

Team MSRA Keypoints Detection

Bin Xiao¹, Dianqi Li², Ke Sun¹, Lei Zhang², Jingdong Wang¹

¹Microsoft Research Asia ²Microsoft

Outline

- Top-Down Pipeline
- Human Pose Estimation
 - General pipeline for human pose estimation
 - Our simple baseline method for human pose estimation
- Experiment Results

Top-Down Pipeline

- Person Detector
 - Re-implement Mask-RCNN
 - Backbone: Xception¹
 - Powered by FPN² and Deformable Convolution³
 - Data: COCO only
 - Performance(on COCO test-dev dataset):
 - Box AP(person): 60.9
 - Box AR(person): 72.59
- Simple Baseline Network for Human Pose Estimation

1. Chollet, Francois. "Xception: Deep Learning with Depthwise Separable Convolutions." *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2017.

2. Lin, Tsung-Yi, et al. "Feature Pyramid Networks for Object Detection." *CVPR*. Vol. 1. No. 2. 2017.

3. Dai, Jifeng, et al. "Deformable Convolutional Networks." *Proceedings of the IEEE International Conference on Computer Vision*. 2017.

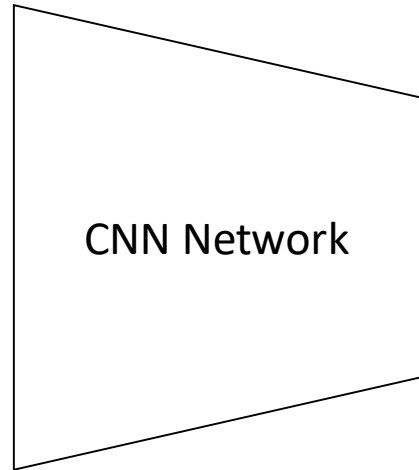
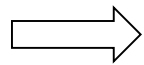
Outline

- Top-Down Pipeline
- Human Pose Estimation
 - General pipeline for human pose estimation
 - Our simple baseline method for human pose estimation
- Experiment Results

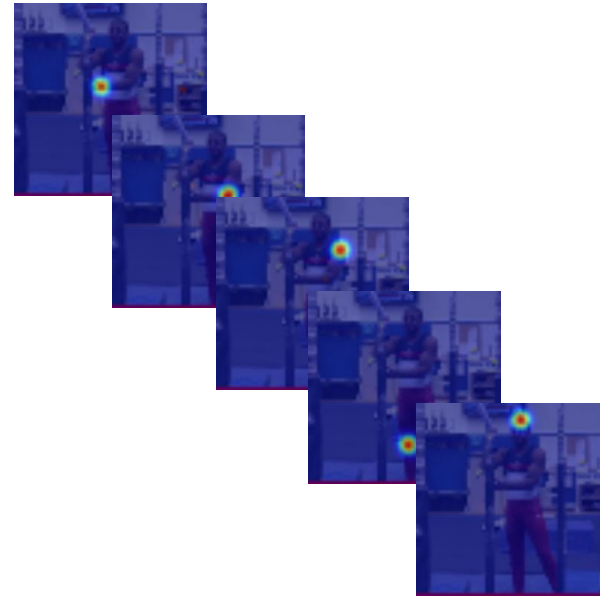
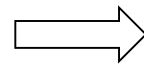
General Pipeline for Human Pose Estimation



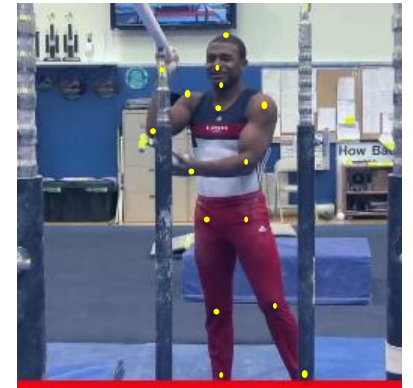
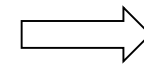
256x256



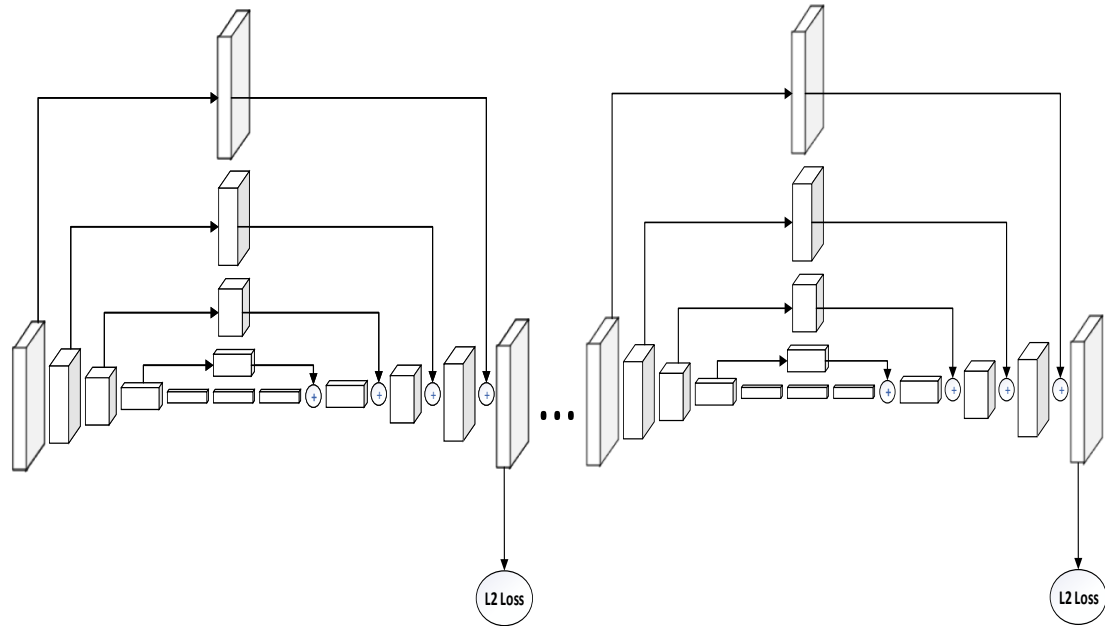
?



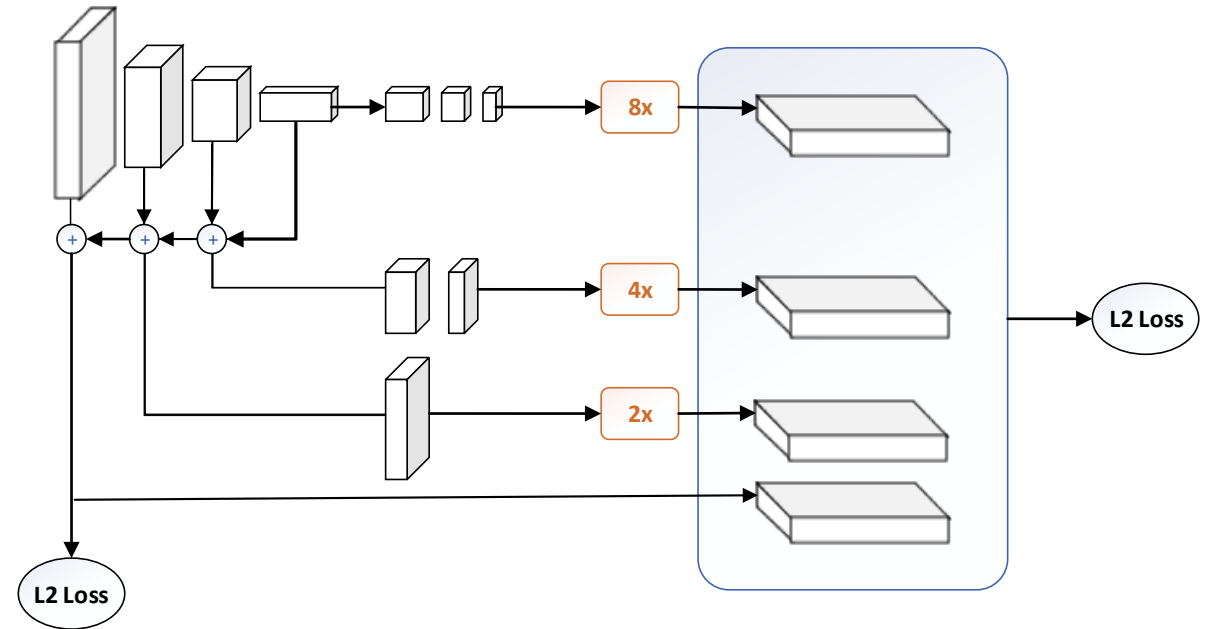
64x64



State-of-the-art networks



Stacked Hourglass



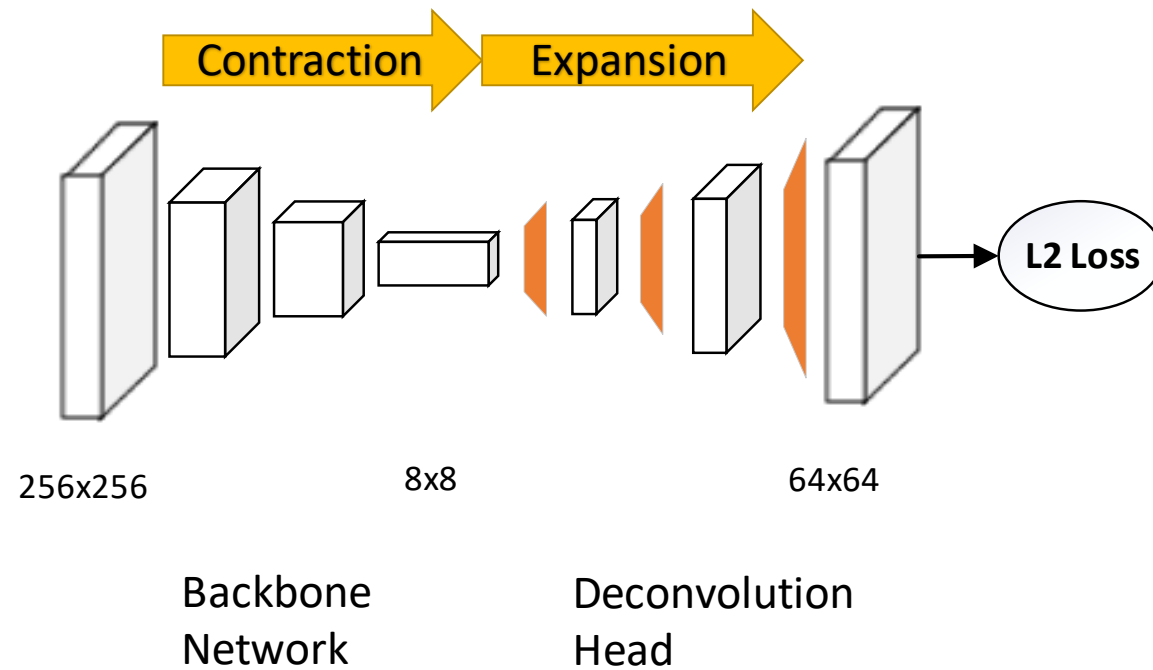
CPN

1. Newell, A., Yang, K., Deng, J.: **Stacked hourglass networks for human pose estimation**
2. Chen, Y., Wang, Z., Peng, Y., Zhang, Z., Yu, G., Sun, J.: **Cascaded pyramid network for multi-person pose estimation**

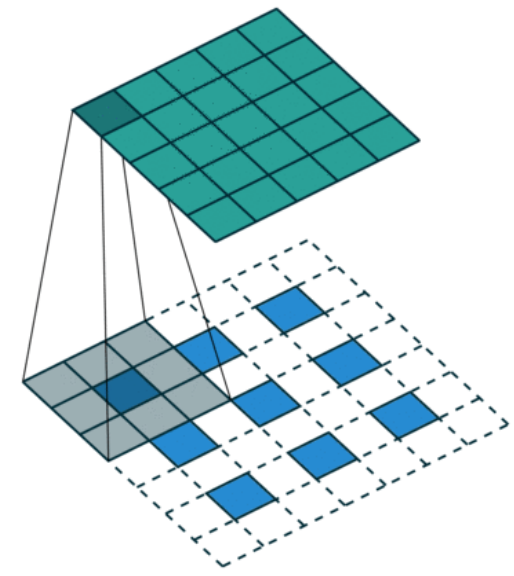
Human Pose Estimation

- How good could a simple method be?

Our simple baseline



Deconvolution Module



Summary

	Stacked Hourglass	CPN	Ours
Bilinear Up-sampling	√	√	×
Deconvolution layers	×	×	√
Skip layer feature concatenation	√	√	×
Multi-stage architecture	√	×	×
Multi-supervision	√	√	×

Outline

- Top-Down Pipeline
- Human Pose Estimation
 - General pipeline for human pose estimation
 - Our simple baseline method for human pose estimation
- Experiment Results

Results on COCO validation dataset

Method	Backbone	Input Size	AP
8-stage Hourglass	-	256x192	66.9
CPN	ResNet-50	256x192	68.6
Ours	ResNet-50	256x192	70.4 ↑3.8

COCO Submission

Method	AP(COCO test-dev)
Our Simple Baseline Network(<i>ResNet50</i>)	70.2
+256 × 192 → 384 × 288	71.3↑ 1.2
+ <i>ResNet50</i> → <i>ResNet152</i>	73.7↑ 2.4
+Training with External Dataset	75.4↑ 1.7
+Models Ensemble	76.5↑ 1.1

Summary

- A simple and effect baseline method is proposed.
- State-of-the-art results are achieved.
- We hope such baseline network would benefit the field by easing the idea development and evaluation.

Human Pose Estimation



The sequences are from PoseTrack dataset.

Teams



Bin Xiao



Dianqi Li



Ke Sun



Lei Zhang



Jingdong Wang

Thank you!



Code and Models is available at <https://github.com/Microsoft/human-pose-estimation.pytorch>