

Pose Proposal Networks

Taiki Sekii^[0000–0002–1895–3075]

Konica Minolta, Inc.

taiki.sekii@koniaminolta.com

Abstract. We propose a novel method to detect an unknown number of articulated 2D poses in real time. To decouple the runtime complexity of pixel-wise body part detectors from their convolutional neural network (CNN) feature map resolutions, our approach, called pose proposal networks, introduces a state-of-the-art single-shot object detection paradigm using grid-wise image feature maps in a bottom-up pose detection scenario. Body part proposals, which are represented as region proposals, and limbs are detected directly via a single-shot CNN. Specialized to such detections, a bottom-up greedy parsing step is probabilistically redesigned to take into account the global context. Experimental results on the MPII Multi-Person benchmark confirm that our method achieves 72.8% mAP comparable to state-of-the-art bottom-up approaches while its total runtime using a GeForce GTX1080Ti card reaches up to 5.6 ms (180 FPS), which exceeds the bottleneck runtimes that are observed in state-of-the-art approaches.

Keywords: Human pose estimation · Object detection

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

Fig. 1. Sample multi-person pose detection results by the ResNet-18-based PPN. Part bounding boxes (b) and limbs (c) are directly detected from input images (a) using single-shot CNNs and are parsed into individual people (d) (cf. § 3).

