Illumination-Enhanced Crowd Counting Based on IC-Net in Low Lighting Conditions

Haoyu Zhao[1], Weidong Min[2,3(✉)], and Yi Zou[1]

[1] School of Information Engineering, Nanchang University, Nanchang 330031, China
[2] School of Software, Nanchang University, Nanchang 330047, China
minweidong@ncu.edu.cn
[3] Jiangxi Key Laboratory of Smart City, Nanchang 330047, China

Abstract. The low lighting in some extreme conditions always affect the accuracy of the crowd counting and other vision tasks. The existing methods mainly rely on the generalization ability of deep-learning model to count the crowd number. But in extremely low lighting conditions, these methods are not efficient. To alleviate this issue, this paper proposes a novel approach, named Illumination-aware Cascading Network (IC-Net). The IC-Net can handle the low lighting conditions and generate a high-quality crowd density map. It contains two submodules, i.e., the Illumination Fusion Module and the Feature Cascading Module. The Illumination Fusion Module can fuse the low-illumination feature and the illumination enhanced feature to highlight the missing feature in darkness. The Feature Cascading Module is a cascading model and used to further express the illumination feature. It can generate the high-quality density map. In addition, a new dataset is collected, named Low Light Scenes Crowd (LLSC) dataset, which all come from extremely low illumination conditions. Experimental results on LLSC and benchmark show that the proposed method outperforms the existing state-of-the art methods in such extreme conditions.

Keywords: Crowd counting · Low lighting conditions · Deep learning

1 Introduction

Crowd counting, applied to many domains, such as security systems, urban planning, and video surveillance, is an interesting and useful technology [1,2]. It aims to count the number of people in an area [3–5]. With the development of deep learning, CNN-based methods achieve the amazing performance on many tasks, including crowd counting. [6] firstly proposed multi-column network to handle

Supported by the National Natural Science Foundation of China (Grant No. 62076117 and No. 61762061), the Natural Science Foundation of Jiangxi Province, China (Grant No. 20161ACB20004) and Jiangxi Key Laboratory of Smart City (Grant No. 20192BCD40002).

Y. Peng et al. (Eds.): ICIG 2021, LNCS 12888, pp. 222–234, 2021.
https://doi.org/10.1007/978-3-030-87355-4_19

this issue. [7,8] mainly relied on the generalization capability of the deep-learning model. Following the train-and-test pattern, these models can get some good performance on benchmarks. But such power has its limits, especially facing the extreme conditions, such as low lighting, perspective distortion, and dense crowd.

In addition, illumination feature is very significant to vision tasks. So, the deep-learning methods is hard to use in low lighting environment. Several work had been done to solve this issue in many domains. [9] proposed a deep neural network for low lighting field restoration. [10] proposed an illumination recovery model to transform severe varying illumination to slight illumination. [11] introduced an illumination-aware Faster R-CNN for object detection. To solve the crowd counting task in low lighting scenes, [12] proposed a deep spatial regression model to handle the appearance variations and the illumination various problems. [13] combined the audio information as auxiliary feature for crowd counting in low lighting environments. But audio feature has limitations when in large and open place. Thus, the existing crowd counting methods cannot directly be used in low lighting scenes.

In order to alleviate such problem, this paper proposes a novel and end-to-end approach, named Illumination-aware Cascading Network (IC-Net). It contains two submodules, i.e., the Illumination Fusion Module and the Feature Cascading Module. The Illumination Fusion Module can fuse the low-illumination feature and the illumination enhanced feature to highlight the missing feature in darkness. The Feature Cascading Module is a cascading model and used to further express the illumination feature. It can generate the high-quality density map. Due to lacking such challenging dataset, this work collects a new dataset, named Low Light Scenes Crowd (LLSC) dataset. The images come from extremely low illumination conditions in outdoor and indoor. The experiments based on the self-collected dataset and benchmark show that the proposed approach outperforms the existing methods.

The main contributions of this study are summarized as follows:

1. This work proposes a novel IC-Net for crowd counting, which can handle the low lighting conditions and generate a high-quality crowd density map.
2. To fuse the illumination feature, the Illumination Fusion Module and the Feature Cascading Module are proposed. They can highlight the missing illumination information in darkness and further express the CNN feature.
3. A new dataset which contains multiple scenes in low lighting conditions is proposed. The IC-Net can get good performances on self-collected dataset and benchmark.

2 Related Work

Crowd counting has significant applications in people's daily life. Many excellent methods based on deep-learning approach have been proposed to solve this problem. Some work also tried to solve this issue in extreme conditions, such as apparent perspective distortion, dense crowd, and illumination variations.

2.1 Crowd Counting Based on Deep Learning

Existing methods mainly based on deep-learning structure [14,15] to solve the crowd counting. Zhang et al. [6] proposed a simple but effective Multi-column Convolutional Neural Network (MCNN) to estimate the crowd density map. Zeng et al. [16] proposed an improved multi-scale CNN. Different from the multi-column network, Wu et al. [17] present a featured channel enhancement block for crowd counting. Cao et al. [18] introduced an encoder-decoder approach to extract multi-scale features and generates high-resolution density maps. In addition, considering the scale variation problem, [19–21] proposed novel network structures with structured features and fixed small receptive fields. [22–24] tried to use map-estimation networks to count the highly dense crowds in images. Some work also tried to solve the over-fitting for crowd counting. Such as Shi et al. [25] designed a new learning strategy to produce generalizable features by the means of deep negative correlation learning.

2.2 Crowd Counting Methods for Low Lighting Scenes

Due to the importance of illumination feature to vision-based tasks, some researchers also explored some methods to count the crowd in low lighting scenes. Hu et al. [13] introduced a novel task of audiovisual crowd counting, in which visual and auditory information are integrated for counting purposes. Wu et al. [26] proposed an adaptive scenario discovery framework for counting crowds with varying densities, which can deal different environments. Zhao et al. [27] designed a depth embedding module as add-ons into existing networks, which aims to solve the scale and illumination variety. Some work also built new benchmark for crowd counting. Wang et al. [28] collected a large-scale dataset which contains many low lighting scenes. It can also improve the train-test pattern models' accuracy.

According to the above analyses, most of existing methods for crowd counting are based on deep-learning structure. And they mainly relied on the generalization capability of convolutional neural network. When facing some extreme conditions, such as low lighting scene, these methods would not get the satisfactory results. In spite of some work, such as [13] tried to use extra audio information to assist the vision feature. The audio feature has limitations in large and open place.

By contrast, this work proposes an illumination enhanced method for crowd counting. To the best of our knowledge, this is the first study which directly improves the illumination feature in low lighting scenes for crowd counting. And a challenging dataset is also collected to show the efficiency of the proposed model.

3 The Proposed IC-Net for Crowd Counting

To solve the crowd counting issue in extremely low lighting conditions, this paper proposes a novel approach, named Illumination-aware Cascading Network (IC-Net), as the Fig. 1 shows. The IC-Net is an end-to-end training structure which

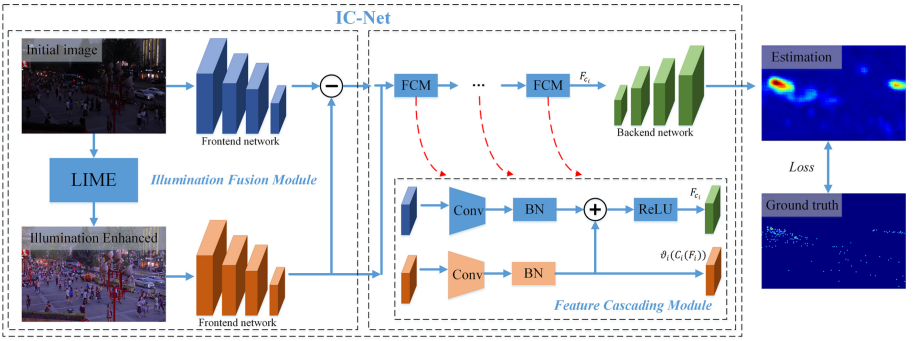

Fig. 1. The crowd counting based on Illumination-aware Cascading Network (IC-Net).

contains two submodules, i.e., the Illumination Fusion Module and the Feature Cascading Module. It can handle the low lighting conditions and generate a high-quality crowd density map. The input crowd image is converted to the density map by [6]. The density map is used to estimate the crowd number.

3.1 Illumination Fusion Module (IFM)

The structure of the IFM is shown in Fig. 1. The input crowd image is taken in low lighting conditions. The image often suffers from low visibility and the crowd is hidden in the darkness. The poor illumination quality will significantly degenerate the performance of many computer vision tasks [10], including crowd counting. Due to lack of enough vision information, it is difficult for convolutional neural network to handle such images. To get the hidden information about the crowd, the initial image is dealt with LIME [29]. LIME belongs to Retinex-based category, which aims to enhance a low light image by estimating its illumination map.

As the Fig. 2 shows, the LIME has a good performance in the outdoor and indoor scenes. The top-left corner in red box is the illumination enhanced images, which have significant increases in brightness. To prove the improvement of illumination can indeed increase the accuracy of the model, this work also conducted related experiments, which can be found in Experiments.

After getting the illumination enhanced operation, IFM puts the low lighting image and the illumination enhanced image into the frontend network at the same time. The frontend network is constructed with convolution, ReLU, and Max-Pooling operations. It is used to get the initial image feature $F_{low-illumination}$ and $F_{enhanced-illumination}$. The enhanced image feature contains more detailed information than the low illumination feature. Such as the people in dark environments would not be detected. When convolutional neural network does the convolution operation, these features will be lost.

Fig. 2. The illumination enhanced results using LIME.

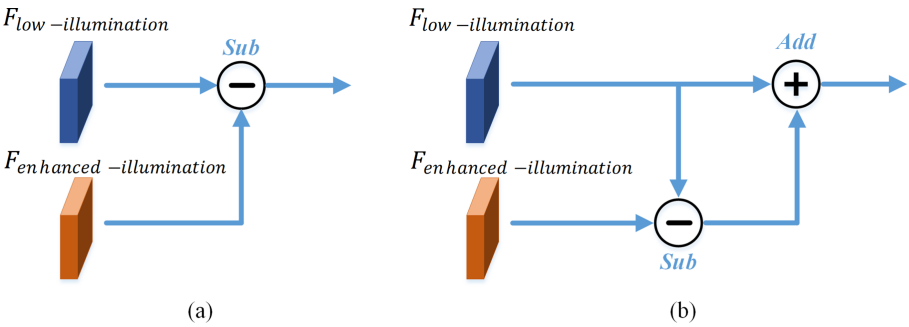

Fig. 3. Two different solutions to fuse the illumination features.

To highlight the missing illumination information in the original image feature, this work tries to fuse the two illumination features $F_{low-illumination}$ and $F_{enhanced-illumination}$. As the Fig. 3 shows, two different solutions are considered. The solution (a) and (b) can be described as Eq. (1) and Eq. (2).

$$F_f = sub(F_{low-illumination}, F_{enhanced-illumination}) \tag{1}$$

$$F_f = add(sub(F_{enhanced-illumination}, F_{low-illumination}), F_{low-illumination}) \tag{2}$$

The $sub()$ operation is used to do the subtraction between the corresponding features. The $add()$ operation is used to do the addition between the corresponding features. F_f is the fused image feature. To get the difference between the low-illumination and the enhanced illumination images, $sub()$ operation can get such difference. The $add()$ operation tries to supplement the original information of the image. In the experiment, the solution (a) is found having better performance. This work analyzes that the overlay of the supplement feature and the original feature will bring some interferences.

3.2 Feature Cascading Module (FCM)

After the IFM, the fused image feature F_f and the illumination-enhanced feature F_i are integrated by several FCMs. F_i donates the $F_{enhanced-illumination}$. The structure of the FCM can be seen in Fig. 1. This module is a cascading model and is used to further combine the illumination feature and the image feature adaptively. The two features are dealt with convolution (C_f and C_i) and batch normalization (ϑ_f and ϑ_i) operations. The batch normalization can accelerate the network learning rate. Then, the two output features are dealt with $add()$ and ReLU operations. The whole process can be described as Eq. (3).

$$F_{c_i} = ReLU(add(\vartheta_f(C_f(F_f)), \vartheta_i(C_i(F_i)))) \tag{3}$$

The F_{c_i} donates the output feature of i module. This work set six FCMs to extract the image. In the final FCM, the feature $\vartheta_i(C_i(F_i))$ doesn't enter into the backend network. The output feature F_c can be got by Eq. (4).

$$F_c = \sum_{i=1}^{n} F_{c_i}, (n = 6) \tag{4}$$

To get the final estimation results, the feature F_c is sent into backend network to recover the size of the feature. In addition, the Mean square loss function is used to train the model. The estimation density map $\hat{M}_e$ and the ground truth density map c can be calculated with Eq. (5).

$$loss = \sum (M_g - \hat{M}_e)^2 \tag{5}$$

4 Experiments

The IC-Net for crowd counting is implemented under the Windows 10 and Pytorch 1.4.0 experimental environment. The hardware environments are Inter Xeon E-2136 3.3 GHz and Quadro P5000.

4.1 Evaluation Metrics

The two standard evaluation metrics to test the IC-Net is used, i.e., Mean Absolute Error (MAE) and the Root Mean Squared Error (RMSE) [30,31]. They are defined as Eq. (6) and Eq. (7).

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i| \tag{6}$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2} \tag{7}$$

The parameter N represents the total number of the test images, y_i is the ground-truth number of people inside the whole i image and $\hat{y}_i$ is the estimated number of people.

4.2 Models' Performances with the Illumination Enhanced Image

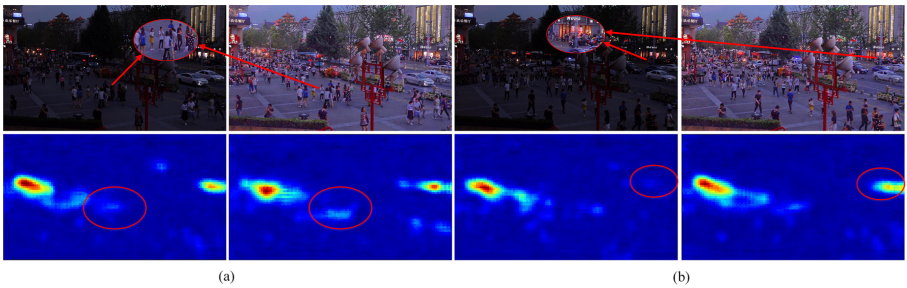

Fig. 4. The different experimental results of the darkness and the brightness.

The Importance of Illumination Information. As the Fig. 4 shows, two scenes (a) and (b) are tested. The first and the third columns are initial low lighting images, the second and the fourth columns are illumination enhanced images. The second row is the corresponding density map results generated by CANNet [8] for each image. The people in low lighting images are almost invisible. Such as the areas with red oval in the first row, these people are easily ignored by the convolutional neural network. The illumination enhanced images recover such detailed information. The images in the second row show that the people can be detected in illumination enhanced image. But they cannot be detected in low lighting image. This result explains the importance of the illumination for the vision tasks, including crowd counting.

Model Performances on Extended Datasets. In order to test the influence of the illumination enhanced images, this work tests the CANNet, CSRNet [7] and the MSR-FAN [32] on the ShanghaiTech dataset [6]. ShanghaiTech dataset includes part A and part B. Part A has 482 images and part B has 716 images. The two parts are divided into training data and testing data. This work directly improves the brightness of the images in Par A and Part B. The illumination enhanced images and the initial images are put together to train the model.

Table 1. The training results on extended datasets.

Methods	Part A		Part B	
	MAE	RMSE	MAE	RMSE
The initial images				
CSRNet [7]	68.2	115.0	10.6	16.0
CANNet [8]	62.3	100.0	7.8	12.2
MSR-FAN [32]	**59.9**	**94.6**	**7.6**	**11.3**
The illumination enhanced images and the initial images				
CSRNet [7]	67.3	113.6	9.2	15.3
CANNet [8]	60.6	100.2	7.1	12.3
MSR-FAN [32]	**57.1**	**98.1**	**6.0**	**11.1**

The Table 1 shows that the performance of these model all gets better results when extended the dataset with illumination enhanced images. The CANNet gets the MAE with 62.3 in the initial dataset and gets the MAE with 60.6 in the extended dataset. The MSR-FAN boosts 2.8 in MAE. The structures of these networks have not changed, but the performances of the networks have improved. This work assumes that the illumination enhanced images in extended dataset bring some neglected feature for network training step.

The Performances of IC-Net in Different Datasets. To test the IC-Net in extremely low lighting conditions, this work collected a new dataset, named Low Light Scenes Crowd (LLSC) dataset. It contains 780 images and they all come from DISCO dataset [13]. The DISCO dataset contains 1,935 images and the corresponding audio clips, and 170,270 annotated instances. These images include many kinds of scenes and conditions. This work chooses all the low lighting images from the DISCO and the audio clips are useless in this work. In these scenes, the illumination information is weak and some people are invisible, which are difficult to count. So, the LLSC is a challenging dataset.

To verify the impact of the model, some state-of-the-art methods are also tested on DISCO dataset. This paper compared with MCNN [6], AudioCSRNet [13], CANNet, and CSRNet. AudioCSRNet is also a novel method which tried to combine audio feature to assist the crowd counting task.

The Table 2 shows the experimental results on LLSC dataset and DISCO dataset. The IC-Net gets the MAE with 20.50 and RMSE with 29.08, which is the best model on LLSC. The AudioCSRNet gets the MAE with 21.46 and RMSE with 29.43 on LLSC. On the DISCO dataset, the IC-Net can get the MAE with 13.01 and RMSE with 26.98. The AudioCSRNet gets the MAE with 13.34 and RMSE with 27.20, which is lower than IC-Net. Except the low lighting images, the DISCO contains many bright scenes which can be handled by network. The LIME would not improve such images' illumination. The two same images are sent into IC-Net and get the final density map. Due to the help of audio feature, the AudioCSRNet performs better than MCNN, CANNet and the CSRNet. It can illustrate that the illumination feature can indeed help deep-learning model to get higher accuracy.

Table 2. The experimental results on the LLSC and DISCO datasets.

Methods	LLSC dataset		DISCO dataset	
	MAE	RMSE	MAE	RMSE
MCNN [6]	68.30	71.08	42.89	66.36
CSRNet [7]	26.81	35.72	17.66	33.35
CANNet [8]	24.08	31.98	15.12	29.85
AudioCSRNet [13]	21.46	29.43	13.34	27.20
Ours	**20.50**	**29.08**	**13.01**	**26.98**

Except from the comparison on the benchmarks, the visualized results between AudioCSRNet and the IC-Net on several images in LLSC are also introduced in Fig. 5. The first line is the initial images, the second line is the illumination-enhanced images, the third line is the ground truth, the fourth line is the results of AudioCSRNet, and the fifth line is the results of IC-Net. The 'gt' donates the ground truth number of the crowd. The 'es' donates the estimation number of the crowd. It can be found that the IC-Net has a good performance.

4.3 Ablation Studies

Considering that different structures of network have different performances [33–35], this work does some ablation studies to prove the efficiency of the IC-Net. In IFM, two different solutions, as the Fig. 3 shows, come forward to fuse the illumination features. To get the best accuracy of IC-Net, the two solutions, $IC_{(a)}$ and $IC_{(b)}$, are tested on the LLSC.

From the Table 3, it can be seen that the $IC_{(a)}$ performs better than $IC_{(b)}$. So, the solution (a) in Fig. 3 is employed in IC-Net to fuse the illumination features.

Table 3. The results tested on LLSC of two solutions.

Methods	MAE	RMSE
$IC_{(a)}$	**20.50**	**29.08**
$IC_{(b)}$	23.46	31.53

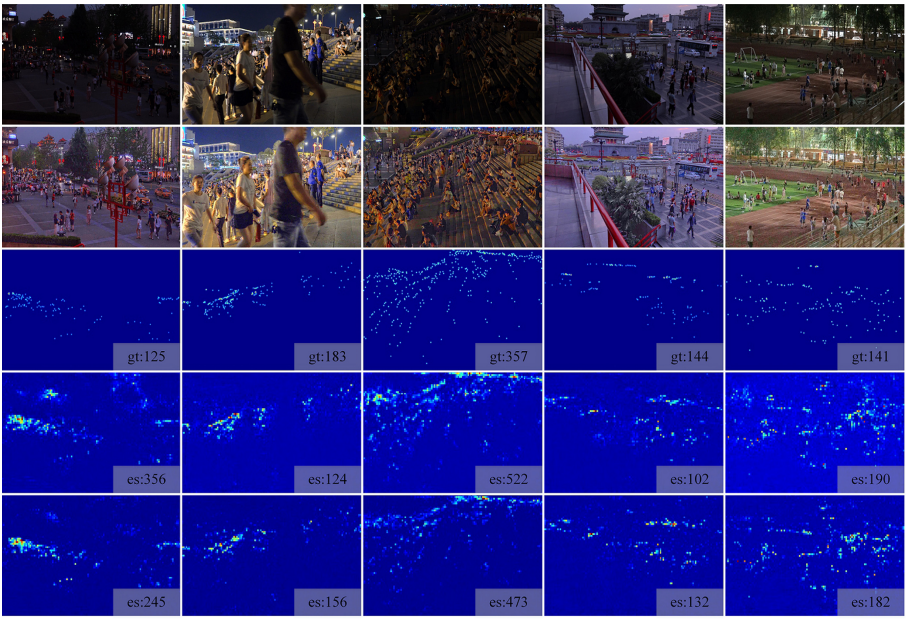

Fig. 5. Visualized results of the AudioCSRNet and IC-Net on several images in LLSC.

In addition, IC-Net contains several FCMs. To find the best number of the FCM, this work also tests the $IC_{(3)}$, $IC_{(4)}$, $IC_{(6)}$, and $IC_{(7)}$ on LLSC dataset. $IC_{(6)}$ donates that six FCMs are employed in IC-Net. From the Table 4, it can be found that the $IC_{(6)}$ performs best. So, the number of the FCM are set as six in IC-Net.

Table 4. The performances of IC-Net with different number of FCMs.

Methods	$IC_{(3)}$	$IC_{(4)}$	$IC_{(6)}$	$IC_{(7)}$
MAE	25.53	22.98	**20.50**	24.08
RMSE	34.93	33.61	**29.08**	35.36

5 Conclusion

In this work, a novel approach named Illumination-aware Cascading Network (IC-Net) is proposed. The IC-Net can handle the low lighting conditions and generate a high-quality crowd density map. It contains two submodules, i.e., the Illumination Fusion Module and the Feature Cascading Module. The Illumination Fusion Module can fuse the low-illumination feature and the illumination enhanced feature to highlight the missing feature in darkness. The Feature Cascading Module is a cascading model and used to further express the illumination feature. It can generate the high-quality density map. Experimental results show that the proposed method outperforms the existing state-of-the art methods in such extreme conditions.

In the future, more work will be done to improve the accuracy of the crowd counting model in extreme conditions.

References

1. Liu, Y., Wen, Q., Chen, H., et al.: Crowd counting via cross-stage refinement networks. IEEE Trans. Image Process. **29**, 6800–6812 (2020)
2. Gao, J., Wang, Q., Li, X.: PCC Net: perspective crowd counting via spatial convolutional network. IEEE Trans. Circ. Syst. Video Technol. **30**(10), 3486–3498 (2019)
3. Sindagi, V., Patel, V.: A survey of recent advances in CNN-based single image crowd counting and density estimation. Pattern Recogn. Lett. **107**, 3–16 (2018)
4. Huang, S., Xi, L., Zhang, Z.: Body structure aware deep crowd counting. IEEE Trans. Image Process. **27**(3), 1049–1059 (2018)
5. Zhang, S., Li, H., Kong, W.: Object counting method based on dual attention network. IET Image Process. **14**(8), 1621–1627 (2020)
6. Zhang Y., Zhou D., Chen S., et al.: Single-image crowd counting via multi-column convolutional neural network. In: Proceedings of IEEE Conference Computer Vision and Pattern Recognition (CVPR), Las Vegas, pp. 589–597. IEEE Xplore (2016)
7. Li Y., Zhang X., Chen D.: CSRNet: dilated convolutional neural networks for understanding the highly congested scenes. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1091–1100. IEEE Xplore, Utah (2018)
8. Liu, W., Salzmann, M., Fua, P.: Context-aware crowd counting. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, pp. 5094–5103. IEEE Xplore (2019)
9. Lamba, M., Rachavarapu, K., Mitra, K.: Harnessing multi-view perspective of light fields for low-light imaging. IEEE Trans. Image Process. **30**, 1501–1513 (2021)
10. Hui, C., Yu, J., We, F., et al.: Face illumination recovery for the deep learning feature under severe illumination variations. Pattern Recogn. **111**, (2021). https://doi.org/10.1016/j.patcog.2020.107724
11. Li, C., Song, D., Tong, R., et al.: Illumination-aware faster R-CNN for robust multispectral pedestrian detection. Pattern Recogn. **85**, 161–171 (2019)
12. Yao, H., Han, K., Wan, W., et al.: Deep spatial regression model for image crowd counting. arXiv:1710.09757 (2017)

13. Hu, D., Mou, L., Wang, Q., et al.: Ambient sound helps: audiovisual crowd counting in extreme conditions. arXiv:2005.07097 (2020)
14. Yang, H., Liu, L., Min, W., et al.: Driver yawning detection based on subtle facial action recognition. IEEE Trans. Multimedia **23**, 572–583 (2021)
15. Wang, Q., Min, W., He, D., et al.: Discriminative fine-grained network for vehicle reidentification using two-stage re-ranking. Sci. China Inf. Sci. **63**(11), 1–12 (2020)
16. Zeng, L., Xu, X., Cai, B., et al.: Multi-scale convolutional neural networks for crowd counting. In: Proceedings of IEEE International Conference on Image Processing (ICIP), Beijing, pp. 465–469. IEEE Xplore (2017)
17. Wu, X., Kong, S., Zheng, Y., et al.: Feature channel enhancement for crowd counting. IET Image Process. **14**(11), 2376–2382 (2020)
18. Cao, X., Wang, Z., Zhao, Y., Su, F.: Scale aggregation network for accurate and efficient crowd counting. In: Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y. (eds.) ECCV 2018. LNCS, vol. 11209, pp. 757–773. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-01228-1_45
19. Liu, L., Qiu, Z., Li, G., et al.: Crowd counting with deep structured scale integration network. In: Proceedings of IEEE International Conference on Computer Vision (ICCV), Seoul, pp. 1774–1783. IEEE Xplore (2019)
20. Qiu, Z., Liu, L., Li, G., et al.: Crowd counting via multi-view scale aggregation networks. In: Proceedings of IEEE International Conference on Multimedia and Expo (ICME), Shanghai, pp. 1498–1503. IEEE Xplore (2019)
21. Zhang, L., Shi, M., Chen, Q.: Crowd counting via scale-adaptive convolutional neural network. In: Proceedings of IEEE Winter Conference on Applications of Computer Vision (WACV), Lake Tahoe, pp. 1113–1121. IEEE Xplore (2018)
22. Lokesh, B., Srinivas, S., Venkatesh, R.: CrowdNet: a deep convolutional network for dense crowd counting. In: ACM International Conference on Multimedia, Amsterdam, pp. 640–644. ACM (2016)
23. Sam, D., Surya, S., Babu, R.: Switching convolutional neural network for crowd counting. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, pp. 4031–4039, IEEE Xplore (2017)
24. Li, H., Zhang, S.H., Kong, W.: Crowd counting using a self-attention multi-scale cascaded network. IET Comput. Vis. **13**(6), 556–561 (2019)
25. Shi, Z., Le, Z., Cao, X., et al.: Crowd counting with deep negative correlation learning. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, pp. 5382–5390. IEEE Xplore (2018)
26. Wu, X., Zheng, Y., Ye, H., et al.: Counting crowds with varying densities via adaptive scenario discovery framework. Neurocomputing **397**, 127–138 (2020)
27. Zhao, M., Zhang, C., Zhang, J., et al.: Scale-aware crowd counting via depth-embedded convolutional neural networks. IEEE Trans. Circ. Syst. Video Technol. **30**(10), 3651–3662 (2020)
28. Wang, Q., Gao, J., Lin, W., et al.: NWPU-crowd: a large-scale benchmark for crowd counting and localization. IEEE Trans. Pattern Anal. Mach. Intell. **43**(6), 2141–2149 (2020)
29. Guo, X., Li, Y., Ling, H.: LIME: low-light image enhancement via illumination map estimation. IEEE Trans. Image Process. **26**(2), 982–993 (2017)
30. Xiong, F., Shi, X., Yeung, D.: Spatiotemporal modeling for crowd counting in videos. In: Proceedings of IEEE International Conference on Computer Vision (ICCV), Venice, pp. 5161–5169. IEEE Xplore (2017)
31. Zhang, C., Li, H., Wang, X., et al.: Cross-scene crowd counting via deep convolutional neural networks. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, pp. 833–841. IEEE Xplore (2015)

32. Zhao H., Min W., Wei X., et al.: MSR-FAN: multi-scale residual feature-aware network for crowd counting. IET Image Process. 1–10 (2021). https://doi.org/10.1049/ipr2.12175
33. Shami, M., Maqbool, S., Sajid, H., et al.: People counting in dense crowd images using sparse head detections. IEEE Trans. Circ. Syst. Video Technol. **29**(9), 2627–2636 (2019)
34. Zhang, Y., Chang, F., Wang, M., et al.: Auxiliary learning for crowd counting via count-net. Neurocomputing **273**, 190–198 (2018)
35. Wang, L., Yin, B., Guo, A., et al.: Skip-connection convolutional neural network for still image crowd counting. Appl. Intell. **48**, 3360–3371 (2018)