Complementary Outfit Recommendation Using Deep Learning

S.Ananya[1], Vanshika Sandeep[2], Dr.G.Paavai Anand[3], N.Mahesh[4]

[1]Student, [2]Student, [3]Professor, [4]Student
[1]B-Tech, Computer Science Engineering,
[1]SRM Institute of Science and Technology, Chennai, India.

Abstract: Developing a system that provides a complementary outfit based on user queries can be challenging due to its complexity and subjectivity. At present, the methodologies available are used to recommend outfits to the users according to meta-data and user's location. So we present an empirical study on the application where we make use of Deep Convolutional Neural Networks (DCNN) to the task of classifying apparel and user's input from their wardrobe along with Convolutional Auto-encoder which is capable of finding similar product images with the aim to remove the usage of product tags and improve the recommendations of e-commerce platforms. Choosing the right pair of clothes according to the customer's choice and preference is of utmost importance. Sometimes people find it difficult to keep up with the trend so in such cases this project will be very helpful for them. Also, this helps the user to save their time and energy which implies that the user need not spend much time worrying about "what goes well with what". Hence we have come up with a system that is capable of resolving such difficulties and whose primary purpose is to make the user's work easier and beneficial.

IndexTerms **– outfit matching, fashion recommendation, fashion, convolution neural networks, autoencoder, deep learning.**

I. INTRODUCTION

In today's era, Fashion has taken topmost priority and has influenced all the people throughout the world. It brings power and cultivates the unique style in us. Both trend and style changes over time. As a result, fashion varies from time to time depending on several factors like season, culture, trend, comfort and occasion. For instance, Fashion in the 1970s for both men and women included bell-bottomed jeans since the preference was given for flared trousers [13]. On the contrary, in the present scenario, people's predilection for slim fit or low waist jeans is on the rise. Moreover, many styles tend to re-emerge after ages, like that of high waist jeans, loose sweatshirts and over-sized coolers which are currently penetrating into the fashion world and seem to be trending too.

Lately, online shopping has become a good way to approach fashion for the people. Since it can be accessed at any time and anywhere this enables people to use it even during their busy schedule. Hence making it easier for people to know more about fashion. The task of opting or suggesting the right kind of clothes is very rarely unique. Fashion helps in discovering the right choice of taste and passion for clothing. Therefore outfit recommendation plays an increasingly important role in the online retail market. The aim of the outfit recommendation is to encourage the interest and engagement of people in online shopping by suggesting trendy outfits that might be of interest to them. This project shows how the fashion sector can be improvised in the most simplest and efficient way. Thus we concentrate on providing such a convenient and user-friendly application.

Having followed the fashion sector closely so far, we have come across two important domains that haven't been explored much. One would be providing clothing recommendations for images of clothes in the user's wardrobe. Most e-commerce websites have catalogues put up by various retailers so there is no room for closet attire. So if a person wants to know what matching outfit might go well with their closet attire, it is ideally not possible. Added to this, these websites require the user to type keywords of an item that they are most likely to purchase and tries to retrieve similar products based on meta-data or tags that are stored in the database. For example, if a user is searching for a "Blue Turtleneck Shirt" the ideal scenario would be that of displaying shirts that are both Blue and Turtleneck. After which it displays a shirt that is either Blue only or those that are Turtleneck but not Blue. These are some close matches to the user's expectations. But what if some of these tags or meta-data are incorrect, after all, it is typed by humans and usually in case of e-commerce websites it is done by retailers. Since the website is prone to human errors, there is a chance that search queries might not fetch proper results. Moreover, in some cases, retailers try to add extra content as tags which could be irrelevant to the product's description just to make sure that their item always pops up at the top of the website irrespective of whatever is being searched by the user. To avoid such situations, we have come up with an approach whose main contributions are as follows:

- Our work shows the importance of and need for providing complementary outfit recommendations based on the user's input.

- With our system, the input query is an image and not only does it provide matching recommendation for clothes available online but also for wardrobe apparel taken by the user.

- We have automated the entire system's operation using Computer Vision and Neural Networks to avoid search by tags and other possible human errors.

II. LITERATURE SURVEY

There is no previous work related to our proposed system. Hence, we will briefly review the other contributions or related research in the Fashion domain which includes fashion compatibility, Classification of clothing attire and location-based shopping, etc.

Gupte S et al [1] has proposed an approach for learning fashion compatibility between apparels to recommend matching outfit which is based on a Siamese network used for extraction of features followed by a fully-connected network used to learn a performance metric for fashion. The embedding developed by the Siamese network is inspired by the important role color plays in determining fashion compatibility with color histogram features. Further, network training is formulated as a maximum a posteriori (MAP) problem where Laplacian distributions are assumed for Siamese network filters to facilitate sparsity and matrix-variate normal distributions are assumed for metric network weights to effectively exploit correlations between the input units of each linked layer.

Lao B et al [3] came up with a system where they mainly focus on four tasks (1) Multiclass classification of clothing type- Classifying the clothing type is the multi-class classification problem of predicting a single label that identifies the type of clothing in the picture. They have made use of the Apparel Classification with Style (ACS) Dataset, which contains 89,484 photos from boundary boxes to encapsulate the clothing of a person. (2) Clothing attributes classification- Clothing attribute classification is used for assigning attributes like color or pattern to a clothing article. (3) Clothing retrieval of nearest neighbors- Retrieval of clothing involves the task of identifying the most appropriate or closely related objects in clothing to a question item. (4) Clothing objects detection- Detection of clothing objects involves detecting the different regions of the clothing objects present in a given image. Consider an image of a person wearing a complete outfit, analyzing clothing artefacts involves predicting bounding boxes that would capture the distinct clothing items such as the shirt, pants, and shoes. Here, they deal with advancing the prevalent techniques and algorithms to improve the forecasts and hints of good recommender structures by making use of Convolution Neural Network. However, they are not providing customized matching outfit to the user.

Lin Y et al [4] proposed a method for explainable outfit recommendation with joint outfit matching and comment generation wherein they have used two neural outfit recommendations (NOR), one for outfit matching and another one for comment generation. For outfit matching, they proposed a convolutional neural network with a mutual attention mechanism to extract visual features. The visual features are then decoded into a rating score for the matching prediction. For abstractive comment generation, they proposed a gated recurrent neural network with an across-modality attention mechanism to transform visual features into a concise sentence. The research drawbacks include the fact that NOR never produces negative comments to justify why an outfit doesn't suit, which is because most of the comments in the dataset are positive. In addition, since short comments account for a large percentage of the dataset, NOR appears to produce short comments.

Zuva T et al [2] proposed a location-based shopping recommender system for mobile users wherein they recommend similar products based on the user's input image. Based on the user's location the products are suggested from the near-by shops that are available. In summary, the main contribution of this paper is to show how image retrieval, image content, and camera-enabled smart mobile device with GPS capabilities can be used to realize a location-based shopping recommender system for mobile users. This does not make the system user-friendly and also there is no usage of neural networks.

Kang W et al [5] proposed a new task called 'Complete the Look' which aims to suggest visually compatible products based on images from the scene. For this mission, they have developed an approach to collect training data, and suggest a new way to learn the quality of the scene-product from photos of fashion or interior design. Their approach measures compatibility, globally and locally, through CNNs and mechanisms of attention. In the future, for more fine-grained compatibility matching, they aim to implement object detection techniques to identify key objects.

III. RESEARCH METHODOLOGY

The proposed system comprises two main modules i.e. Image Classifier and Matching Recommendation. The image classification is carried out using Convolution Neural Network (CNN) which is used for classifying the type and color of clothes provided by the user. The second module is the matching recommendation which makes use of Convolutional Auto-encoder to find a similar pair of clothes and provides a matching outfit accordingly. Hence the two modules make up an efficient and desired fashion recommender system. To enhance the recommendation, we make use of Generative Adversarial Networks (GAN) as an add-on to the system to generate new outfits that might go well with the input image given by the user. For better understanding, the workflow diagram is shown below.

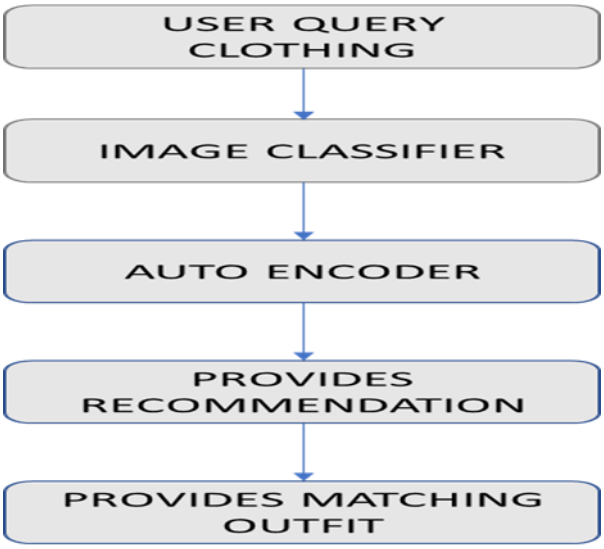

Fig 1: System Workflow

From the above diagram, it is understood that the user sends a query image to the system as an input which is then recognized by the image classifier and is further compared with the catalogue using an autoencoder to provide recommendation and displays a matching outfit to the user as an output. The entire working of our system is segregated into 3 parts:

3.1 Image Classifier

A Convolutional Neural Network also known as ConvNet or CNN is a Deep Learning algorithm that can take in an input image, assign important learnable weights and biases to various aspects or objects in the image and differentiates one from the other. The pre-processing required in a ConvNet is much lower as compared to other classification algorithms. For these reasons we make use of Convolution Neural Network for image classification and have modified it to satisfy our purpose. The model is built by coupling two CNNs back to back where one is for identifying the type of clothes and the other is used for recognizing the color of the input image. For example, using the first Convolution Neural Network the system identifies if the given image is a pant or a shirt while the other identifies the color and the texture of the clothing and gives a {category, color} pair as output. This method is more convenient and beneficial to classify the multiple outputs since there is no need to provide images from all the classes that we are trying to predict as it learns to classify each of the classes separately. It is not required to provide data for all the possible combinations of {category, color} classes since it has the ability to generalize better. In addition, both CNNs are trained simultaneously which reduces space and time complexity. For the classifier to function decently we have used 16 classes with 4 categories and 11 colors. The images comprise of men wear and were augmented so as to make the system more efficient and useful while predicting the class labels.

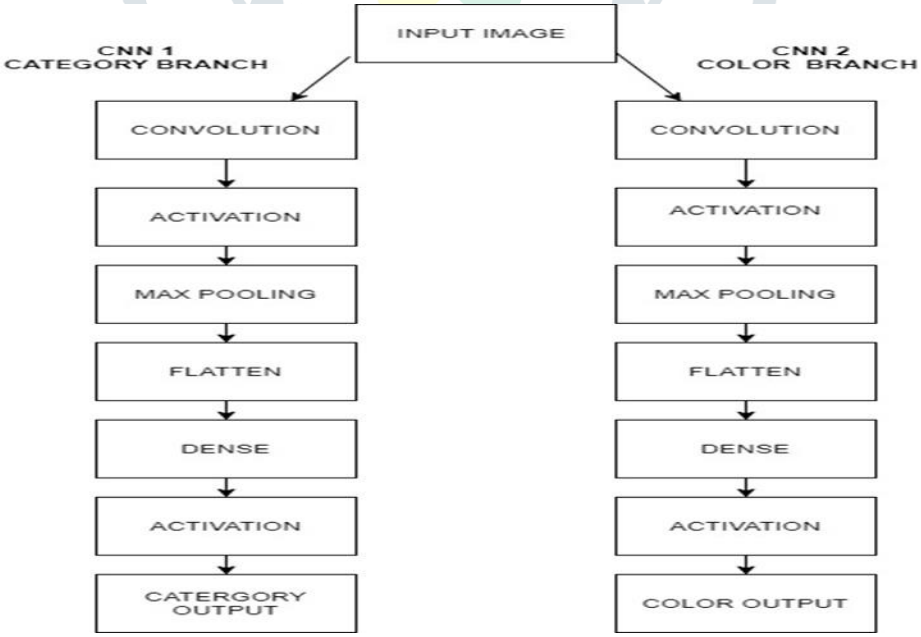

Fig 2: Convolution Neural Network

3.2 Matching Recommendation

Once the user input is predicted by the image classifier it is then provided to the Convolutional Autoencoder. An Autoencoder is a special form of neural feedforward networks, where the input is the same as the output. They compress the input into a lower-

dimensional code and then recreate this representation's output. The code is a compact "description" or "compression" of the data, often called the representation of latent-space. Autoencoder comprises both an encoder and a decoder. It compresses the input data with the encoder then decompresses the encoded data with the decoder such that the output is a perfect reconstruction of the original input data. Being an unsupervised learning algorithm, they do not need to train on specific labels. To be more specific, they are self-supervised, since the training data produce their own labels.

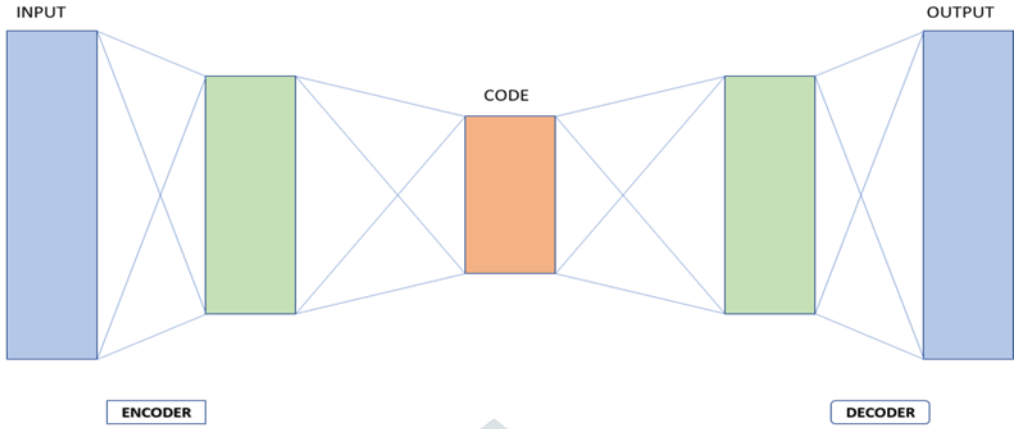

Fig 3: Autoencoder Architecture

With Convolution, Pooling layer and Activation they become Convolutional Auto-encoder which can be used for finding similar images. If we lower the encoding dimensionality, then convolutional auto-encoder can learn the most salient features of data during training and ignores noise. We train the model on catalogue images with the objective of minimizing reconstruction loss. Then the system makes use of reconstructed images and applies k-nearest neighbors (KNN) algorithm on image embedding to retrieve top K similar images from the catalogue which is an important working aspect. The catalogue comprises men's wear which includes the following: formal pants, formal shirts, casual t-shirts and jeans of some prominent colors. After comparing it with the catalogue, the matching outfit is fetched and displayed to the user as an output.

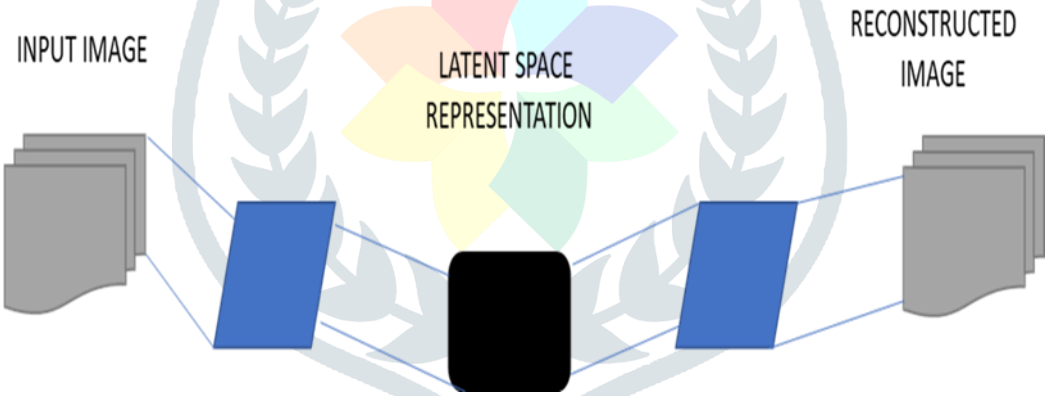

Fig 4: Convolutional Autoencoder

3.3 Generative Adversarial networks

In this system, a Generative Adversarial network is used as an additional feature to the system to come up with new outfits that would go well with the input. Using GAN's system generates matching apparel for the input image, and the main objective is to produce a costume with different kinds of patterns that would match the input image. For example, the system makes use of patterns like striped or checked in case of a shirt and ripped or rugged in case of jeans to come up with new recommendations apart from catalogue. This results in developing a more personalized recommendation system.

IV. RESULTS AND DISCUSSION

Both the modules were evaluated separately and we present the results through the accuracy graph which is plotted over a series of epochs.

4.1 Classifier

For the classification task, two Convolution Neural networks were trained for category and color over 50 epochs. The dataset comprised of 8351 fashion images which were split into 8:2 ratios for training and validation respectively.

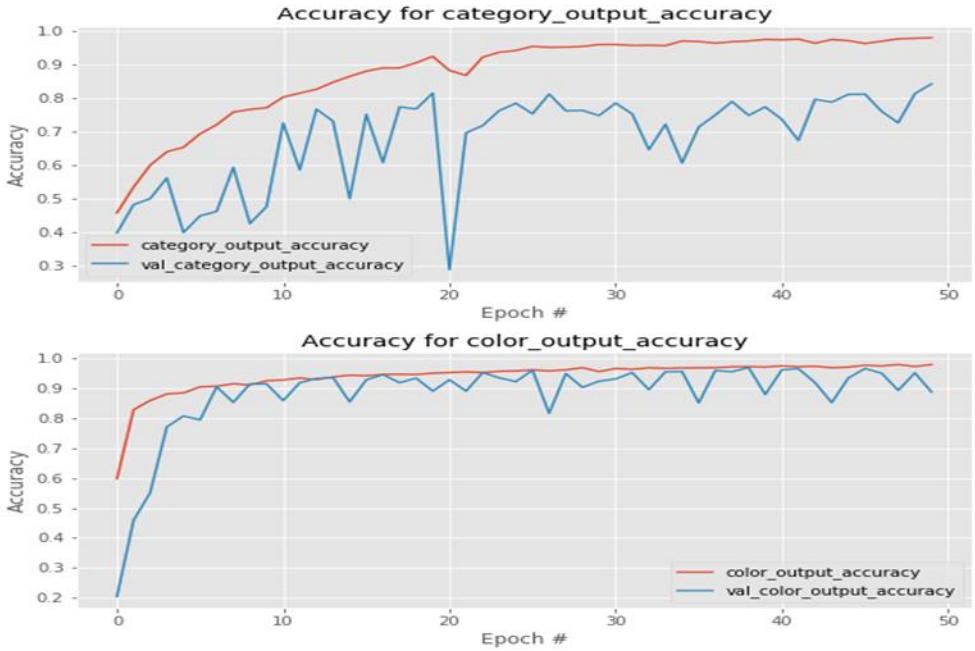

Fig 5: Classifier Accuracy

It is evident from the above graphs that we get training and validation accuracy somewhere around 97% and 85% respectively for category CNN and on the other hand we get a score of around 98% and 90% respectively for color CNN. Even though the color CNN seemed to have performed better, we have attained an overall exceptional result.

4.2 Recommender

For the sole purpose of providing a matching outfit, we have used Convolutional Autoencoder which was trained over 500 epochs. We made use of 2809 clothing images which was split into 8:2 ratios for training and validation respectively and obtained the following graph.

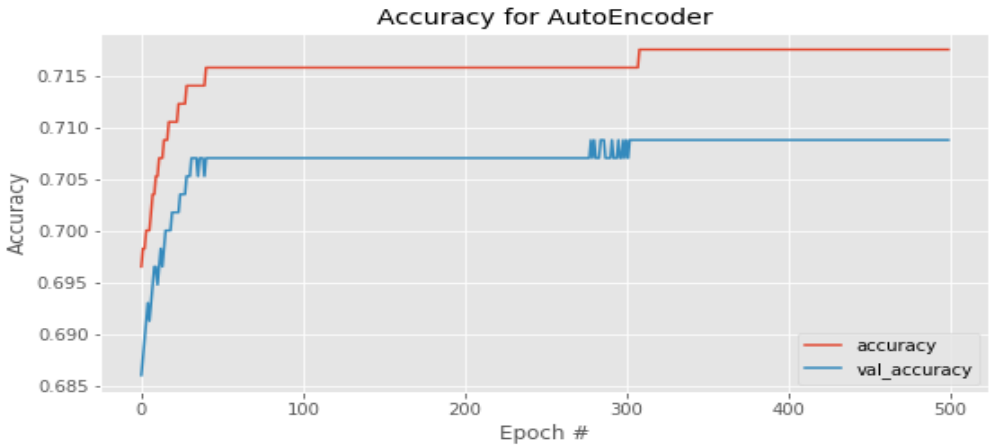

Fig 6: Convolutional Autoencoder Accuracy

The accuracy for training and validation set turns out to be around 72% and 70% respectively. However, the model has been maintaining a consistent accuracy of 70%. Thus the two modules have resulted in a satisfactory outcome.

4.3 Survey

Extensive experiments show that the performance of our model was impressive for the given test data. We collected input data from people including both shirt and pant of various categories and colors. Based on the results we conducted a survey for

human evaluation and qualitative assessment to understand system behavior for which we recorded responses from 51 people. We made people rate our system out of five and the following were analyzed:

- The final outcome was satisfactory for majority of the people as none of the ratings were below three which suggests that the system's performance was better than average.

- Further study suggests that the age range of the respondents were between 19 and 50 which indicate that the system was liked by all the age groups.

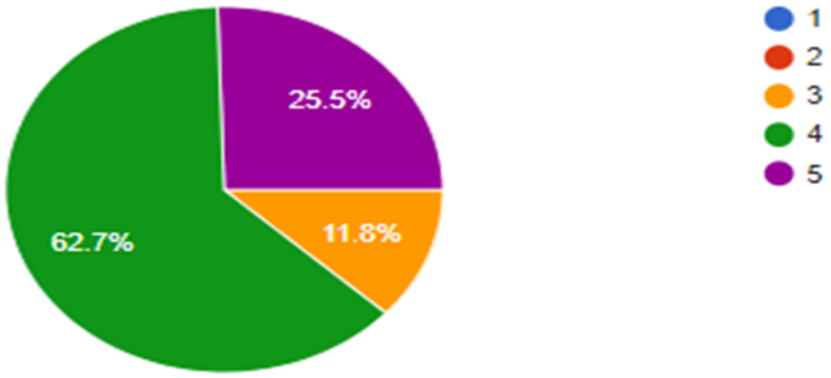

Fig 7: System Rating-On a Scale of 5

We believe that our work can be used in e-commerce websites or any other platform with minimal human effort and errors. With GANs, this application can be used by the textile industry to come up with innovative designs.

V. CONCLUSION AND FUTURE SCOPE

We have presented a novel approach based on deep neural network which provides matching apparel based on user input. Unlike the methods that are available which makes use of semantic data and location to retrieve images for a text-based query, our system focuses on using computer vision to automate the conventional methods. As an add-on to the system, we try to incorporate Generative Adversarial Networks (GAN) for the generation of customized products for the clothes provided which makes the system much more effective and efficient. Thus saving time and resources.

Currently, the fashion recommender focuses on male clothing only, which can be improvised by making it available for all age groups irrespective of gender which makes it more efficient and advanced. Providing fashion recommendations according to the season, region and occasion can be improvised. Also making use of GANs more efficiently would be useful. Along with the outfit, matching accessories can also be recommended.

REFERENCES

[1] Luisa F. Polania, Satyajit Gupte, Learning fashion compatibility across apparel categories for outfit recommendation, September 2019
[2] Tranos zuva, Oludayo o. olugbara, Sunday o. ojo and Seleman m. ngwira , Image content in location-based shopping recommender systems for mobile users, July 2012.
[3] Brian lao, karthik jagadeesh, convolutional neural networks for fashion classification and object detection.
[4] Yujie lin, Pengjie ren, Zhumin chen, Zhaochun ren, Jun ma and Maarten de Rijke, explainable outfit recommendation with joint outfitmatching and comment generation, 2019-03-05.
[5] Wang-Cheng Kang, Eric Kim, Jure Leskovec, Charles Rosenberg, Julian McAuley, Scene-based Complementary Product Recommendation.
[6] Sagar Verma, Sukhad Anand, Chetan Arora, Atul Rai, Diversity in Fashion Recommendation using Semantic Parsing, July 2010
[7] Qingqing Tu, Le Dong, An Intelligent Personalized Fashion Recommendation System, July 2010
[8] Ying Huang, Tao Huang, Outfit Recommendation System Based on Deep Learning, July 2016
[9] Hessel Tuinhof, Clemens Pirker, Markus Haltmeier, Image Based Fashion Product Recommendation with Deep Learning, July 2018
[10] Alexander Schindler, Thomas Lidy, Stephan Karner, Matthias Hecker, Fashion and Apparel Classification using Convolutional Neural Networks, November 2018
[11] Luyang Chen, Fan Yang, Heqing Yang, Image-based Product Recommendation System with Convolutional Neural Network.
[12] Hilary George-Parkin, The Most Influential Fashion Trends, Decade by Decade
[13] History of Fashion – A brief Story of the Evolution of Fashion, February 2020