HIGH SCHOOL EDITION
Journal of Student Research
Volume 11 Issue 1 (2022)

Computer Vision in Fashion Trend Analysis and Applications

Varnika Jain[1] and Catherine Wah[#]

[1]National Public School Koramangala, Bengaluru, Karnataka, India
[#]Advisor

ABSTRACT

Computer vision is a field of artificial intelligence that allows computers to derive relevant information from visual inputs and take actions accordingly. In this paper, we will study the involvement of computer vision in fashion trend analysis and its applications. We begin by understanding the general context of computer vision. We will then discuss the role that computer vision plays in the fashion industry. We will move on to see how over time, computer vision has solved and is still advancing in visual search, fashion style and trend analysis, fashion style compatibility, fashion forecasting, and fashion creativity in the order of increasing complexity. Then, we will discuss the significance of computer vision in the current complex tasks of fashion style analysis, fashion recommendations, and fashion forecasting, while reviewing some of their existing deep learning methodologies. We will then discuss the significance of deep learning in today's approach to solving fashion trend analysis problems. Finally, we will overview the future directions of computer vision, leading us to a novel concept of fashion creativity that still requires further future study to solve potential tasks like cultural inclusivity in the fashion sphere.

1. Introduction

Consider a mechanism that can allow us to understand an extensive dataset of visual information, analyse it, and produce desired outputs on performing highly complex tasks which are difficult for humans to do at an instantaneous speed. This is why today's scientists and researchers have developed a field of artificial intelligence called computer vision. Computer vision is a field that allows computer systems to actually "see" and understand visual data. With time, computer vision requires constant modifications to solve novel tasks of demand in the market.

Now consider an equally evolving field - fashion. Fashion is an integral part of today's society, constantly changing to incorporate new themes, ideas, and desires. We witness news articles, magazines, and digital media every day discussing the recent fashion trends in the market. We even see constant developments in retrieving this information. From studying fashion online to finding clothing articles, to even having style recommenders to match our own tastes, we see ourselves deeply involved in the fashion realm.

In this paper, we study the role of computer vision in the fashion stream, more specifically, in fashion trend analysis and its applications. We begin by understanding the background of computer vision and its implications in Section 2. In Section 3, we study the applications of computer vision in fashion in the order of increasing complexity of the tasks, from visual search to fashion creativity. We then go on to analyse the significance of understanding fashion trend analysis in today's fast-paced society, followed by some existing methodologies towards this front in Section 4. Section 5 discusses the future directions of computer vision, covering topics like fashion creativity and its goal towards social problems like cultural inclusivity.

2. Background on Computer Vision

As humans, we comprehend the three-dimensional structure of the world around us with apparent ease. We can identify the shape, translucency, and subtle patterns of light and shading that play across object surfaces which segment them from the background. We can count the number of people in a portrait and discern their emotions through their facial expressions *(Szeliski, 2010)*.

Computer vision is a discipline that uses statistical methods to extract valuable information from unstructured data - namely digital images, videos, and other visual inputs - using methods involving geometry, physics, and learning theory. Vision requires studying cameras and physical processes such as image formation to obtain simple results from individual pixels, integrating information from various images into a coherent whole, distinguishing groups of pixels so as to deduce shape information, and recognising objects using geometric information or probabilistic techniques *(Forsyth & Ponce, 2012)*. To sum up, computer vision enables computers to see, understand, and analyse visual data.

Computer vision is complex, partly because vision is an *inverse* problem, where the aim is to recover unknown parameters on being given insufficient data to specify the solution completely. Unlike a *forward* problem in computer graphics, like analysing an image by putting together various smaller elements, computer vision tries to depict the visible world in one or more images and reconstruct its properties, like shape, illumination, and colour distributions. What is intriguing is that humans can do this so easily, while computer vision algorithms are so error-prone *(Szeliski, 2010)*.

Computer vision today has various applications, namely, in biometrics, motion capture, optical character recognition, retail, photogrammetry, automotive safety, surveillance, match move, and so forth. *(Szeliski, 2010)*

As part of our review, the reader is recommended to have a minor/basic understanding of the primary methods deployed in the fields of machine learning and artificial intelligence, such as convolutional and deep neural networks, support vector machine approach, conditional random fields, Gaussian distributions, k-fold cross evaluations, submodular function maximization, etc. However, the lack of this knowledge will not affect the reader substantially.

3. Applications of Computer Vision in Fashion

3.1. Visual Search

Online retrieval started with textual cues where one would input text-based data for conducting online searches to obtain desired results. For example, one would have to specify the colour and style of a type of shoe they would want to access online on search engines like Google or shopping sites and the like. The advent of computer vision in fashion has brought us visual search.

Now, what is visual search? Consider a situation where a user wishes to retrieve results of a pair of blue running shoes, similar to one they saw at a friend's house, to buy it online. The user would have to input the shoe type, style, specific colour, and other notable characteristics into a textual retrieval system. The task becomes much simpler and faster if the user can upload a picture of the said pair of shoes for an image-based retrieval. This is where visual search comes into play.

Visual search uses real-world images, like screenshots, photographs, or Internet images, as input data to retrieve desired results. It initially involved human intelligence, i.e., textual attributes for item retrieval. During advances in mobile computing, implementing visual search to extract and match attributes like colour and pattern was a compelling but challenging task due to the high variability and deformability of clothing. The aim was to illustrate potential mobile applications for attribute-based multimedia retrieval of clothing items and image annotation. Researchers proposed attribute vocabulary constructed using human annotations obtained on a novel fine-grained clothing dataset to train an explicit visual recognition system for clothing styles. The attribute vocabulary is organised into groupings, and image-level attribute annotations are obtained via crowdsourcing on platforms like Amazon Mechanical Turk, a marketplace for completing virtual tasks that require human intelligence (*Di et al., 2013)*.

With further advances in computer vision, fashion item retrieval aimed to use visual search with minimum human supervision. The main objective was to eliminate weak labels and noisy annotations to yield precise search outputs. One approach was to directly learn from noisy labels and focus mainly on constructing a model that considers noise; another was a semi-supervised approach, focusing on learning from noisy labelled data with a neural network to combine a small set of clean labels. In building on these approaches, researchers proposed to train a multi-label image classifier on a large dataset with extremely noisy labels, where a small subset of the dataset with human-verified labels was available. This setting could often be used when we collect images from the web or social media and have experts correct some labels *(Inoue et al., 2017).*

With subsequent examinations, researchers have proposed a multimodal search engine for item retrieval, built on two single-modal modules, i.e., two searches are run independently for both image and text queries resulting in two initial sets of results. Then the best matches are selected from the initial pool of results according to blending methods. For input, the baseline style search engine takes two types of query information: an image containing object(s), here, a pair of running shoes, and a textual query used to specify search criteria, like soft-padded and waterproof. For queries that already represent a single object, no object detection is required. If needed, an object detection algorithm is run on the uploaded picture to detect objects of classes of interest; for example, for an entire outfit, sources of interest could consist of a shirt, pants, and a pair of heels. Simultaneously, the engine retrieves the results for a textual query. With all visual and textual matches retrieved, a blending algorithm ranks them depending on the visual feature similarity to the query image as well as contextual similarity - items that appear more often together in the same context - and returns the resulting list of stylistically aesthetically similar objects *(Tautkute et al., 2019).* Visual search has undergone a significant amount of study, and will most definitely be developed based on the discussions above towards understanding consumer needs.

3.2. Fashion Style & Trend Analysis

Let us consider the same shoe example. The consumer would wish to buy a shoe of their desired aesthetic that describes their preference of a particular ‘look’. Or maybe they just want to check out what’s in buzz for the season. Understanding these styles involves gathering information by observation, experience, interviews, and exposure to various events *(Jia et al., 2018).* This is where fashion trend analysis comes into play.

Fashion trends are expressions of socio-cultural aesthetic spheres, primarily employing visual and haptic characteristics applied to clothing and apparel products and are often considered fads or short-term trends. Fashion trends also allow new style experiments; they can introduce people to styling techniques they never thought they would like and types of clothing they would otherwise not have experimented with.

Understanding fashion styles and trends are of great potential interest to retailers and consumers alike. The photos people upload on social media provide a historical and public data source of how people dress across the world and at different times *(Mall et al., 2019).* Fashion styles and trends are used to study visual compatibility and future trend potentials, which will be discussed in later sections.

Simultaneous to the development of visual search methods mentioned above, some computational methods had begun to look at other aspects of recognition problems, such as *when* and *where* an image was captured. This could potentially be useful for historical dating at a large scale or in more everyday scenarios to identify the vintage of objects. An automatic method for dating man-made objects could be quite helpful in organizing and retrieving from these large disorganized collections. The temporal periods of apparel tend to be indicated by particular stylistic elements, for example, the puffed bubble sleeves and the polka-dot printed fabrics of the 60s or the spandex leggings of the 80s.

For this purpose, researchers aimed to estimate the time period of objects using deep networks. As time has a natural ordering, there are additional opportunities for introspection into the structure learned by the deep networks that are not available for arbitrary classification tasks *(Vittayakorn et al., 2017).*

With advances in style and trend analysis, researchers aimed to learn whether they could design automated methods to characterize and predict seasonal and annual fashion trends and detect social events like festivals or sporting events by identifying their respective style elements. Approaches included using recognition algorithms to identify a coarse set of fashion attributes in a large dataset of images and assigning interpretable parametric models of long-term temporal trends to these attributes. These models could capture both seasonal cycles and changes in popularity over time by helping understand existing trends and make accurate, temporally fine-grained forecasts across long time scales. For example, one would find that more people would wear black year on year, but they tend to do so more in the winter than in the summer.

Another trend analysis target was to leverage textual descriptions and captions from social media to investigate the reasons behind short-term spikes caused by social events. For example, one could observe an unusual increase in the colour yellow in Bangkok in early December and associate it with the words “father”, “day”, “king”, “live”, and “dad”, which correspond to the king’s birthday, celebrated as Father’s Day in Thailand by wearing yellow.

One could also predict trends and events not just at the level of individual fashion attributes (such as “wearing yellow”) but also at the level of styles consisting of recurring visual ensembles. These styles are identified by studying photographs from a large database. The predictions of the future popularity of styles would be just as accurate as those of individual attributes. Further, one can run the same event detection framework described above on style trends, allowing us to not only automatically detect social events, but also associate each event with its own distinctive style *(Mall et al., 2019)*.

Fashion style and trend analysis is one of the most important fields of interest for computer vision scientists today since it is the birthstone of subsequent complex fashion tasks, such as style compatibility, recommendations, and fashion forecasting, all of which will be discussed in more depth in future sections. With further study in Fashion trend analysis, we can accomplish these above-stated convoluted tasks.

3.3. Visual Style Compatibility

Now suppose the consumer has decided to buy a pair of blue stilettos. She would want to know what style of clothes it would go along with or whether the shoes are compatible with most of her outfits. Visual style compatibility helps us answer such questions.

Visual style compatibility prediction refers to the task of determining whether a set of items go well together. This could be possible by using fine-grained recognition of subcategories and attributes, for example, “navy blue ankle-strap stilettos,” with a graph that informs which subcategories match together. However, this approach requires significant domain knowledge, and collecting large datasets becomes especially hard in domains like clothing, where fashion collections change every season.

An efficient approach is to learn a feature transformation from the images of the items to a latent space, called style space, wherein images of items from different categories that match together are close in the style space and items that don’t match are far apart. One could retrieve bundles - sets of items from different categories, like shirts, shoes and pants - of compatible objects. The bundle of objects comes from visually distinct categories, so to generate structured bundles of compatible items, one can query the learned latent space and retrieve the nearest neighbours from each category to the query item *(Veit et al., 2015)*.

With further studies in fashion compatibility, a new-found problem statement was to predict how fashionable a person looks in a particular photograph. This would be especially handy to social-media-based fashion style icons. Fashionability is affected not only by the garments the subject is wearing but also by many other factors such as the appeal of the scene behind the person, the image quality, the subject’s age, body characteristics, or even personality. Researchers collected a large-scale dataset - for example, a Fashion144k dataset that consists of 144,169 user posts - by crawling a social website and proposed a conditional random field model that analyses the settings, users and their fashionability and predicts the visual aesthetics related to fashion. It can also analyse fashion trends in the world or individual cities and potentially different age groups and outfit styles. It has significant applications in outfit recommendations *(Simo-Serra et al., 2015)*.

Building on this concept, one would want to choose appropriate and aesthetically pleasing clothing while packing for a journey. This requires considering the weather, the season, and the attraction type of the destination, such as the cultural heritage, the natural scenery, or the natural landscape. For example, one would have to wear modest clothes covering the shoulders and knees when visiting the Taj Mahal in Agra to avoid being disrespectful in a mausoleum. If they're worried about the heat, they could opt for wearing loose-fitting maxi dresses or linen trousers. To appear aesthetically appealing, one will also have to consider the background scenery with the foreground outfit. Currently, an increasing number of travel websites allow people to share their trip photos online and gain popularity. A large number of online photos provide excellent references that travellers can use to plan their trips and decide which clothes will be appropriate and beautiful for their desired destination. This gives rise to novel applications such as location-oriented clothing recommendations by learning the correlation between clothing and location from photos shared on travel websites. Based on the observations, location attributes for climate, weather, attraction type, and colour can be defined. Then a hybrid multi-label convolutional neural network combined with a support vector machine approach can be used to formulate the complex correlation between clothing and location attributes *(Zhang et al., 2017)*.

Another compelling problem statement is to grade outfits from a closet of items and provide an arbitrary score for a combination of these items. This can be possible using a deep neural network system to take variable frequencies of items and predict a score from a large dataset. This could also be coupled with an outfit recommender to serve as a personal closet assistant *(Tangseng et al., 2018)*. Let's consider an example of a college student. Due to financial and space limitations, he might want to keep a limited number of clothing items while also looking fashionable. A solution is to find multi-compatible items from his closet that are visually appealing and create fresh new outfits. A capsule wardrobe - a set of garments that can be assembled into many visually compatible outfits - comes into play here. A curated capsule can help consumers get the best value for their dollars while also being stylish. Capsules require an accurate model of visual compatibility to capture how multiple visual items interact, often according to subtle visual properties. One challenge is that existing compatibility methods assume supervision via labels, limiting scope and precision, and they largely cater only to pairwise compatibility. The second challenge is that capsule generation is a complex combinatorial problem. Of all possible garments, one seeks the subset that maximises versatility and compatibility, and the addition of any one garment introduces multiple new outfit combinations. To permit efficient subset selection over the space of all outfit combinations, researchers developed submodular objective functions capturing the key ingredients of visual compatibility, versatility, and user-specific preference. They devised an iterative approach to allow near-optimal submodular function maximisation so as to expand possible outfits on adding garments *(Hsiao & Grauman, 2018)*.

Building on the above-discussed approaches, we will find that visual style compatibility and fashion recommendations are in great demand for fields of fashion forecasting which will be discussed in later sections.

3.4. Fashion Forecasting

Considering the previous shoe example, the user may wish to understand the potential shoe styles that will be trending in the next season. She might want to know whether the future style statement would be based on previous style trends or whether it will evolve into a mix of different fashion statements. We can generalise the questions as follows. What is the future of fashion? Based on a chronological study of fashion styles throughout the decades, can we find the potential style trend in the next season? Will fashion build on the previous trends cyclically and evolve into something new in the next few years? What does futuristic fashion look like, and how do we achieve this? Does this give us an insight into future consumer interests to promote sales?

Fashion forecasting is the process of making future predictions based on analysis of past and present data early enough to allow production to meet consumer demand. Fashion forecasting involves predicting the colours, fabrics, textures, materials, prints, graphics, beauty, accessories, footwear, street style, and other styles presented on the runway and in the stores for the upcoming seasons.

Researchers now aim to forecast visual style trends by predicting the future popularity of styles discovered from fashion images in an unsupervised manner. Given a list of trending garments, can we predict which will remain stylish in the future? Which old trends are primed to resurface, independent of seasonality? Answering these questions involves hypothesising new mixtures of styles that will become popular in the future, discovering style dynamics, and naming the critical visual attributes that will dominate tomorrow's fashion.

One can predict the future of fashion styles based on images and consumers' purchase data by learning representations of fashion images that capture the garments' visual attributes, discovering a set of fine-grained styles that are shared across images in an unsupervised manner, studying statistics of past consumer purchases, constructing the styles' temporal trajectories and predicting their future trends *(Al-Halah et al., 2017).*

The fashion industry aimed to detect and predict fashion trends and styles over space and time. This could be possible by designing an automated method to characterise and predict seasonal and fashion trends. Mall et al. (2019) presented an automated framework for analysing the temporal behaviour of fashion elements across the globe, which models and forecasts long-term trends and seasonal behaviours. They evaluated their parametric temporal model based on its ability to make out-of-sample predictions.

Fashion forecasting today is one of the most popular fields in today's research towards developing the fashion realm. Further studies in fashion predictions are leading to better data outcomes. The significance of fashion forecasting is astounding, forcing us to develop on the above concepts and introduce new unknown domains, such as fashion creativity, which is discussed more in Section 3.5 and Section 5.

3.5. Fashion Creativity

After analysing fashion styles and trends and forecasting future fashion statements, one would wonder about subsequent computer vision applications in fashion. Considering the shoe example, what if one wanted to create a different shoe type altogether? This involves not only an understanding of what has already been existing but also that of what doesn't. Is there a way we can get technology to generate something novel? Is there a way computers can incorporate creativity to produce an output?

Creativity is the ability to generate and refine ideas. It is the interaction among aptitude, process, and the environment by which an individual or group produces a discernible idea that is both novel and useful as defined within a social context. We shall discuss more on this topic in Section 5 *(Jeon et al., 2021).*

4. Computer Vision in Fashion Trend Analysis

4.1. Significance of Fashion Trend Analysis

Since early history, fashion has played a crucial role in defining one's social status. One's persona can be perceived from their style sense *(Simo-Serra et al., 2015).* Fashion trends indicate shifting cultural attitudes and trends - René König (1973) argued that fashionable behaviour reflects human psychological functioning and receptivity to novel ideas adopted perpetually at points of social stability. Fashion trends are cyclical - they appear and reappear in the fashion landscape. By isolating and analysing these recurrent dress patterns, which carry slightly different meanings in each rendition, one can predict their use in the future *(Lynch & Strauss, 2007).*

Fashion prediction, or fashion trend forecasting, is critical to attract consumers and help retail businesses and designers sell their brands *(K, 2015)* while keeping ahead of the trends amongst consumers. Eminent e-commerce lines are increasing their commitment and investment in automated fashion. Better knowledge of users' unique preferences for product recommendations has become a crucial task *(Stephen et al., 2019)* for retail businesses.

4.2. Existing Methodologies

Fashion trend analysis aims to understand both the dimensions and dynamics of people's preferences (for recommendations), the modelling of which can be challenging due to the need to simultaneously model the visual appearance of products (i.e., style analysis) as well as their evolution over time (i.e., forecasting) *(He & McAuley, 2016)*.

Computer vision tasks involve object detection, recognition and segmentation, leading to successive enhancements in cloth classifications, attribute predictions, and clothing image retrievals. Lately, recommendation and personalisation based algorithms employ techniques driven by collaborative filtering, neural networks, probabilistic graph models, and a combination of these systems - hybrids.

Collaborative filtering works by analysing consumer behaviour and preferences to create predefined patterns for product recommendations. On the other hand, content-based techniques utilise attributes for item recommendation based solely on users' search habits and past feedback. Evolving trends within the community use visual factors that consumers consider when evaluating products. Knowledge-based (hybrid) systems integrate product features, temporal dynamics, and user preferences for product recommendations from all these models *(He & McAuley, 2016; Stephen et al., 2019)*. These techniques are highlighted in later sections.

4.2.a. Style Analysis

4.2.a.i. Fine Tuning for Temporal Estimation

Convolutional Neural Networks (CNN) can be used for temporal estimation of fashion styles, by first evaluating classifiers trained on clothing features from a pre-trained network, then adapting the network to directly predict the time period. Vittayakorn et al. (2017) used two different classification methods, namely, linear Support Vector Machines and Support Vector Regressors. Network fine-tuning adapts networks to tasks such as object detection, pose estimation and action detection, or fine-grained category detection. Researchers were interested in fine-tuning the original object classification model for the temporal estimation task.

4.2.a.ii. Classification based on Product Category

CNNs, due to their classification accuracy, were employed for categorising product images based on their product category. Schindler et al. (2017) suggested a three-fold cross-evaluation for calculating accuracies on a per-image and a per-product scale, taking into account the cumulative maximum of all predicted product images. k-fold cross-evaluation is a data partitioning strategy to effectively use a dataset to build a more generalized model which can perform well on unseen data.

The product classification experiments were conducted using the different CNN architectures presented by the researchers on two different scales. First, a broad evaluation was performed on the small-scale image subsets. Then, the best performing models were evaluated on a large scale image dataset. All models were trained using image data augmentation, including horizontal flipping of the image, shifting it in height and width, and zoom range.

4.2.a.iii. Style Classification based on Minimal Human Supervision

Networks pre-trained on large-scale datasets such as ImageNet are not preferred for variable and subjective fashion-related tasks. Exploiting heterogeneous datasets like Fashion144k that consist of fashion images in various scenes with weak labels and designing a CNN to learn from them can provide better feature representation than ImageNet for fashion style classification.

Inoue et al. (2017) considered the size of the dataset and the quality of the annotations in the dataset for recognition performance. They experimented with multi-label classification and predicted the colours and garments worn by the person in each image. Manually correcting weak labels from the dataset to obtain "clean" labels is essential to evaluate the quality of annotations. Learning a mapping between the noisy and clean labels helps use neural networks to clean the labels and train a prediction model.

Each post in a dataset contains images with metadata including label tags, different user angles, or a zoom-in on various garments. The posts also include short descriptions and tags of the types and colours of the garments. This information is very noisy - not all garments are tagged, or only a part of the information is available.

A multi-label image classifier on a large dataset with extremely noisy labels is trained, where a small subset of the dataset has human-verified labels. For example, this is possible if experts correct some labels from images from the web or social media. An approach where a label cleaning network in combination with a classification network is used on a base CNN can be utilised for this purpose.

4.2.b. Fashion Recommendations

4.2.b.i. Conditional Random Field

A Conditional Random Field (CRF) is a class of statistical modelling methods applied in pattern recognition and machine learning for structured prediction. Unlike a classifier that predicts a label for a single sample without considering neighbouring samples, a CRF takes context into account.

Simo-Serra et al. (2015) collected a novel dataset of thousands of user posts from a clothing-oriented social website. In a post, a user could publish multiple photographs of them wearing a new outfit. Each photograph showed a different angle or zoomed in on different clothing items. The user could decide to add a description and tags of the types and colours of the garments. The user could also reveal their geographic location, which was an essential factor in perceiving fashionability. They analysed each post's number of votes/likes and publishing time. The number of votes was directly proportional to the number of users.

Their model tried to eliminate the temporal dependency by calculating histograms of the votes for each month and fit a Gaussian distribution (a probability distribution that is symmetric about the mean, showing that data near the mean are more frequent in occurrence than data far from the mean. It appears as a bell curve in graph form) to it. They then bin the distribution such that the expected number of posts for each bin is the same. Doing so can eliminate almost all time dependency and obtain a quasi-equal distribution of classes used as a fashionability measure, ranging from 1 (not fashionable) to 10 (very fashionable).

Here, CRF is used to learn the different outfits, types of people and settings. This method uses deep networks over a wide variety of features exploiting Fashion144k images and meta-data to produce accurate predictions of the fashionability of a post.

4.2.b.ii. Visual Temporal Dynamics

Fashion evolution studies require data on visual features or factors that influence users' product preferences and visual temporal dynamics that differentiate fashion styles of different periods. Researchers hypothesised that evolving fashion styles are analysed by consumer purchase choices rather than factors such as 'star ratings' and 'likes' (which were considered in Section 4.2.b.i.)—people only buy items if they are already attracted to their visual appearance. Star ratings and likes can depend on non-visual factors like clothing fit or product defects for an individual consumer. Hence, visual temporal dynamics from implicit feedback datasets, such as purchase histories of clothing and accessories, where visual signals are at play are considered. Systems that were quantitatively helpful for estimating users' personalised rankings (i.e., assigning potential purchases higher ranks than non-purchases), were built which could then be employed for recommendations.

Researchers employed the basic formulation of the Matrix Factorisation method. The preference of a user toward an intel at a particular time was predicted according to global offset, user/item bias terms, and dimensional latent factors describing user and item. Further variable elucidation can be found in *(He & McAuley, 2016)*.

4.2.b.iii. Hybrid Multi-Label Convolutional Neural Networks with Support Vector Machine

Zhang et al. (2017) suggest a method that involves a thorough data observation on photos from several international travel websites to determine location attributes that influence dress, namely, 'Climate: Season', 'Climate: Weather', 'Attraction: Type', and 'Attraction: Color'. A benchmark dataset for location-oriented clothing recommendations is considered which contains thousands of traveller images with travel destination annotations. Unlike methods based only on clothing analysis, the hybrid mCNN-SVM approach correlates clothing and location-related attributes from data like online travel photos for fashion analysis (introduced in Section 4.2.b.i).

The mCNN fully explores the uneven distribution of clothing attributes, while the structured SVM explores the intrinsic clothing-location correlation based on well-defined location attributes. Outfits are recommended on attributes like clothing and location colour schemes.

4.2.b.iv. Personalised Outfit Generator

Unlike conventional item recommendation, fashion recommendation involves creativity for outfit generation process, which requires both innovation and characteristic.

There exist two requirements in fashion outfit generation and recommendation, namely, *compatibility* and *personalisation.* Compatibility is a measurement of how harmonious a set of items is. In previous sections, we have mainly been focused on learning compatibility metrics between pairwise items or predicting the popularity of an outfit. Personalization, on the other hand, represents how the recommendations meet users' personal fashion tastes. In previous discussions, personalization is achieved relying on explicit user input (for example, image or text). However, this type of fashion generation works more like search than recommendation since it requires explicit user queries *(Chen et al., 2019).*

4.2.c. Fashion Forecasting

4.2.c.i. Abstracting

Fashion forecasting is a field that requires analysing temporal data, fashion trends, and style compatibility in order to achieve its goal of predicting fashion. This process of analysing data collected is called abstracting - recognizing similarities or differences across all garments and collections. Fashion forecasters abstract design collections across time periods to identify style transformations and predict the direction of the fashion industry *(Jia et al., 2018).*

4.2.c.ii. Multi-Task Convolutional Neural Networks for Long-Term Forecasting

Mall et al. (2019) suggested an automated, quantitative framework for long-term forecasting and discovery. They used annotations to train a multi-task CNN, such as GoogLeNet, where separate heads predicted separate attributes. For example, one head may predict "short-sleeves" whereas another may predict "mostly green". This training has the effect of automatically producing an embedding of images in the penultimate layer of the network that places similar clothing attributes and combinations of these attributes, thus referred to as "styles", into the same region of the embedding vector space. They took these attribute classifiers and applied them to the full unlabeled set of 7.7 M of people images.

Then they produced a temporal trend for each attribute in user city locations by computing, for each week, the mean probability of an attribute across all photos from that week and city. Per-image probabilities were derived from the CNN prediction scores after calibration using isotonic regression on a validation set. More discussion on this can be found in *(Mall et al., 2019)*.

4.3. How Deep Learning has changed how we approach these problems

From the above analyses, we study applications of various methodologies, involving hybrid multi-label CNNs, deep neural networks, conditional fields, visual, temporal, and consumer-interest attributes, k-fold cross evaluations, etc. These are most definitely not the only ways in which computer vision tasks in style analysis, fashion recommendations, and fashion forecasting take place. Without a doubt, there are multiple more methodologies, old and new, which can aid in achieving the above-stated tasks.

Where previously fashion trend analysis tasks required a predominance of human interaction, such as in the form of crowdsourcing, now these tasks can be accomplished in greater depth through deep learning in computer vision. While artificial intelligence and deep learning methods still contain errors, constant debugging into methods and further advancement in the stream helps us combat technological inaccuracies as well as human errors. Computer vision can also promote task completion in much shorter duration as compared to human methods. With further advancements in deep learning algorithms, we are able to solve more difficult fashion analysis tasks, as can be seen in the increasing complexities in computer vision applications in Section 3.

Developments in deep learning can aid in ground-breaking problem solving for future potential tasks. We have discussed fashion style analysis and recommendations, and have furthered the complexity interest in fashion predictions. Deep learning can also be employed for extremely convoluted tasks, such as fashion creativity, which will further be discussed in Section 5.

5. Computer Vision and Fashion Creativity: Future Directions

With further advances in fashion trend analysis, recommendations, and predictions, we may wonder what the next step toward technology in the fashion sphere is. We are working towards enhancing artificial intelligence in technology, but can we give computers the power of human creativity?

Creativity is an important skill that enables innovation. It is the fundamental base of design thinking, one of the core concepts that define human-computer interaction (HCI), which focuses on the interplay between humans and computers. Creative thinking varies depending on the individual, job, or environment. In fashion design, artistic creativity is crucial to making design results successful. It is highly associated with the quantity and quality of new ideas that designers can generate for a given design task *(Jeon et al., 2021)*. In fact, fashion design itself depends solely on creativity; design and creativity are directly proportional attributes without which both would lose their function.

Creativity involves divergent and convergent thinking. Divergent thinking develops new ideas by referring to various materials in order to augment or modify existing problem statements *(Jeon et al., 2021)*. For example, one can create new fashion design ideas based on existing research, and expand on particular fashion trends, while going one step ahead of fashion forecasting. On the other hand, convergent thinking progressively delimits one's research space and supports finding a design solution that is both novel and adapted to various constraints *(Jeon et al., 2021)*. For instance, one could narrow down their design statement to create a new fashion item or style, focused on immediate need-based attributes of people.

In accordance with the progression of the fashion stream, the inclination of this industry is steered toward efficiency. Consequently, creativity has played the role of providing a method of acceleration for obtaining this goal in the fashion industry. Design products are expected to be convenient, affordable, and comprehensible, all of which involve the notion of efficiency.

On the other hand, design also aims to cultivate an improved sense of well being of the society. These goals are more deeply related to the spiritual dimension than to material wealth. Design is a tool not based on the sole purpose of achieving efficiency. Design is an instrument to be discussed in greater depth, aimed to see the society of the next generation in better conditions *(Taura & Nagai, 2010).*

Design is a process of product evolution that is influenced by the decisions made by society. It involves a problem-solving process, which continually evolves temporally due to society's ever-changing needs, for the sole purpose of progress. One of today's biggest goals of the fashion industry is to promote inclusivity.

Fashion allows individuals to express who they are, demonstrated by their cultural roots. Global cultural inclusivity allows people to understand different perspectives and respect them. It helps eliminate negative stereotypes towards different cultural traditions. Especially in today's time, with greater connectivity and collaborations with people of different backgrounds, it is essential to bring forth cultural collaboration in society. Here, fashion can be influential in achieving this objective.

One way to make this possible is to employ fashion creativity in designing outfits that integrate different cultural aesthetics in otherwise current fashion trends. For example, incorporating cultural styles in formal wear in corporate jobs is a powerful way to accept cultural diversity in the work sphere. Employing Interactive Evolutionary Computation *(Xu et al., 2021)* is one method suggested for incorporating diverse cultural heritages in fashion. Analysing colour themes, patterns, cuts, and materials of traditional wear and merging them with the popular norm using deep networks allows great potential in achieving this target.

6. Conclusion

Computer vision's objective is to "see" visual data. In this paper, we studied its concepts and methodologies in the field of fashion trend analysis and its applications. We saw the role computer vision played in increasingly complex fashion tasks. Then, we reviewed some, and by no means all, deep learning methodologies in today's leading problem-statements of fashion style analysis, fashion recommendations, and fashion forecasting. We then discussed how deep learning impacted our approach to fashion trend analysis problems. Finally, we overviewed a newly introduced concept of fashion creativity, one which requires extensive study to solve potential tasks like cultural inclusivity in the fashion realm.

Acknowledgments

I would like to thank my advisor Catherine Wah for helping me with this project.

References

Al-Halah, Z., Stiefelhagen, R., & Grauman, K. (2017). Fashion forward: Forecasting visual style in fashion. *2017 IEEE International Conference on Computer Vision (ICCV)*. https://doi.org/10.1109/iccv.2017.50

Chen, W., Huang, P., Xu, J., Guo, X., Guo, C., Sun, F., Li, C., Pfadler, A., Zhao, H., & Zhao, B. (2019). POG: Personalized outfit generation for fashion recommendation at Alibaba iFashion. *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. https://doi.org/10.1145/3292500.3330652

Di, W., Wah, C., Bhardwaj, A., Piramuthu, R., & Sundaresan, N. (2013). Style finder: Fine-grained clothing style detection and retrieval. 2013 IEEE Conference on Computer Vision and Pattern Recognition Workshops. https://doi.org/10.1109/cvprw.2013.6

Forsyth, D. A., & Ponce, J. (2012). *Computer vision: A modern approach*. Pearson Education.

He, R., & McAuley, J. (2016). UPS and downs. *Proceedings of the 25th International Conference on World Wide Web*. https://doi.org/10.1145/2872427.2883037

Hsiao, W., & Grauman, K. (2018). Creating capsule wardrobes from fashion images. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. https://doi.org/10.1109/cvpr.2018.00748

Inoue, N., Simo-Serra, E., Yamasaki, T., & Ishikawa, H. (2017). Multi-label fashion image classification with minimal human supervision. *2017 IEEE International Conference on Computer Vision Workshops (ICCVW)*. https://doi.org/10.1109/iccvw.2017.265

Jeon, Y., Jin, S., Shih, P. C., & Han, K. (2021). FashionQ: An AI-driven creativity support tool for facilitating ideation in fashion design. *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. https://doi.org/10.1145/3411764.3445093

Jia, M., Zhou, Y., Shi, M., & Hariharan, B. (2018). A Deep-Learning-Based Fashion Attributes Detection Model. http://arxiv.org/abs/1810.10148

K, A. J. (2015). *Fashion forecasting*. Akhil JK.

König, R. (1973). *The restless image: A sociology of fashion*.

Lynch, A., & Strauss, M. (2007). *Changing fashion: A critical introduction to trend analysis and meaning*. Berg.

Mall, U., Matzen, K., Hariharan, B., Snavely, N., & Bala, K. (2019). GeoStyle: Discovering fashion trends and events. *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*. https://doi.org/10.1109/iccv.2019.00050

Schindler, A., Lidy, T., Karner, S., Hecker, M. (2017). Fashion and apparel classification using convolutional neural networks. *CEUR Workshop Proceedings.* https://arxiv.org/abs/1811.04374

Simo-Serra, E., Fidler, S., Moreno-Noguer, F., & Urtasun, R. (2015). Neuroaesthetics in fashion: Modeling the perception of fashionability. *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. https://doi.org/10.1109/cvpr.2015.7298688

Stephen, O., Maduh, U. J., Ibrokhimov, S., Hui, K. L., Abdulhakim Al-Absi, A., & Sain, M. (2019). A multiple-loss dual-output Convolutional neural network for fashion class classification. *2019 21st International Conference on Advanced Communication Technology (ICACT)*. https://doi.org/10.23919/icact.2019.8701958

Szeliski, R. (2010). *Computer vision: Algorithms and applications*. Springer Science & Business Media.

Tangseng, P., Yamaguchi, K., & Okatani, T. (2018). Recommending outfits from personal closet. *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*. https://doi.org/10.1109/wacv.2018.00036

Taura, T., & Nagai, Y. (2010). *Design creativity 2010*. Springer Science & Business Media.

Tautkute, I., Trzcinski, T., Skorupa, A. P., Lukasz, K., & Marasek, K. (2019). DeepStyle: Multimodal search engine for fashion and interior design. *IEEE Access*, *7*, 84613-84628. https://doi.org/10.1109/access.2019.2923552

Veit, A., Kovacs, B., Bell, S., McAuley, J., Bala, K., & Belongie, S. (2015). Learning visual clothing style with heterogeneous dyadic Co-occurrences. *2015 IEEE International Conference on Computer Vision (ICCV)*. https://doi.org/10.1109/iccv.2015.527

Vittayakorn, S., Berg, A. C., & Berg, T. L. (2017). When was that made? *2017 IEEE Winter Conference on Applications of Computer Vision (WACV)*. Published. https://doi.org/10.1109/wacv.2017.85

Xu, Q., Liu, H., Liu, Y., & Wu, S. (2021). Innovative design of intangible cultural heritage elements in fashion design based on interactive evolutionary computation. *Mathematical Problems in Engineering*, *2021*, 1-11. https://doi.org/10.1155/2021/9913161

Zhang, X., Jia, J., Gao, K., Zhang, Y., Zhang, D., Li, J., & Tian, Q. (2017). Trip outfits advisor: Location-oriented clothing recommendation. *IEEE Transactions on Multimedia*, *19*(11), 2533-2544. https://doi.org/10.1109/tmm.2017.2696825