한국섬유공학회지, Vol. 62, No. 2, 91-100
https://doi.org/10.12772/TSE.2025.62.091

ISSN 1225-1089 (Print)
ISSN 2288-6419 (Online)

딥러닝을 활용한 사용자의 가상 패션 착용 이미지 및 숏폼 생성 프로그램 개발

황준형[1] · 문형민[2] · 최승아[3] · 고준석[1,3] · 박창규[1,3†]

[1]건국대학교 재료공학과, [2]건국대학교 유기나노시스템공학과, [3]건국대학교 재료공학과 첨단소재융합전공

Development of a Deep Learning-based Program for User's Virtual Try-on Images and Short Forms Generation in Fashion

Junhyung Hwang[1], Hyungmin Moon[2], Seunga Choi[3], Joonseok Koh[1,3], and Chang Kyu Park[1,3†]

[1]Department of Materials Science and Engineering, Konkuk University, Seoul 05029, Korea
[2]Department of Organic and System Engineering, Konkuk University, Seoul 05029, Korea
[3]Advanced Materials Program, Department of Materials Science and Engineering, Konkuk University, Seoul 05029, Korea

[†]Corresponding Author: Chang Kyu Park
E-mail: cezar@konkuk.ac.kr

Received February 6, 2025
Revised March 28, 2025
Accepted April 20, 2025

Abstract: The study presents the development of an immersive virtual fitting program leveraging deepfake technology and its extension to video formats. Traditional online shopping environments have been limited by challenges such as the inability to try on clothing, lack of realism, and restricted selection of apparel, thereby constraining the consumer experience. This research integrates advanced deep learning technologies, including StableVITON, DeepFaceLab, and MusePose, to seamlessly synthesize consumer facial images with model images and adapt selected garments appropriately. Furthermore, motion synthesis was implemented to provide virtual fitting results in video format, enabling a more realistic and natural user experience. The results demonstrate that the program achieves high-quality synthesis not only for apparel from specific retailers but also for garments from diverse sources, offering a highly realistic fitting experience. Future developments will focus on optimizing code efficiency and expanding into 3D domains to deliver faster and more precise virtual fitting services.

Keywords: *deep learning, deepfake, virtual fitting*

1. 서 론

1.1. 연구 배경

COVID-19 팬데믹 이후, 온라인 쇼핑몰의 이용률이 급격히 증가했다[1]. *Table* 1에서와 같이 2023년 기준, 배달·포장 음식의 경우 오프라인(전통시장, 대형할인마트, 슈퍼, 백화점, 전문점) 쇼핑몰의 이용률 비율은 41.7%, 온라인(TV홈쇼핑, 인터넷 쇼핑몰, 소셜 커머스, 배달앱) 쇼핑몰의 이용률 비율은 58.3%로 온라인이 오프라인보다 15%p 이상 차이를 보여준다[2]. 서적·문구류의 경우 오프라인 이용률 비율은 46.3%, 온라인 이용률 비율은 53.7%로 온라인이 오프라인보다 약 7%p 높은 비율을 보여준다[3].

Table 1에서 의류(옷, 신발, 가방)의 경우에도 온라인 판매가 눈에 띄게 성장하고 있다[4]. 하지만 의류의 구매경로 통계를 보면 여전히 많은 소비자들이 오프라인을 통해 의류를 구매하고 있는 것으로 나타난다. 오프라인의 비율은 57.8%, 온라인의 비율은 42.2%로 타 품목과는 반대로 오프라인이 온라인보다 약 15%p 더 높은 것으로 나타난다[5].

Figure 1과 같이 의류의 경우 온라인보다 오프라인의 비율이 더 높게 나타나는 이유는 *Figure* 2 Raydiant에서 조사한 The State of Consumer Behavior 통계에서 찾아볼 수 있다. 오프라인 쇼핑을 더 선호하는 이유로는 Because I like to see and touch products before I buy them이 33.1%로 가장 높았고, Because I enjoy the experience of shopping in

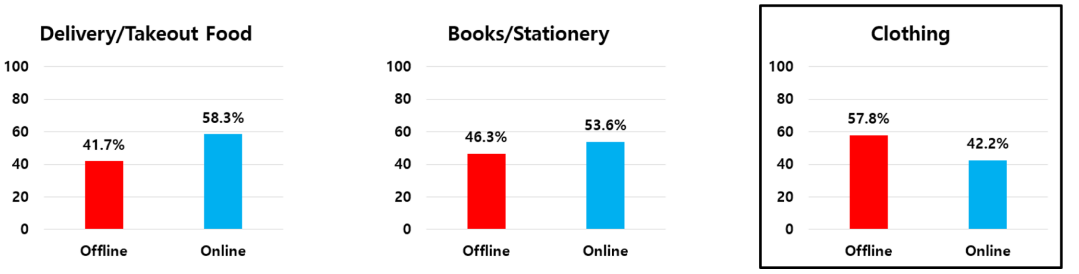

Figure 1. Statistics of purchasing channels by item.

Table 1. Online and offline purchasing channels

Main purchase channel		Delivery/ takeout food (%)	Books/ stationery (%)	Clothing (clothes, shoes, bags) (%)
Offline	Traditional market	8.6	1.0	10.8
	Large discount store	14.1	10.3	13.6
	Supermarket (convenience store)	7.2	1.4	0.7
	Department store	1.2	1.9	16.1
	Specialty store	10.3	31.8	16.6
	Subtotal	41.7	46.3	57.8
Online	TV home shopping	1.3	1.1	6.9
	Online shopping mall	4.8	34.4	25.5
	Social commerce	6.2	17.7	9.5
	Delivery app	46.0	0.5	0.4
	Subtotal	58.3	53.6	42.2
Total		100.0	100.0	100.0

* Based on a 2023 survey conducted in Daegu Metropolitan City.

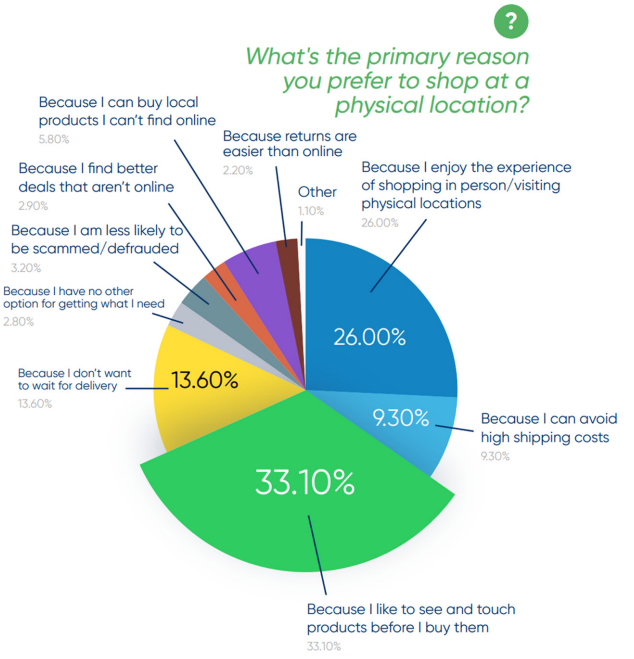

Figure 2. The state of consumer behavior 2021.

person/visiting physical locations가 26.0%, Because I don't want to wait for delivery가 13.6%, Because I can avoid high shipping costs가 9.3%로 뒤를 이었다[6].

온라인 의류 쇼핑몰에서는 착용 전 상품을 직접 확인할 수 없기 때문에, 소비자들은 상품의 착용 이미지, 색상, 그리고 사이즈에 대한 불확실성을 가진다. 이러한 불확실성은 소비자 행동에 새로운 트렌드를 만들어 냈다. 그 중 하나가 Bracketing 쇼핑 방식이다. Bracketing 쇼핑이란 온라인 쇼핑몰에서 다양한 사이즈, 다양한 색상과 디자인의 품목을 구매해서 착용해보고 가장 마음에 드는 제품을 제외하고 나머지를 반품하는 쇼핑 행태를 의미한다. *Figure* 3을 보면 Bracketing 쇼핑 방식은 2017년에는 약 40%의 소비자가 사용해본 적이 있다고 하고, 2021년에는 무려 63%의 소비자가 사용해본 적이 있다고 조사되었다[7,8]. 이는 곧 온라인 쇼핑에서의 시각적인 정보 부족이 소비자의 구매 결정을 어렵게 하는 원인 중 하나임을 보여준다.

이러한 온라인 쇼핑의 시착 불가능의 문제로 인해 여러 플랫폼에서는 가상 피팅 서비스를 시행한 적이 있다. *Figure* 4는 L사의 가상 피팅 서비스이다. 소비자의 신체 사이즈 정

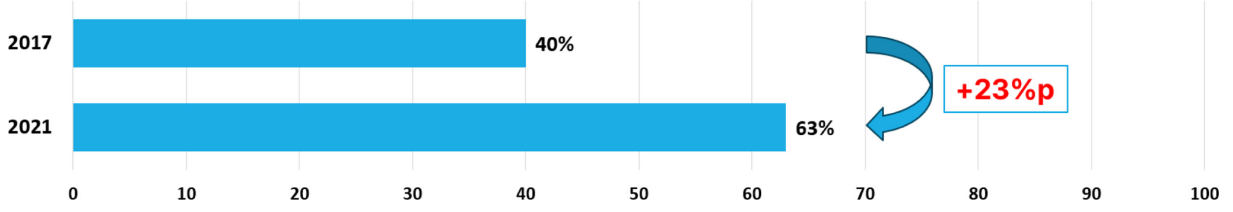

Figure 3. Percentage of consumer utilizing Bracketing purchase method.

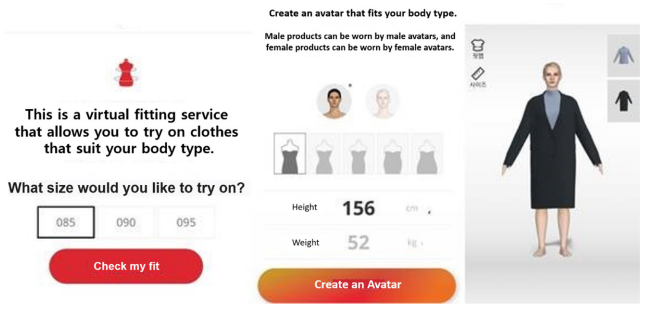

Figure 4. Virtual fitting service of company L.

Figure 5. Virtual fitting service of company A.

보를 기반으로 아바타를 생성한 후 옷을 착용하여 의류 사이즈, 핏, 실루엣 등을 확인할 수 있다. 하지만 CLO, Unity와 같은 특정 소프트웨어를 사용한 서비스이기 때문에 미리 만들어져 있는 3D 의류만 가상 피팅이 가능했고, 기본으로 제공되는 아바타의 얼굴을 사용하기 때문에 소비자가 의상의 어울림을 판단하는 데 어려움이 있다.

Figure 5는 A사의 가상 피팅 서비스이다. 16명의 모델 이미지가 존재하고 여러 포즈가 저장되어 있다. 소비자는 자신이 선택한 모델에게 디지털 방식으로 의류를 착용시켜 가상 피팅을 진행한다. 하지만 이 또한 미리 만들어져 있는 40여 가지의 드레스밖에 착용이 불가하고, 얼굴 또한 소비자의 얼굴이 아니기 때문에 옷의 어울림을 판단하는 데 어려움이 있다.

그 외에도 기존의 가상 피팅 관련 연구에서는 정적인 이미지 합성 중심의 기술 개발에 집중되어 있었다. Jung 등[9]은 중고 의류 거래 플랫폼에서 CNN 기반 합성을 통해 모델에게 정적 가상 피팅을 구현하였으며, Park 등[10]은 GAN을 활용하여 소비자의 사진에 의류를 합성하는 프로젝트를 수행하였다. 그러나 이러한 연구들 역시 소비자의 실제 얼굴을 반영하지 않거나, 동적인 영상 기반의 합성을 시도하지 않았다는 점에서 한계가 존재한다.

또한, 최근 동영상 플랫폼을 중심으로 짧은 길이의 영상인 숏폼 콘텐츠가 빠르게 성장하고 있다. 숏폼은 보통 15초에서 1분 정도 길이의 동영상으로, 유튜브 쇼츠, 틱톡, 인스타그램 릴스, 페이스북 와치 등 다양한 플랫폼에서 소비되고 있다. 기존의 TV 프로그램이나 긴 동영상과는 달리, 숏폼은 단시간에 다양한 정보들을 포함하고 스마트폰에서 간편하게 소비할 수 있다는 장점 덕분에 더욱 성장하고 있다. *Figure* 6은 전 세계 틱톡 월간 사용자 수를 나타낸 그래프이다. 2018년 1월 약 5500만 명을 시작으로 2021년 9월에 약 10억 명의 월간 사용자 수를 달성하였고, 그 이후로도 꾸준히 증가하여 2024년 5월에는 약 10억 4000만 명이 집계되었다[11].

스마트폰이 대중화되면서 전 연령대가 언제 어디서나 고화질 영상을 손쉽게 접할 수 있게 되었고, 특히 짧고 간결한 콘텐츠를 선호하는 Z세대의 트렌드가 숏폼 콘텐츠의 확산을 더욱 가속화시켰다. 또한 소셜 미디어와 결합하여 확산성이 뛰어나 바이럴 콘텐츠로 발전하기 쉬운 구조를 가지고 있어, 패션, 음악, 요리 등 다양한 분야에서 마케팅 도구로 활용되고 있다. *Table* 2는 숏폼 콘텐츠의 하루 이용 시간과 빈도를 나타내었다. 하루 1시간 이상 이용하는 사용자는 약 12%이고, 하루에 10회 이상 접속하는 사용자는 약

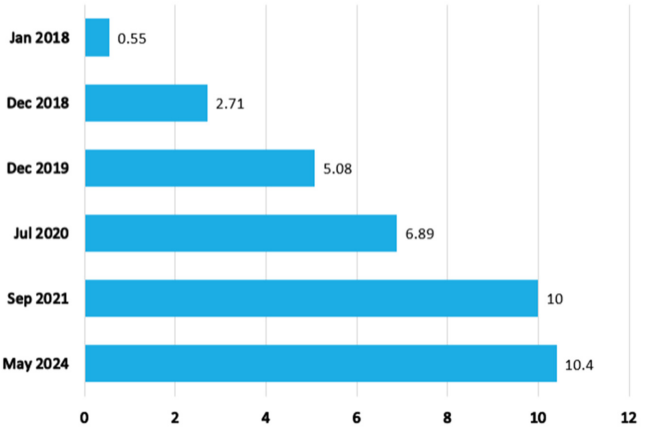

Figure 6. Global monthly tiktok users (hundreds of millions).

Table 2. Daily usage time and frequency of short-form content

Characteristic	Category	Frequency	Percentage (%)
Short-form content-daily usage time	Less than 0.5 hour	65	51.2
	0.5 to 1 hour	46	36.2
	1 to 2 hours	10	7.9
	2 to 3 hours	3	2.4
	More than 3 hours	3	2.4
	Total	127	100
Short-form content-daily usage frequency	Less than 5 times	49	38.6
	5 to 10 times	51	40.2
	10 to 20 times	22	17.3
	20 to 30 times	2	1.6
	More than 30 times	3	2.4
	Total	127	100

Figure 7. Lookbook content.

21%로 많은 시간을 숏폼에 소비하는 것으로 보인다[12].

특히 패션 분야에서는 *Figure* 7과 같은 룩북(Lookbook)이라는 주제가 많은 인기를 끌고 있다. 룩북은 다양한 패션 아이템을 스타일링 하여 보여주는 영상 형식으로, 사용자들이 옷을 입고 스타일을 소개하는 콘텐츠이다. 최근에는 전문 모델뿐 아니라 일반인들도 손쉽게 룩북 영상을 제작하여 공유하는 경향이 증가하고 있다. 이러한 숏폼 룩북 콘텐츠는 패션 트렌드를 보여주는 동시에 사용자가 원하는 스타일을 간접적으로 체험할 수 있게 하여, 소비자들에게 즐거움을 주는 새로운 쇼핑 경험으로 자리잡고 있다.

1.2. 연구 목적

온라인 쇼핑몰의 성장으로 패션 산업에서도 온라인으로 의류를 구매하는 소비자들이 많아졌다. 그러나 빠른 성장만큼 여러 가지 한계점도 확인되었다. 실제로 입어보지 못하는, 즉 시착 불가능의 단점이 있다. 따라서 소비자의 체형에 적합한 모델에 소비자 얼굴을 자연스럽게 합성하는 가상 피팅 프로그램을 개발하는 것을 목표로 한다. 이를 개발하기 위해 딥페이크와 옷 합성을 진행하며, 이는 소비자가 선택한 옷이 소비자에게 잘 어울리는지 판단할 수 있는 지표가 될 것이다.

또한 이미지에 그치지 않고, 동영상으로 확장하여 소비자에게 더욱 자연스럽고 현실감있는 가상 피팅 서비스를 제공할 것이다. 이는 단순히 옷을 구매/판매하기 위해 사용되는 것이 아니라, 숏폼 형식으로 소비자들에게 제공하여 재미 및 흥미를 부여할 계획이다. 동영상 확장으로 인해 소비자들은 선택한 의상을 구매하는 데에 도움이 될 것이고, 이러한 양질의 콘텐츠로 인해 소비자들의 재방문율이 향상될 것이다.

2. 이론적 배경

2.1. 딥러닝(Deep Learning)

딥러닝은 머신러닝의 하위 분야로, 다층 인공 신경망(Artificial Neural Networks)을 활용해 데이터를 분석하고 학습하는 기술이다. 머신러닝의 경우 *Figure* 8과 같이 입력층, 출력층 밖에 존재하지 않지만, 딥러닝은 여러 은닉층이 추가로 구성되어 있다. 이는 추상적이고 고차원적인 특징을 학습할 수 있다는 장점이 있다. 따라서 딥러닝은 비선형 데이터를 처리할 수 있는 능력이 뛰어나 이미지 인식, 음성 인식, 자율주행 등의 분야에서 사용되고 있다.

2.2. 합성곱 신경망(Convolutional Neural Networks, CNN)

CNN은 주로 이미지를 처리하는 데 사용되는 딥러닝 모델이다. 합성곱(Convolution) 연산을 중심으로 동작하며, 이미지나 동영상과 같은 고차원 데이터에서 중요한 특징을 추출하고 분석하는 데 뛰어난 성능을 가진다.

CNN의 동작 원리는 *Figure* 9에 제시되어 있다. 입력 이미지가 Convolutional Layer를 통과하면서 핵심 특징을 추출한다. 이후, Pooling Layer를 통해 크기가 축소되고 중요한 특징만 남겨져 효율적인 연산이 이루어진다. 여러 Convolutional Layer와 Pooling Layer를 거쳐 마지막 과정인 Connected Layer에서는 추출된 특징을 바탕으로 분류 혹은 예측하게 된다.

2.3. 적대적 생성 신경망(Generative Adversarial Networks, GAN)

GAN은 생성기(Generator)와 판별기(Discriminator)라는 두 가지 신경망의 상호 경쟁을 통해 학습하는 구조를 가진

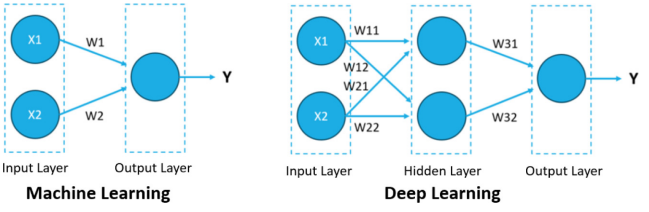

Figure 8. Principles of Machine Learning and Deep Learning.

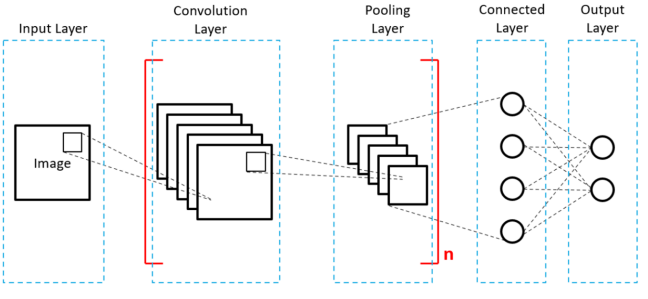

Figure 9. Principle of CNN.

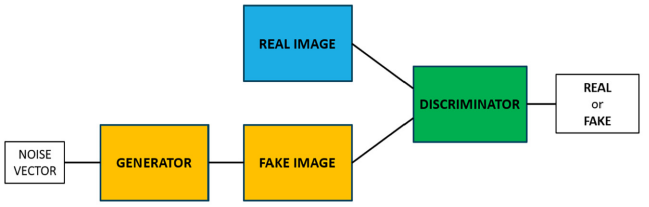

Figure 10. Principle of GAN.

딥러닝 모델이다. *Figure* 10에서 생성기는 노이즈 데이터를 입력받게 되고, 가짜 데이터를 생성해낸다. 생성된 가짜 데이터를 판별기에 입력하게 되면 판별기는 입력받은 데이터가 진짜 데이터일 확률을 계산하게 된다. 판별기에 의해 계산된 확률을 더 높이기 위해 생성기는 더 업그레이드하고, 판별기 또한 가짜 데이터를 더 잘 구별하며 적대적으로 학습하며 데이터를 생성하게 된다.

2.4. 오토엔코더(Autoencoder)

Autoencoder는 Encoder와 Decoder로 이루어져 있는 비지도 학습 방법이다. *Figure* 11을 보면 쉽게 알 수 있듯, Encoder란 입력받은 데이터를 압축하여 주요 특징만을 추출하며, Decoder는 Encoder 과정에서 추출된 데이터를 다시 본래의 형태로 복원시키는 신경망이다.

Autoencoder의 이러한 특징은 딥페이크에서 활용된다. *Figure* 12를 살펴보면, Face A를 Encoder 과정을 통해 A features가 추출된다. 이를 다시 Decoder 과정을 통해 Face A′으로 생성된다. 이때 사용된 Decoder 연산 과정들을 Decoder A라고 한다. Face B도 같은 과정을 거쳐 Decoder B가 정해지게 된다. 이 과정에서 Decoder B 대신 Decoder A의 연산을 진행하게 되면 B features가 Decoder A의 연산에 따라 생성되기 때문에 Swapped Face A의 얼굴이 그려진다.

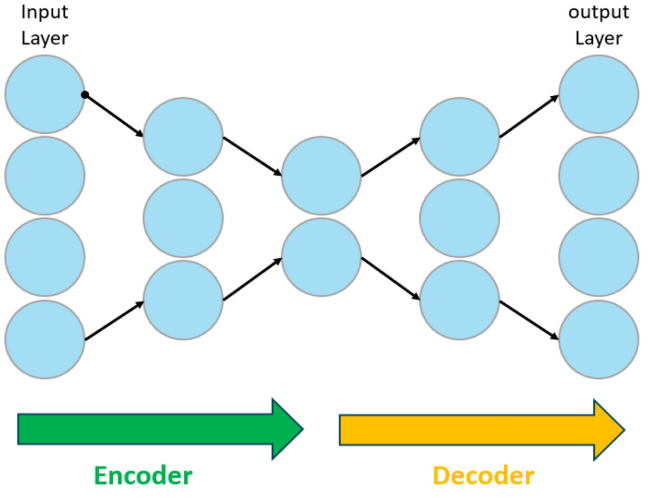

Figure 11. Principle of Autoencoder.

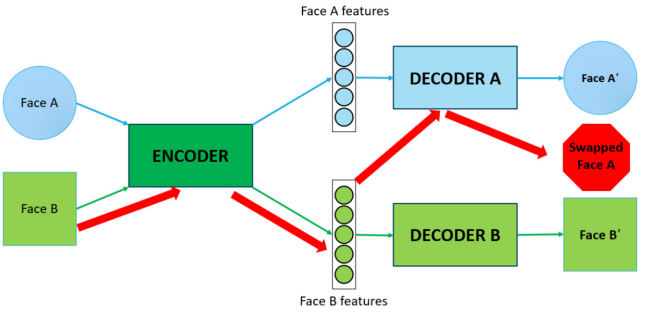

Figure 12. Principle of Deepfake using Autoencoder.

3. 연구 방법 및 설계

3.1. 연구 환경

본 연구는 Python 프로그래밍 언어를 기반으로 수행되었으며, 개발에 사용된 하드웨어와 소프트웨어 환경은 다음과 같다.

- CPU : Intel Core i9-11900K
- RAM : DDR4 32GB * 2
- GPU : NVIDIA GeForce RTX 4070 SUPER
- OS : Windows 10

3.2. 연구 방법

본 연구는 *Figure* 13과 같은 과정을 따른다. 소비자는 두 가지 선택을 할 수 있다. 첫 번째로 옷 합성과 딥페이크만 실시한 이미지 형태의 결과물이다. 이는 소비자가 동영상을 원하지 않고, 간단히 옷의 스타일이 자신에게 어울리는지를 판단할 때 주로 사용된다. 하지만 동영상보다는 덜 자연스럽고, 확실히 판단하기 힘들다는 단점이 존재한다. 두 번째로는 모션 합성 과정을 통해 결과물을 생성하는 것이다. 이는 소비자가 가상 피팅 후에 움직임을 주어서 더욱 자연스러운 상황을 연출할 수 있고, 옷의 스타일이 자신과 잘 어울리는지 판단하는 데 적합하다. 하지만 동영상 작업은 이미지 작업보다 확실히 오랜 시간이 소요된다는 단점이 존재한다.

첫 번째 과정은 소비자가 선택한 옷과 모델 이미지를 기반으로 정보를 추출하는 전처리 과정이다. 이는 뒤에 진행

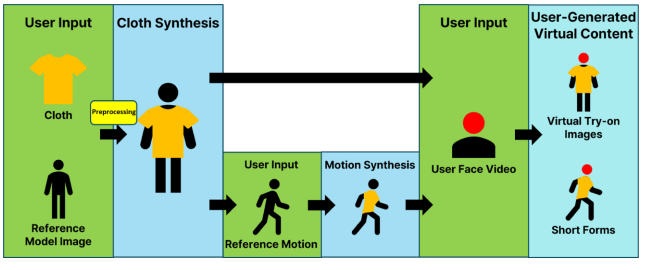

Figure 13. Process of virtual fitting service.

될 옷 합성 과정에 사용될 데이터를 준비하는 과정이다. 먼저 소비자가 선택한 옷에서 U2Net 모델을 사용하여 피사체와 옷을 분리하는 작업을 한다[13]. 이 작업은 옷 합성에 필요한 옷만을 추출하기 위해 진행되는 과정이다. 다음으로 소비자가 선택한 모델 이미지에서 OpenPose, DensePose, Parsing Image를 추출한다[14,15]. OpenPose는 모델의 골격, 자세, 손 모양 등을 인식하고 추적하여 keypoints로 좌표를 추출할 수 있는 오픈소스 라이브러리다. DensePose는 모델의 2D 이미지를 3D 모델로 매핑하여 자세를 추정하는 기술이다. Parsing Image는 모델의 이미지를 머리, 상의, 하의, 왼손, 오른손 등 20가지의 영역으로 분류하는 과정이다. Parsing Image에서 분류된 영역으로 합성될 영역을 인식하는 작업이다. OpenPose, DensePose, Parsing Image 과정은 모두 합성될 옷을 변형시키기 위해 거쳐야 하는 전처리 과정이다.

두 번째 과정은 옷 합성 과정이다. 옷 합성 과정에는 StableVITON의 오픈소스를 활용하였다[16]. 기존의 StableVITON의 훈련모델은 여성 의류 중심의 학습 모델로, 실생활에서 자주 착용하는 셔츠나 후드티 등 남녀 공용 캐주얼 의류에 대한 학습이 누락되어 있었다. 이러한 기존 모델을 그대로 사용하여 실제 쇼핑몰에 적용할 경우 문제점이 크게 부각될 수 있다. 따라서 기존 모델을 그대로 사용하는 대신, 누락된 의류에 대해 추가적으로 학습을 하는 파인튜닝 과정을 진행하였다. 이를 통해 다양한 의류에 적용 가능한 실용적인 합성 결과를 도출하고자 하였다.

Figure 14와 같이 선별된 데이터로 추가적인 모델 훈련을 마친 후에 옷 합성을 진행한다. 첫 번째 과정에서 진행했던 전처리 과정의 추출물을 활용한다. StableVITON은 옷의 이미지를 모델의 자세에 맞게 변형시키고 CNN을 활용해 자연스러운 이미지를 생성한다. 이후 GAN 기반으로 옷과 모델을 합성하며 매끄러운 합성을 가능하게 한다.

세 번째 과정은 모션 합성 과정이다. 모션 합성 과정에서는 MusePose의 오픈소스를 활용하였다[17]. MusePose는 모델의 자세와 모션을 자연스럽게 합성하기 위한 오픈소스로 모델의 자세를 분석하고 입력받은 모션을 분석하여 실제 사람의 모션처럼 생성하는 작업을 한다. 첫 번째 과정에서 생성된 모델 이미지의 OpenPose 데이터를 다시 사용하고, 입력받은 모션의 OpenPose 데이터를 사용한다. 모델 이미지와 모션 동영상의 관절 Keypoints 좌표를 기반으로 모션 합성을 진행한다. 이때 Autoencoder, U2Net 등의 모델을 사용했다.

네 번째 과정은 딥페이크 과정이다. 소비자의 선택에 따라 이미지 기반 딥페이크 합성과, 동영상 기반 딥페이크 합성 과정으로 구분된다. 딥페이크를 하는 과정에서는 DeepFaceLab의 오픈소스를 활용하였다[18]. DeepFaceLab은 CNN을 활용하여 얼굴의 주요 특징을 추출하고, 두 얼굴 사이의 변형을 학습한다. 또한 얼굴의 윤곽을 정확하게 정렬하고 합성에 활용 가능하도록 도와준다. *Figure* 15는 Autoencoder 모델을 활용하여 얼굴을 합성하는 DeepFaceLab의 딥페이크 과정이다.

본 개발에는 위와 같이 다양한 딥러닝 기술이 활용되었다. StableVITON에서는 CNN 기반의 U-Net 구조를 통해 옷 합성 과정을 진행하며, MusePose는 GAN 기반으로 모델의 모션을 자연스럽게 생성한다. 또한 DeepFaceLab은 Autoencoder를 활용하여 얼굴 합성을 진행한다. 각기 다른 기술이 통합적으로 작동함으로써 개발을 수행하였다.

위와 같이 제안된 가상 피팅 숏폼 영상의 시각적 품질 및 소비자 인식을 정량적으로 평가하기 위해 사용자 설문 조사를 실시하였다. 설문은 의상 3종과 모션 2종의 조합으로 된 총 6개의 짧은 영상으로 준비하였다. Clip1과 Clip2는 분홍색 맨투맨을 착용하고 각각 손 허리 포즈와 상의 회전 포즈, Clip3과 Clip4는 초록색 반팔 티셔츠를 각각 손 허리 포즈와 상의 회전 포즈, Clip5와 Clip6은 회색 맨투맨을 각각 손 허리 포즈와 상의 회전 포즈를 취했다. 각각의 Clip을 시청 후 얼굴 합성의 자연스러움, 의류의 착용감 전달력, 모션의 자연스러움, 전체적인 몰입감, 실제 쇼핑에서의 활용 가능성, 총 5가지 항목에 대해 5점 리커트 척도로 평가하게 되었다. 설문은 총 17명의 대학생을 대상으로 온라인으로 진행되었으며, 모든 응답은 익명으로 수집되었다. 또한 추가적인 자유 의견을 수렴하여 정성적 피드백도 함께 분석하였다.

Figure 14. Additional training data.

Figure 15. DeepFaceLab using Autoencoder.

4. 결과 및 고찰

4.1. 옷 합성 결과

옷 합성 과정에서는 StableVITON을 활용하여 소비자가 선택한 옷을 소비자 체형에 가까운 모델에게 합성하는 과정을 수행했다. 다음 *Figure* 16은 소비자가 선택한 모델과 옷의 이미지이다.

Figure 16. Model and clothes selected by the consumer.

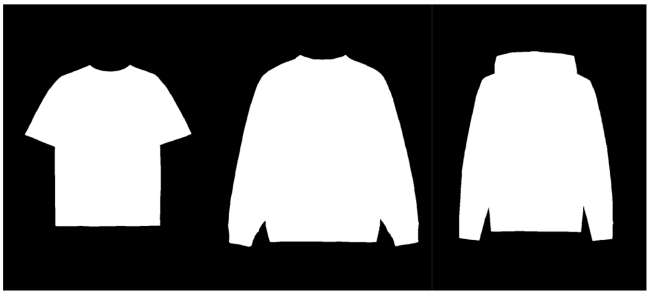

Figure 17. Cloth Segmentation.

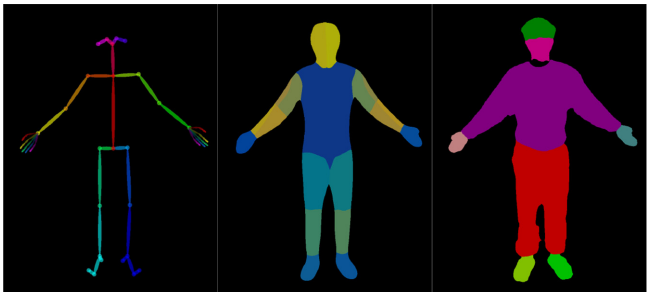

Figure 18. OpenPose, DensePose, Parsing Image.

Figure 19. Clothing synthesis results.

옷의 이미지를 *Figure* 17과 같이 U2Net 모델을 사용한 코드로 피사체와 배경을 분리하는 Cloth Segmentation 과정을 진행하였다.

그 후 모델의 이미지 OpenPose, DensePose, Parsing Image 데이터를 *Figure* 18과 같이 추출하였고, *Figure* 19는 옷 합성의 최종 결과물이다.

4.2. 모션 합성 결과

모션 합성 과정에서는 MusePose를 활용하여 옷 합성된 모델에게 모션을 합성하는 과정을 수행했다. *Figure* 20은 미리 준비되어있는 모션에 OpenPose를 사용하여 좌표를 추출한 동영상 캡쳐 이미지이고, 옷 합성이 완료된 모델에게 모션을 합성한 결과물은 *Figure* 21에서 확인할 수 있다.

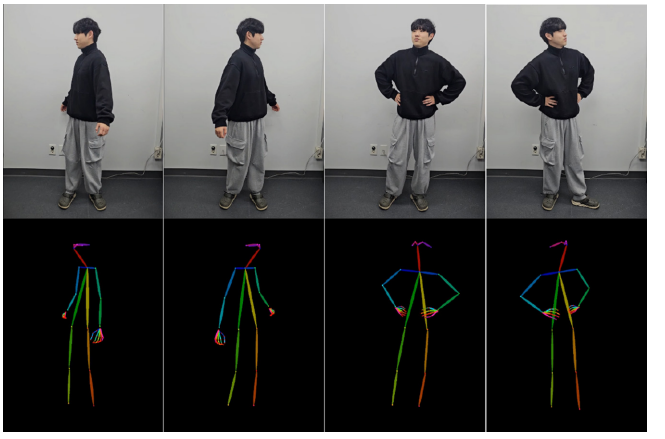

Figure 20. Upper: motion video, Lower: video after OpenPose processing.

Figure 21. Motion synthesis results.

Figure 22. Consumer's face video.

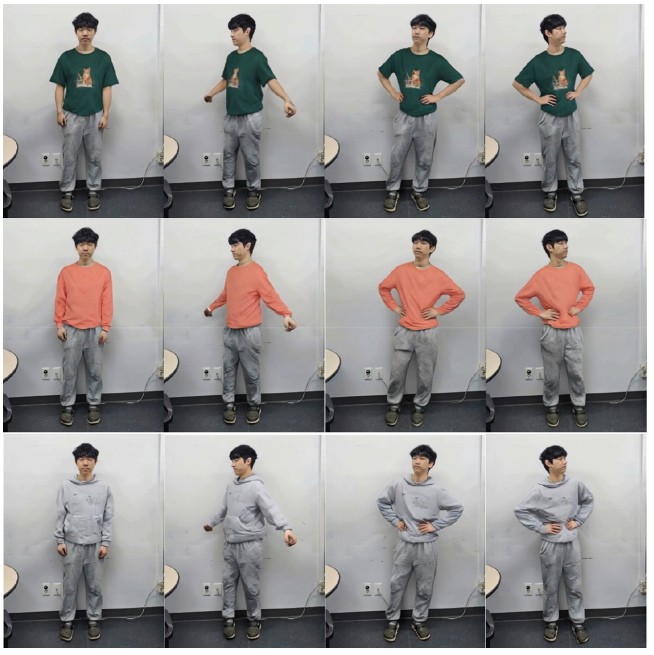

Figure 23. Virtual fitting result video using Deepfake.

4.3. 딥페이크 합성 결과

소비자는 *Figure* 22와 같이 5~8초의 얼굴 동영상을 제공한다.

이후 소비자의 얼굴을 모션 합성이 완료된 모델 동영상에 딥페이크를 실시한다. *Figure* 23은 DeepFaceLab을 활용하여 소비자의 얼굴과 모델 이미지를 자연스럽게 합성한 가상 피팅 동영상 결과물의 캡쳐 이미지이다.

Figure 24. Virtual fitting images using Deepfake.

또한 동영상을 원하지 않아 모션 합성을 선택하지 않은 소비자의 경우에도 *Figure* 22와 같이 얼굴 동영상을 제공하고, DeepFaceLab을 활용해 딥페이크를 실시하여 *Figure* 24와 같은 결과물을 생성했다.

4.4. 설문 조사 결과

설문 결과, 대부분의 항목에서 평균 4점 이상의 긍정적인 평가가 나타났다. 특히 얼굴 합성의 자연스러움과 의류 착용감 표현 항목에서는 4점 이상의 응답이 다수를 차지하였으며, 모션의 자연스러움은 일부 클립에서 비교적 낮은 점수가 기록되었다.

전체적으로 가장 긍정적인 반응을 보인 클립은 Clip3(평균 4.1점 이상)이며, Clip6는 모션과 의류 왜곡 측면에서 비교적 낮은 평가를 받았다. 이는 자유 의견 항목에서도 드러났으며, 일부 설문 응답자는 다음과 같은 피드백을 제공하였다.

- 측면 동작에서 옷의 구김이 명확히 반영되지 않았다.
- 모션이 너무 정적이고 사이언스틱해서 몰입감이 떨어졌다.
- 좌우를 돌 때 의상이 깨지는 듯한 현상이 있었다.

이러한 피드백은 향후 모션 다양성 확대, 옷 주름 처리 개선, 반팔티에서의 팔 노출 표현 보완 등의 개발 방향 설정에 참고할 수 있는 유의미한 자료로 활용된다.

5. 결 론

본 개발은 기존 온라인 의류 쇼핑몰에서 시착 불가능이라는 문제점을 해결하기 위해 진행되었다. 기존의 온라인

Table 3. Mean scores of user evaluations per video clip

Evaluation criteria	Clip1	Clip2	Clip3	Clip4	Clip5	Clip6
Naturalness of the synthesis	3.8	3.8	4.1	3.9	4.1	3.9
Realism of clothing fit	3.9	4.1	4.1	3.9	3.9	3.8
Naturalness of motion	3.8	3.9	4.1	3.9	4.0	3.8
Overall immersion	3.9	3.9	4.2	3.8	3.9	3.7
Practicality for shopping use	4.0	4.2	4.1	4.0	4.2	3.9

의류 쇼핑몰의 경우 소비자가 실제로 옷을 착용해 볼 수 없어 오프라인 매장을 찾아가 착용하고 구매하거나, Bracketing 쇼핑 방식을 사용하여 온라인 의류 쇼핑몰에 피해를 주는 등의 문제가 있었다. 이를 해결하기 위해 일부 플랫폼에서는 가상 피팅 서비스를 제공하였으나, 아바타 기반으로 모델링하여 현실감이 부족하였고, 의상의 수를 제한하여 선택의 폭이 제한적이어서 소비자들이 자신에게 잘 어울리는 옷을 선택하는 데 어려움을 겪었다.

따라서 딥페이크 기술을 활용하여 소비자의 얼굴을 모델에게 자연스럽게 합성하는 가상 피팅 프로그램을 개발하고자 하였다. 이를 통해 소비자들은 자신의 얼굴이 합성된 모델에게 옷을 착용시켜서 잘 어울리는지 판단할 수 있게 된다. 이는 Bracketing의 원인 중 하나인 시각적 정보 부족 현상을 해소함으로써, 온라인 쇼핑의 만족도를 높이고 반품률을 줄이는 데 기여할 수 있다. 또한 특정 온라인 쇼핑몰에 한정되지 않고, 출처가 없는 다양한 의상 이미지도 합성을 가능하게 하여 의상 선택의 제한을 극복하였다. 기존의 이미지 형태로 제공되던 가상 피팅 결과물 또한 동영상 형식으로 확장하여 더욱 자연스럽고 현실적인 가상 피팅 서비스를 제공할 수 있게 되었다.

현재 본 개발은 기술적 가능성을 확인하고 시범적인 프로그램 구현을 목표로 진행되었다. 따라서 몇 가지의 개선사항이 발견되었다. 첫 번째로 현재 합성 과정에서 시간이 많이 소요된다는 문제가 존재한다. 현재 10초 분량의 가상 피팅 동영상을 제작하는 데에 5시간 이상의 시간이 소요되며, 모션 합성이 없는 이미지 형태의 경우에도 약 30분의 시간이 소요된다. 이는 주로 복잡한 모션 합성과 딥페이크 과정에서의 문제점으로, 이를 해결하기 위해 효율적인 알고리즘을 통한 코드 최적화와 장비 개선을 통해 합성 시간을 단축할 계획이다. 두 번째로 정확하지 않은 수치의 문제점이다. 키와 몸무게가 같은 사람일지라도 구체적인 신체 수치는 다르기 때문에 소비자와 모델의 차이점이 나타난다. 따라서 보다 정확한 가상 피팅 시스템 구축을 위해 소비자의 신체 정보를 기반으로 한 2D 모델링 기술을 도입하여, 실제 신체와 더욱 정확하게 일치하는 모델을 생성하는 기술을 개발할 계획이다. 이를 통해, 가상 피팅의 정확도를 한층 더 높일 수 있을 것이다. 마지막으로 딥페이크 설정에 개선이 필요하다. 딥페이크는 얼굴, 즉 이목구비만을 합성하는 프로그램이다. 따라서 소비자의 헤어스타일은 합성되지 않아 어색한 모습을 보인다. 이를 해결하기 위해 딥페이크를 실시하는 과정에서 얼굴에 제한하지 않고 머리 전체를 학습하고 헤어스타일까지 모두 합성을 진행하게 할 예정이다.

다음으로는 앞으로의 개발 계획이다. 현재 프로그램은 x축과 y축으로만 이루어진 2D 이미지, 동영상에 불과하다. 미래를 지향하고자 한다면 3D 영역으로의 확장은 필수가 되었다. 따라서 2D 결과물에서 벗어나 CLO, DAZ와 같은 3D 모델링 프로그램을 활용하여 3D 영역으로 확장을 계획 중에 있다. 향후 3D 가상 피팅 시스템 개발은 소비자에게 더욱 현실감 있는 시착 경험과 정밀한 가상 피팅 시뮬레이션을 제공할 수 있을 것으로 기대된다.

감사의 글: 이 논문은 2023학년도 건국대학교의 연구년 교원 지원에 의하여 연구되었음.

이 논문은 2023년도 정부(산업통상자원부)의 재원으로 한국산업기술진흥원의 지원을 받아 수행된 연구임(P0012770, 2020년 산업혁신인재성장지원사업).

이 논문은 2024년도 정부(산업통상자원부)의 재원으로 한국산업기술진흥원의 지원을 받아 수행된 연구임(RS-2024-00410875, 2024년 산업혁신인재성장지원사업).

References

1. C. Shin and H. Cho, “Time-series Analysis on the Impact of COVID-19 on Online Shopping Purchase”, *E-Trade Review*, 2022, **20**, 97–109.
2. KOSIS, https://kosis.kr/statHtml/statHtml.do?orgId=203&tblId=DT_20308_23_02_101&conn_path=I2 (Accessed January 15, 2025).
3. KOSIS, https://kosis.kr/statHtml/statHtml.do?orgId=203&tblId=DT_20308_23_02_105&conn_path=I2 (Accessed January 15, 2025).
4. J. Ko, “Changes in the Clothing Consumption Behavior After Pandemics”, Master’s Thesis, Yonsei University, Seoul, 2021.
5. KOSIS, https://kosis.kr/statHtml/statHtml.do?orgId=203&tblId=DT_20308_23_02_103&conn_path=I2 (Accessed January 15, 2025).
6. Raydiant, “The State of Consumer Behavior 2021”, Raydiant, 2021.
7. Narvar, “Making Returns a Competitive Advantage”, Narvar, 2017.
8. Narvar, “The State of Returns: Finding What Fits”, Narvar, 2021.
9. I. Jung, J.-M. Lee, and K. Hwang, “Implementation of Secondhand Clothing Trading System with Deep Learning-Based Virtual Fitting Functionality”, *JIIBC*, 2024, **1**, 17–22.
10. S.-H. Park, J.-Y. Ro, S.-Y. Song, S.-H. Shin, and K. Kim, “A Study on Virtual Fitting Service Using GAN”, *KIPS*, 2019, **10a**, 976–979.
11. Backlinko Team, “TikTok Statistics You Need to Know”, BACKLINKO, 2024.
12. S. H. Kim, “A Study on the Motives and Satisfaction of Short Form Content Platform”, Chung-Ang University, Seoul, 2022.
13. X. Qiu, J. Wang, and Y. Li, “U2-Net: Going Deeper with Nested U-Structure for Salient Object Detection”, GitHub Repository, https://github.com/xuebinqin/U-2-Net (Accessed January 15, 2025).

14. Z. Cao, T. Simon, S. Wei, and Y. Sheikh, "OpenPose: Realtime Multi-Person 2D Pose Estimation Using Part Affinity Fields", GitHub Repository, https://github.com/CMU-Perceptual-Computing-Lab/openpose (Accessed January 15, 2025).
15. R. G. Girshick, S. Singh, Y. Farabet, D. R. Johnson, and D. Dolgov, "DensePose: Dense Human Pose Estimation in the Wild", GitHub Repository, https://github.com/facebookresearch/DPT (Accessed January 15, 2025).
16. S. Han, H. Yang, J. Zhang, and X. Wang, "StableVITON: Virtual Try-On Network for Fashion", GitHub Repository, https://github.com/fiendbunny/StableVITON (Accessed January 15, 2025).
17. H. Liao, J. Li, and Q. Zhou, "MusePose: Motion Synthesis for Pose Transfer in Human Pose Estimation", GitHub Repository, https://github.com/wyu-mmm/MusePose (Accessed January 15, 2025).
18. F. Kovalev, D. Iakovlev, and I. V. Melikhov, "DeepFaceLab: Deep Learning Face Swap Framework", GitHub Repository, https://github.com/iperov/DeepFaceLab (Accessed January 15, 2025).