P² Net : Augmented Parallel–Pyramidal Net for Attention Guided Pose Estimation

Luaxuan Hou, Jie Cao, Yuan Zhao, HaiFeng Shen, Jian Tang, Ran He

Center for Research on Intelligent Perception and Computing, Chinese Academy of Sciences
National Laboratory of Pattern Recognition, Chinese Academy of Sciences
University of Chinese Academy of Sciences
AI Labs, Didi Chuxing

Introduction

- Task: The target of human pose estimation is to determine the body parts and joint locations of persons in the image. Angular changes, motion blur and occlusion in the natural scenes make this task challenging, while some joints are more difficult to be detected than others.
- Challenge: In the unconstrained conditions, the visibility of keypoints is greatly affected by wearing, posture, viewing angle, background etc. The large pose variations further increase the difficulty in detection. Therefore, data augmentation plays an important role in pose estimation.

Contributions:
- To the best of our knowledge, we are the first work doing adaptive data augmentation for human pose estimation problem. The data augmentation policy is learnt to suit specific tasks and training data. Therefore the learnt model is more robust in testing. We formulate the problem of searching data augmentation policy in a differentiable form, so that the optimal policy setting can be easily updated by back propagation during training.
- We innovate two fusion structures, i.e. Parallel Fusion and Progressive Fusion, to process pyramid features from backbone network. Both fusion structures leverage the advantages of spatial information affinity at high resolution and semantic comprehension at low resolution effectively.
- We propose a refinement stage for the pyramid features to further boost the accuracy of our network. By introducing dilated bottleneck and attention module, we increase the receptive field for the features with limited complexity and tune the importance to different feature channels. Our ablation study shows this refinement stage has a obvious contribution to the overall performance.
- We conducted extensive experiments on the mainstream datasets and achieved the state-of-the-art single model performance. Remarkably, we got 77.3 AP results in the MS-COCO test-dev dataset with smaller Params and GFLOPs.