Development of a virtual fitting room integrating computer vision, artificial intelligence and virtual reality technologies.

Yanelys Fernández Llerena
Computer Graphic Center
CCG/ZGDV Institute
Guimarães, Portugal
yanelys.llerena@ccg.pt

João Oliveira
Computer Graphic Center
CCG/ZGDV Institute
Guimarães, Portugal
joaooliveira@ccg.pt

Andreia Fernandes Mendes
Computer Graphic Center
CCG/ZGDV Institute
Guimarães, Portugal
andreia.mendes@ccg.pt

Miguel Ângelo Crespo Ferreira
Computer Graphic Center
CCG/ZGDV Institute
Guimarães, Portugal
miguel.ferreira@ccg.pt

José Rubén Pozo Pérez
Computer Graphic Center
CCG/ZGDV Institute
Guimarães, Portugal
jose.perez@ccg.pt

Nuno Sousa
Computer Graphic Center
CCG/ZGDV Institute
Guimarães, Portugal
nuno.sousa@ccg.pt

Yusbel Chávez Castilla
Computer Graphic Center
CCG/ZGDV Institute
Guimarães, Portugal
yusbel.castilla@ccg.pt

Rui Pedro Ribeiro
Computer Graphic Center
CCG/ZGDV Institute
Guimarães, Portugal
rui.pedro.ribeiro@ccg.pt

Inês Caetano
SISTRADE
Software Consulting, S.A.
Porto, Portugal
ines.caetano@sistrade.com

Luís Gonzaga Magalhães
Computer Graphic Center
CCG/ZGDV Institute,
ALGORITMI, University of Minho
Guimarães, Portugal
lmagalhaes@dsi.uminho.pt

Edel Garcia Reyes
Computer Graphic Center
CCG/ZGDV Institute
Guimarães, Portugal
edel.reyes@ccg.pt

Miguel Angel Guevara Lopez
Instituto Politécnico de Setúbal
Setúbal, Portugal
miguel.lopez@estsetubal.ips.pt

2024 International Conference on Graphics and Interaction (ICGI) | 979-8-3315-3599-5/24/$31.00 ©2024 IEEE | DOI: 10.1109/ICGI64003.2024.10923795

***Abstract*—This article presents a virtual fitting room system that improves customer experience by providing personalized image-based recommendations. The prototype designed takes a step forward in the state of the art by contemplating the creation of a customizable avatar based on the user's biometric characteristics and clothing style. Facial recognition models were implemented to predict gender and age, and segmentation and classification techniques are used to extract characteristics from the clothing the user is wearing. The work describes the progress and experiences in the development of some prototype modules and the possible methodologies that are being evaluated to develop a real-time web-based adaptation experience to represent the user's appearance and simulate a combination of multiple garments.**

***Index Terms*—Artificial intelligence, Virtual reality, Computer vision, Deep learning, Fashion recommendation, Machine learning, User experience, Virtual fitting room, Virtual try-on**

I. Introduction

New purchasing trends in the market are causing a greater commitment to intelligent systems. To improve the shopping experience and sales themselves, companies are opting for digital strategies that connect the real world with virtual systems within the interaction facilities used by their customers. These systems provide automated help through recommendation systems adapted to the requirements and information provided by the client, as collected by [1]. With the intensity of work life, people's attention is gradually being displaced by systems that facilitate the search among the large number of products on the market. In the fashion world, the trend is even bigger, as people are attracted to more visually appealing products.

Traditionally, recommender systems evaluate user preferences for specific objects by filtering based on user behavior and object characteristics. These systems analyze user preferences to provide personalized suggestions. A contribution of this work is a personalized clothing suggestion system based on the automatic detection of a person's appearance from a picture of the person. The objective is to identify characteristics of garments and fabrics, as well as extract characteristics from the user's physical appearance. This study employs computer vision (CV) and machine learning (ML) techniques for facial recognition, clothing recommendation using image segmentation, and virtual fitting.

This integration will allow the user to obtain accurate and personalized clothing recommendations, optimizing their shopping experience. Users will be able to view themselves in different styles suggested by the recommendation system and adapted to their physical appearance. Users can also adjust the customized 3D avatar using various sliders corresponding to specific body features. In addition to this, recommendations are made for the fabrics that best adapt to the individual characteristics of the garment. Virtual fitting rooms have become a kind of revolution in the fashion world since choosing clothes can be somewhat exhausting in some circumstances. With

the use of virtual fitting rooms and fashion recommendation systems, this process is greatly simplified. The system has the ability to enhance the overall try-on experience by adding adjustment of body and facial features to correctly represent the user's appearance and integrate real-time fabric simulation along with the ability to match multiple garments.

In the development of this work, different techniques have been applied to accelerate the computational performance of the 3D object visualization algorithms. Cloth simulation is challenging due to the intricate physics of fabric movement, requiring complex mathematical models that account for forces like gravity and friction, demanding significant computational power. Achieving realistic behavior, especially in real-time applications like virtual fittings, requires balancing speed and detail. By leveraging the parallel processing capabilities of graphics processing units (GPUs), each cloth model vertex can be independently simulated, offloading intensive tasks from the CPU and accelerating the process.

On the other hand, different approaches were evaluated to provide better user perception and acceptance of the technology. The introduction of WebGPU addresses the unavailability of compute shaders in WebGL, allowing for more efficient and powerful simulations. This optimization enables more detailed and accurate simulations, providing higher fidelity and more realistic cloth behavior, essential for enhancing the overall try-on experience.

II. Related works

In this section we review current and state of the art literature on clothing recommendation systems and virtual fitting rooms, revealing a wide interest and numerous studies on these topics. Although various techniques have been explored to improve the accuracy and efficiency of these systems, the development of virtual fitting rooms remains a challenge due to the need for more accurate and personalized simulations.

A. CV and ML applied to fashion recommendation

In the literature we found several analyses related to the use of advanced AI and DL techniques to improve user experience in e-commerce platforms [2]. In some cases in the fashion domain, Transfer Learning approaches are used to improve accuracy in product classification and identification, leveraging pre-trained models and adapting them to specific needs.

One example being a study by Bineet Kumar et al. [3], a Transfer Learning system using Visual Geometry Group-19 (VGG-19), InceptionV3 and Convolutional Neural Network (CNN) ResNet-50. This approach employs the cosine similarity function to measure the proximity between detected product features and similar products.

Biresh Kumar et al. [4] use Transfer Learning techniques in addition to Reverse Image Search with the CNN ResNet-50 pre-trained model to extract features from images and measure similarity between feature vectors using cosine metrics. This approach produces five similar product recommendations, demonstrating high accuracy even on images with large differences. In summary, Transfer Learning is a powerful technique to leverage pre-trained models to improve performance and efficiency on new tasks.

Automated systems that seek to improve the user experience through personalized recommendations, using image analysis and user characteristics (such as skin tone, gender, age, etc.) to suggest suitable products, are becoming increasingly common. Challa et al. present FashionNet [5] a fashion recommendation technology designed for online retail platforms. This solution employs a matching network for determining compatibility and a feature network for feature extraction. It uses ML methods such as collaborative filtering and content-based filtering, as well as hybrid strategies. With pretrained CNN ResNet-50 to extract image features and k-nearest neighbors (kNN) to measure similarity between items.

Another case is presented by Rathod et al. [6] where the authors identify the type and color of clothing by analyzing images and recommend the most appropriate clothing for the occasion based on what the user is wearing. Using DL and CV techniques, such as CNN ResNet-50 and Generative Adversarial Networks (GANs), the system offers a virtual fitting in real time, allowing the user to "try on" the clothes in front of a camera.

Ganokratanaa et al. implemented De-Malongsy [7] a mobile application compatible with Android and iOS that employs TensorFlow and TensorFlow Lite for resource-constrained devices, in addition it uses two CNN models for clothing and color classification. The app addresses color identification problems with two specialized models, achieving 78.94% classification accuracy and demonstrating usability. The developers plan to integrate Augmented Reality (AR) and Virtual Reality (VR) technologies in the future.

In recent years, more emphasis has been placed on recommendation systems based on a person's appearance, taking advantage of physical and aesthetic characteristics to recommend the most appropriate styles. Complementing this approach V. Bag et al. [8] propose an innovative system that leverages the user's physical appearance to recommend personalized fashion items. This system identifies the user's skin tone through an input image and uses additional information provided by the user, such as gender, age group, and size. Employing CNN, the system achieved a 92% accuracy rate. Future plans include integrating augmented reality (AR) to enhance the user experience and expanding skin tone identification to include six categories based on the Fitzpatrick system.

In summary, all of these articles share the goal of using advanced AI and DL techniques to improve the accuracy and effectiveness of fashion product classification, recommendation and search, with a strong focus on personalizing and improving the user experience on e-commerce platforms, and in some cases discuss future plans to integrate emerging technologies such as AR and VR to further improve usability and user experience.

This research brings significant value to the field with the integration of a virtual fitting room, which will facilitate the customer experience by allowing them to "try on" clothing virtually, which can reduce the need for returns and increase

customer satisfaction. In this work, different models based on convolutional neural networks (CNN) were evaluated for semantic segmentation and object detection tasks, such as UNET, VGG-16, YOLOv8, among others.

B. Virtual fitting room try-on experience

According to [9] and [10], the degree of engagement and sophistication of a technology plays a crucial role in how a technology is perceived and accepted. In this context, elements such as telepresence, the ability to experience virtual environments as if physically present, experiential value, instrumental value, perceived usefulness and ease of use are essential. Moreover, spatial visual cues and the quality of graphics influence the perception of information and the playfulness of the technology, with 3D graphics being less impactful than 2D graphics

In [11], virtual fitting rooms are categorized into three distinct fitting modes: AR-based fitting with 2D overlays, VR-based fitting, VR-based fitting in immersive environments. AR-based fitting with 2D overlays apply the result of the garment recreation over the obtained silhouette of the user. The overlay may or may not simulate the garment's physics. Simple VR-based fitting simulates the garment's behaviour with the support of a virtual representation of the user. Finally, VR-based fitting in immersive environments recreates the fitting room experience in a completely virtual environment with the support of VR Headsets. The generated environment is similar to 3D representation fitting but includes the user's perspective.

Regarding ease-of-use, AR-based fitting rooms 2D overlay are the most user friendly. By directly using the images of the user obtained by the system, they immediately offer a sense of familiarity between the consumer and the fitting result. However, as mentioned in [9], discrepancies in lightning and poor spatial perception in these types of fitting rooms impact the experience. Both types of VR try-on experiences eliminate these issues but require a good degree of telepresence to obtain the best results.

The experiential value and perceived usefulness are also tied to the simulation quality of a virtual fitting room. As mentioned in [12]–[14], cloth simulations are made through the use of specific processes, namely, the finite element method (FEM), mass-spring system (MSS) or position-based dynamics (PBD). While having different processing weights, all models are computationally heavy. When processing time is crucial to maintain the experiential value, this computational weight becomes a limiting factor in achieving a visually realistic simulation.

III. Methodology

The work described here is split between two main fronts and their corresponding areas of expertise: the Virtual Fitting Room Platform itself, which integrated Mixed Reality techniques, and the Recommendation Module that supports it through the combined usage of CV and ML models and algorithms. On the recommendation side, initial developments were conducted alongside an investigation into the state-of-the-art technologies that supported it. Some architectures and models were explored across three tasks: semantic segmentation, image similarity and human body feature extraction. Semantic segmentation models would locate and identify the clothing article on a provided image, the image similarity models/algorithms would provide a list of items visually similar to said article and the body feature extraction would provide information about the physical appearance of any person featured within the image. These estimations are, in turn, to be fed to the Virtual Fitting Room. With the user providing a photograph of themselves, the detected body features are used to approximate the shape and size of a 3D humanoid avatar to themselves while displaying the virtual re-creations of the recommended clothing items along with the also recommended fabrics/materials. A general overview of the proposed solution can be seen on Fig. 1.

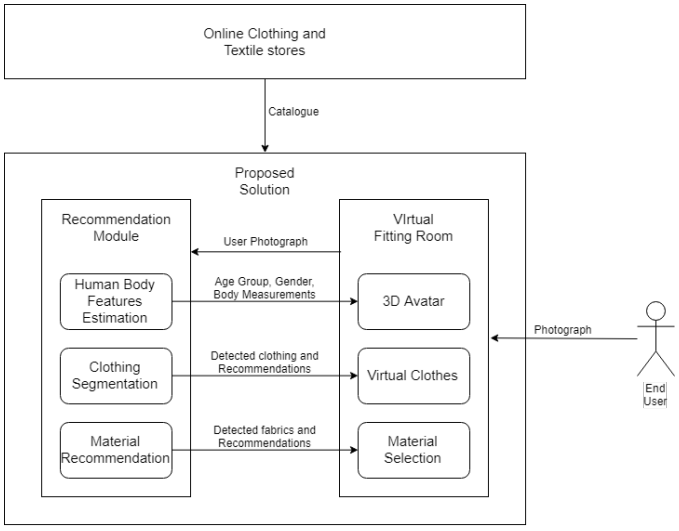

Fig. 1. Overview Diagram of the proposed solution

A. Human Body Feature Extraction

In the case of face recognition, as its name suggests, the area of the image corresponding to the person's face is detected and isolated. This section details the datasets and models analyzed throughout this study for the extraction of the features necessary for the prediction of age group and gender. These predictions are then integrated into the Virtual Fitting Room module, where they are applied together to provide a personalized virtual experience.

1) Datasets: For the estimation of the phenotypical characteristics from a person's face, IMDB-WIKI [15] was used. This is a large set of face images that was created by combining the datasets of IMDB (Internet Movie Database) and certain Wikipedia pages. It contains over 500,000 labeled images of people of different ages and genders, making it a valuable resource for training and testing age group and gender classification models.

2) Models: Several face recognition methods from the OpenCV library were used [16], including Eigenface, Fisherface, Haar-cascade Detection and Local Binary Patterns Histograms (LBPH). These were analyzed for the extraction of the phenotypic characteristics. Haar-cascade Detection [17] was chosen as a starting point, due to its efficiency and low resource consumption, optimized for real-time detection. This method based on the Viola-Joanes framework uses two different face databases (FEI [18] and Yale [19]). These classifiers use Haar-like features that are applied on the image.

Once the face in the image is recognized, this segment will be used as input in the classification methods for both age group and gender. For this purpose, we experimented with Deep EXpectation (DEX) [20], a method using convolutional neural networks (CNNs) with VGG-16 architecture pre-trained with the ImageNet dataset for image classification and then further trained with the IMDB-WIKI dataset. For age prediction, 101 consequent age classes were defined, each representing one year from 0 to 100 years. For gender classification, we applied a similar methodology where, in this case, the model output is simply reduced to 2 classes (male and female).

The model used here, VGG-16 [21], was developed by the Visual Geometry Group (VGG) at Oxford University for image recognition tasks, it is known for its simplicity and ability to extract detailed features, despite its large number of parameters. Its architecture is characterized by 16 trainable weight layers: 13 convolutional layers and 3 fully connected layers. The convolutional layers use 3x3 filters and are grouped into blocks. VGG-16 was pre-trained with the ImageNet dataset with 14 million images belonging to 1000 classes, of fixed sizes of 224*224 and RGB channels.

B. Visual Recommendation

This section details the explored and adopted methodologies for the Visual Recommendation features. This module of the project iterated through various fashion-centered datasets and image segmentation/classification models that will be further detailed in the following sections.

1) Datasets: For the purpose of this project, proprietary datasets are to be used out of catalogue databases made available by partner online stores in order to train segmentation models specialized for each store's offering. However, publicly available datasets were used for testing and might also be used to complement any kind of data imbalances. Examples of such are: DeepFashion [22], Fashion Products Images [23] and Fashionpedia [24]. DeepFashion is a large-scale dataset of clothing images collected at the University of Hong Kong (China), with complex annotations, containing more than 800,000 diverse fashion images, annotated with valuable clothing information, tagged with 50 categories, 1,000 descriptive attributes, bounding boxes and clothing landmarks, and containing more than 300,000 image pairs across domains and poses. Fashion Products Images, on the other hand, is a large dataset of high-resolution product images, professionally photographed and classified with various descriptive product attributes, which were manually entered during the cataloguing process. It also contains descriptive texts commenting on product characteristics and identification, which in turn are stored in a map with all other products in a styles.csv file. This file also shows some of the main product categories and their corresponding display names. Fashionpedia was created by Menglin Jia et al. [24] and consists of 48,825 images featuring clothing being worn by one or multiple persons, annotated with high-detail segmentation masks and split into 27 apparel categories, 19 apparel parts and 294 fine-grained attributes. This dataset is very popular among other state-of-the-art implementations of clothing/fashion detection and was one of the main choices for these implementations.

2) Models & Algorithms: On the clothing recommendation side of the solution, there was first a need to identify the various apparel items present within the image, a semantic segmentation problem. A model based on the U2NET architecture, a salient object detection network first proposed by [25], was one of the first implementations attempted. The idea would be to use the its generated segmentation mask-image in order to "cut out" the various parts of the image pertaining to said clothing articles.

The SOLOv2 instance segmentation framework, an iteration on the SOLO framework that was also conceptualized and developed by [26], has been mentioned in numerous image segmentation implementation surveys and states-of-the-art as the one yielding the most accurate results. It operates with the COCO annotation format, which uses relative coordinates that form a region on the image that is associated with corresponding data. For this case, the regions would pertain to clothing articles and would possess information about clothing type along with other visual characteristics, which the Fashionpedia dataset [24] would be able to provide. Implementation-wise, it was made available within Adelaidet, an open-source toolbox for image recognition developed by [27] derived from Facebook/Meta's Detectron 2 [28].

An implementation was also attempted with FashionFormer, an unified fashion segmentation baseline developed by Shilin Xu et al. [29], that implements Fashionpedia's novel Mask RCNN [24] trained with the aforementioned dataset along with other ResNet models trained on ModaNet [30] and DeepFashion. An example of the out-of-the-box segmentation results can be seen on Fig. 2.

The last implemented model, and that one that was settled upon, was YOLOv8 model that integrates the Ultralytics framework developed by Ultralytics Inc. [31]. It is a proprietary iteration of the popular YOLO family of object detection models and is available, among other variants, as a mask segmentation model. For early testing purposes and given that the purpose of this model was to use a proprietary image dataset build from a partner store's database, a subset of 1084 entries from the Fashionpedia dataset was used, itself also limited to predicting clothing type tags (shirt, skirt, pants, etc). Settings-wise, the training session ran for 180 of 1000 set epochs with a batch size of 8, image size of 640x640 and learning rate of 0.01. A sample of the results can be seen in

Fig. 2. FashionFormer's mask segmentation prediction

Fig. 3.

Fig. 3. YOLOv8's mask segmentation predictions on samples of the validation set.

With this, the clothing recommendation service would consist of a relatively simple textual catalogue filtering with the tags estimated by this model.

In parallel to this exploration of semantic image segmentation, an image similarity estimator was being investigated for a similar effect, on the Textile/Materials side of the solution. The CLIP (Contrastive Language-Image Pre-Training) approach conceptualized by [32] was used to this effect. More precisely, a simple pre-trained VisionTransformer model was used [33] to encode the images and then the Cosine Score function was used to measure the similarity between them.

Regarding the very implementation of this recommendation module, it is expected for it to integrate a Web API hosted by the online store's server and operate within the existing catalogue. This implementation is done as a simple Django-Rest application that loads all the prediction models and logic and make them available across various Python endpoint-methods. Given that encoding the images for the image similarity measuring service is both a resource and time-consuming process, a preliminary bulk encoding functionality was implemented. This makes periodical calls to the main server for a catalogue listing, checking if there were any changes. Should a new item been added, the service proceeds to encode those images and store them along the others on an exported .pkl cross-referenced with a dedicated encoded image database. The functioning of this feature can be seen on Fig. 4.

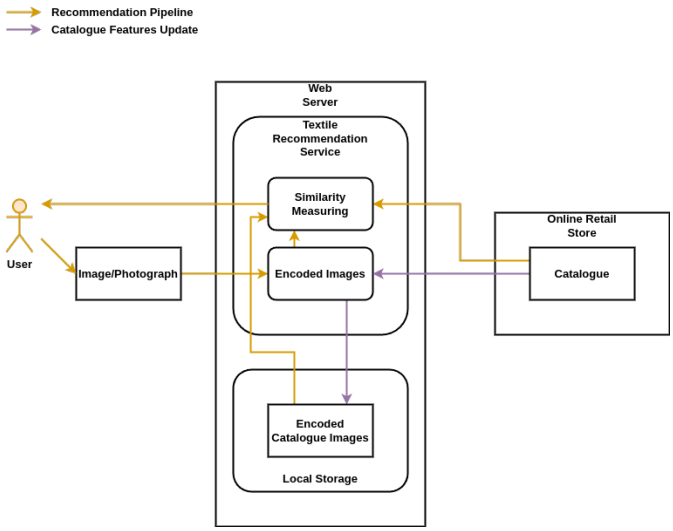

Fig. 4. Demonstrative diagram of how the textile recommendation service is to work. Yellow arrows show the functioning of the main service itself, as the user provides the image that will be compared against the catalogue images and have returned to them an ordered list of catalogue items. Purple arrows show the catalogue

IV. Development of the Virtual Fitting Room

We intend to build a web-based try-on experience with focus on telepresence while allowing the simulation of a combination of multiple cloth pieces.

At the time of writing, the virtual fitting room is still on its initial phase of development. The following subsections will describe considerations, objectives and prospective methodologies for the user representation and the simulation aspects of our try-on experience.

A. User Representation

This paper aims to make use of 3D avatars for a virtual fitting room, since they provide a more realistic and immersive experience compared to 2D avatars. The ability to customize bodily and facial features allows users to create accurate representations of themselves, enhancing personal expression and boosting confidence. This level of detail and personalization is crucial for virtual fitting, as it ensures users can see how clothes will fit and look on something approximated to them. Additionally, 3D avatars facilitate more natural and intuitive interactions, making the virtual fitting process more engaging and effective. Overall, 3D avatars will greatly enhance the accuracy, inclusivity, and user satisfaction of our virtual fitting room.

To achieve this level of avatar customization, we will create a range of pre-defined avatar extremes that users can adjust

through an intuitive slider interface. Users will mix these extremes by adjusting the sliders, allowing for precise personalization of their avatars, which will involve interpolating avatar vertices to ensure smooth and realistic transitions between different features as shown in Fig. 5. This method provides an efficient and user-friendly way to create highly tailored avatars, greatly enhancing the virtual fitting experience. Alternatively, a photo of the user may be used to estimate body morphology with AI, which will be applied to the avatar.

B. Simulation

Cloth simulation quality is highly dependent on the model's resolution. The higher the desired realism, the higher the resolution of the fabric to be simulated, and consequently, the greater computational weight per image.

When it comes to the gaming industry, the simulations can be simplified by fixing the position of certain cloth vertices or by representing the cloth movement through the use of planned animations and rigging. In try-on experiences where the user has free movement, immersive environments for example, this type of practice is not feasible.

Given the considerations, our objective is to apply an unrestricted cloth simulation over the user representation, and, as in [12], [14], use GPU parallel processing capabilities to reduce the simulation effort on the CPU.

To fulfill our objectives, we plan to construct a Position-Based Dynamics (PBD) simulation though the use of GPU Compute Shaders in a web environment using WebGPU within Unity engine.The PBD simulation will use the user representation as collision constraint while trying to maintain the cloth as unrestricted as possible. To allow the simulation of various textiles, the process should allow configurable properties such as grammage, thickness, stretchiness and friction. In scenarios with multiple simulated clothes, we plan to use the previous calculated cloth positions as constraints for the new calculated positions.

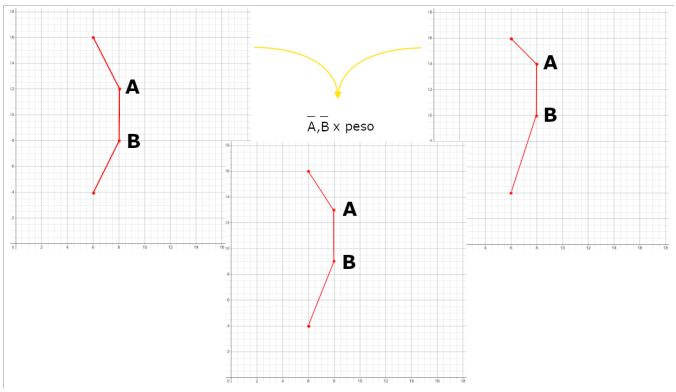

Fig. 5. The two graphs above represent a selection of points from two models. The graph below represents the result obtained from the media of the points from the other graphs (A with A, B with B) according to a weight factor, which is obtained from a slider the user can customize.

V. Results and Discussion

At the time of writing, the various modules of the Virtual Try-On are functional, though not necessarily interconnected. The CV-based services that integrate information into the Virtual Fitting Room have been preliminary evaluated using public datasets to train and test the models. Preliminary pushes on the recommendation side of the Virtual Fitting Room saw various changes in approach. The initially chosen U2NET model, while a relatively popular choice for image segmentation, had no known implementation or applicability for the task of categorizing various types of clothing. Adelaidet's implementation of SOLOv2 was then attempted, but due to various problems derived from the Detectron2 base, the authors were unable to successfully train with the desired Fashionpedia dataset, leading to the abandonment of this framework entirely. Facebook/Meta's own CutLER [34] [35], an image and video segmentation solution, also shown some promise but hardware limitations prevented the training on the single enthusiast-level GPU available. Fashionformer was the first working solution and presented an out-of-the-box and working clothing article semantic segmentation solution, but its usage was held for incompatibilities with the desired API platform. The main, used implementation ended up being Ultralytic's YOLOv8 semantic segmentation model. The results of the training are displayed in Fig. 6 to Fig. 9.

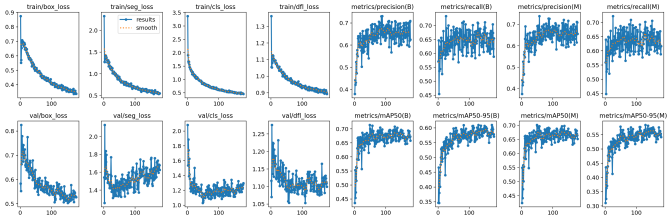

Fig. 6. Overview of the training and validation results of the segmentation model.

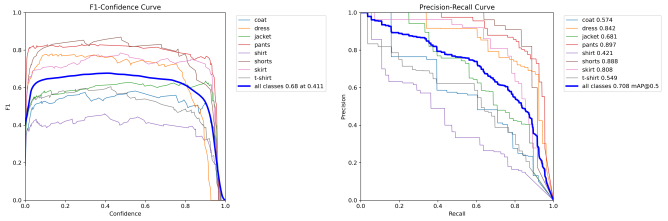

Fig. 7. F1 and Precision-Recall curves of the bounding box results.

Per the aforementioned dataset and training settings, the model achieved an mAP@0.5 of 0.697 for both bounding box and segmentation mask validation results. F1-scores-wise, the bounding boxes achieved an average of 0.68 while the masks achieved 0.67. From this it is possible to conclude that the model, although doesn't seem to achieve ideal values, it nonetheless presents acceptable results when detecting clothing on images with either bounding boxes or segmentation masks.

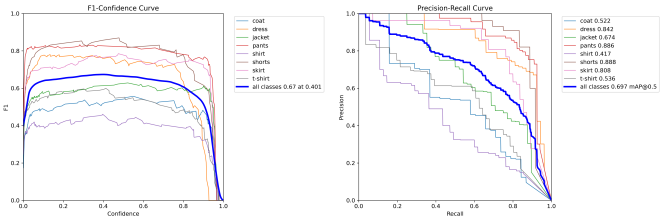

Fig. 8. F1 and Precision-Recall curves of the segmentation mask results.

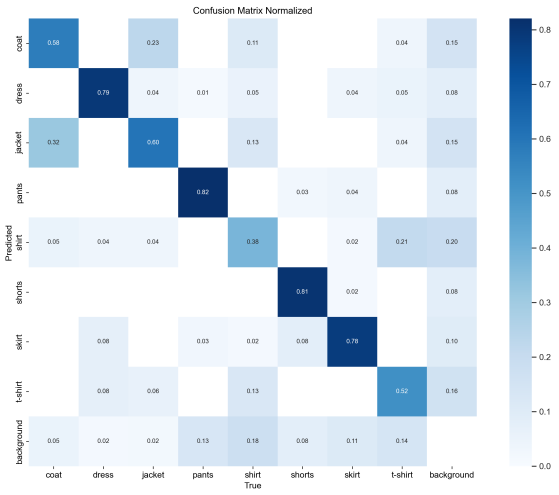

Fig. 9. Normalized Confusion Matrix of the detected classes.

Regarding the detected classes, it is possible to note some dispersion in terms of predictions. The model had the most trouble identifying clothing items of the ”shirt” type, with 38% accuracy while the ”pants” was the class that it most successfully identified with 82% accuracy. This can possibly be attributed to a lack of balance on the dataset or a possible over-abundance of visual variety of ”shirt” items.

In parallel to the clothing side of the CV and ML implementations, the texture recommendation was taking a different and simpler approach. Without an available labelled dataset to train an image classification module, the only solution was to rely on more basic image similarity comparison techniques. Some more impromptu testing of this approach has revealed somewhat accurate choices for similarities, as seen in Fig. 10.

Aditionally, as aforementioned, some preemptive work was done regarding the target usage of this recommendation service as an API embedded on the store’s server. Initial testing on the model revealed that loading the images, encoding them and calculating the similarity between them and the entirety of the test set (about 10 images) took almost an entire minute. This is obviously an issue since a real-world application deals with dozens of thousands of images that the input image would be compared to and not only that, this service would, expectedly, be used by various users at the same time. To try and counter this, the also aforementioned encoding export/import functionality was implemented where the entire catalogue would have its images preemptively encoded and

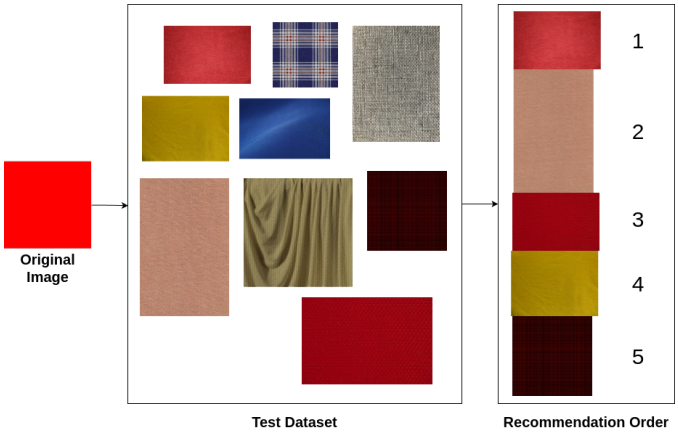

Fig. 10. Test results of the recommendation solution on assorted textile sample images

stored on the filesystem of the server. With this, the input image for this service would naturally still need to be encoded in the moment But for the ones in the catalogue, it would only be a matter of loading the file with the encodings into memory instead of calculating them from scratch. This, at least according to preemptive testing on the smaller test set, considerably reduced the processing time to fractions of a second as only the comparison algorithm needed to run. Additional optimizations to this functionality could pass as a slight reduction on image resolution before the encoding process, potentially saving storage space and processing power requirements.

ACKNOWLEDGMENT

This work has been supported by the European Union under the NextGenerationEU, through a grant of the Portuguese Republic’s Recovery and Resilience Plan (PRR) Partnership Agreement, within the scope of the project TEXP@CT - Innovation Pact for the Digitization of the Textile and Clothing Industries (Project ref. nr. 61 - C644915249-00000025)

VI. CONCLUSIONS AND FUTURE WORKS

This paper describes the first steps in the development of a virtual fitting room for personalized fashion recommendations, implementing state-of-the-art DL methods. Additionally, a textile/material recommendation functionality is also implemented. The virtual fitting room proposal uses a photograph of the user not only gather information regarding the clothes being worn, but also extracts phenotypic characteristics, allowing to provide personalized fashion recommendations that can be displayed on a 3D avatar that is approximated to the user’s likeness. In addition, the system allows visualizing the recommended styles through the virtual fitting room and suggests textiles suitable for the user’s individual characteristics. This solution seeks to improve the fashion shopping experience by combining advanced technologies and personalized approaches. This functionality could also potentially be used for standalone fabric recommendation for textile/material e-commerce. It might be worthy to note that while developing

the clothing recommendation functionality, the image similarity approach was used, out of pure curiosity, with the clothing dataset and also seemed to present some oddly consistent and accurate clothing type recommendations. Further work on a visual-only semantic classifier for clothing articles could yield some potentially valuable results.

Although it displays innovative potential, this work is still very much on early stages of development with recognition and recommendation algorithms, as well as personal data protection, being bound for improvement in the near future. In addition, the integration with emerging technologies such as AR and AI should be further explored in order for this to remain relevant and competitive in the market. With continued advancements, this proposal could transform the fashion shopping experience and set new standards in the digital clothing and fabric industry.

Developing a virtual fitting room project with AI-driven clothing suggestions and customizable 3D avatars can also significantly contribute to the green transition by reducing the environmental impact of the fashion industry. Advanced AI can minimize overproduction and waste by accurately predicting customer preferences, leading to more precise production. Virtual fitting rooms lower the need for physical samples and reduce returns, which cuts down on carbon emissions and packaging waste. This technology fosters sustainable shopping habits by allowing consumers to make informed decisions, thus shifting towards quality over quantity. Overall, adopting such innovations signals a commitment to eco-friendly practices and sets a precedent for other industries, supporting a circular economy and paving the way for a greener future.

REFERENCES

[1] L. Lü, M. Medo, C. H. Yeung, Y.-C. Zhang, Z.-K. Zhang, and T. Zhou, "Recommender systems," *Physics Reports*, vol. 519, pp. 1–49, 10 2012.

[2] A. Kaplan, *Fast Fashion's Fate: Artificial Intelligence, Sustainability, and the Apparel Industry*, pp. 13–30. Springer Nature Switzerland, 2024.

[3] B. K. Jha, S. G. G, and V. K. R, "E-commerce product image classification using transfer learning," *2021 5th International Conference on Computing Methodologies and Communication (ICCMC)*, pp. 904–912, 4 2021.

[4] B. Kumar, A. K. Singh, and P. Banerjee, "A deep learning approach for product recommendation using resnet-50 cnn model," pp. 604–610, IEEE, 6 2023.

[5] N. P. Challa, A. S. Sathwik, J. C. Kiran, K. Lokesh, V. S. D. Ch, and B. Naseeba, "Smart fashion recommendation system using fashionnet," *ICST Transactions on Scalable Information Systems*, 10 2023.

[6] D. Rathod, A. Tundare, P. Shelke, S. Muneshwar, and S. L. Dawkhar, "Fashion recommendation using deep learning," *www.irjmets.com @International Research Journal of Modernization in Engineering*, 2024.

[7] T. Ganokratanaa, S. Chomchaiya, M. Ketcham, N. Sontiyanon, S. Srisawang, and Y. Pornsirianan, "De' malongsy: Apparel style recognition application for fashion enthusiasts with deep learning techniques," pp. 1–4, IEEE, 1 2024.

[8] V. V. Bag, M. B. Patil, V. D. Gaikwad, R. Mohare, S. P. Abhang, and S. Antad, "International journal of intelligent systems and applications in engineering revolutionizing fashion: Fashion era's deep convolutional neural network for outfit recommendations."

[9] R. Batool and J. Mou, "A systematic literature review and analysis of try-on technology: Virtual fitting rooms," *Data and Information Management*, vol. 8, no. 2, p. 100060, 2024.

[10] H. Chen, H. Li, and H. Pirkkalainen, "How extended reality influences e-commerce consumers: A literature review," *Electronic Commerce Research and Applications*, vol. 65, p. 101404, 2024.

[11] Y. Liu, Y. Liu, S. Xu, K. Cheng, S. Masuko, and J. Tanaka *Electronics*, vol. 9, no. 11, 2020.

[12] H. Va, M.-H. Choi, and M. Hong, "Real-time cloth simulation using compute shader in unity3d for ar/vr contents," *Applied Sciences*, vol. 11, no. 17, 2021.

[13] M. H. Hongly Va, Min-Hyung Choi, "Real-time volume preserving constraints for volumetric model on gpu," *Computers, Materials & Continua*, vol. 73, no. 1, pp. 831–848, 2022.

[14] H. Va, M.-H. Choi, and M. Hong, "Efficient simulation of volumetric deformable objects in unity3d: Gpu-accelerated position-based dynamics," *Electronics*, vol. 12, no. 10, 2023.

[15] G. Levi and T. Hassncer, "Age and gender classification using convolutional neural networks," pp. 34–42, IEEE, 6 2015.

[16] G. Bradski, "The opencv library. dr. dobb's journal of software tools," 2000.

[17] R. Padilla, C. F. F. C. Filho, and M. G. F. Costa, "Evaluation of haar cascade classifiers d,"

[18] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. fisherfaces: Recognition using class specific linear projection," 1997.

[19] M. Lal, K. Kumar, R. Hussain, A. Maitlo, S. Ruk, and H. Shaikh, "Study of face recognition techniques: A survey," *International Journal of Advanced Computer Science and Applications*, vol. 9, 06 2018.

[20] R. Rothe, R. Timofte, and L. V. Gool, "Dex: Deep expectation of apparent age from a single image," pp. 252–257, IEEE, 7 2015.

[21] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 9 2014.

[22] Y. Ge, R. Zhang, L. Wu, X. Wang, X. Tang, and P. Luo, "A versatile benchmark for detection, pose estimation, segmentation and re-identification of clothing images," *CVPR*, 2019.

[23] P. Aggarwal, "Kaggle," 2019.

[24] M. Jia, M. Shi, M. Sirotenko, Y. Cui, C. Cardie, B. Hariharan, H. Adam, and S. Belongie, "Fashionpedia: Ontology, segmentation, and an attribute localization dataset," in *European Conference on Computer Vision (ECCV)*, 2020.

[25] X. Qin, Z. Zhang, C. Huang, M. Dehghan, O. Zaiane, and M. Jagersand, "U2-net: Going deeper with nested u-structure for salient object detection," *Pattern Recognition*, vol. 106, p. 107404, 2020.

[26] X. Wang, R. Zhang, T. Kong, L. Li, and C. Shen, "Solov2: Dynamic and fast instance segmentation," *Proc. Advances in Neural Information Processing Systems (NeurIPS)*, 2020.

[27] Z. Tian, H. Chen, X. Wang, Y. Liu, and C. Shen, "AdelaiDet: A toolbox for instance-level recognition tasks." https://git.io/adelaidet, 2019.

[28] Y. Wu, A. Kirillov, F. Massa, W.-Y. Lo, and R. Girshick, "Detectron2." https://github.com/facebookresearch/detectron2, 2019.

[29] S. Xu, X. Li, J. Wang, G. Cheng, Y. Tong, and D. Tao, "Fashionformer: A simple, effective and unified baseline for human fashion segmentation and recognition," *ECCV*, 2022.

[30] S. Zheng, F. Yang, M. H. Kiapour, and R. Piramuthu, "Modanet: A large-scale street fashion dataset with polygon annotations," in *ACM Multimedia*, 2018.

[31] G. Jocher, "Ultralytics." URL: https://docs.ultralytics.com/ Accessed: 2024-06-15.

[32] A. Radford, J. W. Kim, C. Hallacy, A. Ramesh, G. Goh, S. Agarwal, G. Sastry, A. Askell, P. Mishkin, J. Clark, G. Krueger, and I. Sutskever, "Learning transferable visual models from natural language supervision," *Proceedings of Machine Learning Research*, vol. 139, pp. 8748–8763, 2 2021.

[33] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, "An image is worth 16x16 words: Transformers for image recognition at scale," 2021.

[34] X. Wang, R. Girdhar, S. X. Yu, and I. Misra, "Cut and learn for unsupervised object detection and instance segmentation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 3124–3134, 2023.

[35] X. Wang, I. Misra, Z. Zeng, R. Girdhar, and T. Darrell, "Videocutler: Surprisingly simple unsupervised video instance segmentation," *arXiv preprint arXiv:2308.14710*, 2023.