JKSCI

한국컴퓨터정보학회논문지
Journal of The Korea Society of Computer and Information
Vol. 29 No. 12, pp. 129-138, December 2024
https://doi.org/10.9708/jksci.2024.29.12.129

Refining Crowd Pose Annotation Dataset for Accurate Multiple-Person Pose Estimations in Crowd Situations

Jin-Woo Cha*, Chulyoung Kim**, Hyun-Jong Oh***, Da-Jeong Seo****, Jong-Seong Park***, Yoo-Sung Kim*****

*Undergraduate Student, Dept. of Computer Information, Inha Technical College, Incheon, Korea
**Graduate Student, Dept. of Electronics and Computer Eng., Inha University, Incheon, Korea
***Undergraduate Student, Dept. of Information and Communication Eng., Inha University, Incheon, Korea
****Undergraduate Student, Dept. of Artificial Intelligence, Inha University, Incheon, Korea
*****Professor, Dept. of Artificial Intelligence, Inha University, Incheon, Korea

[Abstract]

This paper describes refinement and release of a human pose annotation dataset, which is essential for developing multi-person pose estimators in crowd situations. To achieve this, we first developed an annotation refining tool that can be helpful to put accurate bounding box labels for many people at most and correct keypoint positions for each person in crowd situation images where many occlusions might occur since the crowd density is high. Using this tool, we enhanced the quality of the ground truth annotations for 8,000 test images from the CrowdPose dataset, which is widely used for evaluating the performance of multi-person pose estimators in crowd situations. Our analysis confirms that the performance of the multi-person pose estimator can be more accurately assessed, as the modified dataset contains more person's bounding boxes and more accurate keypoint labels for each person than the previous version. The developed human pose annotation refining tool and the modified dataset will be publicly available at https://github.com/InhaKMS/HuPo-AnT.

▸ **Key words**: Human pose annotation, Crowd situations, Multi-person pose estimation, Person bounding box, Keypoint labels

• Co-First Author: Jin-Woo Cha, Chulyoung Kim, Corresponding Author: Yoo-Sung Kim
*Jin-Woo Cha (chajinwoo.chajinwoo@gmail.com), Dept. of Computer Information, Inha Technical College
**Chulyoung Kim (chulsub0727@gmail.com), Dept. of Electronics and Computer Eng., Inha University
***Hyun-Jong Oh (jong1029@naver.com), Dept. of Information and Communication Eng., Inha University
****Da-Jeong Seo (7803jung@naver.com), Dept. of Artificial Intelligence, Inha University
***Jong-Seong Park (pjsc0903@inha.edu), Dept. of Information and Communication Eng., Inha University
*****Yoo-Sung Kim (yskim@inha.ac.kr), Dept. of Artificial Intelligence, Inha University
• Received: 2024. 10. 11, Revised: 2024. 11. 25, Accepted: 2024. 11. 27.

http://www.ksci.re.kr pISSN:1598-849X | eISSN:2383-9945

[요 약]

본 논문에서는 군중 상황(crowd situations)에서 다중 사람 자세 인식기(multi-person pose estimator)를 개발하는 과정에서 필수적인 사람 자세 주석 데이터 세트를 개선하여 공개하는 연구를 소개한다. 이를 위해 먼저, 군중 밀도가 높아서 가림(occlusion)이 많이 발생하는 군중 상황 이미지에 대해서 가능한 많은 사람의 바운딩 박스(bounding box)와 각 사람의 관절 레이블(keypoint label)을 정확하게 표시하는데 편리하게 사용할 수 있는 주석 개선 도구(annotation refining tool)를 개발하였다. 개발된 도구를 이용하여 군중 상황에서 다중 사람 자세 인식기의 성능 평가용으로 많이 이용되는 기존 CrowdPose 데이터 세트의 8,000개 테스트 이미지의 사람 자세 주석을 교정하여 품질을 개선하였다. 분석 결과에 따르면 수정된 데이터 세트가 이전 버전보다 더 많은 사람의 바운딩 박스를 포함하고 있으며, 또한 각 사람의 관절 레이블을 더 정확하게 표시하고 있어서 다중 사람 자세 인식기의 성능을 더 정확하게 평가할 수 있음을 확인하였다. 개발된 사람 포즈 주석 개선 도구와 수정된 데이터 세트는 https://github.com/InhaKMS/HuPo-AnT에서 공개될 예정이다.

▸ **주제어**: 사람 자세 주석, 군중 상황, 다중 사람 자세 인식, 사람 바운딩 박스, 관절 레이블

I. Introduction

컴퓨터 비전 분야에서 다양한 목적으로 사람의 자세(pose) 및 행위를 인식하는 연구가 활발하게 진행되고 있으며 최근에는 딥러닝(deep-learning) 기술을 이용해서 인식 성능이 많이 증진되고 있다[1-7]. 이러한 기술을 이용하여 최근에는 군중 상황(crowd situations)의 입력 이미지에서 사람을 인식하고, 다중 사람(multi-person)의 자세를 인식하여 군중의 이상 상황을 인식하려는 연구로까지 확대 발전하고 있다[8-12].

군중 상황에서 사람을 인식하고 사람의 자세를 인식하는 딥러닝 네트워크를 학습시키는 과정뿐만 아니라, 개발된 자세 인식기의 성능을 정확하게 측정하여 좋은 모델을 선정하는 과정에서도 다양한 군중 상황의 이미지와 이를 위한 사람의 자세 주석 데이터가 필요하다[5-6, 9-10, 13-15]. 많은 사람이 포함된 군중 상황을 촬영한 이미지에는 기본적으로 출현하는 사람의 수가 많으며, 군중 상황 내에서 각 출현 사람과 촬영 카메라와의 거리 차이로 인해 이미지 내에서의 사람 영역의 크기가 다를 수 있으며, 배경 객체 또는 군중 속의 다른 사람에 의해 가림(occlusion)이 발생할 수 있는 특성이 있다[9-10, 12]. 따라서 군중 상황에서 다중 사람 자세 인식기를 개발하기 위한 데이터 세트에는 이러한 특성들이 잘 표현된 다양한 이미지들을 포함하고 있어야 하며, 지상 실측 주석(ground truth annotation)에는 가능한 각 이미지 내의 많은 사람을 위한 바운딩 박스가 포함되어야 하며 각 사람의 관절이 정확하게 레이블되어 있어야 한다.

군중 상황에서의 사람 자세 인식기를 개발하는 기존 연구에서 많이 사용한 데이터 세트로는 MPII 데이터 세트([13]), MS COCO 데이터 세트([14]), AI 챌린저 데이터 세트([15]), CrowdPose 데이터 세트([10]) 등이 있다. 그러나, 기존 연구 [10]과 [12]에서는 [13, 14, 15]의 데이터 세트가 군중 상황을 대표하기 위한 위의 3가지 특성을 충분하게 포함하고 있지 못하다고 지적했다. 우선 [10]에 따르면 위의 데이터 세트 내의 이미지에 포함된 사람의 수가 적고, 군중의 밀집 정도를 표시하는 Crowd Index가 낮으므로 사람 간의 가림이 거의 없는 이미지들을 대부분 포함하고 있다고 분석하였다. 반면에 CrowdPose 데이터 세트([10])는 Crowd Index가 0부터 1 사이에 균등하게 분포하고 있어서 군중 상황의 다양한 가림 상황을 포함하고 있다고 하였다. 그러나 [12]에 의하면 CrowdPose 데이터 세트도 이미지 내에서 많은 사람이 군집한 부분은 'IS_CROWD' 군중 영역 박스로만 구분하였지, 각 사람을 위한 개별 바운딩 박스를 포함하지 않았기에 이미지 당 평균 약 4명만이 지상 실측 데이터에 포함되었다고 지적하였다. 따라서 CrowdPose의 데이터 세트도 군중 상황에서 다중 사람의 자세를 정확하게 인식하고 평가하는 용도로 사용하기에 부족하다고 하였다[12].

본 연구에서는 군중 상황에서 다중 사람 자세 인식기 개발을 위해서 많이 사용되는 CrowdPose 데이터 세트의 정확성을 개선하여 활용도를 높이고자 한다. 즉, 다중 사람 자세 인식기의 성능을 정확하게 평가할 수 있도록 지원하

기 위해 CrowdPose 데이터 세트의 성능 평가용 데이터인 8,000장의 테스트 이미지를 위한 사람 자세 주석 레이블을 검수하고 정확성을 개선하였다.

이를 위해서 기존 CrowdPose 데이터 세트의 주석 생성을 위해 사용된 도구에 대한 정보가 없었기에 관련 분야의 기존 연구에서 많이 인용하는 [14]의 사람 자세 주석 형식에 맞게 주석 레이블을 생성할 수 있는 도구를 조사하였다. 이렇게 조사한 COCO 주석 생성기([16]), Supervisely([17]), 그리고 Visipedia([18]), 모두 다중 사람의 자세 인식기 개발을 위한 주석을 처음부터 생성하는 용도로는 유용하게 사용될 수 있지만, 기존의 CrowdPose 데이터 세트의 주석을 검수하고 수정하는 용도로 사용하기에는 불편함이 많았다. 특히, 기존의 MS COCO 사람 자세 주석 형식([14])에 맞게 17개의 관절에 대해 정해진 순서대로 처음부터 주석을 생성할 수 있도록 설계되어 있었기 때문에 기존의 CrowdPose 주석을 검수하여 누락된 사람을 위한 바운딩 박스를 추가하거나 기존 사람의 잘못된 관절 레이블을 교정하는 작업을 효율적으로 지원하지 못하였다. 따라서 본 연구에서는 CrowdPose 데이터 세트의 지상 실측 주석을 검수하고 발전시키기 위한 주석 개선 도구를 자체적으로 개발하였다.

개발된 도구를 이용하여 각 이미지 내에서 식별이 가능한 사람을 바운딩 박스 레이블로 표시하여 포함하고, 각 사람에 대해서도 가능한 많은 관절의 위치를 정확하게 레이블링하여 자세 인식에 활용할 수 있도록 정확성을 증진하였다. 개발된 주석 개선 도구를 이용하여 원래의 테스트 이미지 8,000장에 포함된 35,283개의 바운딩 박스내의 총 487,396개의 관절 레이블을 검사하고, 잘못된 바운딩 박스의 삭제, 위치 및 크기 수정, 관절의 위치 수정, 바운딩 박스와 관절 포인트의 추가 등의 교정 작업을 통해 최종적으로 35,993명의 바운딩 박스 내에 총 503,902개의 관절 레이블을 포함하도록 발전시켰다. 또한, 원래의 주석 데이터 세트에서 주석에 오류가 포함된 이미지 샘플을 선정하여 수정 전의 상황과 수정 후의 상황을 비교하는 실험을 통해 개선된 데이터 세트의 정확성과 유용성을 입증하는 비교 실험을 진행하였다.

본 논문의 구성은 다음과 같다. 2절에서는 기존 CrowdPose 데이터 세트를 소개하고 개선 필요사항을 지적한다. 3절에서는 군중 상황의 이미지에 대해 사람의 바운딩 박스와 관절의 주석 레이블을 검수하고 수정하기 위해 개발한 주석 개선 도구를 소개한다. 4절에서는 개발된 도구를 사용하여 발전시킨 CrowdPose 테스트 데이터의 통계, 수정된 사람 자세 레이블의 대표를 예시하고 이들을 이용한 성능 평가 결과의 변화에 대해 분석하였다. 5절에서는 본 연구의 결론과 향후 연구 방향에 관해서 기술한다.

II. Related Works

본 절에서는 군중 상황의 다중 사람 자세 인식기 개발 및 성능 평가의 목적으로 이용되고 있는 CrowdPose 데이터 세트([10])에 대해서 간략하게 소개한다. [10]의 저자들은 인터넷에 공개된 [14, 15] 등과 같은 기존 데이터 세트를 분석하고 다양한 자세를 위한 이미지들을 선별하여 원본 데이터 세트에 포함시켰다. 이러한 원본 이미지에는 소수의 사람이 출현한 이미지뿐만 아니라 여러 사람이 출현한 군중 상황의 이미지가 포함되어 있으며 이를 다중 사람 자세 인식기의 개발 및 성능 평가용 데이터 세트로 사용하기 위해서 사람의 바운딩 박스와 각 사람의 최대 14개의 관절의 위치를 사람 자세 인식을 위한 주석 레이블로 표시하였다.

또한, 이미지 내에서 한 사람의 바운딩 박스 내에 존재하는 해당 사람의 관절 수 대비 다른 사람들의 관절 수의 비율을 *crowd ratio*로 정의하여 가림(occlusion) 정도를 나타내고, 한 이미지에 포함된 모든 바운딩 박스의 *crowd ratio*의 평균값을 해당 이미지의 *Crowd Index*로 정의하여 해당 이미지의 군중 밀집 정도를 표현하였다. *Crowd Index*를 기준으로 분석하면, 기존 [13-15]의 데이터 세트는 0에 가까운, 겹침에 의한 가림이 거의 없는 상황의 이미지들로 대부분 구성되어 있었지만, [10]에서 구축한 데이터 세트에는 *Crowd Index*가 0과 1 사이에 균등하게 분포되어 있어서 가림이 없는 일반 상황의 이미지뿐만 아니라 가림이 많은 상황의 이미지까지 다양하게 포함하고 있기에 ***CrowdPose*** 데이터 세트라고 명명하였다.

그러나 군중 상황이 아니더라도 적은 수의 사람이 근접해서 있는 경우에 가림이 많이 발생할 수 있기 때문에, *Crowd Index*가 높다고 해서 군중 상황을 대표하는 이미지라고 볼 수 없고 이미지에 포함된 사람의 수를 정확하게 파악하는 것이 군중 상황 여부를 결정하는 중요한 요소라고 지적하고 [12]에서는 군중의 밀집 정도를 구분하기 위해 이미지에 포함된 사람의 수에 따라 이미지를 그룹으로 구분하고, 각 그룹 내에서 *Crowd Index*를 이용하여 이미지들을 세 분류하였다. 따라서 사람을 표시한 바운딩 박스가 많고 *Crowd Index*가 높으면 군중의 밀집도가 높은 상황의 이미지이고 사람의 바운딩 박스의 수가 적고 *Crowd Index*가 낮으면 출현 사람이 적고 가림이 적은 일반 상황을 나타

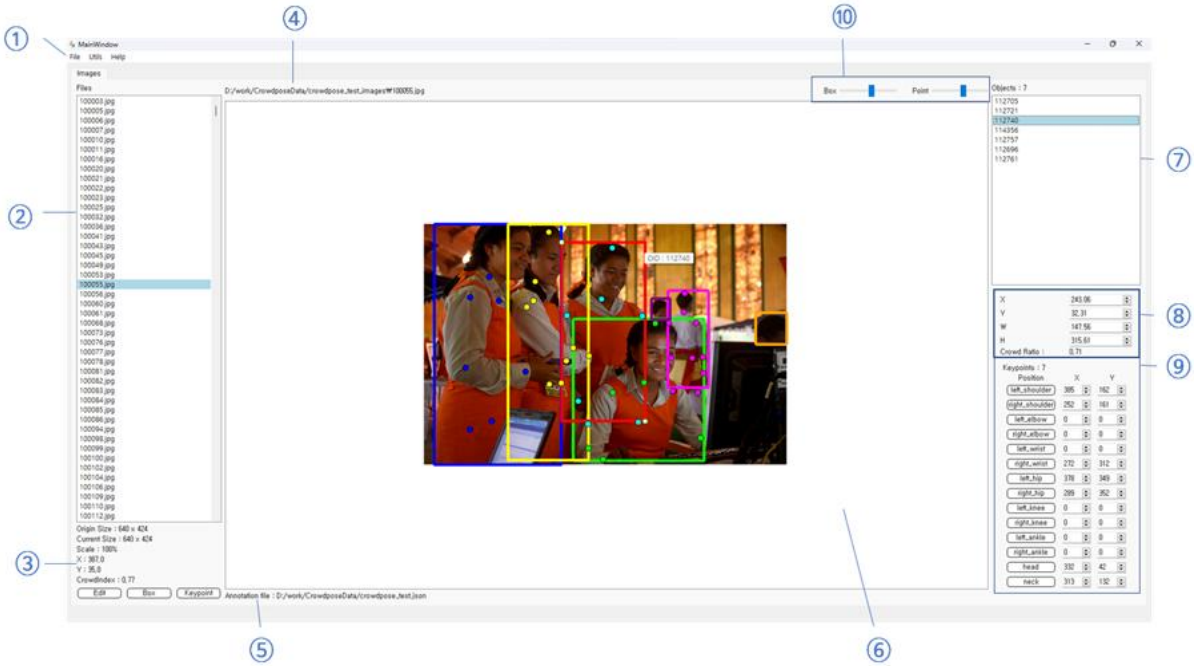


Fig. 1. Configuration of the User Interface of the Developed Annotation Refining Tool

내는 것으로 분류하였다. [12]의 분석 결과에 의하면 출현 사람이 적고 가림이 적은 일반 상황보다 많은 사람이 출현하고 군중 밀도가 높은 경우에 다중 사람의 자세 인식 정확도가 낮아지는 결론을 확인할 수 있었다. 또한, [12]의 분석 결과에 따르면 CrowdPose 데이터 세트의 테스트용 이미지에는 실제로 포함되는 사람보다 적은 수의 바운딩 박스가 주석 데이터에 포함된 경우가 많이 존재하는 것으로 분석하였다. 이러한 이미지에 최신의 객체 검출기(object detector)를 이용하면 지상 실측 데이터에 포함된 사람의 바운딩 박스보다 더 많은 사람을 인식할 수 있음을 확인하였다. 즉, CrowdPose 데이터 세트의 이미지에 포함된 사람들에 대해서 충분하게 지상 실측 레이블을 생성하지 않았기 때문에 군중 상황에서 다중 사람 자세 인식기의 성능을 정확하게 평가하지 못하고 있음을 확인하였다.

이러한 CrowdPose 데이터 세트의 부족함을 개선하기 위한 목적으로 사용하기 위한 도구로써 COCO 주석 생성기([16]), Supervisely([17]), 그리고 Visipedia([18])을 조사하여 분석하였다. 이러한 도구들은 웹 환경에서 사용할 수 있도록 개발되어 편리하게 사람의 자세 주석을 처음부터 생성하는 용도로는 유용하게 사용될 수 있지만, 기존 주석을 검수하고 수정하는 용도로 사용하기에는 불편함이 있었다. 특히, 기존의 CrowdPose 주석을 검수하여 누락된 사람을 위한 바운딩 박스를 추가하고 각 사람의 관절 레이블을 추가하거나 기존 사람의 잘못된 관절 레이블을 교정하는 작업을 효율적으로 지원하지 못하였다. 따라서 본 연구에서는 CrowdPose 지상 실측 주석을 검수하고 개선하기 위한 주석 개선 도구를 자체적으로 개발하여 사용하였다.

III. A Crowd Pose Label Refining Tool

본 연구에서는 기존 [10]의 CrowdPose 데이터 세트의 문제점을 개선하기 위해서 입력 이미지에서 육안으로 확인할 수 있는 사람에 대해서 최대한 바운딩 박스를 정확하게 포함하도록 하고, 각 바운딩 박스 내의 사람의 관절 위치를 쉽고 정확하게 수정하도록 지원하는 주석 개선 도구를 개발하였다. 본 연구에서 사용하는 사람 자세 주석 개선 도구를 개발하기 위해서 기본적으로 Python 3.8.10을 사용했으며, GUI를 구현하기 위해서 PyQt5 5.15.10을, 이미지를 처리하고 표시하기 위해 Pillow 10.2.0을, 그리고 관절 포인트와 바운딩 박스의 기하학적 연산을 위해서 Shapely 2.0.5를 사용하였다.

개발한 도구는 기본적으로 기존 CrowdPose 데이터 세트를 사용하는 기존 연구와의 호환성을 유지하기 위해서 사람별로 14개의 관절을 표시하고, 각 바운딩 박스를 기준으로 crowd ratio를 계산하고, 각 이미지의 바운딩 박스들의 crowd ratio 평균으로 Crowd Index를 계산하여 주석 데이터에 포함하였다. 또한, 주석의 생성 작업을 여러 사람이 동시에 진행하고 이를 통합해서 사용할 수 있도록 지원하는 기능을 포함하고 있으며, 추가로 다양한 목적으

로 진행되는 다중 사람 자세 인식기의 개발 및 성능 평가를 지원하기 위해서 출현 사람의 수, 각 사람의 최소 관절의 수, 이미지의 Crowd Index 등을 기준으로 필터링하는 기능을 지원하고 있으며 다른 군중 상황 데이터 세트를 변환하여 추가하는 기능도 일부 지원하고 있다.

[그림 1]은 개발된 사람 자세 주석 개선 도구의 사용자 인터페이스의 전체 구성을 표시하고 있다. ①은 메뉴바 부분으로 이미지들과 기존 주석 데이터를 선택하는 'File' 메뉴와 다양한 지원 기능을 포함하는 'Utils' 메뉴로 구성되어 있다. ②부분은 선택된 이미지들의 목록을 제시하는 창이고 ③부분은 ②의 이미지 목록에서 선택되어 ⑥의 중앙 작업 창에 나타난 이미지의 원본 크기와 Crowd Index 값, 그리고 줌인 줌아웃 단축키를 이용하여 조절한 이미지의 크기 비율과 현재의 작업 모드와 마우스의 위치를 표시하고 있다. ④부분은 마지막으로 선택한 이미지가 저장된 폴더, ⑤부분은 현재 선택된 주석 데이터의 저장 폴더를 표시하고 있다. ⑥은 중앙 작업 창으로서 현재 선택된 이미지와 해당 이미지 내에 주석이 포함하고 있는 사람의 바운딩 박스와 각 사람의 관절 정보를 표시하고 있다. ⑦은 선택된 이미지 내에 포함된 사람의 바운딩 박스 리스트를 표시한다. ⑧부분은 ⑦의 바운딩 박스 리스트에서 선택된 바운딩 박스의 크기 정보와 *crowd ratio*를 표시한다. ⑨부분은 선택된 바운딩 박스의 사람의 14개 관절의 위치를 표시한다. ⑩은 ⑥은 중앙 작업 창 내에서 바운딩 박스를 표시하는 선의 굵기와 관절의 위치를 표시하는 점의 크기를 조정하는 부분이다.

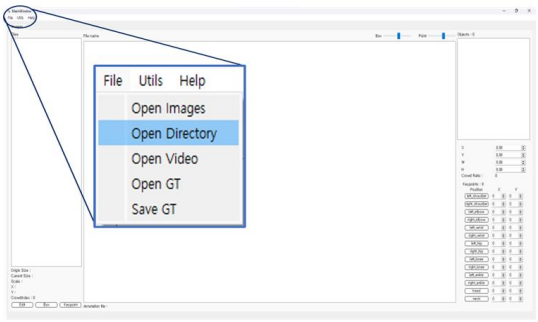


Fig. 2. Drop-down Sub-menus of 'File' Menu

[그림 2]는 개발된 주석 개선 도구의 주요 메뉴인 'File'의 서브 메뉴를 보여주고 있다. 작업할 이미지를 낱개 단위로 선택하기 위한 'Open Images', 이미지들을 포함하고 있는 폴더 단위를 지정하기 위한 'Open Directory', 또한 비디오 데이터를 입력받아서 프레임 단위로 주석을 생성하기 위해 기능 구현 중인 'Open Video' 서브 메뉴가 포함되어 있으며, 그리고 작업에 사용할 주석 데이터를 열기 위한 'Open GT'와 작업 내용을 새로운 주석 데이터로 저장하기 위한 'Save GT' 서브 메뉴가 있다.

[그림 3]은 주석 개선 도구를 이용하여 작업 컴퓨터의 'D:/work/CrowdposeData/crowdpose_test_images' 폴더 내의 이미지들을 선택하고 그중에 '110725.jpg'를 선택한 후 'D:/work/CrowdposeData/crowdpose_test.json' 주석 데이터에 포함된 2명의 사람을 위한 바운딩 박스 '167519'와 '180784' 이외로 하단에 화살표로 표시한 어린이를 위한 노란색의 바운딩 박스 '3186785'를 추가하였고, 오른쪽에 화살표로 표시한 바와 같이 해당 어린이의 오른쪽

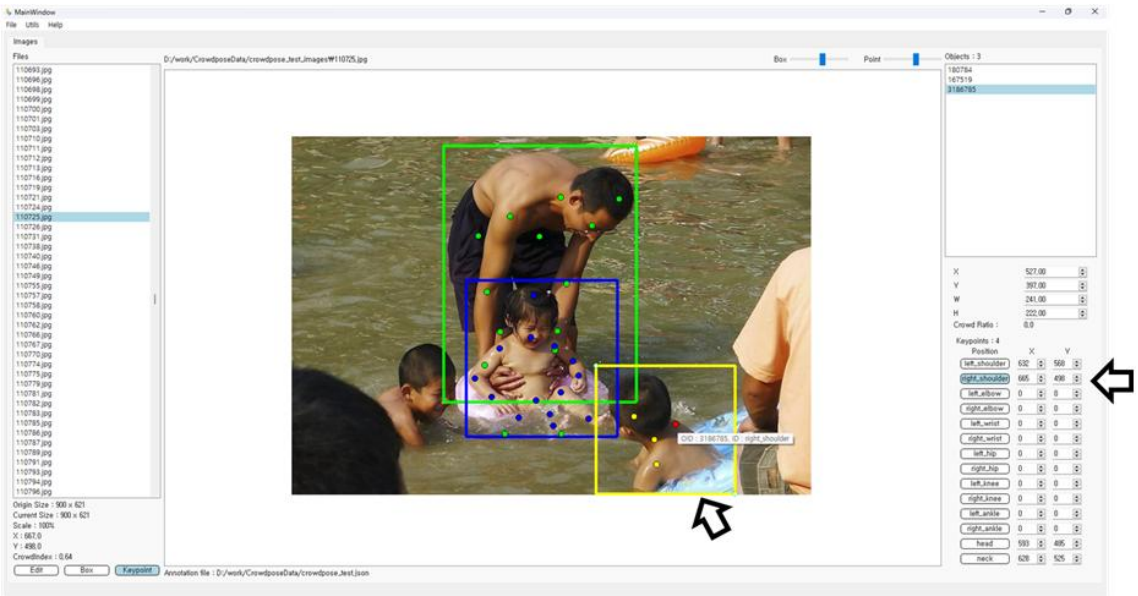
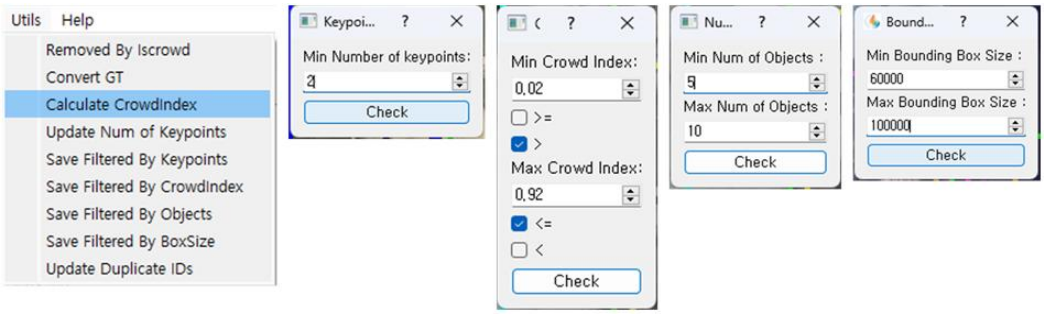


Fig. 3. Snapshot Image of Adding Person's Bounding Box and His Keypoints



(a) (b) (c) (d) (e)

Fig. 4. Drop-down Submenus of 'Utils' and the Related Windows for Filtering Conditions

어깨 관절의 위치를 추가하는 작업 상황을 보여주고 있다.

[그림 4]는 'Utils' 메뉴의 서브 메뉴와 지상 실측 주석 데이터의 필터링 조건을 위한 팝업 창을 예시하고 있다. [그림 4] (a)는 'Utils'의 서브 메뉴를 보여주고 있으며, 'Removed By Iscrowd'는 기존 CrwodPose 데이터 세트의 이미지 내에서 'IS_CROWD'로 표시하여 각 사람을 위한 주석을 생성하지 않은 부분의 표식을 제거하는 기능, MS COCO 데이터 세트 등과 같이 다른 주석 데이터 형식으로 표시된 지상 실측 레이블을 CrowdPose 데이터 세트의 형식으로 변환하는 'Convert GT' 기능, Crowd Index를 다시 계산하는 기능과 레이블이 생성된 관절의 개수를 갱신하는 기능, 그리고 다양한 기준으로 지상 실측 데이터를 필터링하는 기능 등을 포함하고 있음을 보여주고 있다. [그림 4]의 (b)는 주석이 생성된 관절의 최소 개수로 바운딩 박스를 필터링하기 위한 조건 입력 창이고, (c)는 군중 밀집 정도를 표현한 Crowd Index 값으로 이미지를 필터링하기 위한 조건 입력 창이고, (d)는 각 이미지 내의 바운딩 박스의 최소/최대 개수로 이미지를 필터링하기 위한 조건 입력 창이고, (e)는 바운딩 박스의 최소/최대 크기로 바운딩 박스를 필터링하기 위한 조건 입력 창의 예를 표시하고 있다. 이러한 조건을 조합해서 특정 환경 및 목적에 맞는 다중 사람 자세 인식기 개발을 위한 바운딩 박스 만을 포함하는 지상 실측 데이터로 필터링 후 저장하여 해당 목적을 위한 학습 및 성능 평가의 목적으로 사용할 수 있도록 하였다. 따라서 개발된 도구를 사용하면 다목적의 유용한 사람 자세 주석 데이터를 만들 수 있을 것이다.

IV. Effects of Annotation Refinement

본 절에서는 기존 CrowdPose 테스트 데이터의 주석 원본과 본 연구에서 교정한 수정본을 비교하여 수정본의 개선된 통계를 정리하고 증진된 유용성을 분석하였다.

[표 1]은 기존 CrowdPose 테스트 데이터 원본의 정보를 제시하고 있다. 전체 이미지의 76.33%에 해당하는 6,106장에 포함된 사람의 수가 1~5명으로 이미지 당 평균적으로 2.58명의 바운딩 박스만이 포함되어 있으며, 바운딩 박스의 수가 11개 이상인 경우로 많은 사람이 출현하는 이미지는 934장으로 전체 중에 11.68%에 불과하다. 이미지를 출현 사람의 수를 기준으로 3개의 그룹으로 나눈 후, 기존 연구와의 직접적인 비교를 위해서 [12]에서 사용한 군중 밀도에 따른 이미지 분류 체계를 사용하여 각 경우를 세분하였다. 즉, Crowd Index를 기준으로 0.1 이하인 경우는 가림이 거의 없는 상황으로서 다중 사람 자세 인식이 쉬운 'easy'로 표시하였고, 0.1~0.8 사이를 보통의 경우인 'medium'으로, 그리고 0.8 이상의 경우를 가림이 많아서 인식이 어려운 상황으로 'hard'로 구분하였다. 유용성 향성을 비교 분석하기 위해, 원본 주석을 이용한 자세 인식

Table 1. Statistics of the Original CrowdPose Test Data

	1~5 (6,106 images)			6~10 (960 images)			11~20 (934 images)		
	easy	medium	hard	easy	medium	hard	easy	medium	hard
# of images	837	4,246	1,023	40	709	211	32	749	153
# of persons	1,292	11,584	2,898	314	5,439	1,626	419	9,801	1,910
Avg persons/image	1.54	2.73	2.83	7.85	7.67	7.71	13.09	13.09	12.48
Avg Crowd Index	0.01	0.45	0.95	0.04	0.47	1.08	0.03	0.43	1.07
AP by Pose Estimator[12]	0.701	0.613	0.456	0.569	0.540	0.386	0.628	0.515	0.407
AR by Pose Estimator[12]	0.744	0.661	0.522	0.617	0.584	0.441	0.672	0.567	0.460

의 결과와 수정된 버전의 인식 결과를 비교하기 위한 목적으로 [12]에서 제시한 최적의 다중 사람 자세 인식기를 이용하여 인식의 정확도를 비교하였다.

[표 1]에 의하면 이미지에 출현 사람의 수가 1~5명으로 적고 군중 밀도를 표시한 Crowd Index도 0.1 이하로 낮은 'easy' 경우의 837장 이미지를 대상으로 [12]의 자세 인식기를 이용한 실험에서 평균 정확도 AP는 0.701로 비교적 높았지만, 출현 사람의 수가 11~20명으로 많고 Crowd Index가 0.8 이상으로 'hard' 경우의 153장 이미지에 대해서는 0.407의 AP를 보였다.

기존 CrowdPose 테스트 데이터를 검수한 결과 513개의 'IS_CROWD'으로 표시하여 개별 사람을 구분하지 않은 영역이 포함되었으며 이를 제외하면, 총 34,770개의 바운딩 박스에 486,780개의 관절 레이블이 포함되어 있었다. 각 사람의 관절 레이블의 개수를 표시하는 'num_keypoints'의 오류를 수정한 바운딩 박스가 32,243개나 되었으며 원본 데이터에서는 'num_keypoints'가 0인 바운딩 박스가 5,648개이고 수정한 이후에도 0으로 남아 있는 바운딩 박스가 3,696개가 포함되어 있었다. 이는 군중 이미지에서 사람의 바운딩 박스 레이블을 표시하였으나 관절 정보를 하나도 포함하고 있지 않은 상황을 의미하기 때문에 다중 사람의 자세 인식을 위한 성능 평가 데이터로는 사용하기에는 부적절함을 예시하는 상황이다.

본 연구에서는 이러한 문제점을 내포하고 있는 기존 CrowdPose 테스트 데이터를 개발된 지상 실측 주석 개선 도구를 이용하여 수정하였다. 우선 사람을 표시하지 않은 'IS_CROWD' 바운딩 박스와 동물, 동상 등에 표시된 바운딩 박스를 삭제하였으며 'num_keypoints'가 0인 5,648개의 바운딩 박스 중에 1,952개에는 관절 레이블을 추가하고 'num_keypoints'를 수정하여 성능 평가 데이터로 사용할 수 있도록 하였고 최종적으로 관절 레이블을 포함하고 있지 못한 3,696개의 바운딩 박스는 제거하였다.

추가로 군중 이미지에서 최소 4개의 관절을 찾아 레이블을 추가할 수 있는 사람을 최대한으로 찾아서 바운딩 박스를 추가했고, 부정확한 기존 바운딩 박스와 관절의 레이블 정보를 수정하였다. 이러한 검수 및 수정 작업을 통해 만든 수정 데이터 세트에는 총 35,993개의 바운딩 박스에 503,902개의 관절 정보가 포함되어 있다.

[표 2]는 개발된 주석 개선 도구를 사용하여 기존 CrowdPose 데이터를 수정했을 때 변경된 내용을 요약하고 있다. 각 이미지에 출현한 사람의 바운딩 박스가 증가하여 기존 그룹에서 출현 사람이 더 많은 상위 그룹으로 소속이 변경된 이미지가 204장이고, 반대로 상위 그룹에서 하위 그룹으로 소속이 변경된 이미지가 287장이 존재해서 출현 사람 기준의 소속 그룹이 변경된 이미지가 총 491장이다. 특히, 기존 1-5그룹에서 1개의 이미지와 11-20그룹에서 64개 이미지가 각각 새로 분리 구분한 21-100그룹으로 이동했으며, 이 새로 분리한 그룹에는 이미지 당 평균 34.38명이 출현해서 기존 테스트 데이터보다 더 많은 사람을 포함하고 있음을 알 수 있다.

[표 3]은 수정된 CrowdPose 테스트 데이터의 기본 정보와 이를 대상으로 [12]의 자세 인식기를 사용해서 실시한 성능 평가의 결과를 예시하고 있다. 수정된 데이터 세트에서 출현 사람의 수가 1-5명이고 'easy' 경우(Crowd

Table 2. Varying Classification from the Original CrwodPose Test Data to the Modified One

From the original GT	1~5 (6,106 images)				6~10 (960 images)				11~20 (934 images)			
To the modified GT	1-5	6-10	11-20	21-100	1-5	6-10	11-20	21-100	1-5	6-10	11-20	21-100
# of images	5,968	130	7	1	158	800	2	0	6	226	638	64

Table 3. Statistics of the Modified CrowdPose Test Data

	1~5 (6,132 images)			6~10 (1,156 images)			11~20 (647 images)			21~100 (65 images)		
	easy	medium	hard	easy	medium	hard	easy	medium	hard	easy	medium	hard
# of images	833	4,047	1,252	37	807	312	28	471	148	3	57	5
# of persons	1,343	11,444	3,730	297	6,122	2,376	379	6,221	1,846	118	1,968	149
Avg persons/image	1.61	2.83	2.98	8.03	7.59	7.62	13.54	13.21	12.47	39.33	34.53	29.80
Avg Crowd Index	0.01	0.45	1.01	0.05	0.46	1.11	0.07	0.45	1.13	0.08	0.43	1.37
AP by Pose Estimator[12]	0.752	0.693	0.538	0.403	0.536	0.476	0.318	0.49	0.49	0.129	0.26	0.537
AR by Pose Estimator[12]	0.779	0.728	0.584	0.420	0.569	0.517	0.339	0.524	0.526	0.127	0.275	0.565

Index ≤ 0.1)에 속하는 833장의 이미지를 대상으로는 평균 정확도 AP가 0.752이고, 사람의 수가 11-20명이고 'hard' (Crwod Index ≥ 0.8)에 포함되는 148장의 이미지에 대해서는 0.49의 AP를 보였으며 이는 [표 1]의 원본 데이터를 이용한 실험의 평균 정확도 0.701과 0.407보다 높아졌다. 또한 새롭게 분리 추가한 21-100그룹에 포함된 65장의 이미지에 대한 자세 인식 성능 평가 결과에서는 평균적으로 약 0.28 정도의 정확성을 나타낸 것으로 분석되었다. 이는 기존의 다중 사람 자세 인식기가 출현 사람의 수가 많은 군중 상황에서는 일반 상황에서보다 상대적으로 다중 사람의 자세를 정확하게 인식하고 있지 못함을 보여주는 상황으로 해석할 수 있다.

[그림 5]는 CrowdPose 테스트 데이터 중에서 주석 교정의 효과를 대표할 수 있는 이미지 샘플에 대해서 원본 주석과 수정된 주석을 이용하여 최근 연구인 [12]에서 제시한 최적의 다중 사람 자세 인식기를 사용하여 평가한 인식 정확도의 비교 결과를 예시하고 있다. [그림 5] (a) 이미지를 위한 원본 주석에는 군중 영역을 표시하는 'IS_CROWD' 영역을 포함하여 총 14개의 바운딩 박스가 존재하며 총 134개의 관절 레이블이 포함되어 있다. 반면에 수정된 주석에는 출현한 29명 모두에 대해 바운딩 박스 레이블이 부여되었으며 총 254개의 관절 레이블이 포함되었다. 해당 이미지를 [12]의 최적 다중 사람 관절 인식기를 사용하여 실험한 결과, 29명의 출현 사람을 모두 인식하였으며 그들의 406개 관절을 인식하여 위치를 표시하였으며, 기존의 원본 주석을 기준으로는 평균 정확도 AP가 0.584인 반면에 수정 주석을 기준으로는 0.798을 얻을 수 있음을 확인하였다. 이는 출현한 사람을 위한 자세 주석 레이블을 최대한으로 추가하여 성능 평가의 정확도가 증진된 상황을 예시하고 있다. [그림 5] (b) 이미지의 원본 주석에는 총 12개의 바운딩 박스 중에서 6개에 관절 정보가 존재하지 않았으며, 1개의 바운딩 박스는 사람이 아닌

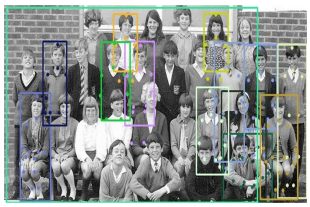
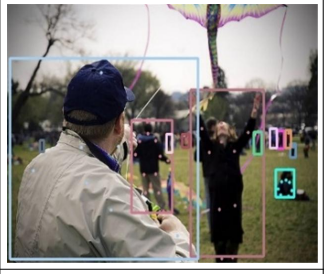
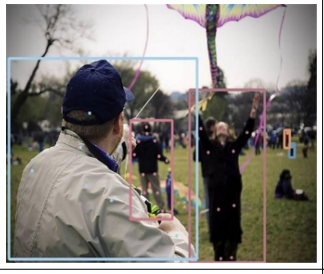
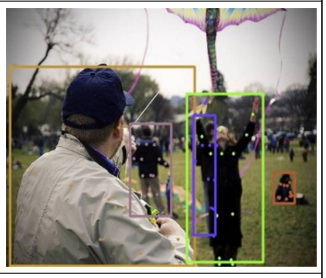
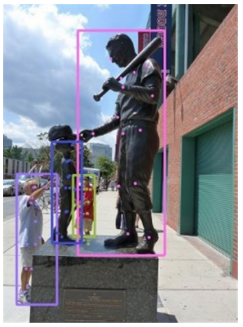
Image ID	With the original GT	With the modified GT	Estimation results
(a) 100315.jpg			
#BBox / #keypoints	14 / 134	29 / 254	29 / 406
AP / AR	0.584 / 0.585	0.798 / 0.828	-
(b) 106947.jpg			
#BBox / #keypoints	12 / 43	5 / 35	5 / 126
AP / AR	0.343 / 0.367	0.386 / 0.380	-
(c) 108357.jpg			
#BBox / #keypoints	10 / 56	5 / 30	4 / 56
AP / AR	0.473 / 0.471	0.217 / 0.320	-

Fig. 5. Performance Comparison of the Original and Modified Annotation Datasets

강아지를 위한 것이며 이 박스 내의 8개의 관절 레이블을 포함하여 총 43개의 관절이 포함되어 있었다. 잘못 레이블링 되어 있는 주석 데이터를 수정한 버전에는 5개의 바운딩 박스에 총 35개의 관절 레이블 만이 존재하였다. 다중 사람 자세 인식기로 인식하였더니 강아지 1마리와 사람 4명이 인식되었으며 총 126개의 관절이 인식되어 인식 정확도가 0.343에서 0.386 정도로 약간 개선되었다. [그림 5] (c)의 이미지를 위한 주석에는 관절이 표시되지 않은 사람 3명을 포함하고 동상 2개를 포함해서 총 10개의 바운딩 박스와 총 56개의 관절 레이블이 포함되어 있다. 잘못 레이블링 되어 있는 주석을 수정한 버전에는 5명 사람만을 위한 바운딩 박스와 총 30개의 관절 레이블이 포함되었다. 같은 사람 자세 인식기를 이용하여 동상 포함한 사람 4명을 인식하였고 56개의 관절을 인식하여 원본 주석 기준으로는 0.473의 인식 정확도를 얻었으며 수정본 주석 기준으로는 0.217의 더 낮아진 정확도를 얻었다. 이는 사용한 다중 사람 자세 인식기에서 사용한 객체 인식기가 이미지 내의 동상을 실제 사람으로 인식하는 오류로 인해서 정확도가 낮아지는 상황으로 분석되었다.

V. Conclusions

본 논문에서는 군중 상황에서의 다중 사람 자세 인식기를 개발하는 과정에서 유용하게 쓰일 지상 실측 주석 데이터를 개발하는 연구를 소개하였다. 이를 위해서 많은 사람이 출현하고 군중 밀도가 높아서 가림이 많이 발생하는 군중 상황의 이미지에 대해서도 가능한 많은 출현 사람을 위한 바운딩 박스를 포함할 수 있고 각 사람의 관절 레이블을 정확하게 표시할 수 있도록 지상 실측 주석 개선 도구를 개발하였다. 또한, 이를 이용하여 기존에 군중 상황의 다중 사람 자세 인식기의 성능 평가용으로 많이 이용되는 CrowdPose 테스트 데이터를 검수하고 교정하여 유용성을 발전시켰다. 분석 결과에 따르면 기존의 원본 주석 데이터보다 수정된 버전에 더 많은 출현 사람을 위한 바운딩 박스가 포함되었을 뿐만 아니라 각 사람의 관절 레이블도 더 많이, 더 정확하게 표시하여서 군중 상황에서 다중 사람의 자세 인식기의 성능을 정확하게 평가하는 용도로 사용할 수 있음을 확인하였다.

현재, 기존 CrowdPose의 학습 데이터 12,000장의 이미지에 대해서도 같은 방법으로 주석을 교정하고 있으며, 향후 사람의 자세 지상 실측 레이블을 포함하고 있는 [19]의 PoseTrack 데이터 세트에서 군중 상황을 대표할 수 있는 이미지와 주석 데이터를 선별 추가하고, 군중 상황의 다중 사람의 바운딩 박스를 포함하고 있는 [20]의 CrowdHuman 데이터에 각 사람의 관절 레이블을 추가하여 더욱 풍부한 다중 사람의 자세 인식 지상 실측 데이터로 발전시키려고 한다.

ACKNOWLEDGEMENT

“This research was supported by the MSIT (Ministry of Science, ICT), Korea, under the National Program for Excellence in SW, supervised by the IITP (Institute of Information & communications Technology Planing & Evaluation) in 2024”(2022-0-01127).

REFERENCES

[1] G. Papandreou, T. Zhu, N.Kanazawa, A. Toshev, J. Tompson, C. Bregler, and K. Murphy, “Towards Accurate Multi-person Pose Estimation in the Wild,” ArXiv:1701.01779v2, Apr. 2017.

[2] H. Fang, S. Xie, Y. Tai, and C. Lu, “RMPE: Regional Multi-Person Pose Estimation,” Proceedings of IEEE International Conference on Computer Vision, Oct. 2017. DOI: 10.1109/ICCV.2017.256

[3] B. Xiao, H. Wu, and Y. Wei, “Simple Baselines for Human Pose Estimation and Tracking,” ArXiv:1804.06208v2, Aug. 2018.

[4] M. Toshpulatov, W. Lee, S. Lee, and A. H. Roudsari, “Human Pose, Hand and Mesh Estimation using Deep Learning: A Survey,” The Journal of Supercomputing, Vol. 78, No. 6, Apr. 2022. DOI: 10.1007/s11227-021-04184-7

[5] M. R. Ronchi, P. Perona, “Bechmarking and Error Diagnosis in Multi-Instance Pose Estimation,” Proceedings of IEEE International Conference on Computer Vision, pp. 369-378, 2017, DOI: 10.1109/ICCV.2017.487

[6] C. Zheng, W. Wu, C. Chen, T. Yang, S. Zhu, J. Shen, N. Kehtarnavaz, and M. Shah, “Deep Learning-Based Human Pose Estimation: A Survey,” ACM Computing Surveys, Vol. 56, No. 11, pp. 1-37, 2018. DOI:10.1145/3603618

[7] H. Fang, J. Li, H. Tang, C. Xu, H. Zhu, Y. Xiu, Y. Li, and C. Lu, “AlphaPose: Whole-Body Regional Multi-Person Pose Estimation and Tracking in Real-TIme,” IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 45, No. 6, pp. 7157-7173, 2023.

[8] M. Zhou, L. Stoffl, M. W. Mathis, A. Mathis, “Rethinking Pose Estimation in Crowds: Overcoming the Detection Information Bottleneck and Ambiguity,” Prceedings of IEEE/CVF International Conference on Computer Vision, pp. 14689-14699, 2023. DOI:

10.1109/ICCV51070.2023. 01350

[9] S. Shao, Z. Zhao, B. Li, T. Xiao, G. Yu, X. Zhang, and J. Sun, "CrowdHuman: A Bechmark for Detecting Human in a Crowd," ArXiv:1805.00123v1, Apr. 2018.

[10] J. Li, C. Wnag, H. Zhu, Y. Mao, H. S. Fang, C. Lu, "CrowdPose: Efficient Crowded Scences Pose Estimation and A New Bechmark," Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 10855-10864, 2019. DOI: 10.1109/CVPR.2019.01112.

[11] M. Bendali-Braham, J. Weber, G. Forestier, L. Idoumghar, and P. Muller, "Recent Trends in Crowd Analysis: A Review," Machine Learning and Applications, Vol. 4, No. 15, 2021. DOI: 10.1016/j.mlwa.2021.100023

[12] C. Kim, Y. Jung, and Y. Kim, "Analyzing the Effects of Human Detection in Top-down Pose Estimation for Crowd Situation Recognitions," Proceedings of International Conference on Computer Science and its Applications(Part of LNEE 1190), Dec. 2023.

[13] M. Andriluka, L. Pishchulin, P. Gehler, and B. Schiele, "2D Human Pose Estimation: New Benchmark and State of the Art Analysis," Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, June 2014. DOI: 10.1109/CVPR.2014. 471

[14] T. Y. Lin, M. Maire, S. Belongie, L. Bourdev, R. Girshick, J. Hays, P. Perona, D. Ramanan, C. L. Zitnick, and P. Dollar, "Microsoft COCO: Common Objects in Context," Proceedings of European Conference on Computer Vision, pp. 740-755, 2014. DOI: 10.1007/978-3-319-10602-1_48

[15] J. Wu, H. Zheng, B. Zhao, Y. Li, B. Yan, R. Liang, W. Wang, S. Zhou, G. Lin, Y. Fu, Y. Wang, and Y. Wang, "AI Challenger: A Large-scale Dataset for Going Deeper in Image Understanding," Proceedings of IEEE International Conference on Multimedia and Expo, July 2019. DOI: 10.1109/ICME.2019.00256

[16] Justin Brooks, COCO Annotator, https://github.com/jsbroks/coco-annotator.

[17] Supervisely, Supervisely, https://supervise.ly.

[18] P. Perona, "Vision of a Visipedia," Proceedings of the IEEE, Vol. 98, No. 8, pp. 1526-1534, Aug. 2010. DOI: 10.1109/JPROC. 2010.2049621

[19] A. Doering, D. Chen, S. Zhang, B. Schiele, and J. Gall, "PoseTrack21: A Dataset for Person Search, Multi-Object Tracking and Multi-Person Pose Tracking," Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 20931-20940, 2022. DOI: 10.1109/CVPR52688.2022.02029

[20] S. Shao, Z. Zhao, B. Li, T. Xiao, G. Yu, X. Zhang, and J. Sun, "CrowdHuman: A Benchmark for Detecting Human in a Crowd," ArXiv:1805.00123v1, Apr. 2018. DOI: 10.48550/arXiv.1805. 00123

Authors

Jin-Woo Cha is a undergraduate student of Department of Computer Information at Inha Technical College. He was a member of KMS Lab at Inha University. He is interested in computer vision and machine learning.

Chulyoung Kim received a B.S. degree in Information and Communication Engineering from Inha University in 2021. He is a graduate student of Department of Electronics and Computer Engineering at Inha. His research interests are computer visions and image processing, especially small object detection in edge computing.

Hyun-Jong Oh is a undergraduate student of Department of Information and Communication Engineering at Inha University. As a member of KMS Lab, he is doing researches on computer visions and analyzing crowd behaviors.

Da-Jeong Seo is a undergraduate student of Department of Artificial Intelligence Engineering at Inha University. As a member of KMS Lab, she is doing researches on computer visions and big-data analysis.

Jong-Seong Park is a undergraduate student of Department of Information and Communication Engineering at Inha University. As a member of KMS Lab, he is doing researches on computer visions and analyzing crowd behaviors.

Yoo-Sung Kim received the B.S. degree in Computer Science from Inha University, Korea, in 1986, his M.S. and Ph.D. degrees in Computer Science from Korea Advanced Institute of Science and Technology(KAIST), Korea, in 1988, and 1992, respectively. Dr. Kim joined the faculty of the Department of Artificial Intelligence at Inha University, Incheon, Korea, in 1992. He is interested in data intelligence, machine learning, and intelligent software systems.