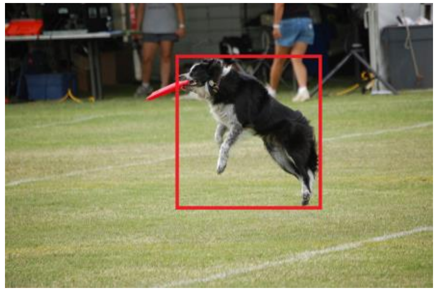
Deep High-Resolution Representation Learning for Visual Recognition

Jingdong Wang
Senior Principal Research Manager
Microsoft Research, Beijing, China

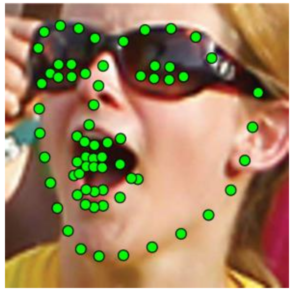
Convolutional neural networks are good at representation learning

Image
classification



Object
detection

Semantic
segmentation



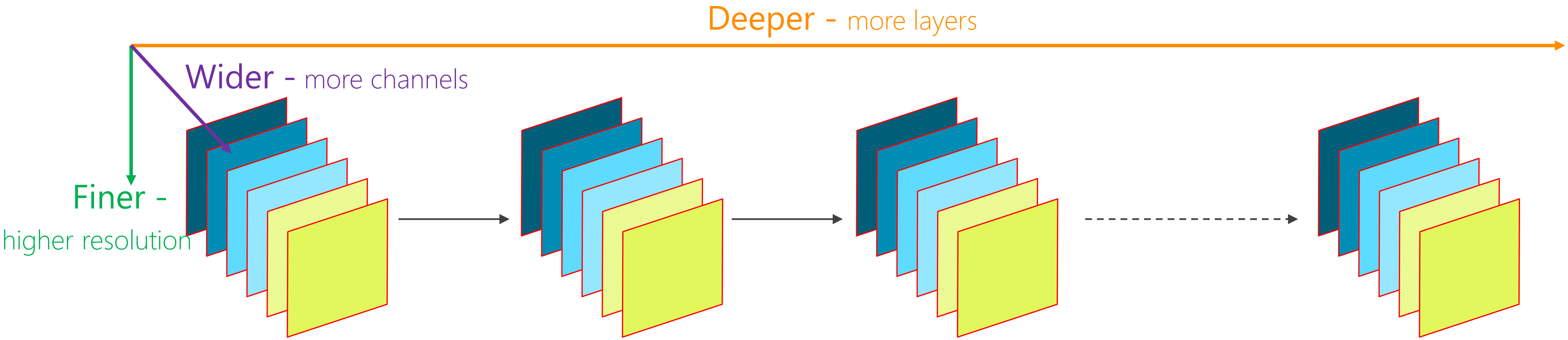
Face
alignment

Pose
estimation

……

Neural architecture design

deeper → wider → finer

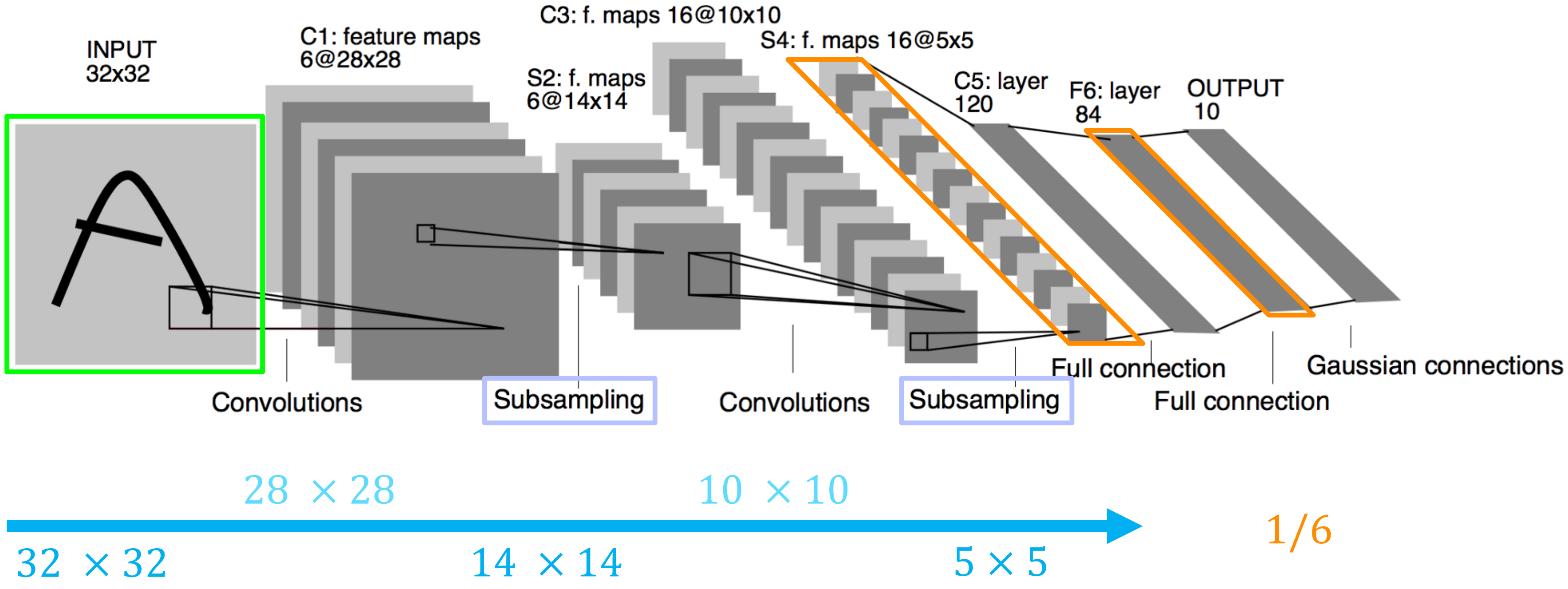


New dimension: go finer towards high-resolution representation learning

Low-resolution representation learning

Connect multi-resolution convolutions in *series*

High-resolution conv. → medium-resolution conv. → low-resolution conv.



LeNet-5, 1998

and same for other classification networks: AlexNet, VGGNet, GoogleNet, ResNet, DenseNet, ……


Low resolution
is enough
image recog.
region-level recog.
pixel-level recog.
global
position-sensitive

Human pose estimation

Human pose estimation

Low resolution is enough

The high-resolution representation is needed

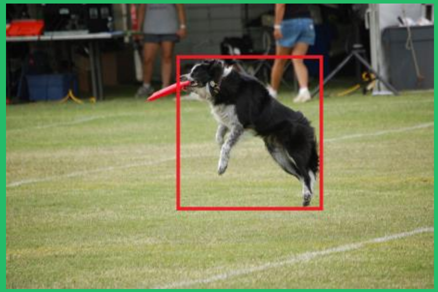


image recog.

region-level recog.

pixel-level recog.

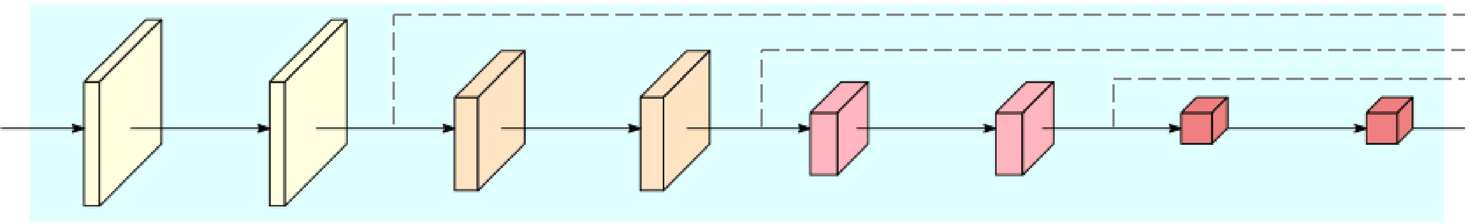
global

position-sensitive

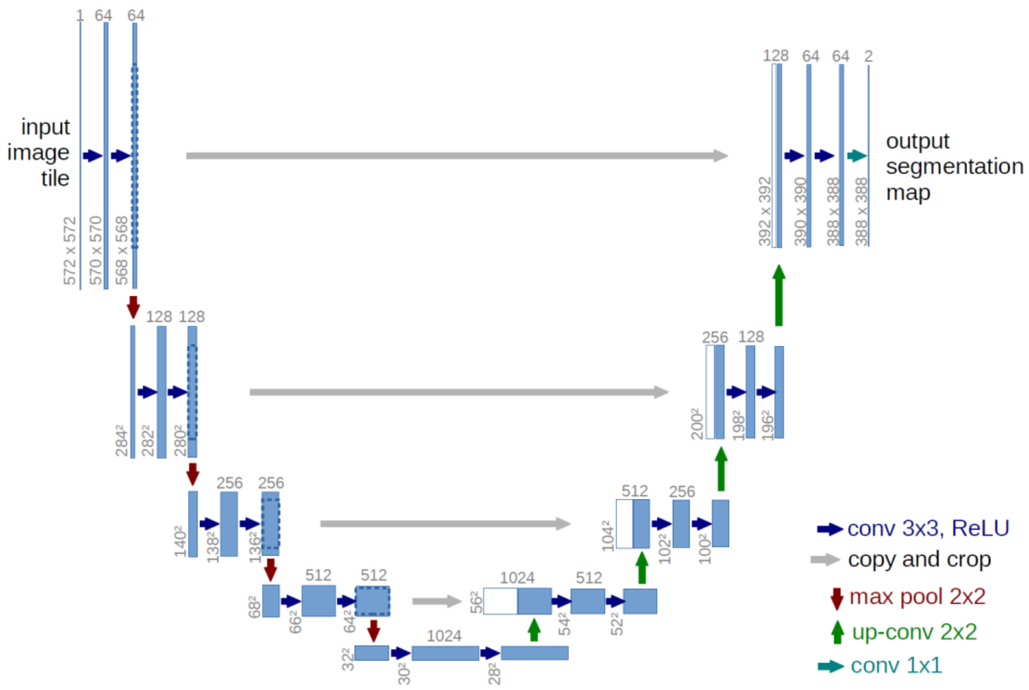
High-resolution representation learning

Previous solutions

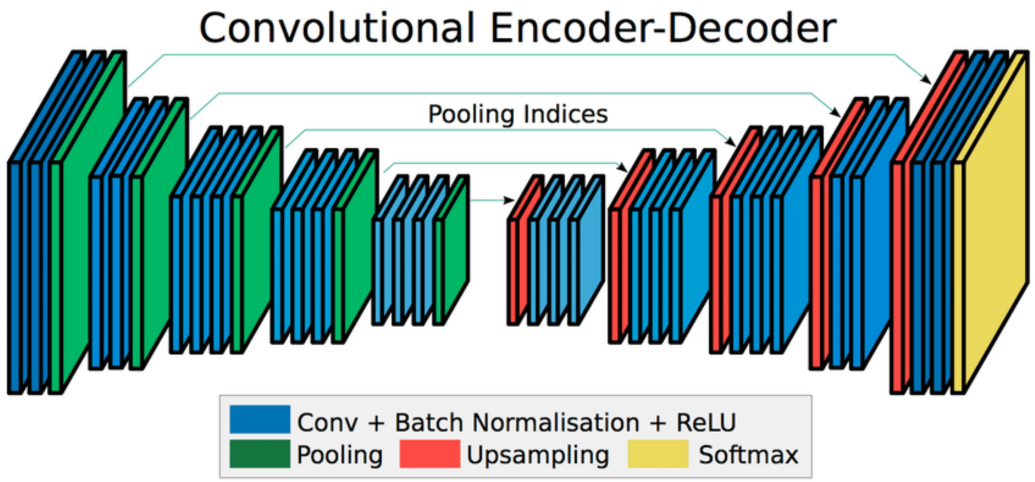
- ❑ Based on low-resolution representations from *classification networks*
- ❑ Recover high resolution from low resolution



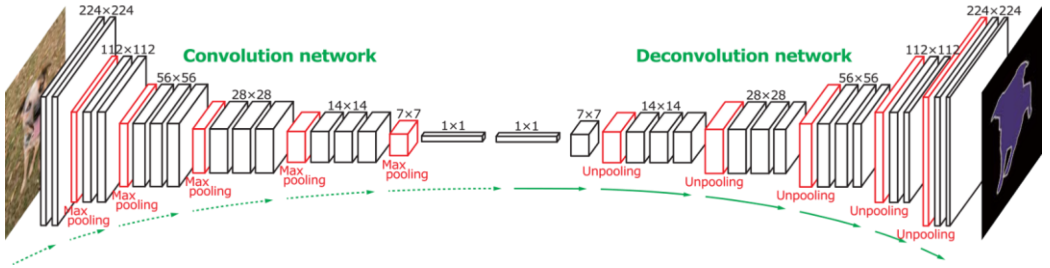
Hourglass, U-Net, Encoder-decoder, DeconvNet, SimpleBaseline, etc



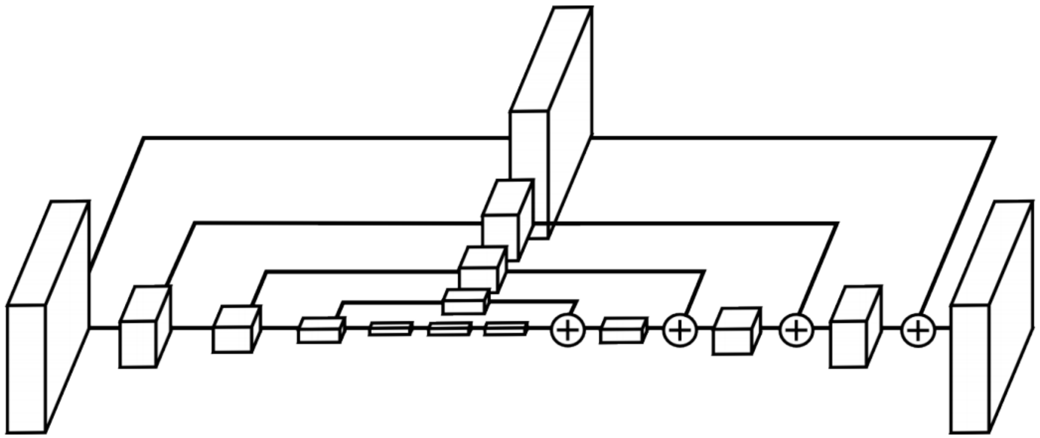
U-Net



SegNet



DeconvNet



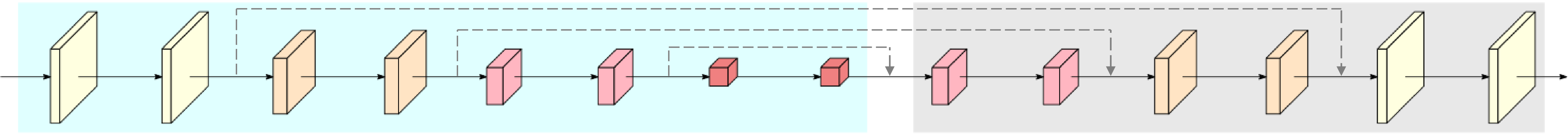
Hourglass

Look different, essentially the same

High-resolution representation learning

Previous solutions

- ❑ Based on low-resolution representations from *classification networks*
- ❑ Recover high resolution from low resolution
- ❑ Drawback: representations are weak due to *location-sensitivity loss*



Hourglass, U-Net, Encoder-decoder, DeconvNet, SimpleBaseline, etc

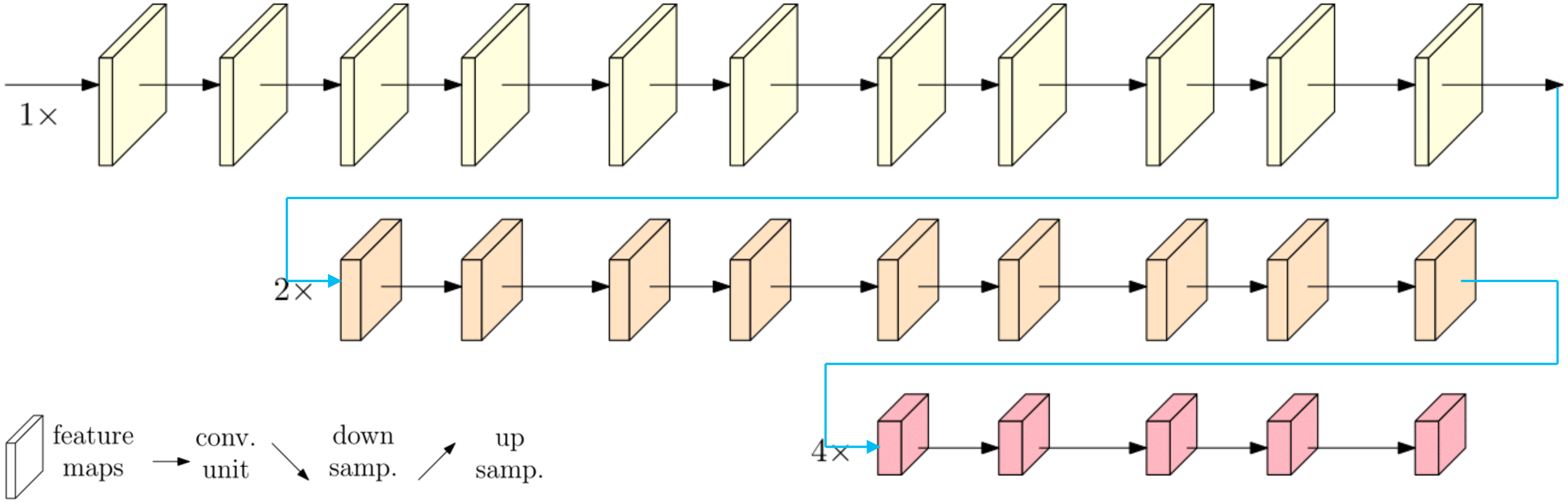
Our work: High-resolution networks (HRNet)

Learn high-resolution representations through high resolution
maintenance rather than recovering

Ke Sun, Bin Xiao, Dong Liu, Jingdong Wang: Deep High-Resolution Representation Learning for Human Pose Estimation. CVPR 2019
Ke Sun, Yang Zhao, Borui Jiang, Tianheng Cheng, Bin Xiao, Dong Liu, Yadong Mu, Xinggang Wang, Wenyu Liu, Jingdong Wang: High-Resolution Representation Learning for labeling pixels and regions
Jingdong Wang, Ke Sun, Tianheng Cheng, Borui Jiang, Chaorui Deng, Yang Zhao, Dong Liu, Yadong Mu, Mingkui Tan, Xinggang Wang, Wenyu Liu, and Bin Xiao: Deep High-Resolution Representation Learning for Visual Recognition (submitted to TPAMI)

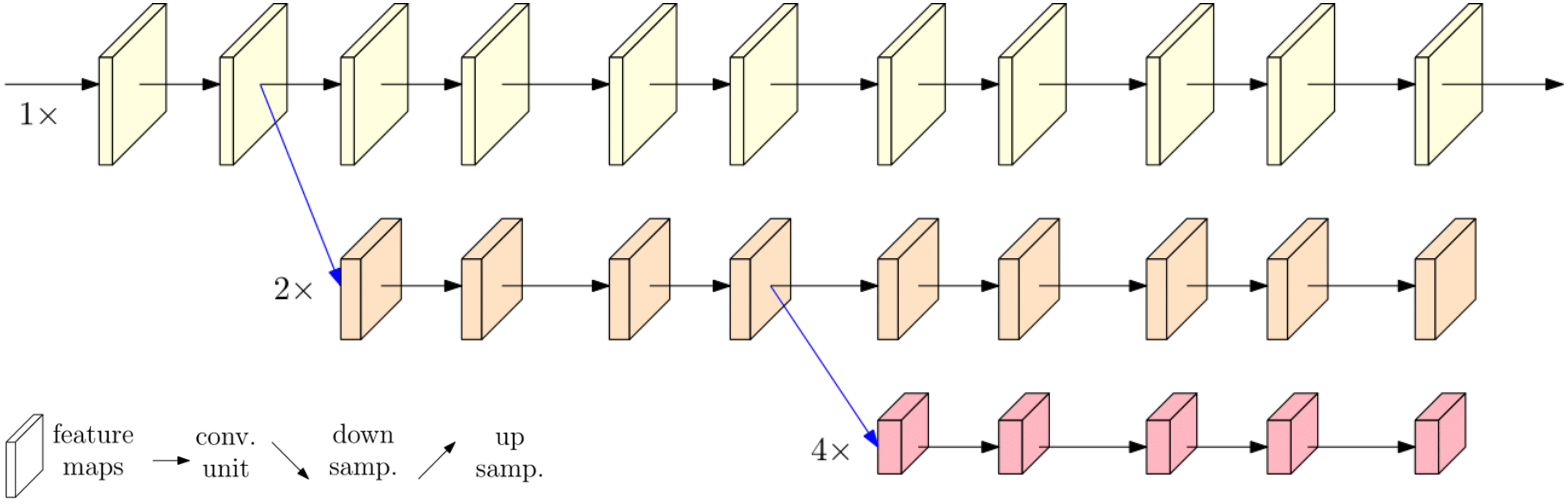
Previous classification networks

Connect multi-resolution convolutions in *series*



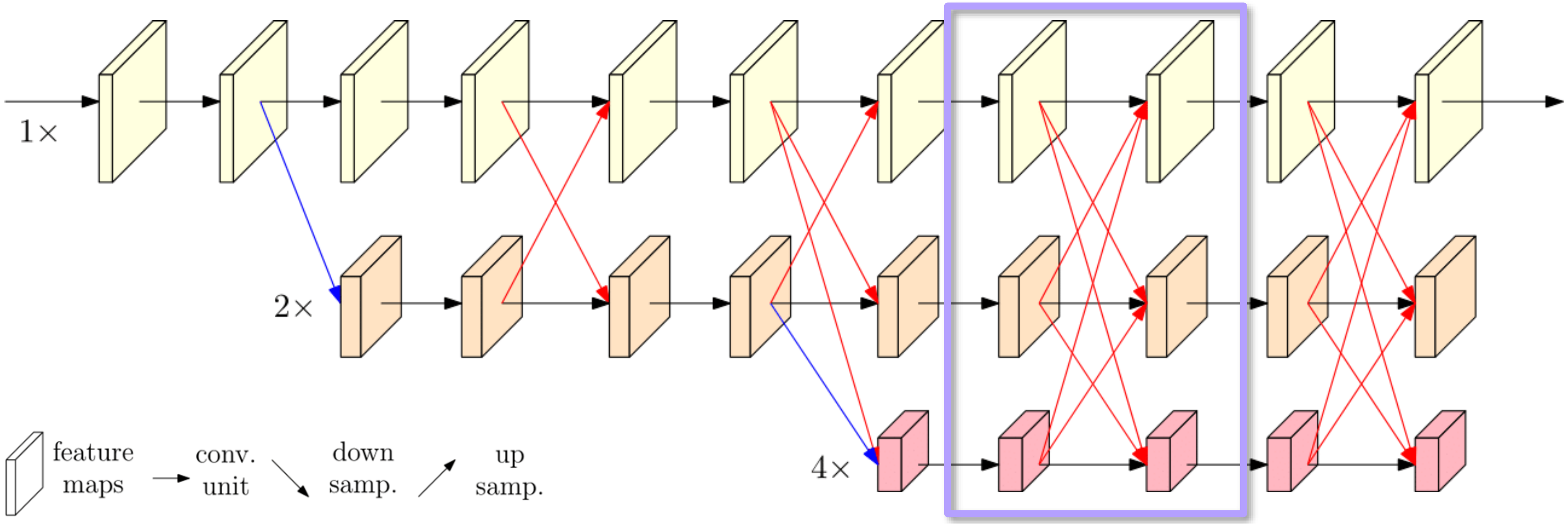
HRNet: high-resolution representation learning

High-resolution networks (HRNet): Connect multi-resolution convolutions in ***parallel***

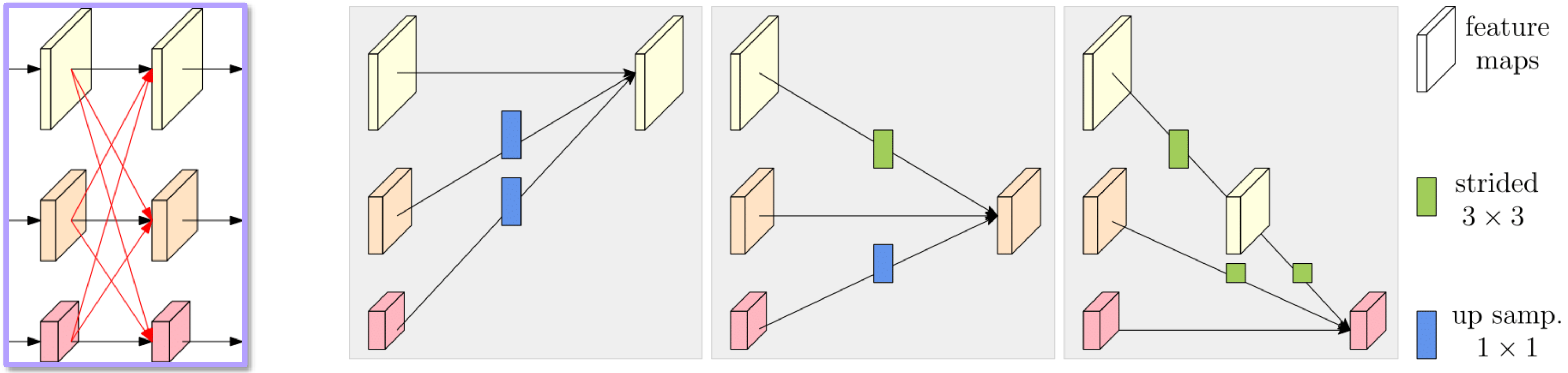


HRNet: high-resolution representation learning

High-resolution networks (HRNet): Connect multi-resolution convolutions in ***parallel*** with ***repeated fusions***



Across-resolution fusion



Down-sample: $\mathbf{stride} - 2\ 3 \times 3$

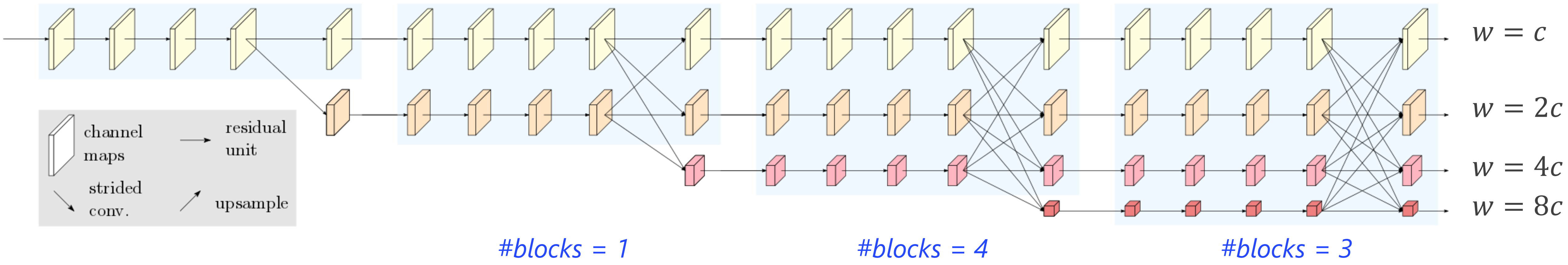
Up-sample: $\mathbf{bilinear} + 1 \times 1$

Fundamental architecture changes

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

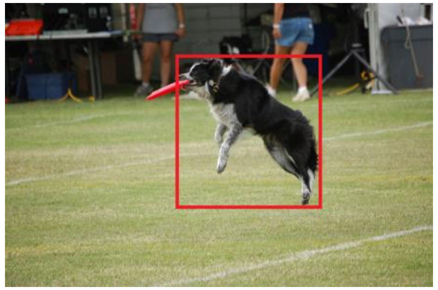
HRNet can learn *high-resolution strong* representations

HRNet instantiation



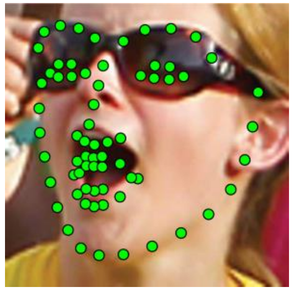
Visual recognition applications

Image classification



Object detection

Semantic segmentation

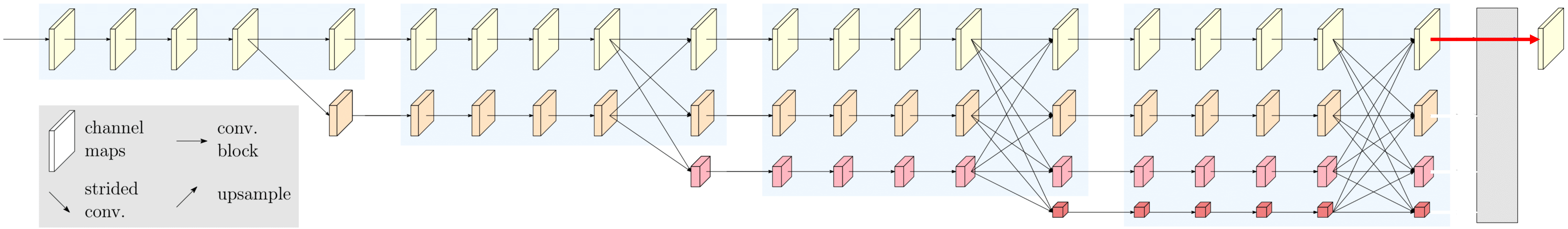


Face alignment

Pose estimation

Human pose estimation

HRNetV1



Datasets and evaluation

Datasets	training	validation	testing	Evaluation
COCO 2017	57K	5000 images	20K	AP@OKS
MPII	13K		12k	PCKh
PoseTrack	292 videos	50	208	mAP/MOTA

COCO: http://cocodataset.org/#keypoints-eval
MPII http://human-pose.mpi-inf.mpg.de/
PoseTrack https://posetrack.net/

COCO

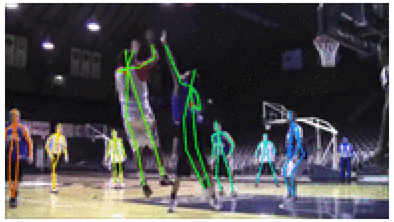
COCO validation

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	50.3M	12.4	71.4	89.3	79.3	68.1	78.1	77.1
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
SimpleBaseline [66]	ResNet-152	Y	256×192	68.6M	15.7	72.0	89.3	79.8	68.7	78.9	77.8
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], *CMU*	-	-	-	-	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], *Google*	-	-	-	-	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], *Facebook*	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] , *Face++*	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], *Face++*	ResNet-Inception	384×288	-	-	73.0	91.7	80.9	69.5	78.1	79.0
SimpleBaseline [72], *Microsoft*	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**

PoseTrack2017: multi-person pose estimation & tracking



PoseTrack Leaderboard

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

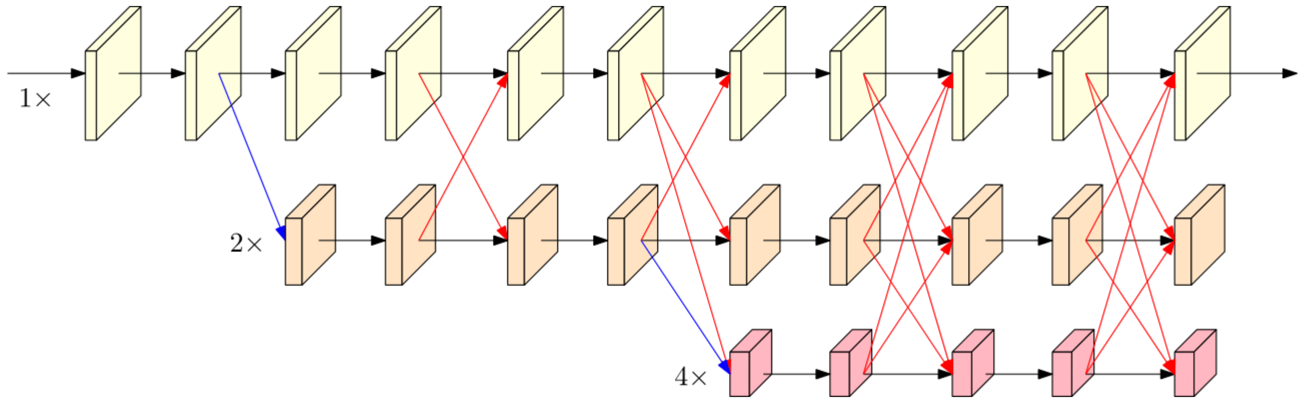
Multi-Person Pose Tracking

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

https://posetrack.net/leaderboard.php
by Feb. 28, 2019

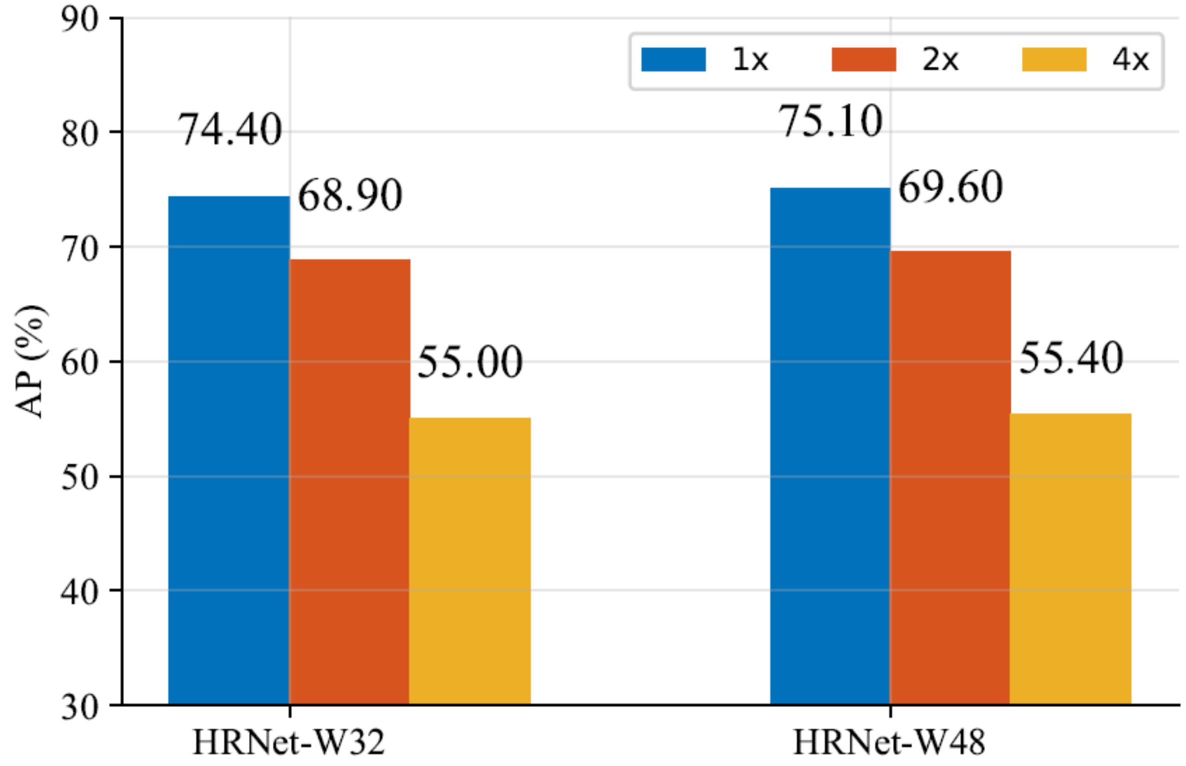
Ablation study: repeated across-resolution fusion

Method	Final exchange	Int. exchange across	Int. exchange within	AP
(a)	✓			70.8
(b)	✓	✓		71.9
(c)	✓	✓	✓	73.4



COCO, train from scratch

Ablation study: high- and low-resolution representations



COCO, train from scratch

Visual recognition applications

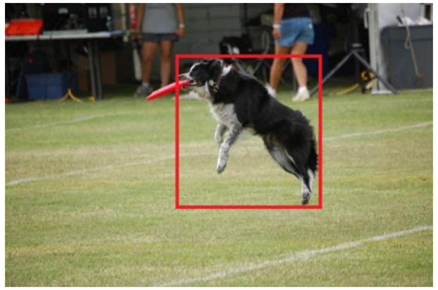


Image classification

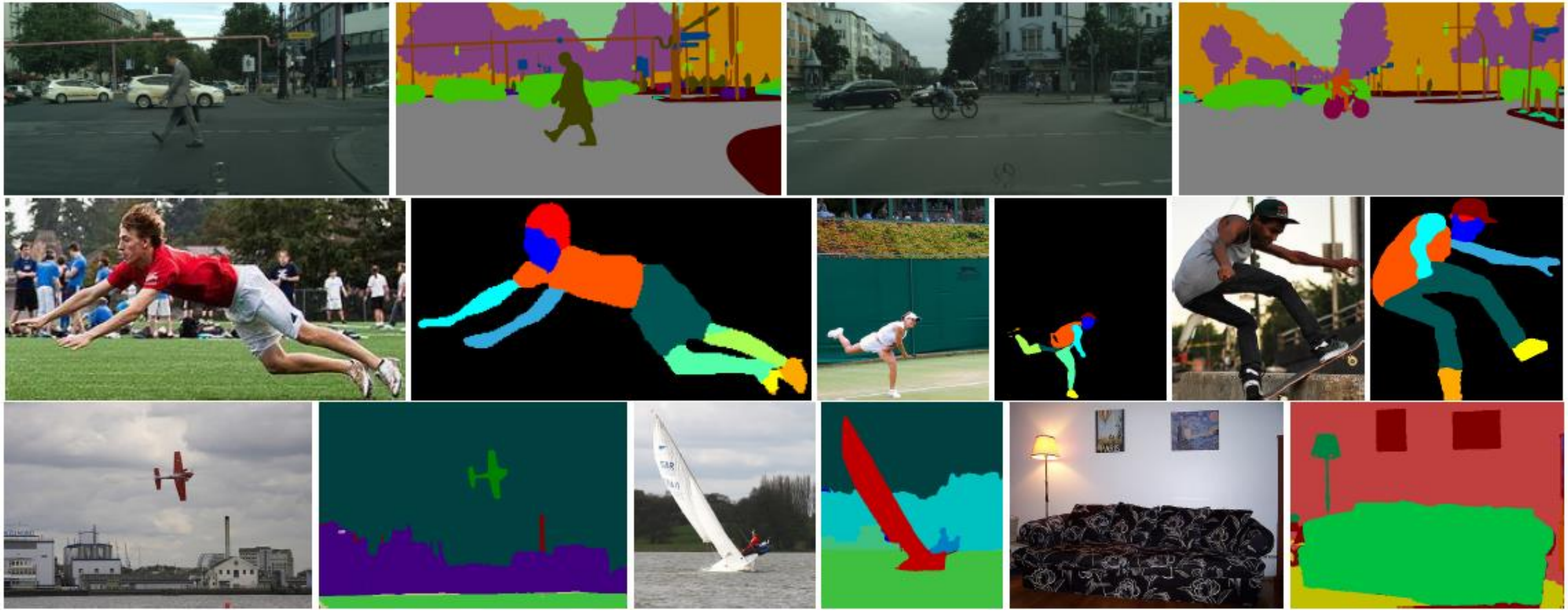
Object detection

Semantic segmentation

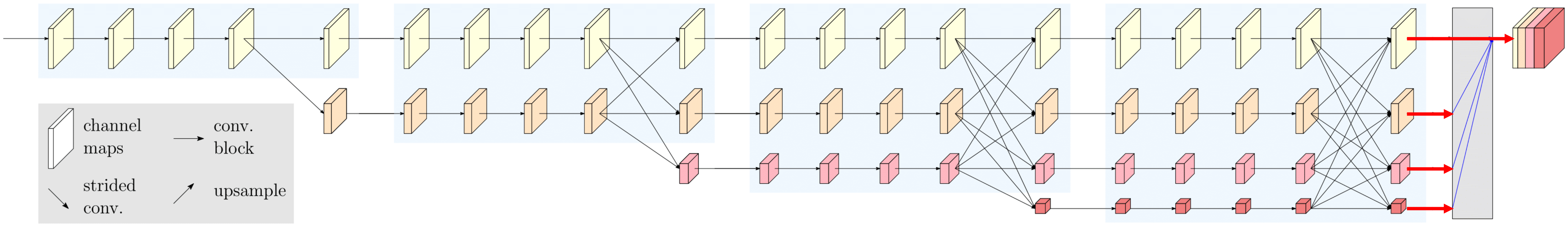
Face alignment

Pose estimation

Semantic segmentation



HRNetV2



Datasets and evaluation

Datasets	training	validation	testing	#classes	Evaluation
Cityscapes	2975	500	1525	19+1	mIoU
PASCAL context	4998		5105	59+1	mIoU
LIP	30462		10000	19+1	mIoU

Cityscapes validation

	backbone	#Params.	GFLOPs	mIoU
U-Net++ [130]	ResNet-101	59.5M	748.5	75.5
DeepLabv3 [14], *Google*	Dilated-resNet-101	58.0M	1778.7	78.5
DeepLabv3+ [16], *Google*	Dilted-Xception-71	43.5M	1444.6	79.6
PSPNet [123], *SenseTime*	Dilated-ResNet-101	65.9M	2017.6	79.7
Our approach	HRNetV2-W40	45.2M	493.2	80.2
Our approach	HRNetV2-W48	65.9M	747.3	**81.1**

Cityscapes test

	backbone	mIoU	iIoU cat.	IoU cat.	iIoU cat.
Model learned on the train+valid set					
GridNet [130]	-	69.5	44.1	87.9	71.1
LRR-4x [33]	-	69.7	48.0	88.2	74.7
DeepLab [13], *Google*	Dilated-ResNet-101	70.4	42.6	86.4	67.7
LC [54]	-	71.1	-	-	-
Piecewise [60]	VGG-16	71.6	51.7	87.3	74.1
FRRN [77]	-	71.8	45.5	88.9	75.1
RefineNet [59]	ResNet-101	73.6	47.2	87.9	70.6
PEARL [42]	Dilated-ResNet-101	75.4	51.6	89.2	75.1
DSSPN [58]	Dilated-ResNet-101	76.6	56.2	89.6	77.8
LKM [75]	ResNet-152	76.9	-	-	-
DUC-HDC [97]	-	77.6	53.6	90.1	75.2
SAC [117]	Dilated-ResNet-101	78.1	-	-	-
DepthSeg [46]	Dilated-ResNet-101	78.2	-	-	-
ResNet38 [101]	WResNet-38	78.4	59.1	90.9	78.1
BiSeNet [111]	ResNet-101	78.9	-	-	-
DFN [112]	ResNet-101	79.3	-	-	-
PSANet [125], *SenseTime*	Dilated-ResNet-101	80.1	-	-	-
PADNet [106]	Dilated-ResNet-101	80.3	58.8	90.8	78.5
DenseASPP [124]	WDenseNet-161	80.6	59.1	90.9	78.1
Our approach	HRNetV2-w48	**81.6**	**61.8**	**92.1**	**82.2**

PASCAL context

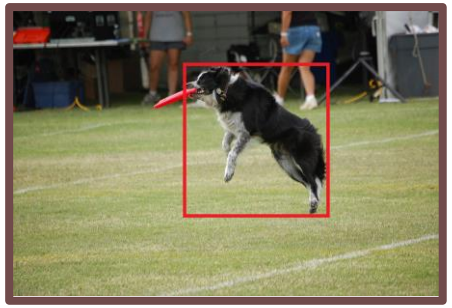
	backbone	mIoU (59classes)	mIoU (60classes)
FCN-8s [86]	VGG-16	-	35.1
BoxSup [20]	-	-	40.5
HO_CRF [1]	-	-	41.3
Piecewise [60]	VGG-16	-	43.3
DeepLabv2 [13], *Google*	Dilated-ResNet-101	-	45.7
RefineNet [59]	ResNet-152	-	47.3
U-Net++ [130]	ResNet-101	47.7	-
PSPNet [123], *SenseTime*	Dilated-ResNet-101	47.8	-
Ding et al. [23]	ResNet-101	51.6	-
EncNet [114]	Dilated-ResNet-101	52.6	-
Our approach	HRNetV2-W48	**54.0**	**48.3**

LIP

	backbone	extra	pixel acc.	avg. acc.	mIoU
Attention+SSL [34]	VGG-16	Pose	84.36	54.94	44.73
DeepLabv2 [16], *Google*	Dilated-ResNet-101	-	84.09	55.62	44.80
MMAN[67]	Dilated-ResNet-101	-	-	-	46.81
SS-NAN [125]	ResNet-101	Pose	87.59	56.03	47.92
MuLA [72]	Hourglass	Pose	88.50	60.50	49.30
JPPNet [57]	Dilated-ResNet-101	Pose	86.39	62.32	51.37
CE2P [65]	Dilated-ResNet-101	Edge	87.37	63.20	53.10
Our approach	HRNetV2-W48	N	**88.21**	**67.43**	**55.90**

Visual recognition applications

Image classification

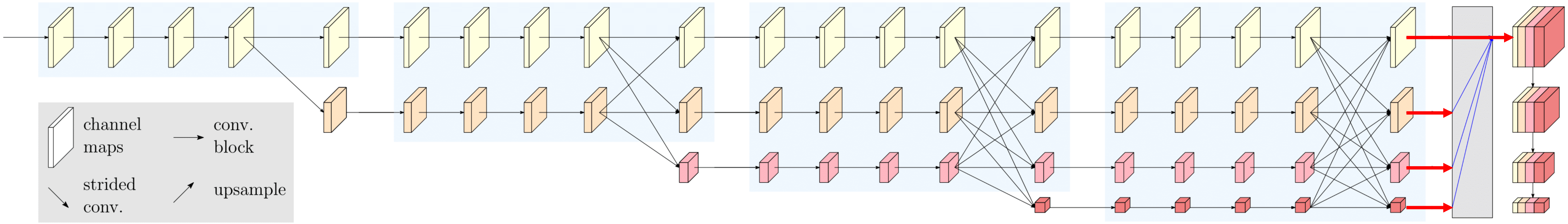


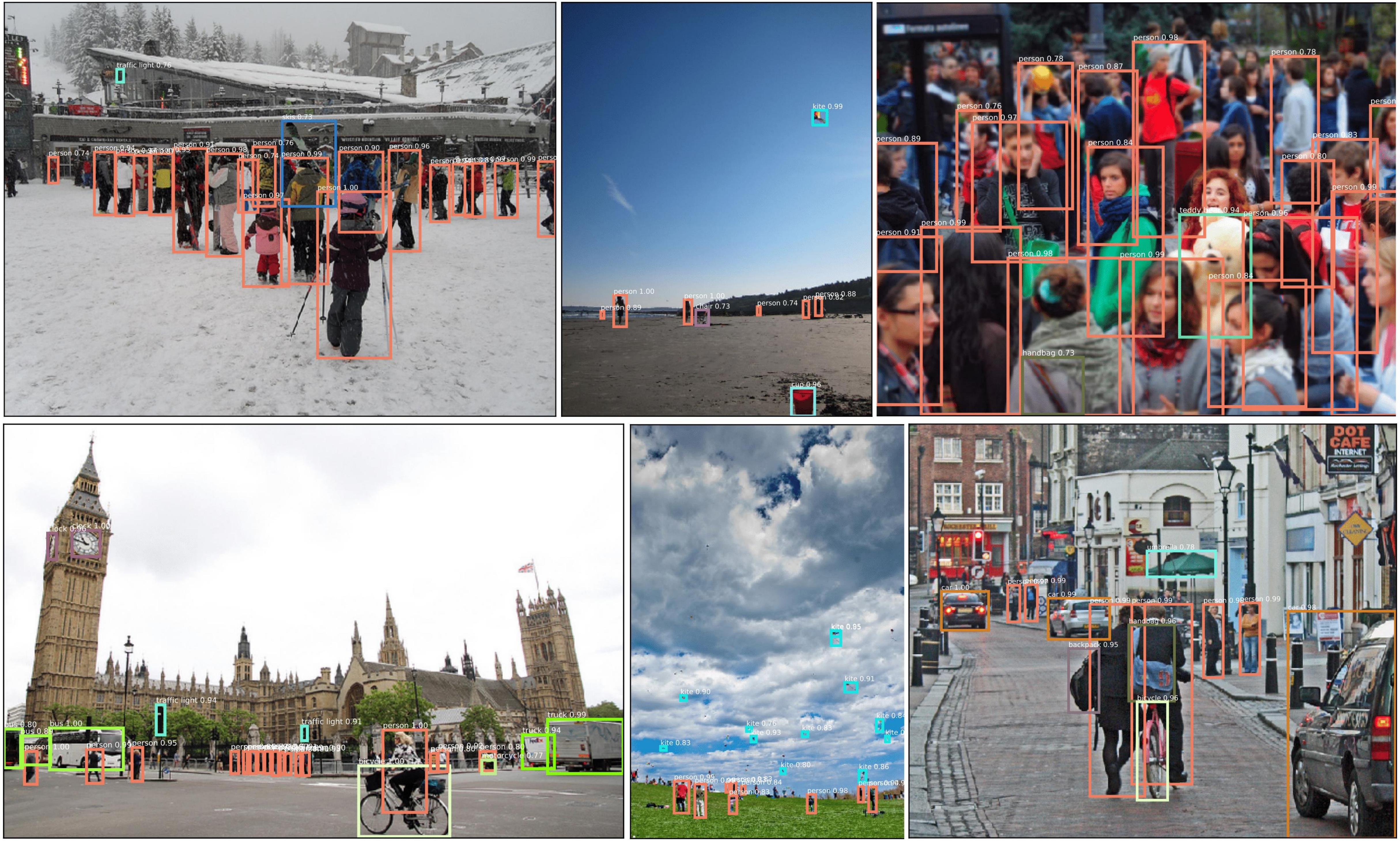
Object detection

Semantic segmentation

Pose estimation

HRNetV2h





traffic light 0.76
skis 0.73
person 0.74
person 0.91
person 0.76
person 0.98
person 0.74
person 0.99
person 0.90
person 0.96
person 0.97
person 1.00
kite 0.99
person 1.00
person 0.89
person 1.00
chair 0.73
person 0.74
person 0.88
person 0.82
cup 0.96
person 0.98
person 0.78
person 0.87
person 0.78
person 0.76
person 0.97
person 0.89
person 0.84
person 0.83
person 0.80
person 0.99
person 0.91
person 0.96
person 0.98
person 0.99
person 0.84
handbag 0.73
clock 0.96
clock 1.00
traffic light 0.94
traffic light 0.91
truck 0.99
truck 0.94
bus 0.80
bus 0.89
bus 1.00
person 1.00
person 0.95
person 0.80
motorcycle 0.77
bicycle 1.00
kite 0.95
kite 0.91
kite 0.90
kite 0.76
kite 0.93
kite 0.83
kite 0.80
kite 0.86
person 0.98
person 0.83
umbrella 0.78
car 1.00
car 0.99
car 0.98
handbag 0.96
backpack 0.95
bicycle 0.96
DOT CAFE
INTERNET

Single model single scale

	backbone	Size	LS	AP	AP_{50}	AP_{75}	AP_S	AP_M	AP_L
DFPR [47]	ResNet-101	512	1 ×	34.6	54.3	37.3	-	-	-
PFPNet [45]	VGG16	512	-	35.2	57.6	37.9	18.7	38.6	45.9
RefineDet [118]	ResNet-101-FPN	512	-	36.4	57.5	39.5	16.6	39.9	51.4
RelationNet [40]	ResNet-101	600	-	39.0	58.6	42.9	-	-	-
C-FRCNN [18]	ResNet-101	800	1 ×	39.0	59.7	42.8	19.4	42.4	53.0
RetinaNet [62]	ResNet-101-FPN	800	1.5 ×	39.1	59.1	42.3	21.8	42.7	50.2
Deep Regionlets [107]	ResNet-101	800	1.5 ×	39.3	59.8	-	21.7	43.7	50.9
FitnessNMS [94]	ResNet-101	768		39.5	58.0	42.6	18.9	43.5	54.1
DetNet [56]	DetNet-59-FPN	800	2 ×	40.3	62.1	43.8	23.6	42.6	50.0
CornerNet [51]	Hourglass-104	511		40.5	56.5	43.1	19.4	42.7	53.9
M2Det [126]	VGG16	800	~ 10 ×	41.0	59.7	45.0	22.1	**46.5**	53.8
Faster R-CNN [61]	ResNet-101-FPN	800	1 ×	39.3	61.3	42.7	22.1	42.1	49.7
Faster R-CNN	HRNetV2p-W32	800	1 ×	39.5	61.2	43.0	23.3	41.7	49.1
Faster R-CNN [61]	ResNet-101-FPN	800	2 ×	40.3	61.8	43.9	22.6	43.1	51.0
Faster R-CNN	HRNetV2p-W32	800	2 ×	41.1	62.3	44.9	24.0	43.1	51.4
Faster R-CNN [61]	ResNet-152-FPN	800	2 ×	40.6	62.1	44.3	22.6	43.4	52.0
Faster R-CNN	HRNetV2p-W40	800	2 ×	42.1	63.2	46.1	24.6	44.5	52.6
Faster R-CNN [11]	ResNeXt-101-64x4d-FPN	800	2 ×	41.1	62.8	44.8	23.5	44.1	52.3
Faster R-CNN	HRNetV2p-W48	800	2 ×	42.4	**63.6**	46.4	24.9	44.6	53.0
Cascade R-CNN [9]*	ResNet-101-FPN	800	~ 1.6 ×	42.8	62.1	46.3	23.7	45.5	55.2
Cascade R-CNN	ResNet-101-FPN	800	~ 1.6 ×	43.1	61.7	46.7	24.1	45.9	55.0
Cascade R-CNN	HRNetV2p-W32	800	~ 1.6 ×	**43.7**	62.0	**47.4**	**25.5**	46.0	**55.3**


sports ball 0.85
boat 0.96
person 1.00
ONE WAY
NO TURNS
traffic light 0.96
traffic light 0.93
car 1.00
bus 0.77
person 0.99
bicycle 0.88

Mask R-CNN

backbone	LS	mask				bbox			
		AP	AP_S	AP_M	AP_L	AP	AP_S	AP_M	AP_L
ResNet-50-FPN	1 ×	34.2	15.7	36.8	50.2	37.8	22.1	40.9	49.3
HRNetV2p-W18	1 ×	33.8	15.6	35.6	49.8	37.1	21.9	39.5	47.9
ResNet-50-FPN	2 ×	35.0	16.0	**37.5**	**52.0**	38.6	21.7	41.6	50.9
HRNetV2p-W18	2 ×	**35.3**	**16.9**	**37.5**	51.8	**39.2**	**23.7**	**41.7**	**51.0**
ResNet-101-FPN	1 ×	36.1	16.2	39.0	53.0	40.0	22.6	43.4	52.3
HRNetV2p-W32	1 ×	36.7	17.3	39.0	53.0	40.9	24.5	43.9	52.2
ResNet-101-FPN	2 ×	36.7	17.0	39.5	54.8	41.0	23.4	44.4	53.9
HRNetV2p-W32	2 ×	**37.6**	**17.8**	**40.0**	**55.0**	**42.3**	**25.0**	**45.4**	**54.9**

More detection and instance segmentation results under FCOS, CenterNet, and Hybrid Task Cascade are available in [1]

[1] Jingdong Wang, Ke Sun, Tianheng Cheng, Borui Jiang, Chaorui Deng, Yang Zhao, Dong Liu, Yadong Mu, Mingkui Tan, Xinggang Wang, Wenyu Liu, and Bin Xiao: Deep High-Resolution Representation Learning for Visual Recognition (https://arxiv.org/abs/1908.07919, submitted to TPAMI)

Visual recognition applications

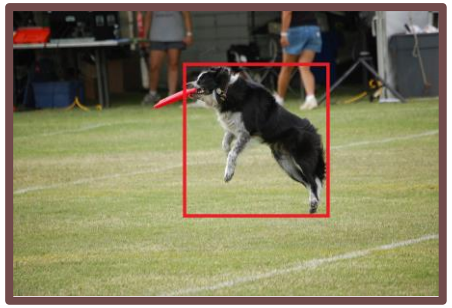


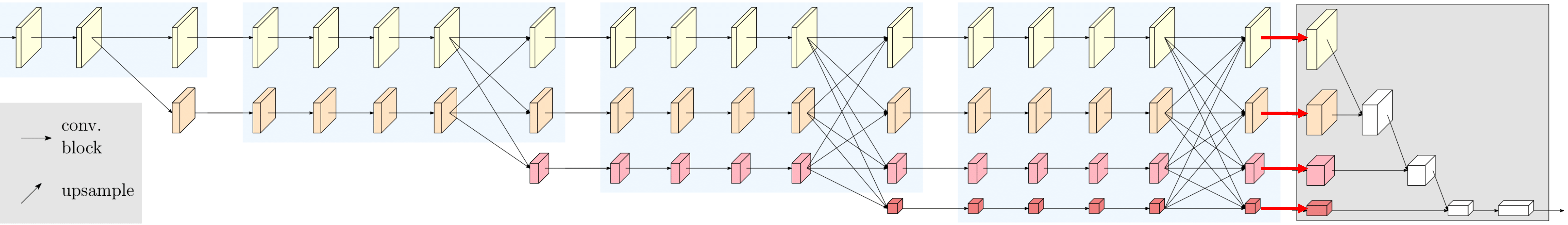
Image classification

Object detection

Semantic segmentation

Pose estimation

HRNets for ImageNet classification



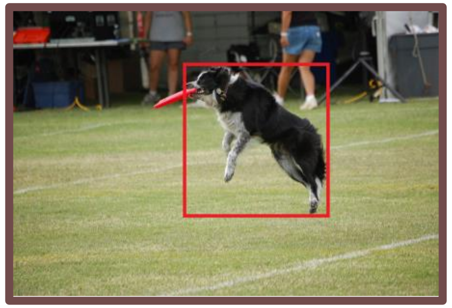
ImageNet classification

	#Params.	GFLOPs	Top-1 err.	Top-5 err.
Residual branch formed by two 3 × 3 convolutions				
ResNet-38	28.3M	3.80	24.6%	7.4%
HRNet-W18	21.3M	3.99	**23.1%**	**6.5%**
ResNet-71	48.4M	7.46	23.3%	6.7%
HRNet-W30	37.7M	7.55	**21.9%**	**5.9%**
ResNet-105	64.9M	11.1	22.7%	6.4%
HRNet-W40	57.6M	11.8	**21.1%**	**5.6%**
Residual branch formed a bottleneck				
ResNet-50	25.6M	3.82	23.3%	6.6%
HRNet-W44	21.9M	3.90	**23.0%**	**6.5%**
ResNet-101	44.6M	7.30	21.6%	5.8%
HRNet-W76	40.8M	7.30	**21.5%**	**5.8%**
ResNet-152	60.2M	10.7	21.2%	5.7%
HRNet-W96	57.5M	10.2	**21.0%**	**5.7%**

Surprisingly, HRNet performs slightly better than ResNet

Visual recognition applications

Image classification



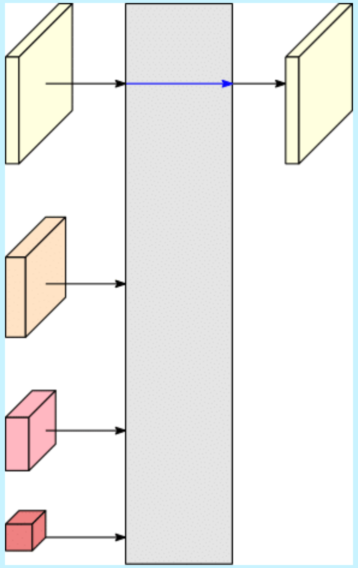
Object detection

Semantic segmentation

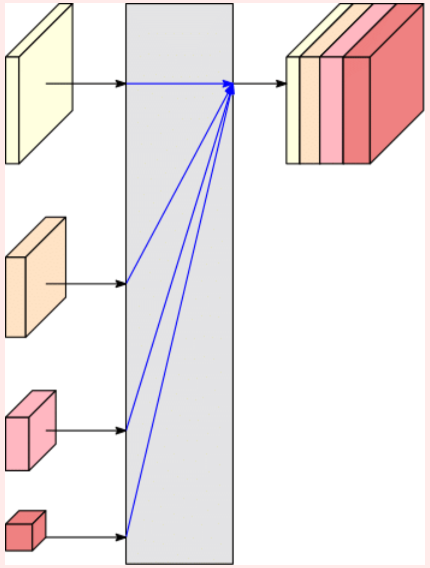
Face alignment

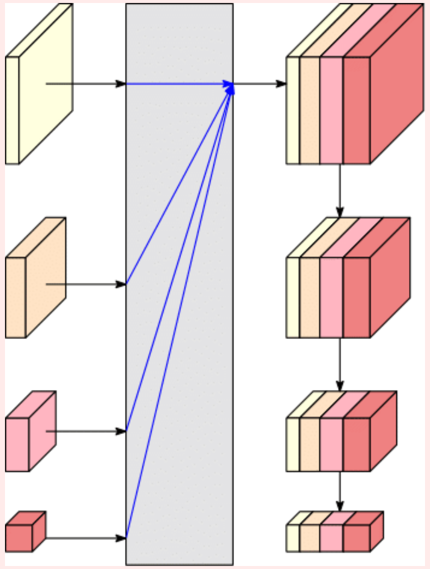
Pose estimation

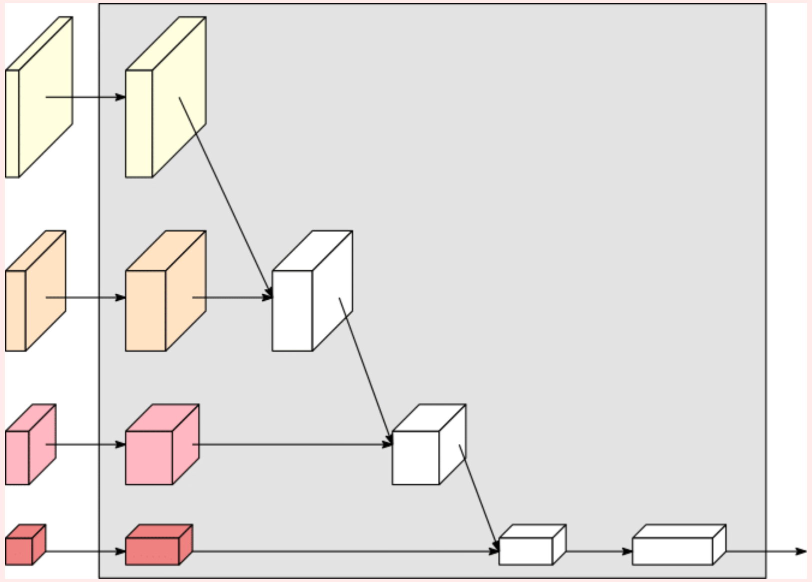
HRNetV1

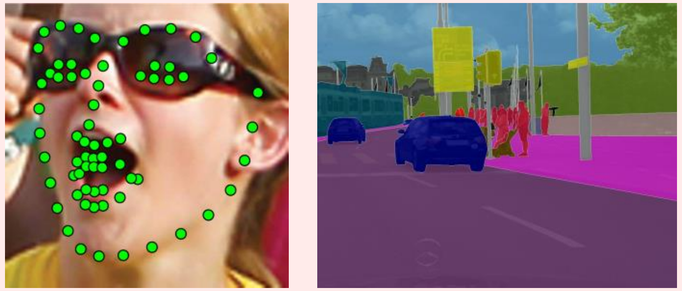


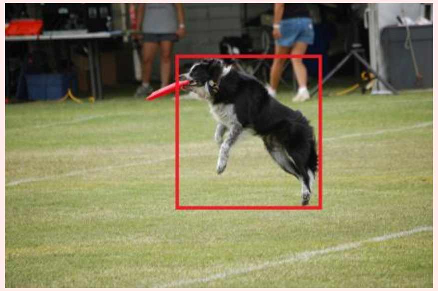
HRNetV2



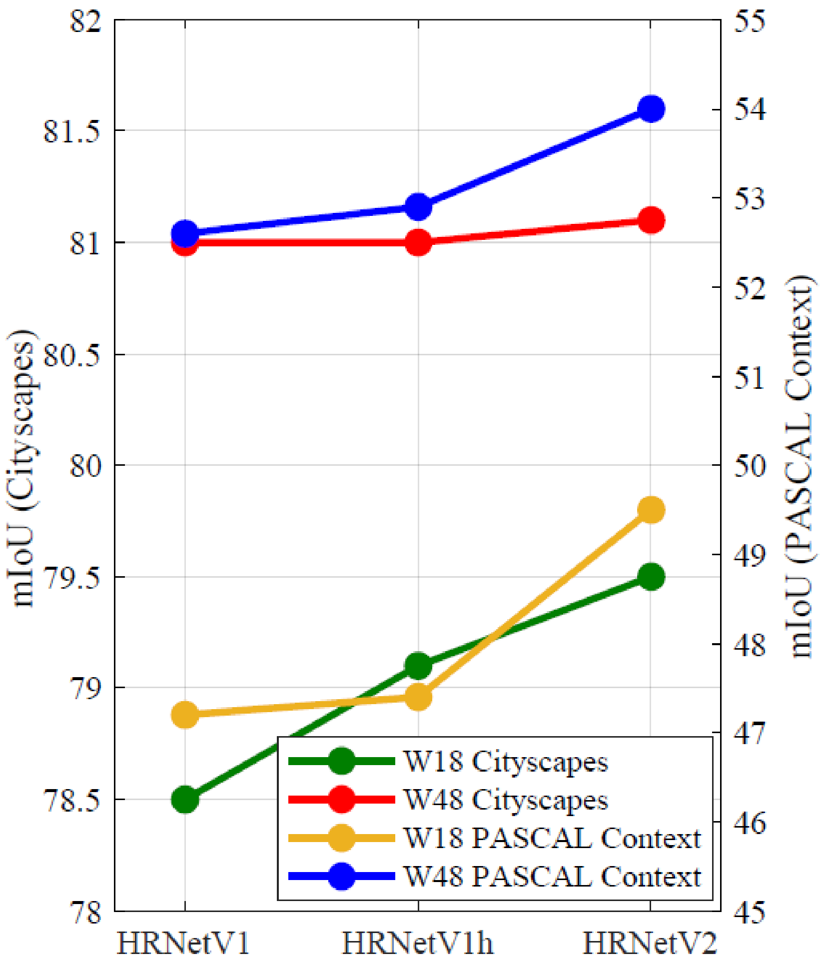




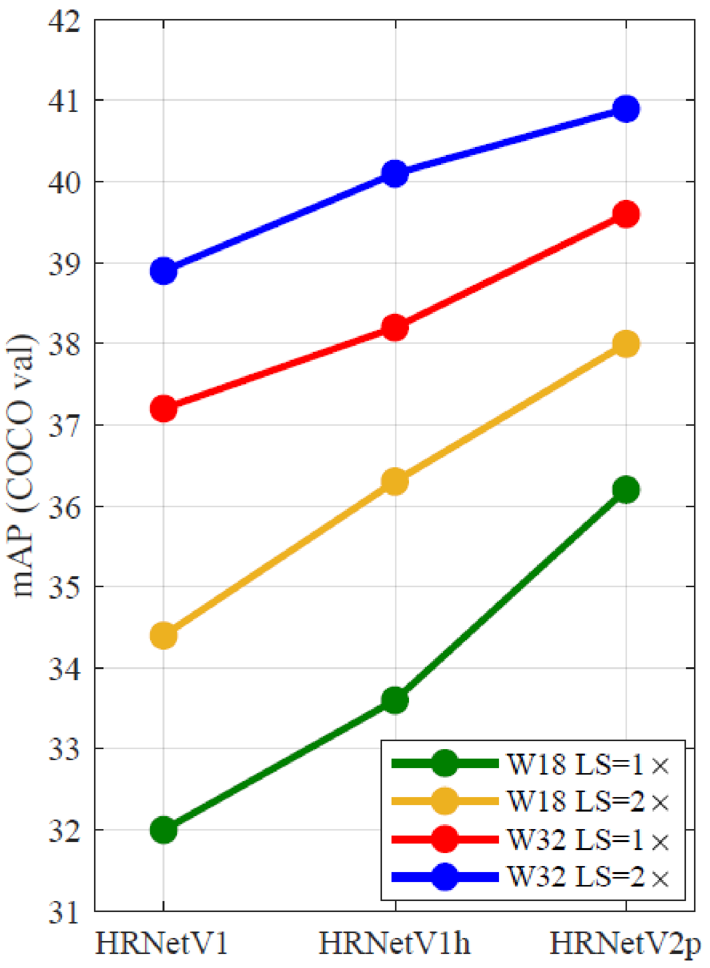




HRNetV1 vs. HRNetV2



Cityscapes and pascal context



COCO detection

Discussions

high-resolution networks
vs
classification networks

neural architecture design (NAD)
vs
neural architecture search (NAS)

High-resolution networks vs classification networks

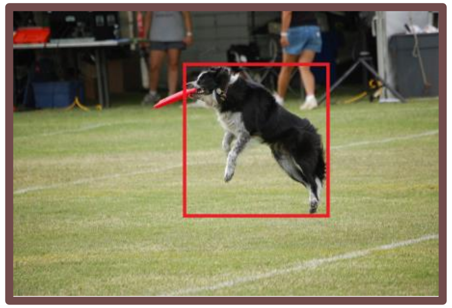




image-level | region-level | pixel-level

Low resolution
High resolution

Recover from low-resolution (ResNet, VGGNet) ✗
High-resolution (our HRNet) ✓

Discussions

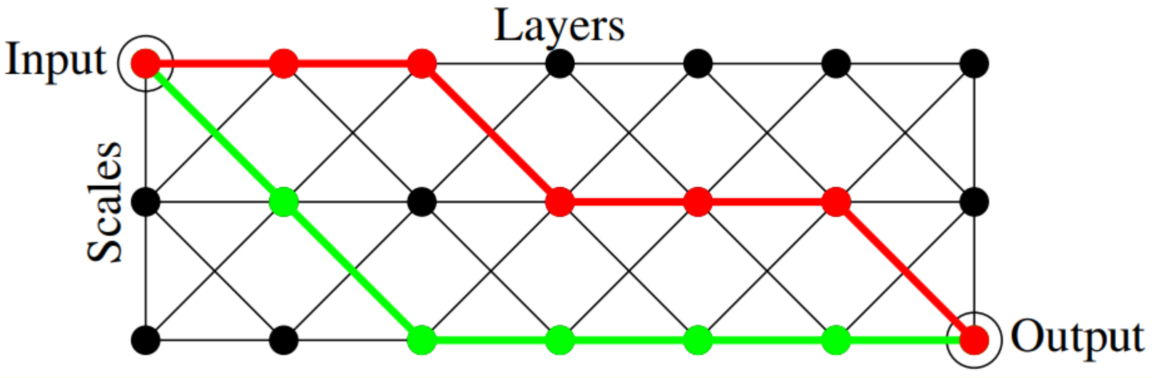
high-resolution networks
vs
classification networks

neural architecture design (NAD)
vs
neural architecture search (NAS)

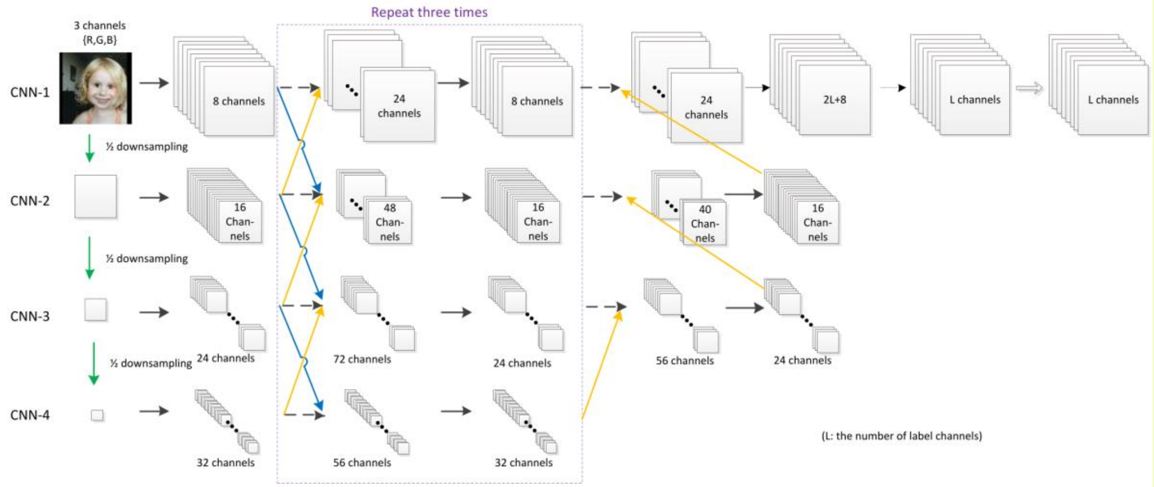
NAD expands search space for NAS

Related networks

Improper fusion frequency

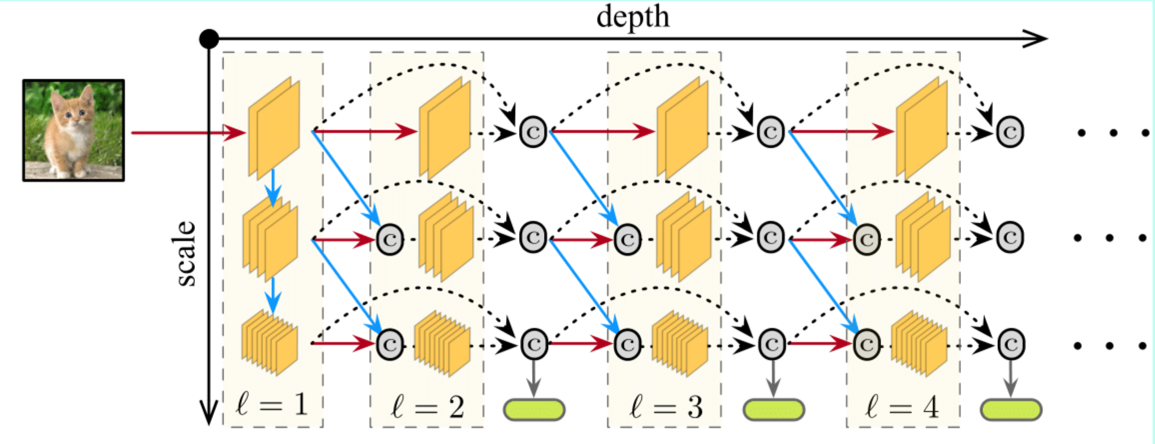


Convolutional neural fabrics

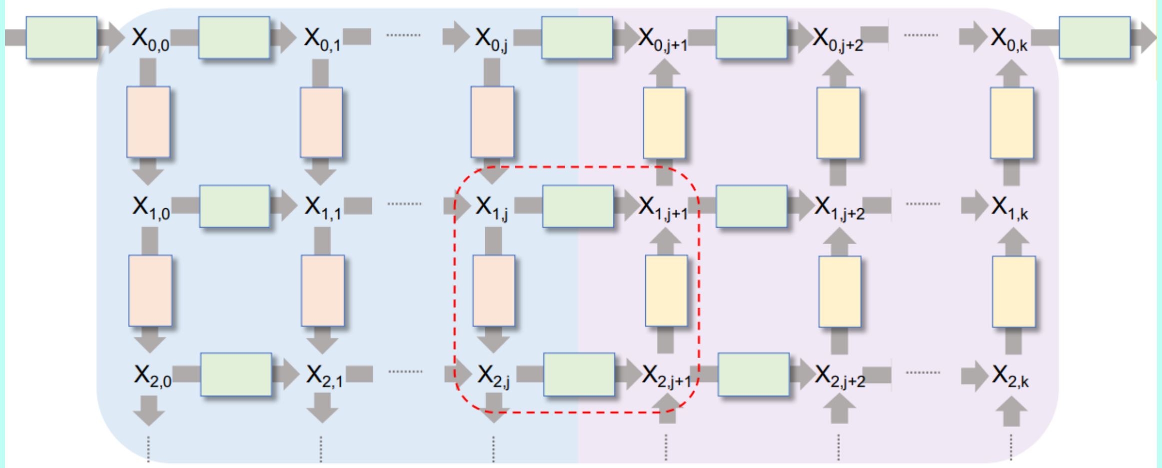


Interlinked CNN

unidirectional fusion

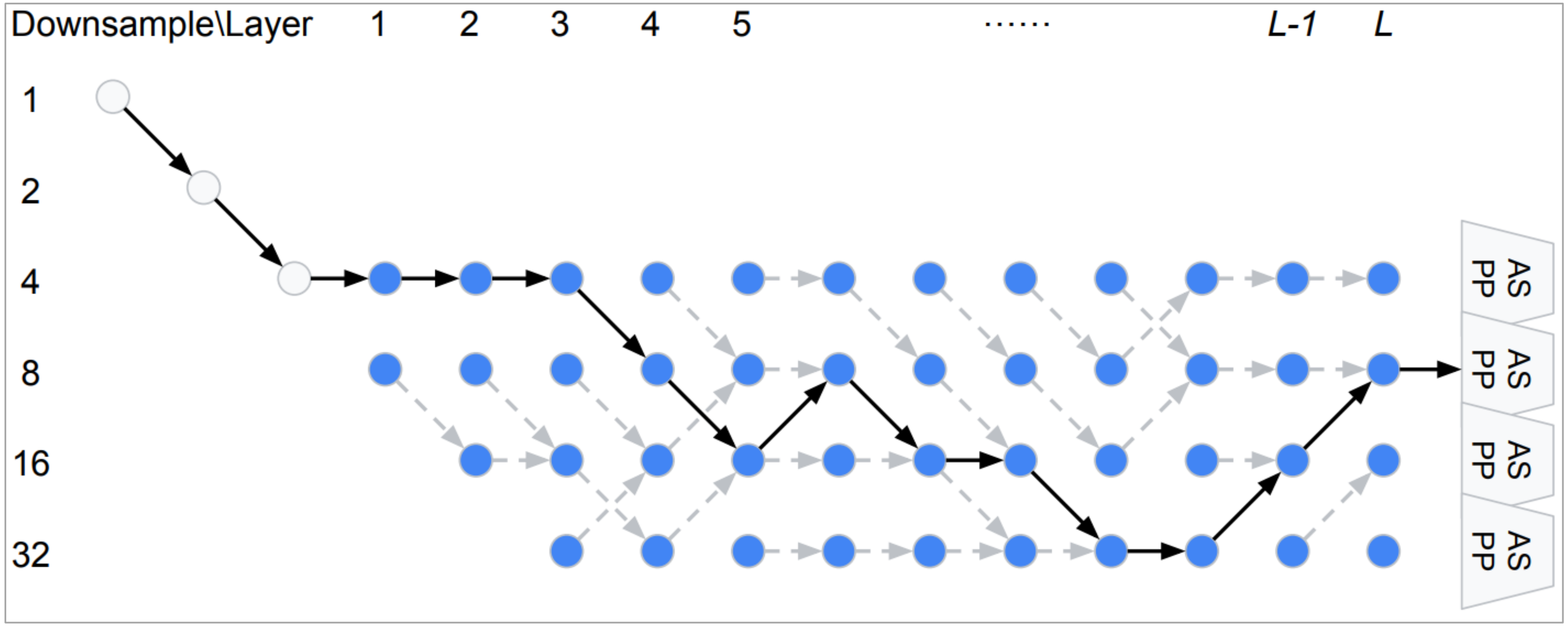


Multi-scale densenet



Gridnet: generalized U-Net

Auto-DeepLab *by Google*



Related to HRNet, but no high-resolution maintenance

Visual recognition applications

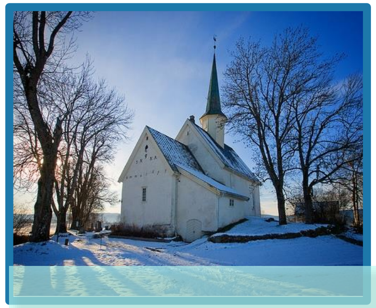
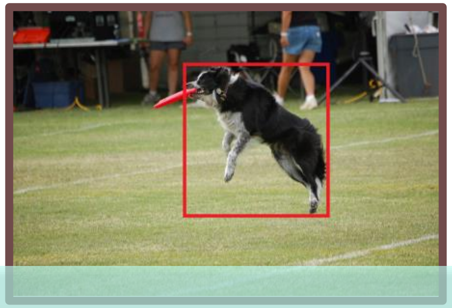


Image classification



Object detection

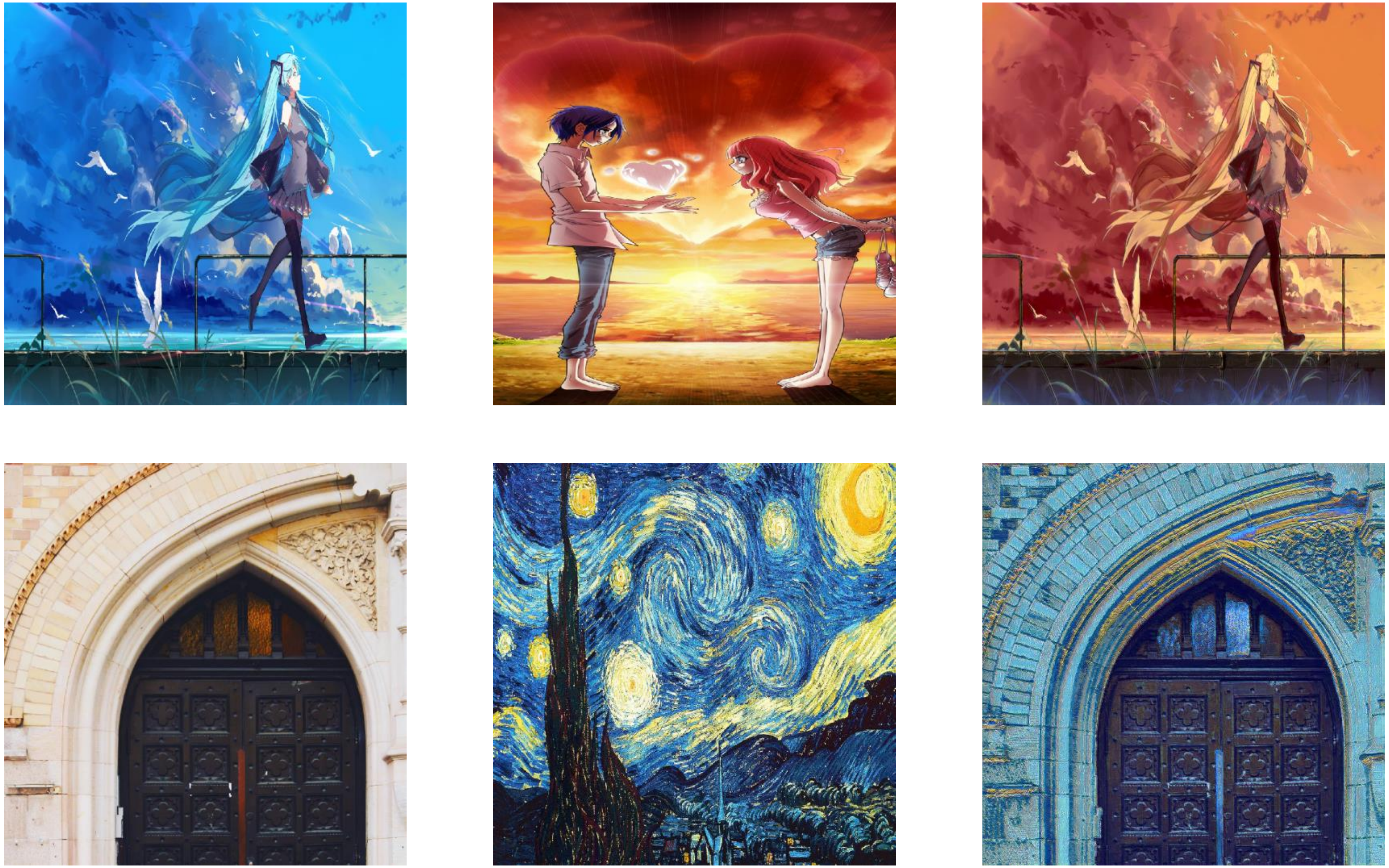
Semantic segmentation

Face alignment

Pose estimation

and …

Photo style transfer



Ming Li, Chunyang Ye, Wei Li: High-Resolution Network for Photorealistic Style Transfer. CoRR abs/1904.11617 (2019)

Image inpainting

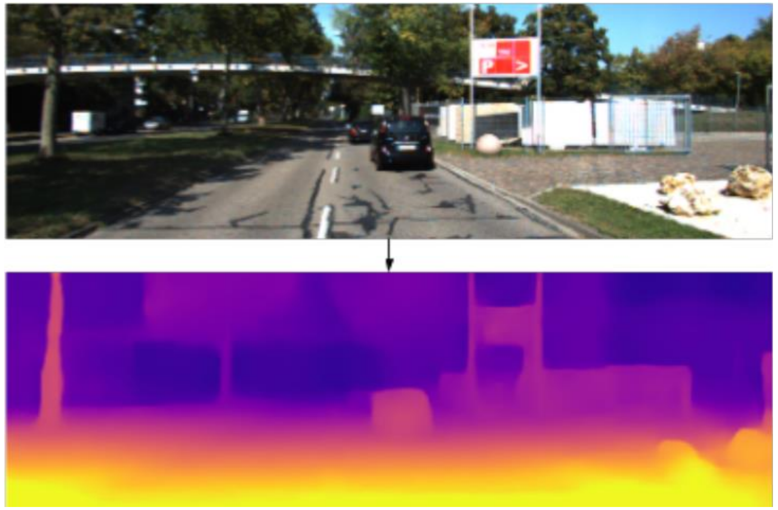
(a) Ground Truth (b) Damaged image (c) Once dilation (d) Twice dilation (e) Final recovering

Progressive Image Inpainting with Full-Resolution Residual Network. Zongyu Guo, Zhibo Chen, Tao Yu, Jiale Chen, Sen Liu. ACM MM 2019

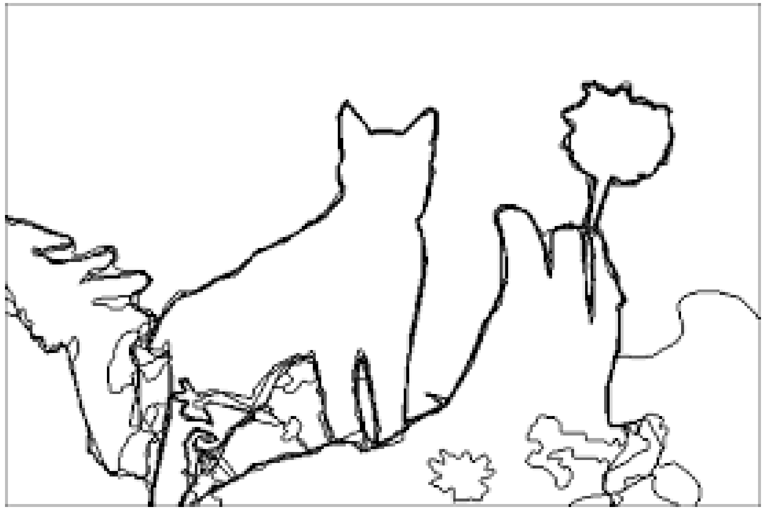
and ...

Super-resolution from LapSRN

Optical flow



Depth estimation



Edge detection

……

leoxiaobin / deep-high-resolution-net.pytorch

Watch 50 | Star 1,529 | Fork 313

Code | Issues 53 | Pull requests 0 | Projects 0 | Security | Insights

The project is an official implementation of our CVPR2019 paper "Deep High-Resolution Representation Learning for Human Pose Estimation" https://jingdongwang2017.github.io/Pr...

human-pose-estimation | deep-learning | coco-keypoints-detection | mpii-dataset | mpii | mscoco-keypoint | deep-high-resolution-net | high-resolution-net

20 commits | 1 branch | 0 releases | 3 contributors | MIT

Branch: master | New pull request | Find File | Clone or download

leoxiaobin Update README.md — Latest commit 00d7bf7 Jun 26, 2019

experiments	add config files for coco resnet	Mar 6, 2019
figures	init	Feb 25, 2019
lib	Merge pull request #39 from sunke123/master	Apr 10, 2019
tools	delete useless code	Apr 11, 2019
.gitignore	init	Feb 25, 2019
LICENSE	Initial commit	Feb 25, 2019
README.md	Update README.md	Jun 26, 2019
requirements.txt	init	Feb 25, 2019

README.md

Deep High-Resolution Representation Learning for Human Pose Estimation (accepted to CVPR2019)

HRNet

High-resolution representation learning. Code for pose estimation is available at https://github.com/leoxiaobin/deep-high-resolution-net.pytorch

https://jingdongwang2017.github.io/Projec... welleast@outlook.com

Repositories 6 People 4 Projects

Pinned repositories

HRNet-Object-Detection
Forked from open-mmlab/mmdetection

Object detection with multi-level r
generated from deep high-resolut
learning (HRNetV2h).

Python ★ 283 Fork 40

HRNet-Image-Classification
Train the HRNet model on ImageNet

Python ★ 242 Fork 36

HRNet-Semantic-Segmentation
High-resolution representation learning (HRNets) for

Python ★ 253 Fork 53

Used in many challenges in CVPR 2019

HRNet-Facial-Landmark-Detection
High-resolution representation learning (HRNets) for facial landmark detection

Python ★ 130 Fork 36

HRNet-MaskRCNN-Benchmark
Object detection with multi-level representations generated from deep high-resolution representation learning (HRNetV2h).

Python ★ 35 Fork 8

HRNet-FCOS
High-resolution Networks for the Fully Convolutional One-Stage Object Detection (FCOS) algorithm

Python ★ 15 Fork 2

Image enhancement challenge (CVPR 2019)

Meitu (美图) adopted the HRNet

Team	Reported Runtime, ms	quantitative results PSNR	SSIM	MOS
Mt.Stars	15	22.348760	0.790668	**2.784**
TeamInception	109	22.414647	0.791206	2.595
BMIPL_UNIST_DW	1500	22.442052	**0.802927**	2.591
BOE-IOT-AIBD	40	21.738756	0.781655	2.554
HIT-Xlab	520	22.143558	0.787604	2.529
TTI	500	22.165396	0.762101	2.527
Geometry	220	21.752385	0.782175	2.419
IVL	6.5	21.369116	0.718913	2.402
HIT-UltraVision	20000	21.921673	0.783971	2.394
ViPr	160	18.685189	0.729987	2.293
MENet	12	18.397164	0.756864	2.253
Rainbow	500	**22.663374**	0.796817	n/a *
MiRL	80	20.966558	0.744575	n/a *

NTIRE 2019 Challenge on Image Enhancement: Methods and Results. , CVPRW 2019

Image dehazing challenge (CVPR 2019)

Meitu (美图)
adopted the HRNet

Team	User (+entry)	PSNR	SSIM
iPAL-AtJ	moonriverLucy	**20.258**	**0.657**
iPAL-COLOR	DH-IRCNN_123_CEDH	19.923	0.653
MT.MaxClear	ucenter52	19.469	0.652
BMIPL-UNIST-DW-1	Sprite+Ours	18.842	0.633
xddqm	Untitled Folder	18.521	0.640
⋮	⋮	⋮	⋮

Hazy Image | IPAL-AtJ | IPAL_COLOR | MT.MaxClear

NTIRE 2019 Image Dehazing Challenge Report, CVPRW 2019

CVPR LONG BEACH CALIFORNIA June 16-20, 2019 L.I.P

Track 2: Single-Person Human Pose Estimation Challenge Winners:

1st:
Kai Su1, 2, *, Dongdong Yu1, *, Xin Geng2 , Changhu Wang1
1ByteDance AI Lab, 2Southeast University

1st:
Bin Xiao, Yifan Lu, Tang Tang, Hao Zhu, Linfu Wen
ByteDance AI Lab

3rd:
Juan Manuel Pérez Rúa, Kaiyang Zhou, Adrian Bualt, Xiatian Zhu, Tao Xiang, Maja Pantic
Samsung AI Center - Cambridge, UK

3rd:
Hong Hu, Feng Zhang, Hanbin Dai, Huan Luo, LiangBo Zhou, Mao Ye
University of Electronic Science and Technology of China(UESTC)

Track 1: Single-Person Human Parsing Challenge Winners:

1st:

Peike Li1, 2, Yunqiu Xu1, Yi Yang1, 2

1Baidu Research, 2CAI, University of Technology Sydney

2nd:

Dongdong Yu1, *, Kai Su1, 2, *, Jian Wang1, Kaihui Zhou1 , Xin Geng2 , Changhu Wang1

1ByteDance AI Lab, 2Southeast University

3rd:

Zhijie Zhang1, Wenguan Wang2, Jianbing Shen2, Siyuan Qi3, Yanwei Pang1 , Ling Shao2

1Tianjin University, 2Inception Institute of Artificial Intelligence,3UCLA

Track 3: Multi-Person Human Parsing Challenge Winners:

1st:

Yunqiu Xu1, Peike Li1, 2, Yi Yang1, 2
1Baidu Research, 2CAI, University of Technology Sydney

2nd:

Meng Zhang1, Xinchen Liu2, Wu Liu2, Anfu Zhou1, Huadong Ma1, Tao Mei2
1Beijing University of Posts and Telecommunications,
2AI Research of JD.com

3rd:

Bingke Zhu1, 2, Xiaomei Zhan1, 2, Yingying Chen1, 2, Ming Tang1, 2,
Hui Li3, Ting Zhang3, Zhaoliang Zhang3, Wenjie Tang3, Jinqiao Wang1, 2
1National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences,
2University of Chinese Academy of Sciences,
3R&D Center, China National Electronics Import & Export Corporation

Cityscapes leaderboard: Rank 1

https://www.cityscapes-dataset.com/benchmarks/

CITYSCAPES DATASET

News | Overview | Examples | Benchmarks | Download | Submit | Citation | Conta

	name	fine	coarse	16-bit	depth	video	sub	IoU class	iIoU class	IoU category	iIoU category	Runtime [s]	code
	HRNetV2 + OCR (w/ ASP)	yes	yes	no	no	no	no	83.7	64.8	92.4	83.5	n/a	yes
	iFLYTEK-CV	yes	yes	no	no	no	no	83.6	64.7	92.1	82.3	n/a	no
	Improving Semantic Segmentation via Video Propagation and Label Relaxation	yes	yes	no	no	yes	no	83.5	64.4	92.2	82.0	n/a	yes
	GALD-Net	yes	yes	yes	yes	no	no	83.3	64.5	92.3	81.9	n/a	yes

https://www.cityscapes-dataset.com/benchmarks/

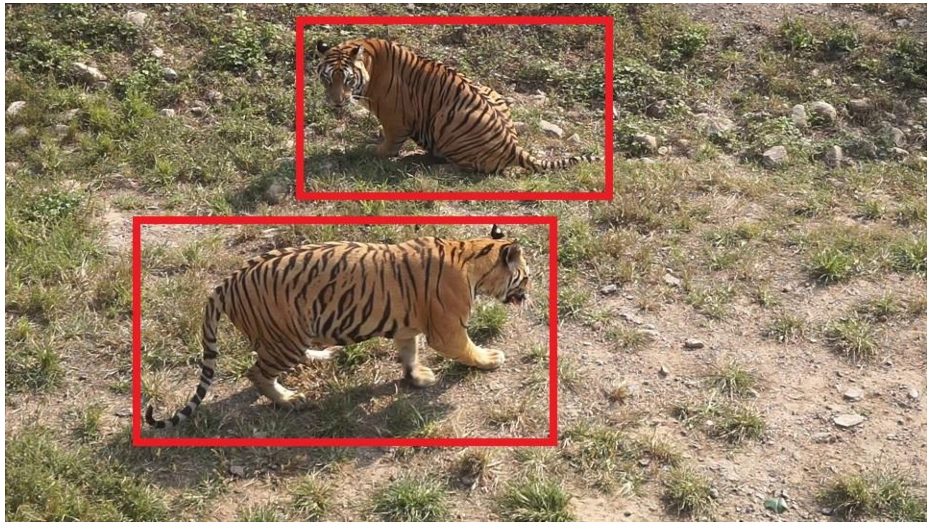
by Aug. 10, 2019

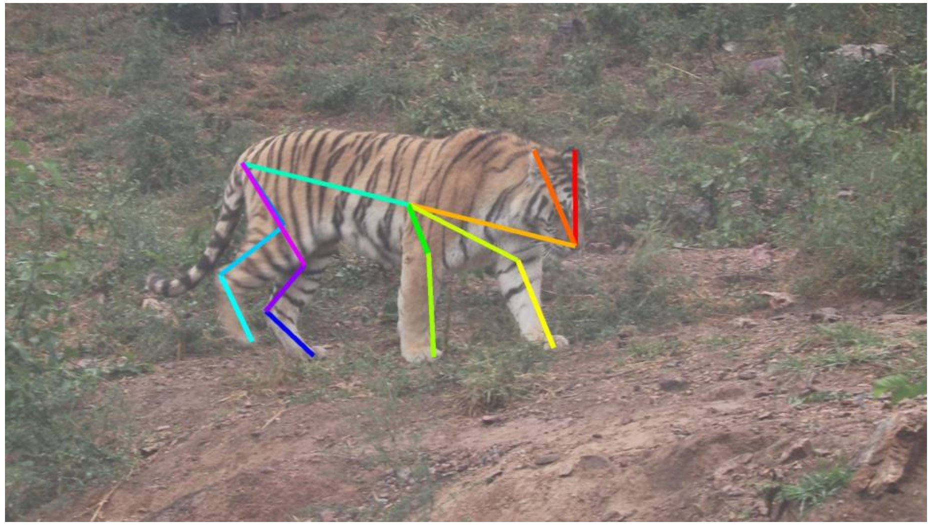
ICCV 2019 Workshop & Challenge on

Computer Vision for Wildlife Conservation (CVWC)

Oct-27, 2019 @ Room-E3, COEX Convention Center, Seoul, Korea

ICCV 2019 Seoul, Korea





One winner solution adopted the HRNet

Summary

Method: Maintain high-resolution representations through the whole process with repeated across-resolution fusions

Result: State-of-the-art performance in position-sensitive vision tasks: semantic segmentation, object detection, facial landmark detection, human pose estimation

Role: *Replace classification networks (e.g., ResNet) for computer vision tasks*

HRNet team


Jingdong Wang


Bin Xiao


Ke Sun


Yang Zhao


Borui Jiang


Tianheng Cheng


Dong Liu


Yadong Mu


Xinggang Wang


Wenyu Liu

Thanks!
Q&A

https://github.com/HRNet