Journal of Medical Systems (2019) 43: 301
https://doi.org/10.1007/s10916-019-1426-y

IMAGE & SIGNAL PROCESSING

Recognition and Clinical Diagnosis of Cervical Cancer Cells Based on our Improved Lightweight Deep Network for Pathological Image

Hongzhu Wang [1] · Chuan Jiang [1] · Kunzhong Bao [1] · Caie Xu [2]

Received: 27 April 2019 / Accepted: 14 July 2019 / Published online: 1 August 2019

Abstract

Accurate recognition of cervical cancer cells is of great significance to clinical diagnosis, but these existing algorithms are designed by low-level manual feature, and their performance improvements are limited an improved algorithm based on residual neural network is proposed to improve the accuracy of diagnosis. Firstly, momentum parameters are introduced into the training model; secondly, by changing the number of training samples, the recognition rate of the algorithm can be improved. Therefore, aiming at the task of object recognition under resource constrained condition, we optimize the design method of the network structure such as convolution operation, model parameter compression and enhancement of feature expression depth, and design and implement the lightweight network model structure for embedded platform. Our proposed deep network model can reduce the parameters of the model and the resources needed for operation under the condition of guaranteeing the precision. The experimental results show that the lightweight deep model has better performance than that of the existing comparison models, and it can achieve the model accuracy of 94.1% under the condition that the model with fewer parameters on cervical cells data set.

Keywords Cervical cancer · Momentum parameter · Clinical diagnosis · Deep learning · Lightweight · Embedded platform

Introduction

With the continuous development of medical image processing and pattern recognition technology, cervical cancer cell detection technology based on machine vision emerges as the requirement [1]. This technology is expected to replace traditional manual screening methods to solve the problems of low efficiency, being greatly influenced by subjective factors and low accuracy in manual screening methods [2]. Since the technology is in the initial stage, there is a general problem, such as low accuracy of recognition, poor sensitivity and specificity of classification.

Intelligent cervical cell image analysis is one of the specific tasks that applies AI technology to achieve modern medical diagnosis [3]. Recently, many researchers are dedicated to the research on intelligent cervical cell image analysis and have made remarkable progress [4]. However, due to the characteristics of realistic cervical cell images as well as cervical cell's complex shapes and structures, the accuracy and real-time performance of existing cervical cell image processing, detection and recognition technologies need to be improved so as to meet the public's pressing demands for intelligent cervical cell image analysis [5]. By the virtue of pattern recognition and machine learning theory, some literature [6–9] provide a thorough analysis to image segmentation, feature extraction, feature selection, feature combination [10] and classification of cervical cell image, and conducts in-depth research on the key technologies of intelligent cervical cell image analysis to promote the accuracy and speed of the intelligent analysis system for cervical cell images [11].

To study on the segmentation of non-overlapping single cervical cell color image, literature [12] has proposed an accurate and fast image segmentation method based on super pixel gap-search MRF (Markov Random Field) model. Based on the pre-segmented super pixels, this model can extend MRF model to label super pixel patches to partition the image into cell nucleus, cytoplasm and background directly. Moreover, due to the proposed gap-search mechanism, a large amount of redundant computation is reduced to solve the

This article is part of the Topical Collection on *Image & Signal Processing*

✉ Caie Xu
G17DHL01@yamanashi.ac.jp

[1] Women's Hospital School Of Medicine Zhejiang University, Hangzhou 310006, Zhejiang, China

[2] Faculty of Engineering, University of Yamanashi, Kofu 4000016, Japan

problem and the proposed method is much faster than pixel-based MRF model. Experimental results show that the proposed segmentation method based on super pixe1 Gap-search MRF can segment the non-overlapping cervical single cell images more accurately and quickly than other segmentation approaches. However, literature [12] is not suitable for overlapping cells, whose segmentation effect is poor.

Therefore, the segmentation of cervical cell overlapping images is explored in literature [13] and this problem is divided into two categories: "partial overlapping" segmentation problem and "true overlapping" segmentation problem. As for "partial overlapping" segmentation of cervical cell image, an improved super pixel and K-means++ is proposed. Firstly, SLIC super pixel algorithm technique is employed to divide the filtered image into super pixel patches. Secondly, 13 dimension features are extracted from the divided super pixels and these super pixels are clustered into 4 classes by K-means++ algorithm. According to the different intensities of cluster's centers, the image is segmented to cervical nucleus, background, candidate cytoplasm and overlapped region. Finally, in the light of the prior knowledge of the cell shape, candidate cytoplasm and overlapped regions, cytoplasm is identified by removing the outside of cell shape. Experimental results show that the proposed method performs better than other classic unsupervised methods for the segmentation of "partial overlapping" cervical color image segmentation. As for "true overlapping" segmentation of cervical cell image, an improved method combines graph cuts and Voronoi diagram algorithms in [14]. Firstly, this method employs graph cuts for scene segmentation to identify background and foreground. Secondly, after detecting nuclei regions of each cell clump as seeds of Voronoi diagram, non-overlapping "rough" cells are segmented from the clump. Thirdly, according to the cell prior shape, the overlapping "compensation" region of each rough cell, which is not in Voronoi cell, is extracted. Finally, "rough" cell region and "compensation" cell region are combined as a whole cell. Experimental results show that the proposed method achieves the same segmentation accuracy as existing best segmentation methods over two challenge datasets of overlapping cervical cell segmentation.

The problem of feature selection and combination for cervical cell images is studied. Margin based adaptive feature selection and combination method is proposed in literature [15]. The basic idea is proposed based on minimizing the distance among instances in a cluster, meanwhile maximizing the distance among instances belong to different clusters. Specifically, the prior probability about the training data is used when solving multi-value weighted vector for feature optimization. In addition, when the feature selection and combination are conducted, the original features are transformed into the new feature space. It thus becomes easier to conduct classification of the instances in the new feature space. Experimental results illustrate that the proposed feature selection and combination method outperforms its counterparts on accuracy when performing the abnormal detection for cervical cell images. Literature [16] studies the outlier detection for cervical cell images. It argues that cervical cancer cell diagnosis system should guarantee both the accuracy and the efficiency. To improve the accuracy of diagnosis system, a three-phase outlier detection model is proposed for cervical cell image analysis. The model separately optimizes the three stages in cervical cancer outlier detection based on a well-segmented cell image. Specifically, it first extracts 160 features to represent the intrinsic characteristics of the cervical cell. Then, a margin-based adaptive feature selection method is used to learn an optimal combination of these features for outlier detection task. After that, the optimized features are fed into a two-stage classified strategy for cancer cell detection.

However, these algorithms are designed by low-level manual feature, and their performance improvements are limited [17]. Accurate recognition of cervical cancer cells is of great significance to clinical diagnosis [18]. Recently, the diagnosis of cervical cancer cells mainly depends on the manual feature, and judgment based on clinical experience may lead to higher errors [19]. There are complexity and difference among cervical cancer cells, and the shape of tissues and organs is irregular. The image data of cervical cancer cells have high-order statistical characteristics, resulting in a large amount of redundant information. In order to improve the recognition rate, deep learning is more and more widely used in the recognition and diagnosis of cervical cancer cells, and has a broad application prospect in the field of biology [20].

Convolutional neural network is the most representative model in deep learning. By constructing a multi-hidden layer model framework and using a large number of training data to simulate the high-level features of human brain function learning, the literature [21] applies convolutional neural network to recognize the cervical cancer cell, which can simplify the procedure of segmentation and extraction of features. With the increase of network layers, the training time is too long and the accuracy will tend to be saturated. On the basis of convolution neural network, the residual neural network is proposed in literature [22]. The advantage of residual neural network is that it can greatly reduce training time and prevent overfitting. However, the stochastic gradient descent algorithm is used to train the network, which will introduce noise sources and make the training parameters fluctuate greatly and the update of parameters unstable, and has a great impact on the recognition rate of the algorithm.

In this paper, an improved algorithm based on residual neural network is proposed to solve these problems. Firstly, by introducing momentum parameters into the training model, the noise interference can be reduced; secondly, by changing the number of training samples, the recognition rate of the algorithm can be improved. Therefore, aiming at the task of object recognition

under resource constrained condition, we optimize the design method of the network structure such as convolution operation, model parameter compression and enhancement of feature expression depth, and design and implement the lightweight network model structure for embedded platform. So that the deep network model can reduce the parameters of the model and the resources needed for operation under the condition of guaranteeing the precision. The experimental results show that the lightweight deep model has better performance than that of the basic model proposed by ResNet, and it can achieve the model accuracy of 93.5% under the condition that the model with l0× fewer parameters on cervical cells data set.

Methods

Deep network usually optimizes the deep network compression of network model by reducing the number of model parameters and the storage space of single parameter. However, simply deleting the structural parameters of the network model without adjusting and optimizing the structure will lead to a significant decline in the performance, thus affecting the actual use of the network model [23]. The object recognition method concerned in this paper is the first level, i.e. image level object recognition. Based on the idea of champion network model in ILSVRC competitions over the years, and combined with the needs of embedded applications, the whole structure of our proposed network model based on modularization is designed, as shown in Fig. 1.

The deep network model structure proposed in this paper adopts modular design method. By orderly linking each module according to the topological structure, the whole structure of the model is finally formed. The DResNet module structure is shown in Fig. 2. The left one is the same dimension of input and output feature diagrams of modules, while the right one is used when reduce the dimension of input feature.

It can be seen from the schematic diagram of the module structure that when the upper module or data reading layer outputs to the lower module, the original feature channel is divided into two identical branches. One branch uses the shortcut connection idea, which directly acts as the front layer $f_i(x)$of the module output. Since the other branch uses bottleneck idea to compress the feature channel first, the compression ratio$w_s$can be set according to the actual situation.

After bottleneck output, the total number of feature channels is segmented according to the proportional coefficientw_s, and convolution feature extraction operations are carried out through convolution kernels of different scales [24]. In this case, convolution kernels with the size of 1 × 1 and 3 × 3 are used to realize the difference of receptive fields between different feature channels, so as to enhance the feature expression ability of small networks. In addition, the feature diagrams of different convolution kernels are output for splicing operation to make them become a complete feature channel. After that, the convolution operation of channel 1 × 1 achieves feature fusion between different receptive field feature channels.

Because of the large redundancy of deep network, the quality of different feature channels is different after 1 × 1 convolution output. So the convolution operation flow based on the idea of feature channel weighting is adopted here. By transforming a single convolution channel into a real value, the quality of the feature channel is measured. The real value is used as the channel weight to measure the features of each channel and then output the feature $f_{se}(x)$. Finally, the features of $f_i(x)$ and $f_{se}(x)$ are added point by point through the idea of residual network, and the feature information processed by the module is output to the next feature extraction module or model classifier to realize the recognition of the cervical cancer cell.

The proposed model structure is based on ResNet network. It mainly improves from three aspects: optimizing convolution operation, compressing model parameters and enhancing feature expression. Each improvement method will be described in detail in next section.

Lightweight deep model for cervical cancer cell

Optimize convolution operation

The standard convolution operation is to add the output results after calculating the feature channels and convolution kernels of the input layer, and take the results as the feature input of the next layer. The standard convolution formula is shown as follows.

$$g(x,y) = \sum_{s=-a}^{a} \sum_{t=-b}^{b} w(s,t) f(x+s, y+t) \quad (1)$$

where w is the convolution kernel; f is the feature image.

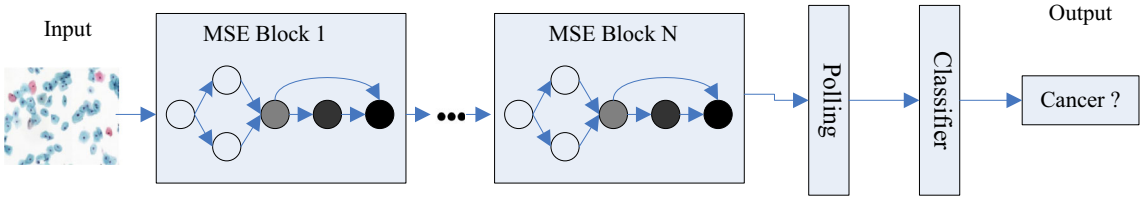

Fig. 1 Our proposed light-weigh deep model structure

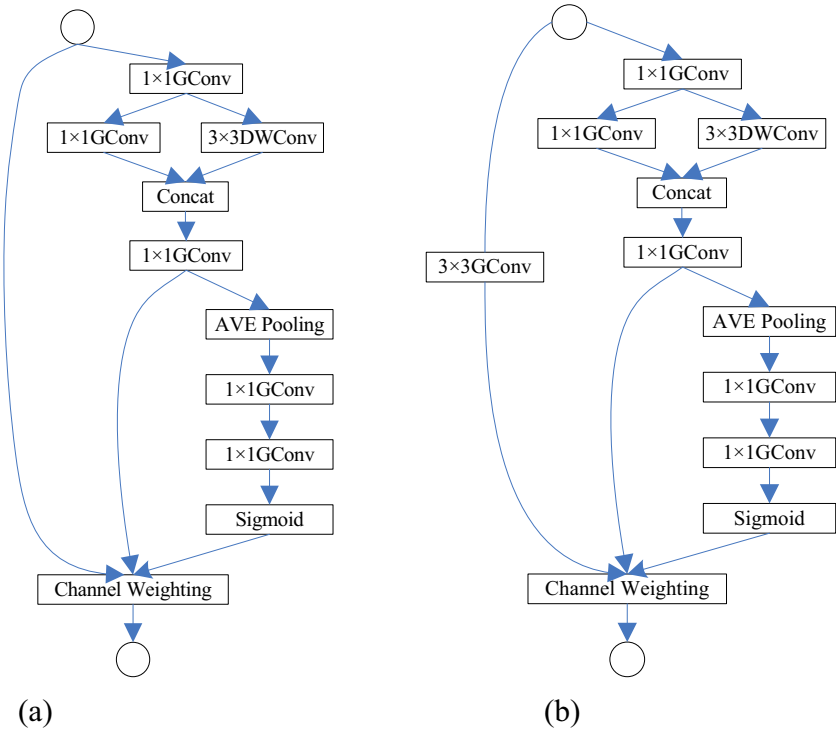

Fig. 2 Diagram of module structure. (**a**) the same dimension of input and output feature; (**b**) Dimensionality reduction

Standard convolution computation can extract image features, but because the number of convolution kernels depends on the input and output channels, there is a strong coupling among them. If it is only from the perspective of input or output channels, it is difficult to achieve better results because of the existence of coupling relationship, so the depthwise separable convolution is used to optimize the convolution operation. The standard convolution operation is divided into Depthwise and Pointwise. And the flow chart is shown in Fig. 3.

In standard convolution operations, it is assumed that the size of the input feature diagram is $M \times D_k \times D_k$, where $M$is the number of input feature and D_k is the size of the input feature. The size of the output feature is $N \times D_f \times D_f$, where N is the number of output feature and$D_f$is the size of the output feature. So the amount of calculation required for standard convolution operation is shown in formula (2).

$$C_{fs} = D_k \times D_k \times M \times N \times D_f \times D_f \tag{2}$$

After depthwise separable convolution, when the sizes of input and output feature diagrams are the same, the calculation can be divided into two parts: deep convolution calculation and point by point convolution calculation. The amount of calculation required is shown in formula (3).

$$C_{fd} = D_k \times D_k \times M \times D_f \times D_f + M \times N \times D_f \times D_f \tag{3}$$

By comparing the standard convolution with the depth separable convolution, it can be found that the depth separable convolution reduces the computational complexity of the

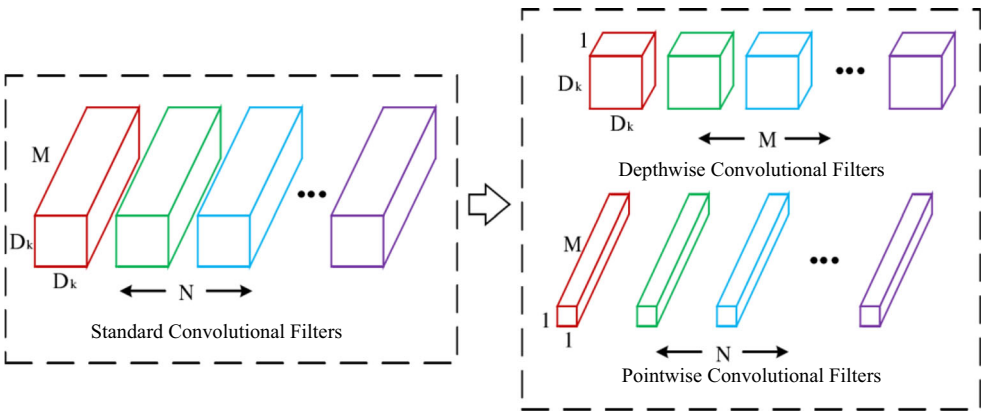

Fig. 3 Depth separable convolution

model by a smaller proportion than the original standard convolution. The calculation formula is shown in formula (4).

$$\gamma = \frac{C_{fd}}{C_{fs}} = \frac{D_k \times D_k \times M \times D_f \times D_f + M \times N \times D_f \times D_f}{D_k \times D_k \times M \times N \times D_f \times D_f} = \frac{1}{N} + \frac{1}{D_k^2} \quad (4)$$

In deep networks, the 1 × 1 convolution carries the information fusion between multi-feature channels. However, because the point convolution operation is also carried out in a traversing way, although the parameters of the model are small, the actual calculation is still relatively large [25]. So this paper is inspired by uses the idea of group convolution, and implements all 1 × 1 convolution operations in the network using group convolution. However, channel mixing operation cannot be adopted after group convolution in this paper. The reasons are written as follows: (1) since the number of channels used in the first three group convolution operations is different, and in the case of different number of channels, the group convolution operation itself has the operation of channel rearrangement, so there will be no edge effect; (2) Mixed channel operation will increase the amount of computation and the memory space used in the network operation. It is contrary to the design requirements of the original hardware resource constraints, and will also lead to the decline of the real-time performance of the model.

Compress model parameters

In order to increase the richness of receptive fields in each layer of the network and reduce the parameters of the model effectively, a model compression method based on Bottleneck's idea is used here. Firstly, the number of feature channels of input 1 × 1 point convolution and 3 × 3 separable convolution is compressed by group convolution, and the compression ratio is controlled by *W*; and then, according to the actual application needs, the number of feature channels of input 1 × 1 and 3 × 3 separable convolutions are set, and the ratio of them is controlled by R. In practical application, the two parameters *W* and *R* can be selected according to the requirements of accuracy and actual hardware. And the schematic diagram is shown in Fig. 4.

Enhance feature expression

Rich receptive fields can improve the performance of the model. The reason is that rich receptive fields guarantee the quality of feature extraction. For the embedded application design network, its features are that the parameters of the model and the hardware requirements for the model operation are relatively low. But this network structure often has a problem, that is, the inadequate feature expression ability of the model. It also leads to the main reason that the performance of the model is not as good as that of the network structure deployed on the server cluster [26]. Therefore, aiming at the weak feature expression ability of lightweight network, the idea of channel weighting proposed by SeNet is used to improve the expressive feature ability of network model. The principle is shown in Fig. 5.

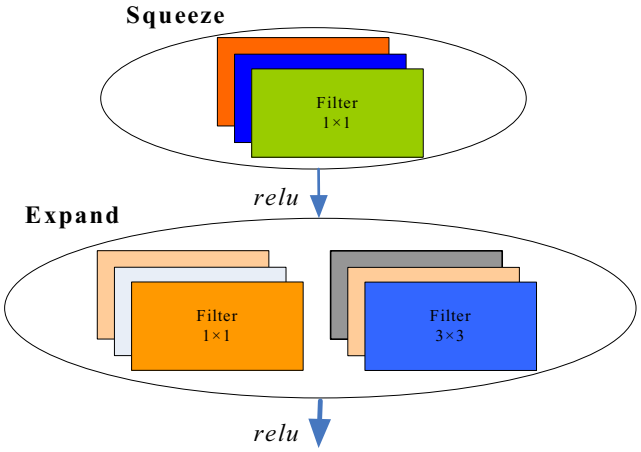

Fig. 4 Model parameter compression method

It can be seen from the schematic diagram that the method is not based on spatial dimension to optimize the model structure, such as Inception, but from the perspective of the relationship between the feature channels to optimize the model structure. By explicitly modeling the interdependence between the channels, the feature response of the channels can be adaptively recalibrated.

Specifically, the original feature channel is first transformed into 1x1xC by global pooling operation; then the proportion of each channel weight is obtained by using the nonlinear activation function; and finally the final feature output is obtained by mapping the proportion to each feature value of the original feature channel. Because the output feature diagram is based on the weighted coefficients corresponding to each feature diagram, it plays an automatic feature differentiation process to a certain extent, and is conducive to the extraction of key features of the object.

The improved loss function

In the data set production and data preprocessing stage, because of the number of samples of different categories, the imbalance of sample proportion may exist in input samples of the classifier.

When a class accounts for a big proportion in the data set, the classifier can often get a good performance. However, the traditional cross-entropy loss function does not distinguish the classes that can be well recognized in the process of model training, which makes the classifier repetitively learn the

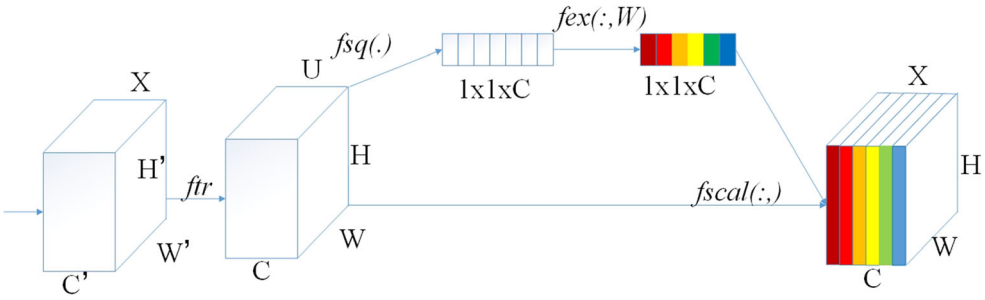

Fig. 5 Channel weighting

samples that have obtained better recognition performance. This also makes it difficult for the classifier to concentrate on the learning of the hard-to-recognize samples, which will increase the training time of the model. Therefore, the focal loss function proposed by RetinaNet based on the idea of hard-to-recognize sample mining is used as the loss function of recognition network training. Focal loss function regards the probability value of each sample class predicted by the network as the weight of loss function. The expression of focal loss function is shown in formula (5).

$$FL(p_t) = -\alpha_r(1-p_r)^{\gamma}\log(p_t) \tag{5}$$

where $(1-p_t)^{\gamma}$is the modulation coefficient, which is used to control the contribution rate of different samples to the loss function; when $\gamma = 0$, it is the standard cross-entropy loss function; and $p_t$is the probability of predicted output sample class in the training process.

Experimental method

Experimental database and environment setup

The experimental environment is divided into two parts: network training and network testing. The network training part is a hardware resource platform based on TITAN X, which uses motherboard XeonES-2450@2.00 GHz CPU with memory 16 GB; the software environment is Ubuntu 14.04 system and the deep learning framework Caffe. The hardware basic platform is based on Jetson TK1 in network testing part, and its display memory is 2 GB; the software environment is Ubuntu 14.04 that is a customized version of Jetson TK1, and the deep learning framework is also Caffe.

The used cell data comes from two datasets with two types of cervical cytology images. The first one is from a publicly available dataset (http://mdelab.aegean.gr/downloads) collected at the Herlev University Hospital by a digital camera and microscope. The image resolution is 0.201 μm per pixel. The specimens are prepared via conventional Pap smear and Pap staining. The Herlev dataset consists of 917 images with ground truth segmentation and classification. There are a total of seven different classes diagnosed by two cyto-technicians and a doctor, so as to maximize certainty of the diagnosis. These seven classes belong to two categories: class 1–3 are normal, and class 4–7 are abnormal. Another database is from self-built database, which is acquired from West China Hospital.

Network model parameter settings

On the basis of the lightweight deep network module proposed above, according to the different hardware environments, the model parameters can be adjusted by setting the values of w and r, where $w_f = 1.0$ and $w_s = 0.5$are used. By connecting each module, the network model structure needed for the experiment is finally formed. Detailed information on model result parameter settings is shown in Table 1 where MB denotes Match_ Block and DB denotes DRes_Block.

After using a common convolution layer, the whole network model directly uses the pooling operation to realize the blurring of the original image and highlight the outline information of the object in the image; then uses four groups of the structure "DB + MB" to extract the high-level semantic feature information on the basis of the previous feature diagram; finally uses the

Table 1 Lightweight model structure parameter setting

Layer	Output size	Kernel size	stride	repeat
image	*224 × 224*	–	–	–
Conv1	*112 × 2112*	*7 × 7*	*2*	*1*
Pool1	*57 × 57*	*3 × 3*	*2*	*1*
MB1	*57 × 57*	*3 × 3*	*1*	*1*
DB1	*57 × 57*	*3 × 3*	*1*	*1*
MB2	*29 × 29*	*3 × 3*	*2*	*1*
DB2	*29 × 29*	*3 × 3*	*1*	*2*
MB3	*15 × 15*	*3 × 3*	*2*	*1*
DB3	*15 × 15*	*3 × 3*	*1*	*4*
MB4	*8 × 8*	*3 × 3*	*2*	*1*
DB4	*8 × 8*	*3 × 3*	*1*	*2*
Pool2	*1 × 1*	–	*global*	*1*

classifier softmax to complete the feature classification and recognize the classification information of the object.

On the basis of completing the model construction, it is necessary to optimize the above network model structure by using reasonable depth network optimization algorithm and appropriate network model training parameters, so as to realize the updating and optimization of the connection parameters between neurons in the model. The parameter setting of model training is shown in Table 2. In this paper, SGD optimization algorithm is used as the calculation method of model optimization.

Experimental results

In order to quickly verify the feasibility of the algorithm and take into account the pertinence of actual cancer-cell recognition applications in cervical cell pathology image, this paper deletes categories from self-built cervical cell dataset. The deleted categories are not related to cervical cell objects, and ultimately the dataset only retains 12 object categories related to cervical cell objects. In order to ensure the fairness of the comparison, the following experimental data are based on the data set. The following experiments are compared in this paper:

a) Comparison between the original ResNet model and our proposed model in this paper.
b) our proposed model is compared with existing lightweight network models such as MobileNet v1, MobileNet v2, ShuffleNet and ResNet-18.

Firstly, our proposed network model is compared with ResNet. And the experimental data are used to verify the performance difference between the deep network structure and the proposed basic network model structure. The comparative experimental data are shown in Table 3 where @.5 means that the channel is cut 0.5 times as much as the original channel.

As can be seen from the Table 3, the size of the original DResNet network model is only 11.2 MBwithout any model tailoring, which is about 1/10 of the size of the ResNet-50 network model. However, the performance of the model is only 0.9% different from that of the ResNet model. Under the condition of further compression, the network model can achieve 7.4% compression of the original structure. In other words, the object recognition accuracy is still 0.5% higher than that of DResNet@.5 under the condition of smaller model size than DResNet@.5. There are also huge differences in the memory usage of the whole model in the running phase. The DResNet network model only needs 533 MB, which is 85% less than the original ResNet network.

Considering the requirements of embedded platform application, this paper makes deep channel-based tailoring for the original DResNet network to test whether the model can meet the application requirements under different hardware resource constraints. As can be seen from Table 4, when the model is tailored to 1.7% of the original model, the performance of the model can still be effectively guaranteed. The object recognition accuracy is similar to that of ShuffleNet model, but the storage space occupied by the whole model is only 1.7 MB.

On the basis of the comparison with ResNet and aiming at the requirements of embedded application, this paper compares the network model with ResNet-18 [12], MobileNet-v1 [18], MobileNet-V2 [22] and ShuffleNet [27]. And the performance comparison curves of those models are shown in Fig. 6.

We can see from the curve that the DResNet network structure has a high convergence speed in the early stage of model training, and its performance is better than other existing network structures designed for embedded applications. At the end of model training, the performance of the four lightweight

Table 2 Model training parameter setting

Parameter	value
Learning rate	0.01
Optimization type	SGD
Learning-rate policy	Multistep
Max Iteration	450 000
Gamma	0.1

Table 3 Model size vs accuracy on cervical cell data set

Model	Size/MB	Compression ratio/%	Accuracy/%	Error/%
ResNet-50	98.5	–	95.4	–
DResNet	11.2	10.4	93.6	−0.9
ResNet-50@.5	65.7	65.9	94.6	−0.7
DResNet@.5	7.4	7.9	93.3	−2.1
ResNet-50@.25	17.4	18.1	93.1	−1.9
DResNet@.25	1.7	2.0	89.9	−5.1

Table 4 Performance comparison of various lightweight network models

Model	Size/MB	Accuracy top1/%	Accuracy top5/%
ResNet-18	44.5	75.4	92.3
ShuffleNet@1 g3	7.2	69.6	90.1
MobileNet	16.7	70.6	92.8
MobileNet v2	17.4	73.3	93.7
DResNet	10.7	71.9	93.5

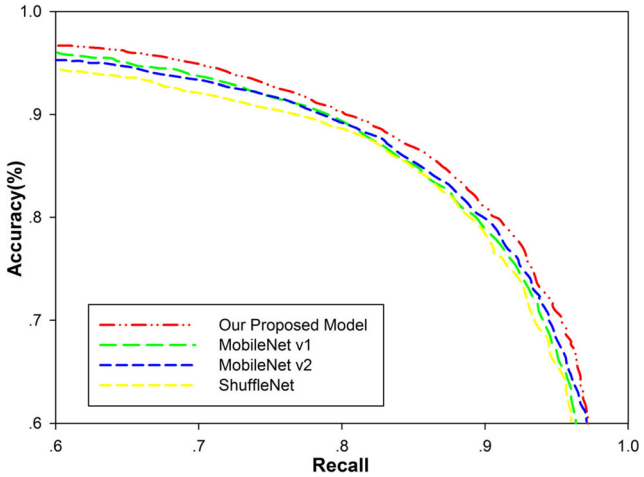

Fig. 6 Performance comparison of several lightweight networks

models converges to approximately equal, but there are still some differences. ShuffleNet network still has a large performance fluctuation at the end of model training, and its performance is slightly lower than that of MobileNet-V2 and MobileNet-V1 networks. The proposed DResNet network structure achieves a slightly higher performance than MobileNet-V1, approximating the accuracy of MobileNet V2 with a difference of only 0.1%. However, the size and convergence speed of the model are obviously better than those of other network models.

Discussion

On the basis of completing the comparison of the whole performance of the model, experiments are carried out to verify the impact of each improvement on the performance of the final model in the process of model design. It can be seen from the table that the algorithm model is only suitable for optimizing the original convolution operation with depth separable convolution. The depth separable convolution adopts the method of dividing the standard convolution into two operations. Although the parameters and calculation amount of the model are reduced to a certain extent, the number of information cross-correlation between each feature channel in the model becomes less. As a result, the performance of the ResNet model is much lower than that of the original model.

By adding squeeze-expand structure, we use the different size of receptive fields generated by different convolution sizes to increase the richness of receptive field sizes in the network model, so that the model can effectively identify the objects of different sizes, and thereby improve the performance of the network. The experimental results show that the performance of the model is improved by 0.8% by adding this method.

The rich receptive field can improve the performance of the model. The reason is that the rich receptive field provides guarantee for the quality of feature extraction. However, if we make full use of high-quality features, whether we can further improve the performance of the network is an argument. Therefore, this paper draws on the Squeeze Excitation network structure based on channel weighting proposed in SE network to enhance the feature expression ability of lightweight model. The experimental results show that the performance of the lightweight model can be greatly improved by this structure, and the lightweight model still has strong feature extraction and expression ability.

In terms of loss function, because the quality of input samples will directly affect the final performance of classifier training, there are many redundant calculations in the training process. The training process has large amounts of data. For example, an image in the data set can be recognized with high confidence in the training process. However, because the network does not have the function of filtering the data set, the model repeats more training of simple samples and cannot focus on learning difficult samples. Therefore, this paper uses the focal loss function proposed in RetinaNet network [28] as the loss function of the classifier, and dynamically adjusts the weight of the loss value through the probability value predicted by the algorithm. It can be ensured that the network is more focused on the learning of difficult samples and the final performance of the model is improved. The experimental results show that the performance of the model is improved by 1.7% through the focal loss function.

Conclusion

In this paper, the lightweight network model structure is designed under the constraints of embedded resources and the DResNet lightweight network structure is proposed. In addition, the actual experimental verification and analysis are carried out for the designed lightweight network structure. Experiments results show that the lightweight network structure can greatly reduce the model parameters on the basis of ensuring the accuracy of the model, thus reducing the computational complexity of model-intensive calculations and the occupation of memory. The model can be effectively deployed with limited hardware and energy resources. However, the proposed lightweight network structure is only for cervical cancer cell classification tasks. In addition, the network compression method only focuses on the optimization of the network model structure. It is not involved in the weight quantization sharing and coding. In future research, it is hoped that the lightweight network structure and medicine diagnosis framework can be integrated effectively and utilize deep network compression methods to achieve cervical cancer cell identification and diagnosis on embedded hardware platforms.

Compliance with ethical standards

Conflict of interest We declare that we have no conflict of interest.

Ethical approval The paper does not contain any studies with human participants or animals performed by any of the authors.

Informed consent Informed consent was obtained from all individual participants included in the study.

References

1. Chiang, Y., Chou, C. Y., Hsu, K. F. et al., EGF upregulates Na+/H+ exchanger NHE1 by post-translational regulation that is important for cervical cancer cell invasiveness [J]. *Journal of Cellular Physiology* 2010, 214(3):810–819.
2. Sokouti, B., Haghipour, S., and Tabrizi, A. D., A framework for diagnosing cervical cancer disease based on feedforward MLP neural network and ThinPrep histopathological cell image features[J]. *Neural Computing and Applications* 24(1):221–232, 2014.
3. Chan, S. W., Leung, K. S., and Wong, W. S., An expert system for the detection of cervical cancer cells using knowledge-based image analyzer [J]. *Artificial Intelligence in Medicine* 8(1):67, 1996.
4. Tang, Z., Wang, S., Huo, J., et al. Bayesian Framework with Non-local and Low-rank Constraint for Image Reconstruction[C]// Journal of Physics Conference Series. 1–12, 2017.
5. Sokouti, B., Haghipour, S., and Tabrizi, A., A pilot study on image analysis techniques for extracting early uterine cervix cancer cell features.[J]. *Journal of Medical Systems* 36(3):1901–1907, 2012.
6. Schilling, T., Miroslaw, L., Glab, G. et al., Towards rapid cervical cancer diagnosis: Automated detection and classification of pathologic cells in phase-contrast images.[J]. *International Journal of Gynecological Cancer* 17(1):118–126, 2010.
7. Nguyen, N. G., Poulsen, R. S., and Louis, C., Some new color features and their application to cervical cell classification[J]. *Pattern Recognition* 16(4):401–411, 1983.
8. Zhao, L., Li, K., Yin, J. et al., Complete three-phase detection framework for identifying abnormal cervical cells[J]. *IET Image Processing* 11(4):258–265, 2017.
9. Bacus, J. W., Cervical cell recognition and morphometric grading by image analysis[J]. *Journal of Cellular Biochemistry* 59(23):12–21, 2010.
10. Tseng, C. J., Lu, C. J., and Chang, C.-C., Application of machine learning to predict the recurrence-proneness for cervical cancer[J]. *Neural Computing & Applications* 24(6):1311–1316, 2014.
11. Long, N. P., Jung, K. H., Yoon, S. J. et al., Systematic assessment of cervical cancer initiation and progression uncovers genetic panels for deep learning-based early diagnosis and proposes novel diagnostic and prognostic biomarkers.[J]. *Oncotarget* 88(65):1094–1136, 2017.
12. Zhang, L., Sonka, M., Lu, L., et al. combining fully convolutional networks and graph-based approach for automated segmentation of cervical cell nuclei[C]// IEEE International Symposium on Biomedical Imaging. IEEE, 112–123, 2017.
13. Zhang, L., Lu, L., Nogues, I., et al. DeepPap: Deep Convolutional Networks for Cervical Cell Classification[J]. IEEE Journal of Biomedical and Health Informatics, 1–12, 2017.
14. Valdes, G., and Interian, Y., Comment on \"deep convolutional neural network with transfer learning for rectum toxicity prediction in cervical cancer radiotherapy: A feasibility study\"[J]. *Physics in Medicine and Biology* 12(1):231–233, 2018.
15. Liu, Y., Zhang, P., Song, Q. et al., Automatic segmentation of cervical nuclei based on deep learning and a conditional random field[J]. *IEEE Access* 24(23):1–17, 2018.
16. Tareef, A., Song, Y., Huang, H. et al., Optimizing the cervix cytological examination based on deep learning and dynamic shape modeling[J]. *Neurocomputing* 33(25):231–237, 2017.
17. Huang, X., Wang, J., Tang, F. et al., Metal artifact reduction on cervical CT images by deep residual learning[J]. *BioMedical Engineering OnLine* 17(1):787–798, 2018.
18. Sornapudi, S., Stanley, R. J., Stoecker, W., et al. Deep Learning Nuclei Detection in Digitized Histology Images by Superpixels[J]. Journal of Pathology Informatics, 2018.
19. Sato, M., Horie, K., Hara, A. et al., Application of deep learning to the classification of images from colposcopy[J]. *Oncology Letters* 12:11):1–11):4, 2018.
20. Song, Y., Zhang, L., Chen, S. et al., Accurate segmentation of cervical cytoplasm and nuclei based on multiscale convolutional network and graph partitioning[J]. *IEEE Transactions on Biomedical Engineering* 62(10):2421–2433, 2015.
21. Phoulady, H. A., Mouton, P. R.. A New Cervical Cytology Dataset for Nucleus Detection and Image Classification (Cervix93) and Methods for Cervical Nucleus Detection[J]. arXiv: Computer Vision and Pattern Recognition, 2018.
22. Bora K, Chowdhury M, Mahanta L B, et al. Pap smear image classification using convolutional neural network[C]// Tenth Indian Conference on Computer Vision. 10287–10296, 2016.
23. Cho, Y. H., and Mangione-Smith, W. H., Deep network packet filter design for reconfigurable devices[J]. *ACM Transactions on Embedded Computing Systems* 7(2):1–26, 2008.
24. Kim, K. H., Hong, S., Roh, B., et al. PVANET: Deep but Lightweight Neural Networks for Real-time Object Detection[J]. arXiv: Computer Vision and Pattern Recognition, 2016.
25. Xia, K., Wang, J., Wu, Y., Robust Alzheimer Disease classification based on Feature Integration Fusion Model for Magnetic[J]Journal of medical imaging and health informatics, 7, 1-6, 2017.
26. Kai-jian Xia, Hong-sheng Yin, Yu-dong Zhang. Deep Semantic Segmentation of Kidney and Space-Occupying Lesion Area Based on SCNN and ResNet Models Combined with SIFT-Flow Algorithm. Journal of medical systems, (2019) 43:2 https://doi.org/10.1007/s10916-018-1116-1.
27. Ma, N., Zhang, X., Zheng, H. T., et al. ShuffleNet V2: Practical Guidelines for Efficient CNN Architecture Design[J]. 31(28): 10392–10402, 2018.
28. Chen T, Lin L, Zuo W, et al. Learning a Wavelet-like Auto-Encoder to Accelerate Deep Neural Networks[J]. 22(12):12–21., 2017