

Simple Multi-Resolution Representation Learning for Human Pose Estimation

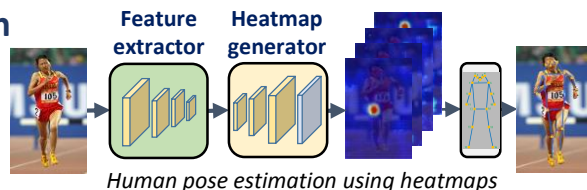
Trung Q. Tran, Giang V. Nguyen, Daeyoung Kim

Korea Advanced Institute of Science and Technology, trungq2019@kaist.ac.kr



- We can infer the wrist location thanks to elbow, shoulder, or human skeleton
- The model needs not only specific features (elbow and shoulder) but also overall patterns (human skeleton)

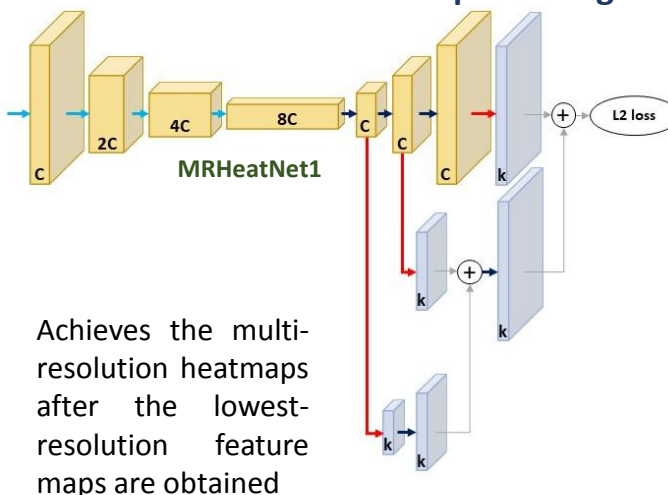
Introduction



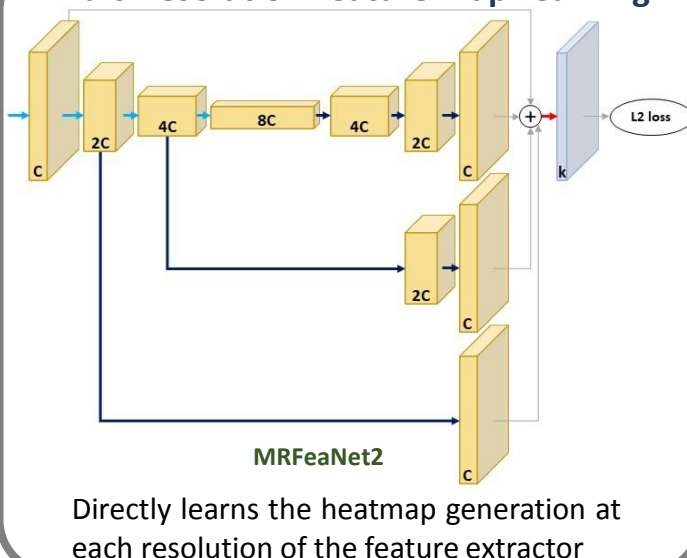
Simple architecture could be ameliorated if it learns the features from multiple resolutions
→ multi-resolution representation learning

→ Residual block → Deconvolutional layer → Convolutional layer Feature maps Heatmaps + Elem-sum

Multi-resolution Heatmap Learning



Multi-resolution Feature Map Learning

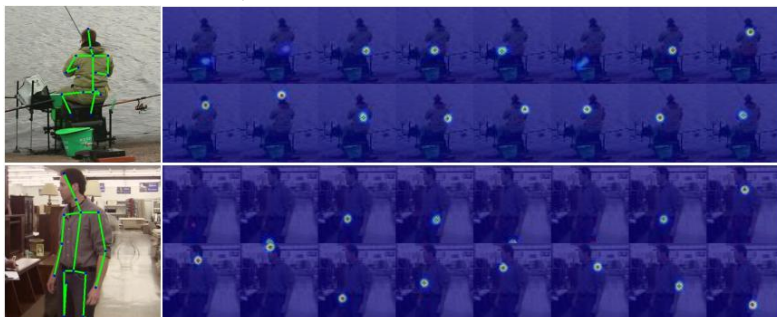


Experimental Results

Method	Backbone	AP
SimpleBaseline	ResNet-50	70.0
MRFeaNet2	ResNet-50	70.4
SimpleBaseline	ResNet-101	70.9
MRFeaNet2	ResNet-101	71.2
SimpleBaseline	ResNet-152	71.6
MRFeaNet2	ResNet-152	71.8

Results on COCO test-dev dataset

Qualitative results on MPII dataset



- Our method can make the predictions for occluded keypoints
- The heatmaps are informative and suitable, where the predictions for occluded keypoints have low confidence

Acknowledgment. This work was supported by the Korea-EU Joint Research Support Project of the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT and Future Planning of Korea (NRF-2016K1A3A7A0395205414); the MSIT(Ministry of Science and ICT), Korea, under the Grand Information Technology Research Center support program(IITP-2020-0-01489) supervised by the IITP(Institute for Information & communications Technology Planning & Evaluation); and the Technology Innovation Program(or Industrial Strategic Technology development Program, 2000682, Development of Automated Driving Systems and Evaluation) funded by the Ministry of Trade, Industry & Energy(MOTIE,Korea).