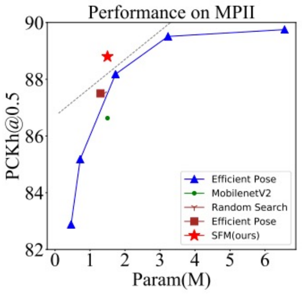
Efficient Human Pose Estimation by Maximizing Fusion and High-Level Spatial

Zhiyuan Ren[1], Yaohai Zhou[2], Yizhe Chen[2], Ruisong Zhou[3], Yayu Gao*

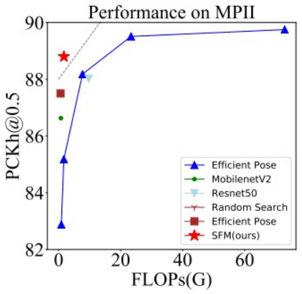
Introduction

We propose an efficient human pose estimation network—SFM (slender fusion model) by fusing multi-level features and adding lightweight attention blocks—HSA (High-Level Spatial Attention).

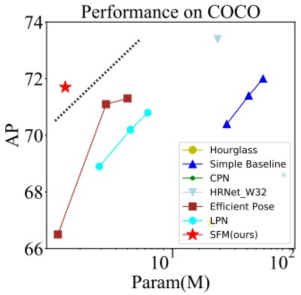
With the help of SFM and HSA, our network is able to generate multi-level feature and extract precise global spatial information with little computing resource compared to other methods



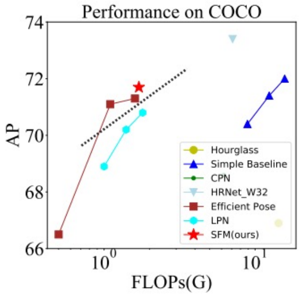
(a) PCKh@0.5 versus Params



(b) PCKh@0.5 versus FLOPs



(c) AP versus Params

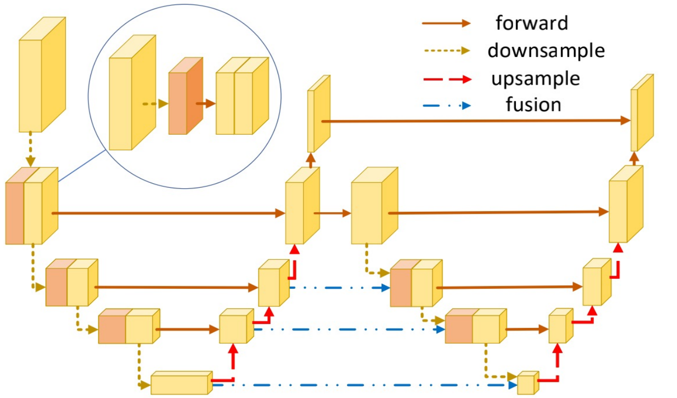


(d) AP versus FLOPs

Experiments on MPII validation and COCO validation. The closer to the top left corner the better the effect.

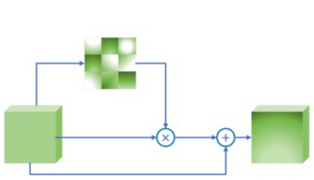
Our Method

SFM: We first expand the general pyramid network to a cascaded pyramid network to extend the network depth so as to increase the fault tolerance of the model. Furthermore, we make full use of the layers in the network and add fusion connections for different scales as much as possible which can contain more effective information and have more complex representing capacity.

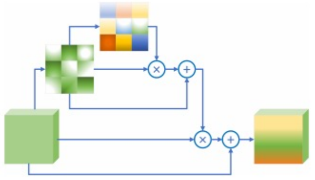


The overview of SFM.

HSA: Its core idea can be understood as computing the attention map of the attention map to get a more precise and learnable attention map. It improves the accuracy of the entire network for long-range feature capture while consuming essentially no computational resources.



(a) Spatial attention



(b) High-level spatial attention(HSA)

Comparison between spatial attention and HSA.

Experiments

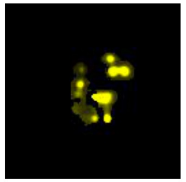
COMPARISONS OF RESULTS ON MPII VALIDATION

Method	Params(M)	GFLOPs	PCKh@0.5	PCKh@0.1
CPMs	31.0	175.0	88.0	-
DLCM	34.0	12.0	88.5	-
SimpleBaseline	34.0	12.0	88.5	-
OpenPose	25.94	160.36	87.60	22.76
EfficientPoseRT	0.46	0.87	82.88	22.76
EfficientPose i	0.72	1.67	85.18	26.49
EfficientPose ii	1.73	7.70	85.18	30.17
EfficientPose iii	3.23	23.35	89.51	30.90
EfficientPose iv	6.56	72.89	89.75	35.63
SFM	**1.50**	**1.70**	**89.0**	**42.0**

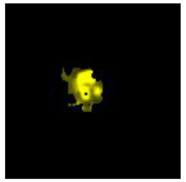
(a) original image

COMPARISONS OF RESULTS ON COCO VALIDATION

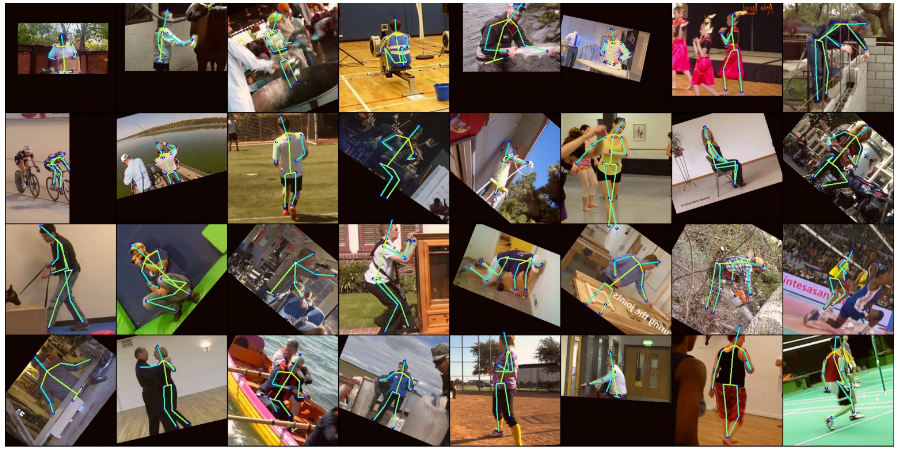
Method	Pretrain(%)	Params(M)(%)	GFLOPs(%)	AP(%)
Hourglass	N	25.1	14.3	66.9
CPN	Y	102.0	6.2	68.6
LPN-50	N	2.9	1.0	68.9
SBline-R50	Y	34.0	8.9	70.4
HRNet-W32	N	28.5	7.1	73.4
EfficientPose-A	N	1.3	0.5	66.5
EfficientPose-B	N	3.3	1.1	71.1
EfficientPose-C	N	5.0	1.6	71.3
SFM	N	**1.5**	**1.7**	**71.7**



(b) attention map in HSA



(c) attention map in general spatial attention



The visualization result of our model on MPII validation dataset.