

## 가상 피팅 모델 ACGPN의 입력 특징 추출 기법 비교: Otsu Mask vs. Salient Region

박혜인<sup>1†</sup>, 박효령<sup>1†</sup>, 지정원<sup>1†</sup>, 김한섭<sup>2</sup>, 김웅식<sup>2\*</sup><sup>1</sup>건양대학교 의료인공지능학과<sup>2</sup>건양대학교 인공지능학과

† 본 논문에 동등하게 기여함.

## I. 서론

온라인 쇼핑은 이용자들이 의류를 직접 착용하지 않고도 편리하게 구매할 수 있다는 장점이 있다. 그러나 착장 결과에 대한 불확실성으로 인해 반품으로 이어질 가능성이 있다 [1]. 이를 해결하기 위해 Virtual Try-On(VTON) [2] 기술이 주목받고 있으며, 사용자의 전신 이미지에 의류를 합성해 실제 착용한 듯한 결과를 제공한다. 하지만 기존 VTON 기술은 합성 과정에서 부자연스러운 맥락적 영역(Artifact)이 발생하는 한계가 존재한다.

최근 Salient Region 기반 특징 추출 기법[3]이 artifact 발생 문제를 해결하는 방법 중 하나로 주목받고 있으며, 시각적으로 중요한 영역을 강조하는 것이 의류 합성 성능 향상에 효과적인 것으로 보고되고 있다 [4]. 본 연구에서는 VTON 시스템에 사용되는 입력 데이터 전처리 방식이 합성 이미지 품질에 미치는 영향을 분석하고자 한다. 이를 위해 기존에 널리 활용되고 있는 마스크 기반 특징 추출 방식과 새롭게 제안된 Salient region 기반 특징 추출 방식을 비교하여, 향후 효과적인 특징 추출 기법 개발을 위한 방향성을 제시하고자 한다.

## 2.2. Salient Region 과 Mask 기반 특징 추출

ACGPN 데이터셋의 Train\_color는 의류의 전체적인 시각 정보를 제공하지만, 본 연구에서는 모델이 보다 시각적으로 중요한 영역에 집중하여 학습할 수 있도록 기존 Train\_color를 대체하고, Salient Region 기법을 적용한 새로운 데이터셋을 구성하였다.

또한 기존 ACGPN은 U-2-Net [7]으로 추출한 Mask를 통해 의류의 외곽과 경계 정보를 반영하는 방식으로 학습을 진행하는 점에 착안하여, 본 연구에서는 이를 확장하여 Otsu, PSPNet, GrabCut, DeepLabV3+ 네 가지의 segmentation 기법을 추가로 적용하고 비교 실험을 수행하였다. 그림 2는 각 기법을 통해 추출한 결과이다.

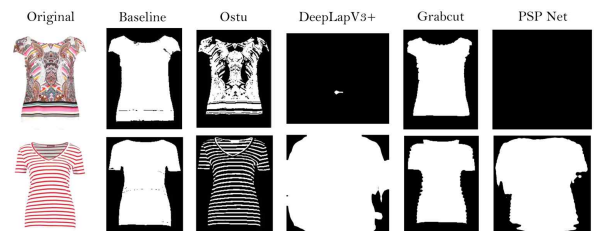


그림 2. Mask 추출 결과

## II. 실험 설계 및 방법

## 2.1. 전체 실험 구조

본 연구에서는 그림 1의 흐름도를 기반으로 DeepFashion [5] 데이터셋에 두 가지 전처리 방식을 적용한 후, 이를 VTON 기술인 ACGPN 모델[6]에 입력하여 결과 이미지를 생성하였다. 생성된 이미지는 정성적 및 정량적 평가 기준에 따라서 품질을 비교하고 평가하였다.

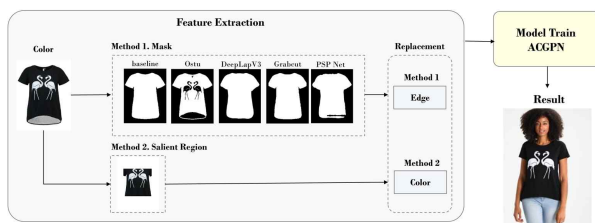


그림 1. 특징 추출 기법 비교 흐름도

정성적 평가 결과, Otsu 방식이 의류 외곽 및 특징을 가장 정밀하게 추출하는 것으로 관찰되었다. 따라서 본 연구에서는 Otsu 방식으로 생성한 마스크 데이터셋을 사용해 ACGPN 모델을 훈련시키고, 최종 합성 이미지 결과를 분석하였다.

## 2.3. 평가 지표

이미지 품질 평가는 SSIM(Structural Similarity Index Measure)과 FID(Fréchet Inception Distance)를 사용하였다. SSIM은 밝기, 명암, 구조를 고려하여 원본과 합성 이미지 간의 구조적 유사성을 평가하며, 수식 (1)로 나타낸다. FID는 생성 이미지와 실제 이미지 간 특징 분포의 거리를 측정해 사실성과 자연스러움을 평가하며, 수식 (2)로 나타낸다.

$$SSIM(x, y) = \frac{((2\mu_x\mu_y + C_1)(2\sigma_x\sigma_y + C_2))}{((\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2))} \quad (1)$$

$$FID = \|\mu_r - \mu_g\|^2 + Tr(\Sigma_r + \Sigma_g - 2(\Sigma_r \Sigma_g)^{\frac{1}{2}}) \quad (2)$$

### III. 실험 결과

본 연구에서는 입력 데이터셋의 특징 추출 방식이 VTON 시스템의 합성 품질에 미치는 영향을 분석하였다. 표 1의 결과에 따르면, Otsu mask 및 Salient region 기법 모두 Baseline 대비 성능 향상을 보이지 않았으며, 과도한 특징 추출이 오히려 품질 저하를 초래할 수 있음을 확인하였다.

표 1. 특징 추출 방식별 성능 비교

Method	SSIM ↑	FID ↓
Baseline	<b>0.860</b>	<b>10.685</b>
Otsu mask	0.848	22.77
Salient region	0.850	16.896



그림 3. 복잡한 패턴 의류에서의 품질 평가

텍스처나 프린팅이 두드러진 의류의 경우, 그림 3과 같이 Salient region 기법이 일정 수준 품질 개선에 기여할 수 있음을 관찰하였다. 반면, Otsu mask를 사용한 경우에는 복잡한 패턴에서도 품질 저하가 발생하였으며, 이는 과도한 단순화로 인한 정보 손실로 해석된다.



그림 4. 특징 추출 기법별 실험 결과 비교

그림 4는 다양한 특징 추출 기법에 따라 생성된 이미지들의 품질 차이를 비교한 사례를 보여준다. 이를 통해 전처리 기법에 따라 결과 이미지의 세부 품질과 부적절한 맥락 양상이 다르게 나타남을 확인할 수 있다.

### IV. 결론

본 연구는 입력 데이터의 특징 추출 방식이 VTON 시스템의 성능에 미치는 영향을 체계적으로 분석하였다. 실험 결과, 특징 추출 기법은 단순 적용을 넘어 데이터의 구조적 복잡성과 특징 분포를 충분히 고려하여 설계되어야 함을 확인하였다. 향후 연구에서는 다양한 의류 특성에 따른 입력 특징 추출 전략의 동적 조정방안을 탐색하고자 한다.

### ACKNOWLEDGMENTS

본 연구는 과학기술정보통신부 및 정보통신기획평가원의 SW 중심대학사업(2024-0-00047) 지원과 대학혁신지원사업의 캡스톤 디자인 교과목 지원을 받아 수행되었음.

### REFERENCES

- [1] Nestler, A., Karessli, N., Hajjar, K., Weffer, R., & Shirvany, R. (2021). SizeFlags: Reducing Size and Fit Related Returns in Fashion E-Commerce. arXiv preprint arXiv:2106.03532.
- [2] Han Yang, Ruimao Zhang, Xiaobao Guo, Wei Liu, Wangmeng Zuo, Ping Luo, "Towards Photo-Realistic Virtual Try-On by Adaptively Generating ↔ Preserving Image Content", 12 Mar 2020
- [3] Borji, A., Cheng, M. M., Jiang, H., & Li, J. (2015). Salient object detection: A benchmark. IEEE Transactions on Image Processing, 24(12), 5706-5722.
- [4] Ji Woo Hong, Tri Ton, Trung X. Pham, Gwahnyeong Koo, Sunjae Yoon, Chang D. Yoo, "ITA-MDT: Image-Timestep-Adaptive Masked Diffusion Transformer Framework for Image-Based Virtual Try-On", arXiv preprint arXiv:2503.20418, 26 Mar 2025.
- [5] Zhu, H., Cao, Y., Jin, H., Chen, W., Du, D., Wang, Z., Cui, S., & Han, X. (2020). Deep Fashion3D: A dataset and benchmark for 3D garment reconstruction from single images. In Computer Vision - ECCV 2020 (pp. 512-530). Springer International Publishing.
- [6] Yang, H., Zhang, R., Guo, X., Liu, W., Zuo, W., & Luo, P. (2020). Towards photo-realistic virtual try-on by adaptively generating-preserving image content. IEEE/CVF conference on computer vision and pattern recognition (pp. 7850-7859).
- [7] Xuebin Qin, Zichen Zhang, Chenyang Huang, Masood Dehghan, Osmar R. Zaiane, Martin Jagersand, "U<sup>2</sup>-Net: Going Deeper with Nested U-Structure for Salient Object Detection", arXiv preprint arXiv:2005.09007, 8 Mar 2022.