NGHIÊN CỨU PHƯƠNG PHÁP THỬ QUẦN ÁO ẢO BẰNG PHƯƠNG PHÁP HỌC SÂU

VIRTUAL TRY-ON CLOTHES USING DEEP LEARNING METHODS

SVTH: Lưu Văn Huy, Dương Sỹ Bình, Mai Hữu Mân, Nguyễn Huy Tưởng

Khoa Công nghệ thông tin, Trường Đại học Bách Khoa, Đại Học Đà Nẵng;
Email: luuvanhuy2012@gmail.com, huytuong010101@gmail.com, binh9adt@gmail.com, maimanhuu@gmail.com

GVHD: Phạm Minh Tuấn

Khoa Công nghệ thông tin, Trường Đại học Bách Khoa, Đại Học Đà Nẵng;
Email: pmtuan@dut.udn.vn

Tóm tắt - Nhiệm vụ của thử ảo dựa trên hình ảnh nhằm mục đích chuyển một quần áo mục tiêu vào vùng tương ứng của một người, thường được giải quyết bằng cách lắp quần áo đó vào phần cơ thể mong muốn và kết hợp quần áo bị cong vênh với người đó. Mặc dù ngày càng có nhiều nghiên cứu được thực hiện, nhưng độ phân giải của hình ảnh tổng hợp vẫn bị giới hạn ở mức thấp. Chúng tôi cho rằng khi độ phân giải tăng lên, các chênh lệch cong vênh ở các khu vực giữa quần áo thử nghiệm và quần áo mong muốn trở nên đáng kể trong hình ảnh tổng hợp sau cùng. Trong mô hình này, chúng tôi kế thừa những kết quả tốt của mô hình đã thành công từ trước VITON và thực hiện một bước chuẩn hoá nhằm cải thiện chất lượng hình tổng hợp sau cùng. Hơn nữa chúng tôi còn thực hiện ứng dụng mô hình của mình xây dựng hệ thống thử đồ ảo, hệ thống của chúng tôi không chỉ với thời gian thử đồ ngắn thể hiện tính ứng dụng thực tế mà còn rất phù hợp trong bối cảnh dịch bệnh Covid -19 đang diễn tạp hay con người ngày càng trở nên bận rộn để chọn một bộ quần áo phù hợp.

Từ khóa - Thử quần áo ảo; học sâu; trí tuệ nhân tạo; thử áo độ phân giải cao; ứng dụng thử đồ ảo;

Abstract - In this report, we study and build a virtual try on model that creates an accurate image of a person wearing clothes. Focus on building models describing the human body shape in 2D and 3D space and how to transform clothes to best fit the human body. This model overcomes the shortcomings of previous studies to improve image quality and is easy to apply in practice. The successful project can be widely applied in fashion business models, making it easy for customers to try on a variety of clothes without going to the store, contributing to the development of the economy in a more modern direction.

Key words - Virtual try on; deep learning; Artifical Inteligent; high quality virtual try on; virtual try on application.

1. Đặt vấn đề

Những năm gần đây chứng kiến nhu cầu mua sắm trực tuyến các mặt hàng thời trang ngày càng cao, nhất là những ảnh hưởng của dịch bệnh Covid-19 khiến cho nhu cầu này tăng tưởng vượt bậc. Doanh số bán hàng may mặc trực tuyến tăng đều trong những năm qua và việc mua sắm thời trang trực tuyến mang lại sự tiện lợi, người tiêu dùng vẫn lo lắng về việc một mặt hàng thời trang cụ thể trong hình ảnh sản phẩm sẽ trông như thế nào khi mặc trực tiếp, không có khả năng mặc thử quần áo là một trở ngại lớn đối với việc mua hàng trực tuyến tuyến. Người tiêu dùng cần một công cụ giúp hỗ trợ họ chọn quần áo phù hợp với kích thước cơ thể mà không cần phải trực tiếp đi đến của hàng để mặc thử. Được thúc đẩy bởi bối cảnh trên, "mặc thử quần áo một cách trực tuyến" là giải pháp ra đời với mong muốn khắc phục những nhược điểm của việc mua đồ trực tuyến. Việc người dùng có thể "mặc thử quần áo một cách trực tuyến" một không chỉ là một trải nghiệm mới lạ mà nó còn nâng cao trải nghiệm mua sắm của họ, thay đổi cách mọi người mua sắm quần áo, thúc đẩy quá trình chuyển đổi số ngày một nhanh hơn.

Thử ảo dựa trên hình ảnh đề cập đến nhiệm vụ tạo hình ảnh là thay đổi mặt hàng quần áo trên người thành một mặt hàng khác, được đưa ra trong một hình ảnh sản phẩm riêng biệt. Mục tiêu của chúng tôi về hình ảnh tổng hợp cuối cùng:

- Tư thế, hình dáng cơ thể của người thử đồ phải được giữ nguyên.
- Sản phẩm quần áo phải được biến dạng một cách tự nhiên theo theo hình dáng cơ thể của người thử.
- Các chi tiết của sản phẩm quần áo phải được giữ nguyên vẹn.
- Hình ảnh tổng hợp phải đầy đủ chi tiết, sắc nét, phù hợp với yêu cầu thực tế.

Những điểm đáng chú ý trong báo cáo của chúng tôi như sau:

- Chúng tôi kế thừa những kết quả thử đồ ảo của các mô hình tiền nhiệm và sử dụng phương pháp chuẩn hoá để giải quyết sự mất mát, sai lệnh các chi tiết quần áo trong quá trình thử nghiệm.
- Mô hình chúng tôi tổng hợp được hình ảnh ở độ phân giải cao hơn so với các mô hình tiền nhiệm mà vẫn giữ được các đặc trưng hoạ tiết trên quần áo.
- Ứng dụng mô hình xây dựng hệ thống thử quần áo ảo phù hợp với bối cảnh thực tế.
- Chúng tôi đã nghiên cứu cho bước phát triển tiếp theo của mô hình với ý tưởng nghiên cứu mô tả hình dáng người thử đồ và tổng hợp

hình ảnh thử đồ trong không gian 3 chiều với hi vọng nâng cao trải nghiệm thử đồ.

Bài báo này có bố cục như sau: Phần 1 trình bày đặt vấn đề. Phần 2 trình bày tổng quan về lý thuyết các công việc liên quan. Phần 3 trình mô hình thử quần áo ảo. Phần 4 trình bày thử nghiệm mô hình thử quần áo ảo. Phần 5 trình bày bàn luận. Phần 6 trình bày hướng phát triển. Phần 7 trình bày ứng dụng Try on.

2. Các công việc liên quan

Đã có rất nhiều bài nghiên cứu về mô hình thử quần áo 2D và 3D. Đối với 3D việc xây dựng một mô hình thử quần áo là khá khó vì phải thu thập dữ liệu bằng thiết bị chuyên dụng và xây dựng một mô hình có khả năng xử lí dữ liệu của không gian 3 chiều cũng rất phức tạp. Tuy vậy vẫn có một số bài nghiên cứu về lĩnh vực này tuy nhiên hiệu quả vẫn chưa thật sự cao như M3D-VITON [4]. Đối với 2D, VITON là bài nghiên cứu tiêu biểu và cũng là cơ bản nhất trong việc thử quần áo ảo. Một số phiên bản cải tiến của VITON phải kể đến như CP-VTON [5] cố gắng cải tiên quá trình biến đổi TPS để đạt kết quả tốt hơn, VTNFP [6] và ACGPN [7] dự đoán thêm segmentation map của người mặc quần áo để nâng cao hiệu quả dự đoán.

3. Mô hình thử quần áo ảo

3.1. Mô hình cơ bản

VITON là mô hình thử quần áo đã có mặt khá lâu và cũng là mô hình đặt nền tảng cho các mô hình thử quần áo ra đời sau này. Trong nhiệm vụ thử quần áo, chúng ta có 3 nhiệm vụ chính: Biểu diễn con người, bẻ cong quần áo và kết hợp con người với quần áo mới.

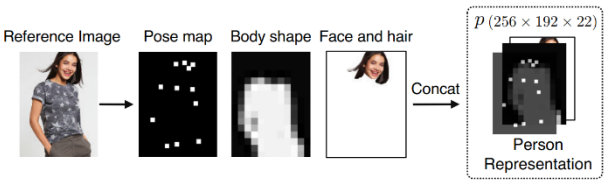
VITON biểu diễn con người bằng 3 thuộc tính chính: Pose map thể hiện tư thể của con người, Body shape thể hiện hình dạng tổng thể và Face, Hair là các chi tiết giữ lại vì nó không ảnh hưởng nhiều tới phần thử đồ ảo nhưng rất khó để sinh ra.

Để bẻ cong quần áo, đầu tiên ta xây dựng một mô hình có khả năng tạo hình ảnh quần áo sau khi mặc và mặt nạ (mask) của nó ứng với thuộc tính người được biểu diễn ở trên. Sau đó dựa vào mặt nạ (mask) của quần áo ban đầu và mặt nạ của quần áo được sinh ta, ước tính TPS transformation (chi tiết ở phần 3.2.3). Dựa vào đó ta có thể bẻ cong được quần áo ban đầu.

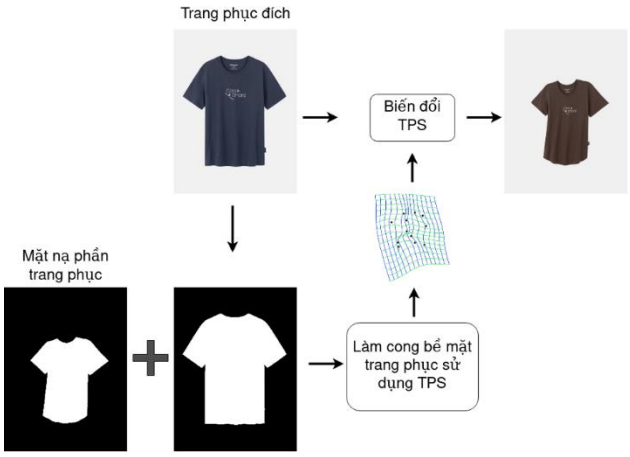
Sau khi bẻ cong quần áo ban đầu, ta có thêm một mạng để sinh ra mặt nạ thành phần (composition mask), mạng này giúp xác định được cần bổ sung bao nhiêu thuộc tính từ quần áo để hoàn thiện trang phục.

VITON mặc dù đạt được kết quả khả quan nhất định, tuy nhiên vẫn sẽ có những khuyết điểm đó là thường không giữ được các chi tiết trên quần áo và không xử lí được với ảnh có độ phân giải cao.

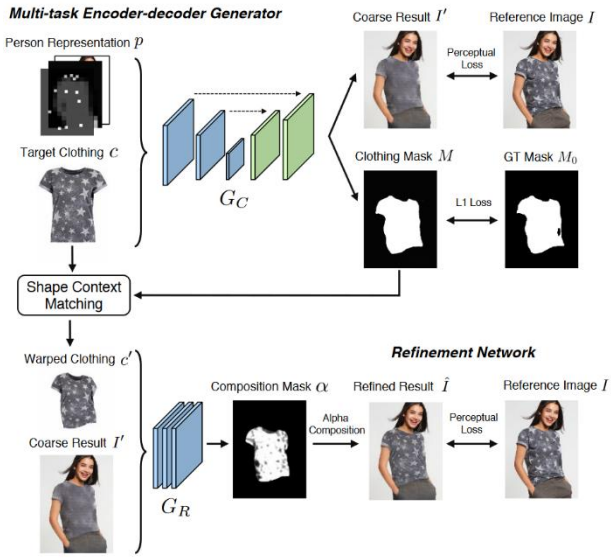
3.2. Mô hình đề xuất



Hình 1: *Biểu diễn mô hình người*



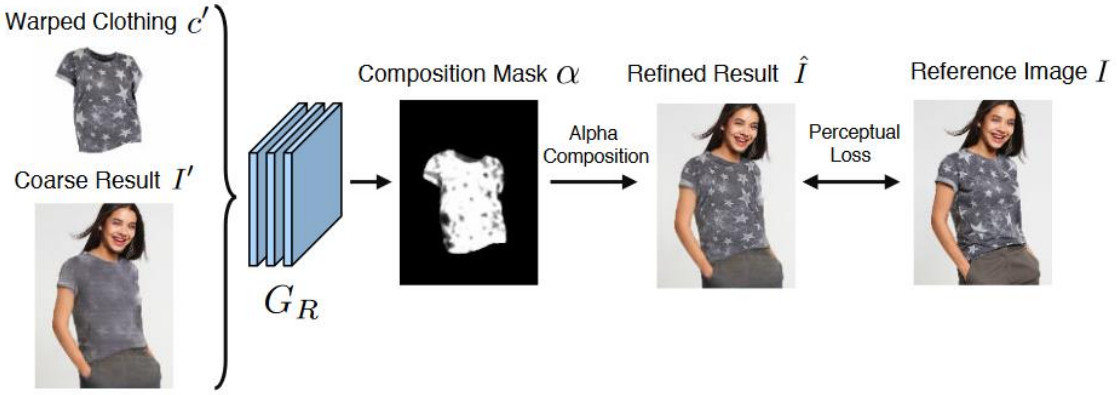
Hình 2: *Bẻ cong trang phục*



Hình 3: *Tổng quan mô hình VITON*

Mô hình cơ bản tuy đã có những kết quả khả quan nhất định, tuy nhiên chỉ xử lí được ảnh ở độ phân giải thấp. Việc tăng độ phân giải lên sẽ làm xuất hiện nhiều vùng bị lệch đi. Đồng thời cách biểu diễn mô hình người ở những mô hình trước vẫn chưa thể hiện được hoàn toàn đặc điểm dáng người và phần cơ thể. Những phương pháp và mô hình dưới đây sẽ khắc phục nhưng nhược điểm trên và tối ưu hóa việc thử quần áo trong ứng dụng thực tế.

3.2.1. Biểu diễn mô hình người



Hình 4: *Mô hình sinh kết quả mô hình VITON*

Hình 5*: Biểu diện hình dáng và tư thế người*

Trong bài toán thử đồ, chúng ta cần biểu diễn mô hình người để có thể thay một trang phục khá c vào, việc này đòi hỏi cần phải biểu diễn sao cho thể hiện được tư thế và các bộ phận cơ thể người, loại bỏ đi những phần quần áo sẽ bị thay thế vì những chi tiết này có thể làm ta nhầm lẫn phần diện tích là cơ thể người và cuối cùng cần giữ lại nhưng chi tiết khó tái tạo như khuôn mặt, tóc, bàn tay. Để làm được điều này chúng ta có thể biểu diễn mô hình người bằng cách sử dụng CIHP [9] để tạo bản đồ phân vùng (segmentation map) cơ thể người và quần áo, từ đó ta có thể tách phần quần áo không cần thiết ra khỏi cơ thể, đồng thời sử dụng Openpose [1] để tạo ta bộ khung tư thế cho cơ thể người.

3.2.2. Tạo phân vùng mới

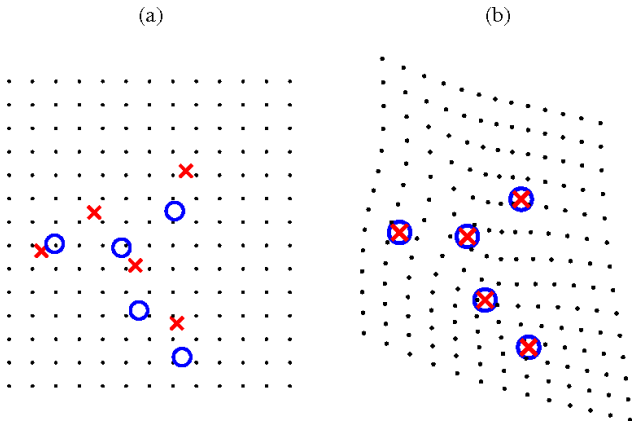
Hình 6: *Quá trình tạo phân vùng mới cho người thử đồ*

Ở bước này, sử dụng mô hình U-Net thuộc kiến trúc encoder-decoder khá nổi tiếng và đạt hiệu suất cao đối với những bài toán phân vùng (segmentation) để xây dựng một mạng sinh mới với đầu vào là bản đồ phân vùng đã loại bỏ đi phần cần thay thế (S_r), Pose Map và trang phục cần thay thế, đầu ra là một phân vùng mới trong trường hợp đã mặc trang phục mới, bước này cho chúng ta hình dung được sau khi thử trang phục thì hình dạng sẽ trông như thế nào và củng tạo nên bộ khung sườn cho việc thay quần áo mới vào mô hình người.

Vì đầu vào đã hoàn toàn xóa bỏ phần quần áo cũ nên ở quá trình huấn luyện, chúng ta chỉ cần hình ảnh người ban đầu và hình ảnh áo người đó đang mặc. Hàm mất mát sẽ là tổng của pixel-wise cross-entropy và mạng đối nghịch phát sinh conditional adversarial giữa S_n và S.

$$L_s = L_{gan} + \lambda . L_{ce}$$

3.2.3. Bẻ cong quần áo

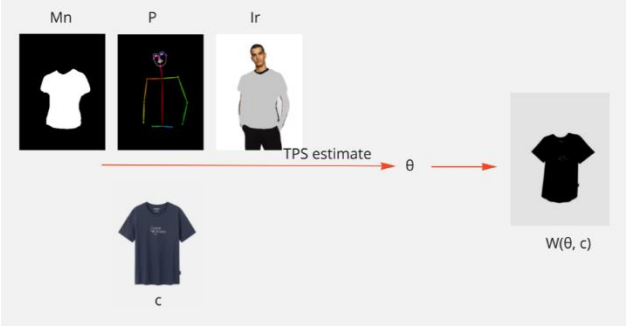


Hình 7: *Ví dụ về biến đổi TPS*

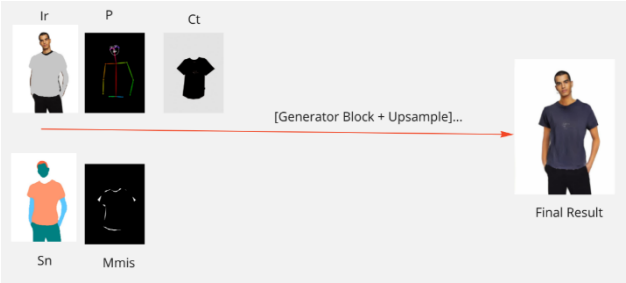
Biến đổi TPS là phương pháp căn chỉnh hình ảnh đã xuất hiện từ lâu. Giả sử chúng ta cần co giãn mặt phẳng để khớp những điểm chấm đỏ tới những điểm chấm xanh, chúng ta cần dự đoán vị trí mới của các chấm đen. Nhưng những điểm này là rời rạc và để co giãn hoàn toàn bức ảnh, vô số điểm còn lại sẽ nội suy ra vị trí của mình.

Biến đổi TPS vẫn là phương pháp chính để bẻ cong quần áo trong hầu hết các bài nghiên cứu. Đầu vào của mô hình là hình ảnh con người đã xóa bỏ đi quần áo cần thay thế (I_r), Pose Map (P) và quần áo sẽ thay vào. Mô hình hồi quy sẽ cho ra bộ trọng số của TPS là θ, từ trọng số θ có thể bẻ cong quần áo để khớp với cơ thể của con

người. Khi huấn luyện chúng ta xem 2 quá trình trên là một mô hình và hàm mất mát được tính ở hình ảnh quần áo đã bẻ cong với bộ trọng số θ.



Hình 8: *Mô hình bẻ cong quần áo*



Hình 9: *Mô hình sinh ra kết quả*

Hàm mất mát là độ lệch giữa quần áo đang được mặc (I_c) so với quần áo đã được bẻ cong thông qua bộ tham số TPS (W(c, θ))

$$L_{warp} = ||I_c - W(c, \theta)||$$

Dường như ở bước này chúng ta đã có thể bẻ cong quần áo và hoàn thành, tuy nhiên sau khi bẻ cong quần áo, chắc chắn sẽ có độ lệch nhất định so mới bản đồ phân vùng (segmentation map) đã được sinh ra ban đầu vì sinh ra một bản đồ phân vùng sẽ dễ dành hơn và chính xác hơn, đồng thời hình ảnh quần áo sau khi biến đổi TPS chắc chắn sẽ không tự nhiên. Vì thế chúng ta cần có thêm một mô hình để khắc phục những sai lệch trên quần áo này.

3.2.4. Mạng sinh - Generator network:

Mạng sinh này ngoài những đầu vào như: quần áo đã được bẻ cong (C), mô hình người đã xóa quần áo cũ (I_r) và bản đồ dáng - Pose Map (P), mô hình còn nhận thêm đầu vào là là một mặt nạ nhị phân (binary mask) Mmis thể hiện những vùng bị lệch giữa mặt nạ quần áo qua biến đổi TPS (M_t) và mặt nạ của quần áo sinh ra ở mô hình sinh phân vùng - segmentation gennerator (M_n). Để tìm được phần bị lệch thì chúng ta cho mặt nạ của quần áo ban đầu biến đổi qua TPS và lấy mặt nạ của phần quần áo đã được sinh ra trên bản đồ phân vùng. Sau đó thực hiện các phép toán trên pixel để lấy phần diện tích bị lệch. Nhờ đó khi kích thước ảnh khá lớn nhưng vẫn giữ được chi tiết của người và quần áo.

Để đơn giản hóa mô hình sinh, phần mã hoá (encoder) được bỏ và mô hình là sự kết hợp giữa các khối dư (residual blocks) và upsampling layers. Quá trình training thực hiện theo SPADE [2] và pix2pixHD [3]

4. Thử nghiệm mô hình thử quần áo ảo

4.1. Bộ dữ liệu

Dữ liệu về con người và quần áo chủ yếu được tìm kiếm và tải về từ nhiều nguồn thông qua công cụ tìm kiếm google.

5. Bàn luận

5.1. Ưu điểm

Mô hình khắc phục được những vùng ảnh bị lệch khi tăng độ phân giải ảnh lên kích thước lớn

Mô hình cho ảnh tự nhiên hơn những mô hình trước.

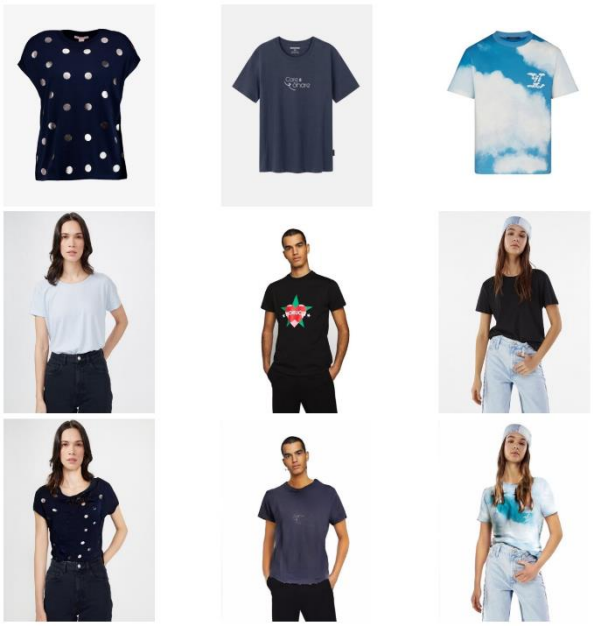
5.2. Nhược điểm

Thử quần áo là một bài toán với input đầu vào rất đa dạng, vì vậy còn nhiều trường hợp mô hình cho ra kết quả sai hoàn toàn. Thời gian thực thi của mô hình còn khá lâu.

5.3. So sánh với mô hình cơ bản

Bảng 1: *So sánh hai mô hình*

Mô hình đề xuất	Mô hình cơ bản
Biểu diễn dáng người qua bộ khung RGB	Biểu diễn dáng người qua mặt nạ
Biểu diễn cơ thể người qua bản đồ phân đoạn	Biểu diễn cơ thể người qua mặt nạ
Mô hình sinh kết quả có chỉnh sửa phần trang phục bị lệch	Mô hình sinh kết quả chỉ làm rõ trang phục



Hình 10: *Kết quả chạy thử trên mô hình. Trên cùng: Áo đầu vào; Ở giữa: Ảnh người đầu vào; Dưới cùng: Kết quả mô hình.*

6. Hướng phát triển

Kết quả thu được từ mô hình đã có thể thử một bộ đồ mới lên một bức ảnh, tuy nhiên trong thực tế khi thử một bộ đồ cần phải xem trên nhiều góc nhìn khác nhau. Để giải quyết vấn đề này, một ý tưởng được đưa ra đó là tái tạo mô hình 3 chiều của người mặc từ ảnh 2 chiều. Đã có nhiều mô hình được đưa ra để giải bài toán tái tạo mô hình 3D cơ thể người, trong đó hướng đi của mô hình PIFU [8] đang cho kết quả triển vọng khi cho phép phục

hồi bề mặt có kết cấu 3D của người mặc quần áo từ một hình ảnh đầu vào duy nhất, có thể số hóa các biến thể phức tạp của quần áo, chẳng hạn như váy ngắn và giày cao gót, bao gồm cả kiểu tóc phức tạp. Hình dạng và kết cấu có thể được phục hồi hoàn toàn ngay cả với các vùng không nhìn thấy được ví dụ như phía sau của người trong ảnh (Hình 11). Mặc dù vậy, mô hình này vẫn cần được nghiên cứu thêm để cải thiện độ chi tiết.

a) *b)*

Hình 11: *Kết quả mô hình 3 chiều cơ thể người được sinh ra sử dụng mô hình PIFu. a) Ảnh đầu vào, b) Mô hình 3 chiều ở các hướng nhìn khác nhau*

7. Ứng dụng Try on

Ứng dụng Try on áp dụng những kết quả thu được từ nghiên cứu lần này, gồm phân vùng hình ảnh cơ thể người, biến đổi ảnh áo quần theo hình dán của cơ thể người và gắn ảnh áo quần lên ảnh người.

Ứng dụng cho phép người dùng sử dụng dáng người được chụp từ máy ảnh điện thoại hoặc thư viện, kết hợp với các quần áo tạo nên kết quả thử đồ cuối cùng.

Ứng dụng được viết bằng Flutter kết hợp API được viết bằng FastAPI, tạo nên thử đồ online Try on.

Tài liệu tham khảo

[1] Z Cao, T Simon, SE Wei, YA Sheikh, et al. Openpose: Realtime multi-person 2d pose estimation using part affinity fields. The IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)

[2] Taesung Park, Ming-Yu Liu, Ting-Chun Wang, and Jun-Yan Zhu. Semantic image synthesis with spatially-adaptive normalization. In Proc. of the IEEE conference on computer vision and pattern recognition (CVPR)

[3] Ting-Chun Wang, Ming-Yu Liu, Jun-Yan Zhu, Andrew Tao, Jan Kautz, and Bryan Catanzaro. High-resolution image synthesis and semantic manipulation with conditional gans. In Proc. of the IEEE conference on computer vision and pattern recognition (CVPR).

[4] Zhao, Fuwei and Xie, Zhenyu and Kampffmeyer, Michael and Dong, Haoye and Han, Songfang and Zheng, Tianxiang and Zhang, Tao and Liang, Xiaodan, M3D-VTON: A Monocular-to-3D Virtual Try-On Network

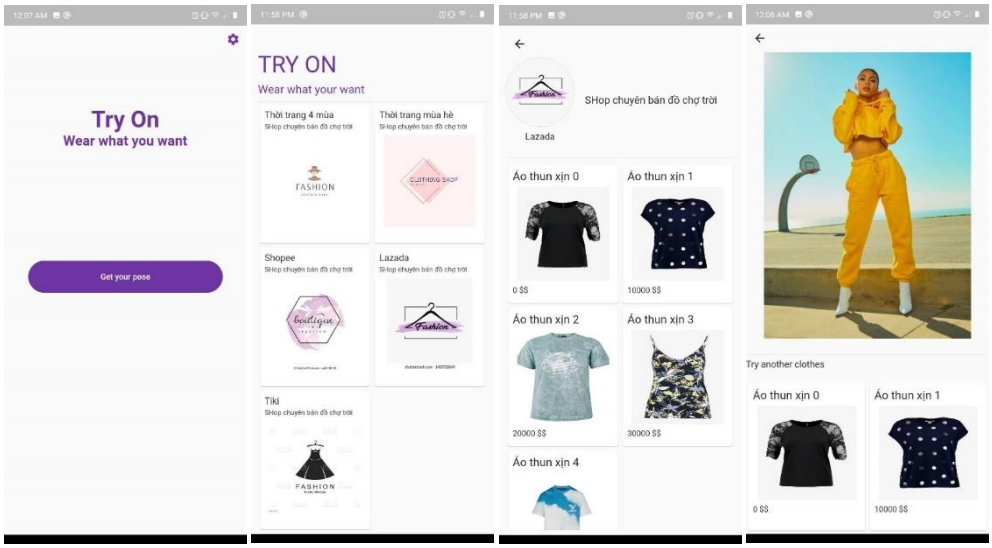
[5] Bochao Wang, Huabin Zheng, Xiaodan Liang, Yimin Chen, Liang Lin, and Meng Yang. Toward characteristic preserving image-based virtual try-on network. In Proc. of the European Conference on Computer Vision (ECCV),

[6] Ruiyun Yu, Xiaoqi Wang, and Xiaohui Xie. Vtnfp: An image-based virtual try-on network with body and clothing feature preservation. In Proc. of the IEEE international conference on computer vision (ICCV),

[7] Han Yang, Ruimao Zhang, Xiaobao Guo, Wei Liu, Wangmeng Zuo, and Ping Luo. Towards photo-realistic virtual try-on by adaptively generating-preserving image content. In Proc. of the IEEE conference on computer vision and pattern recognition (CVPR)

[8] Shunsuke Saito, Zeng Huang, Ryota Natsume, Shigeo Morishima, Angjoo Kanazawa, Hao Li. PIFu: Pixel-Aligned Implicit Function for High-Resolution Clothed Human Digitization. In Proc. of the IEEE International Conference on Computer Vision (ICCV)

[9] Ke Gong, Xiaodan Liang, Yicheng Li, Yimin Chen, Ming Yang, and Liang Lin. Instance-level human parsing via part grouping network. In Proc. of the European Conference on Computer Vision (ECCV), pages 770–785, 2018



Hình 12: *Một số hình ảnh thực tế của ứng dụng*