이미지 기반 가상 착용 이미지 합성 기술 동향

A Survey of Image-based Virtual Try-on Technology

박순찬 (S.C. Park, parksc@etri.re.kr) 감성디지털휴먼연구실 책임연구원
박진아 (J.A. Park, jinahpark@kaist.ac.kr) 한국과학기술원 전산학부 교수
박지영 (J.Y. Park, jiyp@etri.re.kr) 감성디지털휴먼연구실 책임연구원

ABSTRACT

Image synthesis has been remarkably developed in the computer vision domain and various researches have been proposed to generate realistic and high-resolution images. In particular, image-based virtual try-on is an application in fashion domain to simulate wearing clothes. Specifically, using input images of a fashion model and products, an realistic image of the model wearing the provided garments is synthesized. In this paper, we present a comprehensive review of technical trends in image-based virtual try-on technology. We first introduce relevant datasets and discuss their characteristics. Then, we categorize existing image synthesis methods into three main streams: warping-based methods, encoding-decoding-based methods, and diffusion-based methods. Finally, we explore other important research issues in the field of virtual try-on and analyze related researches aimed to tackling those challenges.

KEYWORDS virtual try-on, 가상 착용 이미지 합성, 조건부 이미지 합성, 패션 AI

Ⅰ. 서론

기존 학계에서는 학습된 기계를 통해 사실적인 이미지를 합성하는 기술들이 나날이 발전해 왔다. 특히 신경망을 활용한 기계학습 기술이 크게 발전하면서 제안된 생성적 적대 신경망(GAN: Generative Adversarial Network)은 이미지 합성 기술의 가능성을 크게 넓혔고, 최근에는 노이즈 제거 확산 모델(Denoising Diffusion Model)의 등장으로 언어를 비롯한 다양한 조건에 맞는 사실적인 이미지를 고화질로 합성할 수 있음을 보였다. 나아가 글로벌 AI 기업들이 이미지 합성을 기반으로 하는 다양한 서비스를 공개하면서 이미지 합성이라는 주제는 일반인들도 쉽게 활용할 수 있는 접근성 높은 기술이 되었다.

한편, 패션은 시각적 정보에 크게 의존하는 산업으로 촬영된 이미지를 통하여 상품의 정보를 전달한다. 따라서 공유되는 이미지의 질이 구매자가 소비하게 하는 중요한 요소가 되었고, 패션 기업은 소

* DOI: https://doi.org/10.22648/ETRI.2024.J.3903011
* 본 연구는 과학기술정보통신부가 주관하고 한국지능정보사회진흥원이 지원하는 '인공지능 학습용 데이터 구축 사업(2차)[과제번호: 2020-데이터-위64-1]'와 문화체육관광부 및 한국콘텐츠진흥원의 연구개발진흥사업[과제번호 R2020070002]으로 수행되었음.

비자들을 만족시키기 위하여 막대한 비용을 들여 촬영 스튜디오, 전문 사진사, 전문 패션모델, 후보정 전문가 등을 섭외하여 패션 상업 이미지들을 생산하고 공유해 왔다. 하지만 이러한 과정은 매우 큰 비용을 요구하므로 막대한 예산을 감내할 수 있는 대기업들 위주로 활용되었으며, 상대적으로 상품성은 뒤떨어지지 않으나 촬영에 대한 큰 비용을 감당하지 못하는 작은 기업들은 경쟁력이 뒤떨어질 수밖에 없다. 이에, 자연스럽게 '최근 발전하고 있는 이미지 합성 기술을 활용하여 패션 이미지들을 만들어 낼 수 없을까?'라는 의문이 제기되었고, 이에 다양한 기술이 연구개발되고 있다.

그 중, 이미지 기반 가상 착용 이미지 합성(Virtual-try On) 기술은 그림 1에서처럼 임의의 상품 이미지와 임의의 모델 이미지를 입력으로 하여 학습된 신경망이 해당 모델이 해당 상품을 착용한 것과 같은 이미지를 사실적으로 합성해 내는 기술을 의미한다. 이 기술은 앞서 언급된 상업용 패션 이미지를 촬영하는데 요구되는 막대한 비용을 최소화할 수 있다. 본고에서는 가상 착용 이미지 합성 기술의 연구개발 동향을 소개한다. 연구개발의 가장 기본이 되는 공개 데이터세트를 정리하여 각 데이터세트가 가지는 장단점을 소개한다. 이어 데이터세트들을 바탕으로 만들어진 다양한 연구결과들 및 주요 연구 이슈를 소개하며 가상 착용 이미지 합성 기술들이 지향하고 있는 방향에 대해서 정리한다.

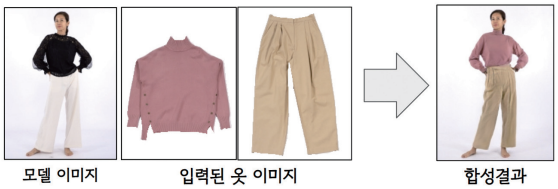


그림 1 가상 착용 이미지 합성 기술의 개요

II. 데이터세트 동향

이미지 합성 기술을 연구개발하기 위해서 양질의 데이터는 매우 중요하다. 하지만 서론에서 언급했던 것처럼 특정 상품과 그 상품을 착용한 패션모델의 이미지를 획득하는 과정은 매우 큰 비용을 요구하고 나아가 촬영된 이미지는 다양한 법적 권리가 존재하기 때문에 연구개발 활용 시 주의가 필요하다. 본 장에서는 합성 기술을 연구개발하기 위한 목적으로 구축되고 공개된 데이터세트에 대해서 표 1[1,5,7,8]과 같이 정리하고, 각 데이터세트의 상세에 관해서 설명한다.

표 1 가상 착용 이미지 합성 기술 연구개발을 위해 활용할 수 있는 데이터세트 및 세부 정보

이름	페어 개수	대응 방법 (상품:모델)	상품 타입	해상도	어노테이션	법적 문제
Zalando[1]	16,253	1:1	상의	256×192	착용정보	문제 존재(활용불가)
Zalando-HD[5]	13,679	1:1	상의	1024×768	착용정보	언급되지 않음
DressCode[8]	53,759	1:1	상의 하의 드레스	1024×768	착용정보	해결
Fashion-HD[7]	117,270	N:1	모자 상의 하의 드레스	1280×720	착용정보 모델영역 상품영역	해결

출처 Reproduced from [1,5,7,8].

1. Zalando 데이터세트

Zalando 데이터세트는 신경망을 이용한 가상 착용 이미지 합성 관련 최초의 데이터세트로 옷 이미지 변형을 활용한 모델 이미지 합성 방법과 함께 소개되었다[1,2]. 해당 데이터세트는 온라인 쇼핑몰에서 크로울링한 16,253개의 모델, 상품 쌍으로 이루어진 데이터세트로 여성 모델의 상의 상품만 포함하고 있다. 모든 이미지는 256×192의 다소 낮은 해상도를 가지고 있으며, 상품과 모델의 연결 관계 외에는 직접 레이블링한 어노테이션은 존재하지 않는다. 한편 기학습된 다른 신경망이 도출한 모델의 자세 결과[3]와 옷 전경 추정 결과는 의사 라벨(Pseudo-label)로서 제공된다. 하지만 온라인 쇼핑몰에서 무단으로 크로울링된 데이터이기 때문에 유포에 있어 법적인 문제가 존재하여 현재 데이터는 공개되지 않고 있다[4].

2. Zalando-HD 데이터세트

기존의 Zalando 데이터세트가 256×192 해상도에 국한되어 있다는 단점을 개선하기 위해 1024×768 이미지 해상도를 가지는 이미지 13,679쌍을 포함하는 Zalando-HD 데이터세트가 한국과학기술원이 발표한 논문에서 공개되었다[5]. 해상도 외 데이터세트 구성은 유사하며 패션모델에 대한 보다 자세한 영역 추출을 가능하게 하는 Look-into-person 데이터[6] 기반 의사 라벨링 결과가 포함된 것이 특징이다. Zalando-HD 데이터세트 역시 크로울링된 데이터이기 때문에 법적인 문제를 내포하고 있다.

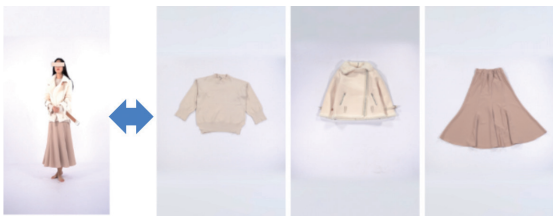


그림 2 하나의 패션모델에 다수의 상품이 대응되어 있는 Fashion-HD

3. Fashion-HD 데이터세트

Fashion-HD는 2021년 한국전자통신연구원이 한국지능정보사회진흥원에서 주관하는 과제에서 취득된 데이터세트로서 대한민국 국민을 대상으로 공개되어 있는 데이터세트이다[7]. 기존 데이터세트가 가지고 있는 해상도 문제와 한 모델이 착용하고 있는 다수 상품에 대한 정보를 구성한 것이 특징이다. 의류는 그림 2와 같이 외투, 상의, 하의로도 구성되어 있으며 모자와 신발에 대한 데이터도 일부 존재한다. 법적인 문제를 극복하기 위해서 상품 디자이너, 전문 사진사, 패션모델의 동의를 얻어 취득한 데이터이며 패션모델 이미지의 경우 비식별화 처리가 되어있어 연구적인 용도의 경우 법적인 문제 없이 활용할 수 있는 데이터세트이다.

4. DressCode 데이터세트

DressCode 데이터세트는 2022년 이탈리아 소재의 Modena and Reggio Emilia 대학에서 발표한 연구에서 소개되었다[8]. YOOX-NET-A-PORTER GROUP의 협력으로 카탈로그에 존재하는 상품 및 패션모델 사진을 취합하여 데이터세트로 구성하였다. 상품의 종류가 상의, 하의, 드레스로 확장되는 것을 특징으로 한 데이터세트이지만 상품과 모델이 1:1로 매칭되어 복합적인 전신 착용 이미지 합성이 어렵다. 예를 들어 티셔츠와 모델이 매칭되어 있을 경우, 해당 모델이 착용하고 있는 바지에 대한 상품 정보는 존재하지 않는다. 53,759쌍의 데이터가

1024×768 해상도로 수집되어 있어 Zalando-HD와 같이 고해상도 이미지 합성 연구에 활용될 수 있다.

Ⅲ. 이미지 합성 기술 연구 동향

본 장에서는 실제 이미지 합성을 수행하기 위한 가상 착용 이미지 합성 연구들의 동향을 소개한다. 관련된 연구들은 앞선 장에서 소개한 데이터세트를 활용하여 수행되며, 크게는 옷 이미지를 변형한 뒤 이를 활용하여 패션모델 이미지를 합성하는 연구와 패션모델과 상품을 분석하여 재합성하는 인코딩-디코딩 기술을 활용하는 연구로 나뉜다. 본 장에서는 관련 이미지 합성 기술의 개요를 간단히 정리한 다음, 옷 이미지 변형 기반 합성 기술 연구, 인코딩-디코딩 기반의 합성 연구, 확산 모델을 활용한 연구 순으로 연구 동향을 소개한다.

1. 이미지 합성 연구의 개요

Ⅱ장에서 설명한 바와 같이 가상 착용 이미지 합성 데이터세트는 옷 이미지 I_c와 해당 옷을 착용하고 촬영한 패션모델 이미지 I_h가 짝을 이루며 구성된다. 그림 3은 이미지 합성 기술이 학습되는 과정을 도식화한 것이다. 주어진 입력 패션모델의 이미지에서 상품과 관련된 부분(예: 옷, 팔, 다리 등)을 삭제시켜 I_h'를 구한 다음 이를 활용하여 이미지를 합성하여 $\tilde{I}_h$를 얻는다. 그 후 $\tilde{I}_h$와 원본 I_h를 비교하여 손실을 계산함으로써 신경망이 학습된다. 이를 간단하게 수식으로 표현하면 가상 착용 이미지 합성을 표현하는 함수 f_{VITON}은 다음과 같이 표현될 수 있다.

$$f_{VITON}(I_h', I_c) = I_h$$

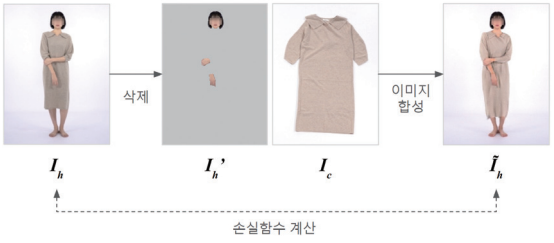


그림 3 모델 이미지를 삭제한 다음 다시 복원하는 방식으로 학습하는 착용 이미지 합성

2. 옷 이미지 변형 기반 합성 연구

옷 이미지 변형이란 입력된 옷 이미지를 목표 패션모델의 신체(체형, 자세, 위치 등)에 맞게 변형하는 과정을 의미하며 단순하게는 와핑(Warping)이라고 명명되기도 한다. 그림 4와 같이 옷 이미지를 변형한 다음 합성에 활용하면 옷 이미지가 가지고 있는 세부적인 디자인들을 합성 결과에 전달할 수 있다는 장점이 있으며, 초기의 연구들이 해당 방법으로 제안되었다[1,2]. 보다 자세하게 전체 신경망은 옷 이미지를 변형하는 신경망과 최종 이미지를 합성하는 신경망으로 구성되고 각 신경망은 개별적으로 학습된 뒤 순차적으로 구동되며 모델 이미지를 합성한다. 이러한 방식을 2-stage 방법이라고 명명한다. 특히 CP-VTON[2]에서는 기하학적 매칭(Geometry Matching)에서 활용되는 TPS(Thin Plate Spline) 변형[9]을 위한 파라미터를 신경망 학습으로 추정하는 방식을 제안하여 입력된 옷 이미지를 자세에 맞게 변형한 뒤 활용함으로써 옷 이미지가 가지고 있는 디테일한 픽셀정보(예: 줄무늬, 호피무늬, 로고 등)

그림 4 입력 상품을 변형시킨 다음 합성에 활용하는 예

그림 5 입력 모델의 자세 정보와 옷 정보를 활용하여 영역을 합성하는 과정

들을 유지하는 착용 이미지 합성이 이루어질 수 있도록 하였다.

제시된 흥미로운 통찰에도 불구하고 그림 3과 같이 패션모델의 많은 정보를 지운 채 체형에 대한 부정확한 정보로 옷 이미지 변형이 일어나기 때문에 변형 결과가 부정확할 수 있다는 단점이 존재한다. 이를 극복하기 위하여 그림 5와 같이 보다 자세한 패션 모델 정보를 획득하기 위해 I_h'가 I_c를 착용했을 경우의 영역(Layout)을 일차적으로 합성하는 단계를 추가하는 방식의 연구들이 발표되었다[5,10-14]. (영역합성)-(옷 이미지 변형)-(패션모델 이미지 합성), 세 단계로 이루어져 3-stage 방법이라고 명명된다. 영역 합성 단계의 추가로 뒤따르는 옷 이미지 변형 기술과 패션모델 합성 기술은 보다 상세한 정보를 활용할 수 있기 때문에 전반적으로 이미지 합성의 정확도가 향상되는 효과가 있음을 보였다.

이렇게 사전정보를 추가하여 정확도를 높이던 방법들에 이어 새로운 옷 이미지 변형 방법을 제안함으로써 성능을 향상시킨 연구들도 존재한다. Appearance Flow[15]란 두 채널로 구성된 정보로서 각 픽셀은 해당 픽셀이 2D 공간에서 이동해야 할 오프셋을 의미한다. 보다 로컬정보를 유지하며 변형이 수행되는 기존의 아핀변환이나 TPS변환과 비교하였을 때, 보다 유연하고 정확한 변형이 이루어지도록 학습을 통해 유도할 수 있다. Appearance Flow를 이용한 연구는 참고문헌 [16]에서 먼저 제안된 다음, 뒤따르는 다양한 연구들에서 널리 활용되고 있다[17-20]. 특히 참고문헌 [18]의 경우는 Appearance Flow에 Attention 개념을 더한 Deformable Attention을 활용해 이미지 합성에 유리한 정보들에 집중하여 학습을 유도하는 방법을 제안하였다.

하지만 앞서 언급한 분리된 단계로 구성된 연구들은 앞서서 구한 사전정보(Prior)를 후속 신경망에게 전달하며 보다 정확한 추정을 수행함을 가정하고 있지만, 분리되어 학습되는 방식은 그 구성 및 학습이 복잡할 뿐만 아니라 앞선 네트워크에서 오류가 발생할 경우 뒤따르는 단계에서 연쇄적으로 오류가 발생하는 문제점을 내포하고 있었다. 이에 최근 단계를 단순화시키는 방향으로 연구가 진행되어 왔다[18-20]. HR-VITON[19]의 경우 3-stage 방법에서 영역 합성과 옷 이미지 변형 과정을 통합하여 2-stage 방법을 제안하였고, 참고문헌 [18,20]의 경우 모든 단계를 통합한 single-stage 방법을 제안하였다. 해당 연구들에서는 통합하여 학습되는 과정이 최종 합성 이미지의 질을 향상시킬 수 있다는 결과를 보였다. 특히 WG-VITON[20]의 경우 single-stage 구성이 전신 착용 이미지 합성을 위한 종단간 학습을 유도하며 파라미터를 크게 절감시키면서도 향상된 성능을 보였다.

3. 인코딩-디코딩 기반 합성 연구

앞서 설명한 기하 매칭 방법을 활용하지 않고 인코딩으로 옷과 모델의 특징을 분석한 뒤 이를 활용하여 착용 이미지를 합성하는 연구들도 존재한다. Zalando Research의 경우 2019년 논문을 통해 입력된 다수 상품을 인코딩한 결과를 바탕으로 Progressive GAN을 구동시켜 고해상도 이미지를 합성하는 연구를 발표하였다[21]. 한편 Amazon Lab126은 2020년 논문을 통해 발표한 O-VITON에서 상품 이미지 없이 모델 이미지만으로 다양한 가상 착용 이미지 합성을 수행할 수 있는 방법을 소개하였다.

인코딩된 잠재공간(Latent Space)의 특징들을 주어진 모델에 맞게 재구성하여 이를 바탕으로 착용 이미지 합성을 수행하였다[22]. 옷 이미지 변형 방법과의 의존성을 해결한다는 장점에도 불구하고, 인코딩-디코딩 방식은 인코딩 과정에서 상품의 세부 특징(로고, 패턴 등)들을 유지하지 못할 수 있기 때문에 여전히 발전이 필요한 상황이다. 최근에는 이러한 문제를 확산 모델을 활용하여 이를 극복하고자 하는 연구도 발표되어 그 가능성을 보여주었다[23].

4. 확산 모델을 활용한 연구

최근 이미지 합성분야에서는 확산 모델(Diffusion Model)이 보다 안정적으로 이미지 합성을 학습하고 수행할 수 있음이 발표되었고, 이에 다양한 기술들이 연구개발되어 비약적인 발전을 이루고 있다[24-27]. 가상 착용 이미지 합성 분야에서는 2023년부터 적용되기 시작하여 다양한 연구개발이 이루어지고 있다[23,28]. 확산 모델이 필연적으로 가지는 비효율성의 단점을 극복하기 위하여 참고문헌 [29]는 잠재 확산 모델(Latent Diffusion Model)을 활용하여 효율성을 개선시키면서도 정확한 착용 이미지 합성방법을 소개하였다.

한편 확산 모델을 가상 착용 이미지 합성 데이터세트에 대해서 학습하는 형태가 아닌, 초거대 데이터를 기학습한 확산 모델들을 활용하는 방법도 검토되고 있다[30,31]. 초거대 데이터를 학습한 기학습 모델은 이미 사실적인 이미지 합성에 필요한 사전 지식(Prior)을 이미 가지고 있기 때문에 이를 착용 이미지 합성에 적절한 형태로 동작할 수 있도록 구성한다면 Ⅱ장에서 설명했던 새로운 데이터 구축에 드는 많은 비용을 절감하면서도 착용 이미지 합성을 수행할 수 있다.

5. 기타 착용 이미지 합성 연구이슈들

가. 합성 이미지 해상도 향상

널리 사용된 Zalando 데이터세트에서 공개된 이미지의 해상도는 256×192이기 때문에 실제 패션 산업에서 활용하는 이미지의 해상도와는 큰 괴리가 존재하였다. 이후 1024×768 이상의 해상도를 가지는 이미지들로 구성되어 고해상도 이미지 합성 연구가 가능해졌다[5,7,8]. 고해상도 합성 방법을 제안한 VITON-HD[5]에서는 고해상도에 보다 강인한 것으로 알려져 있는 의미론적 이미지 합성 신경망[32]을 기반으로 옷 이미지 변형 오류를 개선할 수 있는 의미론적 이미지 합성 방법을 제안하였다. 최근 2023년에는 고해상도 합성을 바로 수행하지 않고 초해상도(Super-resolution) 단계를 최종 단계에 추가하여 합성된 이미지의 해상도를 1024×1024까지 개선시키는 연구들도 소개되었다[33].

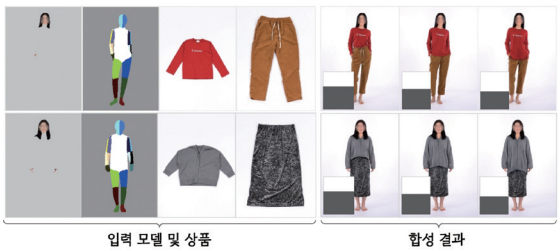
나. 적용 상품의 확장

앞선 Ⅱ장에서 설명했던 것처럼 공개된 데이터세트인 Zalando 데이터세트, Zalando-HD 데이터세트, Dresscode 데이터세트에서는 상품과 패션모델의 매칭을 1:1로 제한하고 있다[2,5,8]. 이는 쇼핑몰들이 가지고 있는 데이터(카탈로그 등)의 구성을 따르기 때문이다. 그 이후 Fashion-HD[7] 데이터세트에서는 보다 범위를 확장하기 위하여 모자, 외투, 상의, 하의, 드레스, 신발을 포함하는 데이터세트를 구축하였고, 상품과 모델의 연결관계가 N:1로 구성되어 복수 개의 상품을 활용하여 전신 착용 이미지 합성을 수행할 수 있는 기틀이 마련되었다. 전신 착용 이미지 합성에 관한 연구로는 Fashion-HD 데이터세트를 일부 활용한 연구들이 전신 이미지 합성을 수행하고 있으며[20], 그 외에도 공개되지 않은

자체 구축된 데이터세트를 활용하여 다양한 상품 타입을 활용한 전신 착용 이미지 합성이 연구개발된 바 있다[21,22].

다. 상품 착용 방법 반영한 이미지 합성

기존 연구들이 합성 이미지의 사실감 향상을 목표로 진행되었다면, 보다 다양한 착용법을 반영하여 합성을 유도하는 방법을 제안한 연구들도 존재한다. WG-VITON[20]은 상 · 하의를 활용하는 전신 착용 이미지 합성 연구에서 다수의 상품을 착용할 때 다양한 착용법이 가능하고, 이를 합성 과정에 반영해야 함을 언급하였다. 예를 들어 티셔츠와 바지가 주어졌을 때, 티셔츠를 바지에 넣어 입을 수도 있고 빼서 바지를 가린 채로 입을 수도 있다. 이를 위하여 그림 6과 같이 합성의 조건으로 착용법을 조절하는 이진 마스크 입력을 추가하여 사용자로 하여금 합성되는 패션 모델의 착용 방법을 조절할 수 있게 하는 방법을 제안하였다. 나아가 최근 연구에서는 패션모델의 키포인트[34]를 참고로 하여 옷이 반영되는 영역을 조절할 수 있도록 하였다[35]. 예를 들어, 긴 팔 티셔츠를 착용할 때 패션모델의 키포인트를 참고하여 손목을 덮는 방식으로 합성할 수도 있으며 손목보다 짧은 소매를 가지도록 합성할 수도 있다. 이런 착용 방법의 문제는 외투 등이 추가되어 착용하는 상품이 많고 다양해질수록 더욱 중요한 이슈가 되며 추후 다양한 상품을 활용하는 합성 연구에서 중요하게 다뤄질 수 있다.



출처 Reprinted from [20].

그림 6 다양한 착용 방법을 가지는 이미지의 합성이 가능한 WG-VITON[20] 수행 결과

라. 합성 성능 평가 방법

가상 착용 이미지 합성 연구들에서는 학계에서 통용되는 이미지 합성 분야에서 널리 활용되는 지표들을 활용하여 정량적 평가를 수행한다. 임의의 패션모델과 임의의 상품을 활용하면 정답 이미지가 존재하지 않기 때문에 이미지들의 잠재공간상에서의 특징의 분포를 비교하는 FID(Frechet Inception Distance)[36]와 KID(Kernel Inception Distance)[37]가 널리 활용된다.

하지만, 착용 이미지 합성은 사실적인 이미지를 합성하는 것이 유일 목표가 아니기 때문에 단순히 이미지에서 표현되는 특징들을 분석하여 추산한 수치만으로 그 성능을 정확하게 평가할 수 없다. 예를 들어, 합성된 모델이 이미지만 보았을 때 매우 사실적이라고 하더라도 입력된 옷을 잘 반영하지 못했다면 착용 이미지 합성은 실패한 것으로 간주되어야 한다. 이러한 특징을 반영하기 위하여 가상 착용 이미지 합성 분야의 연구에서는 사용자 선호도 평가를 수행한다. 보다 자세하게는 A신경망과 B신경망이 같은 패션모델 및 상품 입력으로 합성한 이미지를 피험자에게 제공하고 어느 이미지가 더 이미지 합성을 잘 수행했는지를 질문하고 그 결과를 취합하여 평가한다. 보다 자세하게는 합성된 이미지의 사실감(Realism)과 입력한 옷과의 연관성(Coherence)을 별도의 질문으로 구성하여 두 가지 관점에서 평가를 수행한 연구들도 존재한다[8,19]. 한편, WG-VITON에서는 상 · 하의 상품을 활용할 때는 상의, 하의, 모델을 잘 반영했는지를 복합적으로 평가하여 선호도가 취합되어야 하므로 단순한 선호도 조사가 아닌, 각 이미지에 대해서 1~5점 사이의

점수를 취합하는 실험을 수행하였다[20]. 보다 착용 이미지 합성의 입력 조건이 상품이나 착용법 등을 고려하여 다양해질 경우 사용자 평가 실험은 더욱 섬세하게 고려되어야 하나 이에 대해 추가적인 많은 연구가 필요한 실정이다.

Ⅳ. 결론

본고에서는 가상 착용 이미지를 합성하는 이미지 기반 가상 착용 이미지 합성분야의 연구 동향들을 소개해 보았다. 패션 데이터 구축 및 이를 활용한 이미지 합성 기술 개발의 연구 형태를 벗어나 대규모 데이터를 기학습한 확산 모델을 활용하여 별도의 소모적인 데이터 구축 없이도 원하는 형태의 착용 이미지 합성을 수행하는 방법에 대한 연구가 활발히 진행될 것으로 예상된다.

한편, 현재 실험실 레벨의 다양한 연구개발 결과들이 시장에 존재하는 다양한 형태의 옷과 모델에 대해서 강인하게 동작할 수 있도록 발전시키고 검증하는 과정이 꾸준히 요구되고 있으며, 추후 이러한 과정을 거쳐 발전된 가상 착용 이미지 합성 기술은 패션 시장에서의 비용을 줄이고 구매자로 하여금 다양한 경험을 제공해 줄 수 있는 핵심 기술이 될 것으로 전망한다.

용어해설

이미지 합성 기술 생성적 AI 신경망이 이미지의 특성을 파악하여, 실제와 같은 이미지를 만들어내는 기술을 의미

잠재공간 분석하려고 하는 이미지의 특징이 정리된 공간. 다시 말해 타겟 이미지 각각은 벡터로 정리될 수 있는데, N개 데이터를 모두 분석했을 경우 N개 벡터가 만드는 좌표계상에서의 공간을 의미함

해상도 이미지의 픽셀 크기를 의미하며, 예를 들어 가로 100픽셀, 세로 100픽셀로 이루어진 정사각형 모양의 이미지의 해상도는 100×100으로 표시할 수 있음

약어 정리

FID	Frechet Inception Distance
GAN	Generative Adversarial Network
KID	Kernel Inception Distance
VITON	Virtual Try-on

참고문헌

[1] X. Han et al., “Viton: An image-based virtual try-on network,” in Proc. CVPR, (Salt Lake City, Utah. USA), June 2018.

[2] B. Wang et al., “Toward characteristic-preserving image-based virtual try-on network,” in Proc. ECCV, (Munich, Germany), Sept. 2018.

[3] Z. Cao et al., “OpenPose: Realtime multi-person 2D pose estimation using part affinity fields,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 43, no. 1, 2021, pp. 172-186.

[4] https://github.com/sergeywong/cp-vton/

[5] S. Choi et al., “Viton-hd: High-resolution virtual try-on via misalignment-aware normalization,” in Proc. CVPR, (Virtual), June 2021.

[6] X. Liang et al., “Look into person: Joint body parsing & pose estimation network and a new benchmark,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 41, no. 4, 2018, pp. 871-885.

[7] 박순찬 외, “복수 상품을 활용하는 고화질 패션 착용영상 생성을 위한 데이터세트 Fashion-HD 및 그 활용,” 정보과학회 컴퓨팅의 실제 논문지, 제28권 제1호, 2022, pp. 68-73.

[8] D. Morelli et al., “Dress code: High-resolution multi-category virtual try-on,” in Proc. ECCV, (Tel Aviv, Israel), Oct. 2022.

[9] S. Belongie, J. Malik, and J. Puzicha, “Shape matching and object recognition using shape contexts,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 24, no. 4, 2002, pp. 509-522.

[10] H. Yang et al., “Towards photo-realistic virtual try-on by adaptively generating-preserving image content,” in Proc. CVPR, (Virtual), June 2020.

[11] S. Jandial et al., “Sievenet: A unified framework for robust image-based virtual try-on,” in Proc. WACV, (Snowmass Village, CO, USA), Mar. 2020.

[12] K. Li et al., “Toward accurate and realistic outfits visualization with attention to details,” in Proc. CVPR, (Virtual), June 2021.

[13] A. Chopra et al., "Zflow: Gated appearance flow-based virtual try-on with 3d priors," in Proc. ICCV, (Virtual), Oct. 2021.

[14] H. Yang, X. Yu, and Z. Liu, "Full-range virtual try-on with recurrent tri-level transform," in Proc. CVPR, (New Orleans, LA, USA), June 2022.

[15] T. Zhou et al., "View synthesis by appearance flow," in Proc. ECCV, (Amsterdam, Netherlands), Oct. 2016.

[16] X. Han et al., "Clothflow: A flow-based model for clothed person generation," in Proc. ICCV, (Seoul, Rep. of Korea), Nov. 2019.

[17] Y. Ge et al., "Parser-free virtual try-on via distilling appearance flows," in Proc. CVPR, (Virtual), June 2021.

[18] S. Bai et al., "Single stage virtual try-on via deformable attention flows," in Proc. ECCV, (Tel Aviv, Israel), Oct. 2022.

[19] S. Lee et al., "High-resolution virtual try-on with misalignment and occlusion-handled conditions," in Proc. ECCV, (Tel Aviv, Israel), Oct. 2022.

[20] S. Park and J. Park, "Single-stage virtual try-on for top and bottom clothes with wearing style control," Available at SSRN 4379142 (2023).

[21] G. Yildirim et al., "Generating high-resolution fashion model images wearing custom outfits," in Proc. ICCV, (Seoul, Rep. of Korea), Nov. 2019.

[22] A. Neuberger et al., "Image based virtual try-on network from unpaired data," in Proc. CVPR, (Virtual), June 2020.

[23] A.K. Bhunia et al., "Person image synthesis via denoising diffusion model," in Proc. CVPR, (Vancouver, Canada), June 2023.

[24] J. Sohl-Dickstein et al., "Deep unsupervised learning using nonequilibrium thermodynamics," in Proc. ICML, (Lille, France), Jul. 2015.

[25] J. Ho, A. Jain, and P. Abbeel, "Denoising diffusion probabilistic models," in Proc. NeurIPS 2020, (Virtual Only), Dec. 2020, pp. 6840-6851.

[26] P. Dhariwal and A. Nichol, "Diffusion models beat gans on image synthesis," in Proc. NeurIPS 2021, (Virtual Only), Dec. 2021, pp. 8780-8794.

[27] R. Rombach et al., "High-resolution image synthesis with latent diffusion models," in Proc. CVPR, (New Orleans, LA, USA), June 2022.

[28] D. Morelli et al., "LaDI-VTON: Latent diffusion textual-inversion enhanced virtual try-on," arXiv preprint, CoRR, 2024, arXiv: 2305.13501 (2023).

[29] X. Han et al., "Controllable person image synthesis with pose-constrained latent diffusion," in Proc. ICCV, (Paris, France), Oct. 2023.

[30] J. Kim et al., "StableVITON: Learning semantic correspondence with katent diffusion model for virtual try-on," arXiv preprint, CoRR, 2023, arXiv: 2312.01725.

[31] Y. Choi et al., "Improving diffusion models for virtual try-on," arXiv preprint, CoRR, 2024, arXiv: 2403.05139.

[32] T. Park et al., "Semantic image synthesis with spatially-adaptive normalization," in Proc. IEEE CVPR, (Long Beach, CA, USA), June 2019.

[33] L. Zhu et al., "TryOnDiffusion: A tale of two UNets," in Proc. CVPR, (Vancouver, Canada), June 2023.

[34] T.-Y. Lin et al., "Microsoft coco: Common objects in context," in Proc. ECCV, (Zurich, Switzerland), Sept. 2014

[35] C.Y. Chen et al., "Size does matter: Size-aware virtual try-on via clothing-oriented transformation try-on network," in Proc. CVPR, (Vancouver, Canada), June 2023.

[36] M. Heusel et al., "Gans trained by a two time-scale update rule converge to a local nash equilibrium," in Proc. NIPS 2017, (Long Beach, CA, USA) Dec. 2017.

[37] M. Bińkowski et al., "Demystifying mmd gans," in Proc. ICLR, (Vancouver, Canada), Apr. 2018.