ICPR2020

Spatial-related and Scale-aware Network for Crowd Counting

Lei Li[1] , Yuan Dong[1],Hongliang Bai[2]

[1]Beijing University of Posts and Telecommunications

[2]Beijing Faceall Co

CONTENTS

1 Introduction

The definition and difficulties of Crowd Counting

Introduction

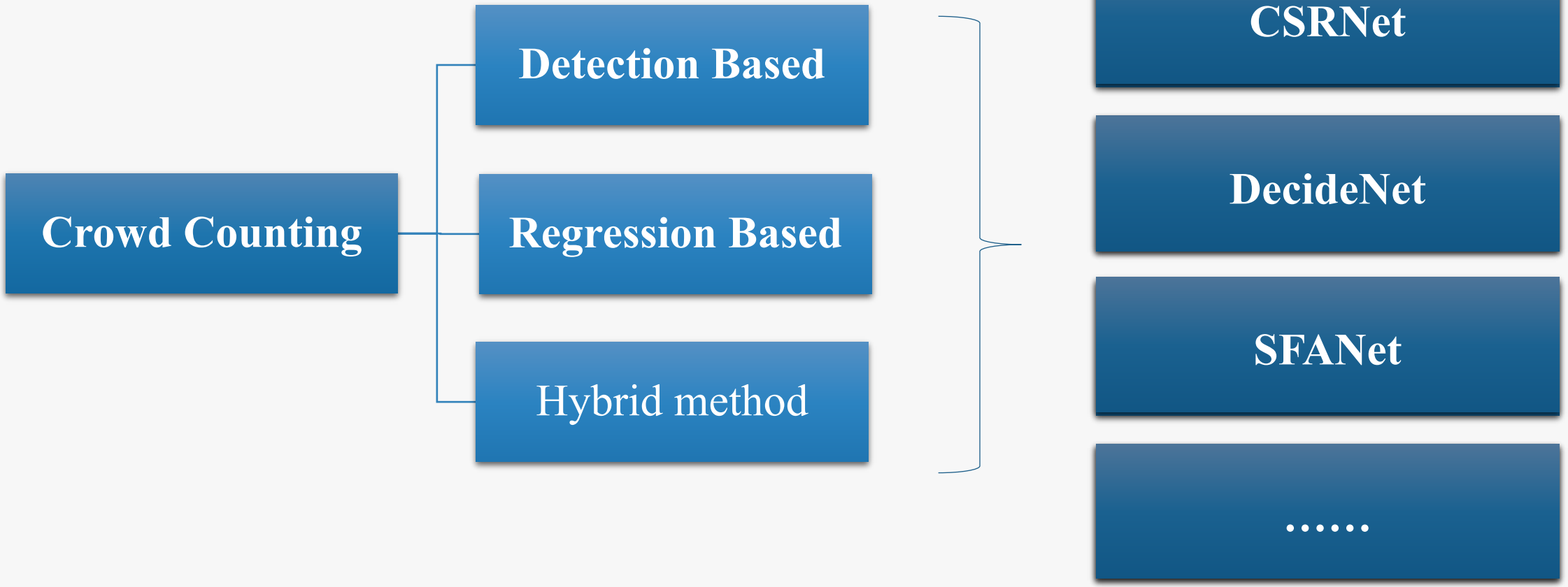
01 Definition

Crowd counting is a task to count the number of people in pictures including sparse and cluttering scenes.

02 Difficulties

Challenge of crowd counting task lies in the variations of scale, different perspective, cluttering, background noise and occlusions.

Introduction



2 Methodology

Motivation and proposed methods

Motivation

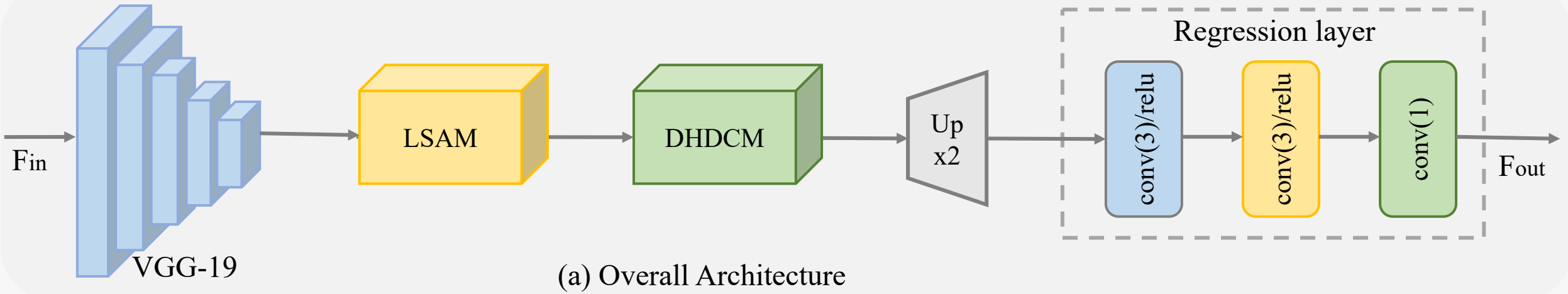
Scale variation

People have different scales since the different distance

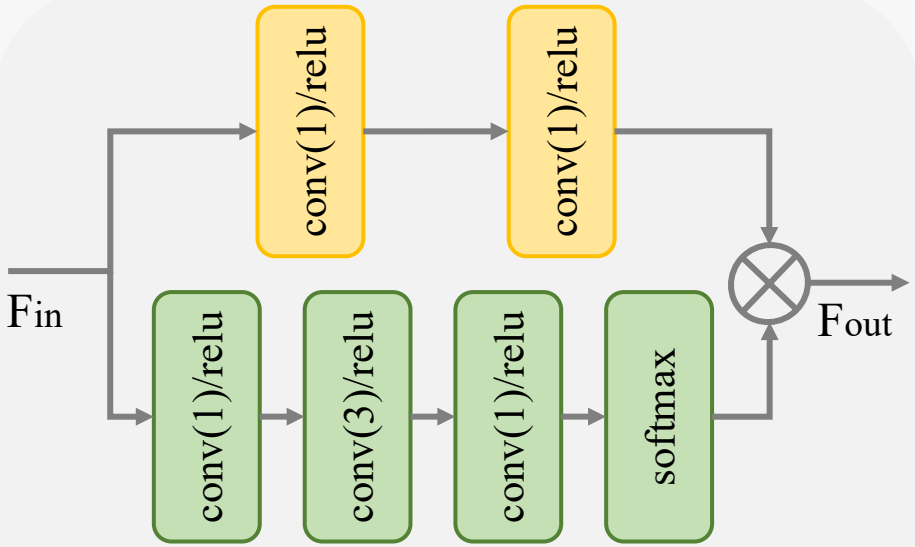
Spatial relations

Backgrounds and the human crowds should have significantly different responses

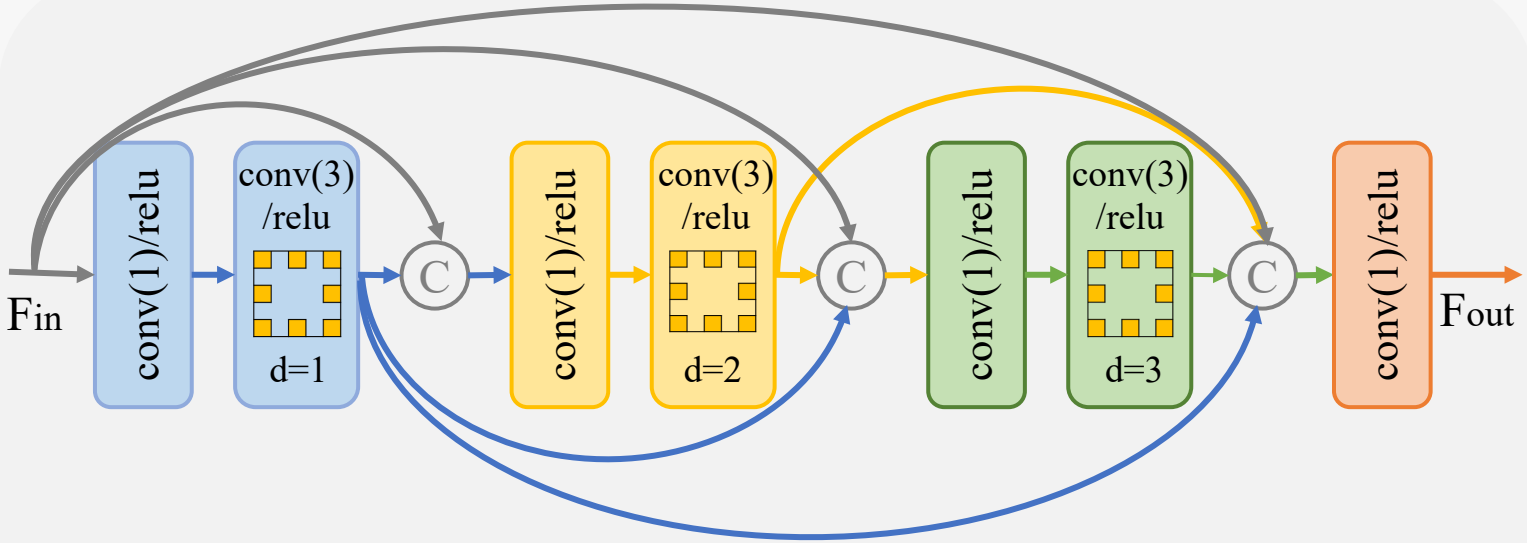
Overview of the proposed framework



(a) Overall Architecture

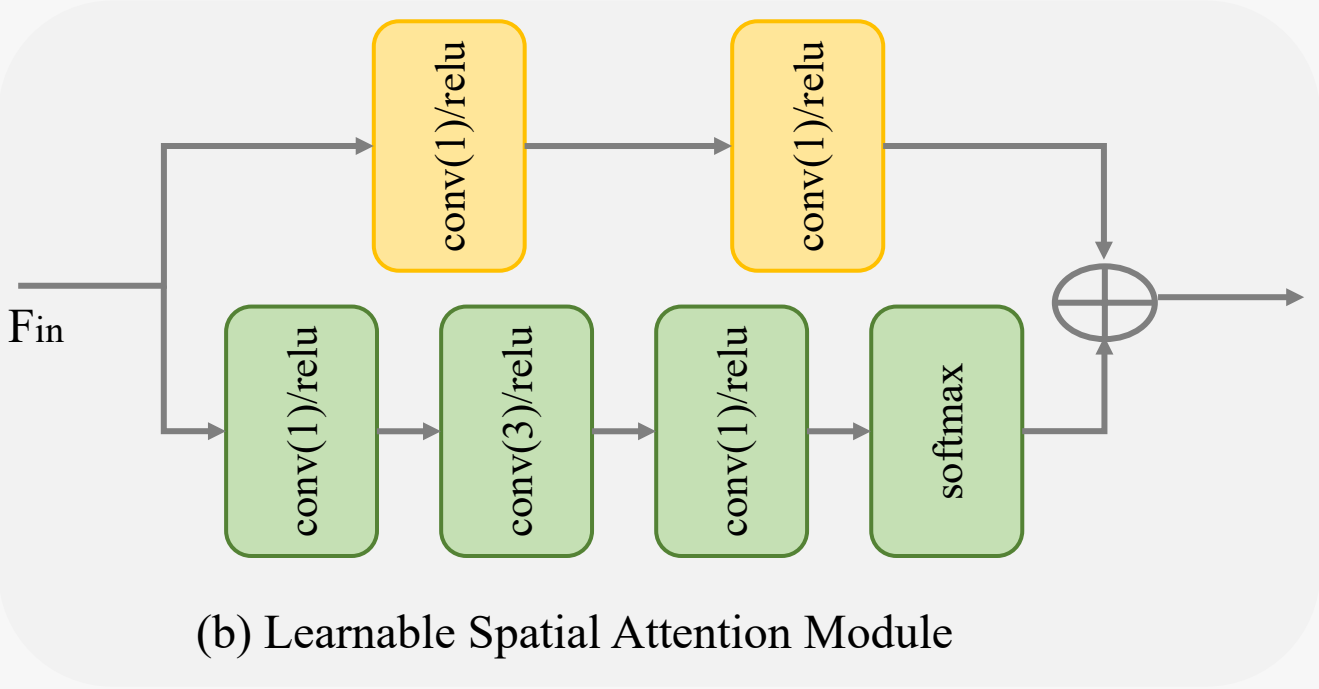


(b) Learnable Spatial Attention Module



(c) Dense Hybrid Dilated Convolution Module

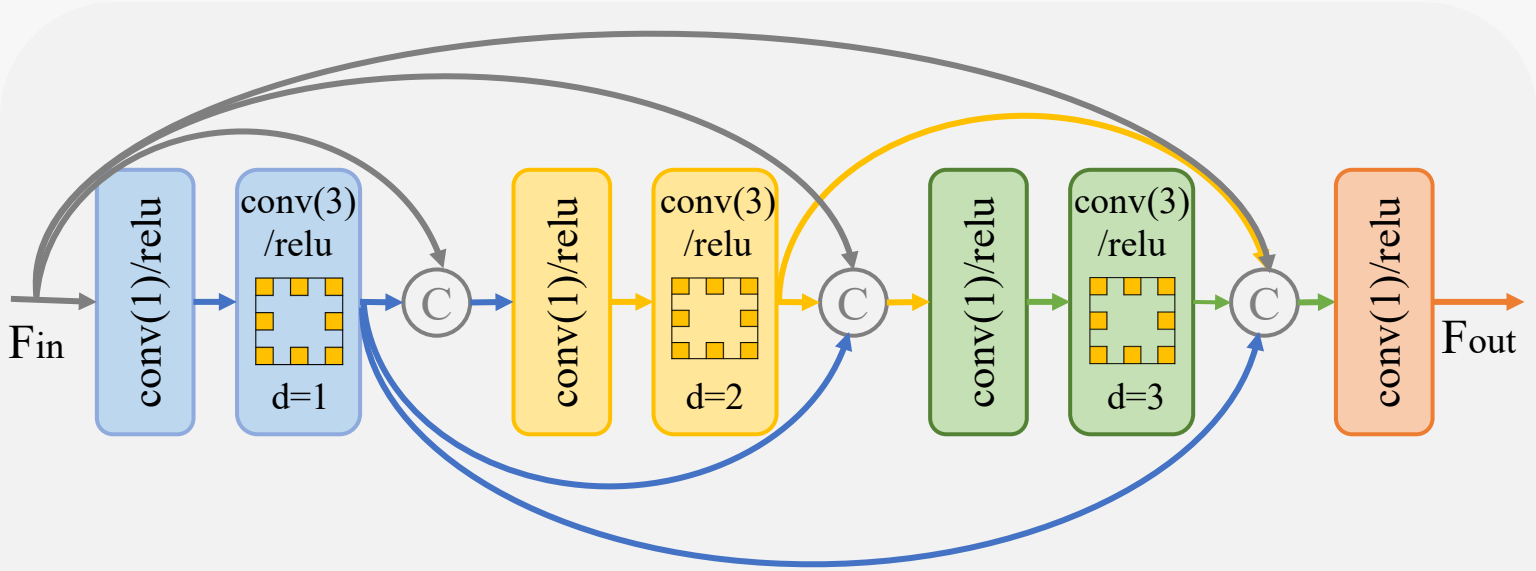
LSAM



(b) Learnable Spatial Attention Module

- To get the spatial relations in the feature maps, we use a learnable convolution operation to capture the spatial attentions. The lower branch consists of three convolutional layers and a softmax layer. The convolutional layers with a kernel size of 1x1, 3x3 and 1x1 are designed to get the spatial attentions.

DHDCM



(c) Dense Hybrid Dilated Convolution Module

- To solve the problem of scale variation, we use dilated convolutional layers to get the most representative features of people across different scales. Inspired by ASPP, we come up with a module to densely connect a set of dilated convolutional layers, so that the generated multi-scale features can cover a denser scale range without significantly increasing the parameters.

3

Results & conclusions

Ablation study and final conclusions

Spatial attentions by LSAM

Scale variance by DHDCM

Evaluation Criteria

1

MAE

$$MAE = \frac{1}{K}\sum_{k=1}^{K}|N_k - C_k|$$

2

MSE

$$MSE = \sqrt{\frac{1}{K}\sum_{k=1}^{K}|N_k - C_k|^2}$$

K is the number of test images, N_k and C_k are the ground-truth count and the estimated count for the k-th image respectively.

Effect of LSAM and DHDCM

ABLATION STUDY OF LSAM AND DHDCM ON UCF-QNRF BENCHMARK.

LSAM	DHDCM	MAE	MSE
-	-	98.6	172.8
✓	-	96.2	168.5
-	✓	95.4	164.3
✓	✓	**93.2**	**158.2**

- Taking MAE as an example, LSAM and DHDCM can respectively bring improvements of 2.4 points and 3.2 points compared with the baseline method. Relatively speaking, DHDCM brings greater promotion. Combining these two modules, our improvement is more obvious since we can simultaneously take the background and scale influence into consideration.

Comparison with SOTA methods

COMPARISON WITH THE STATE OF THE ART METHODS ON UCF-QNRF BENCHMARK.

Method	MAE	MSE
MCNN[1]	277	426
SwitchCNN[6]	228	445
CL[2]	132	191
S-DCNet[28]	104.4	176.1
SFCN[29]	102.0	171.4
Ours	**93.2**	**158.2**

COMPARISON WITH THE STATE OF THE ART METHODS ON UCF_CC_50 BENCHMARK.

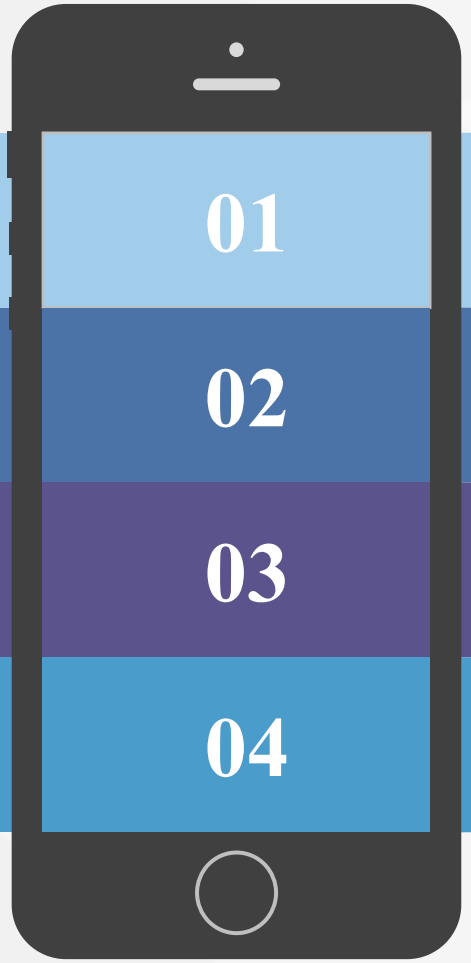
Method	MAE	MSE
MCNN[1]	377.6	509.1
SwitchCNN[6]	318.1	439.2
CSRNet[9]	266.1	397.5
SANet[7]	258.4	334.9
ADCrowdNet[12]	257.1	363.5
BL+[26]	229.3	308.2
Ours	**220.5**	**302.9**

COMPARISON WITH THE STATE OF THE ART METHODS ON SHANGHAITECH BENCHMARK.

Method	PartA		PartB	
	MAE	MSE	MAE	MSE
MCNN[1]	110.2	173.2	26.4	41.3
SwitchCNN[6]	90.4	135	21.6	33.4
IC-CNN[27]	68.5	116.2	10.7	16.0
CSRNet[9]	68.2	115.0	10.6	16.0
BL[26]	64.5	104.0	7.9	13.3
BL+[26]	62.8	101.8	7.7	12.7
ADCrowdNet[12]	63.2	98.9	7.7	12.9
Ours	**60.95**	**97.55**	**7.5**	**12.6**

- All the results in the three datasets prove that the proposed LSAM and DHDCM are effective to optimize the performance, not only in dense scenes but also in sparse scenes.

Conclusions



- Learnable Spatial Attention Module can get spatial attentions
- Dense Hybrid Dilated Convolutional Module can solve the scale variation problem
- Both modules can be transferred to other related task
- The state-of-the-art results on all three datasets.

Lei Li[1] , Yuan Dong[1], Hongliang Bai[2]

[1]Beijing University of Posts and Telecommunications
[2]Beijing Faceall Co