المدرسة العليا للإعلام الآلي - 08 ماي 1945 – بسيدي بلعباس
Ecole Supérieure en Informatique
- 08 Mai 1945- Sidi Bel Abbes

MEMOIRE

En Vue de l'obtention du diplôme de **Master**

Filière : **Informatique**

Spécialité : **Ingénierie des Systèmes Informatiques (ISI)**

Thème

Artificial intelligence approaches applied in the fashion industry

Présenté par **:**

- Mr Bourahla Karim
- Mr Benzeghli Nor El Islam

Soutenu le : **00/00/2021**

Devant le jury composé de :

- M/Mme/Mlle XXX	Président
- **M Khaldi Belkacem**	Encadreur
- M/Mme/Mlle XXX	Examinateur
- M/Mme/Mlle XXX	Examinateur

Abstract

Fashion has been one of the greatest ways for us to express our individuality. It can be defined as the way we wear our clothes, put on apparels and accessories which all takes place in a particular period, place and context. It is one of the most important business industries in the world right now so it's only fitting to do a recap of the fashion evolution in the computer vision field in this thesis we are going to explore the different topics covered in the literatures including fashion detection, fashion analysis, fashion synthesis and fashion compatibility. Where for each topic we cover its subtopics, datasets used, state of the art methods and evaluation metrics for each of them.

Keywords:

Fashion, computer vision, artificial intelligence, recommendation, synthesis, analysis, detection, deep learning.

Acknowledgement:

We would like to express our special thanks and gratitude to the following people for helping with this master thesis:

First of all to **our families** who supported us through the bad and through the good on this noble path and in life overall we cannot be grateful enough for them and words cannot describe the amount of support we had and still having.

To all the staff of our great school ESI SBA which was like family to us we cannot ask for better ones.

To our thesis framer **Dr. Khaldi Belkacem** who is one of the best professors in our university, who guided us to the correct path of realizing this thesis and gave us valuable advices and directions to perfect this domain.

To our Director of Studies in our school **Dr. Amar Bensaber Djamel** and **Dr. Benslimane Sidi Mohammed** for their great effort they've put in this school along our 5 years of rich study.

Finally, we would like to give our thanks to our friends and colleagues for their support through this journey.

Bourahla Karim and Benzeghli Nor El Islam.

Contents

Table of Figures

Table of Tables

Abbreviations

AI	Artificial Intelligence.
CNN	Convolutional Neural Network.
GAN	Generative adversarial network.
R-CNN	Region Based Convolutional Neural Networks.
SVM	Support Vector Machine.
DL	Deep Learning.
ML	Machine Learning.
ResNet	Residual Neural Network.
PGN	Part Grouping Network
LSTM	Long short-term memory
CRF	Conditional random field
SSIM	Structural similarity
IoU	Intersection over union
NE	Normalized Error
FID	Fréchet Inception distance (FID).
LPIPS	Learned Perceptual Image Patch Similarity
MSE	Mean squared error
SRC	Sparkman's Rank-Order Correlation
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error

General Introduction

I. Introduction and motivation:

Fashion is a source of art allowing us to look unique and original [1]. It is one of the biggest industries in the world, always bringing new trends and outfits.

Fashion trends are the ways to create new styles of clothing and appearances until they become accepted and fashionable that a lot of people would copy and make it viral.

Nowadays the fashion industry is invading sales in the e-commerce domain, the revenue of this market was approximately 1.46 trillion U.S. dollars. According to the Consumer Market Outlook, it will reach almost 2.25 trillion U.S. dollars by 2025.[2]

A portion of the world's quickest developing economies are in Africa, and there's a growing market for both luxury fashion and fashionable mid-market clothing and accessories. Unlike China African fashion industries focus more on its own internal demands.

Africa currently accounts for 1.9% of global trade, in the next five years, Africa's textile industry could generate $15.5 billion revenue. Research shows that with easy shipping access to European and USA Africa has an advantage over Asian manufacturers.

It takes three weeks for a shipping container to travel from West Africa to Western Europe and a month to travel to the eastern seaboard of the United States.[3]

In design deals, the suggestion innovation, as an arising innovation, has pulled in wide consideration of researchers. As it's broadly known, the conventional method of clothing suggestion relies upon manual activity.

To be explicit, salesmen need to prescribe a piece of clothing to clients to pursue them in buying.

Notwithstanding, it is hard for salesmen to understand clients' genuine needs and preferences so the domain of Artificial Intelligence interferes to bring us the best and most accurate results.

This is where fashion meets computer vision allowing us not to only detect fashion items and trends but to analyze them, synthesize them and even recommend some new or existing fashion items .

II. Plan of the thesis

Chapter 1 Background:

In this chapter we are going to cover the background of this study alongside with artificial intelligence basic concepts and their relation to computer vision and fashion respectively.

Chapter 2 F**ashion Detection:**

In this chapter we are going to cover 3 main topics which are: landmark detection, fashion parsing and item retrieval alongside the state of art methods for each topic.

Chapter 3 Fashion Analysis:

In this chapter we are going to cover 3 main topics which are: attribute recognition, style learning and popularity prediction alongside the state of art methods for each topic.

Chapter 4 Fashion Synthesis:

In this chapter we are going to cover 3 main topics which are: style transfer, pose transformation and physical simulation alongside the state of art methods for each topic.

Chapter 5 Fashion Recommendation:

In this chapter we are going to cover 3 main topics which are: fashion compatibility, outfit matching and hairstyle suggestion alongside the state of art methods for each topic.

Conclusion:

A conclusion of the overall study on fashion and computer vision.

Chapter 1 Background:

In this chapter, we will detail how algorithms of Artificial Intelligence will work and the steps of the website creation. Since there are multiple models & algorithms in AI we will compare each algorithm and choose the optimal one for our problem based on rapidity and accuracy, for the website we will base generally on the backend since the whole process and work is there rather than just the design itself.

1.1 Artificial Intelligence:

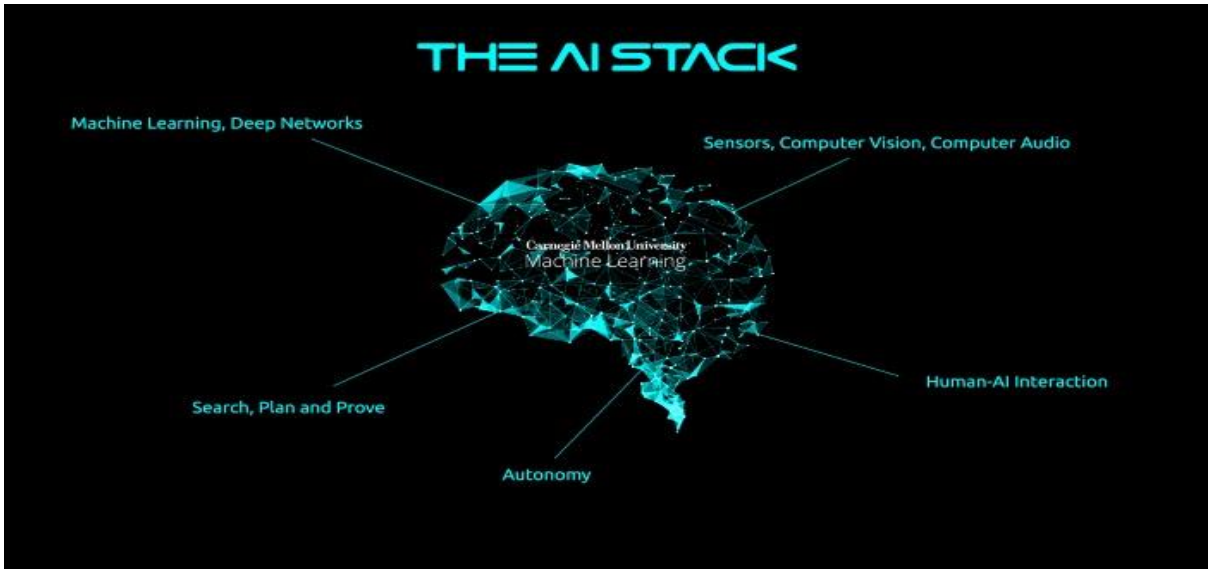

Fig.1: Artificial Intelligence and its main categories [5].

Artificial intelligence (AI) is the imitation of human intellect in robots that are trained to think like people and replicate their activities. The word can also refer to any machine that displays characteristics associated with the human mind, like training and problem-solving [5].

Artificial intelligence's ideal characteristic is its ability to assess and execute activities that have the best chance of achieving a given goal. Machine learning is a subset of artificial intelligence that refers to the idea that computer systems can automatically learn from and adaptable to changes without the assistance of humans. Deep learning approaches facilitate automated learning by absorbing massive quantities of unstructured data such as text, pictures, or video.

Artificial intelligence is founded on the idea that human intellect may be characterized in such a manner that a computer can simply imitate it and carry out tasks ranging from simple to complicated. Simulating human reasoning is one of artificial intelligence's goals. To the degree that they can be concretely described, researchers and developers in the field are making remarkably quick progress in simulating processes such as learning, reasoning, and perception. Some think that in the near future, inventors will be able to create systems that can learn and reason about any subject faster than humans can. Others, on the other hand, remain skeptical, claiming that all cognitive activity is loaded with value judgements based on human experience.

As technology improves, old artificial intelligence criteria become obsolete. Machines that calculate fundamental calculations or detect text using optical character recognition, for example, are no longer called artificial intelligence because these operations are now regarded standard computer functions.

AI is constantly improving to help a wide range of businesses. A multidisciplinary method based on mathematics, computer science, linguistics, psychology, and other disciplines is used to wire machines [5].

1.1.1 Artificial Intelligence Applications:

Artificial intelligence has a plethora of uses. The technology may be used in a variety of businesses and areas. In the healthcare business, AI is being studied and employed for administering medications and various therapies in patients, as well as surgical operations in the operating room.

Computers that play chess and self-driving vehicles are two more instances of artificial intelligence devices. Each of these machines must consider the implications of each action they perform, as each action has an influence on the final outcome. In chess, the ultimate goal is to win the game. The computer system in self-driving cars must account for all external data and calculate it in order to respond in a way that avoids a collision.

Artificial intelligence is also utilized in the financial sector to detect and highlight suspicious behaviors in banking and finance, such as irregular debit card usage and big account deposits, which assists a bank's fraud department. Artificial intelligence (AI) applications are now being utilized to assist expedite and simplify trade. This is accomplished by making it easier to predict the supply, need, and pricing of securities [5].

1.1.2 Machine Learning:

Machine learning (ML) is a subset of artificial intelligence that, according to Tom M. Mitchell, a research scientist and data science pioneer, is "the study of computer algorithms that allow computer systems to automatically improve through experience." — One of the ways we intend to develop AI is through machine learning. Working with tiny to big datasets, machine learning examines and compares the data to identify common patterns and subtleties [6].

For example, if you supply a machine learning model with a large number of songs that you love, as well as their audio characteristics (dance-ability, instrumentality, tempo, or genre). It should be able to automate (based on the supervised machine learning algorithms used) and produce a recommender system that will recommend songs to you in the future that you will likely love, similar to what Netflix, Spotify, and other firms do.

For example, if you feed a machine learning software a big dataset of x-ray images together with their descriptions (symptoms, things to think about, and so on), it should be able to help (or perhaps automate) the data analysis of x-ray images later on. The machine learning algorithm

examines each image in the large dataset for common patterns in images with labels that have similar signals.

Additionally, when you load the model with new photos (assuming we use an appropriate ML method for images), it compares its parameters to the examples it has acquired before to reveal how likely the images are to include any of the indicators it has studied previously.

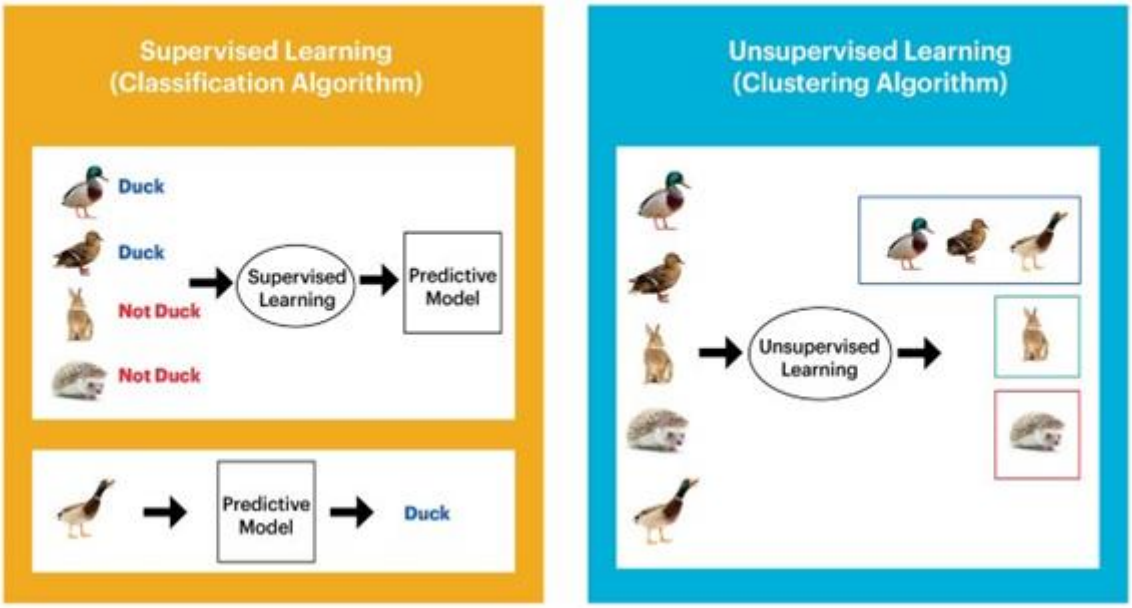

Fig.2: Supervised Learning (Classification/Regression) | Unsupervised Learning (Clustering)[6].

The type of machine learning from our previous example, known as "supervised learning," Classification is the process by which supervised learning algorithms attempt to model correlations seen between target prediction output and the input features, so that we can predict the output values for new data based on those relationships, which it has learned previously datasets served.

Another form of machine learning is **unsupervised learning**, which is a class of machine learning techniques with applications in pattern identification and descriptive modeling. Clustering is the name given to algorithms that do not have output categories or labels on the data (the model trains with unlabeled data). Similar to supervised learning, we have various techniques extracting k-Means, DB Scan, Apriori, and so on.

Semi-supervised learning performs the same functions as supervised learning, except it may train with both labeled and unlabeled data. You'll typically see a lot of unlabeled data and a tiny amount of labeled data in semi-supervised learning. Several studies have discovered that this method can give more accuracy than unsupervised learning while avoiding the effort and expense involved with labeled data. (Sometimes labeling data requires a trained person to do things like transcribe audio recordings or evaluate 3D pictures, which may make producing a properly labeled data set nearly impossible, especially when dealing with the huge data sets that deep learning jobs adore.)

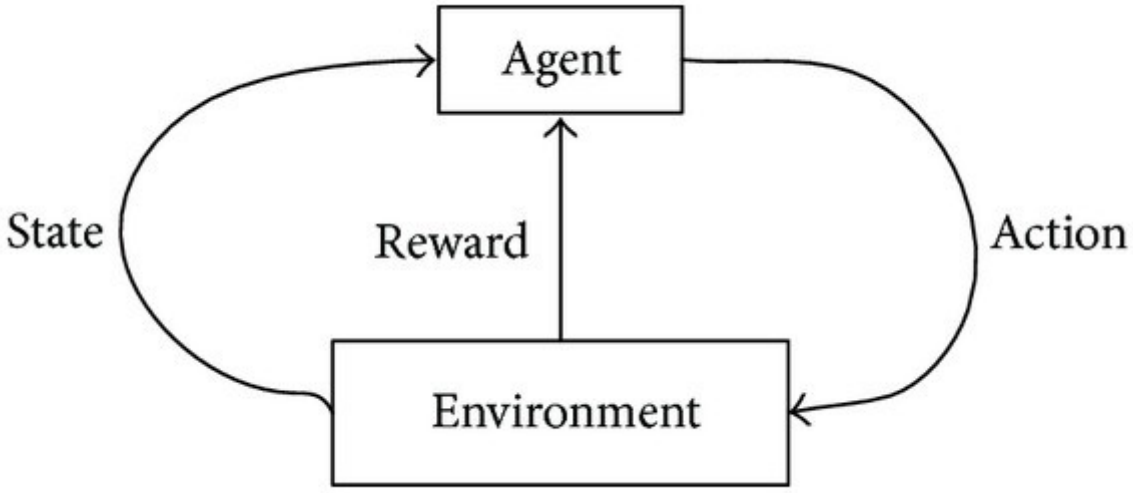

Fig.3: Reinforcement Learning [6].

The third most common form of machine learning, **reinforcement learning**, attempts to use observations gained from the interaction with its environment to adopt behaviors that maximize reward or reduce danger. In this scenario, the reinforcement learning algorithm (referred to as the agent) iteratively learns from its surroundings. Computers attaining superhuman levels and defeating humans in computer games are a wonderful illustration of reinforcement learning.

Machine learning, specifically its sophisticated sub-branches, such as deep learning and different forms of neural networks, may be fascinating. In any event, it is "magic" (Supercomputing Learning Theory), regardless of how difficult it is for the general public to observe its inner workings at times. While some people liken deep learning and neural networks to the human mind, others don't [6].

1.1.3 Deep Learning :

Deep learning is a machine learning technique that mimics how people acquire knowledge. Data science, which encompasses statistics and predictive modeling, incorporates deep learning as a key component. Deep learning is particularly useful for data scientists who are responsible with gathering, analyzing, and interpreting enormous volumes of data; it speeds up and simplifies the process [7].

Deep learning may be viewed of as a technique to automate predictive analytics at its most basic level. Deep learning algorithms are layered in a hierarchy of increasing complexity and abstraction, unlike standard machine learning algorithms, which are linear.

Consider a kid whose first word is "dog" to grasp the concept of deep learning. By pointing to items and repeating the word dog, the child learns what a dog is and is not. "Yes, it is a dog," or "No, that is not a dog," says the parent. As the toddler continues to point to things, he has a better understanding of the characteristics that all dogs have.

Without realizing it, the toddler is clarifying a complicated abstraction — the notion of dog — by creating a hierarchy in which each layer of abstraction is built on the information obtained from the previous layer.

The majority of the modern profound learning architecture is built on artificial neural networks(ANN), extracting features and altering them using several layers of non-linear units.

The output of the previous layer is used as the input for the next layer. What they learn is organized into a hierarchy of concepts, with each level learning to convert its incoming material into a more abstract and composite representation.

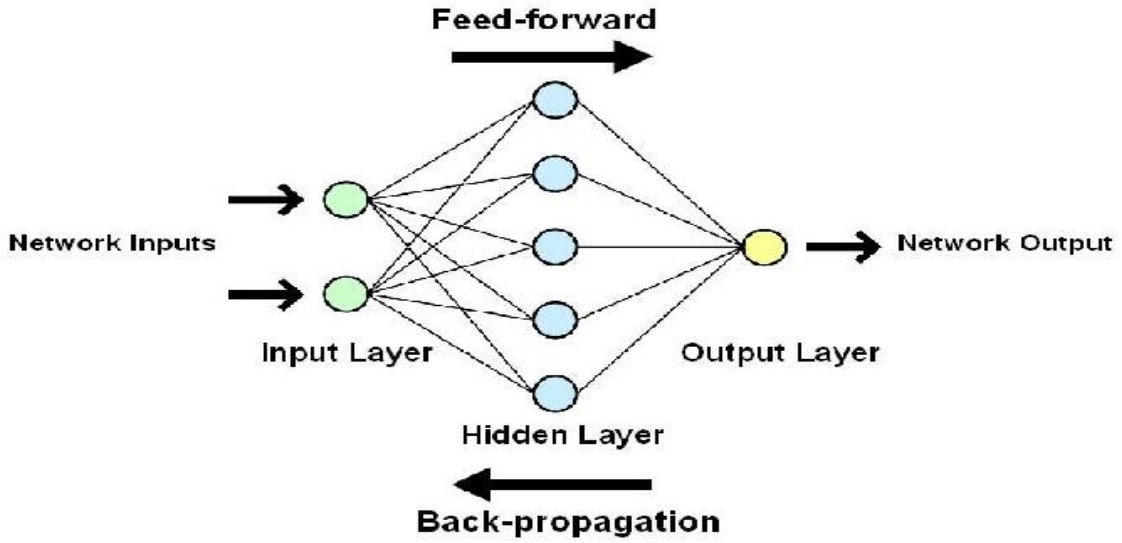

Fig.4: ANN (Artificial Neural Network) [7].

That is, the input for an image might be a matrix of pixels, the first layer could encode the edges and assemble the pixels, the next layer could compose a configuration of edges, the next layer could encode a nose and eyes, the next layer could detect that the picture has a face, and so on.

While some fine-tuning may be required, the deep learning method learns which characteristics to place in which levels all on its own!

1.1.1.1 Convolutional Neural Networks:

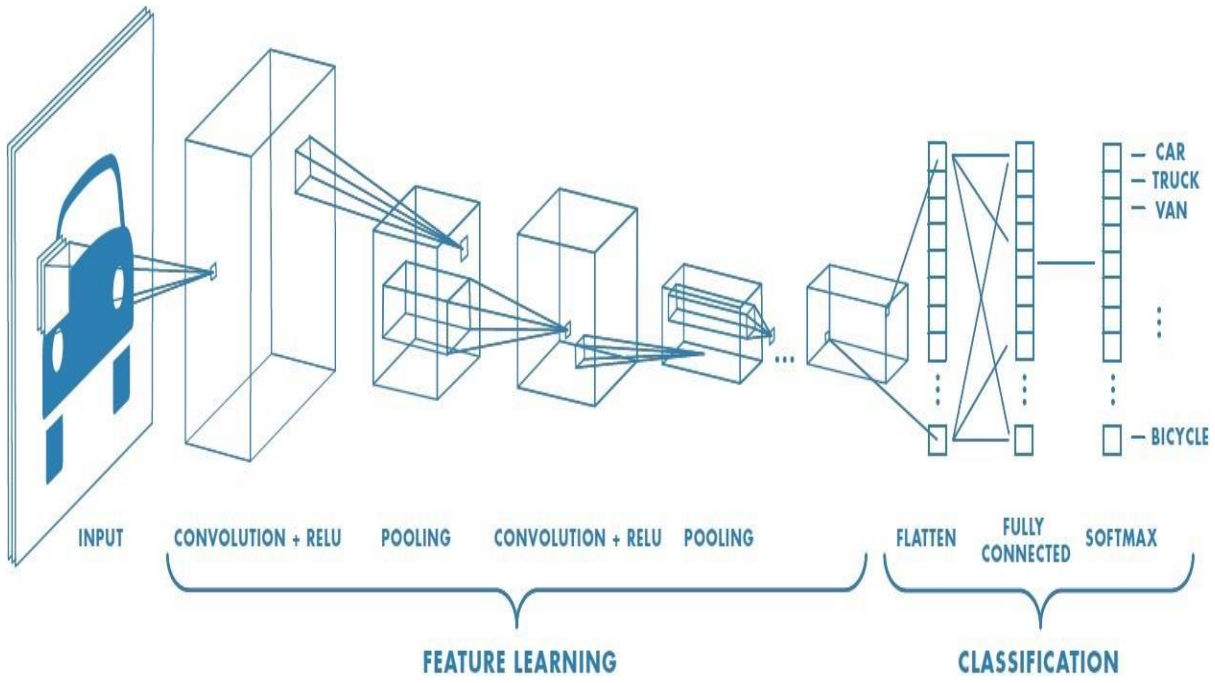

Fig.5: CNN (Convolutional Neural Network) [7].

A Convolutional Neural Network (ConvNet/CNN) is a Deep Learning system that can accept an input picture, assign significance (learnable weights and biases) to various attributes in the image, and distinguish between them.

When compared to other classification methods, the amount of pre-processing required by a ConvNet is significantly less. While basic techniques need hand-engineering of filters, ConvNets can learn these filters/characteristics with plenty of training [7].

A ConvNet's architecture is based on the structure of the Visual Cortex and is similar to the connection network of Neurons in the Human Brain. Individual neurons can only respond to stimuli in a small area of the visual field called the Receptive Field. A number of similar fields can be stacked on top of each other to fill the whole visual field.

There are a variety of CNN architectures available, many of which have played a role in developing algorithms that power and will continue to power AI in the near future, such as ResNet, VGGNet, and so on.

1.2 Computer Vision

Computer vision is a branch of computer science that aims to replicate elements of the human vision system's complexity, allowing computers to recognize and interpret things in pictures and videos in the same manner that people do. Recently computer vision could only do restricted tasks [8].

Artificial intelligence has made significant strides in recent years, surpassing humans in several tasks relating to object detection and categorization, thanks to improvements in artificial intelligence and developments in deep learning and neural networks.

Fig.6: YOLO Multi-Object Detection And Classification [8].

The quantity of data we create today, which is subsequently utilized to train and improve computer vision, is one of the driving forces behind its rise. Along with a massive volume of visual data (about 3 billion photos are uploaded on the internet every day), the computer power needed to evaluate it is now available. Object identification accuracy rates have risen as the field of computer vision has developed with new hardware and algorithms. At less than a decade, today's systems have increased from 50% accuracy to 95% accuracy, making them more efficient than humans in rapidly responding to processing images.

1.4.1 Applications Of Computer Vision

One of the areas of Machine Learning where basic principles are already being implemented into significant products that we use on a daily basis is computer vision [8].

I. **Self Driving Cars:**
   Machine Learning is used in picture applications by more than just IT businesses.
   Self-driving cars use computer vision to understand their environment. Cameras around the car record video from various angles and send it to computer vision software, which analyses the pictures in real time to locate road edges, read traffic signs, and recognize other cars, objects, and people. The self-driving car can then navigate its way through city

streets and highways, avoiding collisions and (hopefully) safely transporting its passengers to their destination [8].

II. **Facial Recognition:**

Face identification applications, which use computer vision to match photographs of people's faces to their identities, are another area where computer vision plays a key role. Facial characteristics in pictures are detected by computer vision algorithms, which then compare them to databases of face profiles. Facial recognition is used by consumer products to verify their owners' identities. Face recognition is used in social networking apps to identify and tag individuals. Face recognition technology is also used by law enforcement organizations to identify offenders in video feeds [8].

III. **Augmented Reality:**

Computer vision is particularly crucial in augmented and mixed reality, which allows computer devices like smartphones, tablets, and smart glasses to overlay and embed virtual items on real-world pictures. AR gear detects things in the real environment using computer vision to decide where a virtual object should be placed on a device's display. Computer vision algorithms, for example, may assist AR applications in detecting planes such as tabletops, walls, and floors, which is a crucial element of defining depth and dimensions and putting virtual items in the real environment [8].

1.4.2 Challenges of Computer Vision:

It turns out that assisting computers in seeing is really difficult. Creating a machine that sees as we do is a surprisingly tough endeavor, not just because it's difficult to make computers do it, but also because we don't fully understand how human vision works.

Understanding the perception organs, such as the eyes, as well as the interpretation of perception inside the brain, is necessary for studying biological vision. Much progress has been achieved, both in terms of documenting the process and identifying the system's tricks and shortcuts, albeit there is still a long way to go, as with any study involving the brain [8].

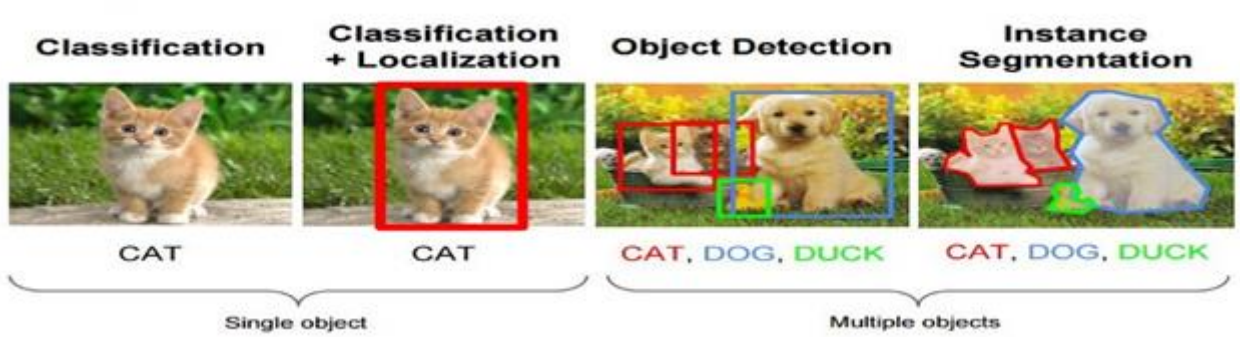

Fig.7: Computer Vision Tasks [8].

1.5 Fashion

Fashion is the embodied identity's cultural creation." Designers and couturiers describe "high fashion" as all forms of self-fashioning, including street style. Fashion relates to how things are created when it comes to manufacturing. Fashion is defined as the most popular style of clothing or conduct at any given period, according to the most prevalent definition [9].

Fig.8: Fashion designing (free stock image credits to pexels.com)[1]

Fashion design is the art of combining design, aesthetics, and natural beauty to create garments and accessories. Fashion design has evolved over time as a result of a number of cultural and societal factors. Fashion designers must occasionally anticipate shifting client desires due to the amount of time it takes to bring a product to market. Fashion designers strive to achieve both utility and aesthetic appeal. Consider who will be wearing the item and when they will be wearing it [9].

They may choose from a variety of materials and combinations, as well as colors, patterns, and styles. Many everyday clothes fall into a small number of traditional styles, but the most unique

[1] https://www.pexels.com/es-es/foto/persona-dibujando-vestido-en-papel-de-impresora-con-tablero-de-clip-1478477/

pieces are reserved for special occasions, such as evening wear or party dresses. Haute couture and bespoke tailoring are two places where you may find one-of-a-kind clothing. For most clothing, especially casual and everyday wear, mass-market fashion is becoming the standard.

Fashion and computer vision inter-relationship can be summarized in the following figure below (figure 9) which best describes the most important topics covered in research papers.

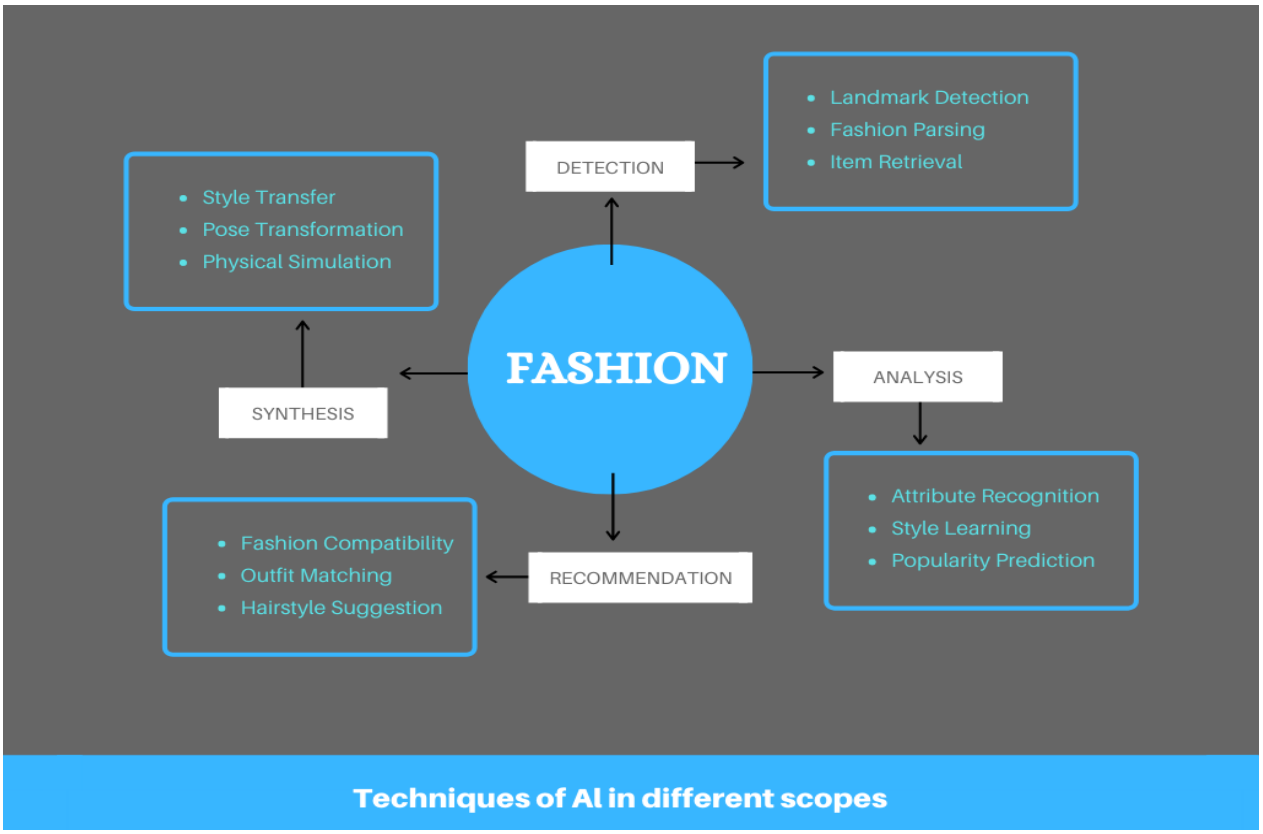

Fig.9: Most important Fashion topics covered in research papers

Four main topics including fashion detection, fashion analysis, fashion synthesis and fashion recommendation power the fashion industry researching area for the following titles we are going to cover them in details, analyze the current methods and evaluate them.

Chapter 2

Fashion Detection

Fashion detection aims to detect the human body part in the input image to determine the clothing region firstly, so it's basically the first step in any further extending work. In this section we are going to cover 3 tasks which are: landmark detection, item retrieval, and fashion parsing.

2.1 Landmark detection:

Fashion landmark detection predicts the key regional areas in clothes items like the edges of neck area, hemline, and sleeve. These landmarks describe the functional areas of clothes and also capture their bounding boxes. Landmark detection is a more challenging task than human pose estimation because clothes contain more deformation than the local regions of fashion landmarks have more significant regional and appearance variations than human body parts, Fig 10 shows the key differences:

Fig.10: Constrained and Unconstrained Fashion Landmark Detection

2.1.1:Datasets used in Landmark detection process:

Dataset name	Publish time	Number of photos	Number of landmark annotations
DeepFashion-C [10]	2016	289,222	8
Fashion Landmark Dataset (FLD) [11]	2016	123,016	8
Unconstrained Landmark Database [12]	2017	30,000	6
DeepFashion2 [18]	2019	491,000	Can reach more than 30

Table.1: Most popular datasets used fashion landmark detection task.

2.1.2 State of the art methods:

The idea of fashion landmarks was first proposed by Liu et al [11], they introduced deep fashion alignment (DFA) framework which is a deep learning model based on three stages shown in figure 11 below:

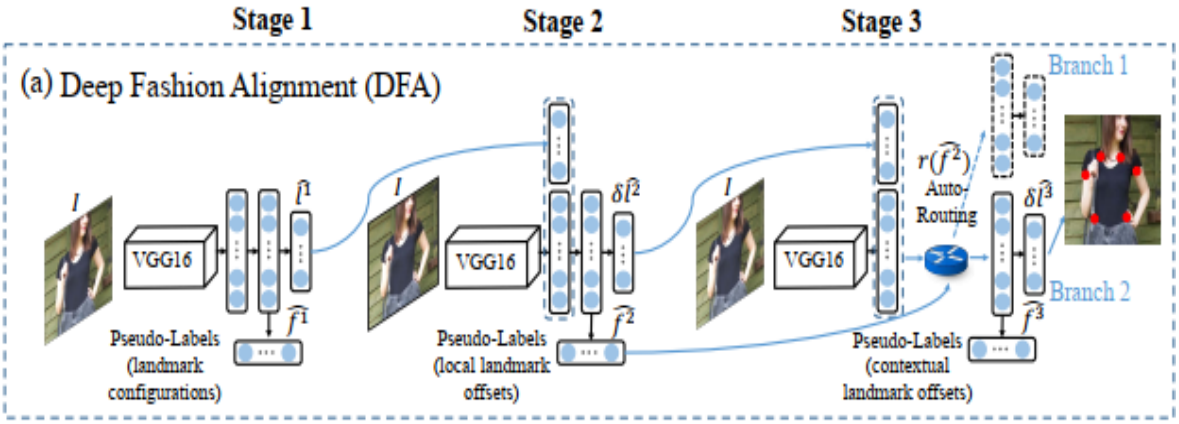

Fig.11: Deep Fashion Alignment(DFA) [11]

Therefore, each stage fulfills the expectations of the past. Yan et al. [12] further added more constraints on the clothing bounce frame, which is computationally expensive and irrelevant in practice. The proposed deep Landmark network (DLAN) integrates a specific expanded convolution and, more importantly, an overlapping spatial transformer, in which bounding boxes and clothes landmarks are iteratively evaluated and prepared from start to finish.

Liu et al [6] and Yan et al [7] both depended on the recurrence model. Later work Wang et al[17] showed that the recurrence model is very indirect and difficult to improve. They recommended providing deterministic guidance for the location of each feature, rather than returning directly to the location of the feature. Furthermore, they also considered fashion language to help reason about the location of landmarks. For example, "Left neck? Left waist? The left hem "association in the motion chain of the human body part is used as a requirement for the garment-related part to show the topology of their language. The human anatomical joints are obtained in an RNN(Recurrent Neural Network)

As for Ge et al [18] ,the adaptive benchmark Deepfashion2 was introduced for four tasks, namely: pose estimation clothing retrieval, human segmentation and clothes detection, covering the most critical style detection tasks. They built a powerful Match R-CNN model that relies on Mask R-CNN ,Kaiming He et al[19] to solve the problems of the previous tasks..

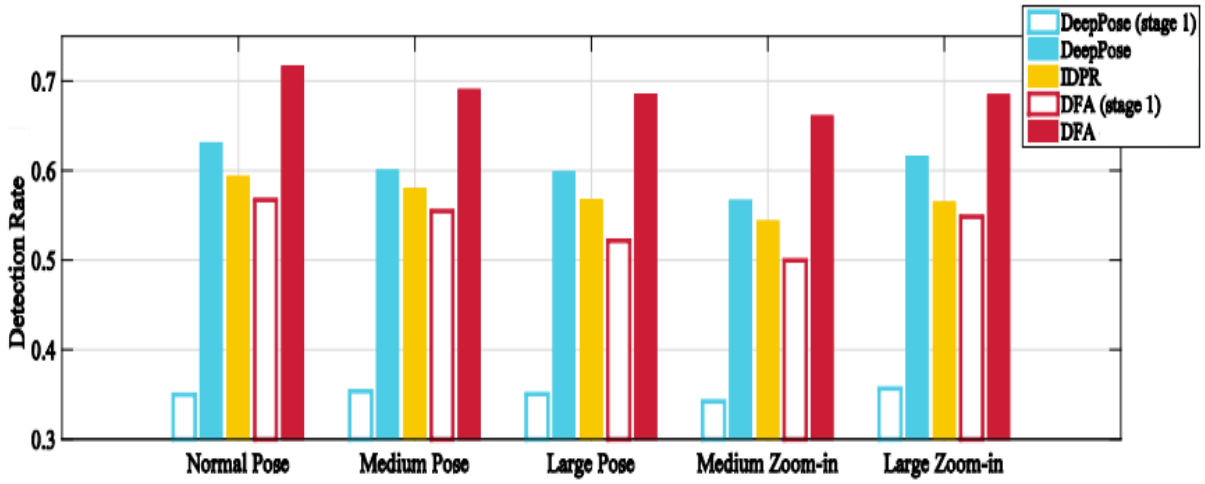

Fig.12: Detection rate based on camera position [11]

Method	Torso	Head	U.arms	L.arms	U.legs	L.legs	Overall
Pose Machines.[15]	93.1	83.6	**76.8**	68.1	42.2	**85.4**	72.0
DFA .[11]	90.8	87.2	70.4	56.2	80.6	75.8	74.4
IDPR[16]	92.7	87.8	69.2	55.4	82.9	77.0	75.0
+p.labels	93.5	88.5	72.3	59.0	83.9	78.7	77.0
+auto-routing	**94.1**	**88.9**	74.3	61.5	**85.1**	80.4	**78.6**

Table.2: Comparison between some recent papers concerning landmark detection accuracy [16] metric (NE)[21].

L:Left , R:Right , Hem :Hemline

Dataset	Method	L. Collar	R. Collar	L. Sleeve	R. Sleeve	L. Waistline	R. Waistline	L. Hem	R. Hem	Avg.
DeepFashion -C [10]	DFA [11]	0.062	0.063	0.065	0.062	0.072	0.070	0.065	0.066	0.066
	DLAN [12]	0.057	0.061	0.067	0.064	0.070	0.069	0.062	0.062	0.064
	AttentiveNet [17]	0.041	0.040	0.049	0.044	0.050	0.052	0.053	0.055	0.048
	Global-Local [13]	**0.031**	**0.032**	**0.042**	**0.043**	**0.036**	**0.037**	**0.044**	**0.047**	**0.039**
FLD [11]	DFA [11]	0.048	0.048	0.091	0.089	–	–	0.071	0.072	0.068
	DLAN [12]	0.053	0.054	0.070	0.073	0.075	0.074	0.069	0.067	0.067
	AttentiveNet [17]	0.046	0.047	**0.062**	**0.061**	0.063	0.069	0.063	**0.052**	0.058
	Global-Local [13]	**0.038**	**0.039**	0.067	0.067	**0.057**	**0.060**	**0.061**	0.062	**0.056**

Table.3: Evaluation of the accuracy of landmark detection methods using the Normalized Error

2.1.3:Evaluation of landmark detection state of the art methods

Fashion landmark detection methods produce the positions of landmarks (i.e., functional important points) in apparel pictures. The normalized error (NE), which is defined as the l2 distance in normalized coordinate space between detected and ground truth landmarks, is the most often utilized assessment metric in fashion landmark identification benchmarks. Typically, smaller NE values imply better outcomes. Table 3 compares the performance of prominent techniques on benchmark datasets. Furthermore, the performance of the same technique varies among datasets, but the rank is typically consistent.

2.2 Fashion Parsing:

Fashion parsing is a particular type of semantic division, where the classes are one of the clothes items, for example, shirt , short , pants etc. Clothes parsing has been effectively considered in computer vision as one of the challenges because of its importance to the fashion industry to create new trends , and furthermore as a result of its gigantic worth in reality application. Dressing is a fundamental piece of our way of life. The categorization of clothes classes requires a high level of understanding of the clothes semantics due to the large possible number of garment items and their variations.

2.2.1 Datasets used in Fashion Parsing process:

Dataset Name	Publish Year	Number of Pictures	Number of classes	Source
Fashionista dataset[14]	2012	158,235	56	Chicotopia.com
MHP(v1.0 [22], v2.0[23])	2017,2018	4,980 ; 25,403	18 ; 58	N/A
DeepFashion2 [18]	2019	491,000	13	DeepFashion
Fashionpedia[24]	2020	48,000	46	Flickr,Free license photos

Table.4: Most popular datasets used datasets for Fashion Parsing task

2.2.2 State of the art methods:

The first work concerning clothes parsing was introduced by Yamaguchi et al [14] , firstly they use an image segmentation algorithm to obtain super pixels it gives around a few hundred to a thousand region per image , then they did a human pose estimation and then the clothes labels are predicted in a Conditional Random Field(CRF) model as shown in figure 13 below .

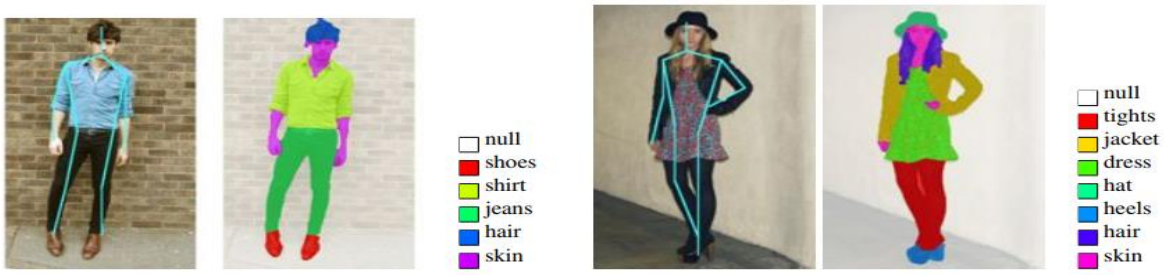

Fig.13: clothes parsing by Yamaguchi et al [14]

Their work, in any case, for the most part centered around a compelled parsing issue, where test pictures were parsed by the user demonstrating portrayed clothes items. To defeat this restriction, [20] Yamaguchi et al proposed dresses parsing with a recovery based methodology. For a given picture, comparative pictures from a parsed dataset were first recovered, and afterward the closest neighbor parsing were moved to the eventual outcome by means of thickness.

Unlike prior approaches, which only addressed single-person parsing tasks, Zhao et al. [23] proposed a deep Nested Adversarial Network with three Generative Adversarial Network (GAN)-

like sub-networks for semantic noticeability prediction, instance-aware clustering. These three sub-networks made a learning process side by side.

Hierarchical graphs were studied for human parsing tasks in 2019. Wang et al. [25] characterized the human body as a hierarchical structure of multi-level semantic elements and used three techniques (direct, top-down, and bottom-up) to collect human parsing information for improved parsing efficiency.

2.2.3: Evaluation of fashion parsing state of the art methods:

For evaluating the fashion parsing task there are a lot of accuracy metrics that can be used we mention some of them which are :

1.Average recall: recall is defined as the ratio tp / (tp + fn), where tp is the number of true positives and fn is the number of false negatives. It is intuitively the classifier's capacity to find all positive samples.

2.Average precision: AP (Average precision) is a widely used metric for assessing the accuracy of object detectors such as Faster R-CNN, SSD, and others. The average accuracy value is computed for recall values ranging from 0 to 1.

3. Intersection over Union (IoU): IoU calculates the amount of overlap between two borders. We utilize this to determine how much our projected border overlaps with the actual boundary (the real object boundary).

2.3: Item Retrieval:

As fashion e-commerce has developed over the years, there has been a strong demand for new solutions to assist clients in quickly finding their favorite fashion goods. Despite the fact that many fashion online buying sites support keyword searches, many visual characteristics of fashion items are difficult to convert into words. As a result, numerous research communities are focusing on developing cross-scenario image-based fashion retrieval tasks for matching real-world fashion goods to online purchasing images. The purpose of picture-based fashion item retrieval is to retrieve comparable or identical products from the gallery given a fashion image query.

2.3.1: Datasets used in Item Retrieval process:

Dataset Name	Publish Year	Number of Pictures	Source
Dress like a star [41]	2017	7,000,000	Youtube.com
DeepFashion2 [18]	2019	491,000	DeepFashion
Ma *et al.* [29]	2020	180,000	DeepFashion [11]

Table.5: Most popular datasets used datasets for Item Retrieval task

2.3.2 State of the art methods:

Building deep neural network designs to address the fashion apparel retrieval challenge has become popular as deep learning technology has advanced. Using attribute-guided learning, Huang et al. [26] created a Dual Attribute aware Ranking Network (DARN) to represent in-depth features.

DARN incorporated semantic attributes and visual similarity restrictions into the feature learning stage at the same time, while simulating the domain asymmetry.

Wang et al. [30] used a Siamese network with two layers for the deep feature representation, which contains copies of the Inception-6 network with identical weights. They also had a strong contrastive loss to reduce overfitting caused by several positive pairs (having the same product) that were observed visually distinct, and employed a single multi-task fine-tuning strategy to build a better feature representation by modifying the parameters of the Siamese network with product and generic photos from ImageNet[27]. In the clothing retrieval methods mentioned above, retrieval is based solely on query photos that reflect users' desires, with no consideration given to the possibility that users may want to contribute additional keywords.

The most recent work concerning Item retrieval was presented by Lin et al[31] which achieved the highest score in the 2019 challenge. They introduced an unsupervised embedding learning method for training a CNN model, as well as a combination of existing retrieval methods learned on diverse datasets to fine-tune the retrieval results as shown in figure 14 below:

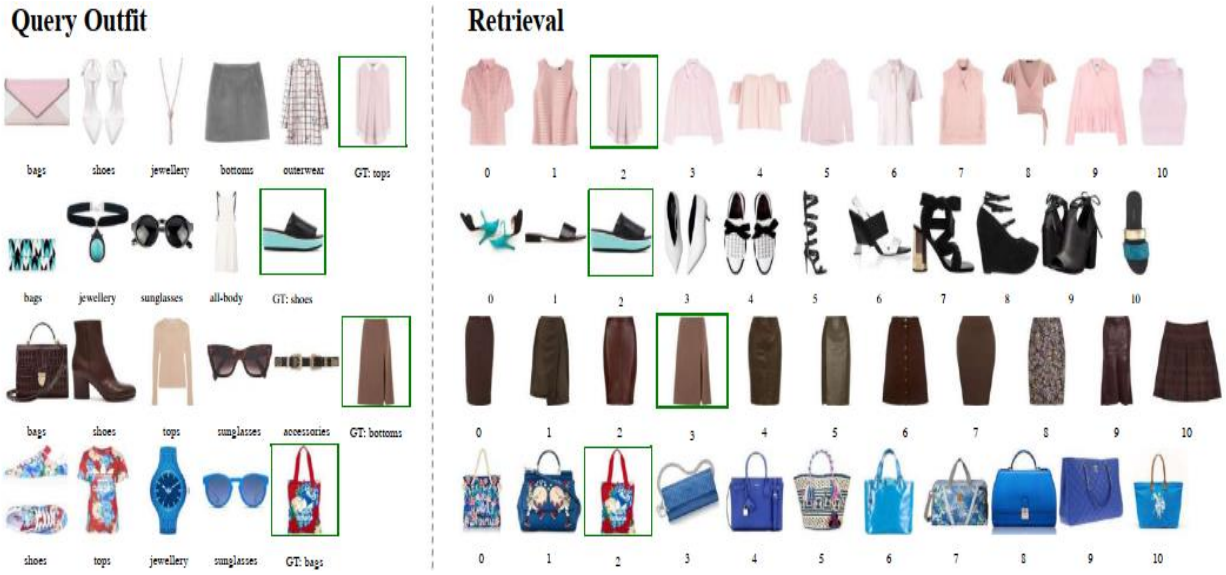

Fig.14: Item retrieval example presented by Lin et al [31]

***FITB**:fill in the blank task (item search)* ***Compat.AUC**: Outfit compatibility task(outfit matching)*

		Polyvore Outfits-D		Polyvore Outfits	
Method	**Feature**	**FITB Accuracy**	**Compat. AUC**	**FITB Accuracy**	**Compat. AUC**
Siamese-Net[32]	ResNet18	51.80	0.81	52.90	0.81
Type-Aware[32]	ResNet18+Text	55.65	0.84	57.83	0.87
SCE-Net average[33]	ResNet18+Text	53.67	0.82	59.07	0.88
CSA-Net + outfit ranking loss [31]	ResNet18	**59.26**	**0.87**	**63.73**	**0.91**

Table.6: Comparison Of different methods on the Polyvore-Outfits Dataset

2.3.3: Evaluation of item retrieval state of the art methods:

For evaluating the item retrieval task there are some accuracy metrics that can be used we mention some of them which are:

1.Top-N retrieval accuracy: standard accuracy of the true class being equal to any of the N most probable classes predicted by the classification model.

2.Normalized Discounted Cumulative Gain (NDCG@*k*), is a ranking quality metric. It is frequently used to evaluate the performance of online search engine algorithms or similar applications in the field of information retrieval. DCG evaluates the utility, or gain, of a document based on its position in a search engine result list, using a graded relevance scale of items in the result set. The gain is accumulated from the top to the bottom of the result list, with lower ranks discounting the gain of each result.

Chapter 3

Fashion Analysis

Fashion is not about what people wear, it reflects character and other social characteristics. Intelligent fashion analysis of models selected by people with great potential in the fashion industry, specific marketing, analytical sociology, etc. has been gaining attention in recent years.

In this section we are going to cover 3 tasks which are: attribute recognition, style learning, and popularity prediction.

3.1 Attribute Recognition:

Recognition of clothing attributes is a problem of classifying several features, the purpose of which is to determine which elements of clothing are related to the attributes between a set of attributes as shown in figure 15 below:

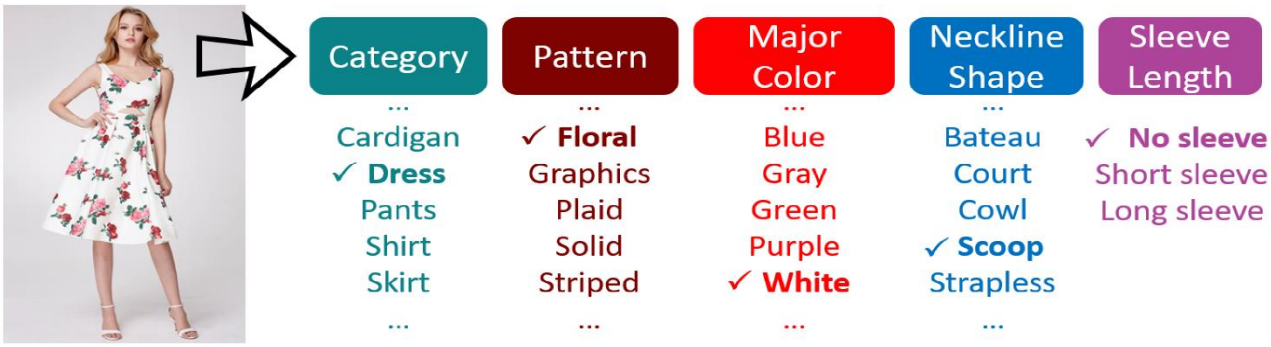

Fig.15: Fashion attribute recognition example

3.1.1:Datasets used in Attribute Recognition process:

Dataset	Publish time	# of photos	# of categories	# of attributes	Sources
DeepFashion-C [10]	2016	289,222	50	1,000	Google Images
Fashion 200K [35]	2017	209,544	5	4,404	Lyst.com
Hidayati et al. [34]	2018	3,250	16	12	shopping sites
CatalogFashion-10x [38]	2019	1M	43	N/A	Amazon.com

Table.7: Most popular datasets used datasets for Attribute Recognition task

3.1.2 State of the art methods:

The basic model of CRF was presented in [36] Yamaguchi et al. The model evaluated the appearance of the specific location in relation to the human body and the compatibility of the elements and features of the clothes, which were trained using a learning frame with a maximum margin.

Stimulated by the huge differences between images taken in controlled and uncontrolled environments, Chen et al. [37] studied cross-domain feature mining. They extracted information from clean clothing for images obtained from online stores and then adapted to an unlimited environment using deep domain customization access.

Han et al. [40] Derived from the spatial representation of each attribute by multiplying the attributes by spatial representation.

	Category		Texture		Fabric		Shape		Style		All	
Method	**Top-3**	**Top-5**	**Top-3**	**Top-5**	**Top-3**	**Top-5**	**Top-3**	**Top-5**	**Top-3**	**Top-5**	**Top-3**	**Top5**
DARN [26]	59.48	79.58	36.15	48.15	36.64	48.52	35.89	46.93	66.11	71.36	42.35	51.95
FashionNet [10]	82.58	90.17	37.46	49.52	39.30	**49.84**	39.47	48.59	66.43	73.16	45.52	54.61
AttentiveNet [17]	**90.99**	**95.78**	**50.31**	**65.48**	**40.31**	48.23	**53.32**	**61.05**	**68.70**	**74.25**	**51.53**	**60.95**

Table.8: Evaluation of the accuracy of attribute recognition methods with the top-N classification metric[21].

3.2.1.3:Evaluation of attribute recognition state of the art methods

Attribute recognition is often evaluated with the top N accuracy is not any different metric, but it's just standard accuracy of the true class being equal to any of the N most probable classes predicted by the classification model. Top 1 accuracy is the accuracy where the true class matches with the most probable classes predicted by the model, which is the same as our standard accuracy. Top 2 accuracy is the accuracy where true class matches with any one of the 2 most probable classes predicted by the model. Top 3 accuracy is the accuracy where true class matches with any one of the 3 most probable classes predicted by the model.

3.2 Style Learning

For fashion experts, fashion style analysis is extremely important. There is also a problem with diverse classification standards, which rely primarily on the subjective experiences of experts who have no quantitative criterion. Recognition of clothing photos in the fashion style helps Ecommerce clothing recovery and suggestions. It is a difficult process, as clothes photos of the same type may seem different. Existing techniques of identification of fashion style use deep neural networks to categorize pixel or regional imagery, style learning can be summarized by figure 16 below as described by Wang et al[41].

Fig.16: Style Learning example by Wang et al[41]

3.2.1:Datasets used in Style Learning

Dataset	Publish time	Number of photos	Sources
Fashion Data [42]	2016	590,234	Polyvore.com
Street Fashion Style (SFS) [43]	2017	293,105	Chictopia.com
Geostyle [44]	2019	7.7M	Street Style, Flickr100M
FashionKE [45]	2019	80,629	Instagram

Table.9: Most popular datasets used datasets for Style Learning task

3.2.2: State of art methods in Style Learning

Specifically, Simo-Serra and Ishikawa [46] adapted a joint positioning and arrangement structure dependent on the Siamese organization. The proposed structure had the option to accomplish

extraordinary execution with highlights being the size of a SIFT(scale-invariant feature transform) descriptor.

Vaccaro et al. [42] introduced an information driven style model that took in the correspondences between undeniable level style depictions (e.g., "valentine's day" and low-level plan components (e.g., "red pullover" via preparing polylingual point demonstrating. This model adjusted a characteristic language preparing procedure to learn idle design ideas together over the style and component vocabularies.

Further, intriguing work for learning the client driven style data in view of events, clothing classes, and qualities was presented by Ma et al. [45]. Their fundamental objective is to get familiar with the data about "what to wear for a particular event?" from online media, e.g., Instagram. They fostered a contextualized design idea learning model to catch the conditions among events, clothing classifications, and traits.

Style Analysis. An exploration pioneer in programmed style examination was introduced by Hidayati et al. [40]. They researched style at ten unique periods of New York Fashion Week by dissecting the rationality (to happen as often as possible) and uniqueness (to be adequately extraordinary from other design shows) of visual style components.

Afterward, Gu et al. [43] introduced QuadNet for investigating style from street photographs. The QuadNet was a characterization and highlight installing learning network that comprises four indistinguishable CNNs, where the common CNN was together upgraded to perform various tasks characterization loss function and a neighbor-compelled quadruplet loss function.

Also, Chang et al [47] accomplished a fascinating work on what individuals decide to wear in various urban communities.

The suggested "Style World Map'' framework made use of a collection of geo-labeled road design pictures from Lookbook.nu, a picture-oriented online media website. They devised a measurement based on deep neural networks to select the potential famous outfits for each city, named the prize-gathering Steiner tree (PCST) issue, and named the location issue of notorious design objects of a city as the prize-gathering Steiner tree (PCST) issue.

Mall et al. [44] provided a comparative idea, establishing a framework for assessing global style in thin credits and style in comparison to city and season. They also analyzed what events signify for people to dress for, for example, "new year in Beijing in February 2014" suggests red upper clothes. The goal of general style investigation is to determine when a style was created, and temporary assessment is a fascinating job. Despite the fact that visual analysis of design styles has been extensively studied, this topic has received little attention from the exploration community.

3.2.3 Evaluation of Style Learning state of the arts methods

Precision, recall, and accuracy are the assessment criteria used to assess the performance of current fashion style learning techniques. Basically, precision measures the fraction of significant retrieved results, recall represents the number of relevant retrieved results, and accuracy measures the percentage of right recognition. Despite the fact that numerous approaches to fashion temporal analysis have been offered, there has been no systematic comparison of them. It is an open question which of the ways performs better than others.

3.3:Popularity Prediction

Accurately predicting fashion trends is not only important for fashion brands pursuing global marketing activities, but also for people who decide what to wear based on specific circumstances (blogs, trends etc..).can better predict fashion popularity and further predict future trends, which will greatly affect the US$ billion fashion industry.

The fashion industry evolves with time. The popularity of two types of fashion trends (velvet and off-shoulder) has changed over time, as shown in the graph, with the numbers derived from the GSFashion dataset [47] as shown in figure 9 below:

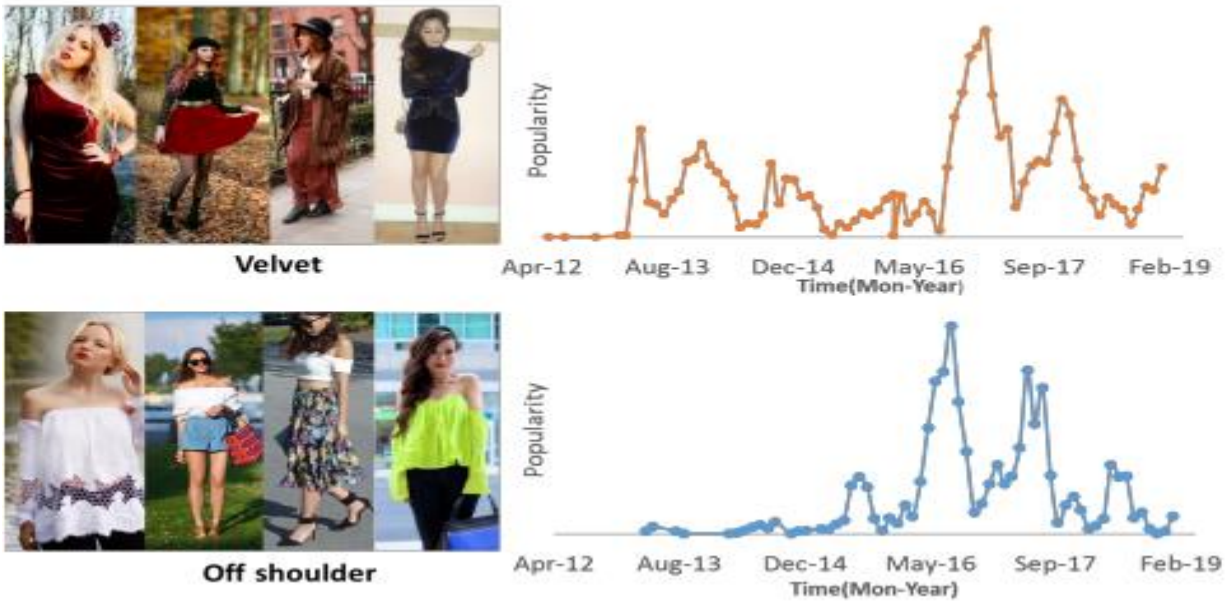

Fig.17: Popularity of 2 different trends over time [47]

3.3.1:Datasets used in Popularity Prediction

Dataset	Publish time	Number of photos	Source
TPIC17 [48]	2017	680,000	Flickr.com
Lo *et al.* [49]	2019	380,000	lookbook.nu
SMPD2019 [50]	2019	486,000	Flickr.com

Table.10: Most popular datasets used datasets for Popularity Prediction task

3.3.2: State of the art methods in Popularity Prediction:

To get around the necessity for rating history for the query, which previous efforts couldn't handle when there was none or only a few, Rothe et al. [51] proposed regressing visual queries to a latent space obtained using matrix factorization for known subjects and ratings. In addition, for attractiveness prediction, they used a visual regularized collaborative filtering approach to infer inter-person preferences.

To additionally think about the facial shape, Gao et al. [52] planned to perform various tasks learning structure that considered appearance and facial shape at the same time and mutually scholarly facial portrayal, milestone area, and facial appeal score.

They demonstrated that learning with milestone limitations is viable for facial engaging quality expectation. For building adaptable channels to gain proficiency with the planning versatile for various traits inside a profound modular, Lin et al. [53] proposed a quality mindful convolutional neural network (AaNet) whose channel boundaries were controlled adaptively by facial credits. They likewise viewed the cases without trait names furthermore, and introduced a attribute-aware convolutional neural network (AaNet), which figured out how to use picture setting data for creating trait-like information.

There was a Social Media Prediction (SMP) challenge held by Wu et al. [50] in ACM Multimedia 2019 which focused on the work zeroed in on foreseeing future snaps of new web-based media posts before they were posted in friendly feeds. The partook groups need to construct another algorithm based on comprehension and learning procedures and naturally anticipate fame to accomplish better exhibitions.

3.3.3 Evaluation of Popularity Prediction state of the art methods:

The most commonly used measures to evaluate popularity detection accuracy are: MAE which is a measure of the difference in error between two observations depicting the very same phenomena, MAPE which in statistics is a metric of a predicting method's forecasting accuracy, SRC measures the strength and direction of association between two ranked variables and MSE calculates the average of the error squares.

Chapter 4

Fashion Synthesis

Fashion synthesis is a complex job that includes putting a baseline clothing on a set of data which is in an indeterminate pose and/or is wearing a different outfit. We can picture what a person would look like with different cosmetics or dress styles based on an image of them This can be accomplished by creating an almost real image so similar to reality. In this section we're going to cover 3 major fashion synthesis tasks which are physical transformation, style transfer and pose transformation.

4.1: Style Transfer

Converting an image as input into a similar output image, like a real photo into a cartoonish drawing, a non-makeup face photo into a cosmetic transformed image, or clothing attempted on inputted image from one style to a different one, is known as style transfer. Image processing applications for style transfer include things like facial cosmetics and virtual try-ons.

Figure 18 shows the style transfer task by Wang et al[54]:

Fig.18: Style transfer task using 2 input images by Wang et al[54] .

4.1.1: Datasets used in the Style Transfer process

Dataset	Dataset name	Publish time	# of photos	Sources
Makeup	LADN [55]	2019	635	The Internet
	Makeup-Wild [56]	2020	772	The Internet
Virtual Try On	DeepFashion [10]	2016	78,979	Forever21
	VITON [57]	2018	32,506	N/A
	FashionTryOn [58]	2019	28,714	Zalando.com
	Video Virtual Try-on [59]	2019	791 videos	fashion model catwalk

Table.11: Most popular datasets used datasets for Style Transfer task

4.1.2: State of the art methods in Style Transfer

Pix2pix [59], a universal solution for style transfer, is the most popular style transfer task. It not only learns the mapping from the input image to the output image, but also the mapping from

the output image to the input image. Not only is there an output image, but there is also a loss function to train the mapping.

Facial Makeup.

Finding the appropriate make-up for a specific human face is difficult, given the reality that make-up fashion varies from face-to-face because of the distinct facial features. Studies on the way to robotically synthesize the results of without or with make-up on one's facial look have aroused a hobby recently.

To cope with this issue, Alashkar et al. [54] trained a deep neural community primarily based on totally make-up advice from examples and understanding base guidelines jointly. The suggested make-up fashion became then synthesized at the challenge face. Different make-up patterns bring about sizable facial look changes, which brings demanding situations to many sensible applications, consisting of face recognition.

Li et al. [58] later brought a bi-degree adversarial network structure, in which the primary opposed scheme became to reconstruct face photographs, and the second one became to hold face identification.

For achieving higher make-up and de-make-up overall performance, Gu et al. [55] targeted on local facial info switch and designed a LAN(local adversarial network) which contained more than one and overlapping LANs..

The most recent work concerning facial makeup transfer was introduced by Ngyuen et al [60] as shown in figure 11 which shows some great visual results almost identical to the real make up process.

Fig.19: Images before and after putting makeup on [60].

Virtual Try-On.

In the following, we overview current strategies and datasets for addressing the trouble of producing photographs of humans in apparel with the help of focusing at the patterns.

Han et al. [57] applied a rough-to-great strategy. Their framework, VIrtual Try-On Network (VITON)targeted on attempting an in-keep apparel photo on someone photo. It first generated a rough tried on end result and anticipated the masks for the apparel item. Based on the masks and coarse end result, a refinement neural network for fine tuning the end result of the first stage output image.

However, [57] fails to deal with huge deformation, especially with extra texture info, because of the imperfect form-context matching for aligning garments and frame form. Therefore, a brand-new version referred to as Characteristic-Preserving Image-primarily based totally on Virtual Try-On Network (CP-VTON) [40] became proposed. The spatial deformation may be higher dealt with the aid of using a Geometric Matching Module, which explicitly aligned the input apparel with the frame form.

There had been numerous stepped forward works primarily based totally on CP-VTON. Different from the preceding works which wished the in-keep apparel photo for digital attempt-on, FashionGAN [62] provided goal attempt-on apparel photo primarily based totally on textual content description and version photo respectively. Given an input photo and a sentence describing a distinct outfit, FashionGAN became capable of "redressing" the person.First, a segmentation map was created with a GAN in accordance with the description. Then, the output photo became rendered with every other GAN guided with the aid of using the segmentation map. M2E-TON became capable of attempt on apparel from human$A$ photo to human$B$ photo, and humans can carry out in distinct poses.

Considering the runtime efficacy, Issenhuth et al. [63] proposed a parser loose digital attempt-on neural networks. It designs a teacher-pupil structure to lose the parsing process for the duration of the inference time for enhancing efficiency.

Viewing the attempt-on overall performance from distinct perspectives is likewise vital for digital attempt-on challenge, Fit-Me [64] became the primary focus to do digital attempt-on with arbitrary poses. They designed a coarse-to-great structure for each pose transformation and digital attempt-on.

Further, FashionOn [65] carried out the semantic segmentation for targeted element-degree getting to know and targeted on refining the facial element and apparel place to give extra sensible results. They succeeded in keeping targeted facial and apparel data, carry out unusual posture, and additionally solve the human limb occlusion trouble in CP-VTON.

Similar structure to CP-VTON for digital attempt-on with arbitrary poses was provided with the aid of using Zheng et al's [58]. They in addition made frame form masks prediction at the start of

the primary degree for pose transformation, and, withinside the 2nd degree, they provided an attentive bidirectional GAN to synthesize the very last end result.

For pose-guided digital attempt-on, Dong et al. [59] in addition stepped forward VITON and CP-VTON, which tackled the digital attempt-on for distinct poses. Han et al. [66] proposed ClothFlow to be aware of the apparel areas and version the arrival goes with the drift among supply and goal for shifting the arrival certainly and synthesizing novel end result.

The most recent work concerning the virtual try-on process was introduced by Luo et al[61]

as shown in figure20 which shows an overall good visual performance when redressing input image with less to no deformities at all.

Fig.20: Luo et al Virtual Try-on results (2021) [61].

4.1.3: Evaluation of Style Transfer state of the art methods:

The majority of style transfer evaluations are based on personal judgment or user research. That being said, the volunteers provide grades to the results, like "very terrible," "terrible," "okay," "good," and "outstanding." The quality of the outputs is then quantified by calculating the percentages of each grade. In addition, objective evaluations for virtual try-on exist, such as inception score (IS) or structural similarity (SSIM). IS is a quantitative method for evaluating image synthesis performance. If the model can generate visually various and conceptually significant images, the grade will be greater. SSIM, from the other side, is used to determine the resemblance between both the input images and the target image.

4.2 Pose Transformation

The purpose of pose transformation is to take a reference image and turn it into a target pose with only key points. to combine a silhouette person image in various postures while maintaining personal information, figure 21 shows a few examples of pose change. Pose changing, in particular, is a challenging task. Because the input and output are not perfectly coordinated, this is a difficult process.

Fig.21: Example of Pose Transformation (left figure) and Pose Estimation (right figure)

4.2.1 Datasets used in the Pose Transformation process

Dataset name	Publish time	Number of photos	Sources
Human3.6M [67]	2014	3.6M	professionally made
DeepFashion [10]	2016	52,712	Shopping sites
Balakrishnan *et al.* [39]	2018	N/A	YouTube

Table.12: Most popular datasets used datasets for Pose Transformation task

4.2.2 State of the art methods in Pose Transformation process

PG2 [68], A 2 phased GAN(Generative adversarial network), made an early approach at this problem. In the first phase, a rough image was created under the desired pose, which was then polished in the second phase. With 2 benchmark datasets, Fig22. (a)-(b) shows the intermediate and final outcomes. However, the results were incredibly blurry, particularly when it came to texture details.

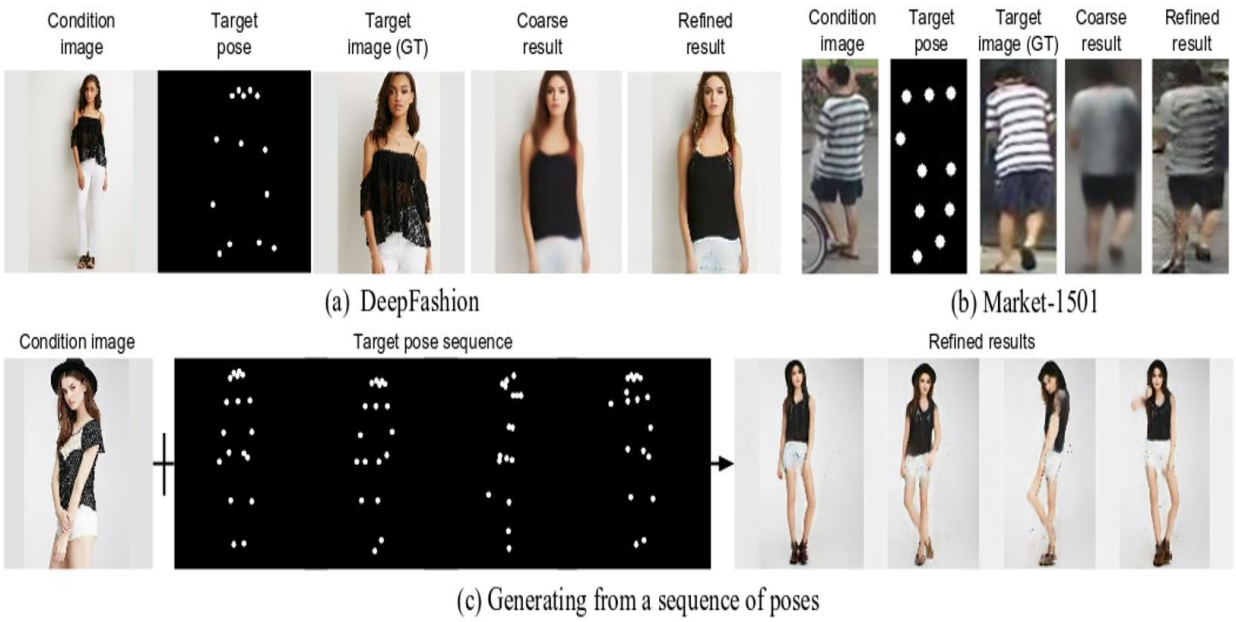

Fig.22: Comparison between DeepFashion and Market-1501 datasets on pose transformation.

Likewise, body part extraction masks were used to control image production in [39]. The suggested framework included four major components that could be trained together: source picture segmentation, positional transformation, foreground synthesizing, and background synthesizing.

Meanwhile, some studies defined the problem in terms of variational auto-encoders (VAE). They were able to effectively simulate the body shape; but, because they created outputs from compressed characteristics sampled from the distribution of data, their results have been less accurate to the appearance of source photos.

To Enhance the quality of the output, Song et al. [69] introduced a novel method for breaking down the challenging mapping into 2 doable subtasks: semantic parsing transformation and visual generation.

To begin, it changed the pose in semantic parsing maps to improve non-rigid deformation learning. The actual end results were then generated by integrating semantic-aware human input to the previously synthesized semantic maps.

The most recent work concerning the pose transformation task was proposed by Cui et al[70] which shows a great output quality with less deformation when changing the posture of input images ash shown in figure 23 below:

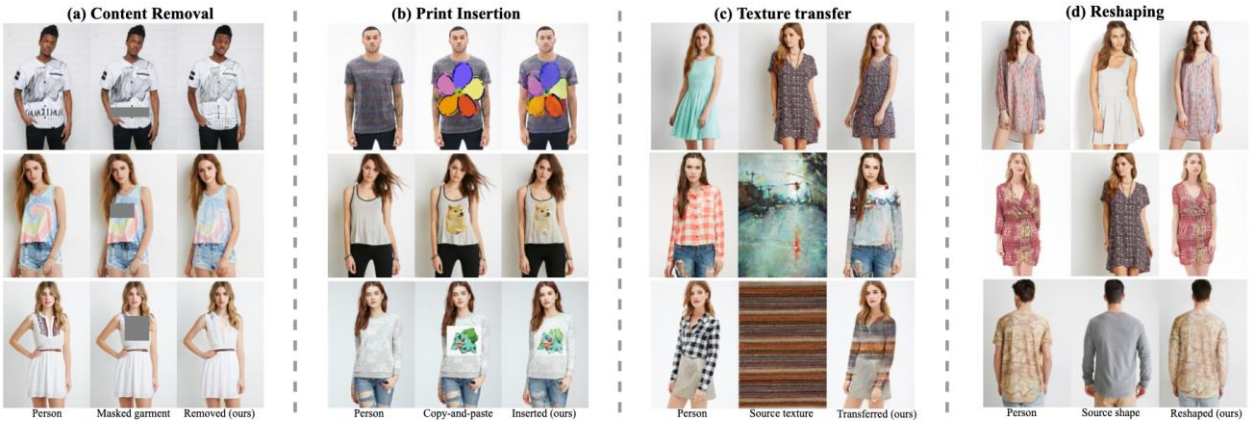

Fig.23: Results of pose transformation by Cui et al [70]

Compared method	Task	Prefer others vs [65]
GFLA[71]	pose transfer	47.73% vs **52.27%**
ADGAN[72]	pose transfer	42.52% vs **57.48%**
ADGAN[72]	virtual try on	19.36% vs **80.64%**

Table.13: User study results comparing [70] with the earlier pose transformation methods.

4.2.3 Evaluation of Pose Transformation state of the art methods:

Pose transformation methods are usually evaluated by SSIM (structural similarity metric) and Fréchet inception distance (FID) ,Learned Perceptual Image Patch Similarity (LPIPS) metric and Intersection-over-Union (SIoU) as shown in table 14.

	Size	SSIM	FID	LPIPS	sIoU
Def-GAN* [73]	82.08M	-	18.46	0.233	-
Intr-Flow* [74]	49.58M	-	16.31	**0.213**	-
GFLA* [71]	14.04M	0.713	**10.57**	0.234	57.32
DiOr [75]	24.84M	**0.725**	13.10	0.229	**58.63**

Table.14: Comparison of some recent Pose Transformation methods

4.3 Physical Simulation

Physical simulation is essential for even more realistic fashion synthesis efficiency. The above-mentioned synthesis works are limited in their modeling of physical deformation, such as shadow, fold, or hair features, to the 2D space. There are physical simulation efforts based on 3D datasets for improving synthesis performance with variable details (clothing-body dynamics).

Wang et al. [75] determined the essential physical properties of the intended clothing and keyframes (indicated in yellow). qualities and applied them to other frames with various poses, as illustrated in figure 24 below.

Fig.24: Physical simulation as represented by Wang et al [75]

4.3.1 Datasets used in Physical Simulation process

Dataset name	Publish time	Number of photos	Sources
DeepWrinkles [76]	2018	9,213	N/A
Santesteban *et al.* [77]	2019	7,117	SMPL
TailorNet [78]	2020	55,800	Simulated by designers
Sizer [79]	2020	N/A	self-collected

Table.15: Most popular datasets used datasets for Physical Simulation task.

4.3.2 State of the art methods in Physical Simulation process

The typical method for developing and modeling realistic clothing is to employ computer software to create 3D models and display the resulting visuals. Wang et al[75] In real textile samples, bending is possible. In order to understand about the physical qualities of clothing Guan et al. [79] created a pose-dependent model based on distinct human body forms and positions. Simulating

clothing deformation For the purpose of imitating ordinary clothing on fully clad people in movement.

To begin, it isolated various clothes from the body in order to estimate the dressed body structure and pose beneath the apparel. Then, using 4D images, it monitored the 3D malformations of the garment over time to assist imitate physical clothing deformations in diverse postures.

Lähner et al. [76] suggested a novel structure comprised of two complementary modules to improve the authenticity of the clothing on the human body: (1) Relying on the 3d scanning of clothed humans in movement, a statistical model learnt to align garment templates, and a linear sub model factored in human body shape and posture. (2) A cGAN enhanced normal maps with greater geometric information and emulated the physical world.

Santesteban et al. [77] developed a two-level training-based clothing animation approach for extremely effective virtual try-on simulation to advance the physical simulation with non-linear deformations of clothes. There had been two fundamental processes: it first applied global body-shape dependent deformations to the garment and then predicted dynamically wrinkle deformations depending on physical appearance and pose.

Furthermore, Wang et al. [75] presented a semi-automatic technique for constructing clothing animation, which encoded essential information about the apparel shape and then learned to recreate the garment structure with physical propulsion based on the essential clothing description and target body movement and proposes a technique for recovering a 3D mesh of a fabric with 2D physical deformations using just a single-view picture.

It initially preprocessed a single-view picture, a human-body dataset, and a clothing-template database as input, performing garment parsing, human body modeling, and feature assessment. The original outfit registration and garment parameter detection were then synthesized for rebuilding body and clothes models with physical characteristics.

The most recent work concerning fashion physical simulation was introduced by Li et al[80] they built three new multi-GPU algorithms: SpMV for temporal integration, matrix assembly, and collision detection. Their method is intended for sparse linear systems with variable layouts, which are commonly employed in robust cloth modeling. They assessed performance on complicated fabric meshes. using over a million triangles and saw nearly linear speedups

on machines with four or eight GPUs.

4.3.3 Evaluation of Physical Simulation state of the art methods

There are few quantitative comparisons between physical simulation works. Most of them merely compute qualitative findings inside their research (e.g., per-vertex mean error) or compare vision with cutting-edge techniques. As shown in figure 25 below

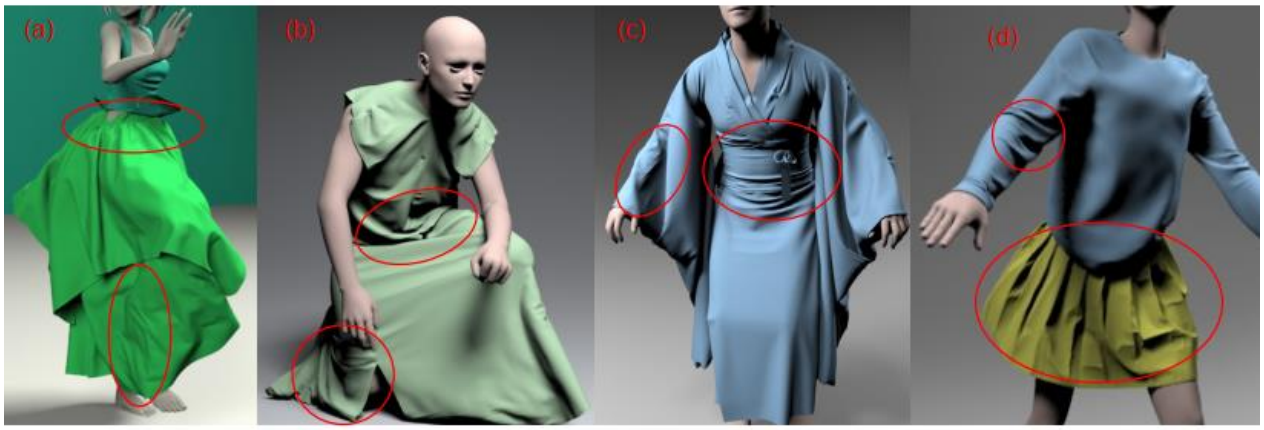

Fig.25: Complex cloth simulation by Li et al[80]

Chapter 5

Fashion Recommendation

Since not everyone is a fashion designer, Fashion recommendation has come in place in order to help people to choose the perfect fashion products according to their lifestyle and choices, fashion recommendation is split into three main parts: fashion compatibility, outfit matching, hairstyle suggestions.

5.1 Fashion Compatibility

Fashion compatibility is where the visual outfits are compatible between all of its parts like dress, shoes, hats, ...etc, Additionally, it merits referencing that the fundamental contrast between item retrieval and fashion recommendation is that the previous learn the visual similitude between similar outfits. Conversely, the last learns both visual similitude and visual similarity between various apparel types.

5.1.1 Datasets used in Fashion Compatibility process:

The most utilized source for fashion compatibility datasets is the Polyvore style site. It's anything but a web-based shopping site, where the style items contain rich explanations, e.g., clothing tone, text portrayal, and multi-see outfit images. We list the benchmark datasets for design similarity in Table 16.

Dataset	Publish Time	Number of outfits	Number of categories	Source
FashionVC [81]	2017	20,726	2	Polyvore.com
Vasileva et al. [32]	2018	68,306	19	Polyvore.com
PolyVore-T [83]	2019	19,835	5	collected from Polyvore
IQON3000 [84]	2019	308,747	6	The fashion web service IQON

Table.16: Most popular datasets used datasets for fashion compatibility task.

5.1.2 State of the art methods in Fashion Compatibility:

Song et al. [82] presented Conditional Similarity Networks (CSNs), which learned non-linear component embeddings that fused different ideas of likeness inside a common inserting utilizing a common component extractor. The CSNs resolved the issue of a standard trio installation that treated all trios similarly and disregarded the wellsprings of closeness.

incorporated visual and relevant modalities of design items by utilizing the autoencoder neural model to look for the non-linear idle similarity space.

For utilizing classification corresponding relations to show similarity, Yang et al. [83] proposed an interpretation based neural style similarity model which contained three sections: (1) initial planned everything into a dormant space through two CNN for visual and textual modality, (2)

encoded the classification integral relations into the idle space, furthermore, (3) limited an edge-based positioning model to improve both item embeddings and connection vectors together.
For making the fashion compatibility task easier to use, Wang et al. [84] presented a diagonal process for giving data about which item made the outfit contrary. They introduced a start to finish complex examination organization to foresee the similarity between various items at various layers and utilize the backpropagation inclination for determination.
In addition, Wang et al. [84] considered client inclinations to present a customized similarity demonstrating plan GP-BPR. It used two parts, general similarity demonstrating and individual inclination displaying, for assessing the item and user-item collaborations, respectively.

5.1.3 Evaluation of Fashion Compatibility state of the art methods:

The most common metric for evaluating the performance of fashion compatibility efforts is the area under the receiver operating characteristic curve (AUC). The AUC determines the likelihood that the reviewed work will recommend a greater level of compatibility for the positive set than for the negative set. The AUC values vary from 0 to 1.

5.2 Outfit Matching

Fig.26: Examples of outfit matching tasks [43].

Each outfit usually consists of several complementing pieces, such as shirts, pants, footwear, and accessories.
The complementing fashion elements, as seen in Fig. 26, are crucial to a beautiful ensemble. Moreover, there are three main reasons why fashion matching is difficult to achieve. First and foremost, fashion is a delicate and subjective notion. Second, there are several qualities that may be used to describe fashion.
Third, fashion item compatibility is a broad concept that encompasses a wide range of categories and connections. This topic has sparked a lot of attention in recent years, resulting in a big number of algorithms and methods.

5.2.1 Datasets used in Outfit Matching process:

Because different articles have different settings and most outfit matching datasets aren't publicly available, practically every work created their own. Table 17 shows the benchmark datasets for outfit matching.

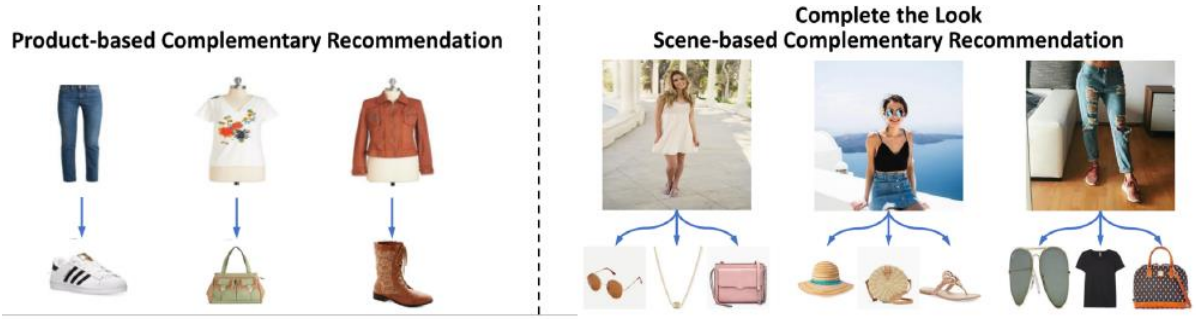

Fig.27: A study of complementary recommendations based on product and scene [85].

Dataset	Publish time	Number of outfits	Number of categories	Sources
Styles and Substitutes [88]	2015	773,465	N/A	Amazon.com
He et al. [90]	2016	598.353	N/A	Tradesy.com, Amazon.com
POG [91]	2019	1.01M	80	iFashion

Table.17: Most popular datasets used datasets for outfit matching task.

5.2.2 State of the art methods in Outfit Matching:

Kang et al. [86] recently launched "Complete the Look," which aims to propose fashion products that complement the current setting. They used Siamese networks and category-guided attention methods to assess global compatibility (i.e., the compatibility between the scene and product pictures) as well as local compatibility (i.e., the compatibility between each scene patch and the product image). Figure 27 shows a comparison of complimentary recommendations based on products and scenes.

Following that, a line of study based on metric-based studies advocated modeling item-to-item compatibility based on co-purchase behavior. they used Amazon.com co-purchase data to train a Siamese CNN to learn style compatibility across categories and generated suitable goods using a robust nearest neighbor retrieval.

The study by Guan et al [79] used a Low-rank Mahalanobis Transform to map compatible items to embeddings close in the latent space to discover the connections between the appearances of pairs of items.

Later they integrated visual and historical user feedback data. Visual signals were included in Bayesian Personalized Ranking using Matrix Factorization as the fundamental classifier in the study.

The majority of past research has been on top-to-bottom matching. An ensemble, in another view, usually comprises more accessories, such as shoes and handbags. To overcome this problem, they modified the typical triplet neural network to accommodate several occurrences instead of three.

Sanchez-Riera et al [87] created i-Stylist, a customized fashion recommendation system that retrieved clothing options based on the analysis of user photos. As a fully linked graph, the i-Stylist arranged the deep learning attributes and clothing properties of the user's clothing items. The probability distribution of an item's likability in shopping websites was later calculated using the user's personalized graph model.

5.2.3 Evaluation of Outfit Matching state of the art methods :

AUC is the most commonly used metric for outfit matching algorithms, as it is the assessment procedure for fashion compatibility. Some approaches are also tested using NDCG, as well as FITB (fill in the blank) accuracy. Unfortunately, both in terms of datasets and evaluation metrics, there is no common standard for outfit matching. As a result, we are unable to provide a comparison of various ways.

5.3 Hairstyle Suggestion:

A hairstyle is crucial to someone's physical appearance. With a different haircut, people might appear radically different. The perfect hairstyle may bring out natural beauty and elegance while enhancing the best facial features and concealing defects. However, selecting the appropriate hairstyle necessitates caution, as not all hairstyles are suitable for all facial features.

5.3.1 Datasets used in Hairstyle Suggestion process

Benchmark datasets for evaluating the performance of hairstyle suggestion systems are provided in Table 14. Although this proposed method is not for hairstyle suggestion, it is worth noting that Hairstyle30k [94] is by far the largest dataset for hairstyle-related challenges.

Dataset name	Publish time	Number of photos	Sources
Yang et al. [92]	2012	84	Hairstylists from 3 salons
Beauty e-Experts [93]	2013	1,605	Professional hairstyle websites
Hairstyle30k [94]	2017	30,000	Different web search engines

Table.18: Most popular datasets used datasets for hairstyle suggestion task.

5.3.2 State of the art methods in Hairstyle Suggestion:

Many articles concentrating on how to model and depict hairstyles with computer graphics or how to segment hair automatically have been developed recently.

Only a few research have been conducted to determine the best haircut for one's face. Following is a survey of the literature on hairstyle recommendations. By learning the association between facial forms and effective hairstyle examples, the pioneering work by Yang et al [92] identified appropriate hairstyles for a specific face.

The statistical learning step and the composition step were both included in the suggested example-based architecture. The goal of the statistical learning stage was to discover the best hairstyle for a given face image using a Bayesian inference-based model that calculated hairstyle probability distributions. They recommended using the ratio of line segments as the feature vector for describing the shape of each face and the alfa-matting-based approach for indicating the hair area in the image. The most acceptable hairstyle from the statistical learning step was then placed over a particular facial image to create a suitable hairstyle for the face.

The Beauty e-Experts system [93] was later created by Liu et al. to automatically recommend the most appropriate facial hair doing and makeup, as well as synthesize the visual effects.

They postulated using the extraction process of facial features and clothing features to learn various tree-structured super-graphs simultaneously to cooperatively model the hidden patterns among high-level beauty attributes (such as eye shadow color and hair length), mid-level beauty-related attributes (such as eye shape and mouth width), and low-level image features.

A facial image synthesis module was also proposed to synthesize the attractiveness features suggested by the multiple tree-structured super-graphs models.

5.3.3 Evaluation of Hairstyle Suggestions state of the art methods:

Liu et al. [93] computed the NDCG, which assesses how nearly the ranking of the top-k recommended styles is to the optimal ranking. Yang et al. [92] conducted a usability test to examine the effectiveness of their proposed approach. However, due to uneven benchmarks for different articles, we are unable to make comparisons between different hairstyle suggestion algorithms.

Conclusion

Attention has been drawn to the fashion and computer vision field for many years, every year there are different methods proposed to tackle new and old fashion challenges including fashion analysis, fashion detection, fashion synthesis and fashion recommendation. In this paper we covered the new and old subjects for the matter in details, as the fashion industry is becoming more intelligent and more reliable and even lowering the gap between the customer and the product by introducing new ways to interact with clothes and apparels.

Due to this over-growing attention to this field of study and the importance of its impact on the global industry, researches on intelligent fashion topics are getting more interest every year covering more topics and making more profits as we revolutionize the fashion industry

5.References:

[1] Kaiser, Susan B. Fashion and cultural studies. A&C Black, 2012.

[2] Anon. s. d. « Revenue of the Global Apparel Market 2012-2025 ». *Statista*. Accessed 2 april 2021 (https://www.statista.com/forecasts/821415/value-of-the-global-apparel-market).

[3] Anon. s. d. « EMERGING MARKET FACTS – Fashion Africa Sourcing Trips ». Accessed 3 april 2021 (https://www.fashionafricasourcingtrips.com/about/emerging-market-facts/).

[4] Anon. 2021. « Computer Vision ». Wikipedia. Accessed 3 april 2021 (https://en.wikipedia.org/w/index.php?title=Computer_vision)

[5] Frankenfield, Jake. s. d. « How Artificial Intelligence Works ». Investopedia. Accessed 18 april 2021 (https://www.investopedia.com/terms/a/artificial-intelligence-ai.asp).

[6] Iriondo, Roberto, et Roberto Iriondo. s. d. « Machine Learning (ML) vs. Artificial Intelligence (AI) - Crucial Differences – Towards AI — The Best of Tech, Science, and Engineering ». Accessed 18 april 2021 (https://towardsai.net/p/machine-learning/differences-between-ai-and-machine-learning-1255b182fc6).

[7] Anon. s. d. « What Is Deep Learning and How Does It Work? » SearchEnterpriseAI. Accessed 19 april 2021 (https://searchenterpriseai.techtarget.com/definition/deep-learning-deep-neural-network).

[8] Mihajlovic, Ilija. 2020. « Everything You Ever Wanted To Know About Computer Vision. Here's A Look Why It's So Awesome. » Medium. Accessed 20 april 2021 (https://towardsdatascience.com/everything-you-ever-wanted-to-know-about-computer-vision-heres-a-look-why-it-s-so-awesome-e8a58dfb641e).

[9] Steele, Valerie. s. d. « Definition of Fashion ». LoveToKnow. Accessed 22 april 2021 (https://fashion-history.lovetoknow.com/alphabetical-index-fashion-clothing-history/definitionn-fashion).

[10] Liu, Ziwei, et al. "Deepfashion: Powering robust clothes recognition and retrieval with rich annotations." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016.

[11] Liu, Ziwei, et al. "Fashion landmark detection in the wild." *European Conference on Computer Vision*. Springer, Cham, 2016.

[12] Yan, Sijie, et al. "Unconstrained fashion landmark detection via hierarchical recurrent transformer networks." *Proceedings of the 25th ACM international conference on Multimedia*. 2017.

[13] Lee, Sumin, et al. "A global-local embedding module for fashion landmark detection." *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops*. 2019.

[14] Yamaguchi, Kota, et al. "Parsing clothing in fashion photographs." *2012 IEEE Conference on Computer vision and pattern recognition*. IEEE, 2012.

[15] Ramakrishna, Varun, et al. "Pose machines: Articulated pose estimation via inference machines." *European Conference on Computer Vision*. Springer, Cham, 2014.

[16] Chen, X., & Yuille, A. (2014). Articulated pose estimation by a graphical model with image dependent pairwise relations. *arXiv preprint arXiv:1407.3399*.

[17] Wang, W., Xu, Y., Shen, J., & Zhu, S. C. (2018). Attentive fashion grammar network for fashion landmark detection and clothing category classification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 4271-4280).

[18] Ge, Y., Zhang, R., Wang, X., Tang, X., & Luo, P. (2019). Deepfashion2: A versatile benchmark for detection, pose estimation, segmentation and re-identification of clothing images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 5337-5345).

[19] He, K., Gkioxari, G., Dollár, P., & Girshick, R. (2017). Mask r-cnn. In *Proceedings of the IEEE international conference on computer vision* (pp. 2961-2969).

[20] Yamaguchi, K., Kiapour, M. H., Ortiz, L. E., & Berg, T. L. (2014). Retrieving similar styles to parse clothing. *IEEE transactions on pattern analysis and machine intelligence*, *37*(5), 1028-1040.

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

[22] Li, J., Zhao, J., Wei, Y., Lang, C., Li, Y., Sim, T., ... & Feng, J. (2017). Multiple-human parsing in the wild. *arXiv preprint arXiv:1705.07206*.

[23] Zhao, J., Li, J., Cheng, Y., Sim, T., Yan, S., & Feng, J. (2018, October). Understanding humans in crowded scenes: Deep nested adversarial learning and a new benchmark for multi-human parsing. In *Proceedings of the 26th ACM international conference on Multimedia* (pp. 792-800).

[24] Jia, M., Shi, M., Sirotenko, M., Cui, Y., Cardie, C., Hariharan, B., ... & Belongie, S. (2020, August). Fashionpedia: Ontology, segmentation, and an attribute localization dataset. In *European conference on computer vision* (pp. 316-332). Springer, Cham.

[25] Wang, W., Zhang, Z., Qi, S., Shen, J., Pang, Y., & Shao, L. (2019). Learning compositional neural information fusion for human parsing. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 5703-5713).

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

[27] Deng, J., Dong, W., Socher, R., Li, L. J., Li, K., & Fei-Fei, L. (2009, June). Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition* (pp. 248-255). Ieee.

[28] Garcia, N., & Vogiatzis, G. (2017). Dress like a star: Retrieving fashion products from videos. In *Proceedings of the IEEE International Conference on Computer Vision Workshops* (pp. 2293-2299). [37] Liao, L., He, X., Zhao, B., Ngo, C. W., & Chua, T. S. (2018, October). Interpretable multimodal retrieval for fashion products. In *Proceedings of the 26th ACM international conference on Multimedia* (pp. 1571-1579).

[29] Ma, Z., Dong, J., Long, Z., Zhang, Y., He, Y., Xue, H., & Ji, S. (2020, April). Fine-grained fashion similarity learning by attribute-specific embedding network. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 34, No. 07, pp. 11741-11748).

[30] Wang, X., Sun, Z., Zhang, W., Zhou, Y., & Jiang, Y. G. (2016, June). Matching user photos to online products with robust deep features. In *Proceedings of the 2016 ACM on international conference on multimedia retrieval* (pp. 7-14).

[31] Lin, Y. L., Tran, S., & Davis, L. S. (2020). Fashion outfit complementary item retrieval. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 3311-3319).

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

[33] Tan, R., Vasileva, M. I., Saenko, K., & Plummer, B. A. (2019). Learning similarity conditions without explicit supervision. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 10373-10382).

[34] Hidayati, S. C., You, C. W., Cheng, W. H., & Hua, K. L. (2017). Learning and recognition of clothing genres from full-body images. *IEEE transactions on cybernetics*, *48*(5), 1647-1659.

[35] Han, X., Wu, Z., Huang, P. X., Zhang, X., Zhu, M., Li, Y., ... & Davis, L. S. (2017). Automatic spatially-aware fashion concept discovery. In *Proceedings of the IEEE international conference on computer vision* (pp. 1463-1471).

[36] Yamaguchi, K., Okatani, T., Sudo, K., Murasaki, K., & Taniguchi, Y. (2015). Mix and Match: Joint Model for Clothing and Attribute Recognition. In *BMVC* (Vol. 1, No. 2, p. 4).

[37] Chen, Q., Huang, J., Feris, R., Brown, L. M., Dong, J., & Yan, S. (2015). Deep domain adaptation for describing people based on fine-grained clothing attributes. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 5315-5324).

[38] Heilbron, F. C., Pepik, B., Barzelay, Z., & Donoser, M. (2019, October). Clothing Recognition in the Wild using the Amazon Catalog. In *ICCV Workshops* (pp. 3145-3148).

[39] Balakrishnan, G., Zhao, A., Dalca, A. V., Durand, F., & Guttag, J. (2018). Synthesizing images of humans in unseen poses. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 8340-8348).

[40] Hidayati, S. C., Goh, T. W., Chan, J. S. G., Hsu, C. C., See, J., Wong, L. K., ... & Cheng, W. H. (2020). Dress with style: Learning style from joint deep embedding of clothing styles and body shapes. *IEEE Transactions on Multimedia*, *23*, 365-377.

[41] Wang, Z., & Quan, H. (2019, July). Fashion Outfit Composition Combining Sequential Learning and Deep Aesthetic Network. In *2019 International Joint Conference on Neural Networks (IJCNN)* (pp. 1-7). IEEE.

[42] Vaccaro, K., Shivakumar, S., Ding, Z., Karahalios, K., & Kumar, R. (2016, October). The elements of fashion style. In *Proceedings of the 29th annual symposium on user interface software and technology* (pp. 777-785).

[43] Gu, X., Wong, Y., Peng, P., Shou, L., Chen, G., & Kankanhalli, M. S. (2017, October). Understanding fashion trends from street photos via neighbor-constrained embedding learning. In *Proceedings of the 25th ACM international conference on Multimedia* (pp. 190-198).

[44] Mall, U., Matzen, K., Hariharan, B., Snavely, N., & Bala, K. (2019). Geostyle: Discovering fashion trends and events. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 411-420).

[45] Ma, Y., Yang, X., Liao, L., Cao, Y., & Chua, T. S. (2019, October). Who, where, and what to wear? Extracting fashion knowledge from social media. In *Proceedings of the 27th ACM International Conference on Multimedia* (pp. 257-265).

[46] Simo-Serra, E., & Ishikawa, H. (2016). Fashion style in 128 floats: Joint ranking and classification using weak data for feature extraction. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 298-307).

[47] Chang, Y. T., Cheng, W. H., Wu, B., & Hua, K. L. (2017, October). Fashion world map: Understanding cities through streetwear fashion. In *Proceedings of the 25th ACM international conference on Multimedia* (pp. 91-99).

[48] Wu, B., Cheng, W. H., Zhang, Y., Huang, Q., Li, J., & Mei, T. (2017). Sequential prediction of social media popularity with deep temporal context networks. *arXiv preprint arXiv:1712.04443*.

[49] Lo, L., Liu, C. L., Lin, R. A., Wu, B., Shuai, H. H., & Cheng, W. H. (2019, September). Dressing for attention: Outfit based fashion popularity prediction. In *2019 IEEE International Conference on Image Processing (ICIP)* (pp. 3222-3226). IEEE.

[50] Wu, B., Cheng, W. H., Liu, P., Liu, B., Zeng, Z., & Luo, J. (2019, October). Smp challenge: An overview of social media prediction challenge 2019. In *Proceedings of the 27th ACM International Conference on Multimedia* (pp. 2667-2671).

[51] Rothe, R., Timofte, R., & Van Gool, L. (2016). Some like it hot-visual guidance for preference prediction. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 5553-5561).

[52] Gao, L., Li, W., Huang, Z., Huang, D., & Wang, Y. (2018, August). Automatic facial attractiveness prediction by deep multi-task learning. In *2018 24th International Conference on Pattern Recognition (ICPR)* (pp. 3592-3597). IEEE.

[53] Lin, L., Liang, L., Jin, L., & Chen, W. (2019, August). Attribute-Aware Convolutional Neural Networks for Facial Beauty Prediction. In *IJCAI* (pp. 847-853).

[54] Wang, H., Xiong, H., & Cai, Y. (2020). Image Localized Style Transfer to Design Clothes Based on CNN and Interactive Segmentation. *Computational Intelligence and Neuroscience*, *2020*.

[55] Gu, Q., Wang, G., Chiu, M. T., Tai, Y. W., & Tang, C. K. (2019). Ladn: Local adversarial disentangling network for facial makeup and de-makeup. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 10481-10490).
[56] Jiang, W., Liu, S., Gao, C., Cao, J., He, R., Feng, J., & Yan, S. (2020). Psgan: Pose and expression robust spatial-aware gan for customizable makeup transfer. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 5194-5202).
[57] Han, X., Wu, Z., Wu, Z., Yu, R., & Davis, L. S. (2018). Viton: An image-based virtual try-on network. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 7543-7552).
[58] Zheng, N., Song, X., Chen, Z., Hu, L., Cao, D., & Nie, L. (2019, October). Virtually trying on new clothing with arbitrary poses. In *Proceedings of the 27th ACM International Conference on Multimedia* (pp. 266-274).
[59] Dong, H., Liang, X., Shen, X., Wu, B., Chen, B. C., & Yin, J. (2019). Fw-gan: Flow-navigated warping gan for video virtual try-on. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 1161-1170).
Conference on Artificial Intelligence (pp. 3721-3727).
[60] Nguyen, T., Tran, A. T., & Hoai, M. (2021). Lipstick Ain't Enough: Beyond Color Matching for In-the-Wild Makeup Transfer. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 13305-13314).
[61] Ge, Y., Song, Y., Zhang, R., Ge, C., Liu, W., & Luo, P. (2021). Parser-Free Virtual Try-on via Distilling Appearance Flows. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 8485-8493).
[62] Zhu, S., Urtasun, R., Fidler, S., Lin, D., & Change Loy, C. (2017). Be your own prada: Fashion synthesis with structural coherence. In *Proceedings of the IEEE international conference on computer vision* (pp. 1680-1688).
[63] Issenhuth, T., Mary, J., & Calauzenes, C. (2020). Do not mask what you do not need to mask: a parser-free virtual try-on. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XX 16* (pp. 619-635). Springer International Publishing.
[64] Hsieh, C. W., Chen, C. Y., Chou, C. L., Shuai, H. H., & Cheng, W. H. (2019, September). Fit-me: Image-based virtual try-on with arbitrary poses. In *2019 IEEE International Conference on Image Processing (ICIP)* (pp. 4694-4698). IEEE.
[65] Hsieh, C. W., Chen, C. Y., Chou, C. L., Shuai, H. H., Liu, J., & Cheng, W. H. (2019, October). FashionOn: Semantic-guided image-based virtual try-on with detailed human and clothing information. In *Proceedings of the 27th ACM International Conference on Multimedia* (pp. 275-283).
[66] Han, X., Hu, X., Huang, W., & Scott, M. R. (2019). Clothflow: A flow-based model for clothed person generation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 10471-10480).
[67] Ionescu, C., Papava, D., Olaru, V., & Sminchisescu, C. (2013). Human3. 6m: Large scale datasets and predictive methods for 3d human sensing in natural environments. *IEEE transactions on pattern analysis and machine intelligence*, *36*(7), 1325-1339.
[68] Ma, L., Jia, X., Sun, Q., Schiele, B., Tuytelaars, T., & Van Gool, L. (2017). Pose guided person image generation. *arXiv preprint arXiv:1705.09368*.
[69] Song, S., Zhang, W., Liu, J., & Mei, T. (2019). Unsupervised person image generation with semantic parsing transformation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 2357-2366).
[70] Cui, A., McKee, D., & Lazebnik, S. (2021). Dressing in Order: Recurrent Person Image Generation for Pose Transfer, Virtual Try-On and Outfit Editing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 3940-3945).
[71] Ren, Y., Yu, X., Chen, J., Li, T. H., & Li, G. (2020). Deep image spatial transformation for person image generation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 7690-7699).
[72] Men, Y., Mao, Y., Jiang, Y., Ma, W. Y., & Lian, Z. (2020). Controllable person image synthesis with attribute-decomposed gan. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 5084-5093).
[73] Siarohin, A., Sangineto, E., Lathuiliere, S., & Sebe, N. (2018). Deformable gans for pose-based human image generation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 3408-3416).
[74] Li, Y., Huang, C., & Loy, C. C. (2019). Dense intrinsic appearance flow for human pose transfer. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 3693-3702).
[75] Wang, T. Y., Shao, T., Fu, K., & Mitra, N. J. (2019). Learning an intrinsic garment space for interactive authoring of garment animation. *ACM Transactions on Graphics (TOG)*, *38*(6), 1-12.
[76] Lahner, Z., Cremers, D., & Tung, T. (2018). Deepwrinkles: Accurate and realistic clothing modeling. In *Proceedings of the European Conference on Computer Vision (ECCV)* (pp. 667-684).

[77] Santesteban, I., Otaduy, M. A., & Casas, D. (2019, May). Learning-based animation of clothing for virtual try-on. In *Computer Graphics Forum* (Vol. 38, No. 2, pp. 355-366).
[78] Patel, C., Liao, Z., & Pons-Moll, G. (2020). Tailornet: Predicting clothing in 3d as a function of human pose, shape and garment style. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 7365-7375).
[79] Tiwari, G., Bhatnagar, B. L., Tung, T., & Pons-Moll, G. (2020). Sizer: A dataset and model for parsing 3d clothing and learning size sensitive 3d clothing. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part III 16* (pp. 1-18). Springer International Publishing.
[80] Guan, P., Reiss, L., Hirshberg, D. A., Weiss, A., & Black, M. J. (2012). Drape: Dressing any person. *ACM Transactions on Graphics (TOG)*, *31*(4), 1-10.
[81] Li, C., Tang, M., Tong, R., Cai, M., Zhao, J., & Manocha, D. (2020). P-cloth: interactive complex cloth simulation on multi-GPU systems using dynamic matrix assembly and pipelined implicit integrators. *ACM Transactions on Graphics (TOG)*, *39*(6), 1-15.
[82] Song, X., Feng, F., Liu, J., Li, Z., Nie, L., & Ma, J. (2017, October). Neurostylist: Neural compatibility modeling for clothing matching. In *Proceedings of the 25th ACM international conference on Multimedia* (pp. 753-761).
[83] Yang, X., Ma, Y., Liao, L., Wang, M., & Chua, T. S. (2019, July). Transnfcm: Translation-based neural fashion compatibility modeling. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 33, No. 01, pp. 403-410).
[84] Wang, X., Wu, B., & Zhong, Y. (2019, October). Outfit compatibility prediction and diagnosis with multi-layered comparison network. In *Proceedings of the 27th ACM International Conference on Multimedia* (pp. 329-337).
[85] Song, X., Han, X., Li, Y., Chen, J., Xu, X. S., & Nie, L. (2019, October). GP-BPR: Personalized compatibility modeling for clothing matching. In *Proceedings of the 27th ACM International Conference on Multimedia* (pp. 320-328).
[86] Kang, W. C., Kim, E., Leskovec, J., Rosenberg, C., & McAuley, J. (2019). Complete the look: Scene-based complementary product recommendation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 10532-10541).
[87] Sanchez-Riera, J., Lin, J. M., Hua, K. L., Cheng, W. H., & Tsui, A. W. (2017, January). i-Stylist: Finding the right dress through your social networks. In *International Conference on Multimedia Modeling* (pp. 662-673). Springer, Cham.
[88] McAuley, J., Targett, C., Shi, Q., & Van Den Hengel, A. (2015, August). Image-based recommendations on styles and substitutes. In *Proceedings of the 38th international ACM SIGIR conference on research and development in information retrieval* (pp. 43-52).
[89] Ni, Y., & Fan, F. (2011). A two-stage dynamic sales forecasting model for the fashion retail. *Expert systems with applications*, *38*(3), 1529-1536.
[90] He, R., & McAuley, J. (2016, February). VBPR: visual bayesian personalized ranking from implicit feedback. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 30, No. 1).
[91] Dong, X., Song, X., Feng, F., Jing, P., Xu, X. S., & Nie, L. (2019, October). Personalized capsule wardrobe creation with garment and user modeling. In *Proceedings of the 27th ACM International Conference on Multimedia* (pp. 302-310).
[92] Yang, W., Toyoura, M., & Mao, X. (2012, January). Hairstyle suggestion using statistical learning. In *International Conference on Multimedia Modeling* (pp. 277-287). Springer, Berlin, Heidelberg.
[93] Liu, L., Xing, J., Liu, S., Xu, H., Zhou, X., & Yan, S. (2014). Wow! you are so beautiful today!. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, *11*(1s), 1-22.
[94] Yin, W., Fu, Y., Ma, Y., Jiang, Y. G., Xiang, T., & Xue, X. (2017, October). Learning to generate and edit hairstyles. In *Proceedings of the 25th ACM international conference on Multimedia* (pp. 1627-1635).