관절 적응형 Gaussian Mixture 히트맵 회귀법을 이용한 하향식 사람 자세 추정에 관한 연구

왕준기[O], 조정찬[**], 최상일[*]
[O]단국대학교 데이터지식서비스공학과,
[**]가천대학교 AI・소프트웨어학부,
[*]단국대학교 컴퓨터공학과
e-mail: jaden6171998@gmail.com[O], thinkai@gachon.ac.kr[**], choisi@dankook.ac.kr[*]

Study of the Gaussian Mixture Joint-Adaptive Heatmap Regression for Top-Down Human Pose Estimation

Ong Zhun-Gee[O], Jungchan Cho[**], Choi Sang-il[*]
[O]Dept. of Data and Knowledge Service Engineering, Dankook University,
[**]School of Computing, Gachon University,
[*]Dept. of Computer Engineering, Dankook University

● 요 약 ●

본 논문은 딥러닝 사람 자세 추정 모델이 사람의 관절 키포인트를 예측하는데 관절의 2차원 면적에 의해 키포인트별 σ, 즉, 표준 편차를 가지는 가우시안 커널(Gaussian Kernel)을 예측하는 방법을 제안한다. 각 관절 키포인트에 대해 다른 σ를 가지는 정답 히트맵(Ground Truth Heatmap)과 제안한 Gaussian Mixture Block를 모델에 추가해서 관절의 크기를 맞는 히트맵을 예측한다.

키워드: 사람 자세 추정(Human Pose Estimation), 히트맵 회귀(Heatmap Regression)

I. Introduction

Heatmap Regression Human Pose Estimation(HPE)은 주어진 RGB 이미지에서 사람의 키포인트(keypoint)를 예측하는 것이다. K개의 키포인트에 대해서 K개의 히트맵을 예측하는 것이 목표이며 정답 히트맵과 예측된 히트맵 간의 Euclidean distance를 최소화하도록 학습한다. 정답 히트맵은 키포인트의 실제 좌표를 중심으로 하는 2차원 가우시안 커널(Gaussian Kernel)에 의해 생성되며 모든 가우시안 커널은 동일한 σ(표준 편차)를 가지고 있다. 히트맵 H_i의 픽셀값들은 해당 위치가 키포인트 i일 확률로 나타낸다.

본 논문은 하향식 HPE에 대한 키포인트 적응형 히트맵 회귀법을 집중하고 있다. 각 관절에 대한 정답 히트맵은 σ가 다른 가우시안 커널로 생성되며 모델은 키포인트의 종류에 따라 적합한 Gaussian mixture 히트맵을 예측하는 것이 목표다. 이 논문에서는 실험결과로부터 Gaussian Mixture 히트맵이 모델 성능에 대한 영향을 보여준다.

II. Preliminaries

2.1 하향식 사람 자세 추정 (Top-Down HPE)

하향식 HPE는 주어진 입력 이미지에서 "사람"이란 객체를 사람 탐지기(human detector)가 검출하고 잘라내서 HPE 모델의 입력 크기와 동일한 크기로 조정하는 다음에 HPE 모델을 통과하여 키포인트를 예측한다.

2.2 관절 적응형 히트맵 회귀법

상향 식 HPE인 SWAHR[1] 논문에서 scale map의 학습을 통해서 입력 이미지에서 사람의 비율과 키포인트의 종류에 맞는 히트맵을 예측했다. 이에 성능을 개선되어 그 당시에 SOTA를 달성하였으며 관절 적응형 히트맵의 가능성을 보여준다.

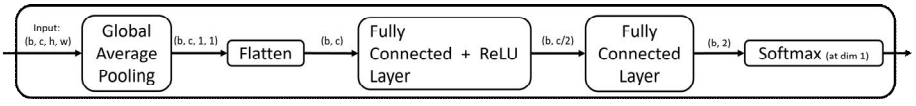

Fig. 1. Gaussian Mixture Block

Table 1. Sigma Set of Gaussian Kernel

관절 종류	가우시안 커널의 σ	
	Set 1	Set 2
코	1.1	0.85
눈	1.1	0.84
귀	1.3	0.99
어깨	1.9	1.49
팔꿈치	1.8	1.42
손목	1.7	1.32
엉덩이	2.2	1.73
무릎	2.0	1.56
발목	2.0	1.58

Table 2. Experiment Result

σ	AP	AR
2.0	70.8%	74.0%
1.5	71.4%	74.7%
Set 1	54.9%	60.1%
Set 2	66.8%	70.2%
GM** Set 1	60.8%	65.6%
GM** Set 2	69.7%	72.8%

** GM: Gaussian Mixture

III. The Proposed Scheme

본 논문에서 사용된 모델은 Simple Baseline[2]이다. 특징 추출기(feature extractor)인 ResNet-50[4]로 추출된 특징맵은 Transpose Convolution layer를 통과하기 전에 Gaussian Mixture Block을 통과한다. Gaussian Mixture Block의 구조는 Fig. 1과 같다.

실험에서 COCO Dataset 2017 데이터셋을 사용하였다. 학습 데이터셋은 총 118287장 이미지에서 149813개 사람 인스턴스를 추출하여 구성되며 검증 데이터셋은 5000장 이미지에서 6352개 사람 인스턴스로 구성된다. Epoch, learning rate 등 실험 파라미터는 [2]와 같이 설정하고 모델을 학습하여 실험하였다. 기본 Gaussian Mixture 사용하지 않은 σ는 2.0, 1.5가 되며 Gaussian Mixture Block을 추가해서 Set 1, Set 2 σ에 대한 실험결과는 [Table 2]에 보여준다. 관절의 종류 구분 없이 σ를 동일하게 2.0이나 1.5를 사용하는 것보다 개별 설정하는 것은 성능이 떨어지는 것을 확인하였다.

IV. Conclusions

본 논문에서는 Simple Baseline 모델을 Gaussian Mixture Block을 추가하여 히트맵을 예측하는 결과를 보여준다. 키포인트 적응형 σ의 사용을 인해 모델의 학습이 어려워지고 성능이 떨어지는 것을 보여준다. 추후 연구에서는 성능에 악영향 없이 관절 적응형 히트맵을 잘 예측할 수 있는 네트워크 구조나 알고리즘을 연구할 예정이다.

ACKNOWLEDGEMENT

이 논문은 2022년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원 (IITP-2022-00155227, 문맥정보를 이용한 딥러닝 기반의 의료 진단에 활용 가능한 ICT-BIO 융합 기술 개발)과 2021년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원(IITP-2017-0-00091, 멀티 모달 딥러닝 기반의 바이오 헬스케어 데이터 분석 기술 개발)과 2020년도 ICT혁신인재4.0 사업의 연구결과로 수행되었음.(IITP-2020-0-01824-003)

REFERENCES

[1] Luo, Zhengxiong et al. “Rethinking the Heatmap Regression for Bottom-up Human Pose Estimation.” 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2021): 13259-13268.

[2] Xiao, Bin et al. “Simple Baselines for Human Pose Estimation and Tracking.” ECCV (2018).

[3] Lin, Tsung-Yi et al. “Microsoft COCO: Common Objects in Context.” ECCV (2014).

[4] He, Kaiming et al. “Deep Residual Learning for Image Recognition.” 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2016): 770-778.