

# Virtual Try-On 기반 유사도 실험을 통한 데이터 분석 연구

김선우<sup>1</sup>

<sup>1</sup> 고려대학교 SW·AI 융합대학원 석사과정

kimsw39@korea.ac.kr

## A Data Analysis Study Based on Similarity Experiments of Virtual Try-On

Sun-woo Kim.

<sup>1</sup>Dept. of Big Data Convergence, Korea University

### 요 약

Virtual Try-On(VTO) 기술은 증강현실(AR)과 인공지능(AI)을 기반으로 사용자가 실제로 옷을 입지 않고도 가상의 피팅을 경험할 수 있도록 하는 혁신적인 기술이다. 그러나 현재 상용화된 VTO 시스템은 이미지 왜곡, 색상 손실, 텍스처 변형 등의 문제로 인해 실제 제품과의 시각적 불일치를 자주 유발하며, 이로 인해 사용자 신뢰도와 만족도 저하의 주요 원인이 되고 있다. 본 연구는 이러한 문제를 정량적으로 분석하고 기술적 개선 방안을 제시하는 것을 목표로 한다. 이를 위해 실제 착용 이미지와 AI로 생성된 가상 피팅 이미지 간의 유사도를 구조적 유사도 지수(SSIM), 신호대잡음비(PSNR), 딥러닝 기반 VGG16-Cosine 유사도 지표를 활용하여 비교·분석하였다. 특히, 상용 모델인 WeshopAI를 활용한 실험에서는 50% 이하의 유사도를 보이는 사례가 다수 확인되었으며, 이는 생성된 이미지가 실제와 시각적으로 상당한 차이를 보인다는 점을 의미한다. 이러한 품질 저하는 AI 생성 알고리즘의 한계, 원본 입력 데이터의 왜곡, 해상도 저하, 색상 정확도 부족 등 복합적인 요소에서 기인하는 것으로 판단된다. 예컨대 의류의 질감이나 주름, 복잡한 패턴이 정확하게 반영되지 않고, 색상이 실제 제품과 달리 과장되거나 흐릿하게 표현되는 경우가 많았다. 이로 인해 소비자는 제품에 대한 오해를 가질 수 있으며, 이는 구매 전환율에도 부정적 영향을 미칠 수 있다. 따라서 본 논문의 기술적 접근은 VTO 기술의 실용성과 완성도를 동시에 제고하며, 궁극적으로는 온라인 사용자에게 보다 정확하고 신뢰할 수 있는 가상 서비스 환경을 제공할 수 있을 것이다.

### 1. 서론

Virtual Try-On(VTO)기술은 최근 온라인 쇼핑 환경에서 급격히 주목받고 있는 첨단 기술 중 하나로, 증강현실(AR)과 인공지능(AI), 3D 모델링 기술을 활용하여 사용자가 실제 옷을 입지 않고도 의류, 액세서리, 화장품 등을 가상으로 착용할 수 있게 해주는 혁신적인 솔루션이다[1][2]. 온라인 쇼핑 시장이 지속적으로 성장함에 따라, 소비자들 사이에서 실제 매장을 방문하지 않고도 다양한 제품을 체험해보고 구매할 수 있는 기술적 요구가 증가하고 있으며, 이에 따라 VTO 기술 역시 빠르게 발전하고 있다[3].

그러나 기술 발전에도 불구하고 Virtual Try-

On(VTO) 서비스의 현실성과 정확성에 있어 여전히 많은 한계점과 문제점들이 존재한다. 대표적으로, 실제 의상과 가상 환경에서 생성된 의상의 시각적 품질 차이, 옷의 형태 및 텍스처 변형, 인물의 포즈에 따른 부정확한 매핑 등이 지속적으로 문제로 제기되고 있다[4]. 특히, AI 기반의 데이터 생성 과정에서 발생하는 왜곡과 생성 오류는 서비스의 품질을 저하시켜 소비자의 신뢰도를 떨어뜨리고, 결과적으로 기술의 실용성에도 부정적 영향을 미치게 된다[5].

본 연구는 이러한 문제점을 해결하고 Virtual Try-On(VTO) 서비스의 품질 향상을 도모하기 위하여, 실제 데이터와 AI가 생성한 데이터 간의 유사도 분석에 초점을 맞추었다. 본 논문의 목적은 실제 사용자

가 기대하는 품질 수준을 충족하기 위한 필수 조건으로, 실제 사진과 AI 로 생성된 가상 이미지를 비교할 때 최소한 50% 이상의 유사도를 유지해야 한다는 가정하에 정량적인 실험을 수행하고 결과를 분석하는데 있다.

## 2. 관련연구

최근 Virtual Try-On(VTO)기술은 생성형 인공지능의 발전과 함께 빠르게 진화하고 있다. 특히 멀티모달(Multimodal) 기반의 생성형 AI 기술이 활발히 연구되며, 이미지뿐 아니라 텍스트 및 비디오 데이터를 융합하여 더욱 현실감 있고 정확한 가상 착용 이미지를 생성하는 방향으로 발전하고 있다. 최근 이미지 기반의 다양한 생성형 AI 모델들이 개발되면서 Virtual Try-On(VTO) 기술에서 발생하는 의류의 주름, 변형, 질감 손실 등의 문제를 해결하기 위한 다양한 접근 방식이 연구되고 있다. 최근 연구된 기법 중 하나인 기하학적 워핑(Geometric Warping)과 추가적인 피팅(Fitting) 단계 도입(Lee et al., 2022)은 의류 텍스처와 인체의 세부 특징을 더욱 정확히 유지하는데 효과적임을 보였다. 이를 통해 주름과 같은 세부적인 특징의 보존과 착용 현실감이 크게 향상되었다.

더불어 StyleGAN 기반의 글로벌 어피어런스 플로우(Global Appearance Flow) 추정 기법(He et al., 2022)이 주목받고 있다. 이 기술은 의류의 왜곡 문제를 최소화하고, 사용자의 신체 포즈와 더욱 정확하게 맞춘 자연스러운 의류 착용 이미지를 생성하는 데 탁월한 성과를 보여주었다. 특히 GAN(Generative Adversarial Network)을 활용한 이미지 합성 기술의 발전은 실제 촬영 이미지와 거의 구분하기 어려운 수준의 가상 피팅 결과물을 제공하고 있다.

색상 관리 기술 역시 Virtual Try-On(VTO) 서비스 품질 향상을 위한 중요한 연구 분야이다. 증강현실(AR)의 색상 조화 기술(Marino et al., 2021)을 VTO에 적용함으로써 가상 의류의 색상을 실제 사용자의 환경과 자연스럽게 융합시키는 기술이 개발되고 있다. 더불어 하이퍼스펙트럴 이미지(Hyperspectral Imaging)를 활용한 스펙트럼 기반 색상 관리 기법(Díaz-Barrancas et al., 2020)을 도입하여 색상의 충실도를 더욱 높이고 현실감 있는 결과물을 생성하는 연구가 활발히 진행 중이다.

## 3. 유사도 비교 시스템 실험 분석

Virtual Try-On(VTO)에서 왜곡 현상의 주된 원인은 입력 데이터 처리의 어려움 때문이다. 특히 의류의 비선형적 변형, 주름, 늘어짐, 복잡한 레이어 구조 등의 데이터 처리가 가장 까다롭다. 인체의 다양한 포즈와 실루엣 변화, 예를 들어 측면 포즈나 팔을 벌린 자세 등도 의류 매핑을 어렵게 만든다. 또한 의류의 텍스처나 패턴 보존이 힘들고, 머리카락이나 팔 등에

의한 가림(occlusion) 처리 역시 왜곡을 유발한다(Ahn, 2018). 이러한 문제들은 보다 나은 서비스를 제공하기 위해 고난도의 처리 기술을 요구한다. 실제로 남자와 여자를 구분하기 어려우며, 계절별 옷의 구분도 쉽지 않다. 이러한 차이를 효과적으로 분석하기 위하여, 본 논문에서는 구조적 유사도와 픽셀을 이용한 품질 유사도, 딥러닝 모델을 이용한 유사도 등 세 가지 유사도 측정 방법을 이용하여 원본과 AI 생성 이미지 간의 유사도를 비교 분석하였다.

먼저, SSIM(Structural Similarity Index Measure) 유사도는 이미지의 구조적 유사도를 측정하는 지표이다. 인간의 시각 시스템이 이미지의 구조적 정보에 민감하다는 점을 착안해 개발되었다(Wang et al., 2004). 유사도의 범위는 0%(완전히 다름) ~ 100%(동일함)이며, 밝기, 대비, 구조의 변화를 고려하여 계산한다. SSIM 유사도는 이미지 압축, 화질 평가 등에 주로 사용되고 있다.

다음은 원본 이미지와 처리된 이미지 간의 픽셀 차이를 기반으로 한 품질 측정 지표인 PSNR 이다. PSNR(Peak Signal-to-Noise Ratio)은 원본 이미지와 처리된 이미지 간의 픽셀 차이를 기반으로 한 품질 측정 지표이다. PSNR 은 두 영상 간의 노이즈 영향, 왜곡 정도, 압축 손실을 수치로 비교할 수 있다. 그리고 계산이 빠르고 간단하여 대규모 실험 혹은 모델 학습 중간 평가에 매우 효율적이다.

마지막으로 딥러닝 모델인 VGG16(Simonyan & Zisserman, 2015)을 활용하는 방식이다. VGG16은 '2014년 Oxford Visual Geometry Group(VGG)'에서 발표한 대표적인 합성곱 신경망(CNN) 모델로, 이미지 특징(feature)을 정교하게 추출한다 이것은 이미지의 고차원 특징을 추출하고, 이 특징 벡터 간의 Cosine 유사도를 계산하여 유사도를 측정하는 방식이다. 이 방식은 CNN 기반의 사전 학습 모델이 인식하는 이미지 특징의 유사성 측정을 통해 색상, 질감, 물체 등 이미지의 의미적 내용을 비교할 수 있게 한다. SSIM 과 PSNR 와 달리 사람이 느끼는 시각적 유사도를 반영할 수 있다. CNN 으로 추출한 high-level feature 를 기반으로 하기 때문에 의미적(semantic) 유사성 분석 가능하다.

본 연구에서는 AI 생성 이미지와 실제 이미지 간의 유사도를 정량적으로 평가하고 시각화하는 종합 분석 시스템을 개발했다. 이 시스템은 다양한 이미지 유사도 평가 방법을 통합하고, 결과를 직관적으로 분석할 수 있는 대시보드 기능을 제공한다. 실험 결과는 SQLite 데이터베이스에 저장되며, 각 결과에는 다음 정보가 포함된다:

- 실험 ID 및 타임스탬프
- 실제 이미지와 AI 이미지 경로
- SSIM, PSNR, VGG 점수 및 평균 점수
- 결과에 대한 연구자 코멘트

분석 프레임워크는 다음 세 가지 주요 기능을 제공한다.

- 1) 유사도 비교: 실제 이미지와 AI 생성 이미지를 업로드하고 유사도 계산

- 2)결과 조회: 실험 결과를 필터링 및 검토  
3)결과 요약: 실험의 통계적 요약과 시각화 제공

#### 4. 데이터분석 및 유사도 실험 결과

본 연구에서는 상용 VTO 플랫폼인 WeshopAI 를 사용하여 실험을 진행하였다. 앞서 소개한 세가지 지표를 가지고 백분율을 기준으로 50%를 기준으로 테스트하였으며, 얼굴과 다리로 인한 오차율과 바지(하의 옷)는 동일한 옷을 입고 실제 사진을 찍었으며, 촬영된 옷의 구김은 고려하지 않기로 하였다. WeshopAI 모델은 구김에 대해서는 기술적으로 보완이 많이 된 것으로 파악되었다. 구김의 정도 차이는 대략 5%의 차이를 보여 오차범위로 가정하였다.

실험 결과, VGG16-Cosine 유사도는 인간의 시각적 인식과 유사한 경향을 보이며, 다른 유사도 지표보다 상대적으로 높은 일관성을 나타냈다. 일반적으로 SSIM 과 PSNR 은 이미지의 저수준 특성에 민감하며, VGG 점수는 의미론적 유사성을 더 잘 포착한다.

품질과 유사도의 관계는 높은 유사도 점수가 반드시 높은 시각적 품질을 의미하지 않을 수 있다. 이는 정량적 측정과 인간의 주관적 평가 사이의 간극을 보여준다.

분석	SSIM	PSNR	VGG16	평균
평균유사도	27.98	56.51	55.07	46.52
최대유사도	40.37	57.84	73.32	57.18
최소유사도	14.41	55.82	26.63	35.99

(표 1) 유사도 통계요약

실험 결과 중, 첫번째 특이점은 연관성 없는 객체의 생성이다.

유형	백분율(%)
SSIM	25.52%
PSNR	55.82%
VGG16	26.63%



(그림 1) 유사도 가장 낮은 실험결과

(그림 1)를 보면 오른손에 없는 시계가 생성되고 옷의 색감이 다소 손실되는 등 육안으로 확인해도 차이가 두드러진다. 유사도 측정결과 값도 전체적으로 낮

았으며, VGG16-Cosine 유사도값은 26.63%로 실험결과 중 가장 낮은 유사도 수치를 보였다.



(그림 2) 인물이미지와 옷이미지

인물데이터에서 신체와 옷을 명확히 구분하기 위해 하얀 색상의 옷을 입어 배경과 인물, 옷 구분을 명확히 하였다. 그러나 입혀지는 옷의 목 부분 디자인이 서로 다르므로 인해 AI 생성데이터의 의상의 목 부분이 원본과 다른 결과가 나왔다. 원본 옷 이미지데이터가 생성데이터에 변화를 주고 결과적으로 의상의 변형이 일어난다. 아래이미지의 경우 지퍼가 단추로 변경되어 생성되었다.

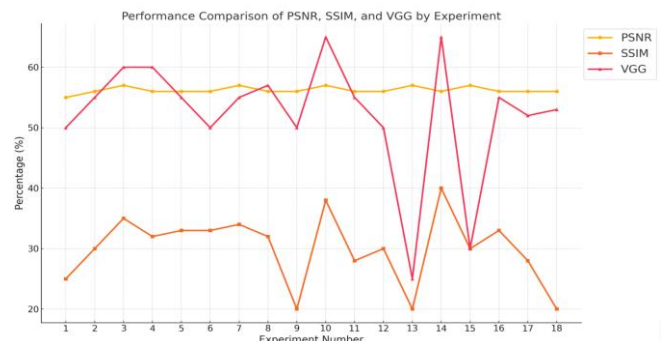
유형	백분율(%)
SSIM	22.34%
PSNR	56.06%
VGG16	59.86%



(그림 3) 객체 생성 오류 AI 생성데이터

두번째 문제점은 대체 대상인 의상을 제외한 다른 부분의 변형이 발생한다는 것이다. 현재 테스트에서는 인체의 부분을 최소화하고, WeshopAI 에 의하여 생성된 이미지는 GAN 네트워크의 특성상 그럴듯한 모습을 보이지만, 육안으로 확인해도 화질이나 구조 및 VGG16-Cosine 유사도에서는 거의 만족할만한 수준이 나오지 않는다.

Performance Comparison Of PSNR, SSIM, And VGG By Experiment



(그림 4) 유사도 추이 및 간격

실험을 통해 SSIM, PSNR, VGG16-cosin 유사도의 간격을 확인하고 그 차이를 확인하였다. 그 결과 색변조가 심하거나 옷의 디자인이 복잡할수록 VGG16-cosin 유사도는 낮았다. SSIM, PSNR, VGG16-cosin 세 가지 지표 모두에서 AI 기반 이미지가 실제 이미지 대비 명확한 품질 저하 현상을 보임을 나타냈다. 특히, 이미지 생성 과정에서 발생하는 색상 손실, 질감 왜곡 및 의상 형태의 변형 등 구체적인 품질 문제들이 관찰되었다. 유사도가 낮게 측정된 이미지일수록 이러한 문제점들이 육안으로도 명확히 식별 가능하였다. 이 결과는 VTO 서비스의 실질적 개선이 필요함을 시사한다.

본 연구에 사용된 소스 및 데이터는 GitHub 저장소 (<https://github.com/sunwoo7441/Image-Similarity-Analyzer>)에 공개되어 있다.

## 5. 결론 및 향후 연구방향

VTO 서비스 품질 개선을 위한 기술적 접근으로, 실제 이미지와 AI 생성 이미지의 유사도를 SSIM, PSNR, VGG16-cosin 을 통해 체계적으로 실험하였다. 실험을 통해 AI 생성 과정에서의 이미지 왜곡과 현실성 저하 등의 문제점을 확인하였다. 이를 통해 VTO 기술이 실제 상용화 과정에서 해결해야 할 주요 기술적 과제들이 분명하게 도출되었다. 이러한 문제점들은 사용자의 기대에 부합하는 데 큰 장애요소로 작용하며, 이를 극복하기 위해서는 다음과 같은 연구 방향이 제안된다. 첫째, 데이터 처리의 최적화된 이미지 처리 알고리즘 개발이 필수적이다. 둘째, 멀티모달 AI 기술을 활용한 다양한 데이터 통합과 고도화를 통해 실제 착용감에 가까운 현실적인 이미지를 제공해야 한다. 셋째, 의류 텍스처 및 세부 구조 보존을 위해 GAN 기반의 고도화된 기하학적 워핑 및 피팅 기술을 더욱 발전시켜야 한다. 마지막으로, 현실적인 색상 관리 시스템의 개발과 적용을 통해 사용자 환경에 더욱 자연스럽게 융합될 수 있는 이미지를 생성해야 한다. 이러한 다각적 접근을 통해 VTO 기술은 사용자 만족도를 획기적으로 높일 수 있을 것이다.

또한 향후 Virtual Try-On 기술의 정밀도를 향상시키기 위해서는 Retrieval-Augmented Generation(RAG) 기반의 외부 지식 통합과 사용자 피드백 시스템의 활용이 필요하다. RAG 는 외부 문서나 예시 데이터를 참조하여 생성 결과의 정확도와 일관성을 높일 수 있으며, 사용자 피드백은 이미지 생성 과정에서 발생하는 오류를 반복적으로 수정하고 보완할 수 있는 방향성을 제시한다. 이번 실험을 통해 이미지 기반의 서비스에서 사용자피드백 및 만족도의 비정형데이터를 수집하고 다양한 접근을 통해 원본이미지를 유지하면서 다양한 포즈에 대한 생성형 AI 서비스를 위한 데이터 마이그레이션에 대해서 연구방향으로 나아가려 한다. 또한 LPIPS (Learned Perceptual Image Patch Similarity) 등과의 비교 분석을 통해 완성도를 높일 예정이다.

추가 유사도 지표 통합하여 더 다양한 이미지 유사도 측정 방법을 시스템에 통합하여 평가의 정확성을

높일 수 있다. 자동화된 유사도 측정과 평가자의 주관적 판단 사이의 상관관계를 분석하여 더 의미 있는 평가 지표를 연구할 예정이다. 이러한 상호작용 기반 학습 방식은 실제 사용자의 만족도를 반영하는 맞춤형 AI 모델 개발에 효과적이며, 유사도 기반 정량 분석 외에도 주관적 품질 개선을 위한 중요한 요소로 작용할 수 있다.

## 참고문헌

- [1] Park, S., Park, J., & Park, J. (2024). Trends in image-based Virtual Try-On (VTO) image synthesis technology. *Electronics and Telecommunications Trends*, 39(3), 107–115.
- [2] Park, S., & Park, J. (2024). Virtual Try-On (VTO) image synthesis reflecting reference wearing style. *Proceedings of the KCGS Conference 2024*, 70–71.
- [3] ETNews. (2024, December 18). Global AI patent application trends: Snap's 'Virtual Try-On image generation solution'. ETNews.
- [4] Ahn, H. (2018). Image-based Virtual Try-On system: Full automatic system design and its performance. *Journal of the Korean Society of Computer Games*, 31(3), 37–45.
- [5] Rahman, M. M., Tuan, D. T., & Ahn, H. (2020). An improved VTON (Virtual-Try-On) algorithm using a pair of cloth and human image. *Journal of the Korea Industrial Information Systems Research*, 25(2), 11–18.
- [6] Lee, S., Lee, S., & Lee, J. (2022). Towards detailed characteristic-preserving Virtual Try-On. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2235–2239.
- [7] He, S., Song, Y.-Z., & Xiang, T. (2022). Style-based global appearance flow for Virtual Try-On. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 3470–3479.
- [8] Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., et al. (2021). Learning transferable visual models from natural language supervision. *Proceedings of the 38th International Conference on Machine Learning*, 139, 8748–8763.
- [9] Liang, L. (2024). KAG: Boosting LLMs in Professional Domains via Knowledge Augmented Generation. Ant Group Knowledge Graph Team, Zhejiang University.
- [10] Ravi, N., et al. (2024). SAM 2: Segment Anything in Images and Videos. Meta FAIR.
- [11] Ranftl, R., Bochkovskiy, A., & Koltun, V. (2021). Vision Transformers For Dense Prediction. Intel Labs.
- [12] Marino, E., Bruno, F., & Liarokapis, F. (2021). Color harmonization, deharmonization and balancing in augmented reality. *Applied Sciences*, 11(9), 3915.
- [13] Dong, S., & Höllerer, T. (2018). Real-time re-textured geometry modeling using Microsoft HoloLens. *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, 231–237.