High-Resolution Representation Learning for Human Pose Estimation

Bin Xiao

Microsoft Research Asia

Microsoft

Outline

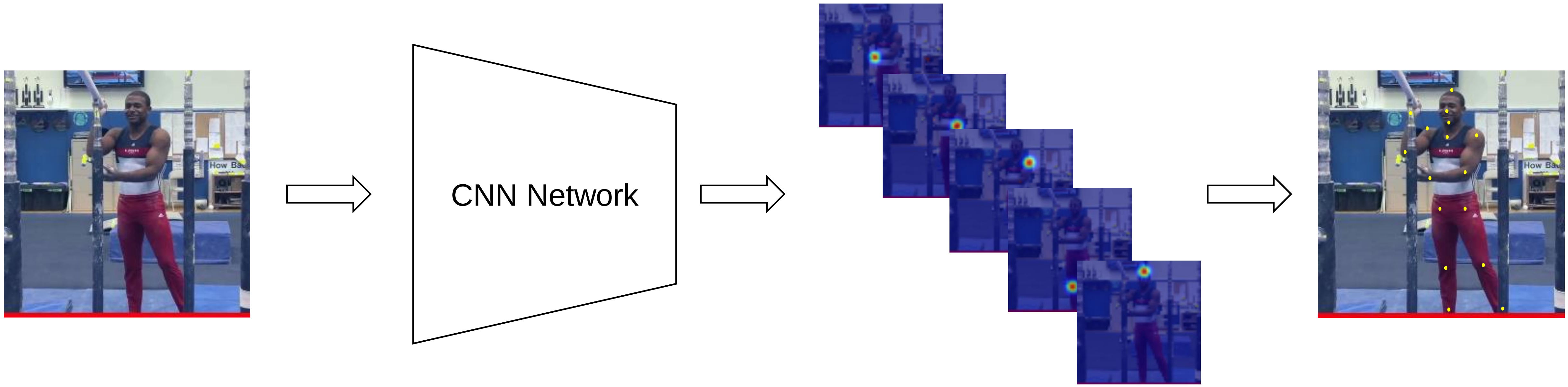
- Human Pose Estimation
  - Top-down vs. bottom-up
  - General pipeline for single-person pose estimation
- High-Resolution Net for Pose Estimation
- Results

Top-down vs. Bottom-up

- Person detection → Single-person pose estimation

- Keypoint detection → Grouping keypoints

General Pipeline for Person Pose Estimation



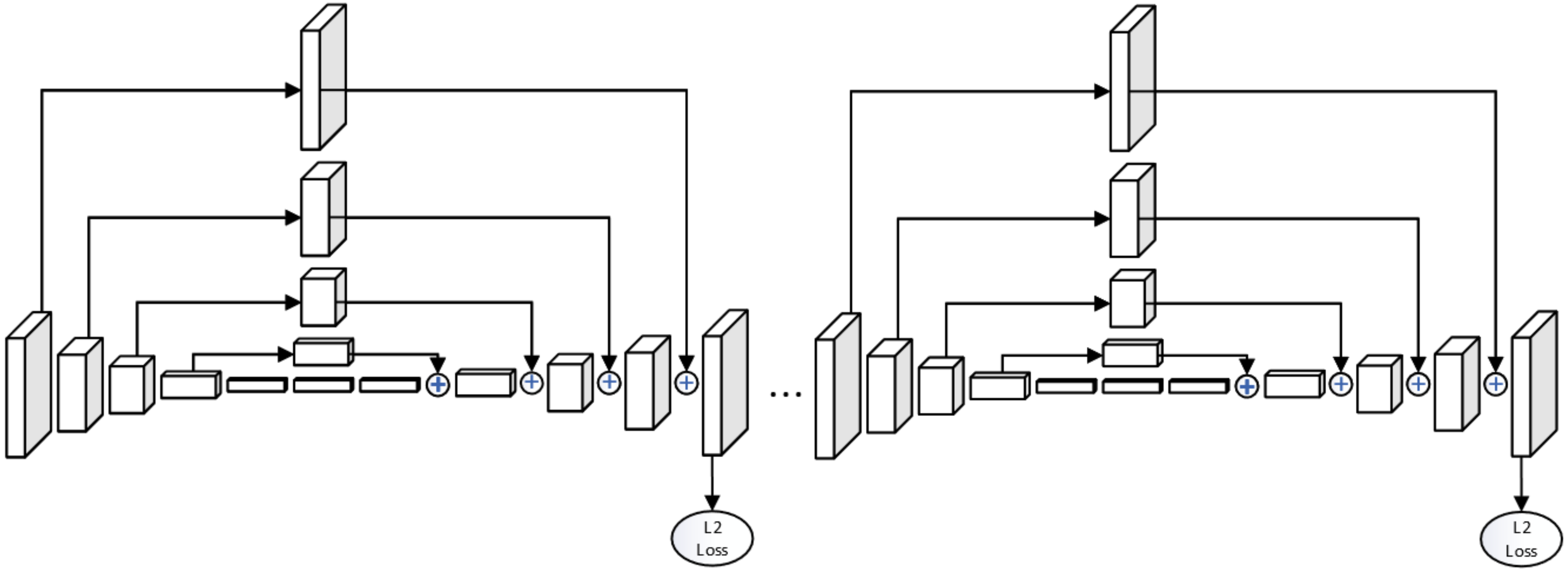
CNN Network

256x256

?

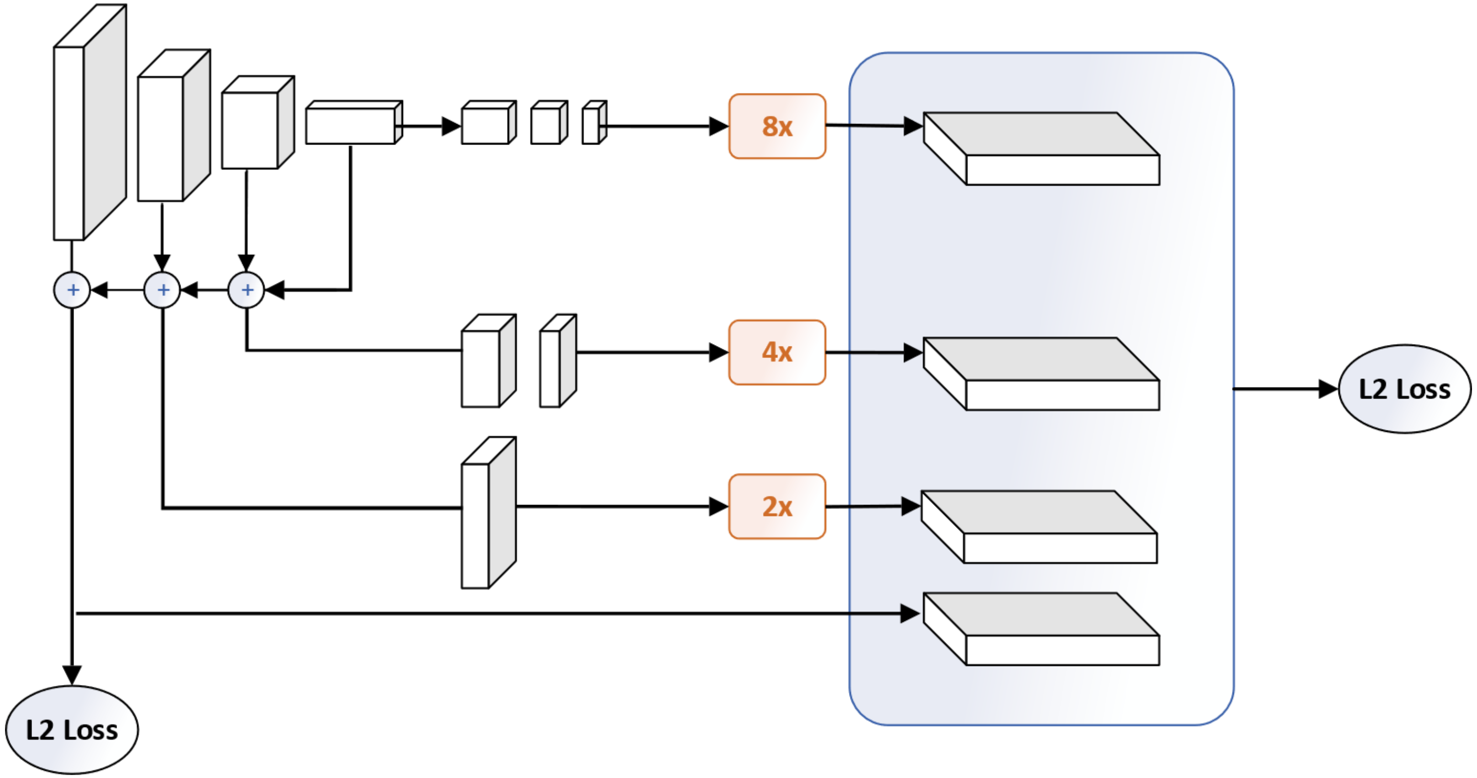
64x64

State-of-the-art networks(Stacked-Hourglass)



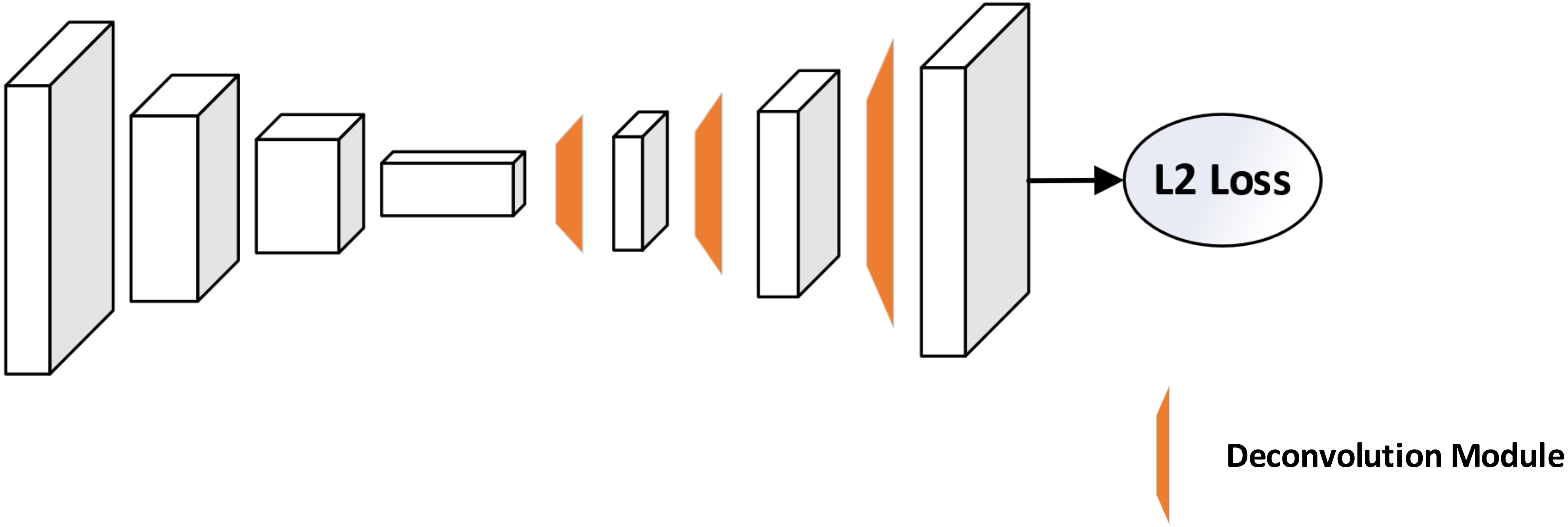
Newell, A., Yang, K., Deng, J.: **Stacked hourglass networks for human pose estimation**

State-of-the-art networks(CPN)



Chen, Y., Wang, Z., Peng, Y., Zhang, Z., Yu, G., Sun, J.: **Cascaded pyramid network for multi-person pose estimation**

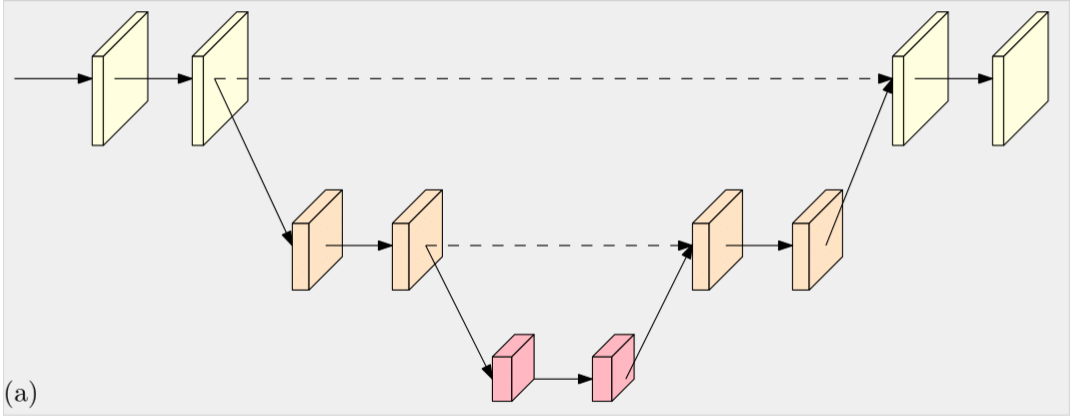
State-of-the-art networks(SimpleBaseline)



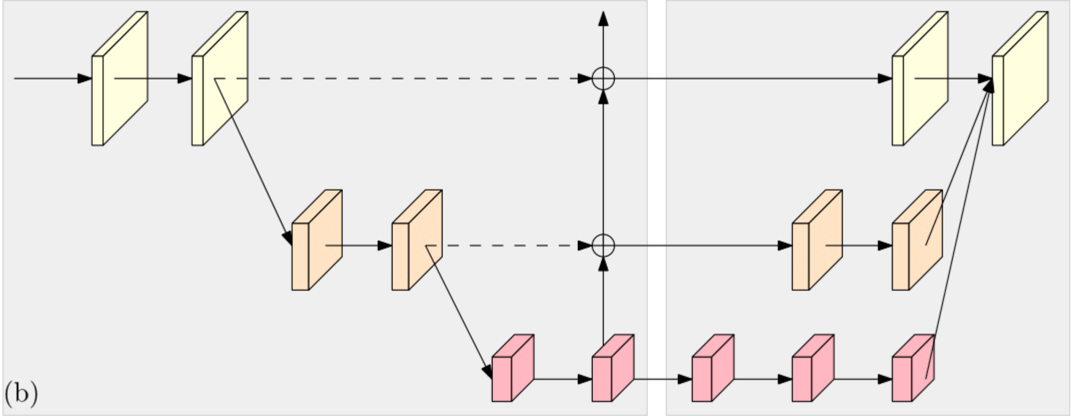
Xiao, B., Wu, H., Wei, Y.: **Simple Baselines for Human Pose Estimation and Tracking**

Microsoft

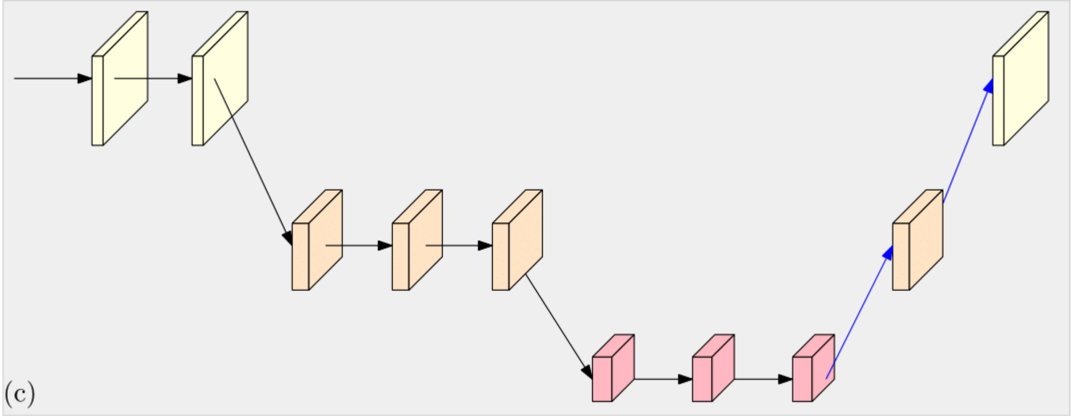
State-of-the-art networks



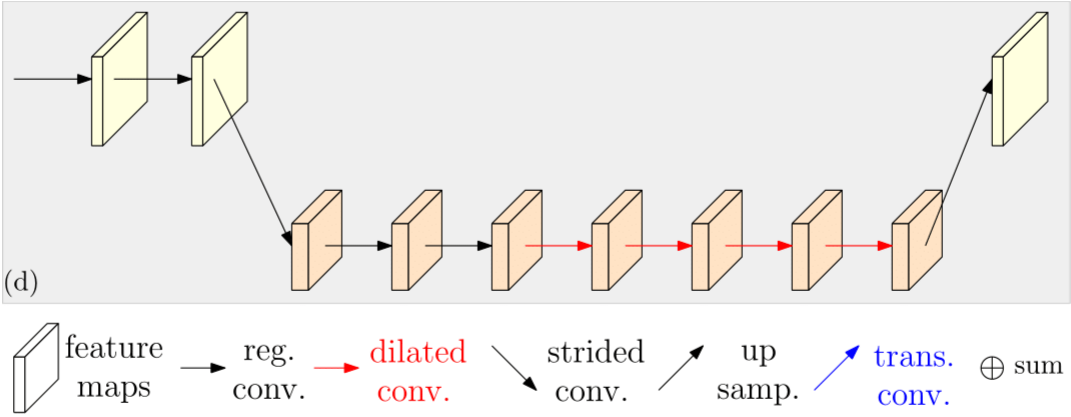
Hourglass



CPN



SimpleBaseline

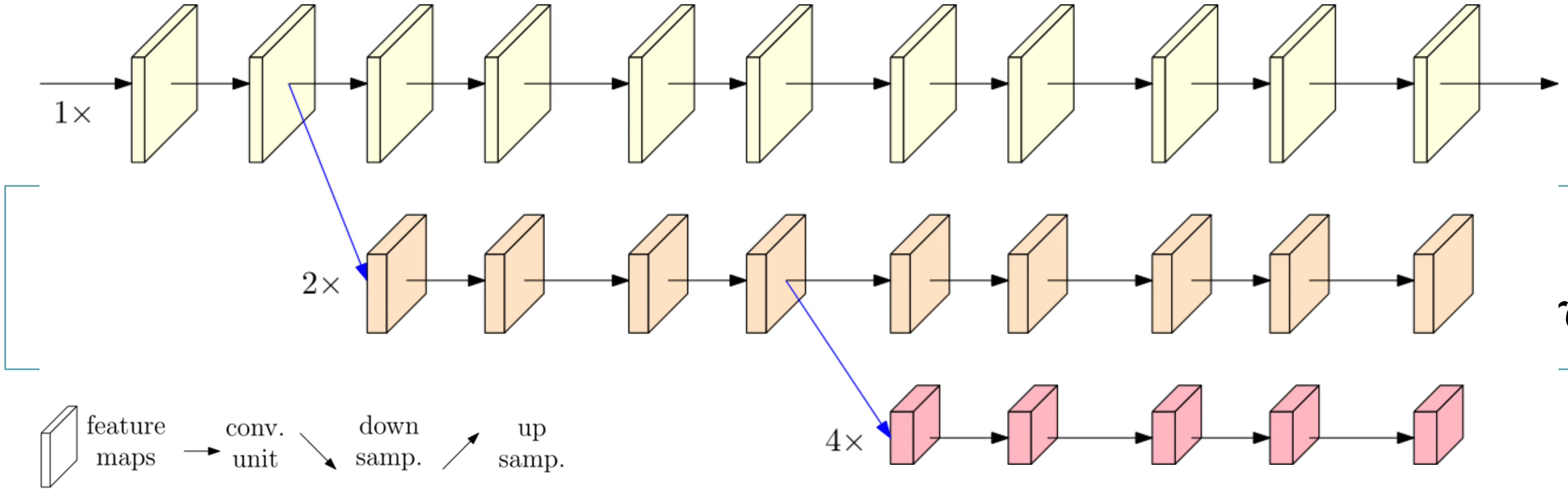


Summary

- Connect high-to-low resolution convolutions in *series*
- *Recover* high-resolution representations *from low-resolution representations*
- Multi-scale *different* level feature fuse(*low level and high level*)

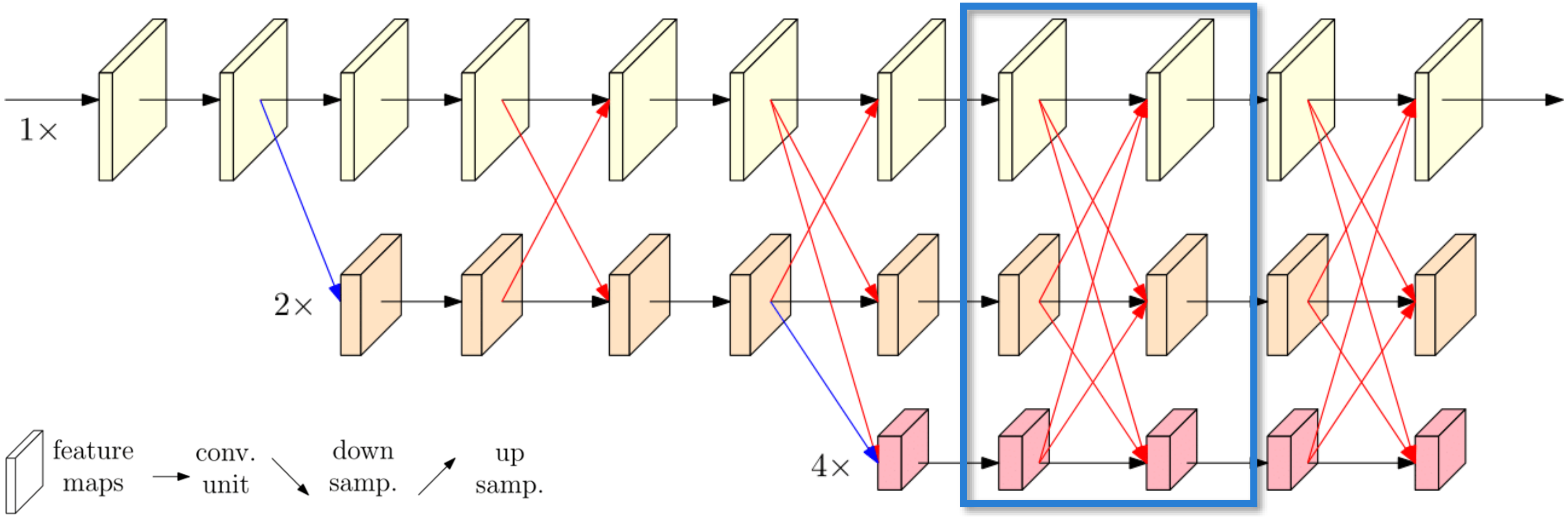
High resolution, but not strong representation

HRNets: high-resolution maintenance



rocess

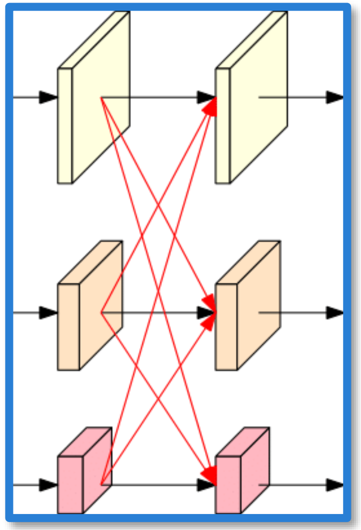
HRNets: repeated multi-scale fusion

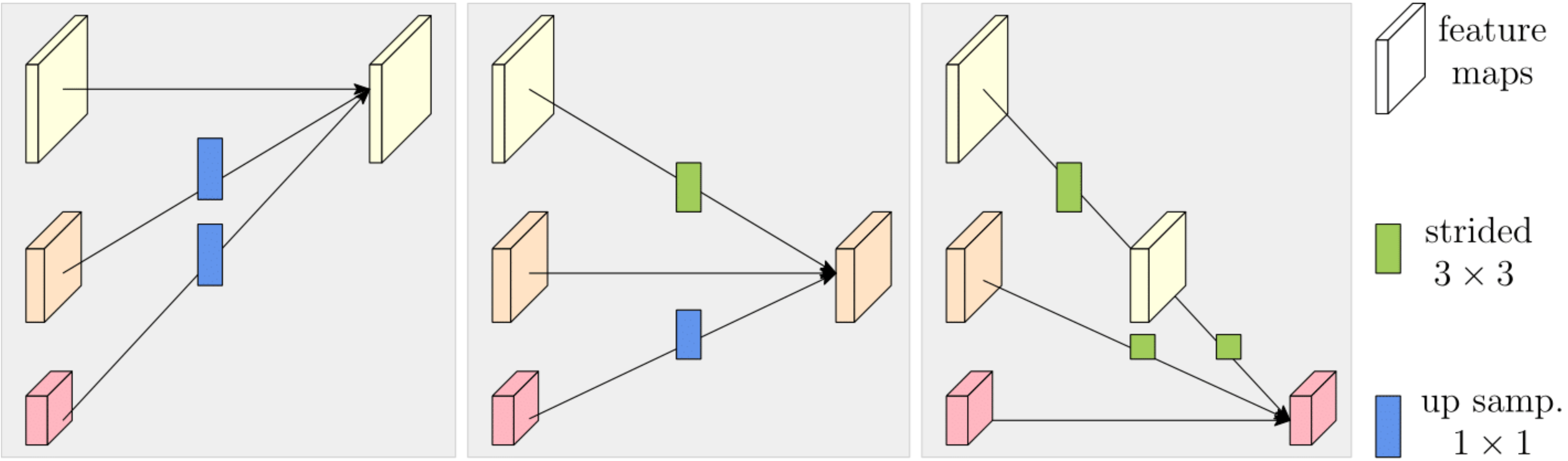


- Repeat fusions across resolutions
- Strengthen high-resolution and low-resolution representations

Microsoft

Multi-scale fusion

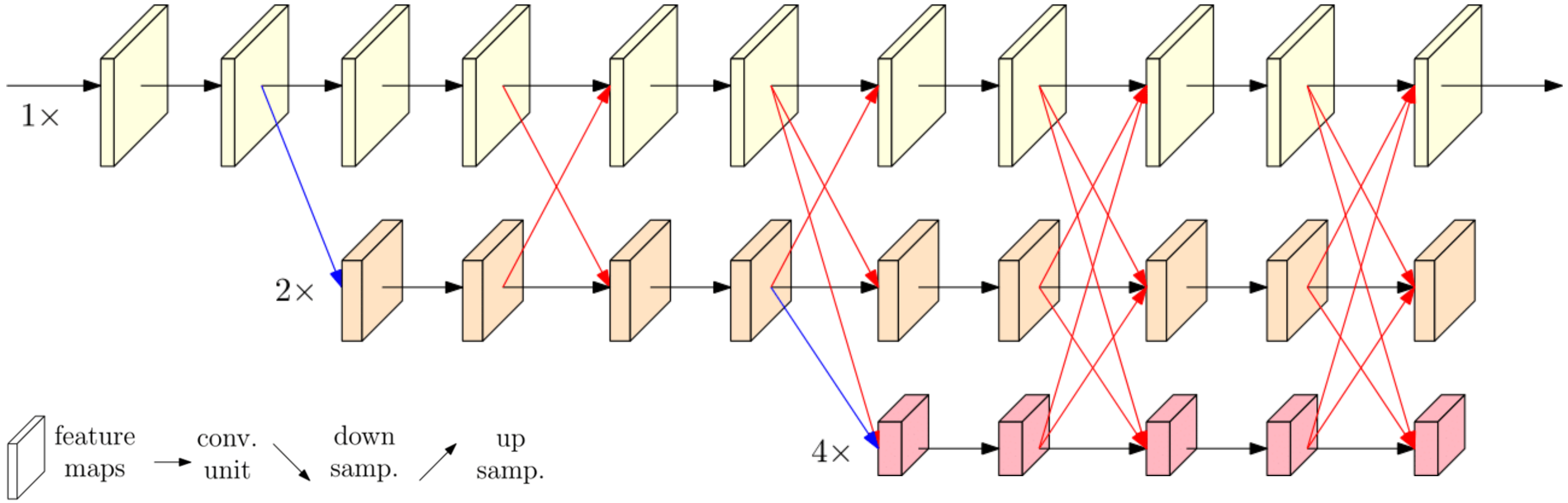




Down-sample: 3 × 3

Up-sample: 1 × 1

HRNet: repeated multi-scale fusion



Summary

parallel

- Connect high-to-low resolution convolutions in ~~*series*~~

Maintain *through the whole process*

- ~~*Recover*~~ high-resolution representations ~~*from low-resolution representations*~~

similar *low resolution and high resolution*

- Multi-scale ~~*different*~~ level feature fuse(~~*low level and high level*~~)

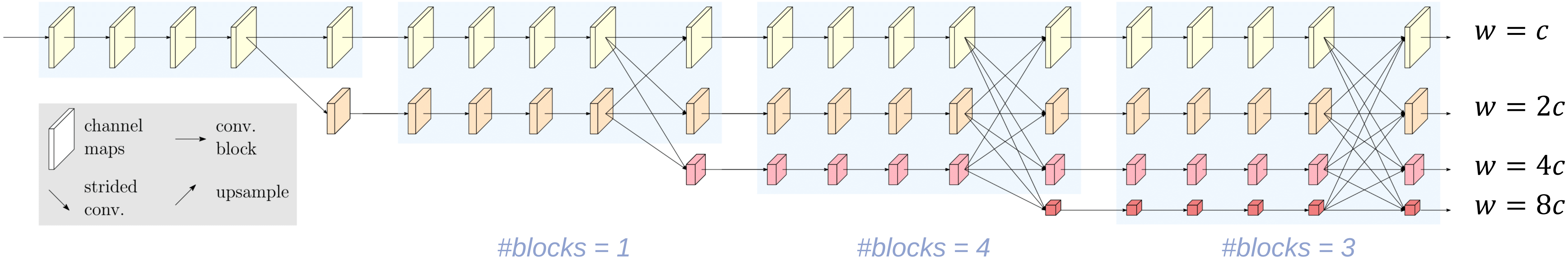
Summary

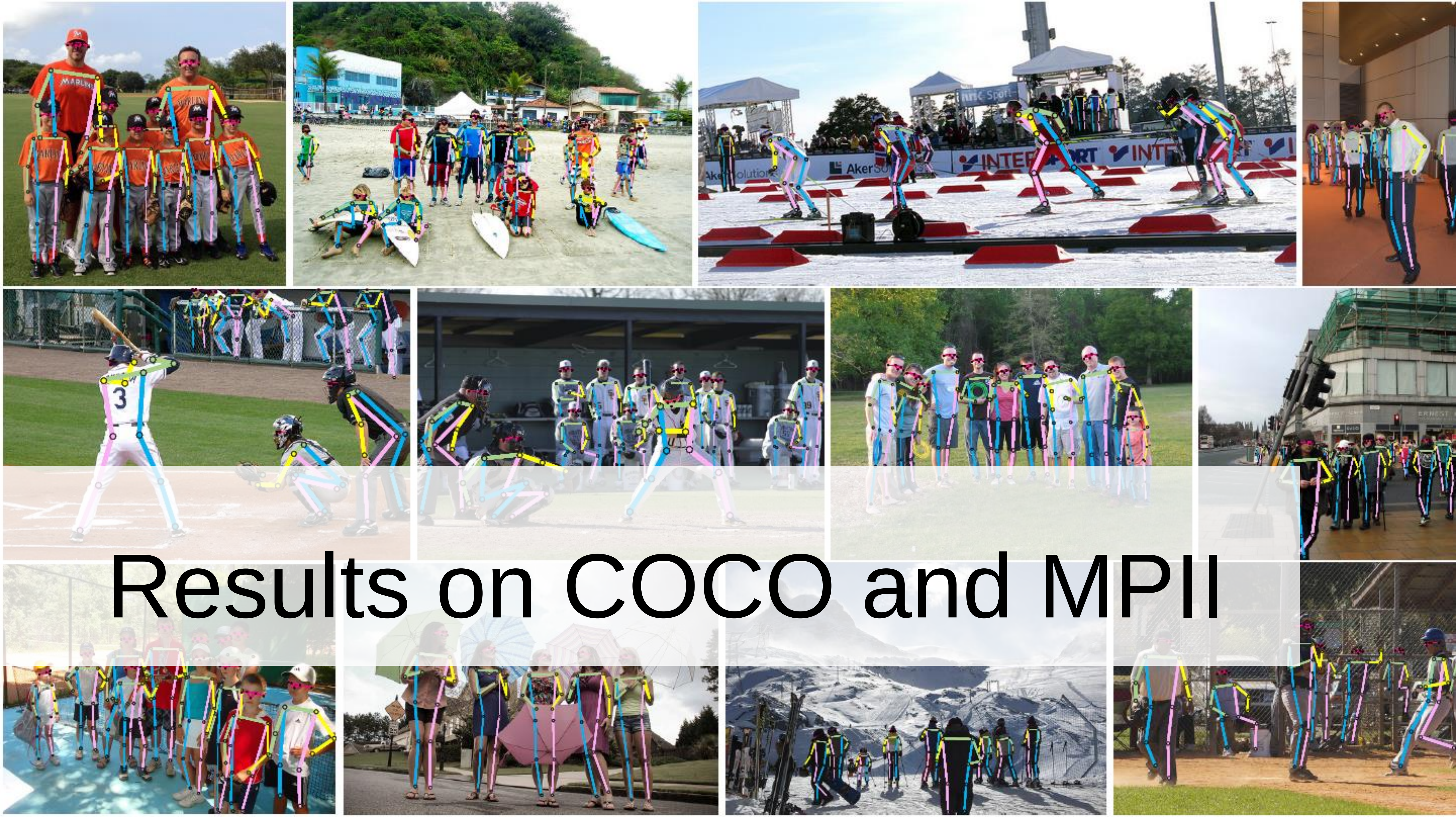
- Connect high-to-low resolution convolutions in *parallel*
- *Maintain* high-resolution representations *through the whole process*
- Repeat fusions across resolutions to strengthen high- & low-representations

High resolution, and strong representation

Microsoft

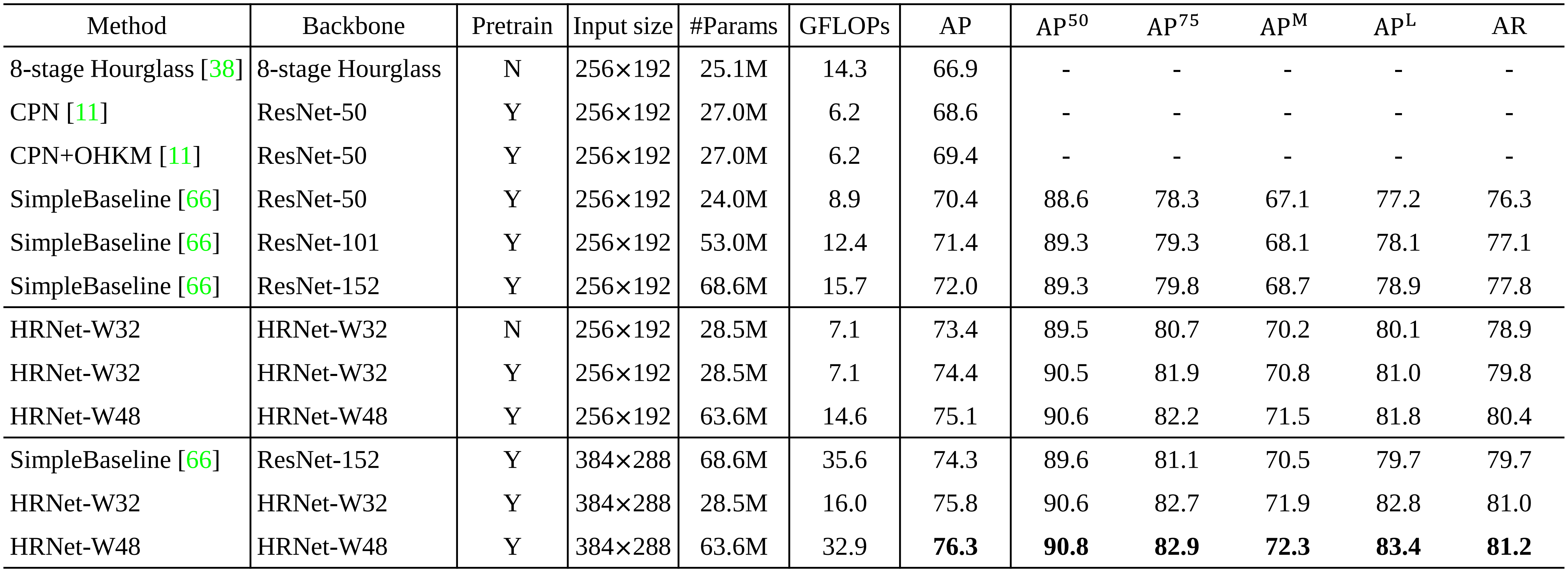
HRNet instantiation





Results on COCO and MPII

COCO validation



Method	Backbone	Pretrain	Input size	#Params	GFLOPs	AP	AP^{50}	AP^{75}	AP^{M}	AP^{L}	AR
8-stage Hourglass [38]	8-stage Hourglass	N	256×192	25.1M	14.3	66.9	-	-	-	-	-
CPN [11]	ResNet-50	Y	256×192	27.0M	6.2	68.6	-	-	-	-	-
CPN+OHKM [11]	ResNet-50	Y	256×192	27.0M	6.2	69.4	-	-	-	-	-
SimpleBaseline [66]	ResNet-50	Y	256×192	24.0M	8.9	70.4	88.6	78.3	67.1	77.2	76.3
SimpleBaseline [66]	ResNet-101	Y	256×192	53.0M	12.4	71.4	89.3	79.3	68.1	78.1	77.1
SimpleBaseline [66]	ResNet-152	Y	256×192	68.6M	15.7	72.0	89.3	79.8	68.7	78.9	77.8
HRNet-W32	HRNet-W32	N	256×192	28.5M	7.1	73.4	89.5	80.7	70.2	80.1	78.9
HRNet-W32	HRNet-W32	Y	256×192	28.5M	7.1	74.4	90.5	81.9	70.8	81.0	79.8
HRNet-W48	HRNet-W48	Y	256×192	63.6M	14.6	75.1	90.6	82.2	71.5	81.8	80.4
SimpleBaseline [66]	ResNet-152	Y	384×288	68.6M	35.6	74.3	89.6	81.1	70.5	79.7	79.7
HRNet-W32	HRNet-W32	Y	384×288	28.5M	16.0	75.8	90.6	82.7	71.9	82.8	81.0
HRNet-W48	HRNet-W48	Y	384×288	63.6M	32.9	**76.3**	**90.8**	**82.9**	**72.3**	**83.4**	**81.2**

COCO test-dev

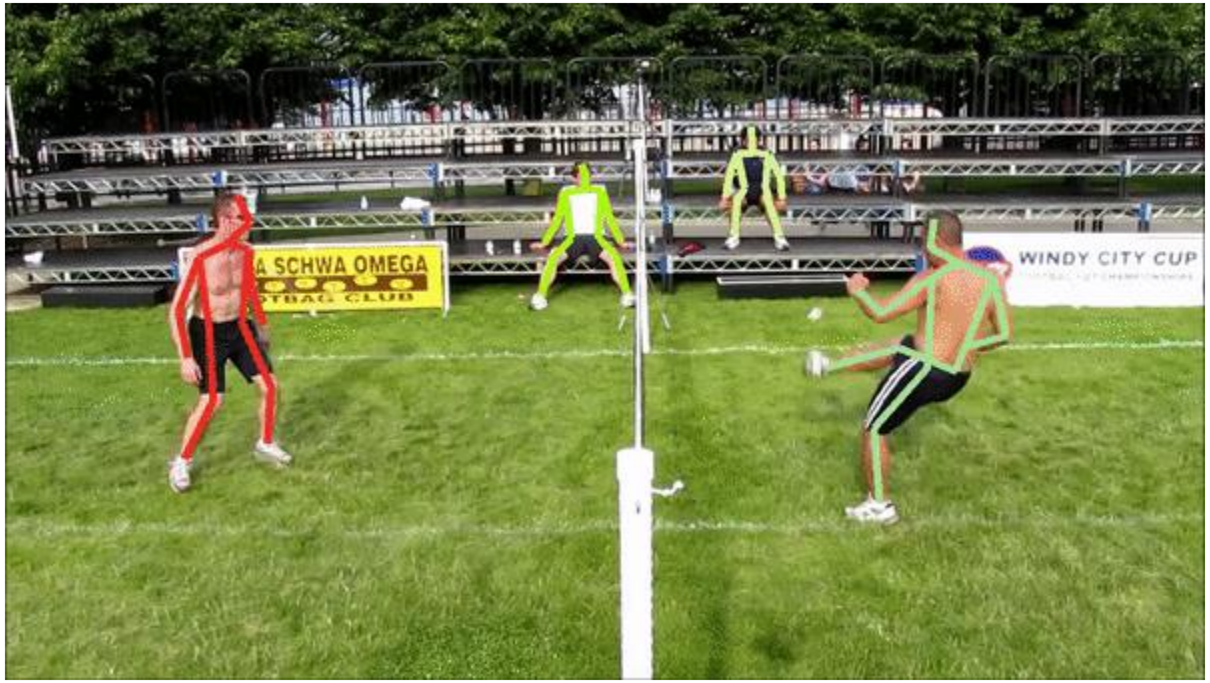
method	Backbone	Input size	#Params	GFLOPs	AP	AP^{50}	AP^{75}	AP^{M}	AP^{L}	AR
Bottom-up: keypoint detection and grouping										
OpenPose [6]	-	-	-	-	61.8	84.9	67.5	57.1	68.2	66.5
Associative Embedding [39]	-	-	-	-	65.5	86.8	72.3	60.6	72.6	70.2
PersonLab [46]	-	-	-	-	68.7	89.0	75.4	64.1	75.5	75.4
MultiPoseNet [33]	-	-	-	-	69.6	86.3	76.6	65.0	76.3	73.5
Top-down: human detection and single-person keypoint detection										
Mask-RCNN [21]	ResNet-50-FPN	-	-	-	63.1	87.3	68.7	57.8	71.4	-
G-RMI [47]	ResNet-101	353×257	42.0M	57.0	64.9	85.5	71.3	62.3	70.0	69.7
Integral Pose Regression [60]	ResNet-101	256×256	45.0M	11.0	67.8	88.2	74.8	63.9	74.0	-
G-RMI + extra data [47]	ResNet-101	353×257	42.6M	57.0	68.5	87.1	75.5	65.8	73.3	73.3
CPN [11]	ResNet-Inception	384×288	-	-	72.1	91.4	80.0	68.7	77.2	78.5
RMPE [17]	PyraNet [77]	320×256	28.1M	26.7	72.3	89.2	79.1	68.0	78.6	-
CFN [25]	-	-	-	-	72.6	86.1	69.7	78.3	64.1	-
CPN(ensemble) [11]	ResNet-Inception	384×288	-	-	73.0	91.7	80.9	69.5	78.1	79.0
SimpleBaseline [72]	ResNet-152	384×288	68.6M	35.6	73.7	91.9	81.1	70.3	80.0	79.0
HRNet-W32	HRNet-W32	384×288	28.5M	16.0	74.9	92.5	82.8	71.3	80.9	80.1
HRNet-W48	HRNet-W48	384×288	63.6M	32.9	**75.5**	**92.5**	**83.3**	**71.9**	**81.5**	**80.5**
HRNet-W48 + extra data	HRNet-W48	384×288	63.6M	32.9	**77.0**	**92.7**	**84.5**	**73.4**	**83.1**	**82.0**

MPII test

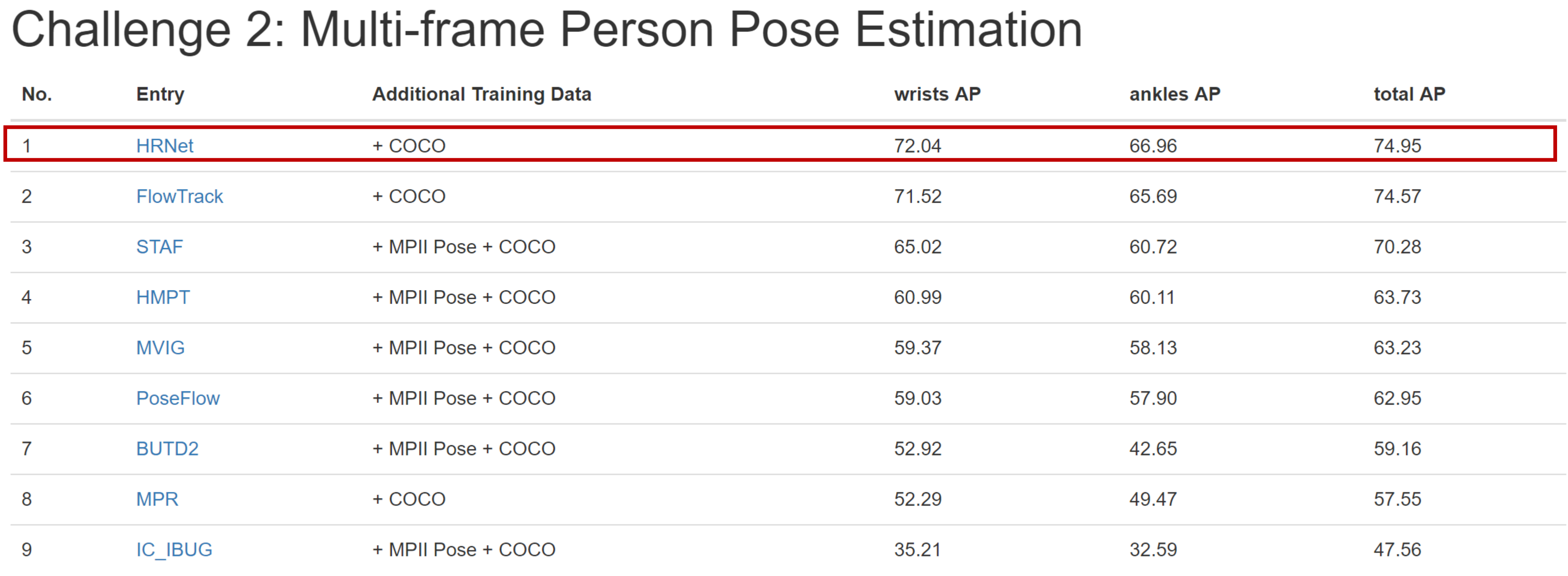
Method	Hea.	Sho.	Elb.	Wri.	Hip	Kne.	Ank.	Total
Insafutdinov et al. [27]	96.8	95.2	89.3	84.4	88.4	83.4	78.0	88.5
Wei et al. [69]	97.8	95.0	88.7	84.0	88.4	82.8	79.4	88.5
Bulat et al. [4]	97.9	95.1	89.9	85.3	89.4	85.7	81.7	89.7
Newell et al. [40]	98.2	96.3	91.2	87.1	90.1	87.4	83.6	90.9
Sun et al. [58]	98.1	96.2	91.2	87.2	89.8	87.4	84.1	91.0
Tang et al. [63]	97.4	96.4	92.1	87.7	90.2	87.7	84.3	91.2
Ning et al. [44]	98.1	96.3	92.2	87.8	90.6	87.6	82.7	91.2
Luvizon et al. [37]	98.1	96.6	92.0	87.5	90.6	88.0	82.7	91.2
Chu et al. [14]	98.5	96.3	91.9	88.1	90.6	88.0	85.0	91.5
Chou et al. [12]	98.2	96.8	92.2	88.0	91.3	89.1	84.9	91.8
Chen et al. [10]	98.1	96.5	92.5	88.5	90.2	**89.6**	86.0	91.9
Yang et al. [77]	98.5	96.7	92.5	88.7	91.1	88.6	86.0	92.0
Ke etal. [31]	98.5	96.8	92.7	88.4	90.6	89.3	**86.3**	92.1
Tang et al. [62]	98.4	**96.9**	92.6	88.7	**91.8**	89.4	86.2	**92.3**
SimpleBaseline [72]	98.5	96.6	91.9	87.6	91.1	88.1	84.1	91.5
HRNet-W32	**98.6**	**96.9**	**92.8**	**89.0**	91.5	89.0	85.7	**92.3**

PCKh@0.5

Application to Human Pose Tracking(PoseTrack)



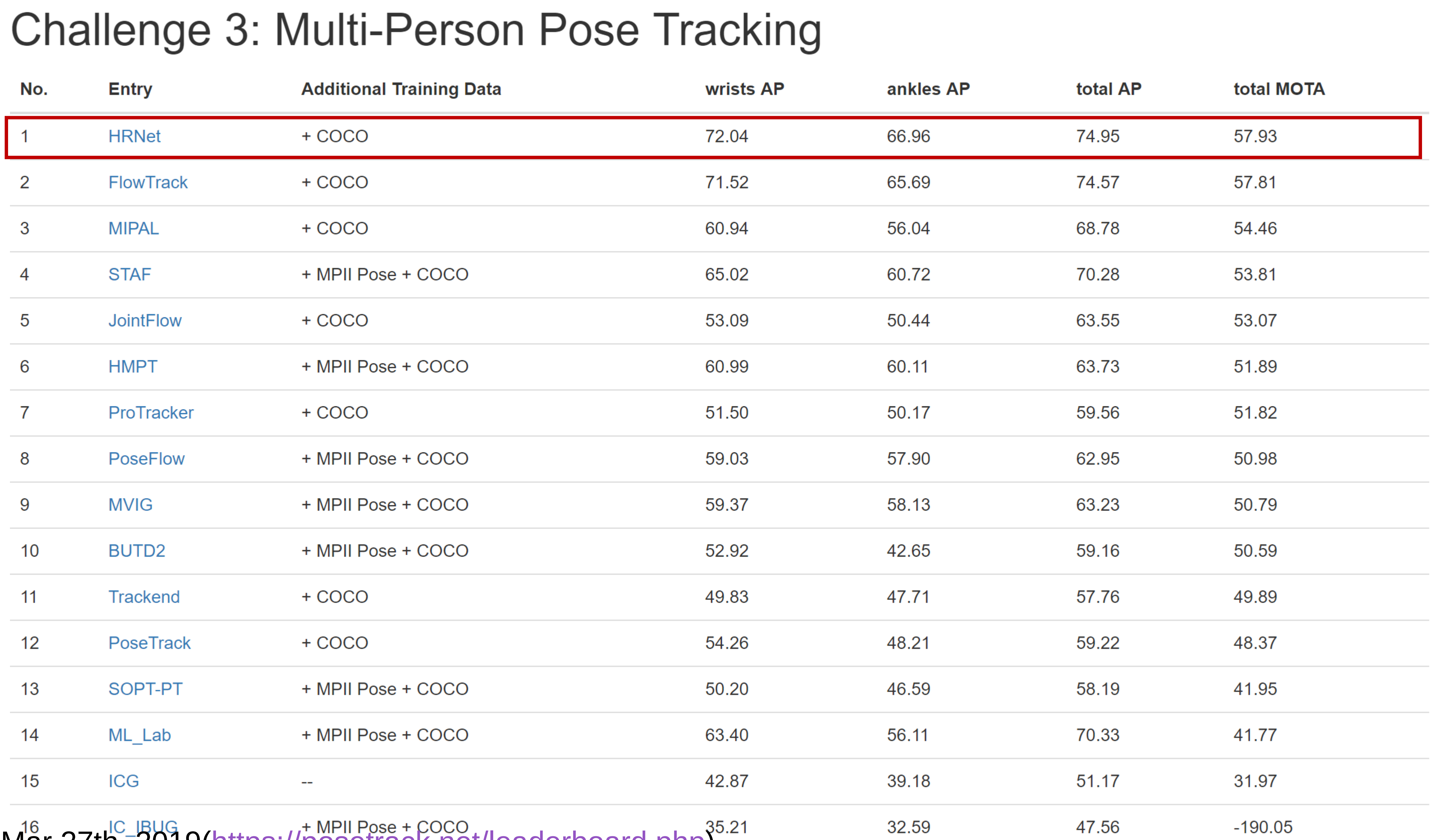
PoseTrack 2017 Leaderboard



Challenge 2: Multi-frame Person Pose Estimation

No.	Entry	Additional Training Data	wrists AP	ankles AP	total AP
1	HRNet	+ COCO	72.04	66.96	74.95
2	FlowTrack	+ COCO	71.52	65.69	74.57
3	STAF	+ MPII Pose + COCO	65.02	60.72	70.28
4	HMPT	+ MPII Pose + COCO	60.99	60.11	63.73
5	MVIG	+ MPII Pose + COCO	59.37	58.13	63.23
6	PoseFlow	+ MPII Pose + COCO	59.03	57.90	62.95
7	BUTD2	+ MPII Pose + COCO	52.92	42.65	59.16
8	MPR	+ COCO	52.29	49.47	57.55
9	IC_IBUG	+ MPII Pose + COCO	35.21	32.59	47.56

From April 1, 2019(https://posetrack.net/leaderboard.php)

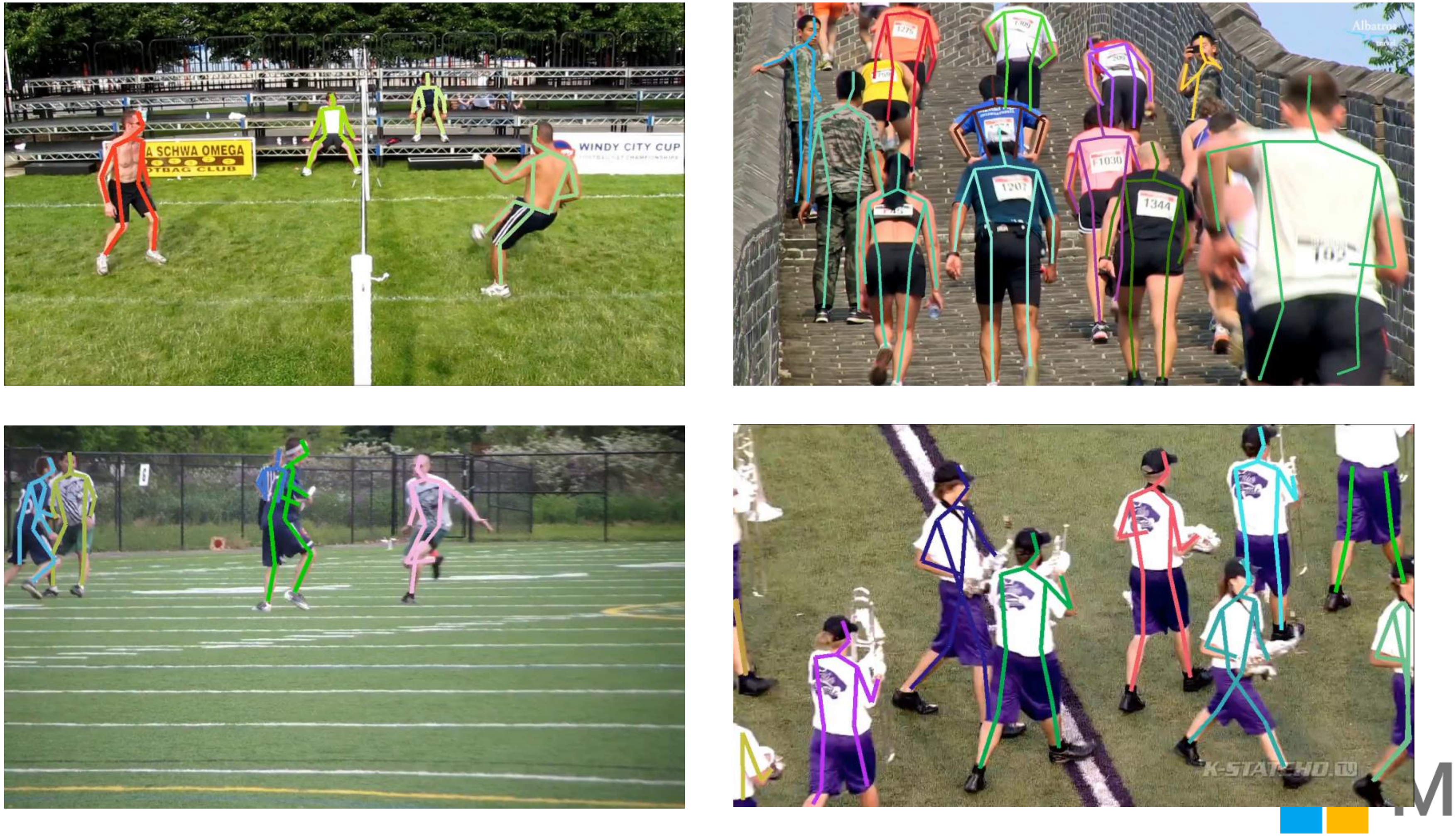


Challenge 3: Multi-Person Pose Tracking

No.	Entry	Additional Training Data	wrists AP	ankles AP	total AP	total MOTA
1	HRNet	+ COCO	72.04	66.96	74.95	57.93
2	FlowTrack	+ COCO	71.52	65.69	74.57	57.81
3	MIPAL	+ COCO	60.94	56.04	68.78	54.46
4	STAF	+ MPII Pose + COCO	65.02	60.72	70.28	53.81
5	JointFlow	+ COCO	53.09	50.44	63.55	53.07
6	HMPT	+ MPII Pose + COCO	60.99	60.11	63.73	51.89
7	ProTracker	+ COCO	51.50	50.17	59.56	51.82
8	PoseFlow	+ MPII Pose + COCO	59.03	57.90	62.95	50.98
9	MVIG	+ MPII Pose + COCO	59.37	58.13	63.23	50.79
10	BUTD2	+ MPII Pose + COCO	52.92	42.65	59.16	50.59
11	Trackend	+ COCO	49.83	47.71	57.76	49.89
12	PoseTrack	+ COCO	54.26	48.21	59.22	48.37
13	SOPT-PT	+ MPII Pose + COCO	50.20	46.59	58.19	41.95
14	ML_Lab	+ MPII Pose + COCO	63.40	56.11	70.33	41.77
15	ICG	--	42.87	39.18	51.17	31.97
16	IC_IBUG	+ MPII Pose + COCO	35.21	32.59	47.56	-190.05

From Mar 27th, 2019(https://posetrack.net/leaderboard.php)

Multi-Person Pose Tracking



Deep representation learning for visual recognition

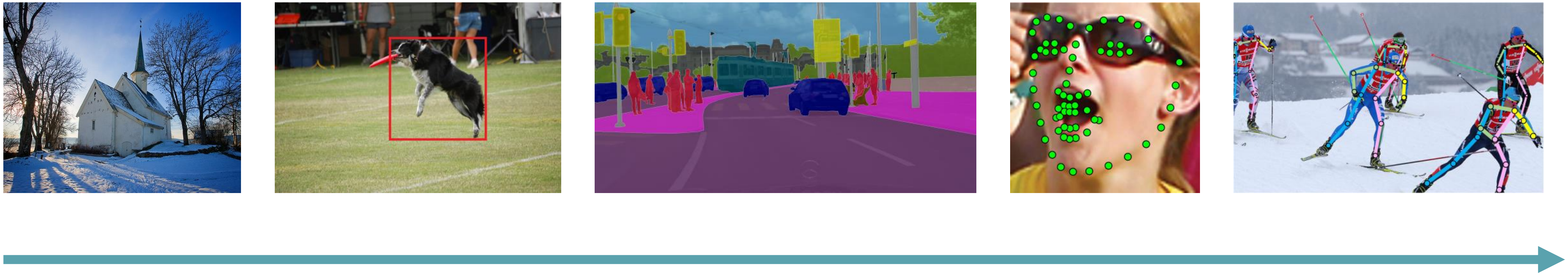


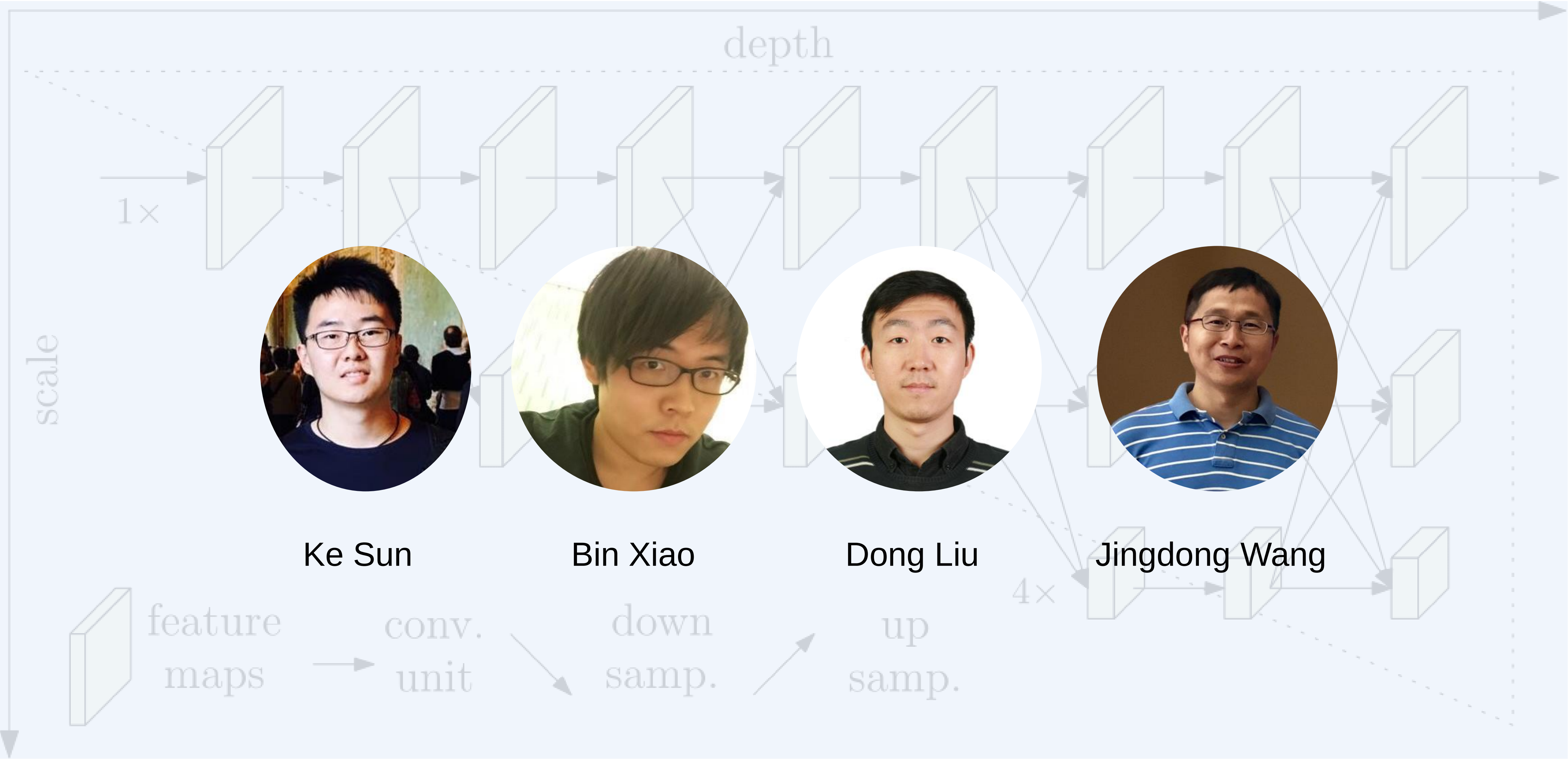
image-level

region-level

pixel-level

Project page
https://jingdongwang2017.github.io/Projects/HRNet/

Team



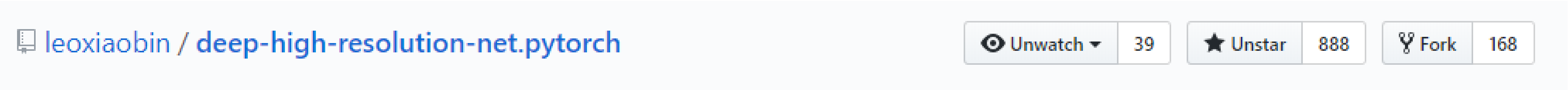
Ke Sun

Bin Xiao

Dong Liu

Jingdong Wang

Thank you!



leoxiaobin / deep-high-resolution-net.pytorch

- Code and Models is available at https://github.com/leoxiaobin/deep-high-resolution-net.pytorch