International Journal of Trend in Scientific Research and Development (IJTSRD)
Volume 4 Issue 6, September-October 2020 Available Online: www.ijtsrd.com e-ISSN: 2456 – 6470

Research on People Counting Method Based on the Density Map

Qian Li, Peng Wei, Minjuan Shang

Beijing Wuzi University, Beijing, China

ABSTRACT

With the rapid development of science and technology and the rise of deep learning, in recent years, image-based people counting has become a hot research topic in the field of computer vision. The method based on the density map retains the spatial information of the crowd in the picture, and on this basis, the network model of deep learning training is also used to improve the accuracy of the people counting. This paper mainly analyzes the traditional counting methods, introduces the common datasets used in people counting. We mainly conducted research and analysis on recent years' people counting methods of generating high-quality population density map through CNN method, the end-to-end convolutional neural network model surprised us with the accuracy of the people counting. Finally, this paper prospects the future research trend of people counting.

KEYWORDS: *people counting, computer vision, density map, convolutional neural network (CNN)*

How to cite this paper: Qian Li | Peng Wei | Minjuan Shang "Research on People Counting Method Based on the Density Map" Published in International Journal of Trend in Scientific Research and Development (ijtsrd), ISSN: 2456-6470, Volume-4 | Issue-6, October 2020, pp.1516-1523, URL: www.ijtsrd.com/papers/ijtsrd33599.pdf

IJTSRD33599

1. INTRODUCTION

China has a large population, so it is practical and challenging to carry out people counting in public places. Considering China's 2015 Shanghai Stampede at the Bund and the novel Coronavirus (COVID-19) outbreak at the end of 2019, it is of great significance to reduce casualties, prevent and control cluster infections and prevent the spread of the epidemic by timely and non-contact monitoring of crowd gathering level and corresponding security and prevention measures. Therefore, the use of advanced technology for people counting has a huge role in the above events, however, the people counting also faces many challenges, such as the uneven distribution of population, severe occlusion of population, varying illumination, scene variability, head scale transformation, image sharpness, and so on.

The advantage of the image-based people counting technology is that it is not easy to find out and does not need to be touched. Compared with traditional technology methods, the unobserved feature makes the detection method not easy to be resisted. Such as manual counting, people have not good feelings in the process of artificial collection, and even consumes a lot of manpower and material resources when the number of people is huge. The characteristic of no contact is undoubtedly improved for the health and safety of people in a special period. The required statistical information does not require the cooperation of the statistical objects or intervention of other staff members, making it a natural, intuitive, rapid, and safe way to count the number of people. Its advantages are mainly concentrated in the following aspects: no intrusion, perfect hardware foundation, quick and convenient acquisition, and good expansibility. Due to the features of the contactless and quick response in the detection process, such detection and counting technology using surveillance videos have a good guarantee in terms of security, privacy, and accuracy [1].

2. Related Technologies

2.1. Research status of People Counting Algorithm

2.1.1. Feature-based Detection Approach for People Counting

The initial people counting mainly focused on the number of detection based on the sliding window. The people counting algorithms of feature-based detection can be divided into two categories, the detection based on the whole body and the detection based on the local body parts. The main steps are as follows, First traverse images by sliding window, in the whole-body detection based on the human body, extracts related features such as the skin color, Hog, edge, and other features related to the human body from the window, and trains a classifier. However, as the image is a two-dimensional plane, the overall detection is not conducive to the processing of occlusion in image detection, and the accuracy is low. Therefore, to reduce the "occlusion" problem and improve the accuracy, a local body detection is proposed based on the human body, which for the human body is the most notable features of the head. At the beginning of head detection, Hough transforms and circle-like detection are used to detect the contour of the head, to further improve the accuracy of the detection, puts forward the combination of head and shoulders, similar in shape "Ω". Shu Wang, et al.[1]designed a new edge feature to extract and enhance the contour of

the head and shoulders. The basic idea is to predict the head and shoulder contour by filtering the edge pattern of the edge image, which is called En-Contour. This new feature significantly improves the HOG+LBP algorithm.

Besides, feature learning is mostly based on machine learning algorithms, such as Boosting, AdaBoost[2], SVM support vector machine, and so on. Sliding window detection mostly requires strong individual characteristics, that is, obvious box labeling, so this detection algorithm is more suitable for individual detection. However, when the number of people increases, the labeling overlaps greatly and the error caused by occlusion will increase, thus reducing the accuracy of people counting.

2.1.2. Pixel-based regression Approach for People Counting

Compared with the algorithm based on detection, the people counting algorithms based on pixel regression have better performance and higher accuracy in dealing with the "occlusion" problem. The main steps are foreground segmentation, feature extraction, regression model learning, and people counting mapping. Feature extraction can include texture features, gradient feature, edge count, etc. Commonly used regression methods include Bayesian linear regression, piece-wise linear regression, ridge regression, and Gaussian process regression. Its basic idea is to learn the mapping relationship between feature extraction and image calibration number and to learn regression function to estimate crowd density and number statistics [3][4].

Chen et al. [5] proposed a regression learning method based on back-propagated information, that is, the model automatically generates a back-propagated information framework. This regression counting framework can also be described as a weighted regression method, which constructs a high-level "supervision" to assign weights to each training sample according to the error of back-propagated. The back-propagated error is used as the weight indicating the importance of each sample, and then the weighted regression model is learned. Such back-propagation information can be used for both low-level features and intermediate semantic attributes to enhance the sample mining using weights, to improve the regression performance of low-level features and intermediate attributes. But the back-propagated information is generated and used only during training, while the learning weight is directly used for testing.

2.1.3. Density map-based Approach for People Counting

The people counting method based on the density map is one of the methods based on density estimation, and the other method is to directly output the input image to estimate the head number. Relatively speaking, the method based on density map estimation is the people counting method which is widely used at present. The density map contains more information and retains the spatial information of population distribution, which is of great help for crowed density estimation, emergency analysis, and abnormal situation detection. The working procedure is usually to mark the head in the image and generate the corresponding crowd density map according to the data of the head position. The Gaussian kernel density function is a common density map generation function. Then, the characteristic mapping relationship between the original image and the density map is learned, and the people counting is finally achieved by integrating the density map.

V. - Q. Pham et al. [6]proposed a patch-based people counting method in public scenes. This method firstly divides the image into many pieces (called patches in the paper),using the random forest framework to learn the nonlinear mapping between "characteristics of patches" and "relative positions of all objects in patches", in which the patch characteristics used are relatively simple and independent of the scene. Then through the Gaussian kernel density estimation to generate a patch density figure, at last count.

2.2. Introduction to commonly used datasets

There are six public datasets in the people counting in Table I.UCSD dataset [7] is the first public dataset in the people counting. It is obtained from the surveillance video installed above the sidewalk and consists of a total of 2000 frames of the video sequence. The number of people in the image of UCSD dataset is small, about 24 people in each frame on average. The collection scene of Mall dataset [8] is a shopping mall where the crowd changes a lot. There are 2000 frames of images in total, and the number of people in each frame is relatively small. Beijing_BRT is a dataset collected at the Beijing BUS rapid transit platform [9]. The time collected at the platform varies greatly, so the illumination variation of the images in the dataset is relatively rich, which is suitable for passenger flow statistics in the traffic field. And what makes it different from other datasets is that its image size is narrow. The first three datasets are collected under only one single scene, while WorldExpo'10[10] is a dataset from 2010 Shanghai World Expo that has multiple perspectives and scenes. The number of people varies from 1 to 250. UCF_CC_50 is also a cross-scene population statistical dataset [3]. The author collected and produced a dataset containing only 50 pictures from multiple scenes such as concerts, sports festivals, and marathons. Although the total number is small, the 50 pictures contain a large number of people, with an average of about 1280 people per image, which is a very challenging dataset. Shanghai Tech is a comparative dataset [11], the author divided it into two parts. Part_A is randomly selected from the network, and the crowd is relatively dense. The number of people in one image can reach up to 3139. The image resolution is all not fixed in Part_Aand UCF_CC_50.And Part_B collected data from busy streets in Shanghai Downtown, and the crowd was relatively sparse, with at least 9 people in one image. Except for UCSD dataset, all the above datasets are marked with the point of the human head by MATLAB, and the coordinate position information *(x,y)* is stored in the *mat* file.

TABLE I Summary of common public datasets for people counting

Dataset		Resolution	Num_i	Max	Min	Ave	Total
UCSD		158×158	2,000	46	11	24.9	49,885
Mall		320×240	2,000	53	13	—	62,325
Beijing_BRT		640×360	1,280	64	1	13	16,795
WorldExpo'10		576×720	3,980	253	1	50.2	199,923
UCF_CC_50		Varied	50	4,543	94	1,279.5	63,974
Shanghai Tech	Part_A	Varied	482	3,139	33	501.4	241,677
	Part_B	768×1024	716	578	9	123.6	88,488

Resolution means the resolution of the images in this dataset. NUM Filed denotes the sum of train images and test images. MAX, MIN, and AVE show the maximum, minimum, and average number of people in an image respectively. TOTAL field denotes the sum of the heads in all images.

3. Deep learning-driven people counting based on the density map

Although the local detection in the method based on feature detection improves the accuracy problem caused by "occlusion", its effect is not great. Therefore, the pixel-based regression approach for people counting is proposed, which can reduce the accuracy decline caused by "occlusion". However, the people counting approach based on direct regression will lose the spatial information in the image, while research shows that [4]Spatio-temporal information is helpful to improve the accuracy of people counting. The method based on density map estimation can make up for this deficiency precisely because the density map retains the spatial information. Given an image, the current mainstream method is to estimate the number of people using CNN based on a density map. The following is an analysis of people counting for CNN network model based on density map estimation in recent years.

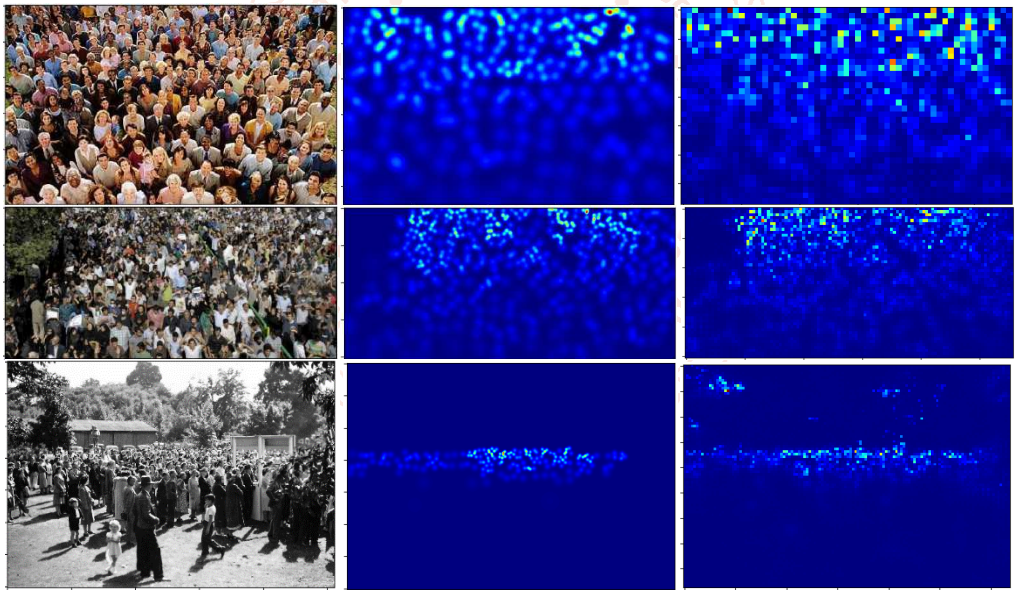

Figure 1 Density map-based people counting

3.1. Analysis of people counting network

Compared with the traditional people counting method, the people counting algorithm based on deep learning can achieve higher counting accuracy, and its mainstream method still focuses on extracting multi-scale characteristic information of human head, among which the representative one is the network structure based on CNN. In 2015, Wang et al. [12] first proposed to apply Alexnet[13] network architecture to people counting and applied a convolutional neural network to people counting for the first time. Since then, scholars have gradually proposed various network structures in this field, such as Switch-CNN[14], SaCNN[15], CP-CNN[16], DR-Resnet [17], CrowdNet[18], etc. Although the convolutional neural network shows higher accuracy in people counting, it also brings a large amount of calculation, such as the popular Mask r-CNN [19] and faster R-CNN[20].

3.1.1. Analysis of people counting algorithm based on MCNN

Multi-column convolution neural network model (MCNN)[11] configured three columns of neural network with different size of convolution kernels. The three parallel sub-networks have the same depth, each column separately extracted the features of human heads with different scales in the image. After features are fused, L, M and S columns of convolutional neural networks are combined, and the convolution kernel with size 1×1 is linearly weighted to generate the final crowed density map. Because the structure of full connection layer is not adopted in the network, a relatively lightweight full

convolutional network with few network parameters is implemented. The loss function chooses Euclidean distance to measure the difference between the ground truth density map and the estimated density map. The loss function is defined as follows:

$$L(\Theta) = \frac{1}{2N}\sum_{i=1}^{N}\left\|F(X_i;\Theta) - F_i^{GT}\right\|_2^2 \qquad (3.1)$$

Where N is the number of training images, which is the batch size of training images. X_i is the input image.$F(X_i;\Theta)$ Represents the estimated density map generated by MCNN. F_i^{GT} is the ground truth density map that result of input imageX_i.$L(\Theta)$ is the loss between estimated density map and the ground truth density map.

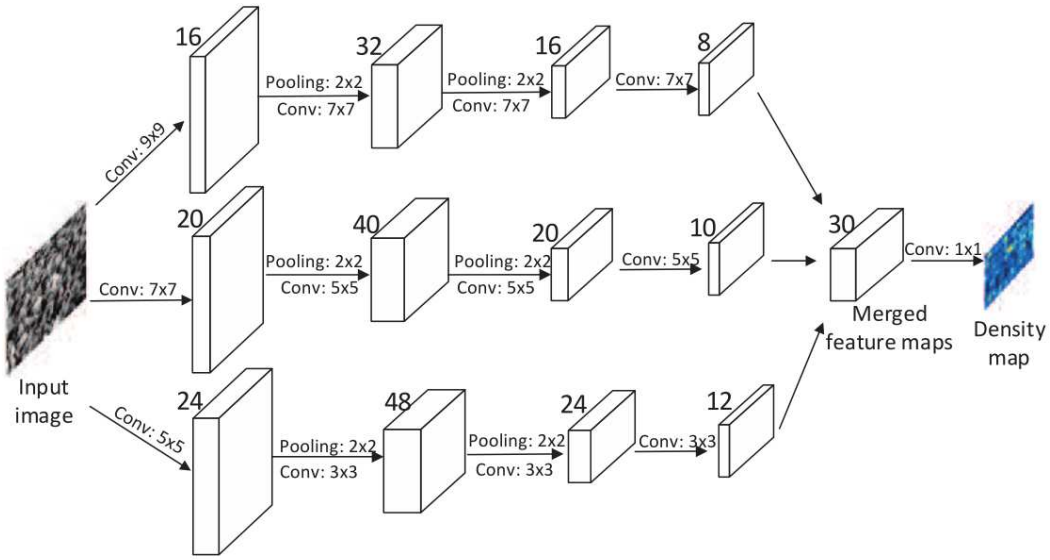

Figure 2 Configuration of multi-column convolutional neural network

3.1.2. Analysis of people counting algorithm based on CSRNet

The CSRNet model [21] is an end-to-end network structure, which is divided into two parts. The front-end of CSRNet is VGG-16 with the full connection layer removed to extract image features, while the back-end network is highlighted, which uses a dilated convolution layer instead of the pooling layer and expands the receptive fields of convolution kernel to generate a high-quality density map. VGG-16 was the runner-up network in ILSVRC2014. Its architecture is simple and flexible. It is a basic CNN network consisting of a convolutional layer and a pooling layer. In VGG-16 of CSRnet, a combination of 10 convolutional layers and 3 pooling layers is adopted, and the image size is no longer limited after the removal of the full connection layer. The back-end network of the CSRNet structure uses a special convolutional layer: dilated convolution. A 2-D dilated convolution can be defined by the following formula:

$$y(m,n) = \sum_{i=1}^{M}\sum_{j=1}^{N} x(m + r\times i, n + r\times j)w(i,j) \qquad (3.2)$$

Where $y(m,n)$ is the output of the dilated convolution with the input $x(m,n).w(i,j)$ is the convolution kernel. Andparameter$r$ is the dilationrate. When $r = 1$, dilated convolution is the normal convolution. $x(m,n)$ is the input image information with length and width of M and N respectively. Through the convolution kernel $w(i,j)$, the output $y(m,n)$ of the dilated convolution is obtained. Experiments show that dilated convolution utilizes sparse convolution kernel to realize alternating convolution and pooling operation, and enlarges the receptive field without increasing network parameters and calculation scale, which is more suitable for crowd density estimation. However, normal convolution operation needs to increase the number of convolutional layers to obtain a larger receptive field, and more data operations are also added.

Configurations of CSRNet			
A	B	C	D
input(unfixed-resolution color image)			
front-end (fine-tuned from VGG-16)			
conv3-64-1 conv3-64-1			
max-pooling			
conv3-128-1 conv3-128-1			
max-pooling			
conv3-256-1 conv3-256-1 conv3-256-1			
max-pooling			
conv3-512-1 conv3-512-1 conv3-512-1			
back-end (four different configurations)			
conv3-512-1 conv3-512-1 conv3-512-1 conv3-256-1 conv3-128-1 conv3-64-1	conv3-512-2 conv3-512-2 conv3-512-2 conv3-256-2 conv3-128-2 conv3-64-2	conv3-512-2 conv3-512-2 conv3-512-2 conv3-256-4 conv3-128-4 conv3-64-4	conv3-512-4 conv3-512-4 conv3-512-4 conv3-256-4 conv3-128-4 conv3-64-4
conv1-1-1			

Figure 3 Configuration of CSRNet

3.1.3. Analysis of people counting algorithm based on Context-Aware Network

The content-Aware network model [22] is similar to CSRnet in that both are end-to-end network structures. Different from CSRnet, it inserts the content-aware structure between the front end and the back end. The structure hierarchy is deeper, which is divided into three parts: the front-end, the Context-Aware block, and the back-end. The starting point is the top 10 layer in VGG-16, the limitations of VGG-16 is that the entire image using the same receptive fields. But this paper combines the use of multiple different size gain characteristics of the receptive fields, and learn the importance of each of these features at each image location. To accurately predict population density, multi-level contextual information is adaptively encoded into the features it generates. Finally, an encoder composed of several dilated convolutional layers in the back-end generatesa density map of the extracted context feature f_I, which generates a good density estimation in the high-density area of the crowd.

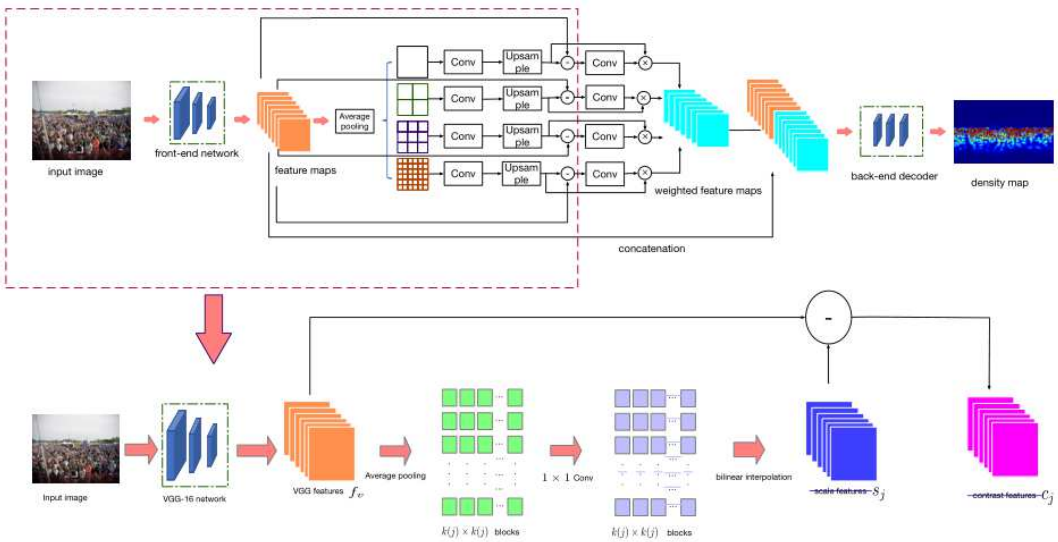

Figure 4 Configuration of Context-Aware Network

3.2. Evaluation and comparison

3.2.1. Evaluation metrics

Mean Average Error (MAE) and Mean Squared Error (MSE) is used to evaluate different methods in most evaluation indexes of people counting. MAE can reflect the error of predicted values, indicate the accuracy of the estimates. MSE is used to describe the robustness of a model. And the smaller MSE is, the higher the accuracy is. MAE and MSE are defined as follows:

$$MAE = \frac{1}{N}\sum_{i=1}^{N}\left|Z_i - Z_i^{GT}\right| \quad (3.7)$$

$$MSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left|Z_i - Z_i^{GT}\right|^2} \quad (3.8)$$

$$C_i = \sum_{l=1}^{L}\sum_{w=1}^{W} z_{l,w} \quad (3.9)$$

where $N$is the number of test images, C_i represents the number of predictors of image X_i, C_i^{GT} represents the people number of ground truth density map, and $z_{l,w}$is the pixel value at (l, w) in the density map with length L and width W.

3.2.2. Experimental analysis of ShanghaiTech dataset

Table II Estimation errors on ShanghaiTech dataset

Method	Part_A		Part_B	
	MAE	MSE	MAE	MSE
MCNN	110.2	173.2	26.4	41.3
CSR-Net	68.2	115	10.6	16
Context-Aware Network	62.3	100.0	7.8	12.2

The above three network models all represent the people counting as a regression crowd density map from the image. The MCNN structure is relatively simple and does not go as deep as other network structures. However, the author's experiment shows that the network with pre-training performs better than the network without pre-training. So its training process requires multi-stage pre-training. We can see from the above table II, the accuracy of CSRNET model and the Context-Aware Network model is better than that of MCNN model. But the network structures of CSRNet and Context-Aware are more complex and in-depth, and both of them need to store a large number of parameters. These methods need a lot of storage and computing resources, which limit their application, especially in embedded devices of limited memory.

3.2.3. Comparative analysis of platform migration performance

Table III. Hardware platform performance comparison on ShanghaiTech dataset

Method	模型内存	显存占用	迭代速度
MCNN	< 1 MB	< 4 GB	0.6 minutes/次
CSR-Net	> 100 MB	> 4 GB	10 minutes/次
Context-Aware Network	50 MB < model <100 MB	> 4 GB	4 minutes/次

We reproduce the three methods. The MCNN model takes up the least memory, and it's running-VRAM and iteration-speed are also the least. Therefore, its hardware platform has the best migration. However, although the CSRNet model and the Context-Aware network model are more accurate, their model storage takes up more memory, requires the higher performance of computer hardware. So the platform migration is relativelypoor. In fixed scenes with certain shooting angles, such as classrooms, small conference rooms, cinemas, etc., the population base is not large, so MCNN with low requirements on hardware platforms can also achieve high accuracy in people counting. But in the context of changing shooting angles and complex scenes, the Context-Aware Network method can achieve the optimal counting effect to ensure the accuracy of people counting.

4. Conclusion

From the current research status, computer vision research gradually tends to be "applied research", and the people counting have strong scene pertinence. The people counting based on image integrates the technology of human detection, foreground segmentation, feature extraction, target detection, tracking, and other fields, and adopts different methods for different environments. The counting method can also be transferred to the counting of cells and bacteria in the medical field [23], the counting of vehicles in the transportation field [24][25], the automatic counting of wheat ears in the agricultural field [26], particle nuclear track counting[27], etc.

In general, people counting gradually tends to be based on density map estimation, which retains more spatial distribution information. With the development and rise of deep learning, crowd density estimation driven by deep learning is of great significance for people counting,

anomaly detection and emergency prediction, and so on. Compared with the traditional method, the accuracy of the convolutional neural network for people counting is greatly improved, and the end-to-end deep learning method is more and more popular in the field of people counting due to its flexible architecture and greatly improved detection accuracy. However, it increases the network complexity, takes up long training time, takes up a large amount of model storage, has higher requirements on hardware conditions, and reduces the platform migration performance. In future research, we can pay more attention to the optimization of network structure and the compression of the model. At the same time, head detection is still the mainstream and most commonly used model in people counting. In the future, attention should be paid to how to make the model distinguish human features more accurately under the overlap of human heads in densely populated areas, that is, under the condition of small head size, to obtain better overall accuracy.

References

[1] S. Wang, J. Zhang, and Z. Miao, "A new edge feature for head-shoulder detection," in 20th IEEE International Conference on Image Processing, Melbourne, Australia, 2013, pp. 2822–2826.

[2] Sheng Yang, Xianmei Liao, UK Borasy, "A Pedestrian Detection Method Based on the HOG-LBP Feature and Gentle AdaBoost," in International Journal of Advancements in Computing Technology(IJACT), 2012, pp. 533–560.

[3] Haroon Idrees, Imran Saleemi, Cody Seibert, and Mubarak Shah, "Multi-Source Multi-Scale Counting in Extremely Dense Crowd Images". In Computer Vision and Pattern Recognition, pages 2547–2554. IEEE, 2013.

[4] K. Chen and J.-K. Kämäräinen, "Pedestrian density analysis in public scenes with spatiotemporal tensor features," IEEE Trans. Intell. Transp. Syst., vol. 17, no. 7, pp. 1968–1977, Jul. 2016.

[5] K. Chen and J.-K. Kämäräinen, "Learning to count with back-propagated information," in Proceedings of International Conference on Pattern Recognition, 2014, pp. 4672–4677.

[6] V.-Q. Pham, T. Kozakaya, O. Yamaguchi, R. Okada, "COUNT forest: Co-voting uncertain number of targets using random forest for crowd density estimation", Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), pp. 3253-3261, Dec. 2015.

[7] B. Chan, Z.-S. J. Liang, and N. Vasconcelos. Privacy preserving crowd monitoring: Counting people without people models or tracking. In Computer Vision and Pattern Recognition, pages 1–7. IEEE, 2008.

[8] Ke Chen, Chen Change Loy, Shaogang Gong, Tao Xiang, "Feature Mining for Localised Crowd Counting", 2012.

[9] X. Ding, Z. Lin, F. He, Y. Wang, and Y. Huang, "A deeply-recursive convolutional network for crowd counting," in Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP), Apr. 2018, pp. 1942–1946.

[10] Cong Zhang, Hongsheng Li, Xiaogang Wang, Xiaokang Yang, "Cross-scene Crowd Counting via Deep Convolutional Neural Networks". In Computer Vision and Pattern Recognition, pages 833–841. IEEE, 2015.

[11] Yingying Zhang, Desen Zhou, Siqin Chen, Shenghua Gao, and Yi Ma. Single-image crowd counting via multi-column convolutional neural network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 589–597, 2016.

[12] C. Wang, H. Zhang, L. Yang, S. Liu, X. Cao, "Deep people counting in extremely dense crowds", Proc. ACM Int. Conf. Multimedia, pp. 1299-1302, Oct. 2015.

[13] A. Krizhevsky, I. Sutskever, G. E. Hinton, "ImageNet classification with deep convolutional neural networks", Proc. Adv. Neural Inf. Process. Syst. (NIPS), pp. 1097-1105, Dec. 2012.

[14] D. B. Sam, S. Surya, and R. V. Babu, "Switching convolutional neural network for crowd counting," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), vol. 1, Jul. 2017, p. 6.

[15] L. Zhang, M. Shi, and Q. Chen, "Crowd counting via scale-adaptive convolutional neural network," 2017, arXiv: 1711.04433. [Online]. Avail-able: http://arxiv.org/abs/1711.04433SaCNN

[16] V. A. Sindagi and V. M. Patel, "Generating high-quality crowd density maps using contextual pyramid CNNs," in Proc. IEEE Int. Conf. Comput. Vis. (ICCV), Oct. 2017, pp. 1879–1888.

[17] X. Ding, Z. Lin, F. He, Y. Wang, and Y. Huang, "A deeply-recursive convolutional network for crowd counting," in Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP), Apr. 2018, pp. 1942–1946.

[18] L. Boominathan, S. S. S. Kruthiventi, and R. V. Babu, "CrowdNet: A deep convolutional network for dense crowd counting," in Proc. ACM Multimedia Conf. (MM), 2016, pp. 640–644.

[19] Kaiming He, Georgia Gkioxari, Piotr Doll´ar, Ross Girshick, "Mask R-CNN". In Computer Vision and Pattern Recognition, pages 2548–2554. 2017.

[20] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: towards real-time Pattern object detection with region proposal networks," IEEE Transactions on Analysis & Machine Intelligence, no. 6, pp. 1137–1149, 2017.

[21] Yuhong Li, Xiaofan Zhang, Deming Chen, "CSRNet: Dilated Convolutional Neural Networks for Understanding the Highly Congested Scenes". In Computer Vision and Pattern Recognition, pages 1091–1100. IEEE, 2018.

[22] Weizhe Liu, Mathieu Salzmann, Pascal Fua, "Context-Aware Crowd Counting". In Computer Vision and Pattern Recognition, pages 5099–5108. IEEE, 2019.

[23] Dongdong Zhang, Pengfei Zhang, Lisheng Wang, "Cell Counting Algorithm Based on YOLOv3 and Image Density Estimation". In: 2019 IEEE 4th

International Conference on Signal and Image Processing (ICSIP), pp. 920–924. July 2019,

[24] Linjun Yao, "An Effective Vehicle Counting Approach Based on CNN". In: 2019 IEEE 2nd International Conference on Electronics and Communication Engineering (ICECE), pp. 15–19, Dec. 2019.

[25] Vehicle Detection and Counting in High-Resolution Aerial Images Using Convolutional Regression Neural Network. In IEEE Access. Dec. 2017, page(s): 2220 – 2230.

[26] Chengquan Zhou, Hongbao Ye, Jun Hu, Xiaoyan Shi, Shan Hua, Jibo Yue, Zhifu Xu, Guijun Yang, "Automated Counting of Rice Panicle by Applying Deep Learning Model to Images from Unmanned Aerial Vehicle Platform". In Sensors and Systems for Smart Agriculture. 2019.06, pp. 1-16.

[27] Heavy Charged Particle Nuclear Track Counting Statistics and Count Loss Estimation in High Density Track Images. In IEEE Transactions on Nuclear Science. Oct. 2014, page(s): 2727 – 2734