Efficient High-Resolution High-Level-Semantic Representation Learning for Human Pose Estimation

Hong Liu, Lisi Guan

Key Laboratory of Machine Perception, Shenzhen Graduate School, Peking University, China

{hongliu, guanlisi}@pku.edu.cn

Ⅰ INTRODUCTION

- **Human Pose Estimation** is a challenging task that requires to locate keypoints of the human body parts. It has wide range of applications in areas such as video analysis, intelligent surveillance, and other systems.
- However, most existing methods suffer from spatial information loss or semantic information mismatch when extracting high-resolution high-level-semantic features.
- We propose a novel Dilation Pyramid Module (DPM), which can enlarge the receptive field multiplicatively to extract high-level-semantic information as subsampling without reducing spatial resolution.

Ⅱ OBSERVATION AND CONTRIBUTION

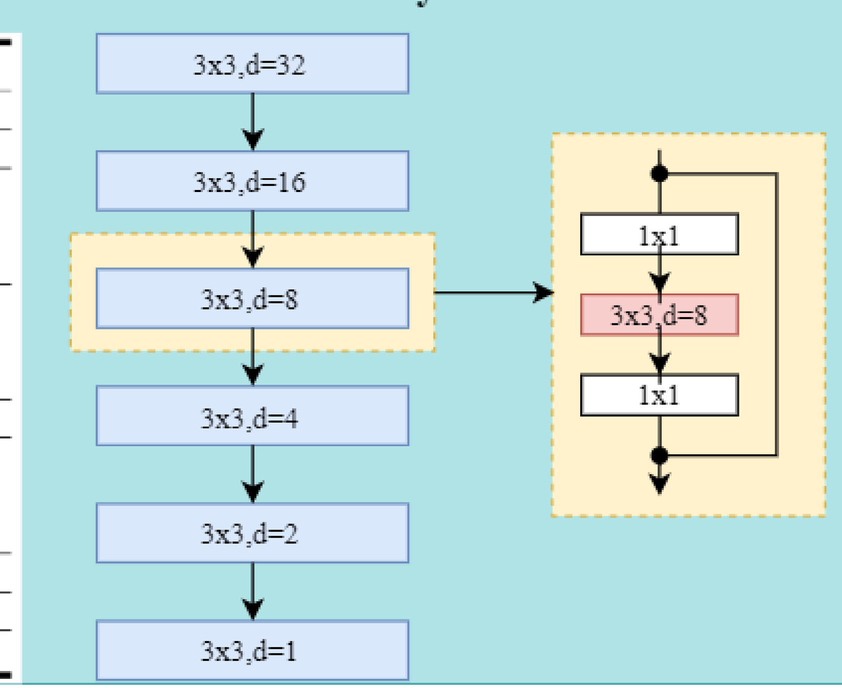
- Ensuring that features contain both high-resolution and high-level-semantic information is important for human pose estimation. Because high-resolution is helpful to reduce quantization error and high-level-semantic information is useful to catch global information.
- High-resolution and high-level-semantic features extracted by existing methods suffer from spatial information loss or semantic information mismatch, This problem is a serious impediment to performance improvement of human pose estimation.
- **Contribution**: To efficiently address these issues, we propose a novel Dilation Pyramid Module (DPM), which can enlarge the receptive field multiplicatively to extract high-level-semantic information as subsampling without reducing spatial resolution.

Ⅲ PROPOSED METHOD

Dilation Pyramid Net

layer name	output size	DPN4_32	DPN4_64
conv0	128x96	7x7,64,stride 2	7x7,64,stride 2
pooling	64x48	3x3 max pool,stride 2	3x3 max pool,stride 2
conv1	64x48	[1x1, 64; 3x3, 64; 1x1,256] ×3	[1x1, 64; 3x3, 64; 1x1,256] ×3
conv2	32x24	1x1,128,stride 2	[1x1,128; 3x3,128; 1x1,512] ×4 1x1,256,stride 2
DPM x4	32x24	[1x1,32; 3x3,32; 1x1,128] ×6	[1x1,64; 3x3,64; 1x1,256] ×6
heatmap	64x48	upsample x2 1x1,17	
#Params	-	0.7 M	3.3 M

Dilation Pyramid Module



- Dilation Pyramid Module (DPM) can enlarge receptive fields multiplicatively without spatial information loss and semantic information mismatch. DPM is composed of N consecutive dilated convolution layers, of which dilation radius is specially designed. DPM is defined as :

$$X_{out} = f_N^{d_N}(f_{N-1}^{d_{N-1}}(...(f_2^{d_2}(f_1^{d_1}(X_{in})))))$$

Where $f_i^{d_i}$ is the i_{th} dilated convolution layer and d_i is the dilation radius of it, d_i is determined by 2^{N-i}, X_{in} and X_{out} are input features and output features, respectively. The kernel size of all the dilated convolution of DPM is set to k x k. The receptive fields of DPM can be formulated as follows:

$$\begin{aligned} RF_{total} &= k + (k-1)*2^1 + ... + (k-1)*2^{N-1} \\ &= (k-1)(1+2^1+...+2^{N-1}) + 1 \\ &= (k-1)(2^N - 1) + 1, \end{aligned}$$

where RF_{total} is the total receptive fields of DPM. DPM can enlarge receptive fields multiplicatively as subsampling and keep spatial resolution unchanged. By default, the kernel size of the dilated convolution used in DPM is set to 3 x 3 for the consideration of parameter consumption and computation cost.

Ⅴ EXPERIMENTS

- **Datasets:** MS COCO human pose estimation dataset
- **Setup**: The proposed Top-Down method, DPN
- **Ablation Studies**

Comparison with Normal Convolution

		DPN1_64	DPN2_64	DPN3_64	DPN4_64
AP	DPM	**64.5**↑4.8	**68.0**↑4.0	**69.2**↑4.9	**69.9**↑5.3
	Normal	59.7	64.0	64.3	64.6

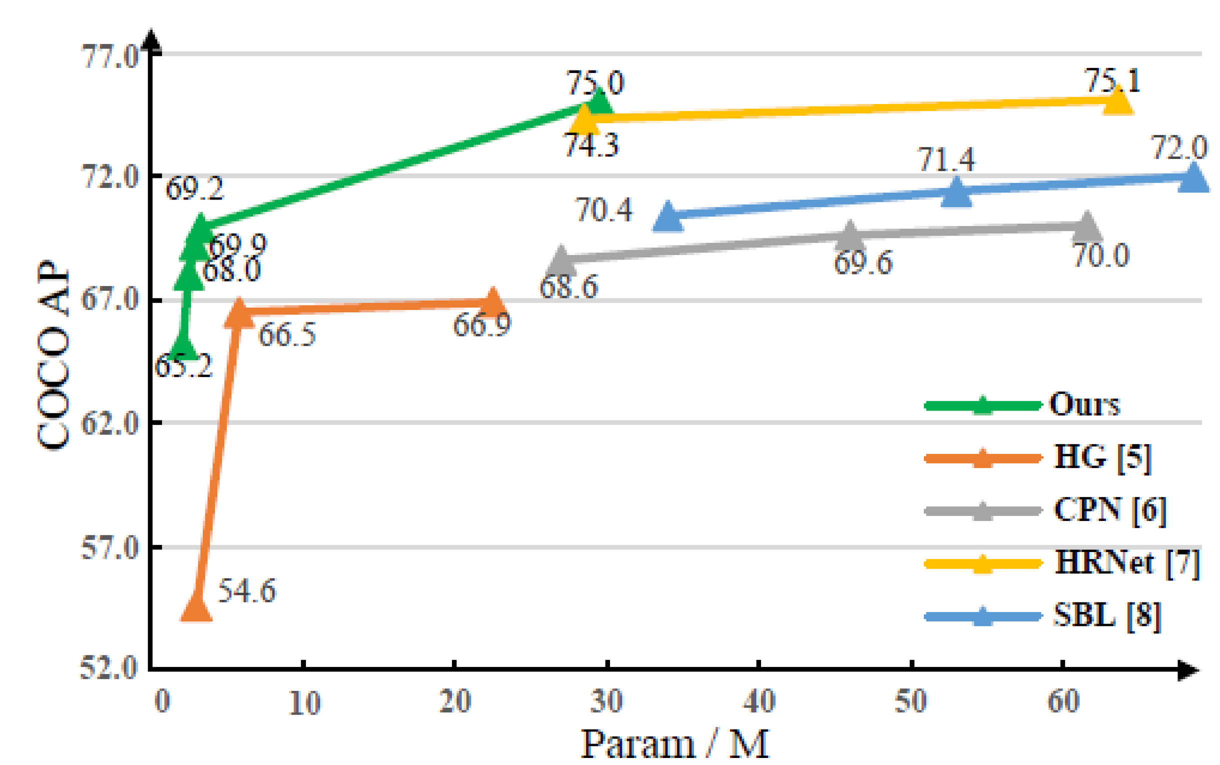
Comparison with Stacked Dilated Convolution

	DPM	D=1	D=2	D=4	D=6	D=8
DPN1_32	**50.6**	40.1	**49.3**	48.7	43.9	38.5
DPN4_32	**62.0**	58.2	**60.6**	56.6	50.0	48.7

- **Comparisons with state-of-arts**

Method	Backbone	#Params	GFLOPs	AP
OpenPose [27]	-	-	-	61.8
Mask-RCNN [28]	ResNet-50	-	-	63.1
Hourglass [7]	8-stage Hourglass	25.1 M	14.3	66.9
CPN [8]	ResNet-50	27.0 M	6.2	69.4
SBL [10]	ResNet-50	34.0 M	8.9	70.4
HRNet-w32 [9]	HRNet-w32	28.5 M	7.1	74.4
HRNet-w48 [9]	HRNet-w48	**63.6 M**	**14.6**	**75.1**
DPN4_32 (ours)	DPN	**0.7 M**	**1.1**	**62.0**
DPN4_64 (ours)	DPN	**3.3 M**	**3.0**	**69.9**
DPN2_hrnet (ours)	HRNet-w32	29.5 M	7.8	**75.0**

- **Comparisons with state-of-arts about efficiency**



- Please feel free to contact us! We are pleased to find new friends! ^ ^ (guanlisi@pku.edu.cn)