Learning a Dynamic High-Resolution Network for Multi-Scale Pedestrian Detection

Mengyuan Ding, Shanshan Zhang , Jian Yang

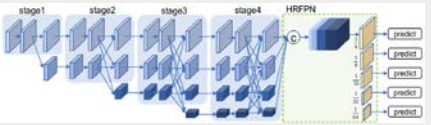
School of Computer Science and Engineering, Nanjing University of Science and Technology, China

Introduction

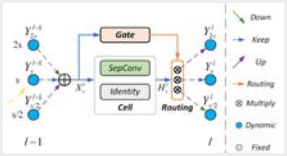
In pedestrian detection, scale variation is one of the key challenges, resulting in unbalanced performance across different scales. Recently, the High-Resolution Network (HRNet) has become popular because high-resolution feature representations are more friendly to small objects. However, when we apply HRNet to pedestrian detection, we observe that it improves for small pedestrians on one hand, but hurts the performance for larger ones on the other hand. Inspired by dynamic routing, we propose a learnable Dynamic HRNet (DHRNet) aiming to generate different network paths adaptive to different scales, and thus improving the performance across different scales consistently.

Related Work

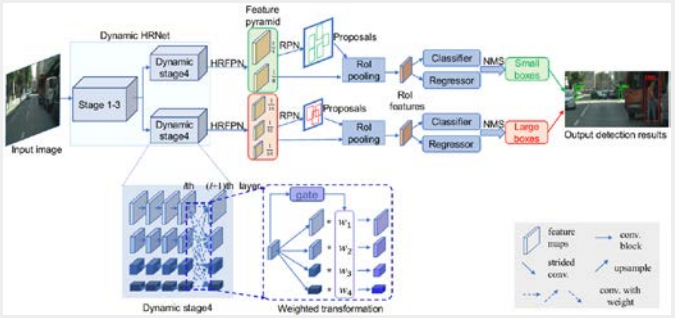
- ***High Resolution Network(HRNet)***[1]



- ***Dynamic Routing***[2]



Methods



Flow chart of our proposed detector based on Faster RCNN and our proposed Dynamic HRNet (DHRNet)

- ***Weighted Transformation Block***

Take each feature map in the *l*th layer as input and then generates four weights (w1; w2; w2; w4) via the gate function;

The generated weights are then multiplied with the rescaled input features to generate new feature maps for later feature fusion

- ***Multi-branch Stage***

To generate scale-specific feature maps.

Experiments

HRNet has 4 stages and conducts fusions across parallel convolutions in the last 3 stages. Therefore, the core problem is to decide the number of branches for different scales as well as which stage to add the soft gate module in.

We conduct ablation studies in three aspects.

- *Stage for the soft gate module to add in*

Stage	MR^{-2}	MR_s^{-2}	MR_m^{-2}	MR_l^{-2}
2	11.08	13.67	3.40	6.62
3	10.58	13.83	3.46	6.53
4	**10.40**	**13.43**	**2.69**	**6.21**
baseline	12.18	15.10	4.87	6.62

- *Number of branches*

Branches	MR^{-2}	MR_s^{-2}	MR_m^{-2}	MR_l^{-2}
1	11.28	13.79	3.50	6.51
2	**10.40**	**13.43**	**2.69**	**6.21**
3	11.12	13.35	3.85	6.49
baseline	12.18	15.10	4.87	6.62

- *Activation function*

Activation	MR^{-2}	MR_s^{-2}	MR_m^{-2}	MR_l^{-2}
Softmax	11.45	14.76	3.38	6.70
Max(0,Tanh)	10.43	**12.55**	3.28	6.46
Sigmoid	**10.40**	13.43	**2.69**	**6.21**

Method	Backbone	MR^{-2}	MR_s^{-2}	MR_m^{-2}	MR_l^{-2}
Adapted-FRCNN[9]	VGG-16	15.4	25.6	7.2	7.9
Adapted-FRCNN+Seg[9]	VGG-16	14.8	22.6	6.7	8.0
OR-CNN[20]	VGG-16	12.8	-	-	-
TLL+MRF[36]	ResNet-50	14.4	-	-	-
ALFNet[19]	ResNet-50	12.0	19.0	5.7	6.6
CSP(with offset)[20]	ResNet-50	**11.0**	16.0	**3.7**	**6.5**
FRCNN+FPN*	ResNet-50	12.77	19.57	5.45	7.03
FRCNN+HRFPN*	HRNet-W18	12.18	**15.10**	4.87	6.62
FRCNN+HRFPN	DHRNet-W18(ours)	**10.40**	**13.43**	**2.69**	**6.21**

Comparisons with the state of the arts on citypersons val set.

Results tested on the original image size (1024x2048 pixels) are reported. Red and blue indicate the best and second best performance. * represents our implementation.

Detection Results

Visualization of detection results on CityPersons validation set (FPPI=0.1; ground truth: green; detection results: red)

Conclusion

In this paper, we propose a Dynamic High-Resolution Network (DHRNet) to alleviate the scale variation problem in pedestrian detection. Specifically, a soft conditional gate module is added to the last stage in HRNet and a multi-branch structure is constructed to generate scale-specific feature maps. The multiple branches share the same parameters except the soft gate module. DHRNet solves the problem that the performance of large-scale pedestrians detection gets worse in HRNet and achieves the consistently better performance at all scales. As a result, the proposed DHRNet achieves competitive accuracy compared to the state-of-the-art methods.

References

[1] K. Sun, et al. “High-resolution representations for labeling pixels and regions,” arXiv preprint arXiv:1904.04514, 2019.

[2] Y. Li, et al. “Learning dynamic routing for semantic segmentation,” CVPR, 2020.

If you have any question, please contact mengyuanding@njust.edu.cn.