

딥러닝을 활용하여 자세 추정에 대한 새로운 접근법에 관한 연구

구종회, 김미경, 차의영*
부산대학교

ninebell@pusan.ac.kr, ddosun@pusan.ac.kr *eycha@pusan.ac.kr

A Study on the New approach about Human Pose Estimation with Deep Learning

Ku Jong Hoe, Kim Mi Kyung, Cha Eui Young*
Pusan National Univ.

요 약

본 논문은 여러 관절을 묶어 사람을 나타내는 새로운 연결 방법을 제안한다. Open Pose[1]와 같은 기존의 관절 인식 방법은 발, 무릎, 골반과 같이 서로 간의 연관성을 기반으로 관절을 묶어 한 사람을 나타낸다. 서로 간의 연관성을 기반으로 관절을 묶을 때 오검출된 관절이 있을 시 연관이 있는 관절도 사용하지 않게 된다. 이 점을 해결하고자 관절과 사람의 중심점을 이용하여 관절을 한 사람으로 묶어내는 방법을 제안한다.

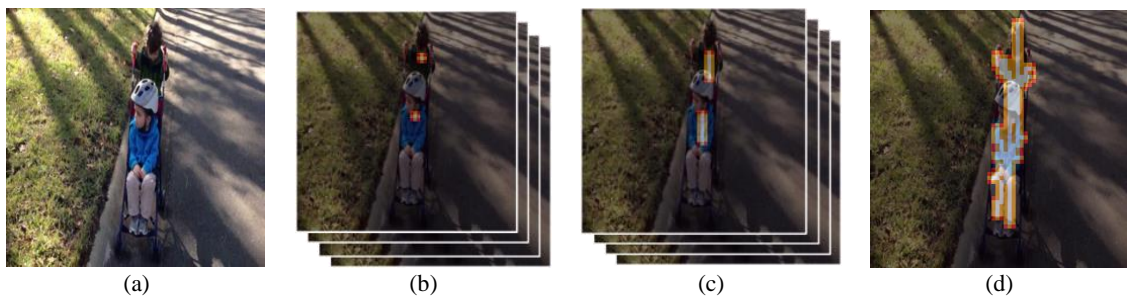


그림 1. (a) 입력 이미지 (b) 예측 관절 (c) 예측 연결도 (d) 관절 검출 결과

I. 서론

딥러닝을 이용한 인체 인식 연구는 많은 관심을 받고 계속해서 진행되고 있다. 인체 인식 중 관절 인식은 제한된 환경에서 관절을 찾거나 관절을 표시하는 장치를 이용하여 검출하였다. 현재는 카메라만을 이용하여 일반적인 환경에서 사람들의 관절을 검출하는 방법에 대하여 연구[1], [2] [3], [4]가 이뤄지고 있다. 이 연구는 최근 뜨고 있는 VR/AR 에서 입력 시스템의 기반이 되고 애니메이션, 영화에서 CG 처리를 위한 용도 등 많은 분야에서 활용된다.

딥러닝을 이용한 다수의 사람에 대한 관절 검출 방법은 Open Pose[1]에서 처음으로 제안하였다. 관절의 위치를 표시하는 Confidence map 과 관절과 관절을 이어주는 뼈를 표시하는 Affinity field 를 이용하여 여러 관절에 대하여 한 사람으로 묶어준다. Confidence map 에서 찾아진 관절 간 연관성을 affinity field 를 통해 연산하여 가장 연관성이 높은 관절들을 서로 이어주는 방법이다. 관절 간의 연관성이 있을 시 하나의 관절 검출이 실패하게 되면 연관이 있는 관절을 검출할 수 없는 큰 문제점이 있다.

본 논문은 이러한 관절의 연관성 의존으로 발생하는 문제를 해결하기 위한 새로운 관절 검출 방식을 제안한다. 기존에 연관성을 기준으로 관절을 묶는 방식이

아닌 사람의 중심점을 찾고 해당 중심점과 관절을 연결하는 center connection map 을 이용한다. 기존에 affinity field 가 관절과 관절의 방향을 나타냈다면 center connection map 은 관절과 사람의 중심점 관계를 나타낸다. 이를 통해 관절 간 연관성이 없어져 오검출로 관절을 검출하지 못해도 다수의 관절을 한 사람으로 묶어내는데 영향을 받지 않고 검출된 관절을 묶을 수 있다.

II. 본론

본 논문의 목표는 MPII Human pose Dataset[5]에서 정의된 관절 17 개에 대하여 단일 RGB 이미지에서 복수의 사람들에 대한 2D 좌표 인식이다. 그림 1. 과같이 입력 이미지를 받아 딥러닝 모델을 통해 Joint Confidence Map, Center Connection Map 을 얻는다. 학습에 사용한 모델은 n-stacked hourglass[6]를 사용하였다. 3 개의 Hourglass 를 사용하였고 각 Hourglass 는 5 번의 다운 샘플링과 업 샘플링을 한다. 기존의 Hourglass 의 bottle neck module 을 Convolution Block Attention Module(CBAM)[2]을 이용한 bottle neck module 로 변경하였다. 모델의 구조는 아래의 그림 2. 와 같다. 모델을 학습시키기 위해 MPII Human pose Dataset[5]을 이용하였다.

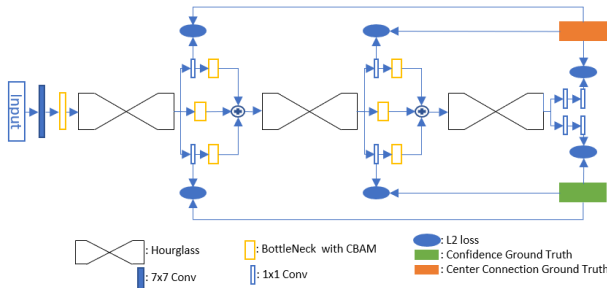


그림 2. 제안 모델 구조

Confidence map 의 Ground truth 는 2D Gaussian kernel 을 이용하여 관절의 위치가 극값을 가지도록 만든다. Center connection map 의 ground truth 는 관절과 사람의 중심점 사이를 1 의 값을 가지게 만든다. 사람의 중심점은 해당 사람의 모든 관절의 평균값을 이용한다. 학습 시, CBAM[2]의 유무에 따라 초기 학습이 차이가 있었다. 학습 이후 confidence map 과 center connection map 을 이용하여 관절을 한 사람으로 묶는 추론 과정을 거친다. 추론 과정에서는 예측된 사람의 관절(d_{jn})과 중심점(d_{cm})의 사이를 center connection map($C_j(p)$)을 이용해 확률을 계산하여 가장 높은 확률을 가지도록 중복 없이 중심점과 관절을 묶어준다. 확률을 계산하는 방법은 아래 식 (1), (2)과 같이 계산한다.

$$E_{jncm} = \frac{1}{\|d_{jn} - d_{cm}\|_2} \int_{u=0}^{u=1} C_j(p(u)) du \quad (1)$$

$$p(u) = (1 - u)d_{jn} + ud_{cm} \quad (2)$$

기존의 논문들은 방향을 이용한 벡터 내적 계산을 통해 관절을 연결하지만, 해당 논문은 벡터 내적 연산이 아닌 connection map 에서의 확률값으로 계산한다.

학습에 사용된 이미지는 총 13,636 개의 이미지를 사용하였고 L2 loss 함수와 Adam optimizer 를 사용하여 300 Epoch 로 학습을 진행하였다. 학습 결과는 그림 3. 과 같다.

IV. 결론

본 논문은 관절의 연관성 없이 다수의 사람에 대한 관절 검출을 연구하였다. 추론 결과인 그림 3.을 봤을 때 학습의 ground truth 와는 유사하게 나오지만, 추론 과정에서 connection map 의 현재의 확률값을 이용한 방식에 단점이 있었다. 기존의 벡터 내적 방식은 방향만을 예측하면 되지만 현재 방식은 방향과 확률값까지 예측해야 하는 단점이 있었다. 또한 특정 부분의 값이 크게 나타나면 해당 위치로 연결이 강제되는 단점이 있다. 이러한 단점만 보완하면 관절의 연관성이 필요 없는 더 나은 자세 추정을 수행할 수 있을 거라 예상한다.

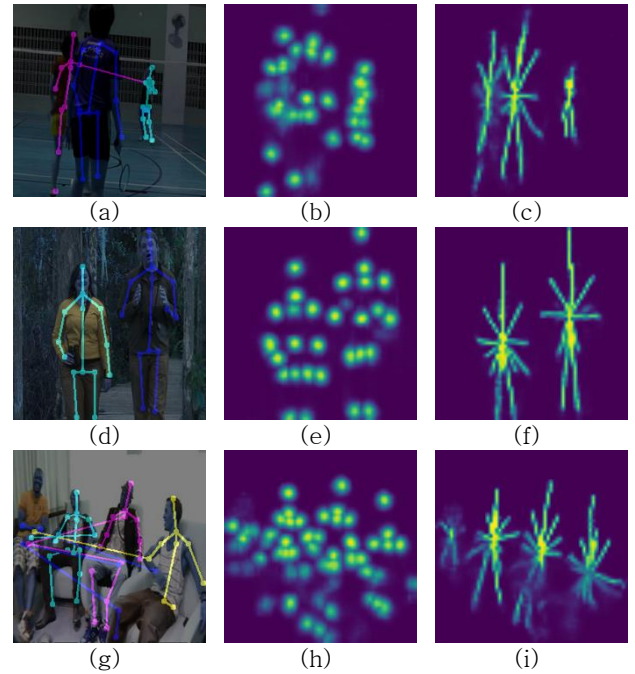


그림 3. (a), (d), (g) 결과 이미지,
(b), (e), (h) joint prediction 예측 이미지,
(c), (f), (i) connection map 예측 이미지

참 고 문 헌

- [1] CAO, Zhe, et al. Realtime multi-person 2d pose estimation using part affinity fields. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2017. p. 7291-7299.
- [2] WOO, Sanghyun, et al. Cbam: Convolutional block attention module. In: Proceedings of the European conference on computer vision (ECCV). 2018. p. 3-19.
- [3] LAW, Hei; DENG, Jia. Cornernet: Detecting objects as paired keypoints. In: Proceedings of the European Conference on Computer Vision (ECCV). 2018. p. 734-750.
- [4] ZHOU, Xingyi; WANG, Dequan; KRÄHENBÜHL, Philipp. Objects as points. arXiv preprint arXiv:1904.07850, 2019.
- [5] ANDRILUKA, Mykhaylo, et al. 2d human pose estimation: New benchmark and state of the art analysis. In: Proceedings of the IEEE Conference on computer Vision and Pattern Recognition. 2014. p. 3686-3693.
- [6] NEWELL, Alejandro; YANG, Kaiyu; DENG, Jia. Stacked hourglass networks for human pose estimation. In: European conference on computer vision. Springer, Cham, 2016. p. 483-499.