Artificial Intelligence Review (2024) 57:67
https://doi.org/10.1007/s10462-024-10703-8

Fashion intelligence in the Metaverse: promise and future prospects

Xiangyu Mu[1] · Haijun Zhang[1] · Jianyang Shi[1] · Jie Hou[1] · Jianghong Ma[1] · Yimin Yang[2]

Accepted: 4 January 2024 / Published online: 20 February 2024

Abstract
With the development of artificial intelligence (AI) and the constraints on offline activities imposed due to the sudden outbreak of the COVID epidemic, the Metaverse has recently attracted significant research attention from both academia and industrial practitioners. Fashion, as an expression of a consumer's aesthetics and personality, has enormous economic potential in both the real world and the Metaverse. In this research, we provide a comprehensive survey of two of the most important components of fashion in the Metaverse: virtual digital humans, and tasks related to fashion items. We survey state-of-the-art articles from 2007 to the present and provide a new taxonomy of extant research topics based on these articles. We also highlight the applications of these topics in the Metaverse from the perspectives of designers and consumers. Finally, we describe possible scenes involving fashion in the Metaverse. The current challenges and open issues related to the fashion industry in the Metaverse are also discussed in order to provide guidance for fashion practitioners, and to shed some light on the future development of fashion AI in the Metaverse.

Keywords Fashion intelligence · Metaverse · Digital virtual humans · Deep learning · Survey

✉ Haijun Zhang
hjzhang@hit.edu.cn

Xiangyu Mu
21B951013@stu.hit.edu.cn

Jianyang Shi
19b951026@stu.hit.edu.cn

Jie Hou
arlo@stu.hit.edu.cn

Jianghong Ma
majianghong@hit.edu.cn

Yimin Yang
yimin.yang@uwo.ca

[1] Department of Computer Science, Harbin Institute of Technology, Shenzhen, Xili University Town, Shenzhen 518055, Guangdong, China

[2] Department of Electrical and Computer Engineering, Western University, 1151 Richmond Street, London, ON N6A 3K7, Canada

1 Introduction

Building a virtual world that is parallel to the real one has been a dream of human beings since ancient times. A science fiction novel written by Neal Stephenson in 1992 (Stephenson 2003) originally described a sun-filled virtual world that offered an alternative to the abysmal real world, and the author referred to this as the Metaverse. Since then, the concept of the Metaverse has continued to appear in films and television works. Over the last decade in particular, the Metaverse has expanded considerably. In addition, many technology companies have also drawn attention to the Metaverse, and some of them has even changed their names to Meta, taken from the first four letters of 'Metaverse'.

Essentially, the Metaverse can be regarded as a virtual world that is both parallel to the real one and interacts with it. It resides in a virtual space that mirrors the natural world, and is independent of the real world. A digital virtual human is an element used by an individual in the real world to move freely in the Metaverse. There is no doubt that clothing plays a vital role in daily life in the real world, as it can implicitly reflect a person's internal characteristics, such as their personality and aesthetics, and social characteristics such as social status and occupation. The dressing of a digital virtual human, which is the mapping of the human user in the real world, will also therefore play an essential role in the Metaverse. A suitable outfit in the Metaverse can not only make a digital virtual human more vivid and concrete, but can also represent the characteristics of the person controlling the digital virtual human.

Fashion artificial intelligence (AI) can affect a wide range of application scenarios in the Metaverse. In fact, the Metaverse is not completely separate from the real world; although it is a world formed in a computer, and is a mapping of the natural world to the virtual world, this virtual world can affect the real one in a straightforward way. In particular, fashion AI can help us carry out certain activities in both the real and the virtual worlds; for example, fashion AI can automatically extract trends from the large amounts of fashion data in the Metaverse, thereby assisting designers to create more data-inspired products. In addition, when consumers are shopping for clothing, they can choose items that suit them more quickly with the help of fashion recommendations. In particular, fashion intelligence can also help us achieve activities that are impossible or difficult to achieve in the real world. For example, compared with real-world fashion data, which are hard to collect, fashion intelligence applications in the Metaverse can utilize these easily accessible online data to make more accurate fashion trend predictions and to help retail companies to characterize market trends. Moreover, fashion editing can help fashion designers to modularize clothing, while fashion generation can simplify the processes used in the clothing industry, from design to ready-to-wear products, as 'what you see is what you get'. Thus, the introduction of the Metaverse means that there is a broader range of application scenarios for fashion intelligence than in the real world.

As an extension of the real world, the Metaverse presents fashion brands with expansive business prospects. Major fashion brands have actively participated in it, impacting the public's horizons with a variety of dazzling and artistic virtual fashion items, and bringing new interests to consumers. In 2019, Amsterdam-based digital fashion company, The Fabricant, launched the world's first digital fashion item, the rainbow dress named Iridescence. Since then, world-renowned fashion brands such as Burberry, Gucci, BVLGARI, etc. have joined the exploration journey of the Metaverse, getting involved in the fields of virtual avatars, virtual clothing and non-fungible token (NFT) creation. Table 1 shows some of the products that fashion brands have launched

Table 1 Fashion brands in the Metaverse

Brand	Country	Metaverse offerings	Keywords
Balmain	France	Unicorn Sneaker NFTs	NFTs
Bulgari	Italy	Bulgari Metaverse, Octo Finissimo Ultra NFTs, and Beyond Wonder Premium Jewellery NFTs	NFTs, virtual space
Burberry	UK	A unicorn, Minny, dressed in Burberry's latest TB Summer Monogram Print Collection	NFTs, virtual doll, apparel
Byredo	Sweden	A perfume for Web3	Web3.0, virtual perfume
Clinique	USA	The first NFT advertisement "Metaverse More Like Us"	NFTs, advertisement
Coach	USA	NFTs Series designed by artists in Amethyst	NFTs
Diesel	Italy	"DIESEL STUDIO" Music NFT Digital Collectibles Project	NFTs
Feng Chen Wang	China	Metaverse fashion show displaying the collection designed by Wang	Virtual fashion show
Gucci	Italy	an experimental space where users will go on a journey through the fashion brand's history through games and NFTs	Virtual space, NFTs
Lacoste	France	Release of 11,212 NFT products	NFTs
LEAF XIA	China	Digital girl LEAF XIA and clothing NFT capsule	NFTs, virtual apparel, virtual character
Lulusmile	China	Metaverse show "PRIMARY Yuanchu" on the Roblox gaming platform, releasing 31 NFT artworks	Virtual apparel, NFTs
PUMA	Germany	Launch of "PUMA and the Land of Games" offering fun sports games and interactive training experiences	Virtual space
TAG Heuer	Switzerland	A new "Connected Calibre E4" smartwatch lets owners display NFTs on the watch's screen	NFTs
Timberland	USA	Release of four virtual shoes in the game Fortnite	Virtual space, virtual apparel
YSL Beauty	France	Collaboration with creative agency Wunderman Thompson to launch YSL Beauty Golden Blocks NFTs Collection	NFTs, virtual space
YES BY YESIR	China	Virtual character "CHUAN" and the first Metaverse fashion show	Virtual character, virtual fashion show

in the Metaverse. These fashion brands' active exploration of the Metaverse not only helps them connect with the exploratory young generation, but also helps them achieve digital transformation and enhance their brand influence.

Our research aims to delve into the latest computer technologies in the fashion Metaverse. Previous reviews of the fashion Metaverse have typically focused on defining the Metaverse and its marketing aspects (Belk et al. 2022; Hadi et al. 2023; Lee and Chen 2011). In contrast, our research focuses on the field of computer science and technology, aiming to provide scholars and computer professionals with an in-depth understanding of computer technology in the fashion Metaverse.

Although there is a large body of literature in the field of fashion intelligence, to the best of our knowledge there has been no systematic investigation of fashion intelligence from the perspective of the Metaverse. To fill this gap, this research aims to provide a comprehensive survey of fashion intelligence in the Metaverse. More specifically, since fashion and people are inseparable, we conduct this survey from two perspectives, and consider digital virtual humans and fashion intelligence technologies. In regard to digital humans, we use body parts as a basis for exploring the standard technologies for generating these avatars, while in regard to fashion intelligence, we summarize the latest developments in the technologies required for fashion intelligence based on the scenes in which designers and customers are located. Finally, we highlight some extant challenges in order to shed some light on future developments in fashion intelligence in the Metaverse.

The remainder of this paper is organized as follows. Section 2 introduces the basic concepts of the Metaverse and the classical tasks associated with fashion AI. In Sect. 3, we present our classification framework and a qualitative analysis of relevant studies. The generation of a digital human is explained in Sect. 4. Section 5 gives an overview of specific methods and techniques used in fashion intelligence, from the perspectives of both designers and customers. We illustrate the challenges faced in the domain of fashion intelligence in the Metaverse in Sect. 6, and conclude the paper in Sect. 7.

2 Terminology and background concepts

2.1 Metaverse

The word 'Metaverse' originates from Neil Stephenson's novel Snow Crash, published in 1992, which described a virtual world parallel to the real world. Each person in the real world had a digital avatar, which was used in the virtual realm of the Metaverse to work, make friends, shop, travel, etc. Currently, there are several definitions of the Metaverse in academia. Mystakidis (2022) views the Metaverse as a persistent multiuser environment based on a fusion of physical reality and digital virtuality. Ning et al. (2021) state that the Metaverse is a multi-technical, social, and super-temporal virtual world that is parallel to the real world. At present, the Metaverse remains at the conceptual stage, and many existing technologies will need to be combined to create this new virtual world and to integrate it with reality. Of these extant technologies, extended reality (XR), digital twins, and the blockchain form the core of the Metaverse.

2.2 Fashion intelligence

As an essential aspect of daily life, fashion can be regarded as a mirror that implicitly reflects people's attitudes. Fashion analysis based on the use of AI has successfully increased the economic benefits of the fashion industry (Nunziatini et al. 2022). Fashion intelligence focuses on the application of AI to the fashion industry; in particular, computer vision technologies such as object detection, object analysis, image retrieval, image generation, etc., are leveraged to improve the efficiency of practitioners in the fashion industry and to enhance the consumer's shopping experience. Hence, determining how best to transform fashion data into relevant computer vision tasks and design-specific models appears to be critical for these real-world applications. In practice, fashion tasks can be roughly divided into three categories: low-level pixel-based fashion computing, which can be used for fashion parsing and landmark detection; a mid-level fashion analysis, which aims at identifying fashion items from images and can be used for fashion detection and fashion attribute prediction; and a high-level understanding of fashion, which involves an overall analysis of the attributes of fashion items at the image level and explores the relationships between fashion items for the tasks such as fashion retrieval, compatibility learning and garment recommendations.

3 Classification scheme and analysis

In this section, we classify current research on fashion intelligence in the Metaverse and conduct an overall analysis of these studies.

3.1 Classification scheme

In the concept of fashion, people and fashion items are inseparable. Fashion items add a unique charm to people, while people become the perfect showcase for fashionable items. Even the most finely crafted fashion items can be affected in their beauty if they are not worn by the right people. Therefore, we surveyed the use of fashion intelligence in the Metaverse, from the generation of digital virtual humans to fashion intelligence technology. The generation of digital virtual humans in the Metaverse focuses on efficiently generating realistic or user-friendly 3D human body models. An avatar can express the user's emotions by wearing fashionable clothing in the Metaverse. Fashion intelligence technology concentrates on analyzing and understanding fashion items, and on facilitating the production and sale of fashion items. For clarity, Fig. 1 provides a classification of fashion intelligence in the Metaverse.

3.1.1 Generation of digital virtual humans

The goal of the Metaverse is to provide users with an immersive virtual world, and the realistic 3D modeling of humans is essential to achieve this goal. Much research has been conducted on the generation of digital virtual humans, most of which has been devoted to automatically generating realistic 3D virtual human models, with the objective of reducing the dependence of digital virtual human modeling on expert manual modeling. In prior

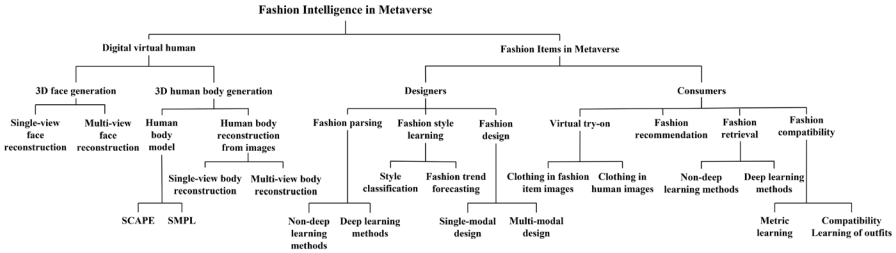

Fig. 1 Research topics associated with fashion in the Metaverse and a taxonomy of these techniques

studies of the human body, the task of generating a digital virtual human has been divided into two subtasks: 3D face generation, and 3D human body generation. Extant techniques related to these subtasks will be elaborated in the following context.

3D face generation 3D face generation is a popular research direction in the field of virtual human generation. In particular, researchers focus on reconstructing 3D faces from 2D images, which can mitigate the constraints on space and equipment required for tasks involved in 3D face reconstruction.

3D human body generation 3D human body generation is an important subtask of 3D object reconstruction. The reconstruction of a 3D human body involves integrating the features of the body into 2D images and deforming a parametric general model of the body to generate a more refined model. The poses of models can be freely changed to match the poses in images.

3.1.2 Fashion intelligence

The use of AI has brought considerable convenience to the fashion industry (Anantrasirichai and Bull 2022; Hosseinnia and Ebrahimi 2022). Increasing numbers of researchers have drawn attention to the mining and analysis of fashion data to achieve an in-depth understanding of fashion elements. In practice, however, designers and customers, the two main groups of fashion practitioners, have different needs in terms of fashion intelligence tasks. Designers prefer technology that can provide them with tools to assist their design processes, while customers expect fashion intelligence to provide them with a better shopping experience. Recently, many techniques have been constantly explored in order to meet or stimulate the needs of both designers and consumers. As a result, the fashion industry is experiencing a strong boom driven by the use of these techniques.

Designers Inspiration is very demanding aspect of the design process. As a result, helping designers to gain inspiration quickly and simplifying the design process are essential. Some fashion intelligence technologies can inspire designers and facilitate the design process; for example, fashion parsing can help designers to extract the regions in which fashion items are located, while fashion style learning can help designers to understand the styles of fashion items and to transfer them to other fashion items, and fashion generation can help designers to generate new fashion items based on styles or sketches.

Customers Convenience and thoughtfulness are the service principles of many famous fashion retailers, which have made them popular with customers. Certain fashion intelligence tasks can provide customers with a more convenient and thoughtful shopping experience. For example, a virtual try-on facility can allow customers to see

the effect of a clothing item without actually trying it on, while a fashion recommendation system can suggest suitable clothes based on the customer's body shape, preferences, and characteristics, and fashion compatibility learning can score the outfits selected by customers using machine learning algorithms.

3.2 Statistical analysis

In this subsection, we conduct a basic statistical analysis of the numbers of publications, the year of publication, and existing experimental datasets in the domain of fashion intelligence. We used a public database called DBLP as our search engine, as it contains most of the studies in this field. The main keywords that were used as input to the search engine were "Metaverse", "fashion", and "digital human". We restricted our attention to works published in high-quality journals and conference proceedings, such as TPAMI, CVPR, ECCV, NeurIPS, etc. Figure 2 shows the distribution of papers published via these outlets, and it can be observed that the top two publication venues were ACM'MM and CVPR. Figure 3 shows the number of yearly publications from 2008 to the present, and an explosive increase in the numbers of publications on the Metaverse can be seen from 2021. In addition, due to the successful application of deep learning (Dong et al. 2022; Wang and Wang 2020; Abbas et al. 2019; Pratama and Wang 2019; Zhou et al. 2023), research on fashion and digital humans has proliferated since 2016.

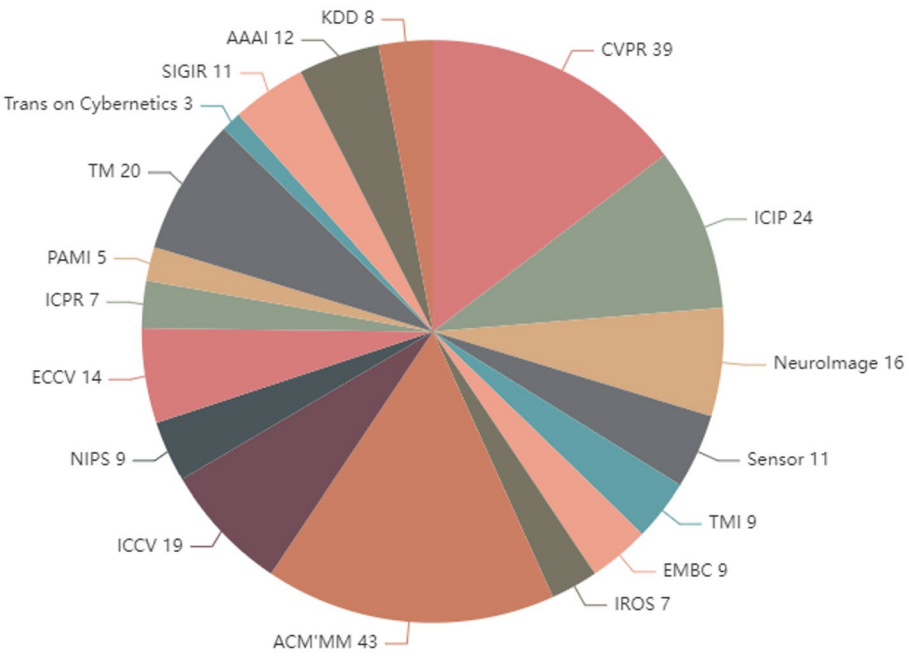

Fig. 2 Outlets publishing articles on fashion in the Metaverse

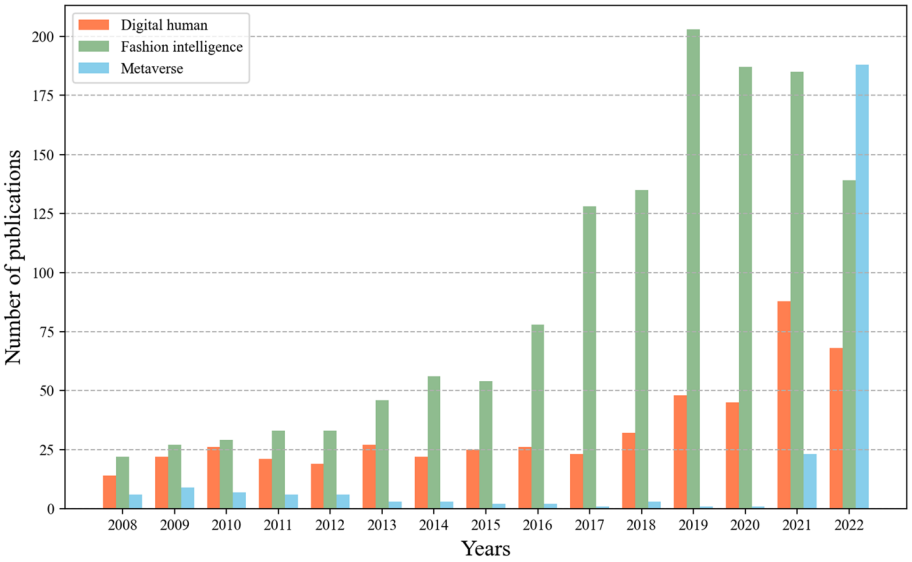

Fig. 3 Number of publications in each year on fashion in the Metaverse

4 Digital virtual humans

As one of the fundamental components in the Metaverse, a digital virtual human is a representation of a digital identity that allows a player to interact others or with computer agents. A digital avatar is a representation of the user's identity, and is the virtual entity that the user is in contact with for the longest in the virtual world. It is therefore natural for users to want to choose the appearance of their digital avatar according to their preferences. In addition, digital virtual humans can wear fashion items in the Metaverse, such as clothes, earrings, bracelets, etc. This section introduces current research on digital virtual human generation technology and the development of this field from two perspectives: 3D face generation, and 3D human body generation.

4.1 3D face generation

A realistic face can reduce the user's sense of disobedience and facilitate a more immersive Metaverse experience. The purpose of 3D face generation is to generate a realistic 3D model of a human face that can be driven by audio, face transformations, etc. The task of 3D face reconstruction involves recreating the detailed features extracted from 2D images in the form of a 3D model, especially in terms of shapes, textures, etc. More specifically, 3D face reconstruction can be separated into two primary types based on the number of 2D photos used for feature extraction: single-view reconstruction and multi-view reconstruction. Single-view reconstruction is usually more difficult than the multi-view process, due to the limitations of a 2D image structure. It is frequently the case that a single image cannot provide all the feature information required for face reconstruction, resulting in the need to predict attributes to achieve the effect of a 3D model. Currently, 3D morphable

models (3DMMs) (Blanz and Vetter 1999) serve as the foundation for 3D face reconstruction. A 3DMM is a parametric face model that can generate virtually any face based on a fixed number of points. Faces can be matched one-to-one in 3D space by linearly adding several orthonormal basis weights. The aim of both single-view and multi-view face reconstruction is to obtain fitting parameters for a 3DMM to obtain realistic faces in 3D space.

4.1.1 Single-view 3D face reconstruction

Over the past decade, there has been a great deal of research on 3DMM single-image fitting. The texture, color, and other features of the face image need to be preserved in the fitted face model as far as possible, and the model must be accurately aligned with the facial contours of the target image. Most traditional methods regard face reconstruction as an optimization problem, where the 3DMM is used to synthesize images based on the unique features of face images, such as facial landmarks, edges, pixel colors, etc. In particular, Choi et al. (2010) proposed a framework that automatically estimated all 3D scene parameters from single- or multi-view images. Kemelmacher-Shlizerman and Seitz (2011) presented a single-image face reconstruction model based on a computation of facial similarity. In addition, global human face similarity and face pose estimation were exploited to overcome the significant differences in shape between the input and reference subjects. Furthermore, Romdhani and Vetter (2005) suggested a multi-feature fitting algorithm to improve the convergence properties of conventional face reconstruction models. However, due to the diversity of face poses and the complexity of image backgrounds, conventional optimization methods are sensitive to initial conditions and parameter changes, making the process of single-image face reconstruction relatively fragile for practical applications. Recent developments in deep learning have allowed many researchers to present new ideas for parameter optimization problems. To address the problem with conventional methods whereby they cannot capture nonlinear expressions to create complex expressions, Ranjan et al. (2018) introduced a convolutional mesh autoencoder (CoMA) that could learn nonlinear representations of human faces. Richardson et al. (2017) presented an end-to-end convolutional neural network (CNN) framework for generating faces in a coarse-to-fine manner. To solve the depth estimation problem in facial reconstruction, Lee et al. (2022) proposed a displacement map generation network (DPMMNet) that generated a displacement map to estimate a detailed geometry.

In addition, several methods have been utilized to recover the 3D information of the face from a single image. Image processing methods such as shape from shading (SFS), UV map, thin plate spline (TPS), and epipolar plane image (EPI) have been applied to single-image face reconstruction. For example, Jin et al. (2022) first introduced 3DMM to reconstruct a smooth face shape and employed landmark-conducted Laplace deformation to fine-tune this shape. An SFS optimization process was then designed to recover the multi-scale geometric details. A position map regression network (PRN) (Feng et al. 2018b) was developed to achieve 3D facial structure reconstruction and dense face alignment. A UV map recording the spatial position of each pixel was fed into a lightweight encoder-decoder for reconstruction of the 3D model. Bhagavatula et al. (2017) proposed a new method of 3D face reconstruction that combined feature extraction with the TPS warping function. EPI is an method of estimating scene depth based on the differences between the image pixels at points in the camera plane and the image plane. Feng et al. (2018a) presented a model-free approach to reconstructing the 3D face model. Their method was trained with a densely connected CNN architecture called FaceLFnet, based on the horizontal and vertical EPIs

of light field images. The authors reported that this method was robust to changes in pose, facial expression, and lighting in face reconstruction tasks.

4.1.2 Multi-view 3D face reconstruction

Unlike single-view face reconstruction methods, multi-view methods do not require strong inductive biases to accomplish model deformation. These approaches can extract facial features from multiple viewpoints in different images to create more detailed 3D models. The efficient fusing of features from multiple views is the key to achieving accurate depth estimation and facial texture recovery. Multi-view methods have attracted considerable attention due to their powerful, fine-grained modeling capability. However, most 3D face reconstruction methods using multi-view face images still rely on generic 3D face models. For example, Wang et al. (2010) proposed a 3DMM-based multi-view face reconstruction method that employed multi-view geometric constraints to eliminate ambiguity from images. Subsequently, an adaptive photometric stereo-based reconstruction method was presented in Roth et al. (2017). Wu et al. (2019a) designed an end-to-end trainable CNN network to set 3DMM parameters. Several image processing methods have also been employed for the task of multi-view face reconstruction. In particular, Li et al. (2022) used an implicit representation to encode the extensive geometric features of faces, which could improve the generalization performance and quality of 3D face reconstruction. It is very likely that view-based 3D face reconstruction methods will have a multitude of applications related to the Metaverse; for example, these methods can greatly lower the thresholds to the large-scale face reconstruction of users and can reduce the computational overhead in the Metaverse.

4.2 Human body generation

A digital virtual human is an indispensable digital identity for each user of the Metaverse. All activities in the virtual world, such as communication, picking up items, etc., must be handled via these digital identities. A beautiful and unique digital avatar, which is manually designed and has a rich level of detail, is welcomed by many users of the Metaverse. However, creating a digital avatar manually for each user is not practical, as this would require a lot of time and effort. The purpose of 3D human body generation is to automatically generate realistic human 3D models, which can reduce the cost of the Metaverse. This section gives an overview of current research on 3D human body generation from two perspectives: generic 3D human body models and human body construction from images. A human body model focuses on representing human bodies in 3D space, whereas human body reconstruction is dedicated to generating similar 3D models from 2D images.

4.2.1 Human body models

Modeling the human body has always been challenging for practitioners in both academia and industry. In the past, creating detailed human models required professional artists to generate models manually or the use of 3D scanning to capture the geometry and texture features of the body. However, these methods are time-consuming, require a high level of expertise of the artists, and are sensitive to site conditions. Fortunately, the human body has standard features in terms of shape and pose, which allow researchers to build a parametric 3D body model based on an analysis of high-quality data representing human features.

This parametric model can create a detailed 3D human body model base on only a few body features, as well as significantly improving the efficiency of body modeling. There exist two commonly used parametric human body models: shape completion and animation for people (SCAPE) (Anguelov et al. 2005), and the skinned multi-person linear model (SMPL) (Loper et al. 2015). Both approaches represent the human body through a set of triangular surfaces $\{f_1, f_2, ..., f_n\}$, where the vertices of each triangle f_i are $\{v_{i,1}, v_{i,2}, v_{i,3}\}$. SCAPE (Anguelov et al. 2005) is a unified parametric 3D human body model that combines body shape and pose information to achieve a human representation. Drawing on Sumner's idea of deformation transfer (Sumner and Popovic 2004), SCAPE employs a 3×3 matrix to represent the deformation of each triangle as a discrete differential gradient field, which can be used to transfer deformation from one model to another. The introduction of SCAPE is regarded as a milestone in the development of 3D human body modeling, and many studies have been devoted to improving the performance of this method. For example, Hasler et al. (2009) designed a model called invariant-SCAPE to solve the problem whereby the triangle deformation in the original SCAPE uses different encodings for the same shape. Hirshberg et al. (2012) proposed an optimized BlendSCAPE model that made the joints of the digital body smoother. In addition, Mebatsion et al. (2012) proposed a simplified SCAPE model (s-SCAPE) to improve the speed of body modeling.

The deformation in SCAPE (Anguelov et al. 2005) depends on the rotational deformation of a triangle patch, which means that human models are unable to be used directly in popular animation software. SMPL (Loper et al. 2015) was proposed to solve this problem. In a similar way to SCAPE, SMPL (Loper et al. 2015) employs pose and shape to model the human body. It uses 10-dimensional values to describe the shape of the body. The parameters can be obtained by principal component analysis (PCA) based on the deformation. To calculate the pose representation, SMPL uses a kinematic tree to represent the 24 joint points of the body. Many studies have been devoted to improving the performance of SMPL. For example, SMPLify (Bogo et al. 2016) is a CNN two-dimensional human pose estimation model in which the SMPL parameters (including body shape and pose parameters) were optimized by minimizing the mean vertex-to-vertex Euclidean error between the synthesized 3D pose and the detected 2D joint points. However, this method does not constrain the shape of the body, and the algorithm easily falls into local optimal solutions, causing reconstruction failure. Based on the SMPLify model (Bogo et al. 2016; Lassner et al. 2017) added more human joint points (91 points) and obtained accurate pose reconstruction results. Corona et al. (2021) proposed a differentiable model for the reconstruction of the body and clothing.

A parametric human body model can be regarded as an essential 3D human body reconstruction technique, in which the aim is to use corresponding parameters as input to construct a precise 3D model of the shape and posture of the human body. SCAPE and SMPL, the two most well-known parametric body models, were developed by leveraging human body datasets to learn the characteristics of the human body shape. Fitting dense 3D point cloud data or depth data of the body to the parameters of a parametric model through point cloud registration, template deformation, etc., is a standard method of reconstructing the human body in fine detail.

4.2.2 Human body reconstruction from images

Human body generation based on 3D scanners requires specialized capturing systems with strict environmental constraints (e.g., large numbers of sensors and controlled lighting) that

are very expensive and cumbersome to deploy. Due to its convenience, image-based 3D human body reconstruction has attracted the attention of many researchers over the last decade. Based on the number of perspectives used for feature extraction, it can be divided into single-view and multi-view methods. Single-view human reconstruction is less restricted by the environment than multi-view approaches, and the corresponding accuracy of the reconstructed 3D model is often lower. In a similar way to 3D face reconstruction, 3D human body reconstruction also requires strong prior models as support. Hence, general body modeling methods such as SCAPE (Anguelov et al. 2005) and SMPL (Loper et al. 2015) are widely used for 3D reconstruction.

Statistical body shape models, as a powerful human prior, allow for convenient disentanglement of pose and shape. Fitting the pose and shape of statistical body shape models to a body in a 2D image is an essential aspect of model-based single-view human body reconstruction. In traditional methods, the prediction of human body model parameters is transformed into a model parameter optimization problem. Initially, annotated 2D landmarks and silhouettes were employed (Guan et al. 2009) as image features to optimize the parameters of the SCAPE model, with promising results. Lassner et al. (2017) used auxiliary landmarks on the body surface and added an estimated silhouette to make the model more accurate. Bogo et al. (2016) annotated keypoints in 2D images and aligned them with keypoints in 3D models to obtain better results. However, optimization problems rely heavily on the initialization effect of the solution, and are prone to local minima. Hence, many researchers have performed pose and parameter regression through network training by mapping the extracted image features to a low-dimensional parameter space. The basic framework used for 3D human body reconstruction for network regression is shown in Fig. 4. Effective feature extraction is critical to ensure an accurate 3D result. Feature extraction strategies such as landmark detection, keypoint detection, body silhouette detection, and semantic segmentation have often been used to improve the fitness of a model.

The parametric general body model can keep the prediction space small in the reconstruction of the human body. However, it leads to the inability of the body model to model the human body in clothing. Therefore, non-parametric methods such as hulls (Wang et al. 2019), point clouds (Qi et al. 2017), triangular meshes (Lin et al. 2019), and voxel grids (Tahir et al. 2021) were used for 3D human body reconstruction. They can predict shape representations directly from images. Natsume et al. (2019) implicitly represented the shape of the human body through the contours and joints of the body pose and then fed the frontal image and its mask into a generative adversarial network (GAN) to infer the texture of the human body to model the clothed human body. Moreover, Krajnik et al.

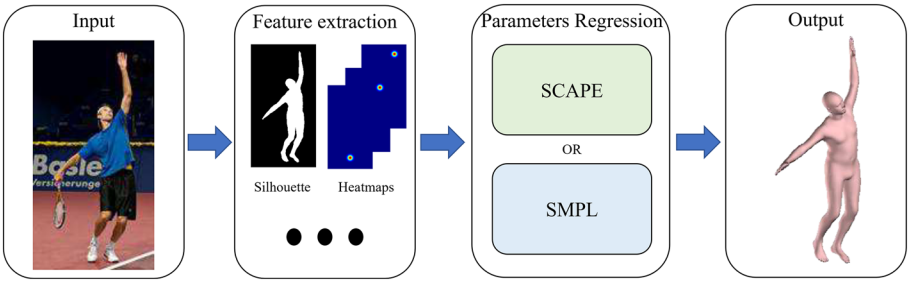

Fig. 4 Pipeline for single-view human body reconstruction

(2022) proposed a novel method to reconstruct each part of the human body independently. It appeared to have smaller errors than other methods, especially in the concave area of the human body.

Unlike single-view body reconstruction, multi-view reconstruction can describe body features from multiple views, which can reduce the error in the prediction of unobservable body parts. Traditional methods use image consistency and depth estimation to establish the correspondence of joints and other feature points between images with different views; however, these methods are easily affected by occlusion. The task of 3D human body reconstruction therefore employs a depth map of 2D images and fuses them to create a united mesh for 3D body generation. With the help of the multi-view calibration capability of deep learning, many of these approaches have overcome the limitations of traditional methods. For instance, Liang et al. (2020) used an image encoder to extract image features and passed these features through multiple regression blocks to predict human body parameters in a stage-by-stage and view-by-view process. Pix2Vox (Xie et al. 2019) involved the use of a decoding encoder to generate corresponding 3D bodies for humans in each view. Saito et al. (2019) designed an end-to-end network to digitize a clothed human body, using a network that employed a pixel alignment implicit function (PIFu) to locally align the pixels in the 2D image with the corresponding context in the 3D body. In addition, Yu et al. (2022) proposed a coarse-to-fine linear learning model that utilized graph convolutional networks to deform templates to the ground-truth mesh.

4.3 Other research between digital human and fashion

In addition to 3D face generation and human body generation, there is a growing interest in research focusing on fashion generation tasks leveraging digital human features. These tasks aim to enhance the appearance and styling of digital humans through the manipulation of various human body details. Notably, makeup transfer and hairstyle generation have emerged as prominent research directions within the realm of digital human fashion-related generation tasks.

4.3.1 Makeup transfer

Facial makeup serves as an effective means to enhance the beauty of an individual's face by adding intricate details to various facial areas. The process of makeup transfer entails seamlessly applying a desired makeup style onto a human face. Preserving key features, including the shape of eyebrows, the size of the mouth, the blush color, and the eye rim hue, presents a significant challenge in facial makeup transfer. Traditional methods (Tong et al. 2007; Guo and Sim 2009; Xu et al. 2013) for makeup transfer primarily rely on comparing before-and-after face images captured under similar lighting conditions and poses. These traditional approaches necessitate intricate image processing techniques, including face alignment, layer decomposition, and appearance correction. Moreover, they primarily leveraged low-level image features and imposed stringent requirements concerning training data and application scenarios.

With the widespread application of deep learning in computer vision, successful advancements have been made in the domain of makeup transfer tasks. Early on, researchers employed CycleGAN (Zhu et al. 2017a) to implement facial makeup style transfer by treating non-makeup face images and makeup face images as content and style images, respectively. However, CycleGAN, which primarily focuses on global

features, encounters difficulties in achieving satisfactory migration due to the significant variations across different regions of the face. To address this limitation, makeup transfer can be enhanced by targeting different facial areas individually. PairedCycleGAN (Chang et al. 2018) and BeautyGAN (Li et al. 2018) incorporate additional loss functions such as identity loss, makeup loss, histogram loss, and style loss to regulate the intricate details of non-makeup faces. Kips et al. (2020) proposed the utilization of a background consistency loss and a color discriminator to mitigate issues arising from changes in skin tone during makeup transfer with reference face images. Deng et al. (2021) tackled spatial misalignment between the input face and reference face by learning face identity features with fusing features from the eyes, skin, and lips regions. Zhang et al. (2019) decomposed facial image features into personal identity features and makeup style features, and achieved local controllable makeup transfer by editing style codes. Furthermore, Sun et al. (2020b) designed four encoders specifically to extract personal identity information, lip makeup style, eye makeup style, and face makeup style, respectively.

4.3.2 Hairstyle learning

The hairstyle of an individual is among the foremost attributes that catch people's attention. A well-designed and customized hairstyle can make a person stand out from the crowd and make a great impression. Hairstyles serve as a means to express one's personal style, taste, and distinct personality traits. Nevertheless, exploring the realm of hairstyles proves to be a challenging endeavor due to the extensive range of available styles, intricate hair textures, and the appearance of hairstyles changing with postures. Consequently, exploring and comprehending hairstyles is a relatively difficult task. Liu et al. (2014a) introduced a groundbreaking approach to hairstyle recommendation that incorporates makeup as a vital factor. Their proposed model utilizes image features, aesthetics, and relevant attributes to determine the most suitable hairstyle that complements the makeup style. To accomplish this, they employed a multi-tree hypergraph model that effectively identifies and selects the hairstyle exhibiting the highest degree of compatibility with the given makeup. Hairstyle-GAN (H-GAN) (Yin et al. 2017) was proposed to edit hairstyles in person images. H-GAN consists of three parts: encoder decoding subnetwork, GAN and recognition subnetwork, and the recognition subnetwork and discriminator of GAN share the same network structure. In response to the persistent challenge of locally generated ambiguities within hairstyles, Gee et al. (2022) introduced a framework for hairstyle generation. This comprehensive approach comprises two fundamental modules: the segmentation module and the generation module. The face segmentation module plays a pivotal role in detecting and extracting both hairstyles and facial features accurately. Leveraging this crucial information, the hairstyle generation module employs an advanced transformer-based GAN to generate high-quality hairstyle images. Notably, this approach not only focuses on generating visually appealing hairstyles but also endeavors to restore intricate details simultaneously. HairstyleNet (Song et al. 2023) was proposed as an interactive hairstyle editing network that allows users to manipulate local or entire hairstyles by adjusting parameterized hair regions. The network encodes roughly hair parameters, face and background images into a latent representation in the hairstyle generation stage, and then generates high-fidelity face images with ideal new hairstyles from the latent codes.

5 Fashion items in the Metaverse

As the two most important roles in the real-world fashion industry, designers and consumers play a vital role in the fashion community of the Metaverse. By creating new fashion items, designers can increase the diversity of the fashion community. As users of fashion items, consumers inject vitality into the fashion community through their evaluations and feedback on fashion items. In this section, we give an overview of certain exciting fashion scenarios in the Metaverse from the perspectives of both designers and consumers. The common methods used in these fashion scenes are summarized and classified, with particular reference to the most representative and novel methods in this field.

5.1 Fashion intelligence for designer innovation

The main objective of fashion designers in the Metaverse is the same as in the real world: to create consumer-preferred fashion products. A system that can facilitate fashion tasks in the Metaverse is crucial in terms of helping fashion designers to design satisfactory products more quickly and efficiently. Metaverse Fashion Intelligence plays a pivotal role in fostering designers' innovation, helping them to conceive exceptional works of fashion. In the following, we describe how carrying out fashion tasks in the Metaverse can help designers.

5.1.1 Fashion parsing

When a designer wants to create a fashion item, browsing existing items of the same type can help in finding inspiration. However, searching for a particular type of fashion item in a multimedia database is difficult for designers. In the Metaverse, fashion parsing can help designers achieve this efficiently.

Fashion parsing involves segmenting fashion items from images containing multiple such items by labeling each pixel in an image. Fashion parsing is a prerequisite for many fashion tasks, as it can identify the individual fashion items in an image for subsequent processing. Due to the diversity of clothing types, fashion parsing is more challenging than general semantic parsing. In addition, the non-rigid characteristics and the deformed structure of clothing on the body in a given image make it necessary to add semantic information to both the clothing and the human body in order to perform high-level judgments in the task of fashion parsing. In general, fashion parsing methods can be divided into two categories: non-deep learning methods, based on traditional techniques, and deep learning methods, which rely on a fully connected network (FCN)-based image segmentation pipeline. In non-deep learning methods, specific prior rules for label inference are added to traditional semantic segmentation models for fashion parsing. In contrast, deep learning methods rely on the robust feature extraction ability of a neural network to fuse information such as the texture, edge, and shape of the clothing, which are used to enhance the performance of the clothing parsing model.

Clothing parsing tasks have been explored for a long time by researchers focusing on clothing recognition in only a few scenarios (Hasan and Hogg 2010) or sketch recognition for clothing design (Chen et al. 2006). However, these works (Hasan and Hogg 2010; Chen et al. 2006) are limited to only a few applications, and the results are usually unsatisfactory in practice. Yamaguchi et al. (2012) put forward an innovative idea for fashion parsing, in which they used superpixels to simplify the task of fashion parsing and combined human

feature estimation to parse clothing. However, their approach requires pixel-level labels in order to carry out model training, which imposes enormous costs in terms of time and manual labor. To address this problem, Liu et al. (2014b) employed multiple well-trained classifiers to parse fashion items from a given image. Drawing on the idea underlying the scheme in Yamaguchi et al. (2012), Dong et al. (2016) proposed Parselet for human pose estimation and used conditional random fields (CRFs) to perform clothing analysis in the unary and pairwise potential. In order to solve the problem in which the performance of a parser is typically limited by the training data, Liu et al. (2015) proposed a fashion parsing algorithm that could be trained on fashion videos. In a later study, Zhao et al. (2016) proposed a clothing co-segmentation (CCS) algorithm to automatically segment and extract clothing regions from given images with natural backgrounds. Although the styles of clothing are ever-changing, most clothing of the same type has similar characteristics, leading to the possibility of parsing garments based on data-driven techniques. In particular, Yamaguchi et al. (2013) proposed a data-driven fashion parsing method that essentially transferred pixel predictions from samples retrieved in response to a query.

Unlike traditional methods, which require prior knowledge in the form of manual segmentation for preprocessing, deep learning methods rely on receptive fields of various sizes in the network to extract the contextual information on the human body and clothing items in an image. Following the developments in deep learning technology, the successful use of FCNs for general semantic segmentation tasks has attracted the attention of researchers working on fashion parsing. Some researchers have performed clothing parsing by adding subsequent processing steps, such as CRF and additional discriminators, to the FCN architecture (Zhang et al. 2022). One group of researchers have focused on building an end-to-end fashion parsing framework by incorporating CRF into parsing neural networks (Fan et al. 2022). Fashion parsing methods based on deep learning generally adopt a dual-path network architecture, as shown in Fig. 5. In this structure, one path employs an FCN to extract the fine-grained content features from images, while the other employs auxiliary modules to enhance the annotation segmentation pipeline. These auxiliary modules improve the accuracy of clothing parsing by extracting the unique semantic information of fashion items. These modules include texture feature maps (Khurana et al. 2018), outfit encoders (Tangseng et al. 2017), edge-preserving modules (Fan et al. 2022), pyramidal aggregation-excitation context modules (Fan et al. 2022), and other network flows.

Fashion parsing is one of the most fundamental problems in fashion computing, as numerous high-level fashion tasks such as virtual try-ons, fashion retrieval, etc. are performed based on the output of fashion parsing. The issue of how to improve the efficiency of fashion parsing while maintaining accuracy is therefore the goal of many researchers. In

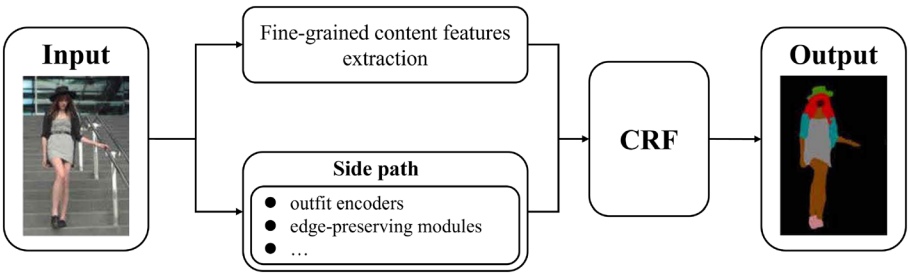

Fig. 5 Basic pipeline for fashion parsing

addition, expanding the categories of items that can be parsed is also an exciting topic in this field.

5.1.2 Fashion style learning

Style is an overall semantic attribute of a fashion item, and is jointly determined by low-level attributes such as color, texture and shape. People who wear different styles of clothing show different temperaments. Style is also an essential factor for designers to consider in their designs. Fashion style learning allows the fashion-assisted design systems in the Metaverse to understand the characteristics of fashion styles in a similar way to humans. In general, fashion style learning can not only help designers to classify fashion styles, but can also predict fashion trends.

Style can be regarded as a semantic description of a fashion item. The classification of fashion styles remains challenging, as items with different fabrics, colors and shapes may belong to the same fashion style. Early studies (Kiapour et al. 2014) used body detection and descriptions for fashion style classification. With the help of deep learning, it is now possible to directly use images of people as input for the task of style classification. Takagi et al. (2017) created a fashion style dataset containing 13,126 images that were classified into 14 categories. They demonstrated the feasibility of fashion style classification through the direct use of a generic classification network. A joint classification and ranking network for weakly labeled data was proposed for style classification in Simo-Serra and Ishikawa (2016), in which global feature extraction was performed on images to measure the similarity between the anchor image and both similar and dissimilar images, and feedback was passed to the classification network for style classification. Identifying clothing style based on local semantic features means that style classification is sensitive to the appearance of clothing items. To address this issue, Yue et al. (2021) developed design issue graphs (DIGs) to provide global and semantic descriptions of clothing styles. As shown in Fig. 6, the semantic representations of fashion style were formed with DIGs and the global features were extracted based on clothes images. However, the precise definition of fashion style remains an ongoing research problem. Although extant style classification datasets

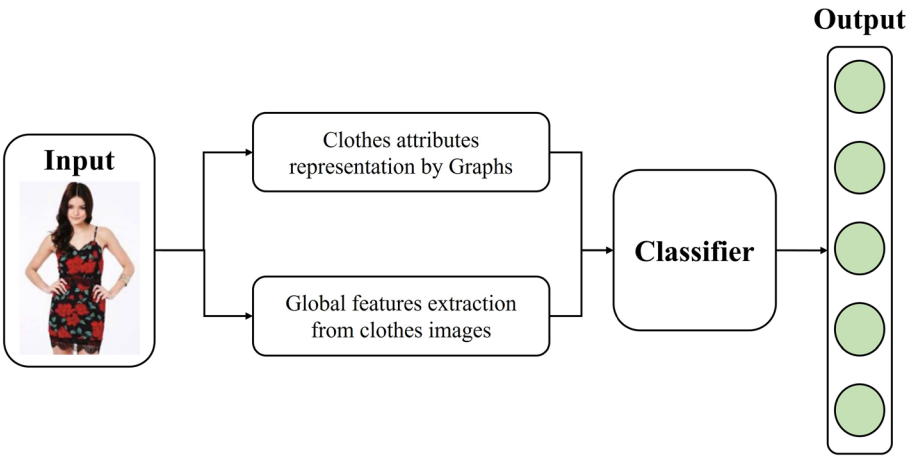

Fig. 6 Fashion style classification method proposed in Yue et al. (2021)

already contain many style categories based on the knowledge of fashion experts, they still cannot cover all styles due to the rapid changes in fashion trends. Furthermore, since no commonly accepted classification criteria for fashion styles have been developed by fashion experts, the same look may be classified into several different styles. Hence, multi-label prediction of styles is also an important direction for future research on style classification.

The prediction of fashion trends is another important application of fashion style representation. The aim in this case is to capture the visual style features of clothing and then to combine historical cross-domain data containing time series to predict future trends. Al-Halah et al. (2017) were the first to propose a fashion style prediction system based on consumer purchase records and images. Later, Zhao et al. (2021) designed a system called NeoFashion to predict trends for fashion designers. In similar research, Gabale and Subramanian (2018) predicted social media trends in India with an improved object detection model. Jin et al. (2021) proposed an end-to-end LSTM encoding-decoding framework for the prediction of clothing trends in various price ranges. Fashion trend forecasting can be seen as a subtask of temporal forecasting. Unlike in ordinary temporal prediction tasks, non-temporal features such as customers' opinions, celebrity outfits, and popular social events may cause sudden changes in fashion trends. Determining how to represent celebrity effects and unexpected events in forecasting is a topic worthy of further discussion in the area of fashion trend forecasting.

5.1.3 Fashion design

In traditional fashion design, designers must spend a great deal of time on carefully selecting colors, fabrics, and textures in order to draw a clothing tile image. Fortunately, computer-aided drawing tools can assist designers in creating clothing templates, which can greatly reduce the workload of designers. However, clothing design requires a wealth of professional knowledge in practice. The Metaverse may lower this barrier to fashion design with the help of AI. It may be that designers and users will be able to specify a few constraints on products in the Metaverse environment, and the system will then instantly generate sketch samples that meet their expectations. Designers will then be able to add further details to these samples to produce a richly textured digital garment. Such a system could greatly improve the working efficiency of designers, and could enable users to create personalized products based on their preferences. Depending on the type of input, fashion design can be divided into single-modal and multi-modal processes.

The aim of single-modal fashion design is to transfer visual elements (such as colors, textures, etc.) from one fashion item to another. However, there are several difficulties with this approach at the transfer stage. First, single-modal fashion designs require high-resolution textural details, and low-resolution fashion items cannot clearly illustrate the effects of style transfer. Second, a fashion design system needs to capture the boundaries of a texture filling accurately. In addition, some parts of fashion items do not need texture padding, such as buttons, zippers, etc. With the help of the controllable generation features of a GAN, many researchers have generated refined and user-controllable fashion items. The loss functions commonly used in single-modal fashion design include the feature loss, style loss $\mathbb{L}_s$, pixel loss $\mathbb{L}_p$, classification loss $\mathbb{L}_c$, texture loss $\mathbb{L}_{tex}$, color loss $\mathbb{L}_{col}$, etc. These losses constrain the images generated by the GAN in terms of style, pixels, texture, etc., and ensure that the generated images do not deviate too far from expectations. TextureGan (Xian et al. 2018) was the first method to allow the user to control the synthesis of fashion items from sketches and textures. Figure 7 showed it reproduced ground-truth handbag,

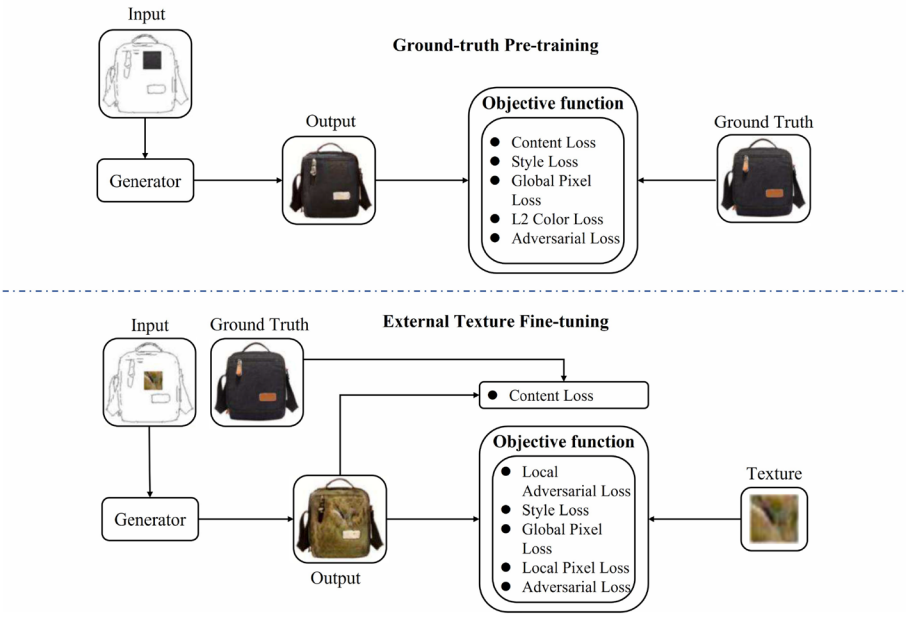

Fig. 7 The pipeline of TextureGan (Xian et al. 2018)

clothes images and supported a broader range of textures by fine-turning the network with the texture database.

A later system called FashionGAN (Cui et al. 2018) was designed based on an end-to-end virtual clothing generation network, in which simple textures and corresponding design sketches were utilized to achieve intelligent clothing design. A fashion generation framework called StyleGan was created by Sbai et al. (2018), which was designed to generate realistic virtual clothing without input. Using a different approach, Jiang et al. (2022) synthesized clothing images by blending them with textures of other items while preserving the global content of the clothing. Recently, Yan et al. (2022a, 2022b, 2022c) focused on the disentanglement of visual attributes, such as the textures and shapes of fashion images, in order to assist designers in accomplishing the task of fashion design. Current research in the field of single-modal fashion design focuses on the refinement, migration, and filling of visual features such as color, texture, shape, etc. Mapping fashion images to a latent space and transferring the mapping matrix can generate new fashion images with similar textures and colors. However, this method cannot edit a single attribute of a fashion item, such as its color or texture. Decoupling the visual attributes of fashion items remains a challenging topic in the area of single-modal fashion design.

Multimodal fashion design combines fashion images with other types of information, such as textual descriptions of fashion items, to generate corresponding fashion images. For example, Zhu et al. (2017b) focused on replacing the clothing of a person with a garment described in the form of text. Their method was implemented in two stages: in the first stage, human parsing was used to generate a reasonable human segmentation map, to maintain the shape of the body and the coherence of the text used to describe the human body, while in the second, a generator was tasked with generating clothing images based on the segmentation map and text descriptions. Zhang et al. (2020a) introduced three attention

layers to the second stage of the network proposed in Zhu et al. (2017b) to obtain more refined clothing details. The generation of clothing images directly from text descriptions is also a research focus in the domain of multimodal design. In particular, an enhanced attentional GAN (e-AttnGAN) (Ak et al. 2020) was proposed to accomplish the task of text-to-image generation. Another system called M6-UFC (Zhang et al. 2021) uniformly leveraged multiple multimodal information to generate new images. The two main research paths in multimodal fashion design involve accurately establishing the mapping relationship between fashion features and text in different spaces and effectively integrating multimodal features, as these can help models to generate more refined fashion items and improve the overall consistency of the generated images.

5.2 Fashion intelligence for enhanced consumer experience

Consumers in the fashion community are expected to have a completely different shopping experience in the Metaverse than in the real world, due to the greater creativity of the Metaverse. The time and distance restrictions of traditional shopping are eliminated, and consumers can shop for impressive clothing at any time, and from anywhere. In the following, we review extant techniques that can be used in shopping scenarios in the Metaverse.

5.2.1 Virtual try-on

If a consumer finds a model in the Metaverse wearing a very attractive outfit, or the clothes in a store catch their eye, it is natural for them to wish to buy such clothing. Trying on clothes directly is an intuitive way for the customer to judge whether clothes suit them. Unlike in the real world, where clothes must be tried on in an offline shop, consumers can wear their favorite clothes at any time, and anywhere, in the Metaverse. Users can freely change their clothes in real time by selecting the clothes they want to try on, and virtual try-on technology is laying the groundwork for these exciting scenarios. The purpose of a virtual try-on is to check the appearance of the target clothing on the user without taking off the clothes that are currently being worn. A virtual try-on can be viewed as a special image-generation task in which images of a model wearing the target outfit are created, under limited circumstances. More specifically, a virtual try-on usually takes two images as input: one is a given model image m_t, which contains the given human body p_0 and the clothing c_0, and the other is a target clothing image c_t. The output of a virtual try-on system is an image m_g, in which the human body p_t is shown wearing the target clothes c_t and the body shape and pose of the model in the input image are preserved. Semantic information about the clothing and models is also fed into the system as one kind of supervision information. A basic virtual try-on framework is illustrated in Fig. 8.

In order to simplify the problem, the backgrounds of the clothing and human body images are usually clear. In a fashion shop, image pairs, i.e., a model wearing the clothes in the target image, are easy to obtain. However, triple image pairs in which the same model has same pose but is wearing different clothes are difficult to collect. This means that each model with different and pixel-wise aligned clothes is usually infeasible. The problem of using unpaired images can be handled in two ingenious ways, as shown in Fig. 9. Many researchers regard the virtual try-on task as an image repairing problem; they first mask the region of the body with the clothes that they want to change, to cover the semantic information of the clothing, and the masked image can then be repaired using the clothing item worn by the model for network training. However, since each person is only matched

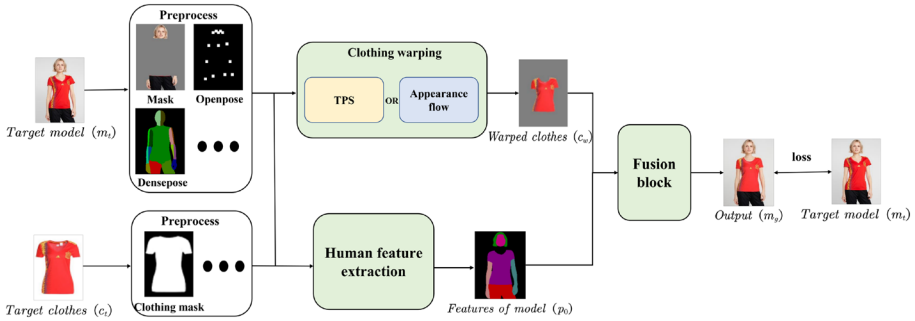

Fig. 8 General framework for virtual try-on models

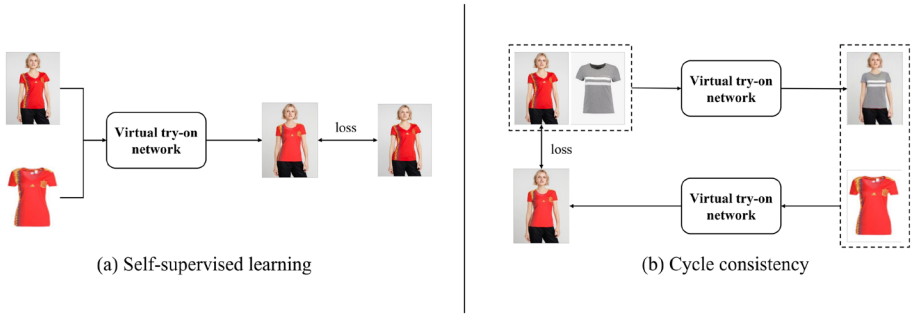

Fig. 9 Two pipelines for virtual try-on systems

to one clothing item during the image reconstruction process, the performance of a virtual try-on model is usually limited, due to the generalization problem. When the target clothing and the clothing on the model have significantly different visual appearances, the virtual try-on system tends to be ineffective. In addition, a cycle-consistent approach is used to train an end-to-end virtual try-on network (Ge et al. 2021a; Kips et al. 2020). The training process of this approach is shown in the right of Fig. 9. The clothes in the input image are replaced with the target clothes, and the clothes in the output image are also replaced with the original clothes in the input image. Several virtual try-on methods have employed the cycle-consistent approach for model training. Notably, Conditional Analogy Generative Adversarial Network (CAGAN) (Kips et al. 2020) is the first method to introduce the cycle-consistent in the virtual try-on task to solve the problem that paired images are difficult to obtain. CAGAN exhibits the capability to implicitly acquire segmentation masks, eliminating the need for costly and time-consuming supervised labeled data. CAGAN also demonstrates its ability to directly predict images that possess the desired attributes, streamlining the process and bypassing the requirement for explicit segmentation information. Furthermore, Ge et al. (2021a) introduced the Disentangled Cycle-Consistency Try-On Network (DCTON) as an improvement upon CAGAN. DCTON generates highly realistic try-on images by effectively distinguishing clothing regions from non-clothing regions. Nevertheless, it is still challenging to simultaneously generate the shape and texture of the clothes, human skin, and non-clothing contents using a cycle GAN. Based on the type of

the target clothing image, the virtual try-on task can be divided into two categories: the target clothing in a fashion item image, and the target clothing in a human body image. In the first category, in the same way as in a traditional virtual try-on task, the system replaces the region of the clothing on the human body with the target clothing from an image that contains only a single fashion item with a clean background. Replacing the region of clothing in an input image through the target clothes on the human body is another category of virtual try-on tasks.

In practice, transforming a real garment from a shop into a photo-realistic garment fitted to a reference image of a person is an important subtask of a virtual try-on. In response to this issue, many researchers have focused on generating natural, realistic transferred garments and retaining more fine texture. They have usually warped the input clothing to align it with the image of the customer using two general methods: a geometric transformation and a warping module. Geometric transformation exploits spatial information to make the deformed clothes more realistic. TPS (Belongie et al. 2002) is a general method of geometric transformation for garment warping. It has been proven to be an effective coordinate transformation model in many computer vision tasks, such as object recognition, virtual try-ons, etc., and is a basic function used to map the representing coordinates. The clothing from an in-shop image is then geometrically transformed to produce the warped clothing image by TPS. VITON (Han et al. 2018) was the first system to exploit TPS for a virtual try-on task, and deformed in-shop clothing to warped clothes with a composition mask. A neural network was also used to learn the transformation parameters of TPS in CP-VITON (Wang et al. 2018). Later, Fenocchi et al. (2022) introduced self- and cross-attention operations to the warping module. They aligned the refined representation of a person and an in-shop garment using two-branch cross-modal attention blocks. In a virtual try-on framework, a generator is typically employed to synthesize the final results, in which a model wears the target garment. The U-net architecture (Ronneberger et al. 2015) is the most widely used type of generator for this task, as it directly shares the features between different layers. However, the basic U-net architecture (Ronneberger et al. 2015) is limited to blurred texture and loss of detail in the generated image of the person. To address these problems, several refinement strategies have been adopted to improve the quality of the final results. For example, realistic details from the deformed clothing have been exploited by a network to render blurred regions (Han et al. 2018). In the same vein, Ge et al. (2021b) used warped clothes, human pose estimation, and reserved regions on the human body as input. They combined Res-UNet with residual connections to preserve the details of the deformed clothes and to generate realistic fitting results.

A virtual try-on task can also replace a target garment on a person with the target model. This task focuses on transferring the clothes worn in the original image C_o onto arbitrary model images m_a, rather than requiring clean product images. However, this task gives rise to different challenges compared to inputting a target garment from a fashion item image with a clear background. For example, identifying and extracting regions of clothing in the input model image m_a becomes essential for a natural result. Due to the differences between the person in the original image C_o and the model image m_a, the problem of aligning the poses of the two bodies is also challenging. In addition, the seamless synthesis between the desired clothing in C_o and the model in the target image m_a is also a factor affecting the success of the virtual try-on task. In view of these issues, researchers have attempted to handle arbitrary poses, clothing extraction, and other challenging problems by developing frameworks with multiple components. For example, Wu et al. (2019b) proposed the M2E-Try-on network to transfer clothes from an original image to an arbitrary person. Since the clothing in the input image contains the pose information of the original person, the pose

alignment module is a critical component in which the pose of the model is aligned to that of the input person. Similarly to the pose transfer module, the pose alignment component aims to modify the viewpoint and the pose of the human in an image. Dense pose conditioning (Güler et al. 2018) and human body segmentation (Raj et al. 2018) are often used to generate pose images in the task of pose transfer. Moreover, a body fitting module (Wu et al. 2019b) and a texture module (Raj et al. 2018) are widely used to facilitate the task of garment transfer learning.

The focus of most research on target clothing in fashion item images and human body images involves warping clothing and splicing it with an image of a human body. However, most current research studies have considered the VITON dataset, which only contains a single type of clothing, and the performance of these systems on a wide variety of garment types is still unpredictable. In addition, the issue of how to alleviate the dependence of virtual dressing tasks on preprocessing, such as fashion segmentation and pose estimation, is also a topic worthy of further research.

5.2.2 Fashion recommendation

Shopping in the Metaverse can overcome space constraints, and can allow customers to enjoy an immersive shopping experience at any time and from anywhere. Since a fashion store in the Metaverse can offer countless fashion products, it represents a paradise for fashionistas who enjoy shopping. However, customers who have difficulty in choosing or who have less time for shopping will have trouble in selecting suitable products when faced with so many, even if these items can be displayed based on their attributes through a fashion retrieval process. To address this issue, fashion recommendation techniques can be adopted to alleviate the burden of choosing products for customers. This type of system can actively recommend suitable products for customers, acting as a shopping guide during their shopping process. Due to the real-time interaction between the customer and the system in the Metaverse, the shopping experience can be greatly improved.

As a specific type of a more general recommendation system (Batmaz et al. 2019), fashion recommender systems have attracted considerable attention from academic researchers and industrial practitioners. The aim in this case is to automatically select clothing that will meet the consumer's preferences or match the customer's needs according to their personal information, dressing scenes, and other information. Compared with a general recommendation system, the task of fashion recommendation has the characteristics of both visual priority and local priority. This means that traditional, general recommendation methods may not be ideal for carrying out fashion recommendation tasks in a straightforward way.

A fashion item with good design has a strong visual expression that is recognized by customers. A recommendation system that considers both the appearance of a product and the user's consumption habits can deliver suggestions that match the customer's preferences. Determining how to represent the visual features of products and how to add them to the recommendation system as essential reference factors are critical aspects of the fashion recommendation task. A CNN framework is a typical means of extracting the visual features of items. As they have excellent feature extraction ability, deep CNNs such as ResNet (He et al. 2016), Caffe (Jia et al. 2014), etc. are widely used to extract the visual features of items at a high level. He and McAuley (2016) first introduced the visual appearance of items into a preference predictor. A network was fitted with an additional layer that could extract the relevant visual features and latent dimensions to provide recommendations.

In the field of fashion recommendation, the style of a fashion item also has a significant impact on customer preference. Liu et al. (2017) proposed a method called DeepStyle (Liu et al. 2017) to characterize user preferences by learning the style features of items. As shown in Fig. 10, the style features was obtained by subtracting item category features from visual features. Style features were then integrated into the widely used BPR (Rendle et al. 2012) framework to generate fashion recommendations.

There is a large body of literature in the domain of fashion recommendation on how to recommend appropriate fashion items to create outfits with existing clothing. This problem can be summarized as a compatibility estimation task, which will be introduced in the later context. However, scenario-oriented and explainable fashion recommendations, among others, are also indispensable aspects of the task of fashion recommendation. Scenario-oriented fashion recommendation recommends suitable outfits for a user based on certain events that the user needs to attend. Liu et al. (2012a) devised a "magic closet" system that suggested the best matching outfits for a special occasion. Zhang et al. (2017) designed a clothing recommendation system that was able to recommend clothing for travelers based on the relevance of the clothing and the destination. When customers receive recommendations from a system, they may also want to know why these clothes were recommended for them. The task of explainable fashion recommendation is more complicated, as it usually involves multiple forms of domain knowledge such as user attributes, regional culture, computer vision, etc. Chen et al. (2019) used an attention model to learn the regions that attracted the customers' attention. They claimed that this method could visually illustrate the reasons for recommending the garment by highlighting the key regions of an image. Using another approach, Lin et al. (2020) explained system-based clothing recommendations by analyzing customer reviews. Tangseng and Okatani (2020) proposed a method of quantifying the impacts of different attributes of clothing. They represented the garment in an image by interpretable features of humans and providing the reason to pick the clothing by the most influential item features. Zhou et al. (2022a, 2022b) introduced outfit generation frameworks to automatically synthesize compatible fashion items when

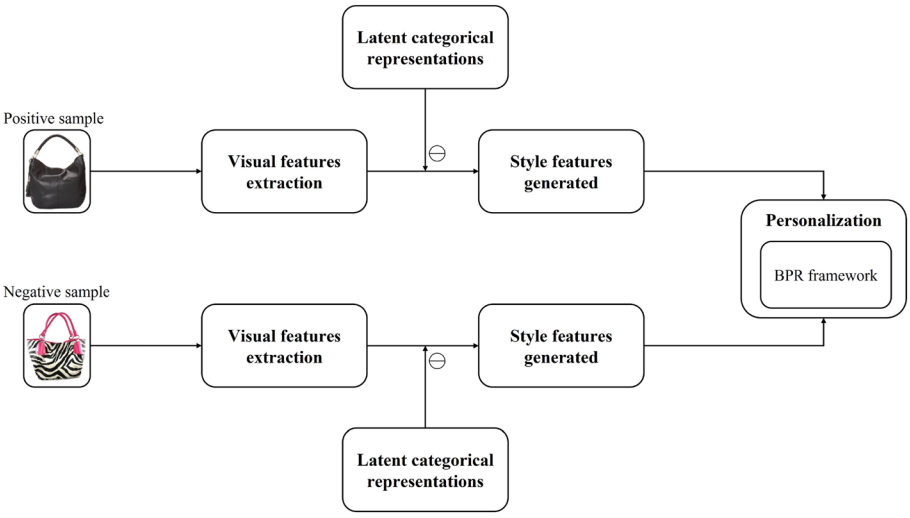

Fig. 10 The pipeline of DeepStyle (Liu et al. 2017)

given an extant item. Collecting the highly correlated factors affecting the customer's purchase intention and adding them to the recommendation network is a critical step in fashion recommendation. However, to create a recommendation model that performs well on the market, the visual similarity must not be considered alone, and the regional culture, the personal attributes of target customers, and social networks are all factors that need to be taken into account in real-life applications.

5.2.3 Fashion retrieval

In the real world, customers may find it tiresome to select the clothes they want from a store which is full of merchandise. However, customers may not encounter this irritating shopping experience in a Metaverse store. When faced with a range of countless products, customers can quickly filter the products based on their attributes at any time, to allow them to pick out suitable products. To address this issue, the task of fashion retrieval involves methods of quickly and accurately searching for a specified item from a massive dataset.

The aim of fashion retrieval is to return accurate and relevant fashion products in response to a query by a customer, thus increasing the convenience of purchasing fashion products. A retrieval system usually retrieves data from the dataset that are similar to the query item based on a comparison of visual similarity. Depending on the scenario in which the query object and the returned object are located, this process can be divided into intra-scenario and cross-scenario fashion retrieval. Intra-scenario image retrieval searches for similar fashion items from the dataset whose images have the same scenario as the query images. In contrast, in the cross-scenario fashion retrieval task, the scenario of the query fashion images is often different from that of the returned fashion images. For instance, users can search for similar fashion items photographed in daily life from an online shopping image dataset or for similar fashion items in online retail fashion images from street photographs. A valid representation of an item is essential for fashion retrieval. Both non-deep learning methods and deep learning methods are effective means of representing the visual features of fashion items.

Non-deep learning fashion retrieval methods can be implemented in two stages. The first stage involves locating and segmenting the region containing a query garment in an image. In the second stage, artificially constructed visual feature representations of segmented garments are captured to enable an image search. Liu et al. (2012b) first proposed a solution to the issue of cross-scene fashion retrieval. An occasion-oriented fashion retrieval approach was also proposed, in which the low-level visual features of clothing and high-level occasion category features were fused with mid-level clothing attributes. A feature representation that was able to characterize the clothing appearance well, using a pose-dependent approach, was used for fashion retrieval (Vittayakorn et al. 2015). These feature representation schemes can facilitate the quantitative analysis of cross-domain clothing image similarity.

Thanks to their powerful feature extraction capabilities, deep learning methods have become the most common solution to the problem of fashion retrieval. A deep network is used to model the similarity of the garments, which is used to determine whether the clothes in two images are the same based on a set of designed rules. The basic pipeline for these deep learning methods is illustrated in Fig. 11.

In the intra-scenario fashion retrieval, the similarity can be calculated without an intermediate image between the query clothes and the candidate clothes, as they reside in the same scenario. As shown in Fig. 11 (Input (a)), the input can be represented by the paired

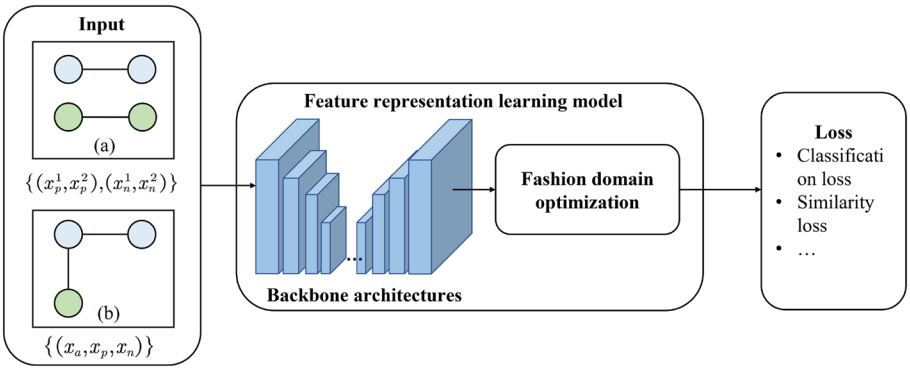

Fig. 11 Basic pipeline for fashion retrieval

data sample $\left\{\left(x_p^1, x_p^2\right), \left(x_n^1, x_n^2\right)\right\}$, where x_p^1 and x_p^2 are two positive samples representing the same or similar fashion items, and $\left(x_n^1, x_n^2\right)$ is a negative data pair representing unmatched fashion items in two images. The similarity learning of binary sample pairs can be regarded as a binary classification task, in which the aim is to minimize the distance between positive sample pairs and maximize the distance between negative sample pairs simultaneously. In view of this, Kiapour et al. (2015) employed a pre-trained CNN with ImageNet to extract feature representations from a query bounding box and the clothing region in shop images. Kinli et al. (2019) proposed densely-connected capsule networks to search for in-shop clothing. Zhao et al. (2022) proposed an anchor-free framework for joint clothing detection and search. The framework grasps more information about the mask area of clothes by predicting the mask of the clothes and extracting the embedding features of the clothes, and improved the retrieval ability of the framework by a matching loss. Gao et al. (2022) utilized the global mask to obtain more comprehensive clothing features rather than being limited to the center point. SEAM Match-RCNN (Godi et al. 2022) connects store images with video sequences to remove the reliance on annotated bounding boxes in previous fashion retrieval datasets.

The difference between cross-scenario and intra-scenario fashion retrieval lies in the scenario of the query and candidate images. In practice, it is challenging to handle the discrepancies between fashion items in different scenarios. One commonly used strategy is the use of domain adaptation techniques, in which triple embeddings are adopted to bridge the discrepancies between domains. As shown in Fig. 11 (Input (b)), a triple data pair $\left\{\left(x_a, x_p, x_n\right)\right\}$ is fed into a deep network to map the samples onto a space. A triplet sample consists of an anchor x_a, a positive sample x_p, and a negative sample x_n. Samples with matching labels are regarded as positive pairs, and those with mismatched labels as negative pairs.

Using this approach, Huang et al. (2015) proposed a dual attribute-aware ranking network (DARN), which consisted of two sub-networks for feature learning. A sub-network was designed for each domain, and semantic attribute learning was exploited for feature representations. The two sub-networks were connected by feeding the features extracted from each into a triplet loss function.

Fashion retrieval can help customers to select coordinating fashion items from a massive dataset of fashion items in the Metaverse based on the attributes and characteristics of the items. Identifying deformed fashion items and mapping cross-domain attributes are

currently research hotspots in the field of fashion retrieval. Exploring the controllability provided by attribute disentanglement and retrieval of unlabeled fashion items is also a worthwhile avenue for future work.

5.2.4 Fashion compatibility

Many people like to ask friends to accompany them when shopping, to help them evaluate the clothes they choose and provide suggestions. However, this may not be possible for someone who has difficulty in finding companions for shopping. Fortunately, there is no such barrier to users when shopping in the Metaverse. The shopping guide offered by a Metaverse store can actively score the clothes chosen by a user in real time, and provide suggestions when a user has difficulty in selecting a match.

The aim of a fashion compatibility system is to estimate how well different types of fashion items match. Learning the compatibility between fashion items forms the basis for many advanced fashion tasks, and represents a challenging task in itself. In practice, it is undesirable to calculate fashion compatibility based solely on visual similarity, as the shapes of different fashion items may be quite different, and the visual properties of two harmonious fashion items, such as their colors and textures, are not necessarily the same. Researchers have developed several models in which harmonic matching is inferred in the fashion domain to enable compatibility learning. Compatibility semantics are usually modeled and characterized based on the deep features of fashion items. Mainstream methods embed fashion items into the underlying representation of the fashion domain through different embedding strategies, and use this underlying representation as the basis for compatibility calculations. Due to their powerful feature extraction ability, deep learning methods of fashion compatibility can map fashion items into deep fashion space, and can learn compatibility based on distance metrics in the mapping space. Hence, the problem of fashion compatibility can be viewed as a specific type of metric learning in which the compatibility of fashion items is determined by computing the independence of their vectors. A further focus for research involves compatibility learning for outfits composed of multiple garments.

The goal of metric learning is to learn a measure of the similarity between two items. In this approach, a pair of items is treated as two feature points x and y in the deep learning space, and a distance function $d(x, y)$ is employed to measure the distance between them. A fashion compatibility system leverages metric learning to learn an embedding space in which the distances between compatible items (positive) are closer than those for non-compatible items (negative). McAuley et al. (2015) were the first to introduce low-rank embeddings to metric learning. Chen and He (2018) added a mixed category metric to the scheme in McAuley et al. (2015), and solved the problem of fashion compatibility by extending the triplet neural network to accept multiple instances in an iterative approach. Sun et al. (2020a) employed the visual semantic fusion model (VSFM) to extract the high-level semantic and visual features of fashion items to learn fashion compatibility. As shown in Fig. 12, a triplet of items with images and texts is inputted into the model, wherein visual embeddings, textual embeddings, and a pioneering triplet loss layer with appropriately tuned weights are merged. To address the issue of difficulty in accessing fashion datasets for supervised learning, a semi-supervised visual representation of fashion compatibility methods (Revanur et al. 2021) was proposed. An encouraging finding was that this method achieved equivalent performance to fully supervised methods. Item-to-item research in metric learning is relatively abundant (Ghojogh et al. 2022); however, limited research has

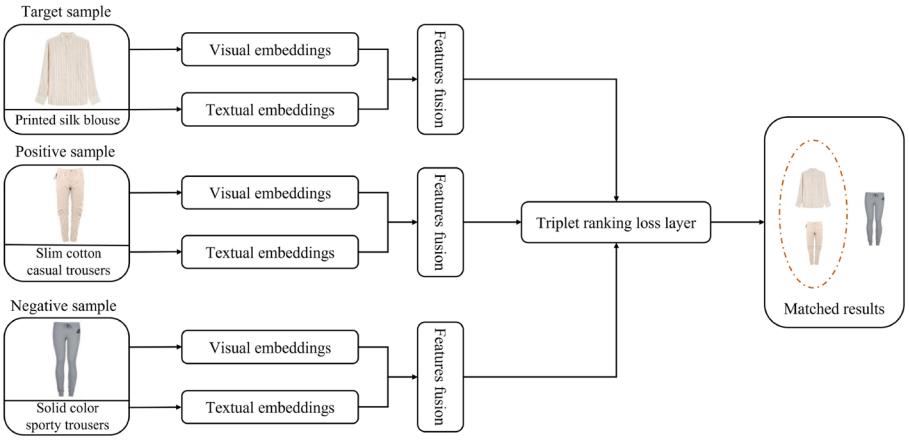

Fig. 12 The pipeline of VSFM (Sun et al. 2020a)

been done on item-to-set metrics. Zheng et al. (2021) proposed a general item-to-set metric for the task of fashion compatibility that used the neighboring importance and intra-set importance to filter out instances that were far away from a set.

In outfit compatibility learning, multiple garments are combined into sets to enable compatibility prediction. The compatibility of outfits is evaluated based not only on the visual similarity and semantic information of the fashion items, but also on the types of fashion items that are necessary to compose outfits. For example, Han et al. (2017) addressed the task of multi-garment compatibility learning by exploiting a bidirectional LSTM model (Le et al. 2019) in which clothing was viewed as a sequence and each item was taken at a step. Hsiao and Grauman (2018) designed a capsule wardrobe that could automatically form outfits from candidate items in a wardrobe to create recommendations. Using another approach, Zhang et al. (2020b) argued that color plays a significant role in clothing compatibility, and used a graph model to model multiple garments. Pang et al. (2021) divided compatibility into three levels and increased the interpretability of fashion compatibility predictions through the use of gradient penalties. In addition, Sarkar et al. (2022) designed a system called OutfitTransformer to capture the global representation of an item set and trained a network using a classification loss.

Fashion compatibility involves calculating the overall harmony of an outfit. Obviously, it is insufficient to treat clothing as a sequence and to focus only on the relationships between items, as this approach will overlook the overall harmony between item sets. Fashion compatibility learning has many important application scenarios in both the physical world and the Metaverse. It can help designers and consumers to select clothing, and can provide quantitative matching assistance for high-level fashion technologies.

5.3 Metaverse marketing

In the Metaverse, individuals have unlimited possibilities for creation, interaction, and exploration. Similar to the real world, the Metaverse contains unlimited business opportunities. A symbiotic and self-reinforcing relationship is formed between marketing and the development of Metaverse. A successful marketing strategy not only generates substantial profits for fashion companies but also allocates a portion of these earnings toward

Metaverse construction. This investment serves to draw an increasing number of consumers into the Metaverse, thus facilitating the expansion of the consumer market. Some emerging work has begun to explore marketing in the Metaverse. Ahn et al. (2022) studied advertising in the Metaverse by analyzing the relationship among consumers, media, and participation behavior. They proposed a bifold triadic relationships model to help readers understand the role of advertising in the Metaverse and its impact on consumer behavior. In the Metaverse, real-time multi-sensory social interaction (RMSI) between people is an important way of communication. Through theoretical logic and extensive field experiments, Hennig-Thurau et al. (2023) found that accessing the virtual world by VR headsets as part of RMSI can generate more interactive values than the 2-dimensional Internet. Hadi et al. (2023) argued that key characteristics of the Metaverse include digital mediation, spatiality, immersion, shared experiences, and real-time interactions. They also explored how the Metaverse is reshaping the understanding of consumer behavior in three areas: consumer identity, social influence, and ownership. Lu and Mintz (2023) aimed to offer marketing guidance from a corporate perspective to reduce uncertainty in the Metaverse. They provided a detailed explanation of Reibstein's 4P (Promotion, Product, Place, and Price) and 5C (Customer, Company, Competitors, Collaborators, and Context) theories (Reibstein and Iyengar 2023) and proposed seven Metaverse marketing strategies.

In addition to the methods mentioned above, many researchers also explored the impact of the Metaverse on consumers from various perspectives. Belk et al. (2022) focused on the research of digital economy and property ownership in the Metaverse, analyzing the motivations of digital art buyers. Giang Barrera and Shah (2023) argued that the user experience in the Metaverse depends on the level of immersion, environmental fidelity, and sociability, and enhancing these three aspects of the experience can attract more consumers. Dwivedi et al. (2023) enumerated various perspectives on the impact of the Metaverse on consumer psychology, consumer well-being, consumer awareness, sensory acceptance, and consumer information flow status. They also provided recommendations for how businesses can meet consumer demands.

5.4 Legal issues in Metaverse

The Metaverse, which remains in its nascent stage, has already gained nearly 100-billion-dollar investments from different companies. However, the absence of a comprehensive regulatory and legal framework presents significant risks. To mitigate these risks and foster further growth, it is essential to establish a rational and robust legal framework that can attract more participants and infuse vitality into the Metaverse. Currently, many scholars have attempted to explore the potential legal issues in the Metaverse. Kostenko et al. (2022) argued that the existing jurisdictional regulations for electronic governance are unapplicable to the regulations of public relations in the Metaverse. Their research and analysis on administrative, civil, criminal, labor, and property laws suggested that Metaverse enterprises, non-governmental organizations, and research institutions should jointly initiate the Metaverse Grand Legal Charter project to standardize public relations in the Metaverse. Cheong (2022b) discussed the powers, obligations, and potential risks that avatars possess in the Metaverse. He believed that real individuals behind the avatars should bear responsibility for the actions of their avatars in the Metaverse. Furthermore, he also extensively examined the risks associated with widely utilized non-fungible tokens (NFTs) within the Metaverse from the perspectives of ownership, financial regulation, taxation, and intellectual property rights (Cheong 2022a). Kasiyanto and Kilinc (2022) further explored the

limitations of directly applying real-world laws to the Metaverse. They believed that property law and intellectual property law would face challenges in terms of legal coverage, choices, and enforcement in the Metaverse. Moreover, they specifically highlighted issues such as data protection, network security, taxation and constraints on unethical behavior.

6 Future prospects and challenges

As an emerging field over the past year, fashion in the Metaverse is attracting increasing amounts of attention from both academia and industry. Many fashion companies have already invested resources into the Metaverse, for example by building virtual spokespersons and running catwalks. Nevertheless, many fashion application scenarios in the Metaverse are still unexplored. This section envisions some novel scenarios involving the Metaverse and highlights the current challenges in this domain.

6.1 Future prospects

6.1.1 Overcoming the physical constraints on fashion items

Beauty and comfort are the two most important factors in fashion clothing design. However, it is difficult to achieve both of these simultaneously due to physical constraints such as gravity, clothing fabrics, etc. For example, when designing a suit, designers add shoulder pads to widen the shoulders to make the body appear tall and straight. However, this limits the movement of the wearer, and prevents them from raising their arms comfortably. Fortunately, it is possible to overcome the constraints on the physical properties of materials in the Metaverse. Garments in the virtual world are a set of data that can make people feel comfortable in various activities. In addition, in the real world, a consumer must wear a heavily padded coat to stay warm, whereas in the Metaverse, clothing can automatically regulate body temperature, which makes it possible to wear lighter-looking clothing in cold places. Clothing in the Metaverse has the potential to overcome the physical constraints of the real world. Under these conditions, designers can boldly use their imagination to create astonishing fashion items that could not be made in the real world.

6.1.2 Convenience of fashion design in the Metaverse

Although designers may be inspired by many things in the real world, they cannot carry out an objective evaluation of clothing at the design stage; they typically first need to create a sketch of a garment and produce a physical sample before they can objectively evaluate it. Obviously, this process wastes a lot of the designer's time. In addition, due to the limitations of printing and dyeing technology, the clothing may not be able to be dyed to the color the designer wants.

Fortunately, every stage of the clothing design process is facilitated in the Metaverse. First, designers working in the environment of the Metaverse are able to easily obtain items that can inspire them. Famous designers do not need to spend several months traveling to find inspiration, as in the real world. In addition, designers can directly put clothing on a virtual model at the design stage for evaluation, which eliminates the step of producing a sample garment in the real world. Second, clothing designed with computer tools in the Metaverse will not show the deviations that occur in the real world, such as

color variations, discrepancies in garment shape, etc. In addition, a modular approach to the design of clothing can be applied in the Metaverse. Designers can employ a computer to preview and evaluate parts of the clothing before the overall design is complete. The Metaverse can therefore shorten the design process and lower the threshold of professional experience required for fashion design, allowing more consumers to join in the design process in an interactive way.

6.1.3 Shopping in the Metaverse

Metaverses are virtual worlds built on networks that can eliminate the physical distances that exist in the real world. Today, consumers generally buy fashion items in two ways: the first is offline shopping, while the other is online shopping and delivered production by express. Both of these methods have drawbacks. Offline shopping requires consumers to spend a lot of time on the road, while online shopping may mean that consumers buy unsuitable items, and several days may be needed to receive them. In contrast, shopping in the Metaverse offers the advantages of both types. Consumers can select and try on their favorite fashion items directly in a Metaverse fashion shop, which allows them to view the fitting in real time. A virtual shopping guide can provide customers with clothing evaluations and recommendations at any time. In addition, consumers can edit the size of the clothes according to their avatar's body, to achieve the most suitable effect. Finally, when consumers have chosen clothing that suits them, they can directly add it to their virtual wardrobe without waiting for delivery, as in the real world. In this way, shopping for clothes in the Metaverse will perfectly combine the advantages of online and offline shopping in the real world, providing the users (or "meta person") with a more comfortable shopping experience.

6.1.4 Expressing emotions and personality through clothing

In the real world, a fashionista can express their mood by the style and color of the clothing that they are wearing. For example, people in a good mood tend to wear brightly colored clothing. However, due to the limitations arising from the physical properties of the fabric, the color and style of clothing cannot be changed according to the mood of the wearer. In contrast, clothing in the Metaverse can overcome these limitations. The Metaverse allows designers to add variable properties to the clothing, such as images and colors, so that consumers can freely edit clothing elements based on their thoughts and emotions. For example, consumers can change their clothing to a warm color to indicate to others that they are in a good mood, while they may change to a cool color to express the idea of keeping strangers away. As a result, clothing in the Metaverse can convey more information and can be used to express the user's personality anywhere, at any time.

6.1.5 Expanding the boundaries of fashion in Metaverse

In the real world, the human body is the main vehicle for fashion items, and the physical features of the human body are one of the most important elements to be considered in clothing design. In the Metaverse, however, avatars can be of various types, and may be human-like, animal-like or even monster-like, meaning that the design of fashion items in the real world may no longer be valid in the Metaverse. For example, clothing designed to cover the private parts of humans would no longer work for avatars that do not have

private parts, such as puppies, Godzilla, aliens, etc. Hence, the Metaverse would significantly enlarge the range of fashion items, enrich the design ideas for fashion items, and expand the boundaries of fashion.

6.2 Challenges

6.2.1 Technology issues

A. Fine modeling of the human body and fashion items The great attractiveness of the Metaverse lies in the fact that it depicts a world that is completely different from the real one, and users can immerse themselves in it to experience an utterly different life. This immersive experience can be realized through the fine modeling of hundreds of objects in the Metaverse. Currently, fine modeling of fashion items and the human body relies on 3D scanning and manual modeling by artists with specialist knowledge, making it impractical to model thousands of objects using these time-consuming methods. Although the use of view-based 3D reconstruction methods can improve the efficiency of this process, the generated models have low accuracy, creating a less immersive experience for users of the Metaverse. Hence, the development of low-cost fine modeling methods for human and fashion items is a significant avenue for future work.

B. Simulation of 3D clothing fabrics Apparel fabric is an essential factor that characterizes the category and style of clothing. The physical properties of fabrics make clothing with the same style visually different. For example, a silk shirt is softer than a cotton shirt, and the details of their textures are also different. Many researchers focus on modeling the texture of clothing materials, and ignore the simulation of the stiffness of the fabrics. Poor simulation of fabric stiffness can make the avatar's clothing deform and swing more rigidegally when exercising, as well as causing unnatural mapping of clothing to different shapes of bodies. The accurate simulation of clothing fabrics in the Metaverse is another challenging topic.

6.2.2 Legal issues

A. Issues of fashion copyright in Metaverse Digitization is a feature that makes it easy to replicate Metaverse fashion items, and it is reasonable to expect that the illegal copying and counterfeiting of fashion items will become more widespread in the Metaverse. Hence, strengthening the copyright protection of fashion items in the Metaverse is an essential topic. In addition, the copyright owner of a given style of fashion item in the physical and virtual worlds is also an issue that needs to be discussed. For example, designer A may design a famous sweater S in the physical world, while designer B may digitize this sweater into the Metaverse. The copyright ownership of sweater S in the Metaverse then becomes controversial. The definition and protection of copyright for fashion items in the Metaverse is a topic that needs to be fully discussed and constrained.

B. Legal issues of avatars A digital avatar serves as the vessel of a real-world individual within the Metaverse. The question of whether identities in the Metaverse should have a one-to-one correspondence with natural persons is a topic worthy of discussion (Yang 2023). An individual from the real world possessing multiple digital avatars may more easily engage in illegal activities within the Metaverse (Cheong 2022b). Coordinating fraud using different avatars would also be facilitated. Furthermore, in the event of a sufficient number of avatars, it may even be possible to manipulate the Metaverse's stock market.

If Metaverse designers restrict each person to a unique avatar, they would inevitably employ distinct characteristics owned by each individual, such as fingerprints, iris patterns, etc., as login credentials. Ensuring the security of natural persons' login information and addressing potential leaks is a critical concern (Wu and Zhang 2023). Moreover, due to the one-to-one binding between Metaverse avatars and real-world individuals, it not only jeopardizes an individual's reputation but also leads to consequences in the real world if an avatar is stolen within the Metaverse.

C. Children's fashion in Metaverse In the Metaverse, the design of fashionable items also needs to adhere to the constraints of social ethics (Joy et al. 2022). Creating a Metaverse environment conducive to the healthy growth of children is a consideration for Metaverse designers (Patruti et al. 2023). For instance, designs for children's clothing should avoid any sexual implications or excessive exposure. Fashion items intended for display in public domains should refrain from incorporating elements like pornography or violence. Therefore, fashion designers within the Metaverse should conscientiously abide by social ethics, avoiding the introduction of products that promote deviant or harmful behaviors. Such constraints not only contribute to upholding harmony within the Metaverse community but also aid in fostering children's proper understanding of aesthetics and values.

6.2.3 Marketing issues

A. Diversity of Metaverse avatars Due to the complexity of consumer's hobbies, different consumers have different attitudes toward digital avatars in the Metaverse. Since the research on human-like digital avatars is relatively extensive, we take it as an example to introduce the development of digital avatar technology in the field of computer science. However, a significant portion of consumers prefers to use non-humanoid avatars such as animals or slimes. Therefore, in the Metaverse design stage, the needs and expectations of users must be fully considered. For instance, we can create a diverse virtual community which offers multilingual support to accommodate different cultures, genders, and abilities. In addition, providing customized fashion services that allow users to shape their virtual lives according to their own desires and interests is also essential. Such Metaverse that can include a variety of different user experiences and preferences can attract more consumers, thus producing more values for business.

B. Uncertainties in Metaverse development Since the concept of the Metaverse has only entered the public consciousness for a short time, the construction of the Metaverse remains in its early stage. There is great uncertainty about the ultimate form of the Metaverse and how consumers will utilize it. From a perspective of computer technology, limitations in computing resources, computer vision and 3D modeling, still prevent us from achieving highly fine modeling and real-time rendering for large-scale virtual worlds. Additionally, the challenges of ensuring real-time communication among hundreds of millions of users in the Metaverse persist due to network quality and data processing speed constraints. From a commercial perspective, how to understand and guide consumer behavior in the Metaverse and how to adjust brand and service marketing to meet consumer needs are also important issues in building a healthy ecology of the Metaverse.

C. Shifting focus between Metaverse and generative AI Recently, the emergence of large models such as ChatGPT (Van Dis et al. 2023) and Stable diffusion (Rombach et al. 2022) has brought amazing results on generative AI. This shift in focus has led people to move their attention from the Metaverse towards generative AI which is also called Artificial Intelligence Generated Content (AIGC). Even Meta, which had previously invested $36

billion in the Metaverse, has shifted the firm's focus away from a Metaverse-first position. This has raised questions among companies about the continued value of researching the Metaverse. From a business perspective, due to technological limitations, the Metaverse still has many uncertainties in the Metaverse, making it challenging to bring substantial profits in a short period. In contrast, AIGC can be embedded into existing services through simple fine-tuning, thereby improving the service experience. However, it is essential to note that the Metaverse and AIGC are not mutually exclusive. The Metaverse focuses on building virtual worlds and optimizing human-computer interaction services, while AIGC specializes in content generation. Currently, research in AIGC can contribute AI-generated contents to the Metaverse, enabling it to offer higher-quality services and making the Metaverse even more attractive.

7 Conclusion

In this paper, we have presented a comprehensive survey of the two main elements of fashion in the Metaverse: digital virtual humans and fashion items. In our study of digital virtual humans, we focused on investigating methods of generating 3D avatars, which can reduce the cost in the Metaverse of generating digital bodies. In addition, we reviewed fashion learning and analysis methods that could assist both fashion designers and consumers in the Metaverse. We also envisioned certain fashion scenarios in the Metaverse and discussed several important open issues associated with the future development of the Metaverse. We believe this survey will be instructive for both academics and industrial practitioners, and will shed some light on the study of fashion tasks in the Metaverse.

Author contributions XM and HZ wrote the main manuscript text. XM, JS and JH collected the investigated literature and prepared the statistics and all figures. JM and YY helped to improve the writing and discussed the contents of the manuscript. All authors reviewed the manuscript.

Declarations

Conflict of interest The authors declare that they have no conflicts of interest.

References

Abbas Q, Ibrahim ME, Jaffar MA (2019) A comprehensive review of recent advances on deep vision systems. Artif Intell Rev 52(1):39–76

Ahn SJG, Kim J, Kim J (2022) The bifold triadic relationships framework: a theoretical primer for advertising research in the metaverse. J Advert 51(5):592–607. https://doi.org/10.1080/00913367.2022.2111729

Ak KE, Lim JH, Tham JY, Kassim AA (2020) Semantically consistent text to fashion image synthesis with an enhanced attentional generative adversarial network. Pattern Recognit Lett 135:22–29. https://doi.org/10.1016/j.patrec.2020.02.030

Al-Halah Z, Stiefelhagen R, Grauman K (2017) Fashion forward: forecasting visual style in fashion. In: IEEE international conference on computer vision, Venice, Italy, October 22–29, 2017. pp 388–397. https://doi.org/10.1109/ICCV.2017.50

Anantrasirichai N, Bull D (2022) Artificial intelligence in the creative industries: a review. Artif Intell Rev. https://doi.org/10.1007/s10462-021-10039-7

Anguelov D, Srinivasan P, Koller D, Thrun S, Rodgers J, Davis J (2005) Scape: shape completion and animation of people. ACM Trans Graph 24(3):408–416

Batmaz Z, Yurekli A, Bilge A, Kaleli C (2019) A review on deep learning for recommender systems: challenges and remedies. Artif Intell Rev 52:1–37

Belk R, Humayun M, Brouard M (2022) Money, possessions, and ownership in the metaverse: NFTs, cryptocurrencies, Web3 and Wild Markets. J Bus Res 153:198–205. https://doi.org/10.1016/j.jbusres.2022.08.031

Belongie SJ, Malik J, Puzicha J (2002) Shape matching and object recognition using shape contexts. IEEE Trans Pattern Anal Mach Intell 24(4):509–522. https://doi.org/10.1109/34.993558

Bhagavatula C, Zhu C, Luu K, Savvides M (2017) Faster than real-time facial alignment: a 3D spatial transformer network approach in unconstrained poses. In: IEEE international conference on computer vision, Venice, Italy, October 22–29. pp 4000–4009. https://doi.org/10.1109/ICCV.2017.429

Blanz V, Vetter T (1999) A morphable model for the synthesis of 3D faces. In: Proceedings of the 26th annual conference on computer graphics and interactive techniques, Los Angeles, CA, USA, August 8–13, 1999. pp 187–194

Bogo F, Kanazawa A, Lassner C, Gehler PV, Romero J, Black MJ (2016) Keep it SMPL: automatic estimation of 3D human pose and shape from a single image. In: 14th European conference, Amsterdam, The Netherlands, October 11–14, proceedings, part V, vol 9909. pp 561–578. https://doi.org/10.1007/978-3-319-46454-1_34

Chang H, Lu J, Yu F, Finkelstein A (2018) PairedCycleGan: asymmetric style transfer for applying and removing makeup. In: 2018 IEEE/CVF conference on computer vision and pattern recognition. pp 40–48. https://doi.org/10.1109/CVPR.2018.00012

Chen L, He Y (2018) Dress fashionably: learn fashion collocation with deep mixed-category metric learning. In: Proceedings of the thirty-second AAAI conference on artificial intelligence, New Orleans, Louisiana, USA, February 2–7, 2018. pp 2103–2110

Chen H, Xu ZJ, Liu Z, Zhu SC (2006) Composite templates for cloth modeling and sketching. In: IEEE computer society conference on computer vision and pattern recognition, 17–22 June, New York, NY, USA. pp 943–950. https://doi.org/10.1109/CVPR.2006.81

Chen X, Chen H, Xu H, Zhang Y, Cao Y, Qin Z, Zha H (2019) Personalized fashion recommendation with visual explanations based on multimodal attention network: towards visually explainable recommendation. In: Proceedings of the 42nd international ACM SIGIR conference on research and development in information retrieval, Paris, France, July 21–25, 2019. pp 765–774. https://doi.org/10.1145/3331184.3331254

Cheong BC (2022a) Application of blockchain-enabled technology: regulating non-fungible tokens (NFTs) in Singapore. Singapore Law Gazette, January

Cheong BC (2022b) Avatars in the metaverse: potential legal issues and remedies. Int Cybersecur Law Rev 3(2):467–494

Choi J, Medioni GG, Lin Y, Silva L, Bellon ORP, Pamplona M, Faltemier TC (2010) 3D face reconstruction using a single or multiple views. In: 20th international conference on pattern recognition, Istanbul, Turkey, 23–26 August. pp 3959–3962. https://doi.org/10.1109/ICPR.2010.963

Corona E, Pumarola A, Alenyà G, Pons-Moll G, Moreno-Noguer F (2021) SMPLicit: topology-aware generative model for clothed people. In: IEEE conference on computer vision and pattern recognition, virtual, June 19–25. pp 11875–11885. https://doi.org/10.1109/CVPR46437.2021.01170

Cui YR, Liu Q, Gao CY, Su Z (2018) FashionGAN: display your fashion design using conditional generative adversarial nets. Comput Graph Forum 37(7):109–119. https://doi.org/10.1111/cgf.13552

Deng H, Han C, Cai H, Han G, He S (2021) Spatially-invariant style-codes controlled makeup transfer. In: 2021 IEEE/CVF conference on computer vision and pattern recognition (CVPR). pp 6545–6553. https://doi.org/10.1109/CVPR46437.2021.00648

Dong J, Chen Q, Huang Z, Yang J, Yan S (2016) Parsing based on Parselets: a unified deformable mixture model for human parsing. IEEE Trans Pattern Anal Mach Intell 38(1):88–101. https://doi.org/10.1109/TPAMI.2015.2420563

Dong L, Zhang H, Yang K, Zhou D, Shi J, Ma J (2022) Crowd counting by using top-k relations: a mixed ground-truth CNN framework. IEEE Trans Consum Electron 68(3):307–316
Dwivedi YK, Hughes L, Wang Y, Alalwan AA, Ahn SJ, Balakrishnan J, Barta S, Belk R, Buhalis D, Dutot V et al (2023) Metaverse marketing: how the metaverse will shape the future of consumer research and practice. Psychol Market 40(4):750–776
Fan J, Wang S, Ma X, Xu A, Ye S, Shi X (2022) Clothing parsing based on context prior and flow alignment pyramid. In: 2022 7th international conference on cloud computing and big data analytics. pp 439–444. https://doi.org/10.1109/ICCCBDA55098.2022.9778856
Feng M, Gilani SZ, Wang Y, Mian AS (2018a) 3D face reconstruction from light field images: a model-free approach. In: 15th European conference, Munich, Germany, September 8–14, proceedings, part X, vol 11214. pp 508–526. https://doi.org/10.1007/978-3-030-01249-6_31
Feng Y, Wu F, Shao X, Wang Y, Zhou X (2018b) Joint 3D face reconstruction and dense alignment with position map regression network. In: 15th European conference, Munich, Germany, September 8–14, proceedings, part XIV, vol 11218. pp 557–574. https://doi.org/10.1007/978-3-030-01264-9_33
Fenocchi E, Morelli D, Cornia M, Baraldi L, Cesari F, Cucchiara R (2022) Dual-branch collaborative transformer for virtual try-on. In: IEEE/CVF conference on computer vision and pattern recognition workshops, New Orleans, LA, USA, June 19–20, 2022. pp 2246–2250. https://doi.org/10.1109/CVPRW56347.2022.00246
Gabale V, Subramanian AP (2018) How to extract fashion trends from social media? A robust object detector with support for unsupervised learning. CoRR. . http://arxiv.org/abs/1806.10787
Gao S, Zeng F, Cheng L, Fan J, Zhao M (2022) Fashion image search via anchor-free detector. In: Proceedings of the 2022 international conference on multimedia retrieval. pp 416–425
Ge C, Song Y, Ge Y, Yang H, Liu W, Luo P (2021a) Disentangled cycle consistency for highly-realistic virtual try-on. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp 16928–16937
Ge Y, Song Y, Zhang R, Ge C, Liu W, Luo P (2021b) Parser-free virtual try-on via distilling appearance flows. In: IEEE conference on computer vision and pattern recognition, virtual, June 19–25, 2021. pp 8485–8493. https://doi.org/10.1109/CVPR46437.2021.00838
Gee S-J, Cho Y-I, Man Q (2022) GAN based hairstyle generation framework for standardization of lightweight-model. In: 2022 13th international conference on information and communication technology convergence (ICTC). pp 754–756. https://doi.org/10.1109/ICTC55196.2022.9952719
Ghojogh B, Ghodsi A, Karray F, Crowley M (2022) Spectral, probabilistic, and deep metric learning: tutorial and survey. CoRR. http://arxiv.org/abs/2201.09267
Giang Barrera K, Shah D (2023) Marketing in the metaverse: conceptual understanding, framework, and research agenda. J Bus Res 155:113420. https://doi.org/10.1016/j.jbusres.2022.113420
Godi M, Joppi C, Skenderi G, Cristani M (2022) MovingFashion: a benchmark for the video-to-shop challenge. In: Proceedings of the IEEE/CVF winter conference on applications of computer vision (WACV). pp 1678–1686
Guan P, Weiss A, Balan AO, Black MJ (2009) Estimating human shape and pose from a single image. In: IEEE 12th international conference on computer vision, Kyoto, Japan, September 27–October 4. pp 1381–1388. https://doi.org/10.1109/ICCV.2009.5459300
Güler RA, Neverova N, Kokkinos I (2018) DensePose: dense human pose estimation in the wild. In: 2018 IEEE conference on computer vision and pattern recognition, Salt Lake City, UT, USA, June 18–22, 2018. pp 7297–7306. https://doi.org/10.1109/CVPR.2018.00762
Guo D, Sim T (2009) Digital face makeup by example. In: 2009 IEEE conference on computer vision and pattern recognition. pp 73–79. https://doi.org/10.1109/CVPR.2009.5206833
Hadi R, Melumad S, Park ES (2023) The metaverse: a new digital frontier for consumer behavior. J Consum Psychol 34:142–166
Han X, Wu Z, Jiang Y, Davis LS (2017) Learning fashion compatibility with bidirectional LSTMs. In: Proceedings of the 2017 ACM on multimedia conference, Mountain View, CA, USA, October 23–27, 2017. pp 1078–1086. https://doi.org/10.1145/3123266.3123394
Han X, Wu Z, Wu Z, Yu R, Davis LS (2018) VITON: an image-based virtual try-on network. In: IEEE conference on computer vision and pattern recognition, Salt Lake City, UT, USA, June 18–22, 2018. pp 7543–7552. https://doi.org/10.1109/CVPR.2018.00787
Hasan B, Hogg DC (2010) Segmentation using deformable spatial priors with application to clothing. In: Labrosse F, Zwiggelaar R, Liu Y, Tiddeman B (eds) British machine vision conference, Aberystwyth, UK, August 31–September 3. pp 1–11. https://doi.org/10.5244/C.24.83
Hasler N, Stoll C, Sunkel M, Rosenhahn B, Seidel H (2009) A statistical model of human pose and body shape. Comput Graph Forum 28(2):337–346. https://doi.org/10.1111/j.1467-8659.2009.01373.x

He R, McAuley JJ (2016) VBPR: visual Bayesian personalized ranking from implicit feedback. In: Proceedings of the thirtieth conference on artificial intelligence, February 12–17, 2016, Phoenix, Arizona, USA. pp 144–150
He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. In: IEEE conference on computer vision and pattern recognition, Las Vegas, NV, USA, June 27–30, 2016. pp 770–778. https://doi.org/10.1109/CVPR.2016.90
Hennig-Thurau T, Aliman DN, Herting AM, Cziehso GP, Linder M, Kübler RV (2023) Social interactions in the metaverse: framework, initial evidence, and research roadmap. J Acad Mark Sci 51(4):889–913
Hirshberg DA, Loper M, Rachlin E, Black MJ (2012) Coregistration: simultaneous alignment and modeling of articulated 3D shape. In: 12th European conference on computer vision, Florence, Italy, October 7–13, proceedings, part VI, vol 7577. pp 242–255. https://doi.org/10.1007/978-3-642-33783-3_18
Hosseinnia Shavaki F, Ebrahimi Ghahnavieh A (2022) Applications of deep learning into supply chain management: a systematic literature review and a framework for future research. Artif Intell Rev 56(5):4447–4489
Hsiao W, Grauman K (2018) Creating capsule wardrobes from fashion images. In: 2018 IEEE conference on computer vision and pattern recognition, Salt Lake City, UT, USA, June 18–22, 2018. pp 7161–7170. https://doi.org/10.1109/CVPR.2018.00748
Huang J, Feris RS, Chen Q, Yan S (2015) Cross-domain image retrieval with a dual attribute-aware ranking network. In: 2015 IEEE international conference on computer vision, Santiago, Chile, December 7–13, 2015. pp 1062–1070. https://doi.org/10.1109/ICCV.2015.127
Jia Y, Shelhamer E, Donahue J, Karayev S, Long J, Girshick RB, Guadarrama S, Darrell T (2014) Caffe: convolutional architecture for fast feature embedding. In: Proceedings of the ACM international conference on multimedia, MM '14, Orlando, FL, USA, November 03–07, 2014. pp 675–678. https://doi.org/10.1145/2647868.2654889
Jiang S, Li J, Fu Y (2022) Deep learning for fashion style generation. IEEE Trans Neural Netw Learn Syst 33(9):4538–4550. https://doi.org/10.1109/TNNLS.2021.3057892
Jin B, Lu B, Wu H, Shi W, Li Y (2021) Fashion style forecasts based on different price ranges. In: IEEE 5th advanced information technology, electronic and automation control conference, vol 5. pp 2296–2302. https://doi.org/10.1109/IAEAC50856.2021.9390629
Jin Y, Li Q, Jiang D, Tong R (2022) High-fidelity 3D face reconstruction with multi-scale details. Pattern Recognit Lett 153:51–58. https://doi.org/10.1016/j.patrec.2021.11.022
Joy A, Zhu Y, Peña C, Brouard M (2022) Digital future of luxury brands: metaverse, digital fashion, and non-fungible tokens. Strateg Change 31(3):337–343
Kasiyanto S, Kilinc MR (2022) The legal conundrums of the metaverse. J Cent Bank Law Inst 1(2):299–322
Kemelmacher-Shlizerman I, Seitz SM (2011) Face reconstruction in the wild. In: IEEE international conference on computer vision, Barcelona, Spain, November 6–13. pp 1746–1753. https://doi.org/10.1109/ICCV.2011.6126439
Khurana T, Mahajan K, Arora C, Rai A (2018) Exploiting texture cues for clothing parsing in fashion images. In: IEEE international conference on image processing, Athens, Greece, October 7–10, 2018. pp 2102–2106. https://doi.org/10.1109/ICIP.2018.8451281
Kiapour MH, Yamaguchi K, Berg AC, Berg TL (2014) Hipster wars: discovering elements of fashion styles. In: 13th European conference, Zurich, Switzerland, September 6–12, 2014, proceedings, part I, vol 8689. pp 472–488. https://doi.org/10.1007/978-3-319-10590-1_31
Kiapour MH, Han X, Lazebnik S, Berg AC, Berg TL (2015) Where to buy it: matching street clothing photos in online shops. In: 2015 IEEE international conference on computer vision, Santiago, Chile, December 7–13, 2015. pp 3343–3351. https://doi.org/10.1109/ICCV.2015.382
Kinli F, Özcan B, Kiraç F (2019) Fashion image retrieval with capsule networks. In: 2019 IEEE/CVF international conference on computer vision workshops, Seoul, Korea (South), October 27–28, 2019. pp 3109–3112. https://doi.org/10.1109/ICCVW.2019.00376
Kips R, Gori P, Perrot M, Bloch I (2020) CA-GAN: weakly supervised color aware GAN for controllable makeup transfer. In: Bartoli A, Fusiello A (eds) Computer vision—ECCV 2020 workshops. Springer, Cham, pp 280–296
Kostenko O, Furashev V, Zhuravlov D, Dniprov O (2022) Genesis of legal regulation web and the model of the electronic jurisdiction of the metaverse. Bratisl Law Rev 6(2):21–36
Krajnik W, Markiewicz L, Sitnik R (2022) sSfS: segmented shape from silhouette reconstruction of the human body. Sensors 22(3):925. https://doi.org/10.3390/s22030925
Lassner C, Romero J, Kiefel M, Bogo F, Black MJ, Gehler PV (2017) Unite the people: closing the loop between 3D and 2D human representations. In: IEEE conference on computer vision and pattern recognition, Honolulu, HI, USA, July 21–26. pp 4704–4713. https://doi.org/10.1109/CVPR.2017.500

Le T, Vo MT, Vo B, Hwang E, Rho S, Baik SW (2019) Improving electric energy consumption prediction using CNN and Bi-LSTM. Appl Sci. https://doi.org/10.3390/app9204237

Lee Y, Chen ANK (2011) Usability design and psychological ownership of a virtual world. J Manag Inf Syst 28(3):269–308. https://doi.org/10.2753/MIS0742-1222280308

Lee J, Lumentut JS, Park IK (2022) Holistic 3D face and head reconstruction with geometric details from a single image. Multimed Tools Appl 81(26):38217–38233. https://doi.org/10.1007/s11042-022-13590-9

Li T, Qian R, Dong C, Liu S, Yan Q, Zhu W, Lin L (2018) BeautyGAN: instance-level facial makeup transfer with deep generative adversarial network. In: Proceedings of the 26th ACM international conference on multimedia. pp 645–653

Li M, Huang H, Zheng Y, Li M, Sang N, Ma C (2022) Implicit neural deformation for sparse-view face reconstruction. Comp Graph For 41(7):601–610

Liang J, Tu H, Liu F, Zhao Q, Jain AK (2020) 3D face reconstruction from mugshots: application to arbitrary view face recognition. Neurocomputing 410:12–27. https://doi.org/10.1016/j.neucom.2020.05.076

Lin C, Wang O, Russell BC, Shechtman E, Kim VG, Fisher M, Lucey S (2019) Photometric mesh optimization for video-aligned 3D object reconstruction. In: IEEE conference on computer vision and pattern recognition, Long Beach, CA, USA, June 16–20, 2019. pp 969–978. https://doi.org/10.1109/CVPR.2019.00106

Lin Y, Ren P, Chen Z, Ren Z, Ma J, de Rijke M (2020) Explainable outfit recommendation with joint outfit matching and comment generation. IEEE Trans Knowl Data Eng 32(8):1502–1516. https://doi.org/10.1109/TKDE.2019.2906190

Liu S, Feng J, Song Z, Zhang T, Lu H, Xu C, Yan S (2012a) Hi, magic closet, tell me what to wear! In: Proceedings of the 20th ACM multimedia conference, MM '12, Nara, Japan, October 29–November 02, 2012. pp 619–628. https://doi.org/10.1145/2393347.2393433

Liu S, Song Z, Liu G, Xu C, Lu H, Yan S (2012b) Street-to-shop: cross-scenario clothing retrieval via parts alignment and auxiliary set. In: 2012 IEEE conference on computer vision and pattern recognition, Providence, RI, USA, June 16–21, 2012. pp 3330–3337. https://doi.org/10.1109/CVPR.2012.6248071

Liu L, Xing J, Liu S, Xu H, Zhou X, Yan S (2014a) Wow! you are so beautiful today! ACM Trans Multimed Comput Commun Appl 11(1s):20–12022

Liu S, Feng J, Domokos C, Xu H, Huang J, Hu Z, Yan S (2014b) Fashion parsing with weak color-category labels. IEEE Trans Multimed 16(1):253–265. https://doi.org/10.1109/TMM.2013.2285526

Liu S, Liang X, Liu L, Lu K, Lin L, Cao X, Yan S (2015) Fashion parsing with video context. IEEE Trans Multimed 17(8):1347–1358. https://doi.org/10.1109/TMM.2015.2443559

Liu Q, Wu S, Wang L (2017) DeepStyle: learning user preferences for visual recommendation. In: Proceedings of the 40th international ACM conference on research and development in information retrieval, Shinjuku, Tokyo, Japan, August 7–11, 2017. pp 841–844. https://doi.org/10.1145/3077136.3080658

Loper M, Mahmood N, Romero J, Pons-Moll G, Black MJ (2015) SMPL: a skinned multi-person linear model. ACM Trans Graph 34(6):248–124816. https://doi.org/10.1145/2816795.2818013

Lu S, Mintz O (2023) Marketing on the metaverse: research opportunities and challenges. AMS Rev. https://doi.org/10.1007/s13162-023-00255-5

McAuley JJ, Targett C, Shi Q, van den Hengel A (2015) Image-based recommendations on styles and substitutes. In: Proceedings of the 38th international ACM SIGIR conference on research and development in information retrieval, Santiago, Chile, August 9–13, 2015. pp 43–52. https://doi.org/10.1145/2766462.2767755

Mebatsion HK, Paliwal J, Jayas DS (2012) Evaluation of variations in the shape of grain types using principal components analysis of the elliptic Fourier descriptors. Comput Electron Agric 80:63–70. https://doi.org/10.1016/j.compag.2011.10.016

Mystakidis S (2022) Metaverse. Encyclopedia 2(1):486–497

Natsume R, Saito S, Huang Z, Chen W, Ma C, Li H, Morishima S (2019) SiCloPe: silhouette-based clothed people. In: IEEE conference on computer vision and pattern recognition, Long Beach, CA, USA, June 16–20. pp 4480–4490. https://doi.org/10.1109/CVPR.2019.00461

Ning H, Wang H, Lin Y, Wang W, Dhelim S, Farha F, Ding J, Daneshmand M (2021) A survey on metaverse: the state-of-the-art, technologies, applications, and challenges. arXiv Preprint. http://arxiv.org/abs/2111.09673

Nunziatini A, Fani V, Bindi B, Bandinelli R, Tucci M (2022) Data-driven simulation for production balancing and optimization: a case study in the fashion luxury industry. In: Winter simulation conference, WSC 2022, Singapore, December 11–14, 2022. pp 2957–2967

Pang K, Zou X, Wong W (2021) Modeling fashion compatibility with explanation by using bidirectional LSTM. In: IEEE conference on computer vision and pattern recognition workshops, virtual, June 19–25, 2021. pp 3894–3898. https://doi.org/10.1109/CVPRW53098.2021.00432
Patruti P, Zbuchea A, Pînzaru F (2023) Fashion joining online gaming and the metaverse. In: Proceedings of the international conference on business excellence, vol 17. pp 1065–1074
Pratama M, Wang D (2019) Deep stacked stochastic configuration networks for lifelong learning of non-stationary data streams. Inf Sci 495:150–174
Qi CR, Su H, Mo K, Guibas LJ (2017) PointNet: deep learning on point sets for 3D classification and segmentation. In: IEEE conference on computer vision and pattern recognition, Honolulu, HI, USA, July 21–26, 2017. pp 77–85. https://doi.org/10.1109/CVPR.2017.16
Raj A, Sangkloy P, Chang H, Hays J, Ceylan D, Lu J (2018) SwapNet: image based garment transfer. In: 15th European conference, Munich, Germany, September 8–14, 2018, proceedings, part XII. Lecture notes in computer science, vol 11216. pp 679–695. https://doi.org/10.1007/978-3-030-01258-8_41
Ranjan A, Bolkart T, Sanyal S, Black MJ (2018) Generating 3D faces using convolutional mesh autoencoders. In: 15th European conference, Munich, Germany, September 8–14, proceedings, part III, vol 11207. pp 725–741. https://doi.org/10.1007/978-3-030-01219-9_43
Reibstein DJ, Iyengar R (2023) Metaverse-will it change the world or be a whole new world in and of itself? AMS Rev. https://doi.org/10.1007/s13162-023-00258-2
Rendle S, Freudenthaler C, Gantner Z, Schmidt-Thieme L (2012) BPR: Bayesian personalized ranking from implicit feedback. CoRR. http://arxiv.org/abs/1205.2618
Revanur A, Kumar V, Sharma D (2021) Semi-supervised visual representation learning for fashion compatibility. In: RecSys '21: fifteenth ACM conference on recommender systems, Amsterdam, The Netherlands, 27 September 2021–1 October 2021. pp 463–472. https://doi.org/10.1145/3460231.3474233
Richardson E, Sela M, Or-El R, Kimmel R (2017) Learning detailed face reconstruction from a single image. In: IEEE conference on computer vision and pattern recognition, Honolulu, HI, USA, July 21–26. pp 5553–5562. https://doi.org/10.1109/CVPR.2017.589
Rombach R, Blattmann A, Lorenz D, Esser P, Ommer B (2022) High-resolution image synthesis with latent diffusion models. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp 10684–10695
Romdhani S, Vetter T (2005) Estimating 3D shape and texture using pixel intensity, edges, specular highlights, texture constraints and a prior. In: IEEE conference on computer vision and pattern recognition, San Diego, CA, USA, 20–26 June. pp 986–993. https://doi.org/10.1109/CVPR.2005.145
Ronneberger O, Fischer P, Brox T (2015) U-Net: convolutional networks for biomedical image segmentation. In: Medical image computing and computer-assisted intervention—MICCAI 2015—18th international conference Munich, Germany, October 5–9, 2015, proceedings, part III, vol 9351. pp 234–241. https://doi.org/10.1007/978-3-319-24574-4_28
Roth J, Tong Y, Liu X (2017) Adaptive 3D face reconstruction from unconstrained photo collections. IEEE Trans Pattern Anal Mach Intell 39(11):2127–2141. https://doi.org/10.1109/TPAMI.2016.2636829
Saito S, Huang Z, Natsume R, Morishima S, Li H, Kanazawa A (2019) PIFu: pixel-aligned implicit function for high-resolution clothed human digitization. In: IEEE/CVF international conference on computer vision, Seoul, Korea (South), October 27–November 2. pp 2304–2314. https://doi.org/10.1109/ICCV.2019.00239
Sarkar R, Bodla N, Vasileva MI, Lin Y, Beniwal A, Lu A, Medioni G (2022) OutfitTransformer: learning outfit representations for fashion recommendation. CoRR. http://arxiv.org/abs/2204.04812. https://doi.org/10.48550/arXiv.2204.04812
Sbai O, Elhoseiny M, Bordes A, LeCun Y, Couprie C (2018) Design: design inspiration from generative networks. In: ECCV 2018 workshops, Munich, Germany, September 8–14, 2018, proceedings, part III, vol 11131. pp 37–44. https://doi.org/10.1007/978-3-030-11015-4_5
Simo-Serra E, Ishikawa H (2016) Fashion style in 128 floats: joint ranking and classification using weak data for feature extraction. In: IEEE conference on computer vision and pattern recognition, Las Vegas, NV, USA, June 27–30, 2016. pp 298–307. https://doi.org/10.1109/CVPR.2016.39
Song X, Liu C, Zheng Y, Feng Z, Li L, Zhou K, Yu X (2023) Hairstyle editing via parametric controllable strokes. IEEE Trans Vis Comput Graph. https://doi.org/10.1109/TVCG.2023.3241894
Stephenson N (2003) Snow crash: a novel. Spectra, London
Sumner RW, Popovic J (2004) Deformation transfer for triangle meshes. ACM Trans Graph 23(3):399–405. https://doi.org/10.1145/1015706.1015736

Sun G, He J, Wu X, Zhao B, Peng Q (2020a) Learning fashion compatibility across categories with deep multimodal neural networks. Neurocomputing 395:237–246. https://doi.org/10.1016/j.neucom.2018.06.098

Sun Z, Liu F, Liu W, Xiong S, Liu W (2020b) Local facial makeup transfer via disentangled representation. In: Proceedings of the Asian conference on computer vision

Tahir R, Sargano AB, Habib Z (2021) Voxel-based 3D object reconstruction from single 2D image using variational autoencoders. Mathematics 9(18):2288

Takagi M, Simo-Serra E, Iizuka S, Ishikawa H (2017) What makes a style: experimental analysis of fashion prediction. In: IEEE international conference on computer vision workshops, Venice, Italy, October 22–29, 2017. pp 2247–2253. https://doi.org/10.1109/ICCVW.2017.263

Tangseng P, Okatani T (2020) Toward explainable fashion recommendation. In: IEEE winter conference on applications of computer vision, Snowmass Village, CO, USA, March 1–5, 2020. pp 2142–2151. https://doi.org/10.1109/WACV45572.2020.9093367

Tangseng P, Wu Z, Yamaguchi K (2017) Looking at outfit to parse clothing. CoRR. http://arxiv.org/abs/1703.01386

Tong W-S, Tang C-K, Brown MS, Xu Y-Q (2007) Example-based cosmetic transfer. In: 15th Pacific conference on computer graphics and applications (PG'07). IEEE, pp 211–218

Van Dis EA, Bollen J, Zuidema W, van Rooij R, Bockting CL (2023) ChatGPT: five priorities for research. Nature 614(7947):224–226

Vittayakorn S, Yamaguchi K, Berg AC, Berg TL (2015) Runway to realway: visual analysis of fashion. In: 2015 IEEE winter conference on applications of computer vision, Waikoloa, HI, USA, January 5–9, 2015. pp 951–958. https://doi.org/10.1109/WACV.2015.131

Wang W, Wang D (2020) Prediction of component concentrations in sodium aluminate liquor using stochastic configuration networks. Neural Comput Appl 32(17):13625–13638

Wang X, Liang W, Zhang L (2010) Morphable face reconstruction with multiple views. In: IEEE second international conference on intelligent human-machine systems and cybernetics, vol 2. pp 250–253

Wang B, Zheng H, Liang X, Chen Y, Lin L, Yang M (2018) Toward characteristic-preserving image-based virtual try-on network. In: 15th European conference, Munich, Germany, September 8–14, 2018, proceedings, part XIII. Lecture notes in computer science, vol 11217. pp 607–623. https://doi.org/10.1007/978-3-030-01261-8_36

Wang H, Yang J, Liang W, Tong X (2019) Deep single-view 3D object reconstruction with visual hull embedding. In: The thirty-third AAAI conference on artificial intelligence, Honolulu, Hawaii, USA, Jan. 27–Feb. 1, 2019. pp 8941–8948. https://doi.org/10.1609/aaai.v33i01.33018941

Wu H, Zhang W (2023) Digital identity, privacy security, and their legal safeguards in the metaverse. Secur Saf 2:2023011

Wu F, Bao L, Chen Y, Ling Y, Song Y, Li S, Ngan KN, Liu W (2019a) MVF-Net: multi-view 3D face morphable model regression. In: IEEE conference on computer vision and pattern recognition, Long Beach, CA, USA, June 16–20. pp 959–968. https://doi.org/10.1109/CVPR.2019.00105

Wu Z, Lin G, Tao Q, Cai J (2019b) M2E-Try On Net: fashion from model to everyone. In: Proceedings of the 27th ACM international conference on multimedia, Nice, France, October 21–25, 2019. pp 293–301. https://doi.org/10.1145/3343031.3351083

Xian W, Sangkloy P, Agrawal V, Raj A, Lu J, Fang C, Yu F, Hays J (2018) TextureGAN: controlling deep image synthesis with texture patches. In: IEEE conference on computer vision and pattern recognition, Salt Lake City, UT, USA, June 18–22, 2018. pp 8456–8465. https://doi.org/10.1109/CVPR.2018.00882

Xie H, Yao H, Sun X, Zhou S, Zhang S (2019) Pix2Vox: context-aware 3D reconstruction from single and multi-view images. In: IEEE/CVF international conference on computer vision, Seoul, Korea (South), October 27–November 2. pp 2690–2698. https://doi.org/10.1109/ICCV.2019.00278

Xu L, Du Y, Zhang Y (2013) An automatic framework for example-based virtual makeup. In: 2013 IEEE international conference on image processing. IEEE, pp 3206–3210

Yamaguchi K, Kiapour MH, Ortiz LE, Berg TL (2012) Parsing clothing in fashion photographs. In: IEEE conference on computer vision and pattern recognition, providence, RI, USA, June 16–21, 2012. pp 3570–3577. https://doi.org/10.1109/CVPR.2012.6248101

Yamaguchi K, Kiapour MH, Berg TL (2013) Paper doll parsing: retrieving similar styles to parse clothing items. In: IEEE international conference on computer vision, Sydney, Australia, December 1–8, 2013. pp 3519–3526. https://doi.org/10.1109/ICCV.2013.437

Yan H, Zhang H, Shi J, Ma J, Xu X (2022a) Toward intelligent fashion design: a texture and shape disentangled generative adversarial network. ACM Trans Multimed Comput Commun Appl. https://doi.org/10.1145/3567596

Yan H, Zhang H, Liu L, Zhou D, Xu X, Zhang Z, Yan S (2022b) Toward intelligent design: an AI-based fashion designer using generative adversarial networks aided by sketch and rendering generators. IEEE Trans Multimed. https://doi.org/10.1109/TMM.2022.3146010
Yan H, Zhang H, Shi J, Ma J (2022c) Texture brush for fashion inspiration transfer: a generative adversarial network with heatmap-guided semantic disentanglement. IEEE Trans Circ Syst Video Technol. https://doi.org/10.1109/TCSVT.2022.3224190
Yang S (2023) Metaverse: a new form of communication integrating reality and virtuality. In: Li F, Junkai L (eds) China's opportunities for development in an era of great global change. Understanding China. Springer, Singapore, pp 325–337. https://doi.org/10.1007/978-981-99-1199-8_19
Yin W, Fu Y, Ma Y, Jiang Y-G, Xiang T, Xue X (2017) Learning to generate and edit hairstyles. Association for Computing Machinery, New York, pp 1627–1635. https://doi.org/10.1145/3123266.3123423
Yu H, Cheang C, Fu Y, Xue X (2022) Multi-view shape generation for 3D human-like body. ACM Trans Multimed Comput Commun Appl. https://doi.org/10.1145/3514248
Yue X, Zhang C, Fujita H, Lv Y (2021) Clothing fashion style recognition with design issue graph. Appl Intell 51(6):3548–3560. https://doi.org/10.1007/s10489-020-01950-7
Zhang X, Jia J, Gao K, Zhang Y, Zhang D, Li J, Tian Q (2017) Trip outfits advisor: location-oriented clothing recommendation. IEEE Trans Multimed 19(11):2533–2544. https://doi.org/10.1109/TMM.2017.2696825
Zhang H, Chen W, He H, Jin Y (2019) Disentangled makeup transfer with generative adversarial network. CoRR. http://arxiv.org/abs/1907.01144
Zhang Y, Li L, Song L, Xie R, Zhang W (2020a) FACT: fused attention for clothing transfer with generative adversarial networks. In: The thirty-fourth AAAI conference on artificial intelligence, 2020, New York, NY, USA, February 7–12, 2020. pp 12894–12901
Zhang H, Yang X, Tan J, Wu C, Wang J, Kuo C-J (2020b) Learning color compatibility in fashion outfits. CoRR. http://arxiv.org/abs/2007.02388
Zhang Z, Ma J, Zhou C, Men R, Li Z, Ding M, Tang J, Zhou J, Yang H (2021) M6-UFC: unifying multi-modal controls for conditional image synthesis via non-autoregressive generative transformers. arXiv e-prints, 2105
Zhang D, Zuo C, Wu Q, Fu L, Xiang X (2022) Unabridged adjacent modulation for clothing parsing. Pattern Recognit 127:108594. https://doi.org/10.1016/j.patcog.2022.108594
Zhao B, Wu X, Peng Q, Yan S (2016) Clothing cosegmentation for shopping images with cluttered background. IEEE Trans Multimed 18(6):1111–1123. https://doi.org/10.1109/TMM.2016.2537783
Zhao L, Li M, Sun P (2021) Neo-fashion: a data-driven fashion trend forecasting system using catwalk analysis. Cloth Text Res J. https://doi.org/10.1177/0887302X211004299
Zhao M, Gao S, Ma J, Zhang Z (2022) Joint clothes image detection and search via anchor free framework. Neural Netw 155:84–94. https://doi.org/10.1016/j.neunet.2022.08.011
Zheng H, Wu K, Park J, Zhu W, Luo J (2021) Personalized fashion recommendation from personal social media data: an item-to-set metric learning approach. In: 2021 IEEE international conference on Big Data (Big Data), Orlando, FL, USA, December 15–18, 2021. pp 5014–5023. https://doi.org/10.1109/BigData52589.2021.9671563
Zhou D, Zhang H, Li Q, Ma J, Xu X (2022a) COutfitGAN: learning to synthesize compatible outfits supervised by silhouette masks and fashion styles. IEEE Trans Multimed. https://doi.org/10.1109/TMM.2022.3185894
Zhou D, Zhang H, Yang K, Liu L, Yan H, Xu X, Zhang Z, Yan S (2022b) Learning to synthesize compatible fashion items using semantic alignment and collocation classification: an outfit generation framework. IEEE Trans Neural Netw Learn Syst. https://doi.org/10.1109/TNNLS.2022.3202842
Zhou D, Zhang H, Ma J, Fan J, Zhang Z (2023) FCBoost-Net: a generative network for synthesizing multiple collocated outfits via fashion compatibility boosting. In: ACM international conference on multimedia. ACM
Zhu J-Y, Park T, Isola P, Efros AA (2017a) Unpaired image-to-image translation using cycle-consistent adversarial networks. In: 2017 IEEE international conference on computer vision (ICCV). pp 2242–2251. https://doi.org/10.1109/ICCV.2017.244
Zhu S, Fidler S, Urtasun R, Lin D, Loy CC (2017b) Be your own Prada: fashion synthesis with structural coherence. In: IEEE international conference on computer vision, Venice, Italy, October 22–29, 2017. pp 1689–1697. https://doi.org/10.1109/ICCV.2017.186