Jie-hao Huang, Xiao-guang Di, Jun-de Wu, Ai-yue Chen, 2020. A novel convolutional neural network method for crowd counting. *Frontiers of Information Technology & Electronic Engineering*, 21(8):1150-1160. https://doi.org/10.1631/FITEE.1900282

A novel convolutional neural network method for crowd counting

Key words: Crowd counting; Density estimation; Segmentation prior map; Uniform function

Corresponding author: Xiao-guang DI
E-mail: dixiaoguang@hit.edu.cn
ORCID: https://orcid.org/0000-0002-5709-6862

Motivation

- Existing methods always use a multi-column convolutional neural network. They always present a complementary effect in predicting a density map. The complementary effect brings a lot of noise to regions without a person, which reduces the quality of the density map and gives unreliable crowd counting.
- The small value in the Gaussian density map is nearly 0. It is possible to make the output of the activation function less than 0 by backpropagation. Therefore, it is hard for the network to converge to the optimal result.

Main idea

- We design a foreground-segmentation network for coarsely segmenting the heads, which can effectively wipe out the mistakenly detected noise in the background, and a crowd-regression network for differentiating head sizes and generating a high-quality density map.
- A novel uniform function is proposed to generate a head-segmentation map and a uniformly distributed density map with cheap single dot annotations.

Method

1. Segmentation prior map and uniform function

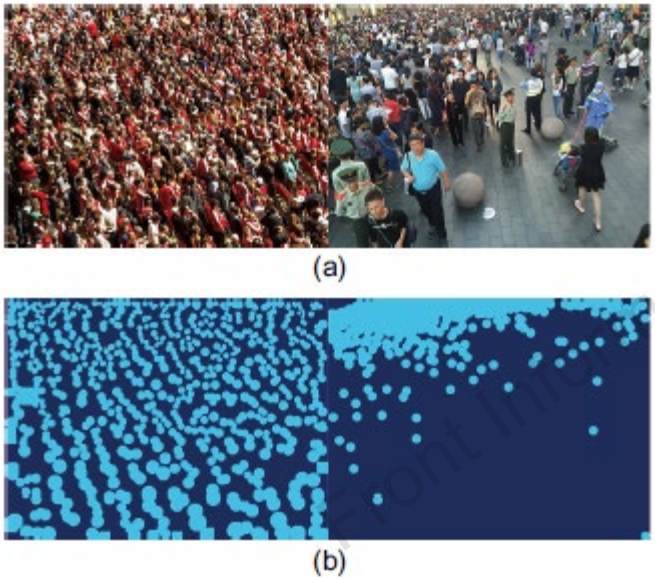


Fig. 1 Raw images (a) and location-level head-segmentation ground truth map (b)

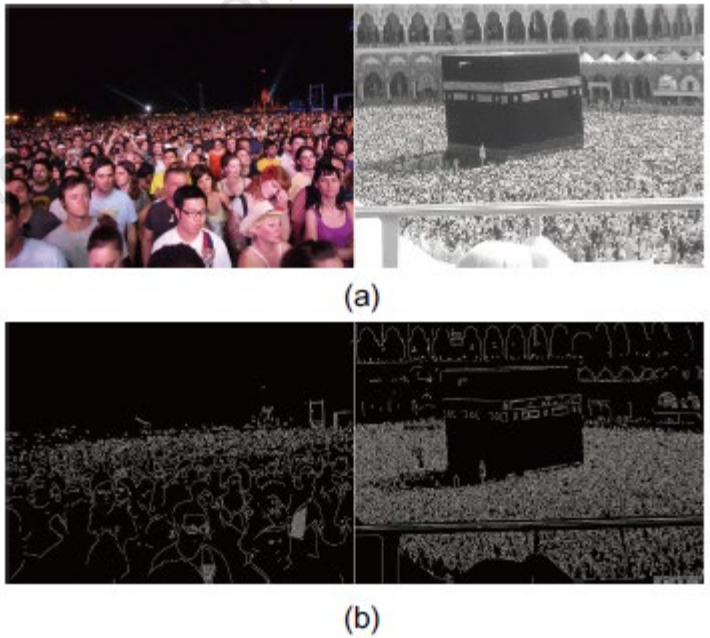


Fig. 2 Original images (a) and the edge prior maps processed by the Canny-edge detector (b)

2. Experimental setup

Table 1 Structure of SAPNet: FS-CNN and CR-CNN

Structure		
First part*	VGG-bone	(conv3-64)*2+max-pooling (conv3-128)*2+max-pooling (conv3-256)*3+max-pooling (conv3-512)*3
	Prior-bone	Conv3-64+max-pooling Conv3-128+max-pooling Conv3-256+max-pooling
Second part	FS-CNN	(dilate-conv3-64)*2 Dilate-conv1-1
	CR-CNN	(dilate-conv3-64)*4 Dilate-conv1-1

* The first part is the same for FS-CNN and CR-CNN

Major results

Table 6 Estimation errors on ShanghaiTech dataset

Method	MAE		RMSE	
	Part *A*	Part *B*	Part *A*	Part *B*
MCNN	110.2	26.4	173.2	41.3
Cascaded MTL	101.3	20.0	152.4	31.1
Switching-CNN	90.4	21.6	135.0	33.4
CP-CNN	73.6	20.1	106.4	30.1
IG-CNN	72.5	13.6	118.2	21.1
ACSCP	75.7	17.7	102.7	27.4
CSRNet	68.2	10.6	115.0	16.0
SAPNet	77.5	9.4	128.8	15.4

ACSCP: crowd counting via adversarial cross-scale consistency pursuit (Shen et al., 2018)

The proposed method outperforms all the previous methods on ShanghaiTech part *B* and gives a competitive result on ShanghaiTech part *A*.

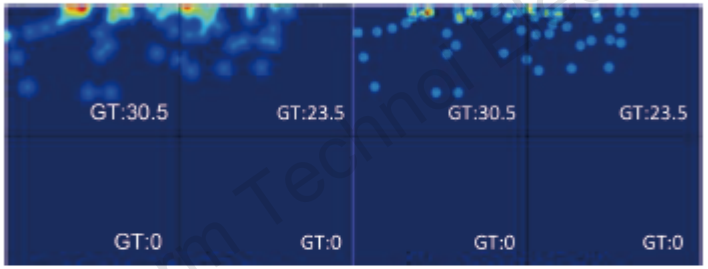
Table 8 Estimation errors on the UCF-CC-50 dataset

Method	MAE	RMSE
Zhang C et al. (2015)'s	467.0	498.5
MCNN	377.6	509.1
Switching-CNN	318.1	439.2
CP-CNN	295.8	320.9
IG-CNN	291.4	349.4
ACSCP	291.0	404.6
CSRNet	266.1	397.5
SAPNet	255.0	327.1

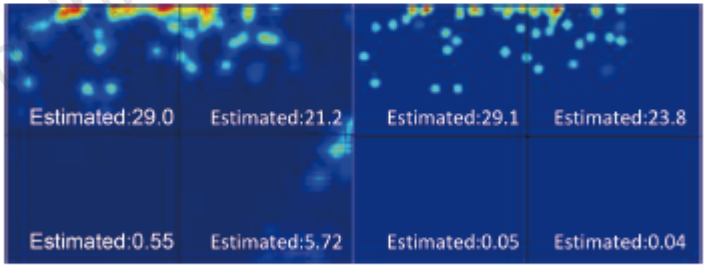
The number of people in the UCF-CC-50 dataset varies greatly, ranging from 94 to 4543. The proposed method outperforms all the previous methods.

Major results

(a)



(b)



(c)

Fig. 3 Original image and head-segmentation map predicted by SAPNet (a), ground truth of two methods CSRNet and SAPNet (b), and two density maps predicted by the two methods (c)

Conclusions

- A foreground-segmentation network is designed for coarsely segmenting the heads, which can effectively wipe out the mistakenly detected noise in the background.
- Two networks take the prior maps as the input, i.e., a Canny-edge prior map and a coarse head-segmentation prior map , which helps quickly recognize effective features and reduce the training complexity.
- A novel uniform function is proposed to generate a head-segmentation map and a uniformly distributed density map.
- We demonstrate our method on four benchmarks and achieve state-of-the-art performances on the ShanghaiTech part *B* and UCF-CC-50 datasets.