Proceedings of the 5th International Conference on Data Intelligence and Cognitive Informatics (ICDICI-2024)
IEEE Xplore Part Number: CFP24VL6-ART; ISBN: 979-8-3503-8960-9

2024 5th International Conference on Data Intelligence and Cognitive Informatics (ICDICI) | 979-8-3503-8960-9/24/$31.00 ©2024 IEEE | DOI: 10.1109/ICDICI62993.2024.10810981

Bridging the Gap Between Online and In-Store Shopping: Fashion Recommendations and Virtual Try-On

Vaishnavi V V
School of Computer Science and Engineering
Vellore Institute of Technology
Chennai, India
vaishnavi.vv01@gmail.com

Narendra G O
School of Computer Science and Engineering
Vellore Institute of Technology
Chennai, India
go.narendra@outlook.com

Aravinda Boovaraghavan
School of Computer Science and Engineering
Vellore Institute of Technology
Chennai, India
aravinda992@gmail.com

L. Jani Anbarasi
School of Computer Science and Engineering
Vellore Institute of Technology
Chennai, India
Janianbarasi.l@vit.ac.in

***Abstract*—The development of extensive clothing databases has significantly advanced the field of clothing recognition and recommendation systems. However, existing datasets often suffer from a limited number of annotations and face challenges when applied to diverse, real-world scenarios. To address these limitations, the proposed research work leverages DeepFashion, a comprehensive large-scale garment dataset renowned for its detailed annotations. The primary objective is to develop an advanced fashion recommendation system combined with a virtual try-on feature using cutting-edge deep learning techniques. This system is designed to enhance the online shopping experience by providing personalized clothing recommendations and enabling users to visualize how the recommended garments would appear on them. By inputting an image, users can receive recommendations of similar clothing items by using powerful feature extraction capabilities of the ResNet50 model and the efficient retrieval mechanism of the Nearest Neighbor algorithm. Following the recommendation phase, the system employs the U2Net model for image segmentation to facilitate the virtual try-on feature. U2Net's two-level layered U-structure architecture performs well in detecting salient objects, producing high-resolution segmented images that isolate the clothing item from the background and the human body in the original input. This segmentation is crucial for accurately overlaying the recommended garment onto the user's image, creating a realistic virtual try-on experience. Following the recommendation phase, the system employs the U2Net model for image segmentation to facilitate the virtual try-on feature. The proposed research aims to address the problems of current fashion recommendation systems by utilizing the extensive and richly annotated DeepFashion dataset and integrating advanced deep learning models.**

Keywords—Fashion Recommendation System; Virtual try-on; Deep Learning; ResNet-50; Segmentation; U2Net; OpenCV

I. INTRODUCTION

The fashion industry has consistently been a dynamic and ever-changing field, influenced by cultural trends, technical breakthroughs, and consumer preferences. The way buyers engage with fashion has changed dramatically in recent years because to the incorporation of deep learning techniques and artificial intelligence. Virtual try-on technology and fashion recommendation systems stand out among these advancements as game-changing tools that improve the shopping experience by making it more interesting and personalized [1].

A person's outer look can be used as a medium to convey their interior perceptions through clothing. It provides details about their interests, convictions, character, occupation, social standing, and outlook on life. As a result, clothing is regarded as an important aspect of a person's external appearance and an indirect means of communication. Several deep learning-based retail systems have been established in recent years to give some consumer convenience. Clothing suggestion systems, for example, frequently provide shoppers either an entire outfit or another fashion item that is compatible with particular items [2]. The outfit-match recommendation system models the style characteristics of clothes to recommend fashion items that make good matches with given clothes, like a shirt and matching trousers. The whole outfit recommendation system suggests a collection of clothes that make up whole outfits with similar attributes, like patterns and colours.

With the recent technological breakthroughs, consumers can now monitor present fashion trends throughout the world, which influences their purchasing decisions. Numerous factors, such as age, gender, culture, individual preferences, geographic location, and interpersonal interactions, all have an impact on consumer fashion choices. The combination of fashion preferences and the previously described characteristics linked with clothing selections may convey visual attributes for a better understanding of customers' preferences. As a result, evaluating consumer preferences and recommendations is beneficial to fashion designers and merchants. As a result, in recent years, e-commerce has emerged as the dominating retail channel. The capacity of recommendation systems to give individualized suggestions and respond swiftly to customer selections has greatly aided the growth of e-commerce sales [3].

Virtual try-on technology complements the recommendation system by allowing users to visualize how recommended clothing items would look on them without physically trying them on. This is achieved through

sophisticated computer vision and augmented reality techniques that create realistic representations of garments on the user's body. The ability to virtually try on clothes addresses several key challenges in the online shopping experience, such as uncertainty about fit and appearance, thereby reducing return rates and enhancing user confidence in their purchases.

In this work, a fashion recommendation system and virtual try-on application is created using deep learning techniques, leveraging the DeepFashion dataset. The main objective is to create a system that not only generates personalized clothing recommendations based on user input images but also allows users to virtually try on the recommended items. By integrating these functionalities, the online shopping experience is enhanced, making it more interactive, convenient, and satisfying for consumers

II. Literature Review

A collaborative recommendation system for the fashion domain has been analyzed by various researchers. Apart from the usual image metrics, a unique metric called the "trend score" has been introduced which takes into account the trendiness of a product by considering the ratings provided by the system's users. Trend patterns include colors, prints, and materials. Apart from recommendations, this score is also used to sort products from each category, from trendies to classic ones [4]. A content-based recommendation approach has been proposed where a relevant image is recommended based on quality queries of the clothes and footwear using the Euclidean similarity score [5]. An alternative method takes the user's preferred style into account. The process of creating the outfit image involved removing the clothing's colour, texture, and shape from both the segmentation mask and the original image. In order to classify the image according to the user's preference, the user's original dataset is built by asking the user to categorize a series of images into a set of styles in order to understand the user's preference beforehand. [6].

The CNN feature vector of a fashion item, or clothes vector, provides two types of information: category and style. The category denotes the shared characteristics of the clothes in the same class, while the style reveals the unique feature of the clothes. Nevertheless, the clothes vector often suggests mismatched clothing because of the inconsistent information on style and category. To address this issue, a style feature extraction (SFE) layer is provided that efficiently breaks down the garments vector into style and category [7]. According to the experimental results, the suggested approach produces state-of-the-art outcomes for link prediction, a performance indicator for a fashionable match. This work proposes the SFE layer, a unique feature layer for clothing style extraction. In terms of link prediction, the proposed SFE layer network performs better than a traditional state-of-the-art network.

A new deep model called FashionNet, is proposed that learns clothing features by simultaneously predicting landmarks and garment qualities, in order to illustrate the advantages of DeepFashion. Next, the learnt characteristics are pooled or gated using the estimated landmarks [8] along with Iterative optimisation. Numerous tests show how useful DeepFashion is and how effective FashionNet is. To support future research, DeepFashion, a fully annotated clothes dataset with large characteristics, clothing signs, and cross-pose correspondences of clothing pairings, has been released. Additionally, evaluation procedures are outlined and benchmark datasets for three commonly recognized apparel detection and retrieval tasks.

A two-stage deep learning framework in is proposed which makes recommendations for fashion images based on input images with comparable styles. A neural network classifier extracts visually-aware feature extractor for making suggestions for fashion products based on image data. This technique is based on visual data, unlike the typical content-based recommendation model, which is mostly focused on descriptive metadata like manually annotated product tags or user reviews. This study proposed a method that uses a modified version of the k-nearest neighbours (k-NN) algorithm for feature space ranking and a trained CNN classifier as an image feature extractor. A visually conscious, data-driven, very straightforward, yet nonetheless powerful recommendation engine was proposed for images of fashion items. The proposed two-stage method uses a CNN classifier to extract features, which are subsequently fed into a similarity recommendation system. By enabling users to contribute a particular fashion image and then presenting comparable products based on the texture and category elements of the uploaded image, it can be utilized in several contexts [9].

III. Proposed Work

In recent years, online shopping has surged, especially in the fashion industry. However, a gap remains between online and in-store experiences due to limited personalized recommendations and the inability to try on garments. Current fashion recommendation systems often rely on inadequate datasets and simplistic algorithms, failing to capture user preferences and garment variations. Additionally, the lack of an accurate virtual try-on feature hinders the ability to visualize how recommended items will look when worn, resulting in a less immersive shopping experience.

This research introduces an integrated fashion recommendation and virtual try-on system that leverages advanced deep learning techniques to enhance the online shopping experience. The system comprises the following components:

- *Data Utilization:* The system utilizes the DeepFashion dataset, a large-scale and richly annotated fashion database, to ensure comprehensive coverage of various garment types and styles.
- *Feature Extraction:* The ResNet50 model is employed for extracting detailed features from garment images, ensuring that the recommendations are personalized and relevant to the user's preferences.
- *Garment Segmentation:* The U2Net model is used for precise segmentation of garments from the background, which is crucial for the virtual try-on process.
- *Virtual Try-On*: After segmentation, the system allows users to virtually try on the recommended garments, providing a realistic visualization of how the clothing would look when worn.

A. Dataset

The DeepFashion database [8] is the largest fashion analytics dataset in image form. DeepFashion is a large-scale clothing database that contains over 800,000 fashion images, from well-posed shop images to modelling images. It is also annotated with rich information like cloth category,

descriptive attributes, bounding box and clothing landmarks. The descriptive attributes provide information on fabric type, color, pattern, and style, while the bounding boxes and landmarks help in precisely locating and identifying clothing parts within an image. Such comprehensive annotations enable the development of advanced algorithms capable of accurate clothing recognition and recommendation.

B. Architectural Design

The proposed work offers an approach to overcome the difficulties in developing a virtual try-on application and an efficient fashion recommendation system. To improve the online shopping experience, clothing goods must be accurately recommended and visualized. Nevertheless, fashion images sometimes have inconsistent backgrounds, lighting, and positions, which presents a challenge for recommendation systems. Therefore, the goal of this effort is to integrate cutting-edge deep learning algorithms to enhance the realism of virtual try-on experiences and the accuracy of clothing recommendations.

Initially, ResNet50 [10] architecture, a powerful convolutional neural network renowned for its capacity is used to extract high-level features from images, for feature extraction. The recommendation results should be more accurate as a result of the detailed attributes extracted from garment images. Furthermore, using the feature embeddings resulted by ResNet50, recommendations are generated using the Nearest Neighbours algorithm. By identifying and recommending to the user the most similar clothing items, this algorithm makes sure that the suggestions are more relevant.

U2Net [11] model is applied to perform the segmentation process. U2Net is a well-known image segmentation model that is very successful at precisely separating objects from their backgrounds. The proposed methodology accurately and reliably separates the clothing items by utilizing U2Net, which is important for the virtual try-on process. The segmented image enhances the accuracy and realism of the virtual try-on experience by concentrating only on the clothing item.

The proposed methodology is broken down into three main stages: Feature Extraction stage, which uses ResNet50 to handle the dataset and extract high-level features; the Recommendation stage, which uses the Nearest Neighbours algorithm to generate recommendations for similar clothing; and the Segmentation and Virtual Try-on stage, which uses U2Net to segment the input image and produce an accurate virtual try-on experience. This integrated method guarantees an uninterrupted and intuitive process, greatly improving the users' online buying experience.

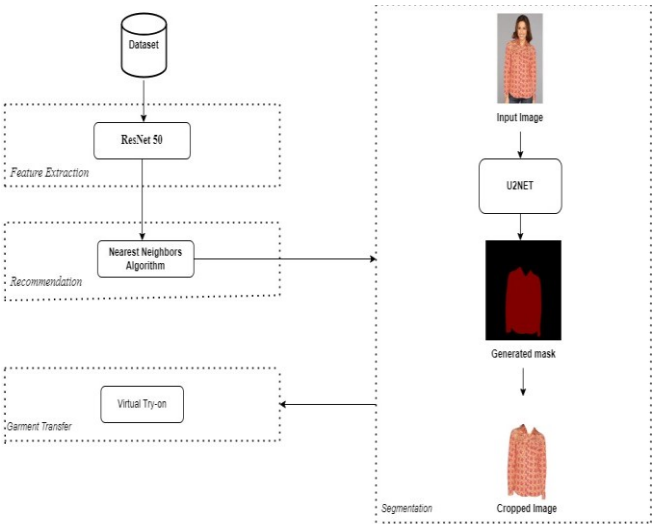

Fig. 1. Overall Architecture

C. Feature Extraction and Recommendation

A comprehensive fashion recommendation system is proposed in this study that makes use of cutting-edge deep learning and machine learning approaches. The ResNet50 model is utilized to gather high-level features from images of clothes, effectively capturing the minute intricacies and patterns that differentiate different fashion items. The Nearest Neighbour algorithm works by figuring out how far apart the embeddings of the input image are from those in the dataset. The user is then suggested images based on which the images with the smallest distances are deemed to be the most similar. This methodology guarantees that the suggestions are exceedingly pertinent and customized to the user's inclinations.

A class of deep neural networks called convolutional neural networks, or CNNs, is frequently employed for the analysis of visual imagery. CNNs use a sequence of convolutional, activation, pooling, and fully connected layers to automatically and adaptively learn the spatial hierarchies of features from input images.

a) Convolutional Layers:

A class of deep neural networks called convolutional neural networks, or CNNs, is frequently employed for the analysis of visual imagery. CNNs use a sequence of convolutional, activation, pooling, and fully connected layers to automatically and adaptively learn the spatial hierarchies of features from input images.

Convolution operations are applied to the input image or feature maps by convolutional layers, which primarily serve to capture local characteristics like edges, textures, and patterns. The convolution operation for a single filter can be expressed mathematically as shown in Eq. 1:

$$(f * g)(i,j) = \sum_{m}^{i} \sum_{n}^{j} f(m,n).g(i-m, j-n) \quad (1)$$

where $'g'$ stands for the filter (or kernel), $'f'$ for the input image, and i,jfor the spatial coordinates. To create the output feature map, the convolution operation entails sliding the filter over the input image and carrying out element-wise multiplication and summing. Convolutional layers typically have three parameters: stride (the step size the filter takes

across the input), padding (adding zeros around the input to regulate the spatial dimension of the output), and filter size (usually 3x3, 5x5, or 7x7).

b) Activation Functions: The network gains non-linearity via activation functions, which helps it to understand complex patterns and representations. Rectified Linear Unit (ReLU) activation function is defined as Eq. 2:

$$\mathrm{ReLU}(x) = \max(0, x) \tag{2}$$

ReLU sets all negative values to zero while maintaining positive values, which helps avoid the vanishing gradient issue and enables the network to learn more quickly and operate more effectively.

c) Pooling Layers: By reducing the spatial dimensions of the feature maps, pooling layers help to lower the network's parameter count and computational load. The network becomes invariant to minor translations as a result of pooling. There are two main types of pooling: average pooling, which uses the average value from each feature map patch, and max pooling, which uses the maximum value from each patch as indicated in Eq. (3).

$$y = \max(x_1, x_2, \ldots, x_n) \tag{3}$$

The typical parameters of pooling layers include the pool size (2x2) and the stride as 1.

a) Fully Connected Layers: In order to identify the input image, fully connected layers, often referred to as dense layers, incorporate the characteristics that the convolutional and pooling layers have learnt. Every neuron in a layer that is fully connected is linked to other neuron in the layer before it, executing an activation function after a linear transformation as shown in Eq. 4:

$$y = \mathrm{Activation}(W.x + b) \tag{4}$$

where the input vector is x, the bias vector is b, and the weight matrix is represented by W. For a given job, the activation function may be ReLU, softmax, or another appropriate function.

D. ResNet Architecture

ResNet (Residual Network) enhances gradient flow during backpropagation and resolves the problem of vanishing gradients in deep networks by adding residual connections. The residual block, the main building block of ResNet, is expressed mathematically as given in Eq. 5:

$$y = F(x, \{W_i\}) + x \tag{5}$$

where , x is the input, $F(x, \{W_i\})$ represents the residual function (a stack of convolutional layers with weights W_i), and y is the output. Residual blocks can be of two types: identity blocks, where the input x is directly added to the output of the residual function, and convolutional blocks, where the input x is modified before being added to the output of the residual function. In particular, ResNet50 is an architecture consisting of a convolutional layer at the beginning and a sequence of stage-organized residual blocks after that. Multiple residual blocks with a predetermined number of filters make up each stage. The network terminates with a fully connected layer that employs a global average pooling layer and a softmax activation function for classification.

E. Detailed Feature Extraction and Recommendation Process

a) Feature Extraction: The user inputs an image I of a clothing item, which is then passed through the ResNet50 model. The ResNet50 model processes I and extracts a high-level feature vector f, capturing essential characteristics of the clothing item:

$$f = \mathrm{ResNet50}(I) \tag{6}$$

b) Nearest Neighbor Algorithm: The Euclidean distance between the feature vectors in the dataset f_{input} and the feature vector of the input image f_{input} is computed in order to identify comparable clothing items as shown in Eq. 7:

$$d(f_{input}, f_i) = \sqrt{\sum_{j=1}^{n} (f_{input,j} - f_{i,j})^2} \tag{7}$$

where, n is dimension of the feature vectors.

The feature vectors with the smallest distances d are identified as the most similar and are recommended to the user, ensuring that the suggestions are highly relevant to the user's preferences.

F. Segmentation

a) U2Net Architecture: U2Net employs a two-level U-structure, effectively capturing both high-level and low-level features from the input image. The process starts with an input image of dimensions 256x256x3 (height, width, and color channels). Multiple convolutional layers with 3x3 filters make up the encoder, which is then followed by ReLU activations. These layers are designed to gradually reduce the image's size while preserving its key characteristics. Max pooling operations (2x2) follow each downsampling step to further reduce the spatial dimensions of the feature maps. The encoder captures the relevant features at various levels of abstraction by arranging itself into blocks that go deeper and deeper. Skip connections (shown as horizontal arrows) transfer feature maps from the encoder to the decoder at corresponding levels. This helps in preserving spatial information and allows the network to learn finer details, aiding in accurate segmentation. The decoder mirrors the encoder structure that performs upsampling operations using Conv2DTranspose layers (2x2) to progressively restore the original image dimensions. Each upsampling step is followed by convolutional layers (3x3) and ReLU activations, gradually reconstructing the segmented object. The final feature map, which provides rich spatial and contextual information, is created by fusing the concatenated feature maps from skip connections and upsampling layers together. The feature maps are flattened and run through a softmax activation function in the final output layer, which is made up of dense layers and produces a segmentation mask. This mask denotes the existence of background areas and salient things, such as clothing. Four channels make up the segmentation mask generated by the U2Net model: backdrop, upper body, lower body, and entire body. The precise segmentation of

various sections of the clothing item is guaranteed by this multi-channel technique.

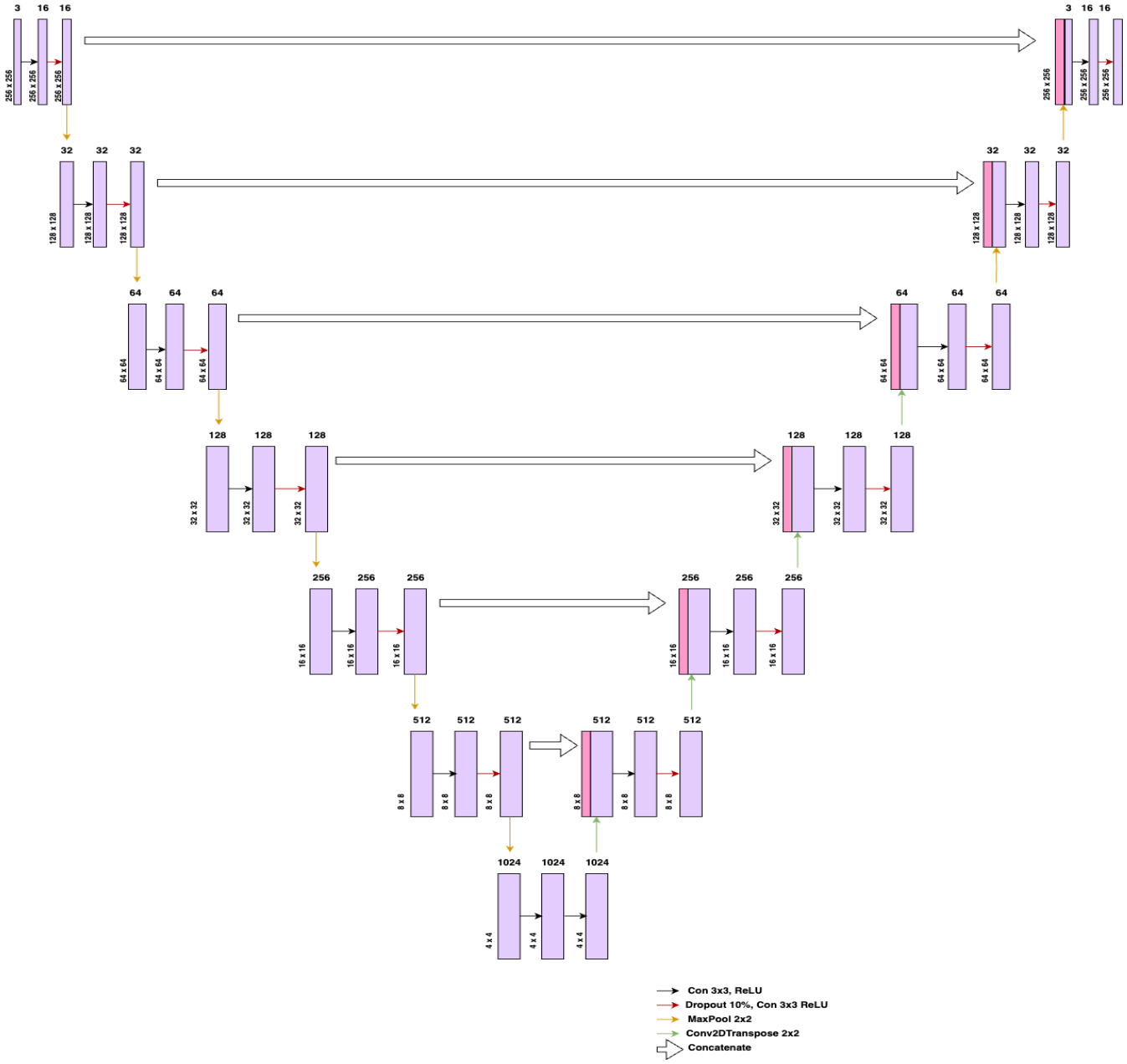

Fig. 2. U2Net Architecture

b) Segmentation Process: Preprocessing is done on the input image to make it the proper size (256x256x3) for the U2Net model. The pre-processed image is fed via the encoder, decoder, and output layers in a forward pass through the U2Net model. The model's output is a segmentation mask that shows the boundaries between the background and the piece of clothing. To improve the accuracy of segmentation, categorical cross-entropy loss is used for every checkpoint. After applying the mask to the input image, the segmented clothing item is the only object that remains after the background and human body have been removed. Subsequent uses of this generated image include in-depth analysis and virtual try-on. The clothing items are properly isolated from the background due to the highly accurate and efficient segmentation procedure made possible by the pre-trained U2Net model. In order to give users a smooth and lifelike virtual try-on experience and improve their engagement with the fashion suggestion system, segmentation is essential.

G. *Virtual Try-on*

The proposed solution increases time efficiency and improves access to garment try-ons by creating a virtual dressing room setting where users can view a virtual representation of themselves in the chosen garment of their choosing. The suggested approach focusses on offering a shopping experience that is both immersive and participatory by allowing consumers to try on the apparel items that are recommended following segmentation.

An open-source computer vision package called OpenCV is used to implement the virtual try-on capability. From producing the segmented clothing image to superimposing it on the user's body image, there are several important steps in this process. A thorough explanation of the procedure, along with the mathematical formulas and transformations used, is provided below.

The segmented clothing image obtained from the U2Net model serves as the input for the virtual try-on. This image is isolated from its background and contains only the clothing item. To ensure the clothing item fits naturally onto the user's body, alignment is necessary. This involves scaling and rotating the segmented image to match the user's body proportions and pose.

a) Scaling: The segmented clothing image is scaled to fit the target region on the user's body. The scaling factor is calculated based on the dimensions of the target region (e.g., chest, torso). Mathematically, the scaling transformation can be represented as Eq. 8:

$$S = \begin{bmatrix} s_x & 0 \\ 0 & s_y \end{bmatrix} \tag{8}$$

where s_x and s_y are the scaling factors along the x and y axes, respectively. The scaled image coordinates (x', y') are obtained from the original coordinates (x,y) using Eq.9:

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = S \begin{bmatrix} x \\ y \end{bmatrix} \tag{9}$$

b) Rotation: A rotation transformation is used to match the user's posture with the segmented clothing image. The orientation of the user's body determines the rotation angle θ. The rotation transformation matrix is given by Eq. 10*:*

$$R = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \tag{10}$$

The rotated image coordinates (x',y') are obtained from the original coordinates (x,y) using Eq. 11:

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = R \begin{bmatrix} x \\ y \end{bmatrix} \tag{11}$$

c) Translation: The final step in alignment is translating the segmented clothing image to the correct position on the user's body. This ensures the clothing item is accurately placed. The translation transformation matrix is given by Eq. 12:

$$T = \begin{bmatrix} 1 & 0 & t_x \\ 0 & 1 & t_y \end{bmatrix} \tag{12}$$

where t_x and t_y are the translation distances along the x and y axes, respectively. The translated image coordinates (x',y') are obtained using Eq. 13:

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} x + t_x \\ y + t_y \end{bmatrix} \tag{13}$$

d) Overlaying the Clothing Image: The user's body image overlays with the segmented clothing image after alignment is complete. This entails creating an even and natural appearance by fusing the image of the person with the image of the clothes. OpenCV functions are used for blending, which combines the pixel values of the clothing and user images based on specified weights as shown in Eq. 14:

$$blended_{image(x,y)} = \alpha.\ clothing_{image(x,y)} + \beta.user_{image(x,y)} + \gamma \tag{14}$$

The outcome is an integrated image that makes the user look as though they are dressed in the suggested item. The user can see this image and decide whether or not to buy the clothing by seeing how it looks on them.

IV. Results and Discussion

a) Experimental Setup: The NVIDIA Tesla P100 GPU was used in Kaggle Notebooks to simulate the suggested methods. Kaggle is a well-known online community for practitioners of machine learning and data science. It includes a number of parts, including an integrated development environment (IDE) for building, launching, and sharing Jupyter notebooks; CPU and GPU accelerators according to computing requirements; and tools for generating and archiving datasets. An accelerator for high-performance computing (HPC) that NVIDIA created for deep learning, scientific computing, and other computationally demanding activities is the NVIDIA Tesla P100 GPU. It is a part of the NVIDIA Pascal architecture, which is a major improvement over earlier GPU designs in terms of efficiency and performance. The proposed method is implemented using PyTorch and OpenCV.

b) Evaluation Metrics: Data compression and other processing stages can reduce the quality of images. This decrease in quality is measured by a subjective metric called the Structural Similarity Index (SSIM). Two images from the same image capture are needed for this thorough reference metric: a reference image and a processed image as ahown in Eq. 15.

$$SSIM(x,y) = \frac{(\mu_x\mu_y + c_1)(\sigma_{xy} + C_2)}{(\mu_x{}^2 + \mu_y{}^2 + C_1)(\sigma_x{}^2 + \sigma_y{}^2 + C_2)} \tag{15}$$

Where,
x and y are the reference and processed images respectively.
μ_x *and* μ_y *are the average pixel values of x and y.*
$\sigma_x{}^2$ *and* $\sigma_y{}^2$ *are the variances of x and y.*
σ_{xy} *is the covariance of x and y.*
C_1 *and* C_2 *are constants to stabilize the division with a weak denominator*

c) Effectiveness of recommendation and try-on: The experimental results of the proposed fashion recommendation and virtual try-on system reveal several key insights into the effectiveness and practicality of integrating advanced deep learning techniques with large-scale fashion datasets. Firstly, the use of the ResNet50 model for feature extraction has proven to be highly effective in capturing the complex details and patterns present in fashion images. This capability allows the system to generate recommendations that are not only visually similar but also align with the user's style preferences. This finding underscores the importance of leveraging deep neural networks with a high capacity for feature extraction, particularly in the context of fashion, where visual aesthetics play a critical role.

The value of using straightforward but effective methods for matching similar items is further demonstrated by using the Nearest Neighbour algorithm for recommendation. The proposed method makes sure that the suggested clothes are in tight alignment with the input image by concentrating on the Euclidean distance between feature vectors. Though this method works well, in order to improve accuracy even more, future studies should look into more complex similarity metrics or hybrid models that incorporate several recommendation algorithms.

The segmentation outcomes derived from the U2Net model emphasise how important accurate object separation is to the virtual try-on procedure. The realism of the virtual try-on experience is improved by U2Net's capacity to produce high-resolution segmentation masks, which allow the system to isolate clothes with the least amount of background interference. This result emphasises how crucial it is to use sophisticated segmentation models in applications where visual accuracy is crucial. It also points out potential areas for development, such as enhancing the segmentation procedure to handle overlapping items or complex backdrops more skilfully.

From a user experience perspective, the integration of the virtual try-on feature represents a significant step forward in bridging the gap between online and in-store shopping. The ability to visualize how recommended items will look when worn provides users with a more immersive and interactive shopping experience. This feature, combined with personalized recommendations, has the potential to increase user satisfaction and engagement with online fashion platforms.

Fig. 2 illustrates the process where a user uploads an input image, which serves as the basis for generating fashion recommendations. The image on the top row represents the original input image, typically showcasing a particular garment the user is interested in. The system then processes this image to generate a set of recommendations, shown in the second row. These recommended images feature garments that match the style, pattern, and overall aesthetic of the uploaded image

The segmentation button, marked "Segment", allows the user to isolate the clothing item from the background and the human model. Once segmentation is complete, the isolated garment images are displayed. This segmentation step is crucial as it ensures that the recommendations are focused solely on the clothing item, free from distractions in the background.

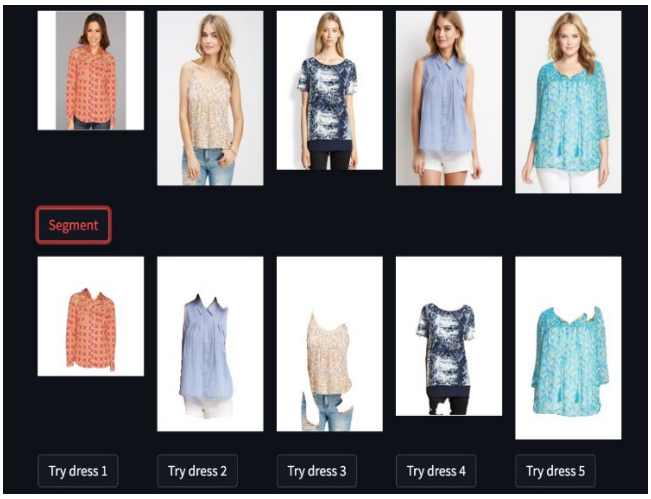

Fig. 3. Recommended images for the input image and resultant images after segmentation

Fig. 3 depicts the recommendations generated for the image uploaded by the user. The nearest neighbor algorithm, along with ResNet50 architecture is involved in recommending similar images.

The segmentation of cloths recommended in Fig. 4(a,b,c). Fig. 4(a) is the input imagewhere as 4(b) shows the generated mask. U2Net model is used to remove the body parts from the input image. The result generated after segmentation is shown in Fig. 4(c).

Fig. 4(a) Input image

Fig. 4(b) Generated mask

Fig. 4 (c) Segmented image

The virtual try-on feature, as indicated by the "Try" buttons, allows users to overlay the segmented garment images onto their own uploaded images. This interactive element enhances the shopping experience by providing a visual representation of how the recommended items might look when worn. This entire process leverages advanced machine learning algorithms and computer vision techniques to deliver highly personalized fashion recommendations, ensuring that users receive suggestions that closely match their style preferences and the patterns of the input image. OpenCv module and Harcasscade algorithm is incorporated for this part. The results had better SSIM value of around 0.735. The high SSIM scores imply that U2Net can result reliable segmentations that are suitable for realistic garment overlay. The user can easily know whether to buy the cloth or not as they have an option to match and check if it is suitable for them.

V. Conclusion And Future Scope

This research presents a novel approach to personalized fashion recommendations by combining image-based feature extraction with nearest neighbor search and virtual try-on capabilities. The system effectively leverages the power of deep learning, specifically ResNet50 and U2Net, to accurately identify relevant fashion items and seamlessly integrate them into virtual try-ons. While the system demonstrates promising results, there is still room for improvement. Future work could focus on enhancing the accuracy of similarity metrics, refining the virtual try-on process, and developing more sophisticated recommendation algorithms. By addressing these areas, the system can provide an even more personalized and immersive online shopping experience. This research study contributes to the advancement of personalized fashion technology, enabling consumers to make informed purchasing decisions and enjoy a more convenient and satisfying online shopping experience.

References

[1] Ding, Y., Lai, Z., Mok, P. Y., & Chua, T. S. (2023). Computational technologies for fashion recommendation: A survey. ACM Computing Surveys, 56(5), 1-45.

[2] He, R., Packer, C., & McAuley, J. (2016, December). Learning compatibility across categories for heterogeneous item recommendation. In 2016 IEEE 16th International Conference on Data Mining (ICDM) (pp. 937-942). IEEE

[3] Hwangbo, H., Kim, Y. S., & Cha, K. J. (2018). Recommendation system development for fashion retail e-commerce. Electronic Commerce Research and Applications, 28, 94-101.

[4] Stefani, M. A., Stefanis, V., & Garofalakis, J. (2019, July). CFRS: a trends-driven collaborative fashion recommendation system. In 2019 10th International Conference on Information, Intelligence, Systems and Applications (IISA) (pp. 1-4). IEEE.

[5] Guillermo, M., Española, J., Billones, R. K., Vicerra, R. R., Bandala, A., Sybingco, E., ... & Fillone, A. (2021, December). Content-based Fashion Recommender System Using Unsupervised Learning. In TENCON 2021-2021 IEEE Region 10 Conference (TENCON) (pp. 29-34). IEEE.

[6] Iso, M., & Shimizu, I. (2021, October). Fashion Recommendation System Reflecting Individual's Preferred Style. In 2021 IEEE 10th Global Conference on Consumer Electronics (GCCE) (pp. 434-435). IEEE.

[7] Shin, Y. G., Yeo, Y. J., Sagong, M. C., Ji, S. W., & Ko, S. J. (2019, September). Deep fashion recommendation system with style feature decomposition. In 2019 IEEE 9th International Conference on Consumer Electronics (ICCE-Berlin) (pp. 301-305). IEEE.

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

[9] Tuinhof, H., Pirker, C., & Haltmeier, M. (2018, September). Image-based fashion product recommendation with deep learning. In International conference on machine learning, optimization, and data science (pp. 472-481). Springer, Cham

[10] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778).

[11] Qin, X., Zhang, Z., Huang, C., Dehghan, M., Zaiane, O. R., & Jagersand, M. (2020). U2-Net: Going deeper with nested U-structure for salient object detection. Pattern recognition, 106, 107404.