

# 가려짐에 강건한 포즈 추정을 위한 이미지 합성 기반 다중 겹침 데이터셋

추연승, 김현식, 박용석  
한국전자기술연구원

piksal@keti.re.kr, hskim@keti.re.kr, yspark@keti.re.kr

## Multi-Overlapped Dataset for Occlusion-Robust Pose Estimation

Yeon Seung Choo, Hyun-Sik Kim, Yong-Suk Park  
Korea Electronics Technology Institute

### 요 약

인체 포즈 추정에서 가려짐은 여전히 도전적인 문제 중 하나이다. 이를 위해 CrowdPose, OCHuman 등 다양한 인체 포즈 데이터셋이 제안되었다. 그러나 이들 데이터셋에서 활용된 데이터셋의 GT(Ground Truth)는 데이터셋 제작자들에 의해 지정되었기 때문에, 실제 가려짐이 발생한 관절 위치가 정확한 관절 위치인지 알 수 없다. 본 논문에서는 이러한 문제점을 해결하고자 기존의 인체 포즈 데이터셋을 개선해 인위적인 인체 겹침에 의한 가려짐이 있는 인체 포즈 데이터셋 대해 제안한다. 본 논문은 기존의 GT 를 그대로 활용하기 때문에 기존의 GT 를 그대로 쓸 수 있다는 장점이 있다. 결과적으로, 본 논문의 방법을 통해 기존의 포즈 추정 성능을 향상시킬 수 있다.

### 1. 서 론

인공지능 기술의 발전에 따라 인체 포즈 추정 기법의 성능 또한 크게 향상되었다. 그러나 여전히 인체의 포즈를 추정하는 과정에서 가려짐은 가장 큰 문제 중 하나이다. 이를 개선하기 위해 다음과 같이 가려짐을 집중적으로 개선하고자 하는 데이터셋 및 방법론에 대한 연구들이 제안되기도 하였다[1-5].

그러나 CrowdPose, OCHuman 등 이러한 가려짐에 집중되어 제안된 기존의 데이터셋들은 제작자에 의해 GT(Ground Truth)가 지정되기 때문에, 정확한 인체 관절 값과 일치하는지 알 수 없다[4-5].

이러한 문제점을 개선하고자 본 논문에서는 기존의 데이터셋을 활용해 GT 를 유지한 채 인위적인 가려짐을 통해 데이터셋을 제작하는 방법에 대해 제안한다.

본 논문에서는 기존의 데이터셋을 통해 임의의 겹쳐짐이 발생한 인체 포즈 데이터셋을 제작하고, 이미지 조화(image harmonization) 기법을 활용해 합성된 영상 간의 조도 차이를 감소시킨다. 이후, 합성된 이미지 간의 포즈 추정 차이를 통해 포즈 추정 성능을 향상시키는 방법에 대해 제안한다. 본 논문에서 제안된 방법을 활용한다면 기존의 GT 정보를 유지한 채로 새롭게 가려짐이 발생한 데이터셋을 제작할 수 있다.

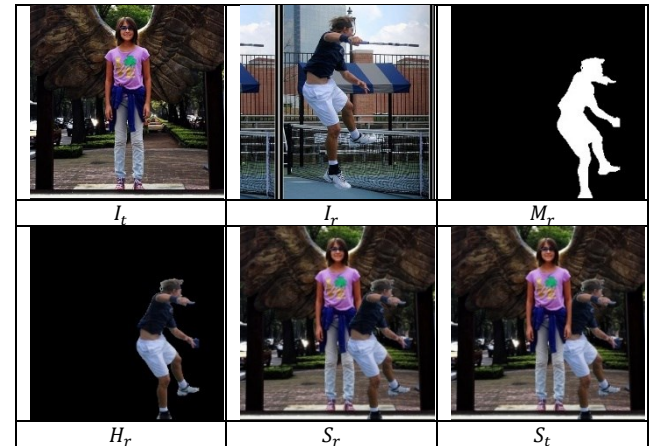


그림 1. 이미지 합성 예시

먼저, 대상 이미지  $I_t$ 와 참조 이미지  $I_r$ 에 대해, 사람 영역 마스크  $M_t$ 와  $M_r$ 을 구한다. 다음으로 사람 영역 마스크  $M_r$ 을 랜덤하게 이동해가며 전체 마스크 영역내에서 겹침 영역 비율을 나타내는 IOU(Intersection Over Union)를 식 (1)을 통해 계산한다.

$$IOU = \frac{M_t \cap M_r}{M_t \cup M_r} \quad (1)$$

### 2. 본론

#### 2.1 이미지 합성

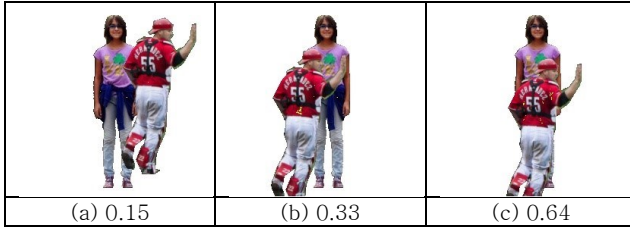


그림 2. IOU 비율에 따른 합성 예시

미리 설정한 IOU 값을 충족하는  $M_r$  이 추출된다면, 사람 영역 이미지  $H_r = I_r \otimes M_r$ 에 대해 대상 이미지  $I_t$ 에 대한 이미지 마스킹을 식 (2)와 같이 수행해 합성 이미지  $S_r$ 를 만든다.

$$S_r = \begin{cases} H_r & \text{if } M_r > 0 \\ I_t & \text{else} \end{cases} \quad (2)$$

다음으로, 합성 이미지  $S_t$ 를  $S_r$ 을 고려해 식 (3)을 통해 만든다.

$$S_t = \begin{cases} I_t & \text{if } M_t > 0 \\ S_r & \text{else} \end{cases} \quad (3)$$

## 2.2 이미지 조화

이미지 조화는 다른 조도에서 촬영된 두 이미지의 합성에서 대상 객체와 배경 이미지 간의 유사한 조도 환경을 유사하게 표현하는 영상 처리 기법 중 하나이다. 본 논문에서는 자연스런 합성을 위해 K. Sofiiuk 등이 제안한 이미지 조화 방법을 적용한다[6]. 그림 3은 이미지 조화를 적용한 예시를 나타낸다.



그림 3. 이미지 조화 기법 적용 예시

## 2.3 포즈 추정

이후, 합성 이미지  $S_t$ ,  $S_r$ 에 대해 포즈 추정 네트워크  $P$ 로 추정된 결과는 서로 같아야 하므로 목적 함수  $L$ 은 식 (4)와 같이 나타낼 수 있다. 결과적으로, 식 (4)를 활용해 어느 포즈 추정 모델에서라도 추가 학습을 수행할 수 있다.

$$L = \frac{1}{N} \sum_{i=1}^N \|P(S_t) - P(S_r)\| \quad (4)$$

## 3. 실험 결과

먼저, 본 논문에서는 Q. Zhou 등이 제안한 AHP dataset을 활용하였다[3]. 또한 본 논문에서 사용한 IOU 범위는 0.45 이상 0.75 이하로 범위를 제한하였다. 그림 4는 실험 결과를 통해 제작된 데이터셋을 나타낸다.



그림 4. 증강 합성 데이터셋의 결과

## 4. 결론

본 논문에서는 기존 데이터셋 합성을 통한 새로운 인체 포즈 데이터셋 제작 방법 및 학습 방법에 대해 제안한다. 제안하는 방법을 통해 기존의 포즈 추정 네트워크와 데이터셋을 따로 수정하지 않고 새롭게 제작한 데이터셋 만으로도 추가 학습을 가능할 수 있다. 따라서, 본 방법은 포즈 추정 이후 성능 고도화를 위한 어느 포즈 추정 방법에서도 활용될 수 있다.

## ACKNOWLEDGMENT

이 논문은 2022년도 정부(산업통상자원부)의 재원으로 한국산업기술평가관리원의 지원을 받아 수행된 연구임 (No. 20009940, 개인 상황 인지를 기반으로 스마트 홈 케어가 가능한 Home Ambient Intelligence Display (HAID) 및 디자인 개발)

## 참 고 문 헌

- [1] R. Khirodkar, S. Tripathi, and K. Kitani, "OCHMR: Occluded Human Mesh Recovery," in *Computer Vision and Pattern Recognition (CVPR)*, pp. 1715-1725, 2022.
- [2] R. Khirodkar, V. Chari, A. Agrawal, and A. Tyagi, "Multi-Instance Pose Networks: Rethinking Top-Down Pose Estimation," in *International Conference on Computer Vision (ICCV)*, pp. 3122-3131, 2021.
- [3] Q. Zhou, S. Wang, Y. Wang, Z. Huang, and X. Wang, "Human De-occlusion: Invisible Perception and Recovery for Humans," in *Computer Vision and Pattern Recognition (CVPR)*, pp. 3691-3701, 2022.
- [4] J. Li, C. Wang, H. Zhu, Y. Mao, H.-S. Fang, and C. Lu, "CrowdPose: Efficient Crowded Scenes Pose Estimation and a New Benchmark," in *Computer Vision and Pattern Recognition (CVPR)*, pp. 10863-10872, 2019.
- [5] S.H. Zhang, R. Li, X. Dong, P. Rosin, Z. Cai, X. Han, D. Yang, H. Huang, and S.-M. Hu, "Pose2Seg: Detection Free Human Instance Segmentation," in *Computer Vision and Pattern Recognition (CVPR)*, pp. 889-898, 2019.
- [6] K. Sofiiuk, P. Popenova, and A. Konushin, "Foreground-Aware Semantic Representations for Image Harmonization," *Winter Conference on Applications of Computer Vision (WACV)*, pp. 1620-1629, 2021.