VIETNAM DATATHON 2023

TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN KHTN ĐHQG-HCM

Virtual Sports Clothes Try-On

Lý thuyết Toán Tổ Hợp

Team code: 66

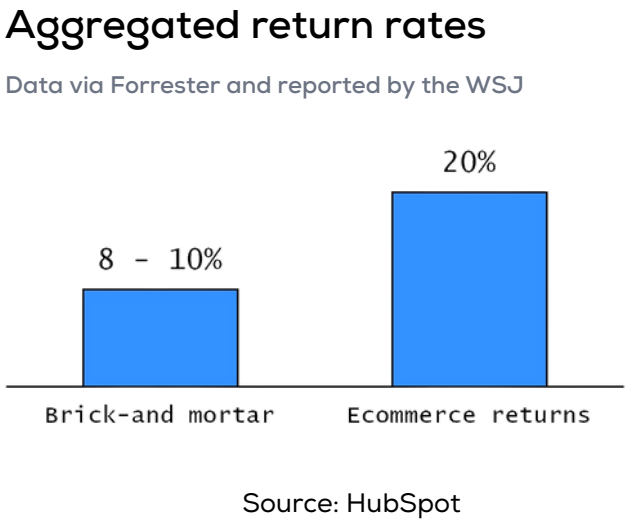
Table of Contents

Introduction

Overview of Problem

Technology has changed the way of shopping among customers. In the busy schedule customers are not ready to spend their time for purchasing goods from shops. They prefer to purchase goods and services from their home. Online shopping allows people to buy things from the comfort of their homes.

But as online sales have increased, so has the number of returns. Since it can be difficult to tell how something looks on you online, many customers have embraced "bracketing" or buying the same product in multiple sizes or colors with the intention of returning whatever they don't like.



Source: HubSpot

That's where virtual try-on technology comes in. Instead of relying on photos of models wearing the clothes, customers can see themselves in the clothes with virtual try-on tools.

Overview of MVP

About our MVP

Our MVP virtual try-on is a web-deployed application capable of processing two-dimensional images. Its primary function involves taking input, which is an image of the user combined with clothing provided by the store available on the app that the user wishes to try on, and producing an output image showing the user trying on that clothing.

Mission

Our mission is to allow customers to visualize themselves in different styles and **sizes** before making a purchase.

Vision

Our vision is to become a user-friendly application, make the virtual try-on experience accessible for everyone, and usher in a new era of immersive, personalized, and convenient retail experiences.

Problem Statement

Define the problem

The challenge that our team aims to solve is the inherent uncertainty and hesitation in online shopping caused by the inability to physically try on or interact with products. This lack of physical interaction often leads to doubts about the fit, style, and overall suitability of the item, resulting in increased return rates, customer dissatisfaction, and a loss of potential sales for businesses.

The approach to address this problem is by enabling customers to imagine themselves in a range of styles before purchasing using AI clothing try-on technology, it eliminates the guesswork in fit and reduces the disappointment of receiving items that don't match expectations. This innovation not only enhances the customer's decision-making process but also benefits retailers by potentially decreasing return rates and contributing to a more sustainable and efficient retail ecosystem.

The Minimum Viable Product which our team proposes focuses on addressing a portion of these issues. Specifically, we'll utilize 2D image-based technology to assist customers in visualizing the colors and styles of clothing when virtually trying them on. This will enable customers to make informed decisions about suitable attire.

Problem Statement

Describe the pain points

- **Fit and Sizing Uncertainty**: Customers often struggle to determine the right size or fit of clothing or shoes without trying them on physically. This uncertainty leads to hesitation in making a purchase, fearing that the item might not fit as expected.

- **High Return Rates:** Due to uncertainties about how a product will look or fit, customers resort to ordering multiple sizes or variations, intending to return those that don't meet their expectations. This results in increased return rates, which are costly for both consumers and businesses.

- **Customer Dissatisfaction and Trust Issues:** When products don't meet expectations based on online representations, it leads to dissatisfaction and undermines trust in the brand or platform, impacting customer loyalty and future purchases.

- **Limited Personalization:** The absence of in-person trials in online shopping can lead to a lack of personalized experiences, leaving customers feeling that their unique preferences or body types are not being catered to effectively.

Problem Statement

Identify Competitors in Vietnam

Competitor #1 — Virtual Try-On by Viettel High Tech

Cons

- Diverse platforms: VTO Mirror, VTO Website, VTO Application
- Recommends the best size for customers
- Create a personalized 3D model for each customer

Pros

- Having limited experience opportunities for customer. Until now, there is only 1 VTO Mirror available at Belluni store in Go Vap District, Ho Chi Minh City, targeting male customers.

Competitor #2 — SmartFashion

Cons

- Enhanced functionality: Try-on and outfit recommendations.
- Simple and user-friendly design.

Pros

- The application is experiencing functionality issues and not operational.
- The application is fee-based.

Solution Overview

A Virtual Try-On solution based on Artificial Intelligence (AI) is designed to enhance the online shopping experience by allowing users to virtually try on clothing, accessories, or other items before purchasing.

In terms of its technical aspect, by utilizing computer vision techniques, the solution is able to accurately detect key body landmarks and measurements, allowing it to simulate the appearance of clothing items on the user's body with remarkable accuracy.

In terms of its fulfillment of business needs, the solution offers robust integration with popular e-commerce platforms. By achieving this, the solution enhances user efficiency by offering personalized recommendations and the ability to virtually try on clothes without the need for users to visit physical stores to try on clothes. Additionally, this approach proves cost-effective for fashion brands as it eliminates the need for brick-and-mortar stores and showrooms, thereby reducing overhead costs like rent and maintenance.

This is a novel approach in the fashion industry, where traditional methods of shopping often require users to spend hours browsing through racks of clothes or waiting in line to try on clothes.

Solution Overview

Given the challenges and expenses associated with 3D model-based approaches, we have decided to adopt 2D image-based Virtual Try-On (VTON) technologies. Our Virtual Try-On solution comprises two key components.

Image Capture or Upload

Users can upload their images or use a live camera to capture their body or face. This step is important for creating a realistic virtual representation of the user.

Clothing Overlay:

The selected clothing items are overlaid onto the segmented regions of the user's image. This process involves adjusting the size, position, and orientation of the clothing to make it appear as if the user is wearing the items.

Solution Overview

In the second component - Clothing Overlay, we use multiple models for easy comparison like CP-VTON+, PF-AFN, and Diffusion (optional),...

CP-VTON+

VITON first used the two-stage architecture:

- Geometric Matching Module (GMM): The try-on clothing is warped to align with the target human.
- Try-On Module (TOM): The warped clothing is blended with the target human image

CP-VTON+ refined VITON for improving the clothing texture transfer, where the clothing area is blended with the warped clothing generated from the original clothing image, not reconstructing through a decoder network.

PF-AFN

PF-AFN uses a novel "teacher-tutor-student" knowledge distillation scheme and further formulates knowledge distillation of the try-on problem as distilling the appearance flows between the person image and the clothes image.

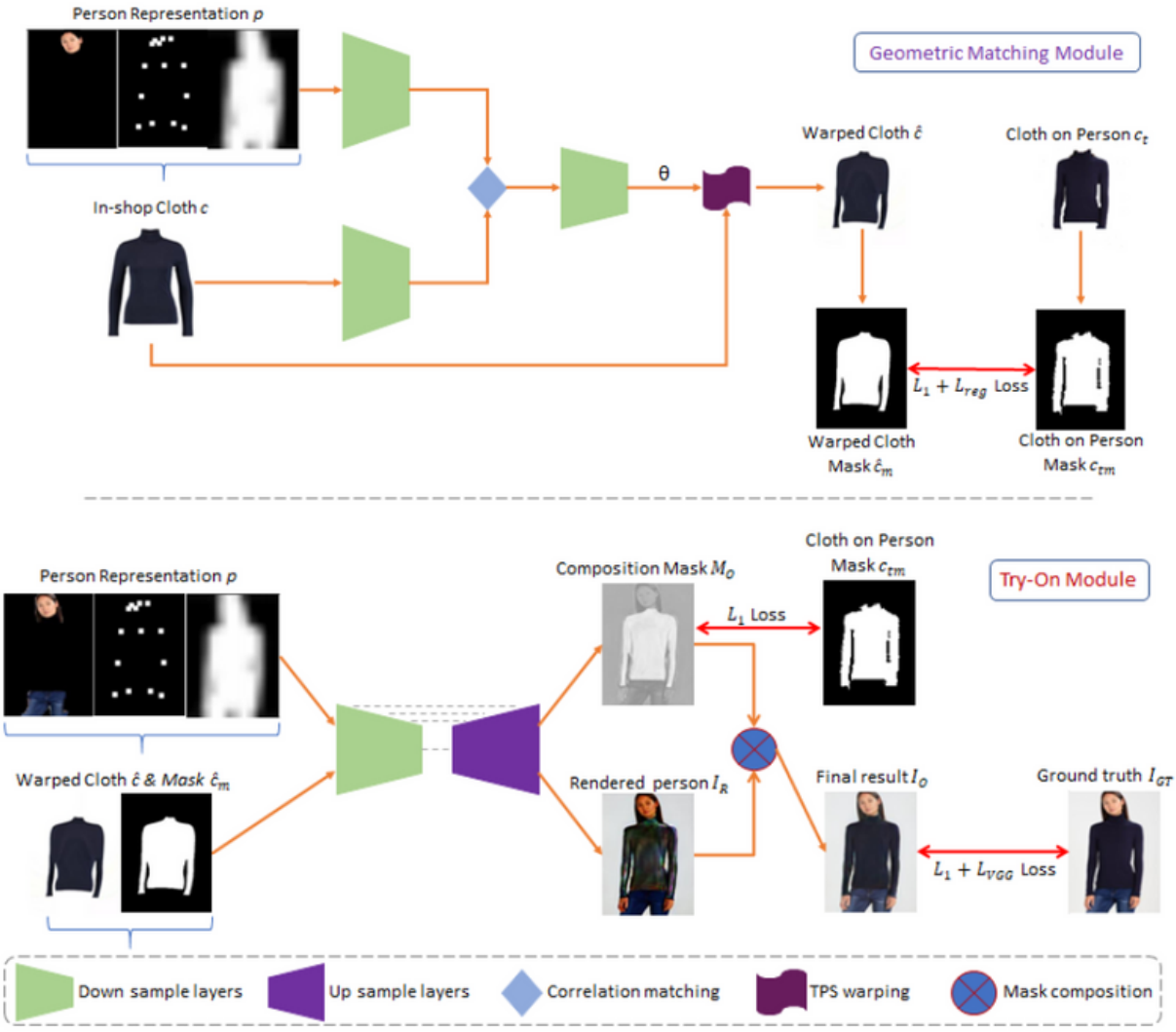
- Appearance flow warping module (AFWM): to establish accurate dense correspondences between the person image and the clothes image
- Generative module (GM): to synthesize the try-on image. Distill the appearance flows to generate a high-quality image.

Methodologies

Our team considers two models to handle the main function - clothes try-on.

CP-VTON+

Clothing Shape and Texture Preserving Image-Based Virtual Try-On



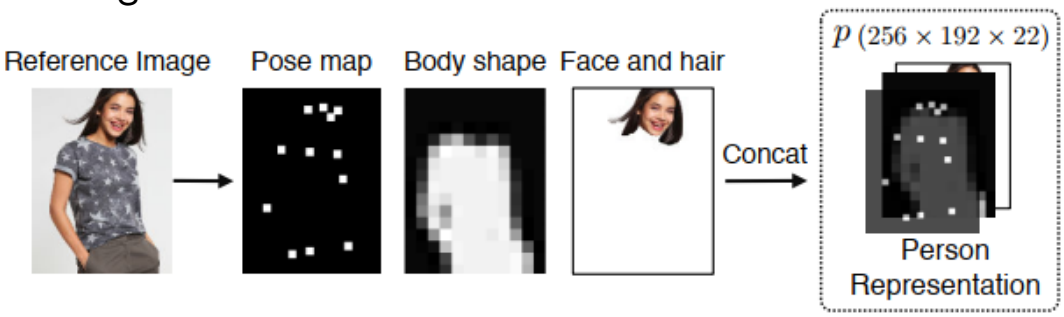
Full pipeline of CP-VTON+ consists of two stages: the try-on clothing is warped to align with the target huma - Geometric Matching Module (GMM), and the warped clothing is blended with the target human image - Try-On Module (TOM).

Methodologies

CP-VTON+

A clothing-agnostic person representation.

Given a reference image I, we extract the pose, body shape and face and hair regions of the person, and use this information as part of input to our generator.

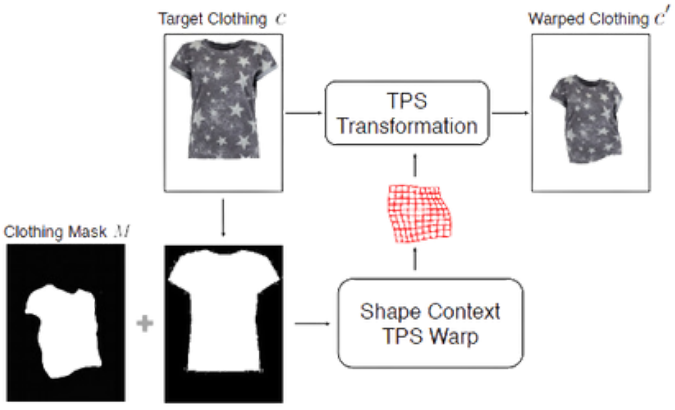


In the VITON dataset, the neck and bare chest area is wrongly labeled as background (like the image above) and the body shape is often distorted by hair occlusion. The model corrects this by adding a new label 'skin' to the label set.

Warping a clothing image

Given the target clothing image and a clothing mask predicted in the first stage, we use shape context matching to estimate the TPS (Thin-Plate Spline) transformation and generate a warped clothing image.

Since the colored texture from try-on clothing does not help in the matching process, our GMM uses a clothing mask MC instead of colored C.

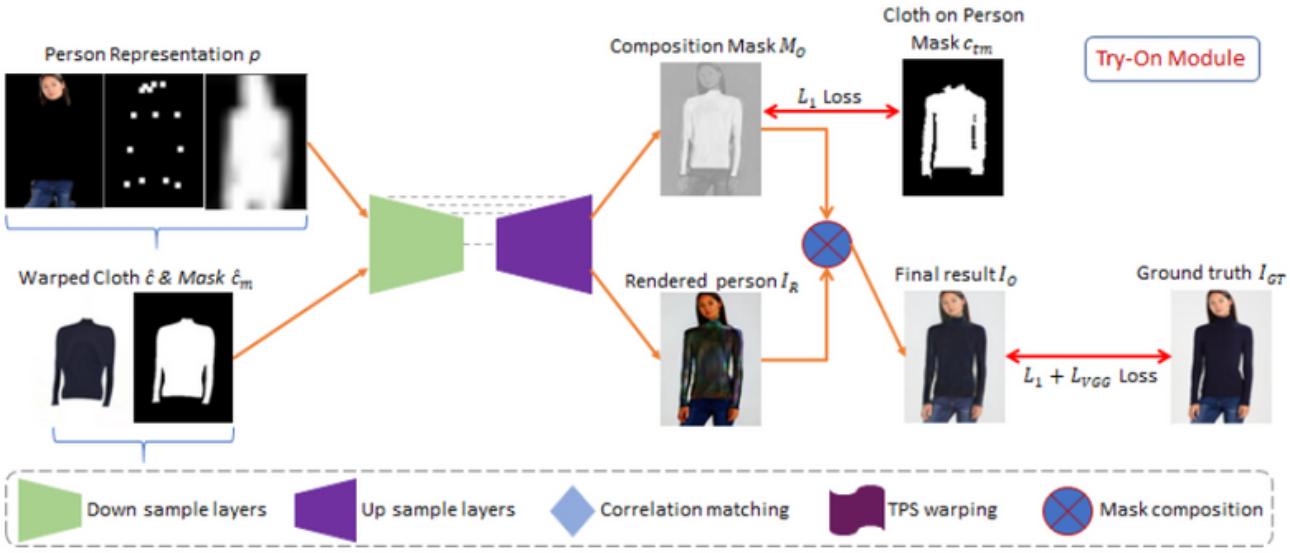


Methodologies

CP-VTON+

Try-on module

The goal of our Try-On module is to fuse wrapped cloth c with the target person and for synthesizing the final try-on result.



Given a concatenated input of person representation p and the warped clothes c, UNet simultaneously renders a person image Ir and predicts a composition mask M. The rendered person Ir and the warped clothes c are then fused together using the composition mask M to synthesize the final try-on result Io:

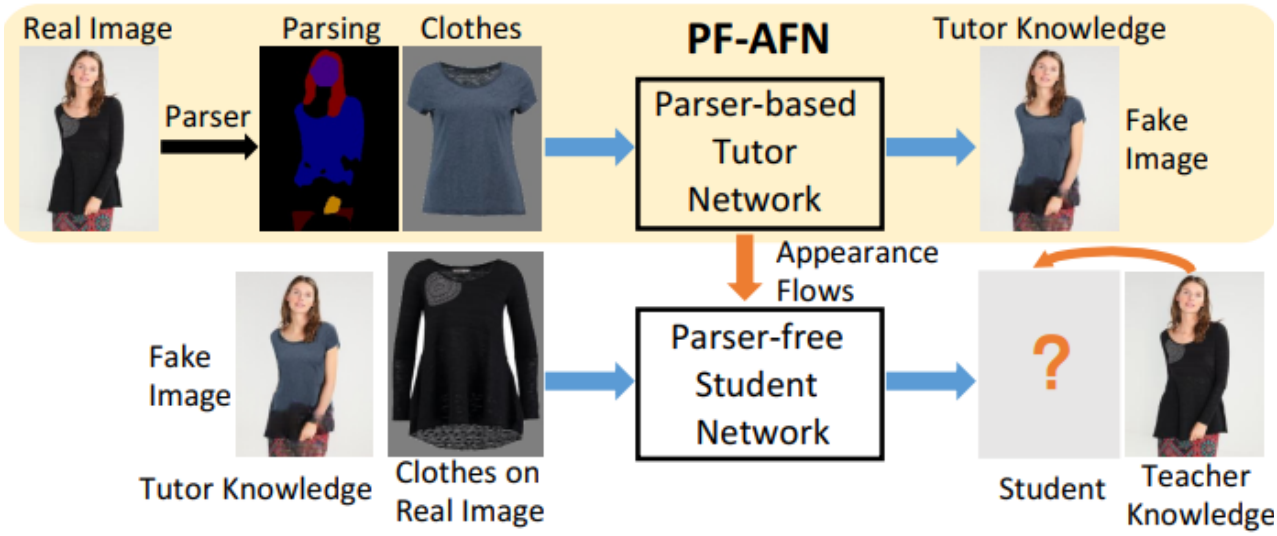
$$I_o = M \odot \hat{c} + (1 - M) \odot I_r$$

Methodologies

The second model considered:

PF-AFN

Parser-Free Virtual Try-on via Distilling Appearance Flows



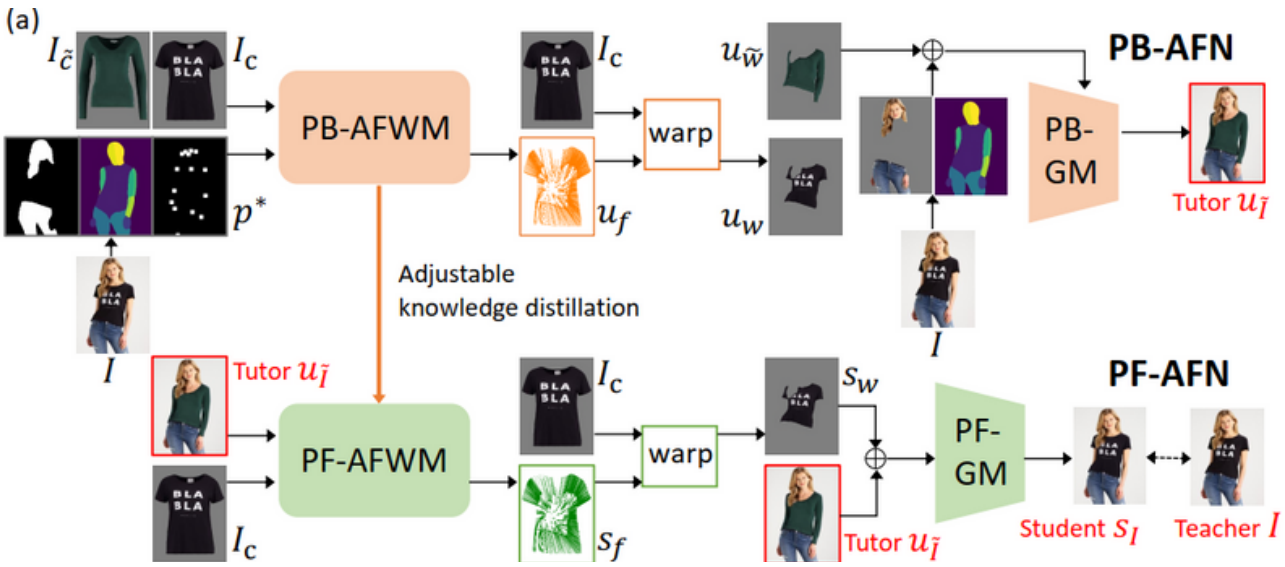
Using "Teacher - Tutor - Student" knowledge distillation scheme, which is able to produce highly photo-realistic images without human parsing:

- Treating the fake images produced by the parser-based method as "tutor knowledge", where the artifacts can be corrected by real "teacher knowledge" extracted from the real person images in a self-supervised way.
- Formulating knowledge distillation in the try-on problem as distilling the appearance flows between the person image and the garment image.

Methodologies

PF-AFN

The training pipeline



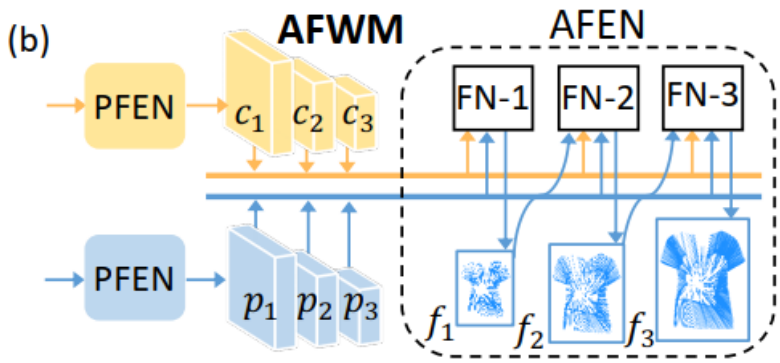
We obtain p∗ from the person image I. Given p∗, the parser-based network PB-AFN randomly selects a different clothes image I(c~) to synthesize the fake image u(I~) as the tutor. We use the tutor u(I~) and the clothes image Ic as inputs to train the parser-free network PF-AFN, where the generated student sI is directly supervised by the real image I.

Methodologies

PF-AFN

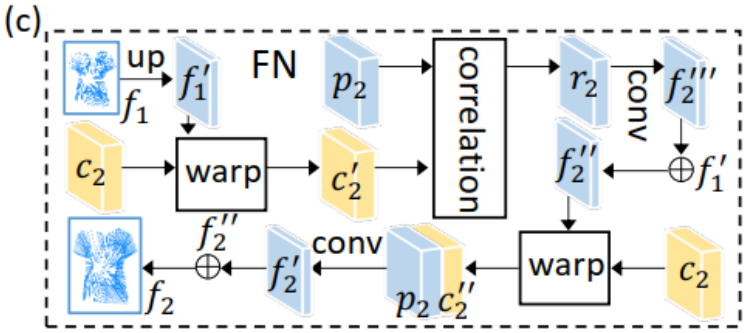
Appearance Flow Warping Module (AFWM)

Both PB-AFN and PF-AFN contain the warping module AFWM, to predict the dense correspondences between the clothes image and the person image for warping the clothes.



The warping module consists of dual pyramid feature extraction network (PFEN) and a progressive appearance flow estimation network (AFEN). PFEN extracts two-branch pyramid deep feature representations from two inputs. Then at each pyramid level, AFEN learns to generate coarse appearance flows, which are refined in the next level.

We carefully design the FN module, which performs pixel-by-pixel matching of features.

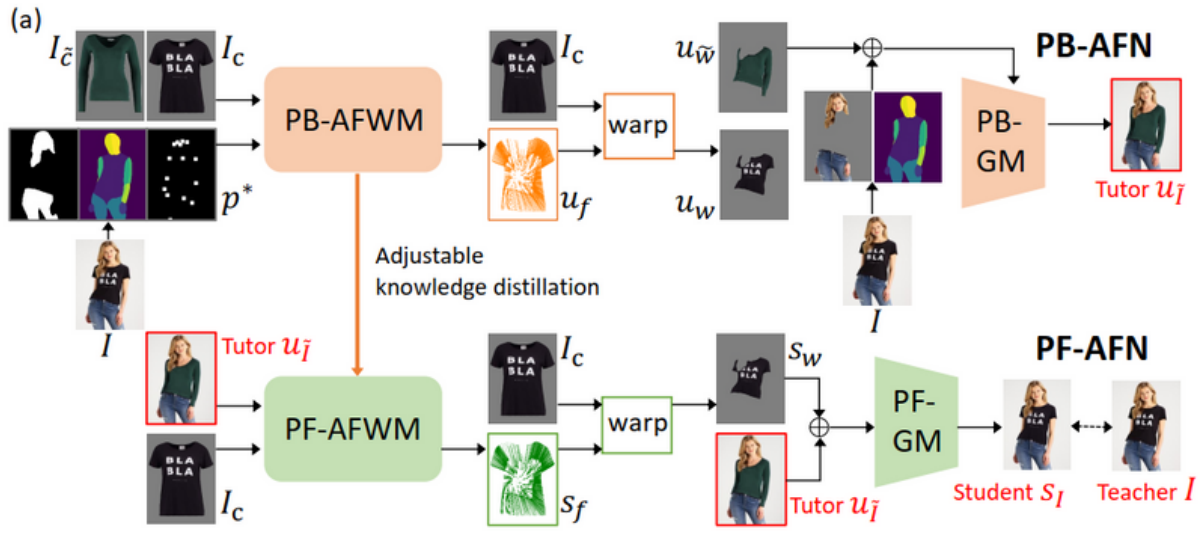


Methodologies

PF-AFN

Generative Module (GM)

Both PB-AFN and PF-AFN contain the generative module to synthesize the try-on image. The parser-based generative module (PB-GM) concatenates the warped clothes, human pose estimation, and the preserved region on the human body as inputs, while the parser-free generative module (PF-GM) concatenates the warped clothes and the tutor image ule as inputs. Both modules adopt the Res-UNet, which is built upon a UNet architecture, in combination with residual connections.

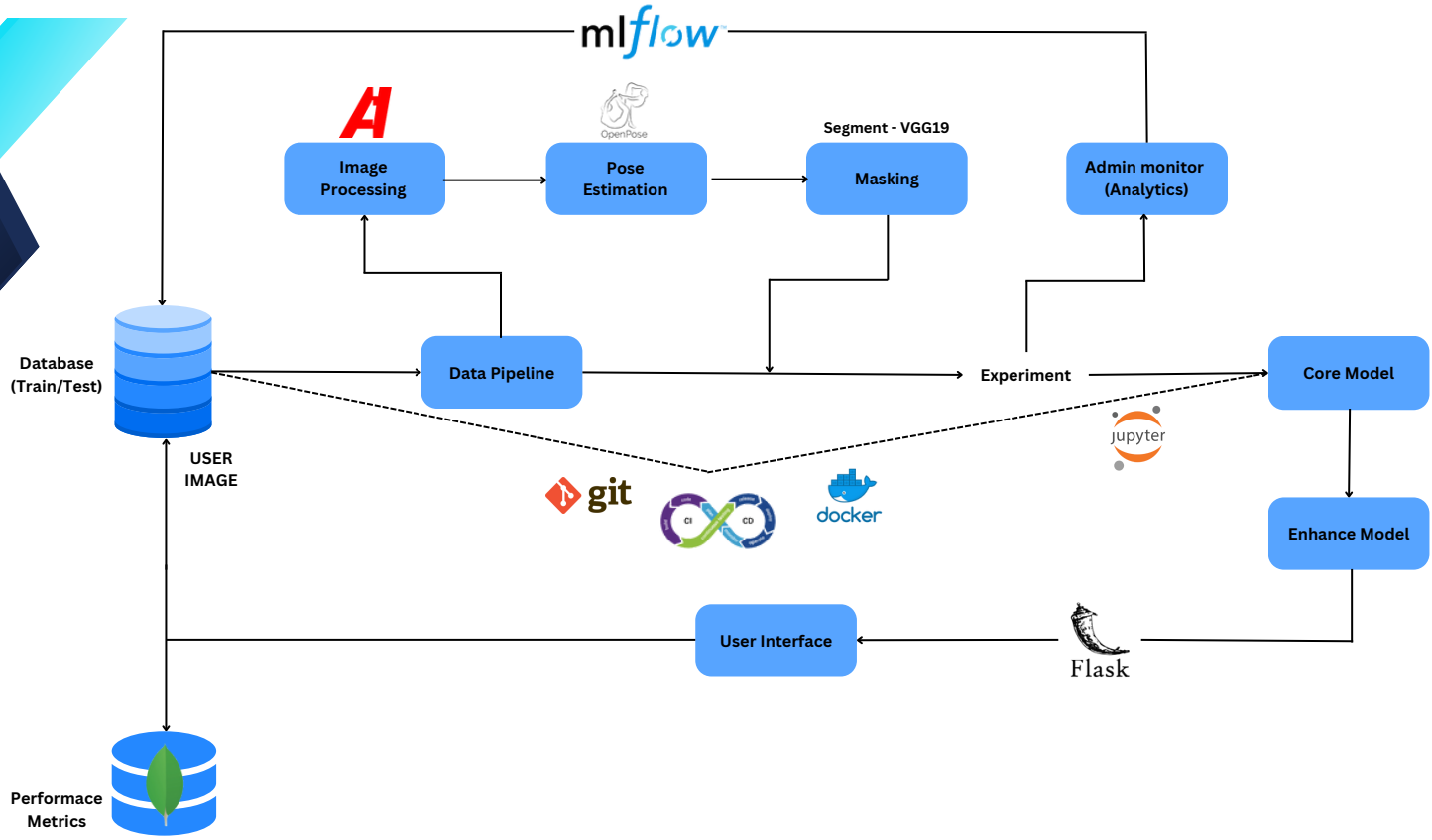


Adjustable Knowledge Distillation

As shown in Fig (a), the inputs of the parserbased tutor network PB-AFN include human parsing results, densepose estimations and pose estimations of the input person. In contrast, the input of the student network PF-AFN is only the fake image and the clothes image. Thus, in most cases, the extracted features from PB-AFN usually capture richer semantic information and the estimated appearance flows are more accurate, thus can be used to guide PF-AFN.

Methodologies

Pipeline & Tech Stack of MVP



We use Dataset 1: Adidas and Nike products and data crawled from websites of Adidas and Nike to fine-tune the pretrained model. Then, output result image of core model (CP-VTON+ or PF-AFN) is used as a raw material for Enhance Model: Stable Diffusion, while promt is not changing clothes of the human , it also forcus on shape the garment of body form. We hope that the final result through model enhancement is better.

Methodologies

First demo of using CP-VTON+

Core Functionality

Product Catalog

The system is designed to provide comprehensive details for each product in the catalog, making it easier for users to find and select the virtual products that best meet their needs.

Each product entry in the catalog includes a range of relevant details such as product name, description, category, color, style, size, and pricing information.

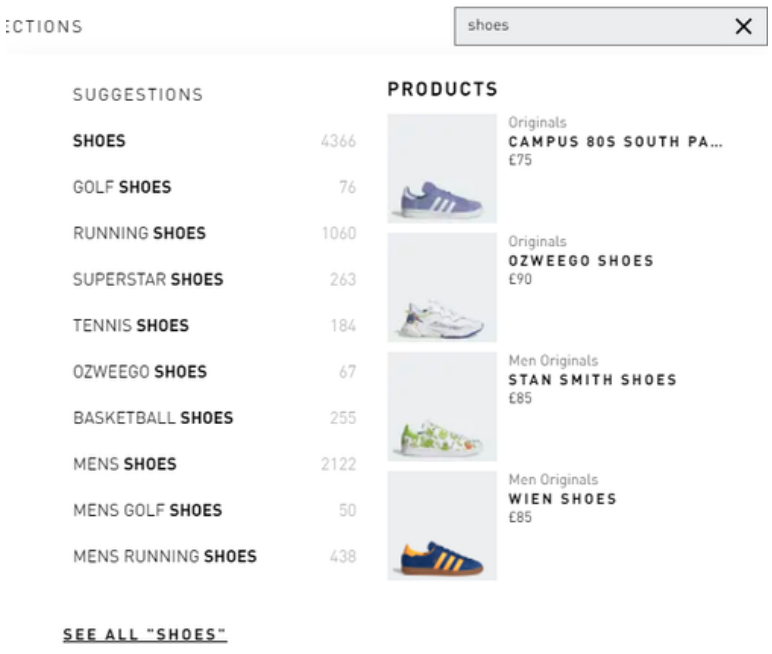
In addition to these details, we also associate product images with each catalog entry, providing users with a visual representation of the virtual products. This helps users to better understand the product and make informed decisions about their purchases.

Core Functionality

Search Engine

The product offers a user-friendly search box that allows users to enter keywords related to the specific product they are seeking.

The system is able to provide users with advanced filtering options to narrow down their search based on various product attributes. These attributes can include product categories, sizes, colors, styles, brands, prices, or any other relevant characteristics. Users can select multiple filters to refine their search results and find products that meet their specific requirements.

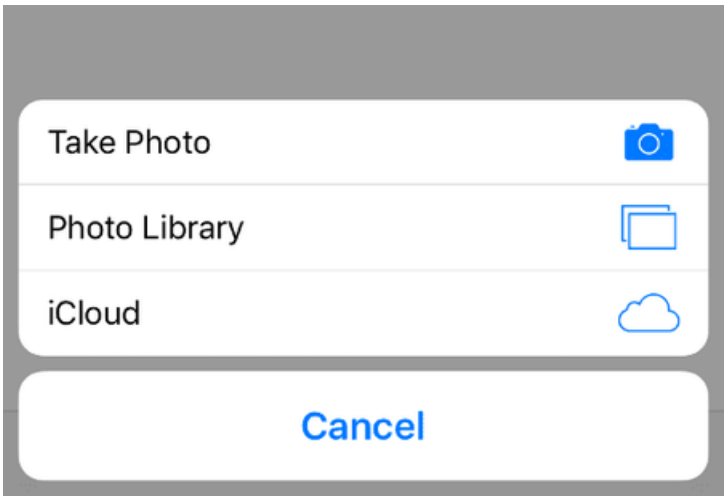


Core Functionality

Image Capture and Upload

The product enable users to capture or upload their own images for try-on. For optimal results during the try-on experience, users have to follow instructions provided on the website.

In addition to providing an accurate virtual try-on experience, the system have to ensure the security and privacy of user data. This includes protecting sensitive information such as user's uploaded images, the result of virtual try-on based on user's images and personal preferences.



Core Functionality

Virtual Try-On Visualization

The system have the ability to seamlessly overlay virtual products onto the user's uploaded image, creating a realistic representation of how the product would look on the user.

The system offers users the ability to preview multiple virtual products simultaneously or in sequence, allowing them to compare different products or variations side by side. This feature helps users make informed decisions and find the perfect product that meets their needs and preferences.

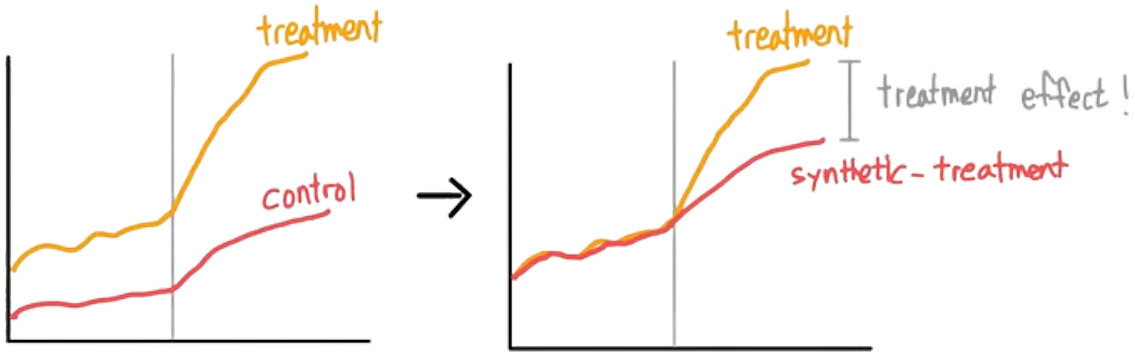
To ensure that the virtual products align perfectly with the user's image, the system performs error handling mechanisms that can handle scenarios where the alignment may not be perfect due to variations in pose, lighting, or image quality.

Performance Metrics

Profit contribution

Storing data on revenue, return rates, access and usage of MVP, number of user accounts, and frequency.

Using the Synthetic Control Method to analyze the effectiveness of using MVP and the impact of improvements or added features.



Customer satisfaction metrics

Using Net Promoter Score (NPS), Customer Satisfaction Score (CSAT) and Customer churn rate (CCR).

AI proxy metrics

Loss function of each stage and overall the model.
Setting baselines and benchmarks for the metrics to understand how the product is improving or declining over time and how it compares to the market or industry standards.

Return on Investment (ROI)

Timeline and Roadmap

Estimated Timeline

Week 1

06 - 12 November 2023

Discovery, Planning & Data Collection

Week 2

12 - 19 November 2023

Model Development & Proposing Idea

Week 3-4

20 November - 3 December 2023

Model Development & Testing

Week 5-6

4 - 17 December 2023

Deployment & Enhancement

Week 7-10

18 December 2023 - 14 January 2024

Product Maintaince, Enhancement & Performance Analysis

Timeline and Roadmap

Main stages

We have two main stages with several tasks:

1. **Development**

- Enhance model Try-on (fine-tune pretrained model)
- Extract Items Informations
- Search engine
- UI design
- Database Administration

2. **Deployment**

- CI/CD pipeline
- Dockerizing model
- Deploying AWS

We will implement the development stage first, with tasks that can be carried out in parallel. And after having the first acceptable model, our team proceeds to the deployment stage, while still considering improving the model if possible.

Timeline and Roadmap

Major milestones

30 November 2023

Fine-tune and choose core model

3 December 2023

Obtain First Acceptable Try-On Model

10 December 2023

CI/CD pipeline

17 December 2023

First-version of Completed MVP

Limitations

Inference time cost

The complexity of the deep learning model used in our Virtual Try On system has a notable impact on inference time. More complex models demand greater computational resources, leading to longer inference times. Additionally, the choice of hardware infrastructure, such as GPUs or specialized accelerators, also affects inference time.

Input Poses

Incorrect or unnatural poses, including variations in angles, can make it challenging to standardize the input data. As a result, this can lead to inaccurate results in the Virtual Try On system.

Most existing models and datasets for pose estimation and body tracking are primarily focused on adult body proportions and movements. Additionally, children's clothing sizes and styles can vary greatly from adults, making it difficult to adapt the virtual try-on feature to accommodate these differences. As a result, these models may struggle to accurately capture and interpret the poses of children.

Future Enhancements

Introducing diverse poses

By introducing a wide range of poses, including those commonly seen in everyday activities or specific to different product categories, users will have the chance to visualize how virtual products appear and fit in various real-life scenarios. This enhancement allows for a more comprehensive and realistic virtual try-on experience.

Recommender system

By integrating a recommender system, we can offer personalized suggestions to users. This system would analyze user data, including browsing history, previous purchases, and feedback, to gain insights into their preferences and style. Utilizing this information, the recommender system would provide recommendations for virtual products that are likely to match their unique tastes and requirements

Conclusion

Our team's MVP encompasses all necessary functionalities to meet customers' demands for searching and virtual trying on clothes, assisting their purchase decisions, and potentially boosting profits for the store. Given the opportunity for further development on a larger scale, refining, and enhancing features, the product could positively impact e-commerce specifically and the retail industry in Vietnam generally.

The Minimum Viable Product that our team proposes leverages state-of-the-art pre-trained models that currently yield promising results and can be further enhanced through the Stable Diffusion Model. We expect this enhancement to improve the output, provide a better experience for customers. Additionally, the model undergoes fine-tuning using data from sportswear collections from Adidas and Nike, creating a specialized product for the clothing lines of these two brands.

References

1. **cp-vton-plus**
   Link: https://github.com/minar09/cp-vton-plus

2. **cp-vton**
   Link: https://github.com/sergeywong/cp-vton

3. **Flow-Style-VTON**
   Link: https://github.com/senhe/flow-style-vton

4. **PF-AFN**
   Link: https://github.com/geyuying/PF-AFN

5. **How to Measure AI Product Performance the Right Way**
   Link: https://medium.com/swlh/how-to-measure-ai-product-performance-the-right-way-2d6791c5f5c3?fbclid=IwAR2fT5dDGEUOnlj64a-IrO5SQrUS9pTdIp6HZQFcgruMDcTA_d9X2SgKBKk

6. **VEESUAL**
   Link: https://www.veesual.ai

7. **Virtual try-on tech is tackling a $280 billion problem in fashion retail and offering customers a different way to shop**
   Link: https://www.businessinsider.com/artificial-intelligence-trends-retail-virtual-try-on-2023-10