VIETNAM NATIONAL UNIVERSITY HCM
UNIVERSITY OF INFORMATION TECHNOLOGY
COMPUTER SCIENCE FACULTY

UIT
TRƯỜNG ĐẠI HỌC
CÔNG NGHỆ THÔNG TIN

IMAGE-BASED VIRTUAL TRY-ON

INSTRUCTOR	Group's Members				
PhD. Ngo Duc Thanh thanhnd@uit.edu.vn	**Truong Thanh Minh** 21520064@gm.uit.edu.vn	**To Anh Phat** 21520085@gm.uit.edu.vn	**Le Chi Cuong** 21520012@gm.uit.edu.vn	**Doan Nguyen Tran Hoan** 21520239@gm.uit.edu.vn	**Le Chau Anh** 215218215@gm.uit.edu.vn

I. INTRODUCTION

In recent years, with the rapid growth of e-commerce platforms and shopping apps, online shopping has become increasingly popular, surpassing traditional retail. However, a significant challenge arises for those seeking to purchase clothing: the inability to try on garments before buying, as one would in a physical store. To address this, the problem of image-based virtual try-on has emerged, aiming to provide customers with an objective evaluation of clothing suitability when shopping online.

II. PROBLEM DESCRIPTION

Input

2 images

Person image

Solid background; Full-body frontal; Person standing; High-resolution (1024x768); Lighting: can clearly see the body of the person.

Clothing image

Solid background; Full clothing frontal; Lighting: can clearly see the details of the clothing.

Output

An image of a person wearing a given clothing item

***Fig. 1**: A virtual try-on result*

III. APPLYING COMPUTATIONAL THINKING

1. Abstraction

The problem involves 3 sub-problems: segmentation map generation (create segmentation map of the person wearing the target cloth), clothing warping (transforming the clothing image based on human's pose) and generate the realistic image by using output of 2 previous sub-problems. Segmentation map generation is abstracted as image segmentation and synthesis, while image warping abstracts the clothing warping problem.

2. Decomposition

Image-based virtual try-on

Input:
- A person image
- A cloth image

Output: An image of a person wearing a given clothing item

Segmentation map generation

Input:
- A person image
- A cloth image

Output: Generated segmentation map of the person wearing the target cloth

Clothing warping

Input:
- A person image
- A cloth image

Output: Warped clothing image based on person's pose

Generate the photo-realistic image

Input:
- Generated segmentation map of the person wearing the target cloth.
- Warped clothing image based on person's pose

Output: An image of a person wearing a given clothing item

Human parsing

Input: A person image
Output: Segmentation Map (Each pixel is labeled with a specific class)

Pose estimation

Input: A person image
Output: Pose representation (keypoints, edges)

Cloth's mask extraction

Input: A cloth image
Output: Cloth's mask

3. Pattern recognition

- **Human parsing:** assigning semantic labels or pixel-level masks to different regions of the human body, such as the head, torso, arms, legs, and other body parts. We solve this problem by **CIHP_PGN.**
- **Pose estimate:** providing information about the positioning and orientation of the human body. The issue at hand is resolved through the utilization of **Openpose**.
- **Generate the photo-realistic image:** Generating a realistic image refers to the process of creating an image that closely resembles a real photograph. To solve this problem, we use **GANs**.

IV. ALGORITHM AND EXPERIMENT

Data

1024x768 virtual try-on dataset zalando-hd. It contains 13,679 frontal-view woman and top clothing image pairs. We split the dataset into a training and a test set with 11,647 and 2,032 pairs, respectively.

Method

The method we use is **HR-VTON**. It's performs wrapping and segmentation map generation simultaneously, and inherently misalignment-free, and can handle the occlusion of clothes by body parts naturally.

Metrics

The metrics used in this study:
- **SSIM (Structural similarity index measure):** used to evaluate the structural similarity between 2 images: the original image and the try-on image.
- **FID (Fréchet Inception Distance):** measures the similarity between the distribution of generated images from the model and the distribution of real images.

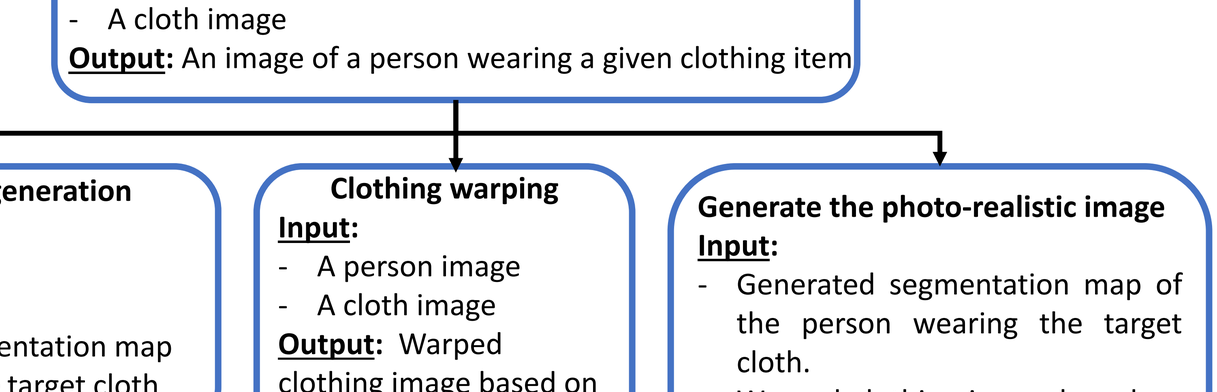

***Fig. 2**: Overview framework*

SSIM	FID
0.892	10.91

***Table 1**: Experimental results*

Conclusion

It completely eliminates the misaligned region and solves the pixel-squeezing problem by handing the occlusion by body parts. So it's practically helpful for real-world virtual try-on applications.