterns can hardly be automated. so, we propose that pattern makers who have the prior knowledge of the brand design and brand pattern play the most significant role in drawing pattern from highly abstracted generated images.”

Some FIR methods adopt various attention mechanisms by using the advances of metric learning. The Visual Attention Model (VAM) (Z. Wang et al., 2017) created an end-to-end network structure by training a two-stream network with an attention branch and a global convolutional branch and then concatenating the produced vectors to improve a conventional triplet objective function. FashionNet (Z. Liu et al., 2016) also trained a network employing a triplet loss. Hard-aware Deeply Cascaded embedding (HDC) (Y. Yuan et al., 2017) used a cascaded mechanism to mine hard instances at several levels by combining a series of models with different complexities. The featured vectors from each sub-network were then scaled by fixed weights and fused to produce retrieval representations. (B. Gajic et al., 2018) focused on optimizing the training process as well as inference time. They emphasized the necessity of appropriate training of simple architecture, trained the network using the triplet loss, and tailored general models to the particular function. (Y. Zhao et al., 2018) developed an adversarial network for Hard Triplet Generation (HTG) to improve the network's capacity to discriminate similar examples of various categories while grouping specific examples of the same categories.

(M. Shin et al., 2019) intended for a competitive FIR performance. They offered a novel approach for converting a query into a representation with the preferred characteristics and a novel concept of feature-level attribute modification. Some deep learning algorithms integrate multiple approaches. The Grid Search Network (GSN) (A. Chopra et al., 2019) posed the training function as a search task, with the goal of finding matches for a given image query in a grid comprising both positive and negative images. To increase retrieval performance, (S. Park et al., 2019) investigated the training techniques and DNNs. It has been demonstrated that better training procedures, data augmentation, and structural modification can improve FIR results. Some FIR methods use attribute modules (J. Huang et al., 2015; Q. Dong et al., 2017; Y. Lu et al., 2017; M. Shin et al., 2019). (S. Park et al., 2019) investigated training strategies and DNNs to improve the retrieval performance. It is proved that better training strategies, better data augmentation, and better structural refinement could achieve better FIR results.

Conventional Recommender Systems are tasked with making predictions based on the similarity (between items in content-based techniques and between users in collaborative-filtering methods), whereas the outfit completion task involves the concept of compatibility between the items that make up an outfit (G. Prato, 2019; Cheng et al., 2021). (Cheng et al., 2020) each outfit generally involves multiple complementary items that stylistically or visually match when worn together (J. Craik, 2009), such as tops, bottoms, shoes, and accessories (J. Craik, 2009). Generating harmonious fashion matching is challenging, as it has been

explained with details in previous section 3, and here, we just recapitulate it up as three main reasons. First, the fashion concept is subtle and subjective (Cheng et al., 2020). Second, there are a large number of attributes for describing fashion (Cheng et al., 2020). Third, the notion of fashion item compatibility generally goes across categories and entails complex relationships (Cheng et al., 2020). Within the fashion recommendations research area, the recommendation of complimentary clothing products has been a central topic in recent years, attracting a lot of attention and resulting in a huge list of algorithms and methodologies. (Jaradat et al., 2021) Complementary item recommendations can be considered as a relaxed version of the Outfit recommendations problem.

3.4.2 Complementary Item Recommendation

(Jaradat et al., 2021; G. Prato, 2019) 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. (G. Prato, 2019) 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 (G. Prato, 2019). 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 (M. Vasileva et al., 2018; X. Han et al., 2017) (the only exception is in (Y. Li et al., 2017), that uses subsets of five elements, one of which is the correct one). (Jaradat et al., 2021) explained that the approaches proposed to address this problem are usually similar to the ones used in similarity-based item recommendation, so this problem can also be modeled 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, (G. Prato 2019) introduced an extended task as Unconstrained Outfit Completion, which is 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, reciprocal ranking. In this approach, a model is given an incomplete outfit and then is asked to predict the missing items, given their categories. According to (G. Prato 2019), 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), or 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; according to cloth ontology (Y. Wen et al., 2018), 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 in (Zhao et al., 2017). They deduce this relationship between fashion items according to the title description, assuming that the title contains the most relevant information of the item. They employed (SCNN) Siamese Convolutional Neural Network to find the compatible pairs of items in a words space, then mapped into an embedded style space. words are the only inputs here, making computation lighter; it also needs a few preprocessing stages without any feature engineering. (A. Veit et al., 2015) introduced item-to-item compatibility modeling as a metric-learning problem based on co-purchase behavior. they employed the co-purchase data from Amazon.com to train a Siamese CNN to learn style compatibility across categories, besides a robust nearest neighbor retrieval to generate compatible items. (J. McAuley et al., 2015) modeled human preference to discover the relationships between the appearances of pairs of items that mapped compatible items to embeddings close in the latent space. In contrast to most approaches based on content-extraction and try to understand similarity or complementary relationships between items like a human brain does, (J. McAuley et al., 2015) design 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 most previous research focused on top-bottom matching issues, (Y. Hu et al., 2015) worked on the personalized issue using a functional tensor factorization technique to model user-item and item-item relations. (R. He et al., 2016) later extended the work of (J. McAuley et al., 2015) by jointly using visual and historical user feedback data. Their proposed approach employed visual features into Bayesian Personalized Ranking with Factorization Matrix as the underlying predictor. (X. Zhang et al., 2017) suggested an occasion-oriented clothing recommender system by considering both principles of wearing properly and aesthetically. For learning the recommendation model that used clothing matching rules among visual features, attributes, and occasions, they employed a unified latent Support Vector Machine (SVM). An end-to-end framework (Y. Li et al., 2017) addressed the problem as a classification task. An outfit composition set was labeled as trendy or not. They proposed a multi-modal multi-instance model that used images and meta-data of fashion items and information across fashion items, observing items in terms of aesthetics and compatibility. using image captioning method (Donahue et al., 2015), (X. Han et al., 2017) suggested a model employing bidirectional LSTM (Bi-LSTM) to treat an outfit as a sequence of fashion items.

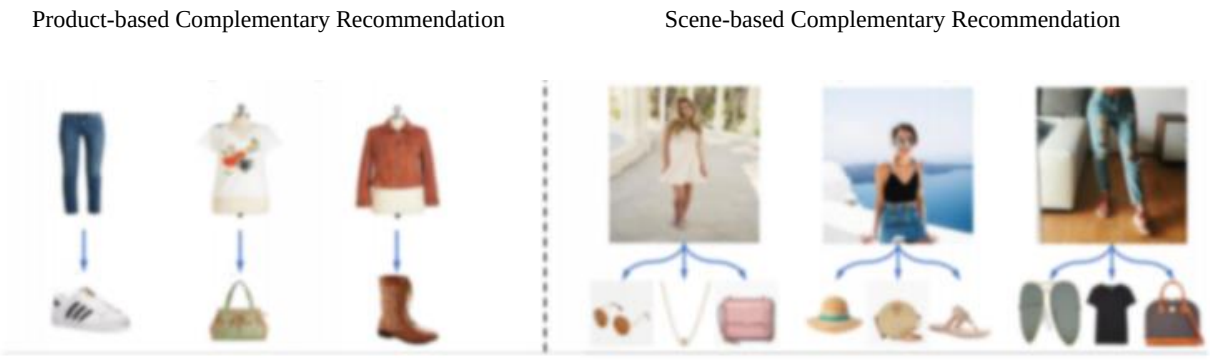

Figure 10: A comparison among product-based and scene-based complementary recommendation
Source: (Kang et al., 2019) W.-C. Kang, E. Kim, J. Leskovec, C. Rosenberg, and J. McAuley. 2019. Complete the Look: Scene-based Complementary Product Recommendation. In CVP

(Kang et al., 2019) developed "Complete the Look," which aims to suggest fashion objects that complement the scene. They used Siamese networks and category-guided attention techniques to measure both local (i.e., compatibility between each scene patch and product image) and global (i.e., compatibility between the scene and product images). Figure 10. shows a comparison of complementary product recommendations and those are based on a given scene.

(Chen et al., 2018) improved the traditional triplet neural network, which typically accepted only three items, to consider more items by proposing a mixed-category metric learning method that is able to have multiple inputs. They also model the intra-category and cross-category items of fashion collocation by feeding both well-collocated and bad-collocated clothing items to the deep neural network. (Hsiao et al., 2018) used natural language processing considered an outfit as a "document," and clothing attribute as a "word," and a clothing style was considered as a "topic." Each outfit matching was addressed by the topic model. The STAMP (Short-term attention/memory priority model for the session-based recommendation) proposed by (Q. Liu et al., 2018) addressed the limitations of previous approaches by considering the user's current actions to generate future recommendations in the same session, in real-time or almost 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 (Wu et al., 2020) introduced a session-based approach with some improvements in the STAMP model (Q. Liu et al., 2018) to produce complementary personalized item recommendations. (Yu et al., 2019) 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 both user interests and fashion compatibility. (Chen et al., 2019) introduced an industrial-scale Personalized Outfit Generation (POG) model. They developed POG on the Dida platform on Alibaba to recommend personalized fashionable

outfits to customers. They utilized user clicks on preferred clothing items to learn users' preferences in integration with their history of purchased items. they employed a transformer encoder-decoder method to model compatibility among clothing items and users' interests. (Polanía et al., 2019) The proposed method generates recommendations for complementary apparel items given a query item on a siamese network used for feature extraction followed by a fully connected network used to learn a fashion compatibility metric.

The majority of personalized outfit recommendation systems presented in researches are based on the idea of providing an entire outfit based on a single clothing item that the client is exploring (Jaradat et al., 2021) or by offering a complimentary item that completes the look (Kang et al., 2019). Furthermore, another classification is based on personal wardrobe (S. Liu et al., 2012). (Jaradat et al., 2021) declared that outfit recommendation can 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 as a concatenated representation of the image and its relative metadata. The style representation can also be deduced and used as input in the next phases (Jaradat et al., 2021). The compatibility among different clothing items for an outfit can be obtained through experts' opinions like fashion designers and stylists (e.g., X. Song et al., 2018) or personal preferences of customers (e,g. F. Harada et al., 2012). It can also be learned by the system itself based on positive/negative samples of compatible items or scoring similarity between latent representations of different items (Jaradat et al., 2021). Understanding personal preferences can be specified both 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 which is used for learning (Jaradat et al., 2021).

3.4.3 Whole outfit recommendation

Figure 11: Example outfit in the Polyvore68K dataset,

Source: (K. Laenen et al., 2020) Attention-Based Fusion for Outfit Recommendation, Katrien Laenen and Marie-Francine Moens, 2020.

Pioneering research on whole outfit recommendation (e.g., The Complete Fashion Coordinator (Tsujita et al., 2010), What Am I Gonna Wear (Shen et al., 2007), and Magic Closet (Liu et al., 2010)) was dependent on user input concerning the cloths they hold in their wardrobes and the occasions they wear them, outfits suggestion are made also based on historical data, deduced preferred style and matched with the occasion For instance, in the Complete Fashion Coordinator proposed by (Tsujita et al., 2010), the user enters pictures of their own clothes along with some information such as the occasion that item was worn there, it is also possible using social networks they get feedback on them. (S. Liu et al., 2012) suggested Magic Closet, a method for retrieving matching items that can be used by online retailers. It matches each cloth item from the user's wardrobe.

(Jaradat et al., 2021) has identified two main sub-categories of research in outfit recommender systems, including models that used Outfits' Compatibility Scoring and Sequential Outfits Representations and Predictors. According to (Jaradat et al., 2021), A vast majority of researches, including (J. McAuley et al., 2015; Veit et al., 2016; Li et al., 2017; Chen et al., 2019; Shin et al., 2019; Vasileva et al., 2020; S. Song et al., 2018., 2018; Bettaney et al., 2019; Kang et al., 2019) 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. In the second highlighted group (e.g., Jiang et al., 2018; Wang et al., 2018; X. Han et al., 2017; Nakamura et al., 2018) used a sequential method to modeling

the fashion outfit. Each piece of clothing represents a time step, and consistent order of clothing categories is used to assure that no single item in an outfit is duplicated or missing. The advantages of Bi-directional Long short-term memory (LSTM) architectures (Graves et al., 2013) enables modeling the interactions among current, past, and future preferred items using forward and backward LSTMs, the global dependencies between the clothing items can be identified. In the following, we are going to introduce some most recent and significant outfit recommender researches.

(K. Laenen et al., 2020) proposed a model to create an outfit (see figure 11.) regardless of the scratch point or an incomplete one. While they indicate that Outfit recommendation deals with two main challenges, including a. item understanding that demands visual and textual feature extraction and combination to make a better understanding and b. item matching concerning the complexity of the Item compatibility relation, they focused on item understanding. Considering the fact that the role of different Item features may differ in determining compatibility based on the types of items that are selected to be matched. Their proposed model received two triplets as input, one triplet of image embeddings, and the second is a triplet of corresponding description embeddings. These triplets are sent to a semantic space considering that semantic space can capture the concepts of image similarity, text similarity, and image-text similarity better. (K. Laenen et al., 2020) used attention mechanism to focus on interesting parts of the input. Neural machine translation has also been introduced in the attention mechanism itself to leverage fine-grained Item features required to the forefront. It has to be said that this is the first time this concept has been used to provide better item understanding in FRS. Towards developing the proposed system, they compared different attention mechanisms to fuse the visual and textual information to find better performance, Including Visual Dot Product Attention, Stacked Visual Attention, Visual L-Scaled Dot Product Attention, Co-attention. The models have been evaluated on the fashion compatibility (FC) task and the fill-in-the-blank (FITB) task. The images have been represented via the ResNet18 architecture pre-trained on ImageNet. The text descriptions are represented with a bidirectional LSTM. All models are trained for ten epochs using the ADAM optimizer. All models are trained for five runs with the aim of relaxing the effect of the negative sampling. The performance is obtained from averaging the performance on the FC task and FITB task in those five runs. The result of this research shows that "the attention-based fusion mechanism can integrate visual and textual information in a more purposeful way than common space fusion" (K. Laenen et al., 2020). According to the author, Fusion of the information in the product image and description to capture the most important, fine-grained product features into the item representation through this research demonstrate that attention-based fusion improves item understanding.

Unlike low-level visual compatibility (e.g., color, texture), high-level semantic compatibility (e.g., style, functionality) cannot be handled purely based on fashion images. (Sun et al., 2020) developed a state-of-

the-art multimodal framework for fashion compatibility learning, which combined semantic and visual embeddings into an integrated deep learning model. a multilayered Long Short-Term Memory (LSTM) is used for discriminative semantic representation learning tasks. Besides, for visual embeddings, a deep Convolutional Neural Network (CNN) is employed. Then semantic and visual information of fashion items is concatenated using a fusion module, which transforms both into a latent feature space. In addition, for measuring fine-grained relationships between fashion items, a new triplet ranking loss with compatible weights is also proposed, which is more in line with human emotions on comprehending the notion of fashion compatibility. Lots of experiments emphasized the effectiveness of the proposed method, which outperforms the novel methods employed on the Amazon fashion dataset.

(Ding et al., 2021) suggested a novel Attentional Content-level Translation-based Recommender (ACTR) framework that models both the instant user intent and the probability of intent-specific transition for each transition. Three types of explicit instant intents have been defined to model the user intent and predict this intention: match, substitute, and other. (Ding et al., 2021) enhanced the item-level transition modeling with several sub-transitions using various content-level features to further utilize the characteristics of the fashion domain and ease the item transition sparsity problem. To model the interaction among the user, the previously interacted item, and the intent, a transnational operation has been employed.

(Wu et al., 2020) developed a Visual and Textual Jointly Enhanced Interpretable (VTJEI) model for fashion recommendations using the product image and historical reviews. The model generates more accurate recommendations and visual and textual explanations by employing combined improvement of both textual and visual information. They also proposed a bidirectional two-layer adaptive attention review model to receive the user's both implicit and explicit preferences. Besides, (Wu et al., 2020) introduced a review-driven visual attention model to create higher degrees of personalized image representation in accordance with the user's preference using the historical review.

A data-driven novel framework for fashion recommendation was introduced by (M Sridevi et al., 2020). this model employed a Convolutional Neural Network and a Nearest Neighbor Based Recommender. The neural networks are first trained, after which an inventory is chosen for generating suggestions, and a database for the objects in the inventory is produced. Based on the supplied image, the closest neighbor algorithm is utilized to identify the most relevant products, and recommendations are produced.

In the deep learning fashion field, (Chen et al., 2020) suggested a novel task. Considering Attribute editing, generating a photorealistic image combines the texture from different image references. As a result, the highly convoluted attributes and the lack of paired data are the main challenges to these tasks. (Chen et al., 2020) propose a new self-supervised model to integrate garment pictures with separated attributes (e.g.,

vest and sleeves) without paired data to address such limitations. The Model training entails two steps, including self-supervised reconstruction learning and generalized attribute manipulations using adversarial learning. A fully supervised training process for the learning process of each image has been used, an encoder-decoder structure for the self-supervised reconstruction training step has been employed,

(Zhao et al., 2020) developed a state-of-the-art deep neural network, Detect, Pick, and Retrieval Network (DPRNet), to break the gap between fashion products from videos from celebrities and audiences interested in their clothing items to address video-to-shop problems. For object detection from videos, they modified the traditional object detector, which automatically picks out the best object regions without duplication, to improve the performance in the video-to-shop task. Looking at the fashion retrieval task, a multitask loss network has been employed on DeepFashion.

(Stefani et al., 2019) developed an aesthetic-aware clothing recommender system that proposed a collaborative fashion recommendation system (CFRS), introducing the trend score metric as a novel metric. The trend score can be viewed by users in addition to other product details to convey more insight about the products to them and is also used for Sorting products from trendiest to classic ones in each category. The System administrator is responsible for products management, and consequently, trend and user management. The administrator is allowed to modify the trends in the three categories: Colors, prints, and materials. The Fashion expert's category consists of fashion magazine editors, fashion designers, fashion bloggers, etc. these fashion experts can rate fashion trends by like or disliking the items. The user's category can see current fashion trends and follow (like) experts, while visitors can't see or rate fashion trends and just can observe.

To address the shortfalls of previous works which focus on the compatibility of two items or represent an outfit as a sequence and fail to make completely using the complex relations among items in an outfit, (Cui et al., 2019) suggested representing an outfit as a graph in a way that each node shows a category and each edge represents the interaction between two categories. In doing so, each outfit will be considered as a subgraph. (Cui et al., 2019) introduced Node-wise Graph Neural Networks (NGNN) To deduce the outfit compatibility from such a graph. In NGNN, the node interaction for each edge varies. for outfit compatibility estimation, an attention mechanism is employed via learning node representations. NGNN can be employed for modeling outfit compatibility from multiple modalities.

(Ramesh, 2018) proposed a model as an event-based outfit recommender system. In their proposed model: first, 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 a nearest neighbor approach for event recognition task (Ramesh, 2018) employed transfer learning to train models for object detection and used RCNN as the meta-architecture for object detection.

Looking at Instagram, a single photo of an influencer may attract followers to become interested in the same products they used inside the image. Existing solutions to reliably recognize styles and brands using Deep CNN models must also deal with complexities relating to fashion domain data. Clothing information included within a single photograph, on the other hand, only represent a small percentage of the vast and hierarchical space of brands and clothing item attributes. To address these challenges, (Jaradat et al., 2018) proposed two novel approaches to use social media textual content in addition to visual classification in a dynamic way. The first is adaptive neural pruning (DynamicPruning), which uses the clothing item category detected from text analysis in posts to activate the clothing attribute classifier's possible range of links. The second approach (DynamicLayers) is a dynamic structure in which many attribute classification layers exist, and an appropriate attribute classifier layer is dynamically activated based on the image's mined text.

(Gorripati et al., 2018) developed a two-stage deep learning framework that recommends fashion clothes according to the visual similarity style of other images in the dataset. Using images as input, try understanding features. By doing so, a neural network classifier has been employed as an image-based feature extractor. The similarity algorithm is fed by a feature extractor as an input for generating recommendation and ranking suggestions. As a commonly used technology in image recognition, a convolutional neural network is utilized to address the major functionality of classification and recommendation.

3.4.5 Capsule wardrobe recommendations

The capsule closet is undoubtedly one of the most popular and pervasive minimalist notions, focused on people's closets (Heger et al., 2016). Whether it's referred to as a capsule closet, a closet detox, an apparel diet, or a minimalist wardrobe, it can help to shape the future of the 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 (Heger et al., 2016). According to (Hsiao et al., 2018), “given an inventory of candidate garments and accessories, the algorithm must assemble a minimal set of items that provides maximal mix-and-match outfits.” A capsule wardrobe is a collection of clothing items that may be composed to create different sets of outfits that are compatible visually. (Hsiao et al., 2018) proposed an approach to automatically create a capsule wardrobe (figure 12.) by defining the task as a subset selection problem. They provided new insights in both efficient

optimizations .)for a combinatorial mix-and-match outfit selection and generative learning of visual compatibility. There are similar examples including (F. Harada et al., 2012; Dong et al., 2019)

Figure 12: Capsule wardrobe concept image

Source: (Hsiao et al., 2018) Creating Capsule Wardrobes from Fashion Images Wei-Lin Hsiao, Edu Kristen Grauman UT-Austin

3.5. Outfits Recommendations: evaluation metrics

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

- **Compatibility Estimation** is a method of assessing a model's ability to discriminate between compatible and incompatible outfits. Typically, a non-compatible instance can be created for each compatible outfit by substituting one item at a randomly selected place with another item from the clothing items glossary (G. Prato, 2019). This generated example is then labeled as non-compatible. Then, the task converts into a binary classification problem where the model is trained with compatible and non-compatible outfits. In this approach, the area under the curve (AUC) of the receiver operating characteristic (ROC) is a typically used evaluation measurement. According to (G. Prato, 2019; Jaradat et al., 2021), The relative Compatibility Estimation task equation has been represented as follows:

$$AUROC = \int_0^1 TPR(FPR^{-1}(x))dx \tag{1}$$

- **Fill in the Blanks** is a common fashion compatibility test (Jaradat et al., 2021). The following is a common formulation: One of the clothing items (e.g., xi) is randomly masked out of an ensemble consisting of x1,....xi1, xi, xi+1....,xn clothing items. The next step is to estimate which item in the whole outfit will go best with the others.

$$Accuracy = \frac{|guessed\ missing\ items|}{|questions|} \tag{2}$$

Where the appropriately recommended items are the guessed missing items, and questions are the subsets of options for recommendation (G. Prato, 2019).

- **Unconstrained Outfit Completion task** all of the metrics utilized in the OUC 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, at which the majority of the items appear only once in the entire datasets, and some items also are labeled twice with various ids. Using a cut-off of k = 6, for instance, precision will be restricted to a maximum of 0.3 in the great majority of queries, as merging of testing groups would be infrequent (G. Prato, 2019). The relative equations have listed as follows:

$$Precision@k = \frac{|guessed\ items|}{k} \tag{3}$$

$$Recall@k = \frac{|guessed\ items|}{|missing\ items|} \tag{4}$$

$$F1@k = 2\frac{Precision@k \cdot Recall@k}{Precision@k + Recall@k} \tag{5}$$

$$RR@k = \begin{cases} \frac{1}{rank\ guessed\ item} & if\ guessed \\ 0 & otherwise \end{cases} \tag{6}$$

$$MAP@k = \frac{\sum_{c=1}^{k} Precision@c \cdot rel(c)}{|missing\ items|} \tag{7}$$

$$\text{where } rel(c) = \begin{cases} 1 & \textit{if recommendation@c is a missing item} \\ 0 & \textit{otherwise} \end{cases} \quad (8)$$

- **Outfits ranking accuracy** can be measured employing metrics as Normalized Discounted Cumulative Gain (NDCG). Applying ranking metrics for the task of 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 (Jaradat et al., 2021). This necessitates specifying guidelines for the user's positive or neutral choices. According to some academics, an 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 (Jaradat et al., 2021).

4. AI-based Recommender System (role of computer vision in fashion recommender systems)

4.1. AI in recommender systems

Various AI techniques have lately been used to recommender systems, improving user pleasure and experience (Q. Zhang et al., 2020). AI allows for a higher quality of recommendation than is possible with traditional methods. This has opened a new era for recommender systems, allowing for deeper insights into user-item relationships, the presentation of more complex data representations, and the uncovering of comprehensive data in demographical, textual, virtual, and contextual data (Q. Zhang et al., 2020).

4.1.1. Artificial intelligence: main models and methods

Artificial intelligence (AI) is a rapidly evolving field with applications ranging from chess to learning systems and illness diagnosis (Luger, 2005). AI approaches are being developed with the purpose of automating intelligent behaviors in six areas: knowledge engineering, reasoning, planning, communication, perception, and motion **(Russell et al., 2016)**. Knowledge engineering, in particular, refers to strategies for knowledge representation and modeling that allow robots to understand and process knowledge. For problem-solving and logical deduction, reasoning techniques are developed. The purpose of planning is to assist machines in setting and achieving a goal. Understanding natural language and communicating with humans are the goals of communication; The purpose of perception is to analyze and process inputs such as images or speech, while the role of motion is to move and manipulate. Aside from motion, approaches from the first five domains can be used to improve and accelerate the creation of recommender systems, which is necessary due to the massive amount of data that needs to be processed.

As illustrated in Figure 13., (Q. Zhang et al., 2020) introduced eight main models and techniques. Deep neural networks, transfer learning, active learning, and fuzzy techniques are interconnected and serve as representations of knowledge and reasoning. Reasoning and planning are associated with evolutionary algorithms and reinforcement learning, whereas communication and perception are addressed by natural language processing, and image perception is approached by computer vision. They also remarked Natural language processing and Computer vision as two main AI application areas in recommender systems amongst those eight methods. According to the scope of this thesis, we are focusing on using Computer vision methods in image-based fashion recommender systems considering the advances in deep learning techniques.

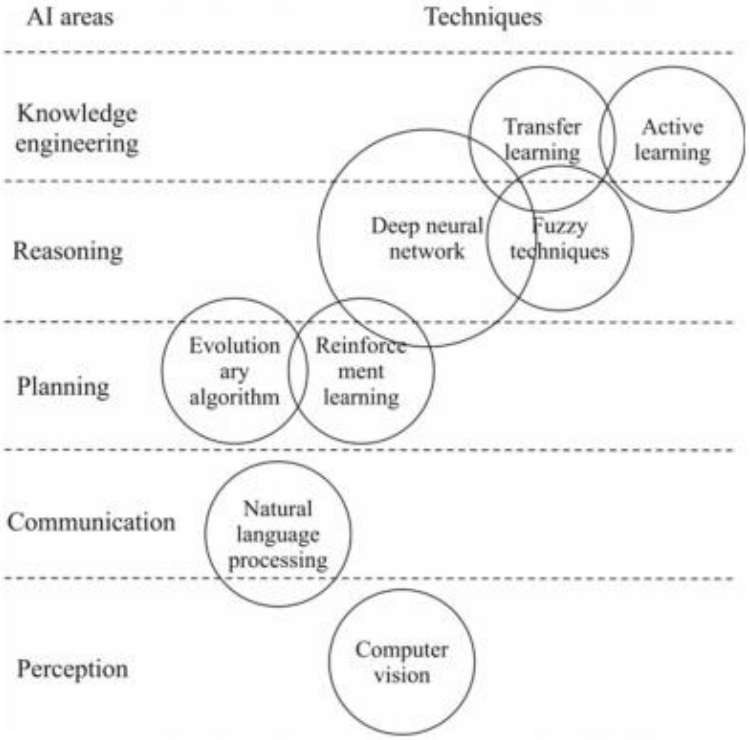

Figure 13: AI areas and techniques

Source: (S. Zhang et al., 2019) Shuai Zhang, Lina Yao, Aixin Sun, Yi Tay, Deep Learning based Recommender System: A Survey and New Perspectives, 2019

4.2 Deep Learning

Deep learning is perceived as a sub-field of machine learning (S. Zhang et al., 2019). DL is mainly based on Artificial Neural Networks (ANNs) (Mahony et al., 2018), a computation paradigm that is inspired by the functioning of the human brain. Deep learning is defined by the ability to learn deep representations from data, which may entail learning several levels of representations and abstractions (S. Zhang et al., 2019). Object identification, motion tracking, action recognition, human position estimation, and semantic segmentation are just a few of the computer vision tasks that have benefited from deep learning (A. Voulodimos et al., 2018).

4.2.1 Role of Deep Neural Networks for Recommender systems

For many online businesses and mobile apps, recommender systems are perceived as essential tools for improving customer experience and boosting sales/services (S. Zhang et al., 2019). Nowadays, an increasing number of companies tend to use deep learning to improve the quality of the recommendation they provide for their customers (S. Zhang et al., 2019). RecSys1, the world's foremost international

conference on recommender systems, began regular workshops for recommender systems, particularly on deep learning, in 2016(S. Zhang et al., 2019).

According to (S. Zhang et al., 2019), when it comes to proposing specific architecture and to the particular scenario to have the most optimum result, it is highly dependent on the issue of the task, domains, and recommender scenarios. They explained that there is no reason not to use deep learning-based techniques for the creation of any recommender system in today's research (and even industry) setting.
(S. Zhang et al., 2019) highlighted the strengths of deep learning-based recommendation models as following:

• **Representation Learning**. Deep neural networks excel at extracting underlying explanatory elements and usable representations from incoming data. Deep neural networks have undeniable advantages for representation learning: (1) It decreases the time and effort required to design handcrafted features. With automatic feature learning from raw data in an unsupervised or supervised fashion, (2) it allows recommendation models to contain multimodal content information such as text, graphics, audio, and even video (S. Zhang et al., 2019).

• **Nonlinear Transformation.** Deep neural networks, unlike linear models, can simulate nonlinearity in data with nonlinear activations such as relu, sigmoid, tanh, and so on. This aspect allows complicated and detailed user-item interaction patterns to be captured. Matrix factorization, machine, and sparse linear models are examples of traditional linear models (S. Zhang et al., 2019).

• **Flexibility**. Many popular deep learning frameworks, such as Tensorflow, Keras, Caffe, MXnet, DeepLearning4j, PyTorch, Theano, and others, have increased the adaptability of deep learning approaches. The majority of these tools are modularly designed and have a strong community and professional support. Development and engineering become much more efficient as a result of good modularization. It's simple to mix different neural architectures to create strong hybrid models or swap out one module for another to capture multiple characteristics and factors simultaneously (S. Zhang et al., 2019).

• **Sequence Modelling.** On a variety of sequential modeling problems, such as machine translation, deep neural networks have shown promising outcomes. In these tasks, RNN and CNN are significant. RNN does this through internal memory states, whereas CNN achieves it through time-sliding filters. Both are widely useful and adaptable when it comes to mining data's sequential structure. Mining the temporal dynamics of user behavior and item change necessitates the modeling of sequential data. Next-item/basket prediction and session-based suggestion are two examples of typical applications (S. Zhang et al., 2019).

4.2.2 Two main categories of deep learning-based recommendation models

To provide a wide perspective on this field, (S. Zhang et al., 2019) classified models based on the types of employed deep learning techniques. Figure 14. summarizes the classification scheme.

These classifications are as follows:

• **Recommendation with Neural Building Blocks** Models is grouped into eight subcategories in this area, in accordance with the eight deep learning models listed above: MLP, CNNs, AE, AM, AN, RNNs, RBM, DRL, and NADE based recommender system. The applicability of a recommendation model is determined by the deep learning technique used. CNN's for example can extract local and global representations from multiple data sources such as textual and visual information; MLPs, can easily model non-linear interactions between users and items, and RNNs can model the temporal dynamics and sequential formation of content information in the recommender system.

• **Recommendation with Deep Hybrid Models** Some recommendation models based on deep learning employ multiple deep learning techniques. Deep neural networks' adaptability allows them to join numerous neural building blocks to create a more efficient hybrid model. There are different possible ways to combine deep learning techniques, but not all have been used.

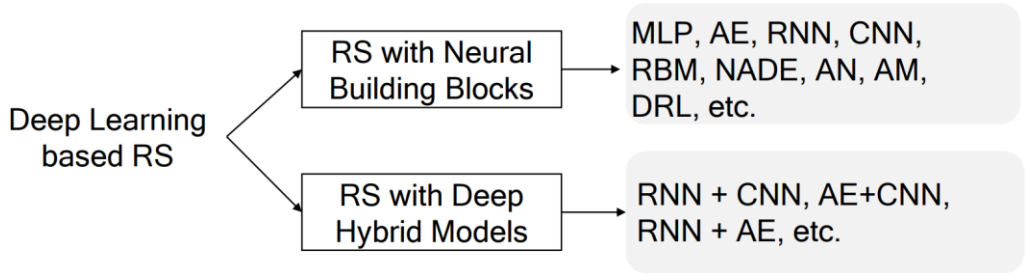

Figure 14: Categories of deep neural network-based recommendation models.
Source: (S. Zhang et al., 2019) Shuai Zhang, Lina Yao, Aixin Sun, Yi Tay, Deep Learning based Recommender System: A Survey and New Perspectives, 2019

4.3. Computer vision

(Wiley et al., 2018) There is an overlap with Image Processing on basic approaches, and some writers use both words interchangeably. Image understanding is the result of the Computer Vision process. Image Processing is concerned with performing computational transformations for images, such as sharpening and contrast, whereas Computer Vision is concerned with creating models and data extractions from images. Computer vision is defined as "the automatic extraction, analysis, and understanding of meaningful information from a single image or a sequence of images" from an engineering perspective. In Computer Vision, the principal element is to extract the pixels from the image to study the objects and thus understand what it contains.

According to (Rani et al., 2020), 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).

in “Fashion Meets Computer Vision: A Survey,” (Cheng et al., 2020) determined main tasks of computer vision in fashion domain into four main categories: detection, analysis, synthesis, and recommendation.

Image processing algorithms and approaches were mostly used in computer vision. The main task of computer vision was to extract the image's features (Rani et al., 2020). The initial stage in completing a computer vision task was to detect color, edges, corners, and objects. These features are human-engineered, and the extracted features and the methodologies employed for feature extraction directly impact the model's accuracy and reliability. (Mahony et al., 2018) for tasks such as image classification, an initial stage called feature extraction was carried out before the advent of DL. In the conventional vision domain, techniques such as SIFT (Scale-Invariant Feature Transform), BRIEF (Binary Robust Independent Elementary Features), and SURF (Speeded-Up Robust Features) plays a critical role in extracting the features from the raw image (Mahony et al., 2018) mostly for object detection.

4.4. Deep Learning in Computer Vision

(Rani et al., 2020) “deep learning is playing a major role as a computer vision tool.” (Mahony et al., 2018). “Deep Learning has pushed the limits of what was possible in the domain of Digital Image Processing.” (Mahony et al., 2018). “Deep Learning (DL) is used in the domain of digital image processing to solve difficult problems (e.g., image colorization, classification, segmentation, and detection).”

(A. Voulodimos et al., 2018) highlighted Convolutional Neural Networks (CNNs), the “Boltzmann family” including Deep Belief Networks (DBNs) and Deep Boltzmann Machines (DBMs), and Stacked (Denoising) Autoencoders as three of the most important types of deep learning models with respect to their applicability in visual understanding. (Mahony et al., 2018) stated, “DL is still only a tool of CV. For example, the most common neural network used in CV is CNN. But what is a convolution? It’s, in fact, a widely used image processing technique (e.g., see Sobel edge detection). (Ganesan et al., 2017) Convolutional Neural Networks have proven to be extremely effective in a variety of computer vision applications, including object recognition, detection, image segmentation, and texture synthesis.

Neural networks were mostly neglected by the machine learning scientists in the 1980s, however; by the late 1990s, a distinct type of deep feedforward network known as a convolutional neural network (CNN) had been developed that is considerably easier to train, besides CNNs are also considerably more generalizable than classic neural networks, which is why they were immediately adopted in speech recognition and computer vision (Q. Zhang et al., 2020).

4.5. Traditional computer vision techniques vs. deep learning

Significant advances in deep learning (DL) are increased hardware technologies such as computing power, memory capacity, power consumption, image sensor resolution, and optics. These have enhanced the performance and cost-effectiveness of vision-based applications, hastening their adoption. DL allows CV engineers to achieve improved accuracy in tasks like image classification, semantic segmentation, object recognition, and Simultaneous Localization and Mapping (SLAM) compared to traditional CV techniques. Because DL applications are trained rather than coded, they require less expert analysis and fine-tuning and can take advantage of the massive amounts of video data accessible today. In contrast to CV techniques, which tend to be more domain-specific, DL methods offer greater flexibility because CNN models and frameworks can be re-trained using a new dataset for every use case.

The problem with this conventional method is that it requires determining which aspects in each image are most essential. The process of extracting features becomes more difficult as the number of classes to categorize grows. To identify which features best characterize distinct kinds of objects, the CV engineer must use his or her judgment and go through a rigorous trial and error process. Furthermore, each feature definition necessitates dealing with many parameters, all of which must be fine-tuned. As a significant advancement, end-to-end learning was developed by DL, in which the machine is simply given a dataset of images that have been labeled, where neural networks recognize the hidden patterns in image classes and automatically calculate the most descriptive and salient attributes for each object class. DNNs have been known to outperform traditional algorithms, although there are trade-offs in computing resources and training time.

Depicted as Figure 15., by substituting hand-crafted feature extraction with knowledge and experience that has been iterated through deep learning architectures, using DL approaches in CV drastically improved the workflow of the CV engineer.

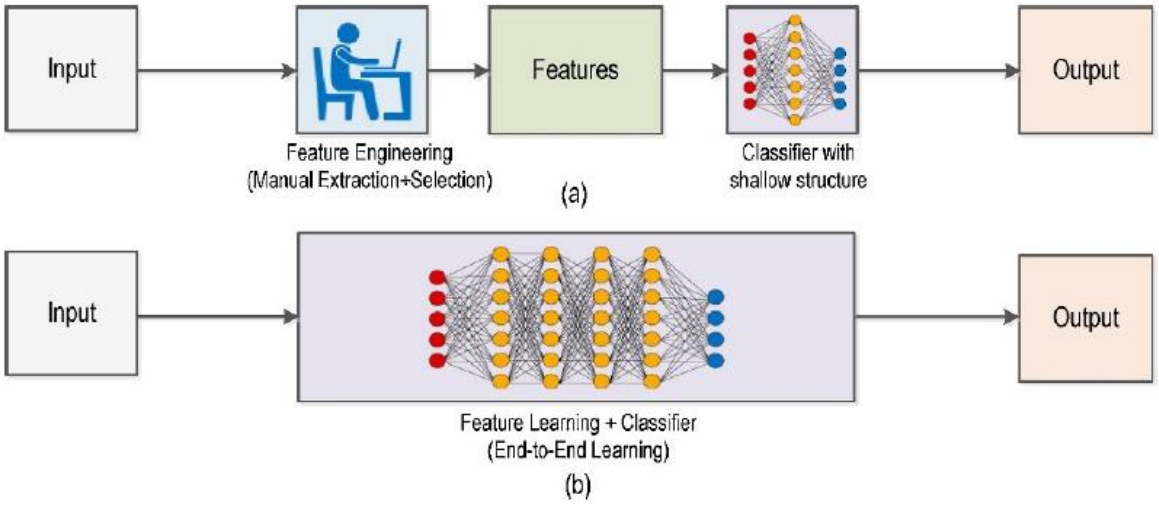

Figure 15: (a) Traditional Computer Vision Workflow vs. (b)Deep Learning Workflow
Source: (Mahony et al., 2018) Niall O' Mahony, Sean Campbell, Anderson Carvalho, Suman Harapanahalli, Gustavo Velasco Hernandez, Lenka Krpalkova, Daniel Riordan, Joseph Walsh, Deep Learning vs. Traditional Computer Vision, 2018

The development of CNNs has had a significant impact on CV in recent years and is credited for a substantial increase in the ability to recognize objects (A. Voulodimos et al., 2018).

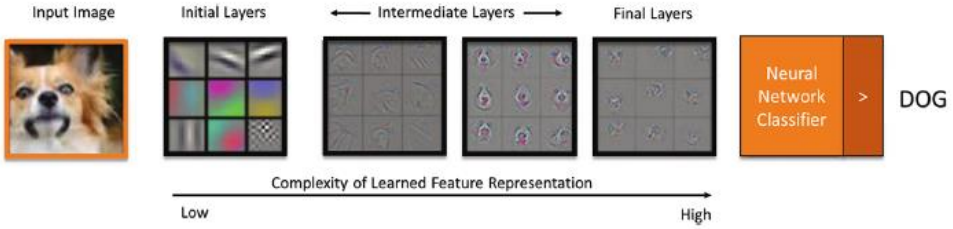

Figure 16: CNN learns low-level features at the initial layers. As going more profound, the features become more complex (Zeiler et al., 2014; Khan et al., 2018; Gkelios et al., 2021)

Convolutional Neural Networks (CNNs) have a hierarchical structure that starts with the input layer and progresses through numerous hidden layers to the output layer representation (Gkelios et al., 2021). The performance of CNNs is directly influenced by the strategy of training samples extraction, represented as Figure 16. the hierarchy of feature representation in CNNs. The initial layers of CNN compress the input image by extracting favorable features like edges and curves, with a focus on local features. Going deeper, the network may extract higher-level properties like hands or legs and combine them into global features. The features extracted in the upper layers of a CNN trained to classify images are useful for image retrieval applications.

Kernels (also known as filters) are used by CNNs to detect features (such as edges) across an image. A kernel is just a set of weighted values in a matrix that has been trained to detect particular features. The fundamental principle behind CNNs is to spatially convolve the kernel on a given input image to see if the feature it's supposed to detect is there (Mahony et al., 2018).

The output of the convolution layer is summed with a bias term and then sent to a non-linear activation function to aid in the learning of kernel weights. Non-linear functions such as Sigmoid, TanH, and ReLU are commonly used as activation functions (Rectified Linear Unit). These activation functions are chosen based on the nature of the data and classification tasks. ReLUs, for example, are considered to have a higher biological representation (neurons in the brain). Thus, it provides sparser, more efficient representations and is less sensitive to the vanishing gradient problem, which results in better outcomes for image recognition tasks (Mahony et al., 2018).

Conventional CV approaches such as SIFT and SURF and traditional machine learning classification algorithms such as Support Vector Machines and K-Nearest Neighbors are frequently integrated with traditional CV techniques to overcome the aforementioned CV challenges and increase performance. SIFT methods and simple color thresholding and pixel counting algorithms are not class-specific, meaning they perform the same for any image. On the other hand, deep neural net features are particular to your training dataset and, if poorly created, are unlikely to function well for images other than the training set. As a result, SIFT and other algorithms are useful for applications such as 3D mesh reconstruction, which don't require specific class knowledge (Mahony et al., 2018).

There are significant trade-offs between traditional CV and deep learning-based approaches. Conventional CV algorithms are well-established, interpretable, and provide optimum performance and power efficiency, whereas DL improves adaptability at the cost of a large number of processing resources (Mahony et al., 2018).

Traditional CV and deep learning techniques have recently been integrated with hybrid methodologies, which incorporate the benefits of both fields. Furthermore, the hybrid approach utilizes half the memory bandwidth and uses fewer CPU resources (Mahony et al., 2018).

DL also introduces limitations. The results of DL vision processing are also influenced by image resolution. The most recent deep learning algorithms may achieve far higher accuracy. Still, at the cost of significant additional math calculations which demand a higher computing power, this makes use of particular hardware essential. Millions of data records are also needed for DL. Because it involves trial and error with multiple training parameters, there is usually a need for a large number of iterations in a given application. The most prevalent approach for minimizing training time is transfer learning. CV techniques are widely

used to enhance training data through data augmentation or reduce data to a specific feature through various pre-processing stages. Before training your model, pre-processing requires modifying the data (typically using classic CV approaches). Data augmentation is a common pre-processing strategy when there is a lack of training data. (G. Marcus et al., 2018) suggested that deep learning must be combined with other techniques if the aim is to obtain artificial general intelligence. G. Marcus et al. illustrated DL algorithms' limitation to learning visual relations, or if specific objects in an image are the same or not. (Mahony et al., 2018) also declared that for some cases, traditional techniques with global characteristics are a better solution. The emergence of DL may offer several possibilities for traditional methodologies to address the many issues that DL provides (e.g., computing power, time, accuracy, characteristics and quantity of inputs, etc.).

According to (Mahony et al., 2018), there are lots of challenging problems in the computer vision field which cannot be easily handled by deep learning, but they can benefit from "conventional" solutions, including Robotics, augmented reality, automatic panorama stitching, virtual reality, 3D modeling, motion estimation, video stabilization, motion capture, video processing, and scene understanding

4.6. Computer vision in Fashion recommender systems

(Hsiao et al., 2018) The fashion domain is a magnet for computer vision. New vision tasks are emerging following with fast growth of the fashion industry towards an online, social, and extremely personalized business domain. Style models (M. Kiapour et al., 2014; E. Simo-Serra et al., 2016; K. Matzen et al., 2017; Hsiao et al., 2017; H. Lee et al., 2017), forecasting trends (A. Halah et al., 2017), interactive search (B. Zhao et al., 2017; A. Kovashka et al., 2012), and recommendation (Y. Hu et al., 2015; A. Veit et al., 2015; S. Liu et al., 2012) all demand visual understanding considering all details with delicacy.

In “Fashion Meets Computer Vision: A Survey,” (Cheng et al., 2020) declared, “Current researches on intelligent fashion (according to Cheng et al., intelligent fashion refers to the fashion technology which is empowered by computer-vision) covers the research topics not only to detect what fashion items are presented in an image but also analyze the items, synthesize creative new ones, and finally creating customized recommendations.” (Cheng et al., 2020) in recent years, computer vision researchers have paid a lot of attention to fashion, mostly expressed visually. From a technical point of view, intelligent fashion is a complex process because, unlike generic objects, fashion items have notable stylistic and design variations. Most importantly, there is a large semantic gap between computable low-level features and the high-level semantic concepts that they encode. (S. Song et al., 2018), published a paper in 2018 that summarized developments in multimedia fashion research and divided fashion tasks into three categories: low-level pixel computing, mid-level fashion understanding, and high-level fashion analysis. Human

segmentation, landmark identification, and human pose estimation are examples of low-level pixel computation. Mid-level fashion understanding aims to recognize fashion images, such as fashion items and fashion styles. High-level fashion analysis includes recommendation, fashion synthesis, and fashion trend prediction. (Ganesan et al., 2017) Object recognition, object detection, image segmentation, and texture creation are just a few of the computer vision tasks that Convolutional Neural Networks have excelled in. In “Fashion Object Detection and Pixel-Wise Semantic Segmentation,” (Mallu 2018) explained numerous possibilities of enhancing the fashion technology with deep learning. One of the key ideas is to generate fashion styles and recommendations using artificial intelligence. Likewise, another significant feature is gathering reliable information on fashion trends, which includes analysis of existing fashion-related images and data.

When dealing with images, localization and segmentation are well known to address in-depth studies on pixels, objects, and labels present in the image. Computer vision has a plethora of techniques of processing an image, out of which localization and segmentation can serve as significant candidates for detailed image analysis. In general, localization involves object detection in an image, while segmentation includes pixel-level analysis. Localization and segmentation are most commonly performed to study information embedded in an image. Such information is useful for the development of deep learning architectures (Mallu, 2018). Segmentation is more detailed and compound in comparison to localization (Mallu, 2018). In segmentation, an image is broken down into different regions where several pixels or groups of pixels represent contrasting areas or objects, e.g., pixels identified as background in an image have a high probability of not containing the object of interest (Mallu, 2018).

4.7. The evolution of CV methods with DL advancements in FRS

(L. Wu et al., 2021) introduced Content-based models and hybrid recommendation models as the two types of the most recent image recommendation solutions via modeling image information. Visual signals are used to generate item visual representations in content-based models, and the customer preference is represented in the visual space (Q. Zhang et al., 2017; J. Wen et al., 2016; Hou et al., 2019; Alashkar et al., 2017; Q. Xu et al., 2018; J. McAuley et al., 2015; Rawatet al., 2016; C. Lei et al., 2016; Q. Liu et al., 2017). Hybrid recommendation models, on the other hand, use item visual modeling to solve the problem of data sparsity in CF (J. Chen et al., 2017; S. Wang et al., 2017; Â. Cardoso et al., 2018; R. He et al., 2016; Yu et al., 2018). (R. He et al., 2016) incorporate visual content to develop a unified hybrid recommendation system as Visual Bayesian Personalized Ranking (VBPR). Each user (item) is displayed in two latent spaces in this method: a visual space projected using CNN-based visual features and a collaborative latent space used to detect users' latent preferences. Given a user-item pair with a corresponding image, the projected preference is learned by integrating users' preferences from two areas. Following the primitive idea of

VBPR, rather than representing users' preference in two spaces, matrix factorization-based models have used the item's visual content as a regularization term, assuring that the learned item latent vector is similar to the visual image representation obtained by CNNs (R. He et al., 2016; L. Wu et al., 2021).

As previously discussed in 3.4.1, Content-Based Image Retrieval (CBIR) techniques have attracted lots of interest among scholars in the fashion recommender system domain. Content-Based Image Retrieval (CBIR) has also been widely used and improved via different computer vision and artificial intelligence approaches (X. Li et al., 2021). According to (Tian et al., 2018; Gkelios et al. 2021), 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. Following the proposed evolutionary eras of Content-based image retrieval techniques, we mapped out the main milestones of proposing fashion recommender systems in the Figure below.

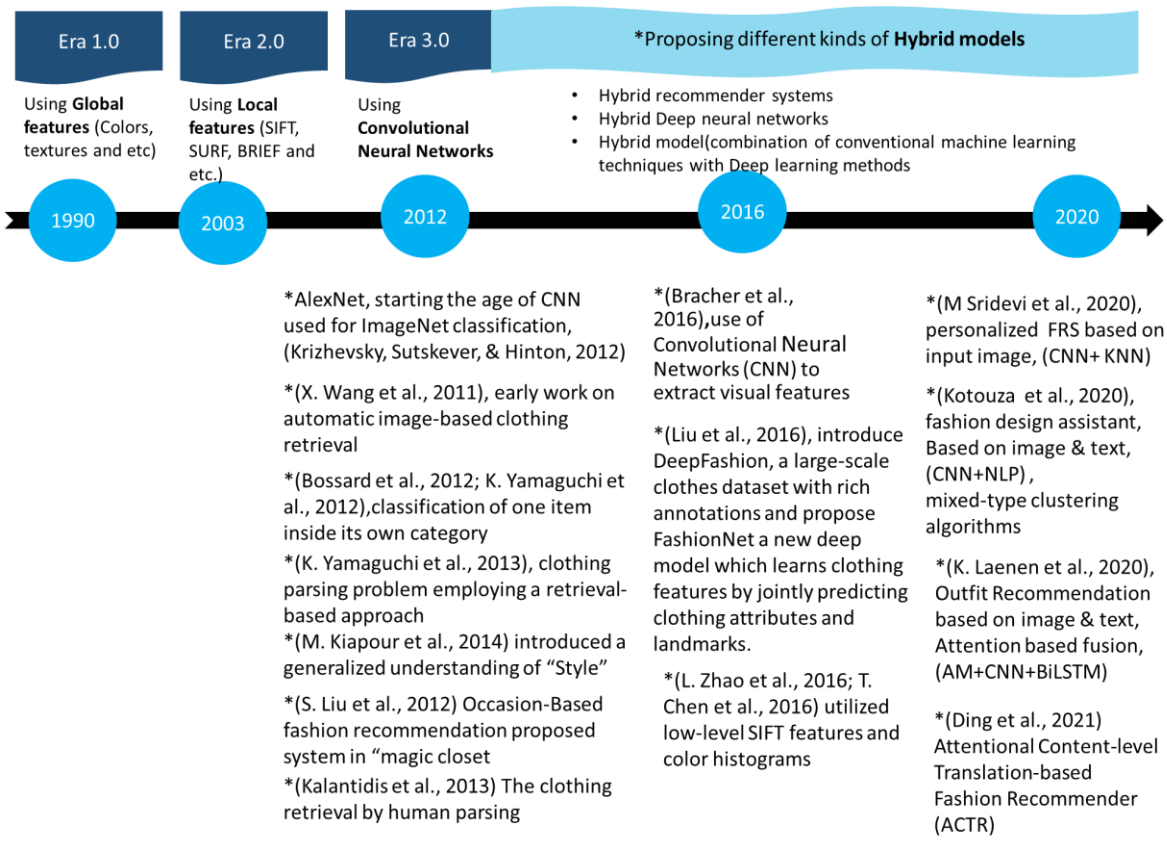

Figure 17: The evolution of CV methods with DL advancements in FRS

Conventional CV methods have been used in the first two eras for image retrieval tasks, particularly for object detection, including feature descriptors (SIFT, SURF, BRIEF, and so on). As it will be discussed in

the next sections, for applications like image classification before the rise of deep learning, feature extraction was used. A growing body of knowledge indicated the significant use of Convolutional Neural networks since 2012, called Era 3.0, in the evolution of Content image retrieval systems in the fashion domain. (T. Iwata et al., 2011) Utilized fashion magazines photographs to generate recommendations. To meet the diverse needs of different users, intelligent clothing recommender systems are studied (S. Liu et al., 2012) based on the principles of fashion and aesthetics. In (Bossard et al., 2012; K. Yamaguchi et al., 2012), the main idea was classifying one item inside its category. Kota Yamaguchi and his colleagues (K. Yamaguchi et al., 2012). Later, several researchers proposed different methods for clothing parsing with various inputs (J. Dong et a, 2015). Kota Yamaguchi and his colleagues developed an effective way to Parsing clothing in fashion photographs (K. Yamaguchi et al., 2012), which is challenging because of the large number of items and the variations in setting and garment appearance layering, and occlusion. In addition, they provided a large novel dataset and tools for labeling garment items for subsequent researches. (K. Yamaguchi et al., 2013) Yamaguchi and his colleagues addressed the clothing parsing problem employing a retrieval-based approach. Then, following this path through the next milestone where the concept of "style" has gotten more attention, particularly by researchers like (M. Kiapour et al., 2014), who tend to provide a more generalized understanding of "Style" and creates valuable data set in this regard. Generating garment recommendations, customer ratings, and clothing was utilized (X. Hu et al., 2014). The history of clothes and accessories, weather conditions were considered in (C. Limaksornkul et al., 2014) to generate recommendations. Recently, there have been impressive advances in computer vision tasks like object recognition and detection, and segmentation (A. Krizhevsky et al., 2012; S. Ren et al., 2015; L. Chen et al., 2016). The revolution started with Krizhevsky et al. substantially improving object recognition on the ImageNet challenge using convolutional neural networks (CNN). This led to research and subsequent improvements in many tasks related to fashion, such as classification of clothes, predicting different kinds of attributes of a specific piece of clothing, and improving the retrieval of images (Z. Liu et al., 2016; Y. Kalantidis et al., 2013; Bossard et al., 2012; T. Xiao et al., 2015; M. Kiapour et al., 2014).

From 2016, research in the field had a significant growth; the main innovation that was developed was the use of Convolutional Neural Networks (CNN) as the primary way to extract visual features from item's images as in (C. Bracher et al., 2016); this method of extracting visual features has become, in a wide variety of different fields of application, giving researchers the chance to possibly identify the complex and non-linear relationship between the visual characteristics of items. (D. Sha et al., 2016) also provided recommendations by extraction of several features from the images to learn the contents such as material, collar, sleeves, etc. (Z. Liu et al., 2016) create a robust fashion dataset of about 800,000 images that contain annotations for various types of clothes, their attributes and the location of landmarks as well as cross-domains pairs. They also design a CNN to predict attributes and landmarks. The architecture is based on a

16-layer VGGNet and adds convolution and fully connected layers to train a network to predict them. (Gatys et al., 2015&2016), describes the process of using image representations encoded by multiple layers of VGGNet to separate the content and style of images and recombine them to form new images. They designed a new parametric model for texture synthesis based on convolutional neural networks. They model the style of an image by extracting the feature maps generated when the image is fed through a pre-trained CNN, in this instance using a 19-layer VGGNet. The idea of style extraction is based on the texture synthesis process that represents the texture as a Gram Matrix of the feature maps generated from each convolutional layer. The style is extracted as a weighted set of gram matrices across all convolutional layers of the pretrained VGGNet when it processes an image. The content is obtained from feature maps extracted from the higher layers of the network when the image is processed. The style and content losses are computed as the mean squared error (MSE) between the original image's features maps and Gram matrices and a randomly generated image (initiated from white noise). (Ganesan et al., 2017) in "Fashioning with Networks: Neural Style Transfer to Design Clothes," used a 19-layer pre-trained VGGNet, Scotia, et al., developed a method for personalizing and creating new personalized clothes based on a user's preferences and by learning the user's fashion preferences from a small set of clothes in their wardrobe. In order to create fresh personalized garments, the neural style transfer algorithm is applied to fashion. In "Personalized fashion recommender system with image-based neural networks" as the most recent study, M. Sridevi et al., 2020, used neural networks to process the images from the DeepFashion dataset and the nearest neighbor backed recommender to generate the final recommendations based on a given input image, to find the most similar one.

4.8. A categorization on DL-based fashion recommender systems

Table 2. Categories of deep neural network-based fashion recommender systems

	Factor	Method	Papers
Input	Side information	Utilize (image/ Image& text)	(Sridevi et al., 2020), (W. Yu et al., 2021), (C. Bracher et al., 2016),(K. Laenen et al., 2020), (R. He et al., 2016), (Goncalves et al., 2019), (Kang et al., 2017), (J. McAuley et al., 2015), (Kalantidis et al., 2013), (He et al., 2018),(S. Keerthi et al., 2018), (Polanía et al., 2019), (X. Geng et al., 2015),(M. Vasileva et al., 2018), (Y. Li et al., 2017), (A. Veit et al., 2017), (R. He et al., 2016), (Kipf et al., 2017; Z. Zhang et al., 2020; L. Wu et al., 2021), (Ganesan et al., 2017), (Gatys et al., 2015&2016)
	Behavior type	user clicking records/ interaction history	(Ding et al., 2021), (N. Kato et al., 2019), (Rodríguez et al., 2020)
		User past feedback	(R. He et al., 2016)
		Sequential pattern of behavior (the most recent purchased items)	(J. Tang et al., 2018)
	Repeat consumption	user's purchased items, purchased/viewed items user's co-purchase data	(Ding et al., 2021) (J. McAuley et al., 2015) (A. Veit et al., 2015)
Model Structure	FRS with single Neural building blocks	P-GANs	(N. Kato et al., 2019)
		GNN	(Kipf et al., 2017; Z. Zhang et al., 2020; L. Wu et al., 2021)
		STAMP	(J. Wu at al., 2019), (Rodríguez et al., 2020)
		NARM	(Rodríguez et al., 2020)
		CNN	(S. Keerthi et al., 2018), (X. Geng et al., 2015), (R. He et al., 2016), (Gatys et al., 2015&2016)
		SCNN	(A. Veit et al., 2015)
		AM+MTL	(Ding et al., 2021)
		CNN+KNN	(Sridevi et al., 2020), (Goncalves et al., 2019), (Kalantidis et al., 2013),
		CNN+WNN	(J. McAuley et al., 2015)
		CNN+SVM	(Ganesan et al., 2017)
		Deep CNN+KNN	(R. He et al., 2016)

		GRU4REC+KNN	(Jannach et al., 2017)
	FRS with Deep Hybrid models	CNN+NLP	(Kotouza et al., 2020), (J. Tang et al., 2018)
		AM+CNN+BiLSTM	(K. Laenen et al., 2020)
		CNN+BDN	(W. Yu et al., 2021)
		CNN+ FCDNN	(C. Bracher et al., 2016)
		CNN+CSNs	(A. Veit et al., 2017)
		SCNN+ CSNs	(M. Vasileva et al., 2018)
		SCNN+GANs	(Kang et al., 2017)
		SCNN+FCDNN	(Polanía, et al., 2019)
		DCNN+FCNN	(He et al., 2018)
		DCNN+MLP	(Y. Li et al., 2017)

Represented in Table 2., the DL-based fashion recommender systems have been categorized based on two main categories, including 1) fashion recommender systems with single neural building blocks and 2) fashion recommender systems with deep Hybrid models. The input factor of each fashion recommender system also has been demonstrated in this table. In following each of these studies has been introduced, briefly.

(Sridevi et al., 2020), proposed a personalized fashion recommender system that generates recommendations for the user based on an input given. (Kotouza et al., 2020) proposed a semi-autonomous decision support system that is focused on the creative part of the design for assisting fashion designers via retrieving, combining, and organizing data from different sources and considering the designer's personal preferences (Ding et al., 2021). In the field of sequential fashion recommendation, Ding et al. proposed a novel Attentional Content-level Translation-based Fashion Recommender (ACTR). They enhance the sequential fashion recommendation model by modeling the user's instant intent and incorporating the item's content-level attributes. In (K. Laenen et al., 2020), the main purpose was to create an outfit regardless of the scratch point or an incomplete one. Fussing the product image and description information to capture the most important, fine-grained product features into the item representation demonstrates that the attention-based fusion improves item understanding. (K. Laenen et al., 2020) proposed a model to create an outfit regardless of the scratch point or an incomplete one. while they indicate that Outfit recommendation is dealing with two main challenges, including a. item understanding that demands visual and textual feature extraction and combination to make a better understanding and b. item matching with respect to the complexity of the Item compatibility relation, they focused on item understanding. Considering the fact that the role of different Item features may differ in determining compatibility based on the types of items that are selected to be matched. Their proposed model received two triplets as input, one triplet of image embeddings, and the second is a triplet of corresponding description embeddings. These triplets are sent to a semantic space considering that semantic space can capture the concepts of image similarity, text similarity, and image-text similarity better. (K. Laenen et al., 2020) used attention mechanism to focus on interesting parts of the input. neural machine translation has also been introduced in the attention mechanism itself to leverage fine-grained Item features required to the forefront. It has to be said that this is the first time this concept has been used to provide better item understanding in FRS. towards developing the proposed system, they compared different attention mechanisms to fuse the visual and textual information to find better performance, Including Visual Dot Product Attention, Stacked Visual Attention, Visual L-Scaled Dot Product Attention, Co-attention. The models have been evaluated on two tasks, including the fashion compatibility (FC) task and the fill-in-the-blank (FITB) task. The images have been represented via the ResNet18 architecture (Ren et al., 2016) pre-trained on ImageNet. The text descriptions are represented with a bidirectional LSTM. The parameters of the ResNet18 architecture and

the bidirectional LSTM are both finetuned. All models are trained for ten epochs using the ADAM optimizer. All models are trained for five runs with the aim of relaxing the effect of the negative sampling. The performance is obtained from averaging the FC and FITB tasks' performance in those five runs. the result of this research shows that “the attention-based fusion mechanism is able to integrate visual and textual information in a more purposeful way than common space fusion” (K. Laenen et al., 2020). (W. Yu et al., 2021), developed aesthetic-aware clothing recommender systems. Proposing the state-of-the-art aesthetic deep model tensor factorization model, optimized with pairwise learning., negative sampling strategies. (C. Bracher et al., 2016) a combined model of both attribute- and image-based architectures, DNN based model. forecast purchase likelihood for the customer–item pair, while the angle between vectors is a measure of item similarity. (R. He et al., 2016) Exploring Visually Similar Items retrieves items similar to a specific user query and is also considered a fashionable item. The fashion similarity is calculated as a K-nearest-neighbours problem. the fashionability is learned from user-item interactions, (Goncalves et al., 2019) proposed an approach to extract style embeddings for use in FRS with a special focus on style information such as textures, prints, material, etc. (Kang et al., 2017) proposed a visually aware approach as an answer for both design and recommendation needs are fashion domain improved a Bayesian personalized ranking (BPR) formulation which employed pre-trained representation. (N. Kato et al., 2019) making patterns and proposed an approach to clothing images generation. (Jannach et al., 2017) recommending the next item in an anonymous Session as session-based recommendations, developing the GRU4REC method with an alternative session-based nearest neighbor method. (J. Wu et al., 2019) Following the STAMP (Short-Term Attention/Memory Priority Model for Session-based Recommendation) model proposed a session-based complementary FR approach to personalized complementary item recommendation in the fashion domain. (Rodríguez et al., 2020) employing the two-stage architecture of neural networks with a clear separation of candidate selection and ranking generation, in session-based recommender field, experimented with a method for re-ranking the most relevant items from the original recommendations, at the aim of improving the similar-item recommendation with using attention network to encode the session information of the user. (J. McAuley et al., 2015) try to understand similarities or complementary relationships between items like a human brain does and understand a complimentary item's suitability. they employed textual and visual attributes of the Amazon data set of users purchased/viewed items and built a general-purpose method. (J. Tang et al., 2018) considering Sequential pattern of behavior (the most recent purchased items) proposed a Convolutional Sequence Embedding Recommendation Model (Caser), Considering each user as a sequence of items have interaction in past and projection for future of the most top-N probable interaction. (Kalantidis et al., 2013) presented a scalable approach to automatically suggest relevant clothing products, given a single image without metadata as a cross-scenario retrieval. (He et al., 2018) proposed FashionNet consists of two components,

a featured network for feature extraction and a matching network for compatibility computation. The outfits with the highest scores are recommended to the users. (S. Keerthi et al., 2018) proposed a two-step deep learning framework that recommends fashion clothes based on the visual similarity style of another image. In (Ganesan et al., 2017), the neural style transfer algorithm is applied to fashion so as to synthesize new custom clothes. they developed an approach to generate new custom clothes based on a user's preference and by learning the user's fashion choices from a limited set of clothes from their closet. (Polanía et al., 2019) the proposed method generates recommendations for complementary apparel items given a query item. A siamese network is used for feature extraction, followed by a fully connected network used to learn a fashion compatibility metric. (A. Veit et al., 2015) used a Siamese CNN architecture to learn feature transformation for compatibility measures between pairs of items. They modeled compatibility based on co-occurrence in largescale user behavior data. In (X. Geng et al., 2015), a deep model which learned a unified feature representation for both users and images was presented. This is done by transforming the heterogeneous user-image networks into homogeneous low-dimensional representations. (M. Vasileva et al., 2018) This approach is based on the idea that an item-level proxy can substitute for outfit compatibility. Instead of considering an outfit as a whole, Vasileva et al. calculated the compatibility between all the pairs of items in an outfit and then averaged the score. A separate space was assigned to each item category pair to compute the compatibility between the items of each category. This approach leads to increases in the algorithm's time complexity (in an outfit containing n items, there are O (n 2) pairs) and loses the relationships among the different pairs in an outfit and the other subsets in an outfit. the approach uses a scoring system on a property (being close in that same space) to force embedding for items to be close in a generally shared space. Instead of just learning the embedding of each item of the dataset in a common shared space, a first embedding space is created by employing the visual features extracted from a CNN and features representing the textual description of the item via a visual-semantic loss; as a second, the authors use a "learned projection which maps the general embedding to a secondary embedding space that scores compatibility between two different item types." The embeddings are then used together with a generalized distance metric to compute compatibility scores between items. (Y. Li et al., 2017) suggested an automatic composition method based on a scoring system for fashion outfit candidates dependent on appearances and meta-data. The scoring module is a multi-modal, multi-instance deep learning system that assesses instance aesthetically and set compatibility. They propose to learn modality embedding and fuse modalities jointly. (A. Veit et al., 2017) in similarity learning, the assumption is commonly made that images are only compared to one unique measure of similarity. mainly because the notions of similarities cannot be captured in a single space. To address this issue, (A. Veit et al., 2017) proposed Conditional Similarity Networks (CSNs) that learn embeddings differentiated into semantically separated subspaces that capture the various notions of similarities. One of the first attempts to incorporate visual content to

develop a unified hybrid recommendation system was Visual Bayesian Personalized Ranking (VBPR) (R. He et al., 2016). GNNs have recently demonstrated impressive performance in graph data modeling using heuristic graph convolutional techniques (Kipf et al., 2017; Z. Zhang et al., 2020; L. Wu et al., 2021). Researchers also suggested constructing a heterogeneous graph of clients, outfits, and items and using hierarchical GNNs to promote customized outfits (X. Li et al., 2020). (Gatys et al., 2015&2016), describes the process of using image representations encoded by multiple layers of VGGNet to separate the content and style of images and recombine them to form new images.

5. Findings

Answering what distinguishes the fashion domain from that of other recommender systems leads to the identification of 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 capable of handling a large corpus of items, a large number of attributes in each item, and a large number of relationships among them, as well as the high-dimensional and semantically complicated features involved. Within answering our first research question, a wide dispersion among literature has been identified on central notions and keywords of FRS that is another reason for extra confusion in understanding FRS domain specifics. We integrate all of these concepts to demonstrate how closely they are interconnected. Finally, through a rational reasoning and deductive process, we illustrate how these notions can be defined as two main principles of application and design; following the principle of design, we show how these two concepts are correlated. We hope this further step helps researchers with providing a consistent, integrated structure for conceptual understanding.

In spite of the fact that literature in this area of research has been scarce and scattered, our effort through conducting a literature review with the aim of identifying and categorizing the main tasks which have been assigned to fashion recommender systems indicates that there are four main tasks which FRSs usually have been developed for answering them including Clothing item-retrieval (similar or identical item recommendation), Complementary Item Recommendation, Whole outfit recommendation, Capsule wardrobe recommendations. The details have been presented in chapter 3.

Although some scholars believe that the retrieval systems cannot be considered a fashion recommender system, the others indirectly reference them. Our review reveals that image retrieval systems have been used as the primary technique in the majority of proposed fashion recommender systems both solely or in combination with other methods; we put it as the first category of tasks defined in fashion recommender systems. Our research also shows that among different image retrieval methods, Content-based image retrieval (CBIR) has mainly been employed in FRS research. According to a fashion image query, cross-scenario image-based fashion retrieval tasks have also got significant attention among many academic communities to retrieve nearly equivalent or the same products from the inventory. Complementary Item Recommendation is the second category that has been a central topic in recent years within the fashion recommendations research area, usually perceived as relaxation of whole outfit recommendation tasks; we introduce the latter as the third main category. It is worth mentioning that the notion of compatibility should be taken into account in addition to similarity in both second and third task definitions. Finally, capsule

wardrobes have been identified as the fourth main task in FRS, however; in some studies, they are respected as a subset of outfit recommender system tasks; we highlighted it as a separated group mainly because the concept of outfit recommendation in this latter group entails a higher level of complexity in terms of algorithmic calculations. The number of outfits can be generated based on a different number of items in an inventory to find the minimum number of things that can provide the maximum set of possible compatible outfits. It also has to be said that some studies work based on wardrobe items, but the concept of a capsule wardrobe has not been considered in most of them.

In the third part, toward answering how image-based fashion recommender systems have been affected by computer vision advancements, we provide a trajectory of the evolvement of computer vision techniques beginning from employing conventional image processing techniques such as SURF, SIFT, etc., to more recent ones utilizing Convolutional neural networks for image representations. It shows there was a big jump in proposing creative concepts like style and then design in this field, mainly since 2012 with employing 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 use deep hybrid models in FR's architectures. Although there are some arguments on limitations of Deep learning methods, our findings show that particularly in image-based FRS, concerning the characteristics of the domain including large dimensionality due to large amounts of attributes, nonlinear nature of the relationship among features, and complex semantics, employing deep learning methods significantly excels using conventional techniques, in some cases using these methods in combination with conventional ones has been recommended.

6. Conclusion

In this Literature review, we have illustrated a big picture on different research approaches towards fashion recommender systems. We introduced the trajectory of studies in fashion recommender systems from the very beginning. The main categories have been defined. We clarified what makes developing fashion recommender systems a necessity for the fashion domain, in this contemporary society, as a competitive advantage leveraging the power of data within employing machine learning methods and AI solutions for different purposes, including marketing, decision making, cross-selling, etc. Representing what makes the fashion domain distinguished from other recommender system domains, we conceptualized the sources of complexity in the fashion domain by illustrating how interconnected these concepts are, as a framework that any fashion recommender system can be defined and understood through it. Focusing on image-based fashion recommender systems, we identified four main tasks in fashion recommender systems, bringing their characteristics to the fore, including cloth-item retrievals, Complementary item recommendation, Outfit recommendation, and Capsule wardrobes. The studies which have been conducted in each category also have been introduced. In addition, we provide the evolvement trajectory of image-based fashion recommender systems, which consists of three main eras, in addition to considerations of the most recent advancements in computer vision and deep learning-based methods. We compared the traditional computer vision techniques versus deep learning. Finally, we categorized the DL-based fashion recommender systems based on employing one single neural network or deep hybrid neural networks with highlighting the methods they used and the input.

7. Implications

Figure 18: The Main aspects of Research implication

Developing a novel fashion recommender system is not doable without obtaining comprehensive insight from this field. Mainly, choosing and designing novel architecture in this field demands domain-specific considerations and understanding of challenges, tasks, and techniques used in the past to be illuminative for researchers to take more insightful steps in the future. In this regard, my great hope is that this thesis was able to create both in-depth and comprehensive insight into fashion recommender systems. To my knowledge, this is the first literature review that has been done in this field, concentrating on image-based fashion recommender systems. The points identified, integrated, and highlighted here should be included in analyzing a fashion recommender system to develop a new one or even replicate those that exist, specifically for image-based modalities that contain the most important features in the fashion domain for visual appearance comprehension. The points identified here provide a big picture of approaches towards fashion recommender and integrate and introduce the main aspects of domain complexities and main tasks to assist fresh researchers in deciding from which perspective they are in to take the first step and design their FRS scenario. For example, any researcher should know the complexities of the fashion domain is about the notion of compatibility in addition to the notion of similarity that other recommender systems are dealing with, and they should consider they are going to design a complementary fashion recommender system or just developing an instance matching system through this thesis. They understand these points and are familiar with different researches that have been performed in each aspect. Choosing the most optimum image processing technique in developing an image-based fashion recommender system is vital; researchers should know how to select their architecture based on the resources are available, or those

should be prepared for a specific scenario, for example, if they are dealing with a small data set or larger ones, there are linear attributes or complex nonlinear relationship among characteristics has to be handled, doing so they should know how to provide a balance between employing traditional techniques in versus most recent ones or sometimes using both of them. For this purpose, they should first find out which techniques exist and what the differences are. Illustrating the path of evolution in image-content-based techniques in this thesis highlights the advent of computer vision approaches mainly improvements after deep learning advancements during recent years. It provides a great perspective for researchers and introduces them to the most prominent researches in each era. Along with this, it is essential to consider that the focus of most architectures that has been developed recently is on proposing deep hybrid models, which sometimes are integrated with hybrid recommender systems. It is worth mentioning that this is the first time this evolutionary path of image-content-based techniques in the fashion recommender domain has been provided.

Going broader answering main questions of this research, through a rational reasoning approach, a complexity notion framework has been introduced to show how main aspects of this complexity are interconnected in fashion recommender systems in addition to introducing the researches which have been worked on these notions with different techniques and approaches. Taking a step further, considering the most recent research done in fashion recommender systems, a classification has been provided based on various combinations of computer vision and deep learning-based techniques employed in two main groups, including single neural building blocks and hybrid ones. I hope this perspective assists both fresh general scholars and experts in being familiar with the distribution of DL-vision-based techniques in research in fashion recommender systems. It has to be added that this is the first time this classification in the fashion recommender systems domain has been done.

8. Future Research

In this section, we suggest some future research directions for fashion recommender systems. Considering the rapid growth of multimedia data, where visual information will be the critical component. More in-depth research in applications of multi-model fusion and multi-task learning in fashion recommender systems are required to model recommender system to be capable of profiling users comprehensively. Besides, while the majority of researches in fashion recommender systems is mainly based on similarity-based retrieval techniques, there is a need for more studies in the development of new functions such as designing clothes, which are highly demanded in future fashion recommender systems. Furthermore, most of the current fashion datasets do not contain outfit compatibility annotations, or they are limited in terms of size and the type of annotations they provide.

Consequently, most researchers built their dataset, which is a labor-costing process, and most of them are not accessible publicly for further research. So, the other future direction for subsequent studies may be focusing on developing automatic annotation methods, constructing large-scale rich annotated data sets for particular task definitions in fashion recommender systems. From an ethical perspective in fashion recommender systems also there is a need for performing the comprehensive study since it has not been studied in almost any of the researches, which have been reviewed through this thesis.

References

Cheng, W. H., Song, S., Chen, C. Y., Hidayati, S. C., & Liu, J. (2020). Fashion meets computer vision: A survey. arXiv preprint arXiv:2003.13988.

Song, S., & Mei, T. (2018). When multimedia meets fashion. IEEE MultiMedia, 25(3), 102-108.

Guan, C., Qin, S., Ling, W., & Ding, G. (2016). Apparel recommendation system evolution: an empirical review. International Journal of Clothing Science and Technology.

Lu, J., Wu, D., Mao, M., Wang, W., & Zhang, G. (2015). Recommender system application developments: a survey. Decision Support Systems, 74, 12-32.

Liu, S., Liu, L., & Yan, S. (2014). Fashion analysis: Current techniques and future directions. IEEE MultiMedia, 21(2), 72-79.

Sutton, R. I., & Staw, B. M. (1995). What theory is not. Administrative science quarterly, 371-384.

Brocke, J. V., Simons, A., Niehaves, B., Niehaves, B., Reimer, K., Plattfaut, R., & Cleven, A. (2009). Reconstructing the giant: On the importance of rigour in documenting the literature search process.

Zorn, T., & Campbell, N. (2006). Improving the writing of literature reviews through a literature integration exercise. Business Communication Quarterly, 69(2), 172-183.

Torraco, R. J. (2005). Writing integrative literature reviews: Guidelines and examples. Human resource development review, 4(3), 356-367.

Salipante, P., Notz, W., & Bigelow, J. (1982). A matrix approach to literature reviews. Research in organizational behavior , 4 , 321-348.

Webster, J., & Watson, R. T. (2002). Analyzing the past to prepare for the future: Writing a literature review. MIS quarterly, xiii-xxiii.

Malone, T. W., & Crowston, K. (1994). The interdisciplinary study of coordination. ACM Computing Surveys (CSUR), 26(1), 87-119.

Pandit, A., Goel, K., Jain, M., & Katre, N. A Review on Clothes Matching and Recommendation Systems based on user Attributes.

Elahi, M., & Qi, L. (2020). Fashion Recommender Systems in Cold Start. In Fashion Recommender Systems (pp. 3-21). Springer, Cham.

Liew, J. S. Y., Kaziunas, E., Liu, J., & Zhuo, S. (2011). Socially-interactive dressing room: an iterative evaluation on interface design. In CHI'11 Extended Abstracts on Human Factors in Computing Systems (pp. 2023-2028).

Goldberg, D., Nichols, D., Oki, B. M., & Terry, D. (1992). Using collaborative filtering to weave an information tapestry. Communications of the ACM, 35(12), 61-70.

Hill, W., Stead, L., Rosenstein, M., & Furnas, G. (1995, May). Recommending and evaluating choices in a virtual community of use. In Proceedings of the SIGCHI conference on Human factors in computing systems (pp. 194-201).

Adomavicius, G., & Tuzhilin, A. (2005). Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. IEEE transactions on knowledge and data engineering, 17(6), 734-749.

Karatzoglou, A., Amatriain, X., Baltrunas, L., & Oliver, N. (2010, September). Multiverse recommendation: n-dimensional tensor factorization for context-aware collaborative filtering. In Proceedings of the fourth ACM conference on Recommender systems (pp. 79-86).

Rendle, S., Freudenthaler, C., & Schmidt-Thieme, L. (2010, April). Factorizing personalized markov chains for next-basket recommendation. In Proceedings of the 19th international conference on World wide web (pp. 811-820).

Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep learning. MIT press.

Quijano-Sánchez, Lara, et al. "Recommend systems for smart cities." Information systems 92 (2020): 101545.

Ricci, F., Rokach, L., & Shapira, B. (2011). Introduction to recommender systems handbook. In Recommender systems handbook (pp. 1-35). Springer, Boston, MA.

Prato, G. (2019). New Methodologies for Fashion Recommender Systems.

Bollacker, K., Díaz-Rodríguez, N., & Li, X. (2016). Beyond clothing ontologies: modeling fashion with subjective influence networks. In KDD workshop on machine learning meets fashion.

Vasileva, M. I., Plummer, B. A., Dusad, K., Rajpal, S., Kumar, R., & Forsyth, D. (2018). Learning type-aware embeddings for fashion compatibility. In Proceedings of the European Conference on Computer Vision (ECCV) (pp. 390-405).

He, R., & McAuley, J. (2016, April). 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 (pp. 507-517).

Wen, Y., Liu, X., & Xu, B. (2018, July). Personalized Clothing Recommendation Based on Knowledge Graph. In 2018 International Conference on Audio, Language and Image Processing (ICALIP) (pp. 1-5). IEEE.

Sullivan, L. (2010). Form follows function. De la tour de bureaux artistiquement.

Zhang, Q., Lu, J., & Jin, Y. (2021). Artificial intelligence in recommender systems. Complex & Intelligent Systems, 7(1), 439-457.

Yu, W., & Qin, Z. (2020, November). Graph convolutional network for recommendation with low-pass collaborative filters. In International Conference on Machine Learning (pp. 10936-10945). PMLR.

Hsiao, W. L., & Grauman, K. (2018). Creating capsule wardrobes from fashion images. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 7161-7170).

Nakamura, T., & Goto, R. (2018). Outfit generation and style extraction via bidirectional lstm and autoencoder. arXiv preprint arXiv:1807.03133.

Liu, S., Feng, J., Song, Z., Zhang, T., Lu, H., Xu, C., & Yan, S. (2012, October). Hi, magic closet, tell me what to wear!. In Proceedings of the 20th ACM international conference on Multimedia (pp. 619-628).

Kalantidis, Y., Kennedy, L., & Li, L. J. (2013, April). 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 (pp. 105-112).

Hadi Kiapour, M., Han, X., Lazebnik, S., Berg, A. C., & Berg, T. L. (2015). Where to buy it: Matching street clothing photos in online shops. In Proceedings of the IEEE international conference on computer vision (pp. 3343-3351).

Vittayakorn, S., Yamaguchi, K., Berg, A. C., & Berg, T. L. (2015, January). Runway to realway: Visual analysis of fashion. In 2015 IEEE Winter Conference on Applications of Computer Vision (pp. 951-958). IEEE.

Liu, Z., Luo, P., Qiu, S., Wang, X., & Tang, X. (2016). Deepfashion: Powering robust clothes recognition and retrieval with rich annotations. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1096-1104).

Kiapour, M. H., Yamaguchi, K., Berg, A. C., & Berg, T. L. (2014, September). Hipster wars: Discovering elements of fashion styles. In European conference on computer vision (pp. 472-488). Springer, Cham.

Simo-Serra, E., & Ishikawa, H. (2016). 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 (pp. 298-307).

Laenen, K., & Moens, M. F. (2020). Attention-based Fusion for Outfit Recommendation. In Fashion Recommender Systems (pp. 69-86). Springer, Cham.

Wong, W. K., Zeng, X. H., Au, W. M. R., Mok, P. Y., & Leung, S. Y. S. (2009). A fashion mix-and-match expert system for fashion retailers using fuzzy screening approach. Expert Systems with Applications, 36(2), 1750-1764.

Veit, Andreas, et al. "Learning visual clothing style with heterogeneous dyadic co-occurrences." Proceedings of the IEEE International Conference on Computer Vision . 2015.

Iwata, T., Watanabe, S., & Sawada, H. (2011, June). Fashion coordinates recommender system using photographs from fashion magazines. In Twenty-Second International Joint Conference on Artificial Intelligence.

Al-Halah, Z., Stiefelhagen, R., & Grauman, K. (2017). Fashion forward: Forecasting visual style in fashion. In Proceedings of the IEEE international conference on computer vision (pp. 388-397).

Ma, Y., Jia, J., Zhou, S., Fu, J., Liu, Y., & Tong, Z. (2017, February). Towards better understanding the clothing fashion styles: A multimodal deep learning approach. In Thirty-First AAAI Conference on Artificial Intelligence.

Fan, W., Qiyang, Z., & Baolin, Y. (2014, October). Refined clothing texture parsing by exploiting the discriminative meanings of sparse codes. In 2014 IEEE International Conference on Image Processing (ICIP) (pp. 5946-5950). IEEE.

Liang, X., Lin, L., Yang, W., Luo, P., Huang, J., & Yan, S. (2016). Clothes co-parsing via joint image segmentation and labeling with application to clothing retrieval. IEEE Transactions on Multimedia, 18(6), 1175-1186.

Jia, Jia, et al. "Learning to appreciate the aesthetic effects of clothing." Proceedings of the AAAI Conference on Artificial Intelligence . Vol. 30. No. 1. 2016.

Yasuda, K., Furuta, H., & Kobayashi, T. (1995, September). Aesthetic design system of structures using neural network and image database. In Proceedings of 3rd International Symposium on Uncertainty Modeling and Analysis and Annual Conference of the North American Fuzzy Information Processing Society (pp. 115-120). IEEE.

Kouge, Y., Murakami, T., Kurosawa, Y., Mera, K., & Takezawa, T. (2015). 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).

Kang, J. Y. M., Johnson, K. K., & Kim, J. (2013). Clothing functions and use of clothing to alter mood. International Journal of Fashion Design, Technology and Education, 6(1), 43-52.

Yang, Y., & Ramanan, D. (2011, June). Articulated pose estimation with flexible mixtures-of-parts. In CVPR 2011 (pp. 1385-1392). IEEE.

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

Li, J., Zhong, X., & Li, Y. (2011). 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 (pp. 57-64). Springer, Dordrecht.

Yamaguchi, K., Hadi Kiapour, M., & Berg, T. L. (2013). Paper doll parsing: Retrieving similar styles to parse clothing items. In Proceedings of the IEEE international conference on computer vision (pp. 3519-3526).

Zeiler Matthew, D., & Rob, F. (2013). Visualizing and understanding convolutional networks. CoRR.—2013.—Vol. abs/1311.2901.—URL: http://arxiv. org/abs/1311.2901.

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., ... & Zheng, X. (2015). Tensorflow: large-scale machine learning on heterogeneous distributed systems (2016). arXiv preprint arXiv:1603.04467, 52.

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., ... & Fei-Fei, L. (2015). Imagenet large scale visual recognition challenge. International journal of computer vision, 115(3), 211-252.

Kang, W. C., Fang, C., Wang, Z., & McAuley, J. (2017, November). Visually-aware fashion recommendation and design with generative image models. In 2017 IEEE International Conference on Data Mining (ICDM) (pp. 207-216). IEEE.

Kato, N., Osone, H., Oomori, K., Ooi, C. W., & Ochiai, Y. (2019, March). Gans-based clothes design: Pattern maker is all you need to design clothing. In Proceedings of the 10th Augmented Human International Conference 2019 (pp. 1-7).

Zhu, S., Urtasun, R., Fidler, S., Lin, D., & Change Loy, C. (2017). Be your own prada: Fashion synthesis with structural coherence. In Proceedings of the IEEE international conference on computer vision (pp. 1680-1688).

Yu, C., Hu, Y., Chen, Y., & Zeng, B. (2019). Personalized fashion design. In Proceedings of the IEEE/CVF International Conference on Computer Vision (pp. 9046-9055).

Sbai, O., Elhoseiny, M., Bordes, A., LeCun, Y., & Couprie, C. (2018). Design: Design inspiration from generative networks. In Proceedings of the European Conference on Computer Vision (ECCV) Workshops (pp. 0-0).

Banerjee, R. H., Rajagopal, A., Jha, N., Patro, A., & Rajan, A. (2018, December). Let AI clothe you: diversified fashion generation. In Asian Conference on Computer Vision (pp. 75-87). Springer, Cham.

Dong, X., Song, X., Feng, F., Jing, P., Xu, X. S., & Nie, L. (2019, October). Personalized capsule wardrobe creation with garment and user modeling. In Proceedings of the 27th ACM International Conference on Multimedia (pp. 302-310).

Raffiee, A. H., & Sollami, M. (2021, January). Garmentgan: Photo-realistic adversarial fashion transfer. In 2020 25th International Conference on Pattern Recognition (ICPR) (pp. 3923-3930). IEEE.

Dubey, A., Bhardwaj, N., Abhinav, K., Kuriakose, S. M., Jain, S., & Arora, V. (2020). AI Assisted Apparel Design. arXiv preprint arXiv:2007.04950.

Reddy, K. S., & Sreedhar, K. (2016). Image retrieval techniques: a survey. International Journal of Electronics and Communication Engineering, 9(1), 19-27.

Jaradat, S., Dokoohaki, N., Corona Pampin, H. J., & Shirvany, R. (2021, September). Workshop on Recommender Systems in Fashion and Retail. In Fifteenth ACM Conference on Recommender Systems (pp. 810-812).

Nandish, C., & Goyani, M. Comparison of different image retrieval techniques in CBIR.

Li, X., Yang, J., & Ma, J. (2021). Recent developments of content-based image retrieval (CBIR). Neurocomputing.

Wang, X., & Zhang, T. (2011, November). Clothes search in consumer photos via color matching and attribute learning. In Proceedings of the 19th ACM international conference on Multimedia (pp. 1353-1356).

Huang, J., Feris, R. S., Chen, Q., & Yan, S. (2015). Cross-domain image retrieval with a dual attribute-aware ranking network. In Proceedings of the IEEE international conference on computer vision (pp. 1062-1070).

Jiang, S., Wu, Y., & Fu, Y. (2016, October). Deep bi-directional cross-triplet embedding for cross-domain clothing retrieval. In Proceedings of the 24th ACM international conference on Multimedia (pp. 52-56).

Kuang, Z., Gao, Y., Li, G., Luo, P., Chen, Y., Lin, L., & Zhang, W. (2019). Fashion retrieval via graph reasoning networks on a similarity pyramid. In Proceedings of the IEEE/CVF International Conference on Computer Vision (pp. 3066-3075).

Zhao, B., Feng, J., Wu, X., & Yan, S. (2017). Memory-augmented attribute manipulation networks for interactive fashion search. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 1520-1528).

Kovashka, A., Parikh, D., & Grauman, K. (2012, June). Whittlesearch: Image search with relative attribute feedback. In 2012 IEEE Conference on Computer Vision and Pattern Recognition (pp. 2973-2980). IEEE.

Guo, Y., Liu, Y., Oerlemans, A., Lao, S., Wu, S., & Lew, M. S. (2016). Deep learning for visual understanding: A review. Neurocomputing, 187, 27-48.

Kato, N., Osone, H., Oomori, K., Ooi, C. W., & Ochiai, Y. (2019, March). Gans-based clothes design: Pattern maker is all you need to design clothing. In Proceedings of the 10th Augmented Human International Conference 2019 (pp. 1-7).

Wang, Z., Gu, Y., Zhang, Y., Zhou, J., & Gu, X. (2017, December). Clothing retrieval with visual attention model. In 2017 IEEE Visual Communications and Image Processing (VCIP) (pp. 1-4). IEEE.

Yuan, Y., Yang, K., & Zhang, C. (2017). Hard-aware deeply cascaded embedding. In Proceedings of the IEEE international conference on computer vision (pp. 814-823).

Gajic, B., & Baldrich, R. (2018). Cross-domain fashion image retrieval. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (pp. 1869-1871).

Y. Zhao, Z. Jin, G.-J. Qi, H. Lu, X.-S. Hua, An adversarial approach to hard triplet generation, in: Proceedings of the European Conference on Computer Vision (ECCV), 2018, pp. 501–517.

M. Shin, S. Park, T. Kim, Semi-supervised feature-level attribute manipulation for fashion image retrieval, CoRR abs/1907.05007. arXiv:1907.05007.

Chopra, A., Sinha, A., Gupta, H., Sarkar, M., Ayush, K., & Krishnamurthy, B. (2019). Powering robust fashion retrieval with information rich feature embeddings. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (pp. 0-0).

Park, S., Shin, M., Ham, S., Choe, S., & Kang, Y. (2019). 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 (pp. 0-0).

Huang, J., Feris, R. S., Chen, Q., & Yan, S. (2015). Cross-domain image retrieval with a dual attribute-aware ranking network. In Proceedings of the IEEE international conference on computer vision (pp. 1062-1070).

Dong, Q., Gong, S., & Zhu, X. (2017, March). Multi-task curriculum transfer deep learning of clothing attributes. In 2017 IEEE Winter Conference on Applications of Computer Vision (WACV) (pp. 520-529). IEEE.

Lu, Y., Kumar, A., Zhai, S., Cheng, Y., Javidi, T., & Feris, R. (2017). 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 (pp. 5334-5343).

Craik, J. (2009). Fashion: the key concepts. Berg Publishers.

Han, X., Wu, Z., Jiang, Y. G., & Davis, L. S. (2017, October). Learning fashion compatibility with bidirectional lstms. In Proceedings of the 25th ACM international conference on Multimedia (pp. 1078-1086).

Li, Y., Cao, L., Zhu, J., & Luo, J. (2017). Mining fashion outfit composition using an end-to-end deep learning approach on set data. IEEE Transactions on Multimedia, 19(8), 1946-1955.

McAuley, J., Targett, C., Shi, Q., & Van Den Hengel, A. (2015, August). Image-based recommendations on styles and substitutes. In Proceedings of the 38th international ACM SIGIR conference on research and development in information retrieval (pp. 43-52).

Hu, Y., Yi, X., & Davis, L. S. (2015, October). Collaborative fashion recommendation: A functional tensor factorization approach. In Proceedings of the 23rd ACM international conference on Multimedia (pp. 129-138).

Donahue, J., Anne Hendricks, L., Guadarrama, S., Rohrbach, M., Venugopalan, S., Saenko, K., & Darrell, T. (2015). Long-term recurrent convolutional networks for visual recognition and description. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 2625-2634).

Chen, L., & He, Y. (2018, April). Dress fashionably: Learn fashion collocation with deep mixed-category metric learning. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 32, No. 1).

Liu, Qiao, et al. "STAMP: short-term attention / memory priority model for session-based recommendation." Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining . 2018.

Yu, Zeping, et al. "Adaptive User Modeling with Long and Short-Term Preferences for Personalized Recommendation." IJCAI . 2019.

Wu, Q., Zhao, P., & Cui, Z. (2020). Visual and textual jointly enhanced interpretable fashion recommendation. IEEE Access, 8, 68736-68746.

Polanía, L. F., & Gupte, S. (2019, September). Learning fashion compatibility across apparel categories for outfit recommendation. In 2019 IEEE International Conference on Image Processing (ICIP) (pp. 4489-4493). IEEE.

Song, X., Feng, F., Han, X., Yang, X., Liu, W., & Nie, L. (2018, June). Neural compatibility modeling with attentive knowledge distillation. In The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval (pp. 5-14).

Harada, F., & Shimakawa, H. (2012). Outfit recommendation with consideration of user policy and preference on layered combination of garments. International Journal of Advanced Computer Science, 2, 49-55.

Tsujita, H., Tsukada, K., Kambara, K., & Siio, I. (2010, May). 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 (pp. 127-132).

Shen, E., Lieberman, H., & Lam, F. (2007, January). What am I gonna wear? Scenario-oriented recommendation. In Proceedings of the 12th international conference on Intelligent user interfaces (pp. 365-368).

Bettaney, E. M., Hardwick, S. R., Zisimopoulos, O., & Chamberlain, B. P. (2020, September). Fashion outfit generation for e-commerce. In Joint European Conference on Machine Learning and Knowledge Discovery in Databases (pp. 339-354). Springer, Cham.

Jiang, Y., Qianqian, X. U., & Cao, X. (2018, September). Outfit Recommendation with Deep Sequence Learning. In 2018 IEEE Fourth International Conference on Multimedia Big Data (BigMM) (pp. 1-5). IEEE.

Wang, J., Ma, Y., Zhang, L., Gao, R. X., & Wu, D. (2018). Deep learning for smart manufacturing: Methods and applications. Journal of manufacturing systems, 48, 144-156.

Graves, A. (2013). Generating sequences with recurrent neural networks. arXiv preprint arXiv:1308.0850.

Sun, G. L., He, J. Y., Wu, X., Zhao, B., & Peng, Q. (2020). Learning fashion compatibility across categories with deep multimodal neural networks. Neurocomputing, 395, 237-246.

Ding, Y., Ma, Y., Wong, W., & Chua, T. S. (2021). Modeling Instant User Intent and Content-level Transition for Sequential Fashion Recommendation. IEEE Transactions on Multimedia.

Sridevi, M., ManikyaArun, N., Sheshikala, M., & Sudarshan, E. (2020, December). Personalized fashion recommender system with image based neural networks. In IOP Conference Series: Materials Science and Engineering (Vol. 981, No. 2, p. 022073). IOP Publishing.

Chen, Y. C., Li, L., Yu, L., El Kholy, A., Ahmed, F., Gan, Z., ... & Liu, J. (2020, August). Uniter: Universal image-text representation learning. In European conference on computer vision (pp. 104-120). Springer, Cham.

Zhao, H., Yu, J., Li, Y., Wang, D., Liu, J., Yang, H., & Wu, F. (2020). Dress like an Internet Celebrity: Fashion Retrieval in Videos. In IJCAI (pp. 1054-1060).

Stefani, M. A., Stefanis, V., & Garofalakis, J. (2019, July). CFRS: A Trends-Driven Collaborative Fashion Recommendation System. In 2019 10th International Conference on Information, Intelligence, Systems and Applications (IISA) (pp. 1-4). IEEE.

Cui, Z., Li, Z., Wu, S., Zhang, X. Y., & Wang, L. (2019, May). Dressing as a whole: Outfit compatibility learning based on node-wise graph neural networks. In The World Wide Web Conference (pp. 307-317).

Ramesh, N., & Moh, T. S. (2018, August). Outfit recommender system. In 2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM) (pp. 903-910). IEEE.

Keerthi Gorripati, S., & Angadi, A. (2018). Visual Based Fashion Clothes Recommendation with Convolutional Neural Networks. International Journal of Information Systems & Management Science, 1(1).

Heger, G. (2016). The capsule closet phenomenon: A phenomenological study of lived experiences withcapsule closets.

Luger, George. "AI: Early history and applications." Artificial Intelligence: Structures and Strategies for Complex Problem Solving. Boston: Addison-Wesley (2005).

Russell, Stuart J., and Peter Norvig. "Artificial intelligence: a modern approach. Malaysia." (2016).

Zhang, S., Yao, L., Sun, A., & Tay, Y. (2019). Deep learning-based recommender system: A survey and new perspectives. ACM Computing Surveys (CSUR), 52(1), 1-38.

O'Mahony, N., Campbell, S., Carvalho, A., Harapanahalli, S., Hernandez, G. V., Krpalkova, L., ... & Walsh, J. (2019, April). Deep learning vs. traditional computer vision. In Science and Information Conference (pp. 128-144). Springer, Cham.

Voulodimos, A., Doulamis, N., Doulamis, A., & Protopapadakis, E. (2018). Deep learning for computer vision: A brief review. Computational intelligence and neuroscience, 2018.

Wiley, V., & Lucas, T. (2018). Computer vision and image processing: a paper review. International Journal of Artificial Intelligence Research, 2(1), 29-36.

Samatha Rani, R., & Laxmi Devi, P. (2020). A Literature Survey on Computer Vision Towards Data Science. IJCRT, 8(6).

Ganesan, A., & Oates, T. (2017). Fashioning with networks: Neural style transfer to design clothes. arXiv preprint arXiv:1707.09899.

Zeiler, M. D., & Fergus, R. (2014, September). Visualizing and understanding convolutional networks. In European conference on computer vision (pp. 818-833). Springer, Cham.

Khan, S., Rahmani, H., Shah, S. A. A., & Bennamoun, M. (2018). A guide to convolutional neural networks for computer vision. Synthesis Lectures on Computer Vision, 8(1), 1-207.

Gkelios, S., Sophokleous, A., Plakias, S., Boutalis, Y., & Chatzichristofis, S. A. (2021). Deep convolutional features for image retrieval. Expert Systems with Applications, 177, 114940.

Marcus, G. (2018). Deep learning: A critical appraisal. arXiv preprint arXiv:1801.00631.

Matzen, K., Bala, K., & Snavely, N. (2017). Streetstyle: Exploring world-wide clothing styles from millions of photos. arXiv preprint arXiv:1706.01869.

Lee, H., Seol, J., & Lee, S. G. (2017). Style2vec: Representation learning for fashion items from style sets. arXiv preprint arXiv:1708.04014.

Al-Halah, Z., Stiefelhagen, R., & Grauman, K. (2017). Fashion forward: Forecasting visual style in fashion. In Proceedings of the IEEE international conference on computer vision (pp. 388-397).

Zhao, B., Feng, J., Wu, X., & Yan, S. (2017). Memory-augmented attribute manipulation networks for interactive fashion search. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 1520-1528).

Mallu, M. (2018). Fashion Object Detection and Pixel-Wise Semantic Segmentation: Crowdsourcing framework for image bounding box detection & Pixel-Wise Segmentation.

Tian, X., Zheng, Q., & Xing, J. (2018, October). Content-Based Image Retrieval System Via Deep Learning Method. In 2018 IEEE 3rd Advanced Information Technology, Electronic and Automation Control Conference (IAEAC) (pp. 1257-1261). IEEE.

Iwata, T., Watanabe, S., & Sawada, H. (2011, June). Fashion coordinates recommender system using photographs from fashion magazines. In Twenty-Second International Joint Conference on Artificial Intelligence.

Bossard, L., Dantone, M., Leistner, C., Wengert, C., Quack, T., & Van Gool, L. (2012, November). Apparel classification with style. In Asian conference on computer vision (pp. 321-335). Springer, Berlin, Heidelberg.

Yamaguchi, K., Kiapour, M. H., Ortiz, L. E., & Berg, T. L. (2012, June). Parsing clothing in fashion photographs. In 2012 IEEE Conference on Computer vision and pattern recognition (pp. 3570-3577). IEEE.

Dong, J., Chen, Q., Huang, Z., Yang, J., & Yan, S. (2015). Parsing based on parselets: A unified deformable mixture model for human parsing. IEEE transactions on pattern analysis and machine intelligence, 38(1), 88-101.

Yamaguchi, K., Hadi Kiapour, M., & Berg, T. L. (2013). Paper doll parsing: Retrieving similar styles to parse clothing items. In Proceedings of the IEEE international conference on computer vision (pp. 3519-3526).

Hu, X., Zhu, W., & Li, Q. (2014). HCRS: A hybrid clothes recommender system based on user ratings and product features. arXiv preprint arXiv:1411.6754.

Limaksornkul, C., Nakorn, D. N., Rakmanee, O., & Viriyasitavat, W. (2014, March). Smart closet: Statistical-based apparel recommendation system. In 2014 Third ICT International Student Project Conference (ICT-ISPC) (pp. 155-158). IEEE.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. Advances in neural information processing systems, 25, 1097-1105.

Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster r-cnn: Towards real-time object detection with region proposal networks. Advances in neural information processing systems, 28, 91-99.

Chen, L. C., Yang, Y., Wang, J., Xu, W., & Yuille, A. L. (2016). Attention to scale: Scale-aware semantic image segmentation. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 3640-3649).

T. Xiao, T. Xia, Y. Yang, C. Huang, and X. Wang. Learning from massive noisy labeled data for image classification. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 2691–2699, 2015.

Bracher, C., Heinz, S., & Vollgraf, R. (2016). Fashion DNA: merging content and sales data for recommendation and article mapping. arXiv preprint arXiv:1609.02489.

Sha, D., Wang, D., Zhou, X., Feng, S., Zhang, Y., & Yu, G. (2016, June). An approach for clothing recommendation based on multiple image attributes. In International conference on web-age information management (pp. 272-285). Springer, Cham.

Wu, L., He, X., Wang, X., Zhang, K., & Wang, M. (2021). A Survey on Neural Recommendation: From Collaborative Filtering to Content and Context Enriched Recommendation. arXiv preprint arXiv:2104.13030.

Wen, J., Li, X., She, J., Park, S., & Cheung, M. (2016, December). 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) (pp. 521-527). IEEE.

Hou, M., Wu, L., Chen, E., Li, Z., Zheng, V. W., & Liu, Q. (2019). Explainable fashion recommendation: A semantic attribute region guided approach. arXiv preprint arXiv:1905.12862.

Alashkar, T., Jiang, S., Wang, S., & Fu, Y. (2017, February). Examples-rules guided deep neural network for makeup recommendation. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 31, No. 1).

Xu, Q., Shen, F., Liu, L., & Shen, H. T. (2018, June). Graphcar: Content-aware multimedia recommendation with graph autoencoder. In The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval (pp. 981-984).

Rawat, Y. S., & Kankanhalli, M. S. (2016, October). ConTagNet: Exploiting user context for image tag recommendation. In Proceedings of the 24th ACM international conference on Multimedia (pp. 1102-1106).

Lei, C., Liu, D., Li, W., Zha, Z. J., & Li, H. (2016). Comparative deep learning of hybrid representations for image recommendations. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 2545-2553).

Liu, Q., Wu, S., & Wang, L. (2017, August). DeepStyle: Learning user preferences for visual recommendation. In Proceedings of the 40th international acm sigir conference on research and development in information retrieval (pp. 841-844).

Chen, J., Zhang, H., He, X., Nie, L., Liu, W., & Chua, T. S. (2017, August). 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 (pp. 335-344).

Cardoso, Â., Daolio, F., & Vargas, S. (2018, July). 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 (pp. 80-89).

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

Ren, Shaoqing, et al. "Object detection networks on convolutional feature maps." IEEE transactions on pattern analysis and machine intelligence 39.7 (2016): 1476-1481.

Yu, Wenhui, et al. "Visually aware recommendation with aesthetic features." The VLDB Journal (2021): 1-19.

Goncalves, D., Liu, L., & Magalhães, A. (2019). How big can style be? Addressing high dimensionality for recommending with style. arXiv preprint arXiv:1908.10642.

Jannach, D., & Ludewig, M. (2017, August). When recurrent neural networks meet the neighborhood for session-based recommendation. In Proceedings of the Eleventh ACM Conference on Recommender Systems (pp. 306-310).

Wu, J. C., Rodríguez, J. A. S., & Pampín, H. J. C. (2019). Session-based complementary fashion recommendations. arXiv preprint arXiv:1908.08327.

Rodríguez, J. A. S., Wu, J. C., & Khandwawala, M. (2020). Two-Stage Session-Based Recommendations with Candidate Rank Embeddings. In Fashion Recommender Systems (pp. 49-66). Springer, Cham.

Tang, J., & Wang, K. (2018, February). Personalized top-n sequential recommendation via convolutional sequence embedding. In Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining (pp. 565-573).

He, T., & Hu, Y. (2018). FashionNet: Personalized outfit recommendation with deep neural network. arXiv preprint arXiv:1810.02443.

Keerthi Gorripati, S., & Angadi, A. (2018). Visual Based Fashion Clothes Recommendation with Convolutional Neural Networks. International Journal of Information Systems & Management Science, 1(1).

Geng, X., Zhang, H., Bian, J., & Chua, T. S. (2015). Learning image and user features for recommendation in social networks. In Proceedings of the IEEE International Conference on Computer Vision (pp. 4274-4282).

Kipf, T. N., & Welling, M. (2016). Semi-supervised classification with graph convolutional networks. arXiv preprint arXiv:1609.02907.

Gatys, L. A., Ecker, A. S., & Bethge, M. (2016). Image style transfer using convolutional neural networks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 2414-2423).

Gatys, L. A., Ecker, A. S., & Bethge, M. (2015). A neural algorithm of artistic style. arXiv preprint arXiv:1508.06576.