International Journal of Computational Science and Engineering

ISSN online: 1742-7193 - ISSN print: 1742-7185
https://www.inderscience.com/ijcse

Cell counting via attentive recognition network

Xiangyu Guo, Jinyong Chen, Guisheng Zhang, Guofeng Zou, Qilei Li, Mingliang Gao

DOI: 10.1504/IJCSE.2022.10055133

Article History:

Received:	02 September 2022
Last revised:	12 October 2022
Accepted:	29 October 2022
Published online:	25 January 2024

Cell counting via attentive recognition network

Xiangyu Guo, Jinyong Chen, Guisheng Zhang and Guofeng Zou

School of Electrical and Electronic Engineering,
Shandong University of Technology,
Zibo, China
Email: xiangyvguo@163.com
Email: sdut_jychen@163.com
Email: sdut_guisheng@163.com
Email: gfzou@sdut.edu.cn

Qilei Li

School of Electronic Engineering and Computer Science,
Queen Mary University of London,
London, UK
Email: qilei.li@outlook.com

Mingliang Gao*

School of Electrical and Electronic Engineering,
Shandong University of Technology,
Zibo, China
Email: mlgao@sdut.edu.cn
*Corresponding author

Abstract: Accurate cell counting in biomedical images is a fundamental yet challenging task for disease diagnosis. The early manual cell counting methods are mainly based on detection and regression, which are time-consuming and prone to errors. Benefiting from the advent of deep learning, convolutional neural network (CNN)-based cell counting has become the mainstream method. Despite the outstanding performance of CNN-based cell counting methods, the complex tissue background in medical images still hinders the accuracy of cell counting. In this paper, to solve the problem of complex tissue background and improve the performance of cell counting, an attentive recognition network (ARNet) is built. Specifically, the ARNet is composed of five convolution blocks and a channel attention (CA) module. The convolution blocks are employed to extract the basic features, and the CA module is introduced to suppress the complex background by recalibrating the weight of each channel to pay more attention to cells. Subjective and objective experiments on synthetic bacterial cells (SBC) dataset and modified bone marrow (MBM) dataset prove that the proposed ARNet outperforms the mainstream methods in accuracy and stability.

Keywords: healthcare; cell counting; attention mechanism; convolutional neural network; CNN.

Reference to this paper should be made as follows: Guo, X., Chen, J., Zhang, G., Zou, G., Li, Q. and Gao, M. (2024) 'Cell counting via attentive recognition network', *Int. J. Computational Science and Engineering*, Vol. 27, No. 1, pp.1–8.

Biographical notes: Xiangyu Guo is pursuing his MS at the School of Electrical and Electronic Engineering, Shandong University of Technology, Zibo, China. His research interests include cell counting, crowd counting and deep learning.

Jinyong Chen is pursuing his MS at the School of Electrical and Electronic Engineering, Shandong University of Technology, Zibo, China. His research interests include computer vision and deep learning.

Guisheng Zhang is pursuing his MS at the School of Electrical and Electronic Engineering, Shandong University of Technology, Zibo, China. His research interests include fake face detection and deep learning.

Guofeng Zou is at the School of Electrical and Electronic Engineering, Shandong University of Technology, Zibo, China. His research interests include computer vision and deep learning.

Qilei Li is currently a PhD student with the School of Electronic Engineering and Computer Science, Queen Mary University of London, London, UK. He received an MS in Signal and Information Processing from Sichuan University. His research interests are computer vision and deep learning.

Mingliang Gao received his PhD in Communication and Information Systems from the Sichuan University, Chengdu, China, in 2013. He is currently an Associate Professor at the School of Electrical and Electronic Engineering, Shandong University of Technology, Zibo, China. His main research interests include computer vision and deep learning.

This paper is a revised and expanded version of a paper entitled 'Cell counting via attentive recognition network' presented at IoT and Healthcare International Conference (IoTHIC-2022), Haliç University, Istanbul, Turkey, 21–22 October 2022.

1 Introduction

The amount of cells could assist in the diagnosis of blood diseases (Venkatalakshmi and Thilagavathi, 2013), determining tumour types (Coates et al., 2015), and learning cellular and molecular genetic mechanisms (Solnica-Krezel, 2005). Therefore, automatic cell counting is vital in microscopy medical images analysis, and it has drawn extensive attention in the domain of computer vision.

The early cell counting methods are mainly based on detection and regression. The detection-based methods aim to discover the centroid position of cells for counting (Liu and Yang, 2015). These methods require a high accuracy of centroid labelling, and perform unsatisfactory in medical images with occlusions (Wan et al., 2022). The regression-based methods (Marsden et al., 2018; Xue et al., 2016) tend to learn a mapping from the input image to cell count. These methods are proven to outperform the detection-based methods in complex scenes. Nevertheless, the regression-based methods only output the cell counts, while cannot label the cell position information, which is also very valuable for medical image analysis (He et al., 2021). Due to the powerful feature extraction and inferring ability of convolutional neural networks (CNNs) in computer vision (Xing et al., 2022), the density map-based counting has become mainstream (Ciampi et al., 2022; He et al., 2020). Meanwhile, the attention mechanism further promoted the development of CNNs (Zhai et al., 2022a; Li et al., 2020; Zhang et al., 2021). The attention-based counting methods minimises the influence of irrelevant information by re-adjusting parameters in different dimensions (Guo et al., 2021).

Although the aforementioned methods enhance the performance of cell counting to some extent, the complex tissue background in medical microscopic images still hinder the further improvement of the counting accuracy. To mitigate the adverse effects of the complex tissue background, we propose the attentive recognition network (ARNet). It consists of five convolution blocks for extracting the low-level features. The first four blocks are employed to extract the basic features, and the last block adopts dilated convolution layers to enlarge the receptive field. Followed that, a channel attention (CA) module is introduced to suppress the background clutter. Specifically, the CA module consists of three process, i.e., global average pooling (GAP) for extruding 2D image to 1D sequence, one dimension convolution operation to adjust the channel weights and a sigmoid activate function to output the optimised weights. In a nutshell, the contributions of this paper are as follows.

1 An ARNet is proposed to enhance the counting performance in cell images with complex tissue background.

2 A CA module is introduced to adjust the weight along the channel dimension, which aims to reduce the background weight and increase the foreground weight.

3 Extensive experiments are conducted to demonstrate the performance of ARNet. Meanwhile, ablation study is performed to prove the effectiveness of the individual components in the proposed model.

The remainder of this paper is structured as follows.Related work is reviewed in Section 2. Proposed method is introduced in detail in Section 3. Details of the experiment and conclusion of this paper are provided in Sections 4 and 5, respectively.

2 Related work

In this section we describe two traditional cell counting methods, i.e., detection-based and regression-based method, and the mainstream methods, i.e., density estimation-based method.

Traditional methods deal with the cell counting by the means of detection (Arteta et al., 2016; Ellouze et al., 2022) and regression (Marsden et al., 2018; Gao et al., 2021). The counting by detection methods employs a detector to locate the cells and then sums up the cells as a count. Arteta et al. (2016) proposed a tree-structured discrete graphical model to detect all the cells in microscopy images. The detection-based methods are difficult to deal with occlusion and shape variations. The regression-based methods have been proposed to enhance the counting performance, which learn a mapping from the medical image to a count. For

example, Lempitsky and Zisserman (2010) built a general learning-based model for object counting tasks, i.e., crowd counting and cell counting.

Recently, the development of CNN has further enhanced the accuracy of cell counting. Cohen et al. (2017) built a redundant counting method to predict the density map, which is helpful to address the complicated object counting task. Falk et al. (2018) employed the U-Net to simultaneously complete the task of cell counting, detection and morphometric. Xie et al. (2018a) developed two parallel fully convolution regression networks, namely FCRN-A and FCRN-B to improve efficiency in an end-to-end manner. He et al. (2021) built an auxiliary CNN model to boost the main regression model. Furthermore, the attention mechanism (Guo et al., 2021, 2022b) has been adopted to the deep learning to improve network performance. Guo et al. (2019) incorporated a self-attention module to U-Net to enhance cell detection accuracy in microscopic images. Jiang and Yu (2020) built a weighted channel module to tackle the occlusions in cell counting.

3 The proposed method

Given a medical microscopic image $X \in \mathbb{R}^{H\times W\times 3}$, a density map $Y \in \mathbb{R}^{H\times W\times 1}$ can be regressed by the CNN. The number of cells is obtained by integrating the pixels on the density map. The process can be represented by,

$$N = \int Y = \int F(X;\theta), \tag{1}$$

where the $F(X;\theta)$ denotes a function of density regression, in which θ represents the parameters to be learned by the network.

3.1 Network design

The architecture of the proposed ARNet is depicted in Figure 1. It consists of five convolution blocks for extracting basic features and a CA module to cope with the tissue background. Specifically, the first three blocks are composed of convolution layers with dilated rate of 1 and max pooling layers to extract the cell features with small size, while the fourth block only consists of convolution layers without max pooling layer to suppress the detrimental effects on output accuracy caused by the pooling operation. The convolution layers with dilated rate of 2 in the final block are employed to enlarge the receptive field to ensure that the network can capture cells with large scale. To address the problem of complex background in medical image, the CA module is built. Lastly, a convolution layer with kernel size of 1×1 is employed to reduce the channels to 1 and output the density map. In a nutshell, the proposed ARNet is formulated as,

$$M = Conv_{1\times 1}(f_{ca}(Block_i(\mathbf{X}))), i \in \{1,2,3,4,5\}, \tag{2}$$

where the f_{ca} denotes a function of CA. $Block_i$ represents the operation of the five convolution blocks.

3.2 CA module

The purpose of CA is to select the targets (cells in this paper) by recalibrating the weight of each channel (Guo et al., 2021; Zhao et al., 2022; Hu et al., 2022) adaptively. As shown in Figure 1, the CA module contains three operations, i.e., GAP, one-dimensional convolution layer and an activation function.

The GAP operation reduces the input 2D image to a 1D array for subsequent parameter adjustments. The fast 1D convolution operation is performed to produce the channel weights. The size of the convolution kernel is an adjustable parameter, which affects the final channel weight generation (note: more details are referred in the ablation study). The sigmoid activation function is adopted to output the optimised weights. Last, element-wise multiplication is performed between the weights and the initial input features to generate the refined feature map. The map can suppress the side effects of background clutter and emphasise the areas where cells exist. Briefly, the CA module can be formulated as (Wang et al., 2020),

$$P = M \otimes \mathrm{Sig}(\mathrm{Conv1d}(\mathrm{GAP}(M))), \tag{3}$$

where P represents the refined feature map, and M is the extracted basic feature map. Conv1d denotes a fast 1-dimension convolutional operation. Sig denotes the Sigmoid function.

3.3 Loss function

The MSE function is adopted as the loss function to optimise the network. It can minimise the Euclidean distance between the ground truth and the prediction. The function is formulated as,

$$Loss = \frac{1}{N}\left\|y - \hat{y}\right\|_2^2, \tag{4}$$

where N denotes the number of test images. y and $\hat{y}$ denote the estimated and the ground truth values, respectively.

3.4 Ground truth generation

The ground truth map M_{gt} is generated by adopting a Gaussian kernel G_{σ_i} convolving a delta function (Zhang et al., 2016; Zhai et al., 2022c).

$$M_{gt} = \sum_{i=1}^{H} \delta(x - x_i) * G_{\sigma_i}(x), \sigma_i = \beta \bar{d}_i, \tag{5}$$

where H denotes the number of cell annotations and x refers to the position pixel. σ_i represents the variance of the kernel. β is a hyperparameter which is set to 0.3. $\delta(x - x_i)$ depicts a target cell.

Figure 1 Architecture of the proposed ARNet for cell counting (see online version for colours)

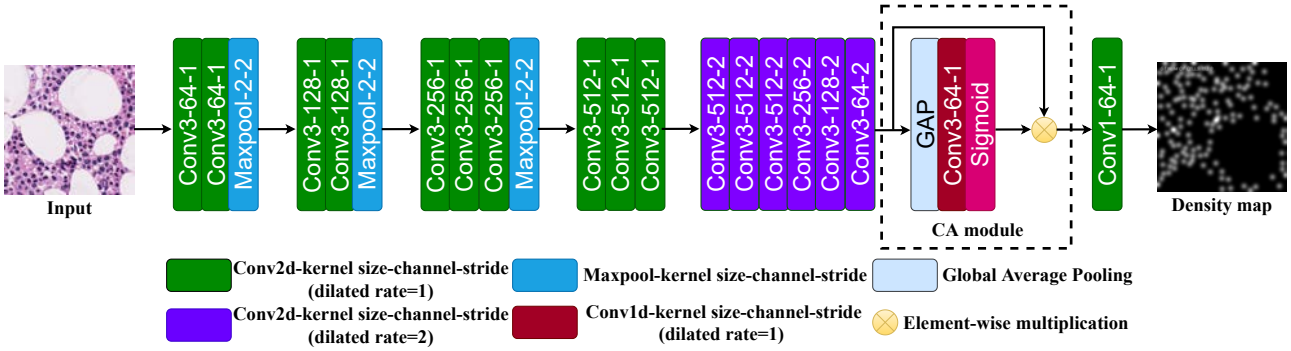
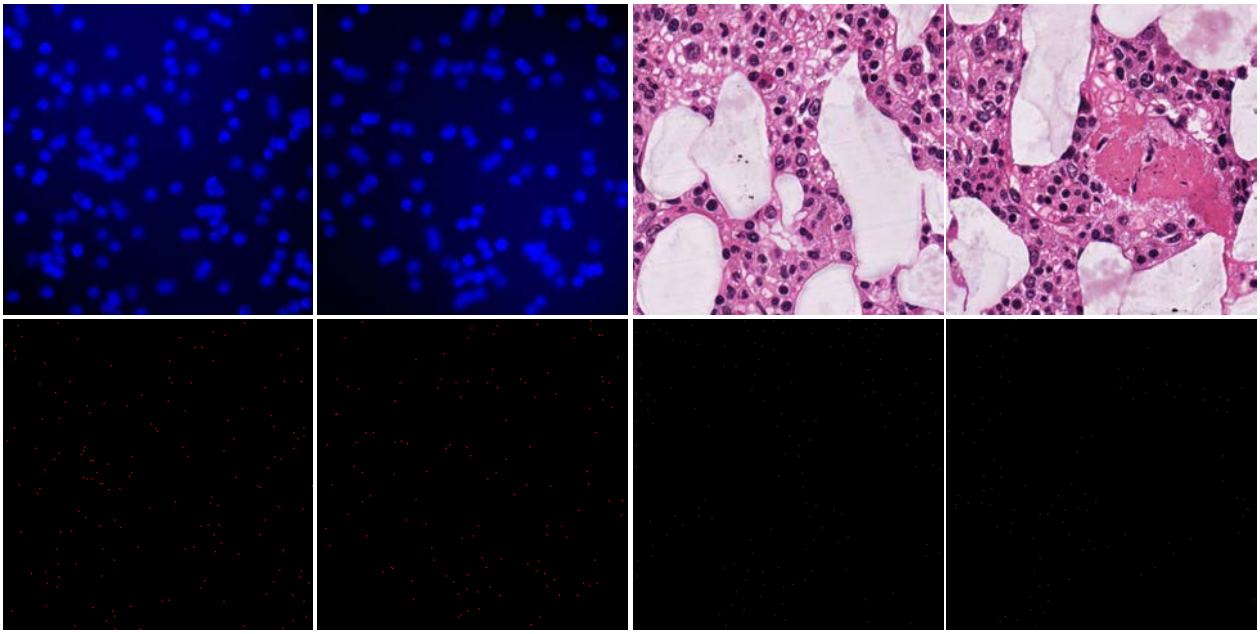


Figure 2 Examples of SBC and MBM datasets (see online version for colours)



Note: The left two images are from SBC dataset, while the right two images are selected from MBM dataset.

4 Experiment and analysis

4.1 Datasets

4.1.1 Synthetic bacterial cells

The synthetic bacterial cells (SBC) dataset was established by Visual Geometry Group (VGG) (Lempitsky and Zisserman, 2010). It includes 200 fluorescent microscopy images with resolution of 256 × 256. Because the dataset is synthetic, the cells almost have the same size, but they cluster together and are heavily obscured. Annotations are automatically executed during the data construction process, which makes the labels error-free. Samples of SBC dataset are present in Figure 2.

4.1.2 Modified bone marrow cells

The modified bone marrow (MBM) cell counting dataset contains 44 RGB microscopy images (600 × 600), which are collected from Kainz et al. (2015). Specifically, the images are obtained from 11 Hematoxylin-Eosin images (1,200 × 1,200) of bone marrow tissue. Each image is cut into four images with the same size. The labelled cells have large-scale variation and non-uniform background, which make the counting task more troublesome. Figure 2 provides some examples of the MBM dataset. More details of the two cell datasets are reported in Table 1.

Table 1 Details of the SBC and MBM datasets

Dataset	*SBC*	*MBM*
Resolution	256 × 256	600 × 600
Train/validation/test	50/50/100	15/15/14
Number of cell	174 ± 64	126 ± 33
Image type	Synthetic	Real

4.2 Implementation details

All experiments are performed based on the PyTorch framework and are implemented on two NVIDIA 3080 GPUs. For data augmentation, we adopt random cropping, horizontal and vertical flipped. To avoid the overfitting, we

set several dropout layers in the end of the network. At the training stage, Adam optimiser (Kingma and Ba, 2015) with a learning rate of 1e-5 is employed to optimise the network. The weight decay is set as 1e-3. The batch size is set as 8 for SBC dataset, and 4 for MBM dataset. The maximum number of training epochs is set as 1,000. During the test phase, 100 and 14 images are randomly selected as the samples for generating experimental results.

4.3 Evaluation metrics

The mean absolute error (MAE) and its related standard deviations (STD) are adopted as the evaluation metrics (Huan et al., 2020; Guo et al., 2022a; Zhai et al., 2022b) which are formulated as,

$$MAE = \frac{1}{S}\sum_{i=1}^{S}\left|C_{gt_i} - C_{est_i}\right|, \tag{6}$$

$$STD = \sqrt{\frac{1}{S-1}\sum_{i=1}^{S}\left(\left|C_{gt_i} - C_{est_i}\right| - MAE\right)^2}, \tag{7}$$

where S represents the number of test samples. C_{gt_i} and C_{est_i} denote the ground truth and estimated cell counts of the i-th images, respectively. The lower MAE and STD value indicate that the model has better counting accuracy and counting stability.

4.4 Comparative analysis

We compare and analyse the proposed ARNet with other cell counting methods. The objective comparison results are depicted in Table 2.

Table 2 Objective comparison results on SBC and MBM datasets

Methods	*SBC*		*MBM*	
	MAE	*STD*	*MAE*	*STD*
U-Net (Falk et al., 2018)	27.8	25.5	48.0	19.0
ResNet-152 (Xue et al., 2016)	7.5	2.2	-	-
StructRegNet (Xie et al., 2018b)	9.8	8.7	12.8	8.6
Mask R-CNN (He et al., 2020)	36.9	19.7	44.4	14.2
Marsden et al. (2018) method	-	-	20.5	3.5
FCRN-A (Xie et al., 2018a)	2.9	*0.2*	21.3	9.4
FCRN (Saxe et al., 2014)	2.8	2.5	8.5	7.6
ARNet (ours)	*2.7*	2.2	*5.0*	*3.2*

Notes: Italic indicates the best performance.

On the SBC dataset, one can see that the ARNet ranks first with a score of 2.7 in MAE, and it achieves 6.8% gains compared with the second-best method, i.e., FCRN-A (Xie et al., 2018a). Meanwhile, the ARNet achieves a competitive score of 2.2 in STD which ranks the second place. Specifically, compared with ResNet-152 (Xue et al., 2016) which also score 2.2 in STD, the proposed method improve the MAE score by 64%. Figure 3 provides some visualised results on SBC dataset.

On the MBM dataset, the proposed ARNet scores 5.0 and 3.2 in MAE and STD, which outperform all the reported methods. Compared with FCRN (Saxe et al., 2014), it improves by 41.1% and 57.9%, respectively. Furthermore, the (Marsden et al., 2018) method also adopts the GAP operation to compress the image to generate new weights. However, there is still a big gap compared with the proposed method in MAE (75.6%). Figure 4 shows some visualised results on MBM dataset.

4.5 Ablation study

Ablation studies are carried out on MBM dataset to explore the benefits of the CA module. The experimental configuration items are as follows.

- '*Baseline*' represents only five convolution blocks are employed as the basic network.
- '*Baseline + CA ($k = n$)*' means adding the CA module to the baseline, and the kernel size of the 1D convolution is set to n.

The ablation experimental results with four configurations are tabulated in Table 3. The item 'baseline' scores 9.6 and 7.6 in MAE and STD, respectively. It obtains the worst performance compared with other items. One can observe that the CA module is beneficial to boosting the counting accuracy and stability. Specifically, the score of item 2 (k = 3) is better than the item 3 (k = 5) in MAE, while the opposite is true for STD. By contrast, the item 4 (k = 1) achieves the best performance in both MAE and STD. The reason is that the size of cells in medical images is generally small, and thus a large convolution kernel would lead to the loss of detail.

Table 3 Comparative results with different configurations on MBM dataset

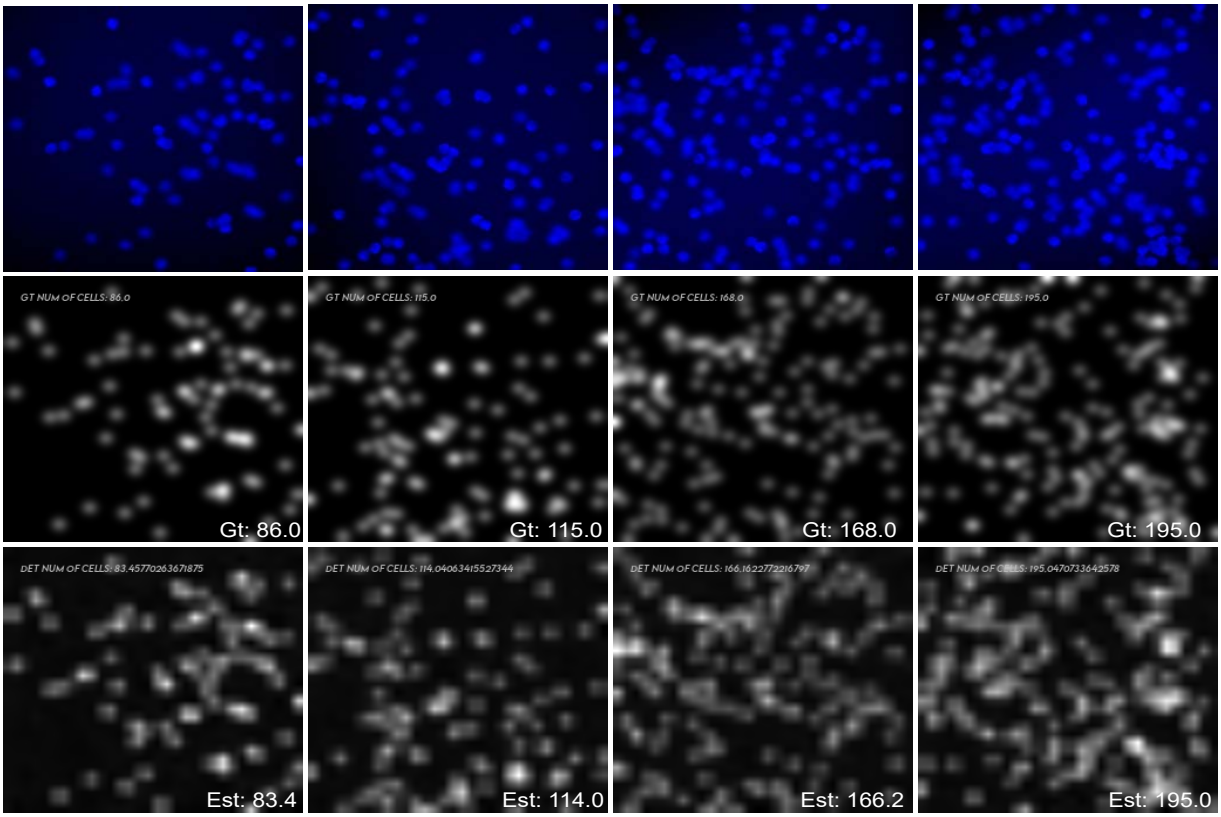
Methods	*MAE*	*STD*
Baseline	9.6	7.6
Baseline + CA (k = 3)	6.6	5.1
Baseline + CA (k = 5)	5.8	5.9
Baseline + CA (k = 1)	*5.0*	*3.2*

Note: Italic indicates the best performance.

5 Conclusions

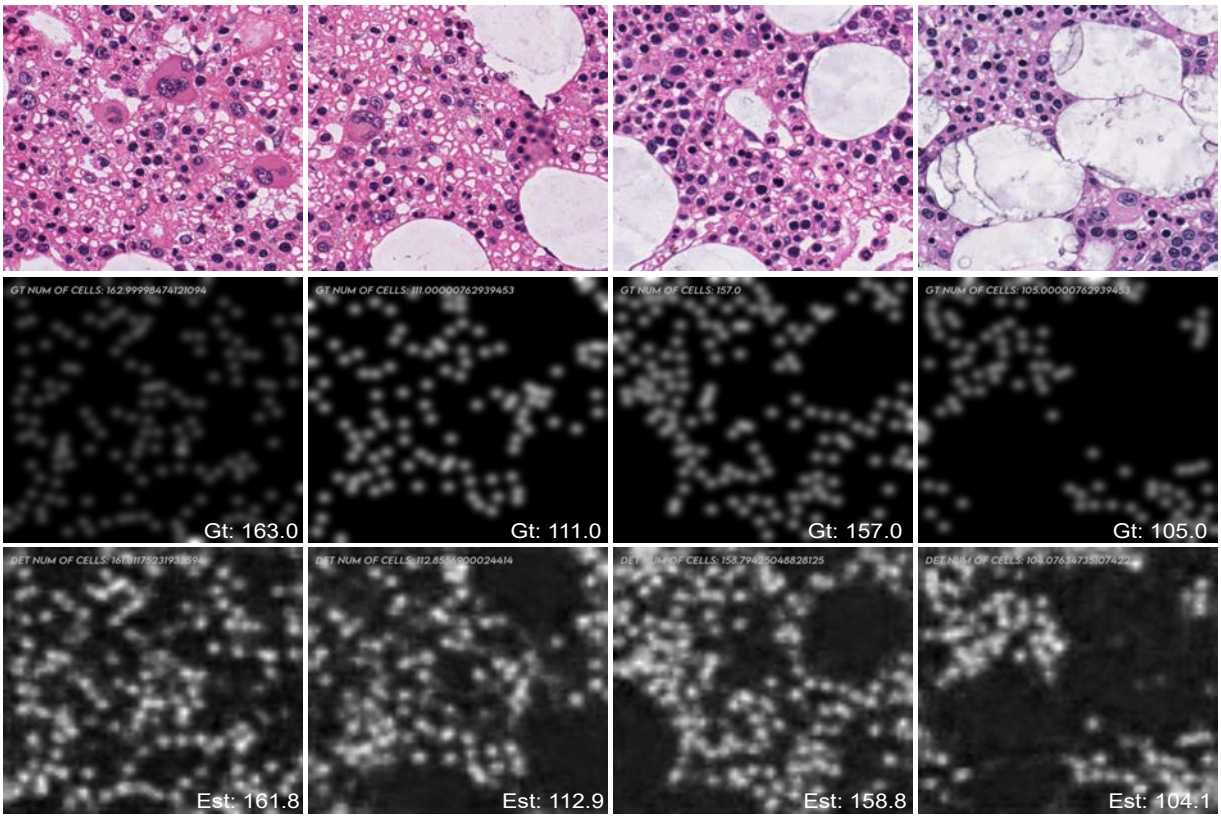
In this paper, we propose an ARNet to suppress the complex tissue background problem so as to improve the performance of cell counting in biomedical images. The ARNet consists of five basic convolution blocks to extract low-level features, and a CA module to deal with the complex tissue background. The CA module aims to re-adjust the weight along the channel dimension through GAP, 1D convolution and activation function. The feature map refined by the CA module can minimise the weights of background and improve the weight of foreground. Experimental results prove that the ARNet outperforms the mainstream methods in cell images with complex tissue background clutter.

Figure 3 Subjective comparison results on SBC datasets (see online version for colours)



Note: 'Gt' and 'Est' denote the ground truth and estimated counts.

Figure 4 Subjective comparison results on MBM datasets (see online version for colours)



Note: 'Gt' and 'Est' denote the ground truth and estimated counts.

References

Arteta, C., Lempitsky, V.S., Noble, J.A. and Zisserman, A. (2016) 'Detecting overlapping instances in microscopy images using extremal region trees', *Medical Image Analysis*, Vol. 27, pp.3–16.

Ciampi, L., Carrara, F., Amato, G. and Gennaro, C. (2022) 'Counting or localizing? Evaluating cell counting and detection in microscopy images', in *VISIGRAPP*.

Coates, A.S., Winer, E.P., Goldhirsch, A., Gelber, R.D., Gnant, M., Piccart-Gebhart, M.J., Thürlimann, B. and Senn, H. (2015) 'Tailoring therapies – improving the management of early breast cancer: St. Gallen international expert consensus on the primary therapy of early breast cancer 2015', *Annals of Oncology*, Vol. 26, No. 8, pp.1533–1546.

Cohen, J.P., Boucher, G., Glastonbury, C.A., Lo, H.Z. and Bengio, Y. (2017) 'Count-ception: counting by fully convolutional redundant counting', *2017 IEEE International Conference on Computer Vision Workshops (ICCVW)*, pp.18–26.

Ellouze, M., Mechti, S., Krichen, M., Ravi, V. and Belguith, L.H. (2022) 'A deep learning approach for detecting the behaviour of people having personality disorders towards COVID-19 from Twitter', *Int. J. Comput. Sci. Eng.*, Vol. 25, No. 4, pp.353–366.

Falk, T., Mai, D., Bensch, R., Çiçek, Ö., Abdulkadir, A., Marrakchi, Y., Böhm, A., Deubner, J., Jäckel, Z., Seiwald, K., Dovzhenko, A., Tietz, O., Bosco, C.D., Walsh, S., Saltukoglu, D., Tay, T.L., Prinz, M., Palme, K., Simons, M., Diester, I., Brox, T. and Ronneberger, O. (2018) 'U-Net: deep learning for cell counting, detection, and morphometry', *Nature Methods*, Vol. 16, No. 1, pp.67–70.

Gao, K., Chang, C. and Liu, Y. (2021) 'Predicting missing data for data integrity based on the linear regression model', *Int. J. Embed. Syst.*, Vol. 14, No. 4, pp.355–362.

Guo, Y., Stein, J.L., Wu, G. and Krishnamurthy, A.K. (2019) 'SAU-Net: a universal deep network for cell counting', *Proceedings of the 10th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics*.

Guo, M-H., Xu, T-X., Liu, J-J., Liu, Z-N., Jiang, P-T., Mu, T-J., Zhang, S-H., Martin, R.R., Cheng, M-M. and Hu, S-M. (2021) *Attention Mechanisms in Computer Vision: A Survey*, arXiv preprint arXiv:2111.07624.

Guo, X., Gao, M., Zhai, W., Li, Q., Pan, J. and Zou, G. (2022a) 'Multiscale aggregation network via smooth inverse map for crowd counting', *Multimedia Tools and Applications* [online] https://doi.org/10.1007/s11042-022-13664-8.

Guo, X., Gao, M., Zhai, W., Shang, J. and Li, Q. (2022b) 'Spatial-frequency attention network for crowd counting', *Big Data*, Vol. 10, No. 5, pp.453–465.

He, K., Gkioxari, G., Dollár, P. and Girshick, R.B. (2020) 'Mask R-CNN', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 42, No. 2, pp.386–397.

He, S., Minn, K.T., Solnica-Krezel, L., Anastasio, M.A. and Li, H. (2021) 'Deeply-supervised density regression for automatic cell counting in microscopy images', *Medical Image Analysis*, Vol. 68, p.1892.

Hu, J., Liu, Y. and Wu, K. (2022) 'Neural network pruning based on channel attention mechanism', *Connect. Sci.*, Vol. 34, No. 1, pp.2201–2218.

Huan, J., Cao, W., Gu, Y. and Qin, Y. (2020) 'A hybrid model of empirical wavelet transform and extreme learning machine for dissolved oxygen forecasting', *International Journal of Embedded Systems*, Vol. 13, No. 1, pp.9–17.

Jiang, N. and Yu, F. (2020) 'Cell counting with channels attention', *2020 IEEE 5th International Conference on Signal and Image Processing (ICSIP)*, pp.494–498.

Kainz, P., Urschler, M., Schulter, S., Wohlhart, P. and Lepetit, V. (2015) 'You should use regression to detect cells', in *MICCAI*.

Kingma, D.P. and Ba, J. (2015) 'Adam: a method for stochastic optimization', *Proceedings of the International Conference on Learning Representations (ICLR)*.

Lempitsky, V.S. and Zisserman, A. (2010) 'Learning to count objects in images', in *NIPS*.

Li, W., Wang, Z., Wu, X., Zhang, J., Peng, Q. and Li, H. (2020) 'CODAN: counting-driven attention network for vehicle detection in congested scenes', *Proceedings of the 28th ACM International Conference on Multimedia*.

Liu, F. and Yang, L. (2015) 'A novel cell detection method using deep convolutional neural network and maximum-weight independent set', in *MICCAI*.

Marsden, M.A., McGuinness, K., Little, S., Keogh, C.E. and O'Connor, N.E. (2018) 'People, penguins and petri dishes: adapting object counting models to new visual domains and object types without forgetting', *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp.8070–8079.

Saxe, A.M., McClelland, J.L. and Ganguli, S. (2014) 'Exact solutions to the nonlinear dynamics of learning in deep linear neural networks', *CoRR*.

Solnica-Krezel, L. (2005) 'Conserved patterns of cell movements during vertebrate gastrulation', *Current Biology*, Vol. 15, pp.R213–R228.

Venkatalakshmi, B. and Thilagavathi, K. (2013) 'Automatic red blood cell counting using hough transform', *2013 IEEE Conference on Information and Communication Technologies*, pp.267–271.

Wan, W-C., Luo, X., Ma, L. and Xie, S. (2022) 'Side-path FPN-based multi-scale object detection', *Int. J. Comput. Sci. Eng.*, Vol. 25, No. 1, pp.44–51.

Wang, Q., Wu, B., Zhu, P., Li, P., Zuo, W. and Hu, Q. (2020) 'ECA-Net: efficient channel attention for deep convolutional neural networks', *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp.11531–11539.

Xie, W., Noble, J.A. and Zisserman, A. (2018a) 'Microscopy cell counting and detection with fully convolutional regression networks', *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, Vol. 6, No. 3, pp.283–292.

Xie, Y., Xing, F., Shi, X., Kong, X., Su, H. and Yang, L. (2018b) 'Efficient and robust cell detection: a structured regression approach', *Medical Image Analysis*, Vol. 44, pp.245–254.

Xing, L., Ling, L. and Wu, X. (2022) 'Lithium-ion battery state-of-charge estimation based on a dual extended kalman filter and bpnn correction', *Connection Science*, Vol. 43, No. 1, pp.2332–2363.

Xue, Y., Ray, N., Hugh, J.C. and Bigras, G. (2016) 'Cell counting by regression using convolutional neural network', in *ECCV Workshops*.

Zhai, W., Gao, M., Anisetti, M., Li, Q., Jeon, S. and Pan, J. (2022a) 'Group-split attention network for crowd counting', *Journal of Electronic Imaging*, Vol. 31, No. 4, p.41214.

Zhai, W., Gao, M., Souri, A., Li, Q., Guo, X., Shang, J. and Zou, G. (2022b) 'An attentive hierarchy ConvNet for crowd counting in smart city', *Cluster Computing* [online] https://doi.org/10.1007/s10586-022-03749-2.

Zhai, W., Li, Q., Zhou, Y., Li, X., Pan, J., Zou, G. and Gao, M. (2022c) 'Da2Net: a dual attention-aware network for robust crowd counting', *Multimedia Systems* [online] https://doi.org/10.1007/s00530-021-00877-4.

Zhang, Y., Zhou, D., Chen, S., Gao, S. and Ma, Y. (2016) 'Single-image crowd counting via multi-column convolutional neural network', *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp.589–597.

Zhang, J., Qiao, J., Wu, X. and Li, W. (2021) 'Vehicle counting network with attention-based mask refinement and spatial-awareness block loss', *Proceedings of the 29th ACM International Conference on Multimedia*.

Zhao, W., Zhao, S., Chen, S., Weng, T-H. and Kang, W. (2022) 'Entity and relation collaborative extraction approach based on multi-head attention and gated mechanism', *Connection Science*, Vol. 34, No. 1, pp.670–686.