Adaptive Scenario Discovery for Crowd Counting

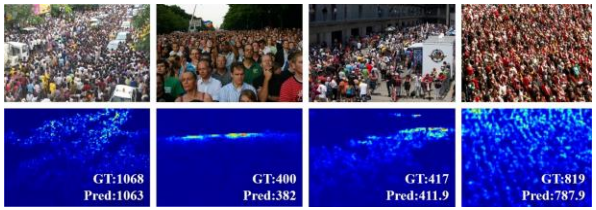
Xingjiao Wu, Yingbin Zheng, Hao Ye, Wenxin Hu, Jing Yang, Liang He

The Problem

Now the architecture of the crowd counting is difficult to represent highly variable crowd images but the images under similar scenario seem to have the same prediction pattern.

we have presented a novel architecture for high-density population counting, which focuses on the implicit discovery and dynamic modeling of scenarios.

Approach – Details

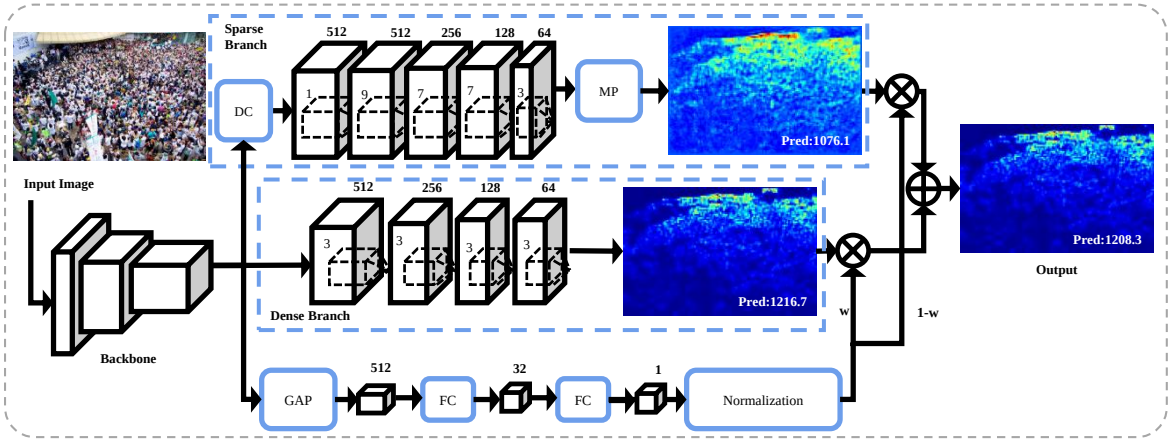


The formula of adaptively recalibrate the weight:

$$w = arctan(sigmod(w)) * \frac{2}{\pi}$$

During training, we employ the stochastic gradient descent (SGD).

Proposed Framework



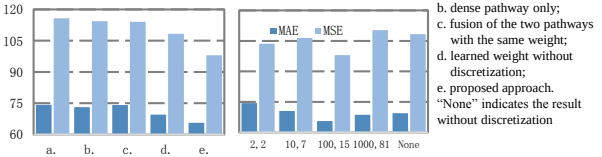
1.The backbone is the VGG-16 model pre-trained from the ImageNet dataset and fine-tuned with the crowd images;
2. The dense branch is designed to model the high congested scenario with a dense crowd, and sparse branch is for the sparse scenario;
3. The third pathways will output w that to adaptively recalibrate the weight of the dense and sparse pathways, therefore we normalize it into the interval of [0,1).

Evaluations

Comparison with state-of-the-arts

Method	Part A		Part B		UCF CC 50	
	MAE	MSE	MAE	MSE	MAE	MSE
Zhang et al.[19]	181.8	277.7	32.0	49.8	467.0	498.5
MCNN[5]	110.2	173.2	26.4	41.3	377.6	509.1
Cascaded-MTL[20]	101.3	152.4	20.0	31.1	322.8	397.9
Switching-CNN[6]	90.4	135.0	21.6	33.4	318.1	439.2
DAN[21]	88.5	147.6	17.6	26.8	234.5	289.6
CP-CNN[7]	73.6	106.4	20.1	30.1	295.8	320.9
Huang et al.[22]	-	-	20.2	35.6	409.5	563.7
D-ConvNet[13]	73.5	112.3	18.7	26.0	288.4	404.7
ACSCP[12]	75.7	102.7	17.2	27.4	291.0	404.6
DecideNet[23]	-	-	20.8	29.4	-	-
SaCNN[11]	86.8	139.2	16.2	25.8	314.9	424.8
CSRNet[8]	68.2	115.0	10.6	16.0	266.1	397.5
ASD [ours]	**65.6**	**98.0**	**8.5**	**13.7**	**196.2**	**270.9**

Ablation study



a. sparse pathway only;
b. dense pathway only;
c. fusion of the two pathways with the same weight;
d. learned weight without discretization;
e. proposed approach.
"None" indicates the result without discretization

Left: effect of varying network architecture, right: effect of scenario discovery w.r.t number of discretization bins and grouped scenarios

Key Related Work

[1] K. Simonyan *et al.*, Very deep convolutional networks for large-scale image recognition. *arXiv:*1409.1556, 2014.
[2] J. Hu et al., Squeeze-and-excitation networks. CVPR, 2018.
[3] D. B. Sam et al., Switching convolutional neural network for crowd counting. CVPR, 2017.
[4] Y. Li et al., CSRNet: Dilated convolutional neural networks for understanding the highly congested scenes. CVPR, 2018.

Conclusions

1. We presented a novel architecture for high-density population counting;
2. T*he implicit discovery and dynamic modeling of scenarios* are effective;
3. Our proposed framework achieves state-of-the-art performance.