



Urine Sediment Recognition Method Based on Multi-View Deep Residual Learning in Microscopic Image

Xiaohong Zhang¹ · Liqing Jiang¹ · Dongxu Yang¹ · Jinyan Yan¹ · Xinhong Lu¹

Received: 12 July 2019 / Accepted: 10 September 2019 / Published online: 23 October 2019
© Springer Science+Business Media, LLC, part of Springer Nature 2019

Abstract

Urine sediment recognition is attracting growing interest in the field of computer vision. A multi-view urine cell recognition method based on multi-view deep residual learning is proposed to solve some existing problems, such as multi-view cell gray change and cell information loss in the natural state. Firstly, the convolutional network is designed to extract the urine sediment features from different perspectives based on the residual network, and the depth-wise separable convolution is introduced to reduce the network parameters. Secondly, Squeeze-and-Excitation block is embedded to learn feature weights, using feature recalibration to improve network representation, and the robustness of the network is enhanced by adding spatial pyramid pooling. Finally, for further optimizing the recognition results, the Adam with weight decay optimization method is used to accelerate the convergence of the network model. Experiments on self-built urine microscopic image data-set show that our proposed method has state-of-the-art classification accuracy and reduces network computing time.

Keywords Urine sediment · Deep residual · Multi-View learning · Squeeze-and-excitation · Convolutional network · Microscopic image

Introduction

With the development of the economy, people's living standards been improving dramatically. People are also paying more attention to their own health. Human's urine composition can effectively reflect the people's health, among which the type and quantity of the Urinary sediment can effectively reflect the health of the kidney [1, 2].

Automatic urine sediment classifier has significant influence on clinic urine analysis. As compared to traditional manual way which is limited by the technical level, deviation error based on vision and low efficiency, it can relieve the doctors of their hard, time consuming manual work and avoid diagnostic error caused by subjectivism. Moreover, it provides quantitative analysis the trad and high efficiency. Meanwhile, digital results and images are convient for long-distance transfer

which is important for long-distance medical treatment and consultation.

The automatic urine sediment classification model has important significance for clinical urine examination. It can solve the shortcomings of traditional eyepiece examination of urine sediment smear, which is affected by technical level, visual deviation, low work efficiency and unable to give quantitative analysis results [3, 4, 5]. It can also liberate medical personnel from heavy repetitive labor, eliminate all kinds of errors caused by subjective factors, and realize fast and accurate quantitative analysis for urine sediment images, and improve the accuracy of diagnosis. In addition, digitized examination results and images can easily realize the remote transfer of images and data, and provide convenience for telemedicine and disease consultation.

In recent years, according to the theory of digital image processing and pattern recognition, scholars at home and abroad have put forward many automatic analysis models through in-depth research and a large number of experiments, which can provide the basis for diagnosis of various kinds of constituents in digital microscopic images: white blood cells, red blood cells, epithelial cells, tubular, sperm, crystals, fungi and so on. They can be well segmented from the background and are identified by decision tree classifier.

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

✉ Xinhong Lu
hide0717@163.com

¹ Affiliated Hospital of Jining Medical University, Jining No.1 people's hospital, Jining 272000, Shandong, China

Due to the variety of components in the urine sediment, its structure is complex, the background of the image is complicated and unclear, and it is easy to cause interference and misjudgment, which makes the difficulty of automatic identification of urine sediment increase and the accuracy is low [6, 7]. Most traditional algorithms use adaptive threshold segmentation and SVM classification algorithms to achieve statistical classification of various types of urinary sediment cell images. In [8], the adaptive two-dimensional canny double threshold is used to segment the image, and the morphological features are used to make the Epithelial and Urinary cast images of the low-power microscope. The most effective magnification area is obtained by the coordinate tracking of the ow-power microscope. The image is segmented with a fixed threshold to extract the 26-dimensional feature of red blood cells and white blood cells under high-power microscope. After normalization, the SVM classifier is used for training and recognition.

The out-of-focus noise and complex background interference of the urine sediment cell image directly affect the accuracy of the recognition. Most algorithms choose to use Gabor filters to preprocess the image so as to obtain better image quality and enhance the accuracy of recognition. In cell image segmentation, the literature [9] proposed a simple yet effective algorithm to improve image segmentation speed and segmentation effect. On the basis of extracting the characteristic value of urinary sediment formation, the most representative 26-dimensional morphology, statistics and texture features are used to improve the SVM algorithm based on kernel function and parameters, and the SVM classifier with high recognition rate is obtained. Literature [10] proposed a feature fusion based on SVM algorithm in the whole urine sediment image processing system, and gradually improved the correct recognition rate of urine cells. Through the SVM algorithm, when the number of training samples is small, the correct classification effect can be obtained, and the classification and promotion ability can be achieved. In the normalized feature-vector matrix, the cross-validation method is used to select the kernel function and parameters [11, 12]. According to the best 26-dimensional feature-vectors of Red cell and White cell, the classifier is designed and the corresponding confusion matrix is obtained. After classifier training, the classifier with the highest recognition rate is obtained, and then demonstrated the higher accuracy of the classifier.

For urine sediment images, there are various problems such as uneven gray scale and large noise [13, 14]. This paper proposes that image segmentation is regarded as the maximum a posteriori probability estimation, and constructs the Bayesian probability model by constructing the energy function to solve the global minimum value of the energy equation. According to the segmentation result of the image, according to the 17 features after dimensionality reduction, a urine sediment image recognition algorithm based on Bayesian network was proposed. The extracted features were used to construct the

conditional probability table. According to the sample data of each attribute, the conditional probability table of the feature was obtained by constructing a Bayesian network. Bayesian network is used to realize the recognition of urine sediment images. The algorithm has high recognition accuracy, strong robustness and fast recognition speed, and has strong engineering application advantages.

Nowadays, the cell image recognition in the traditional image processing is to artificially design features at first, and then use the classification algorithm of machine learning to achieve identification. For each specific project, it is necessary to design special algorithms for preprocessing, segmentation, feature extraction. In the processes of algorithm design, it is quite easy to cause errors, and hard to achieve the ideal recognition precision, which makes cell image classification a hot and difficult issue in the research. The cell image classification method based on the convolutional neural network avoids the complex feature engineering and preprocessing in traditional methods, and has higher accuracy and robustness. Some scholars applies the convolutional neural network method to the automatic identification of body fluid cells, which has the advantages of simple use and high accuracy. Literature [16] summarizes and analyzes the shortcomings of the improved network model based on LeNet-5 and the improved AlexNet-based network model, and then construct a network structure with the best effect on the recognition of urine microscopic cells. Based on AlexNet network, Literature [17] uses the fractional maximum pooling layer instead of the maximum pooling layer to reduce the speed of feature downsampling by the largest pooled layer, to ensure that the expression of the final abstracted feature is not greatly lost. However, directly using the existing algorithm for cell recognition, although there is a little effect, the generalization of urine microscopic images is not strong. In this paper, we mainly focus on the deep residual network to improve the recognition accuracy and the counting error of red blood cell in urine image captured under microscope.

Urine cell recognition is attracting growing interest in the field of medicine vision. A novel urine cell recognition method based on multi-view deep residual learning model is proposed to solve some existing problems, such as multi-view cell change and cell information loss in the natural state. Firstly, the convolutional network is designed to extract the urine cell features from different perspectives based on the residual network, and depthwise separable convolution is introduced to reduce the network parameters. Secondly, Squeeze-and-Excitation block is embedded to learn feature weights, using feature re-calibration to improve network representation, and the robustness of the network is enhanced by adding spatial pyramid pooling. Finally, for further optimizing the recognition results, the Adam with weight decay optimization method is used to accelerate the convergence of the network model. Experiments on the collected urine

microscopic database show that our proposed method has state-of-the-art classification accuracy and reduces network computing time.

Related works

Residual learning

It is assumed that $H(x)$ is a base mapping that is constructed by several stacked layers, and x is used to represent the inputs of these first layers. Assuming that multiple nonlinear layers can approximate complex functions, it is equivalent to assuming that they can approximate residual functions, such as $H(x) - x$ (assuming inputs and outputs are on the same scale). So we make it very clear that these layers are approximated by the residual function $F(x) := H(x) - x$, rather than expecting the stacked layers to approximate $H(x)$. So the original function becomes $F(x) + x$. Although both forms can approximate the expectation function (hypothesis), its learning difficulty may be different [18].

The new idea is motivated by an abnormal drop of accuracy in deep model. As discussed in [21], if the added layer can be constructed as an identity mapping, then the training error of a deeper model should not be greater than the training error of the corresponding shallower model. The problem of reduced accuracy shows that the solver has difficulty in approximating the identity mapping through multiple nonlinear layers. With residual learning reconstruction, if the identity mapping is the best method, the solver can simply drive the weights of multiple nonlinear layers toward zero so as to approximate the identity mapping.

In reality, an identity map cannot be optimal, but the strategy may help to deal with the problem in advance. If the optimal function is closer to the identity function than to the zero map, the solver is more likely to find interference with the identity mapping than to learn a new function. We have experimentally shown that the residual function we have learned generally has a small response, which indicates that the identity mapping provides reasonable preconditioning.

Identical mapping of shortcuts

The residual learning is adopted for each stacking layer. A building block is shown in Fig. 1. Formally speaking, the building block is defined as:

$$y = F(x, \{W_i\}) + x \tag{1}$$

where x and y are the input and output vectors of the considered layers. Function $F(x, \{W_i\})$ represents the residual function of the learning. As shown in Fig. 1, there are two layers, where σ represents ReLU in $F = W_2\sigma(W_1x)$. To simplify the

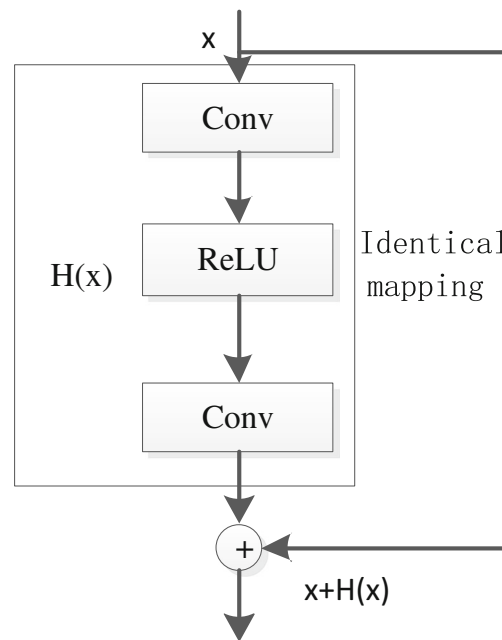


Fig. 1 Identical mapping

comment, we ignore the offset. The operation $F + x$ is done by a shortcut connection and an element-by-element addition. After the increase, a second nonlinear characteristic is adopted in model.

The shortcut connection described in eq. (1) does not introduce additional parameters and complicated calculations. Not only is this attractive in practice, it is equally important in comparing normal and residual networks. With the same number of parameters, depth, width, and computational cost (except for negligible element-by-element additions) [19, 20], we can make a simple comparison of normal and residual networks.

In the formula (1), the sizes of x and F must be the same. If it is different (for example, changing the input and output channels) we can match the dimensions by quickly connecting the linear projection W_s :

$$y = F(x, \{W_i\}) + W_s x \tag{2}$$

We can also use a square matrix W_s in eq. (1). However, we will show through experiments that the identity mapping is sufficient to solve the problem of accuracy degradation and is very cost-effective, so it is only used when matching dimensions for W_s .

The form of the residual function F is flexible, and the experiments in Resnet involve a function F with three layers or more. But if there is only a single layer in F , eq. (1) is similar to the linear layer, so its advantages is very obvious.

It is also noted that although the above symbols are intended to simplify the representation of layers that are fully connected, they apply to convolutional layers. A function $F(x, \{W_i\})$ can represent multiple convolutional layers. The

elements that are added one by one are performed by channel-to-channel on the two feature maps.

Multi-view residual network architecture

In this paper, a multi-view residual network is constructed on the basis of the residual network, and the network is used to extract features on microscopic images from different perspectives. The deep separable convolutional network is combined with the local residual network structure to speed up the model convergence and improve the recognition accuracy. Using the Squeeze-and-Excitation blocks to “screen” the effective feature mechanism, a powerful deep saliency feature in the microscopic image is obtained. Spatial pyramid pooling strategy is added to improve the robustness of the network model. The AdamW optimizer is adopted to increase training speed.

Deep separable convolution and residual network

Generally speaking, as the number of layers of the convolutional network increases, more cell feature information can be extracted, but the accumulation of such a simple network is prone to gradient vanishing and gradient explosion problems, resulting in a decrease in recognition rate. Furthermore, although the depth of the layer can improve the network reconstruction performance, a large number of network training parameters are added, which increases the difficulty of network training [22]. Therefore, this paper proposes a method of combining depth separable convolution and residual networks to solve that it is difficult to converge when the depth of the network is deepened. Among them, deep separable convolution is a form of decomposition of standard convolution. In standard convolution, each input channel must be convolved with a particular kernel, and the result is the sum of the convolution results from all channels. In the case of a depth separable convolution, the deep convolution firstly performs a convolution on each input channel. Furthermore, convolution is performed point by point. Compared with the standard convolution, this convolution structure can greatly reduce the number of parameters and the amount of calculation of the network model without causing significant loss of accuracy.

As shown in Fig. 2, it is assumed that the size of the input feature map is $M \times M \times N$ and the kernel size is $K \times K \times N \times P$; the number of weights required for standard convolution is $W_{SC} = K \times K \times N \times P$ in the case of a step size of 1; and the corresponding number of operations is $Q_{SC} = M \times M \times K \times K \times N \times P$; the total number of weights is $W_{DSC} = K \times K \times N + N \times P$ in the case of deep separable convolution; so the total number of operations is $O_{DSC} = M \times M \times K \times K \times N + M \times M \times N \times P$. Therefore, the amount of the weight and operation reduction calculations can be written as follows:

$$F_W = \frac{W_{DSC}}{W_{SC}} = \frac{1}{P} + \frac{1}{K^2} \quad (3)$$

$$F_O = \frac{O_{DSC}}{O_{SC}} = \frac{1}{P} + \frac{1}{K^2} \quad (4)$$

The deep residual network is shown in Fig. 1. Overlay an identical shortcut connection to speed up network convergence. In the figure, $FI(x)$ is an ideal mapping, $F(x)$ is a residual mapping, and $FI(x) = F(x) + x$. The fitted initial objective function $H(x)$ is transformed into a superposition of the fitted residual map $F(x)$ and the input, so that the direct mapping problem can be transformed into a residual mapping problem.

Squeeze-and-excitation blocks

In order to effectively extract the deep features in the urine image, the Squeeze-and-Excitation block is introduced to screen useful features to improve the sensitivity of the network to information features. The main idea of the Squeeze-and-Excitation blocks (SE) is to improve the expressiveness of the network by explicitly modeling the interdependencies between convolutional feature channels. The mechanism for calibrating each feature channel enables the network to proceed from global information to enhance valuable feature channels and suppress feature channels that are not useful for current tasks. The schematic diagram of the Squeeze-and-Excitation modules is shown in Fig. 3.

For any given transformation: $F_r: X \rightarrow U$. It is assumed that F_r is a standard convolution operation, $V = [v_1, v_2, \dots, v_C]$ represents a set of learned filters, $U = [u_1, u_2, \dots, u_C]$ is denoted as the output of the convolution, and its formula is written as:

$$u_c = v_c * X = \sum_{s=1}^C v_c^s * x^s \quad (5)$$

where $*$ is expressed as convolution operator, $v_c = [v_c^1, v_c^2, \dots, v_c^C]$, $X = [x^1, x^2, \dots, x^C]$, v_c^s is 2D space kernel, so a single channel represented by v_c acts on the corresponding channel x .

the SE network module is adopted to perform feature recalibration, which can be divided into two steps: squeeze and excitation. Specific steps are as follows:

Step 1 (Squeeze) Feature U first generates channel statistics by using global average pooling and squeezing global spatial information into a channel descriptor, which embeds the global distribution of channel feature responses, so that the subsequent network layer obtains the global receptive field information. Statistics $z \in R^C$ are generated by shrinking U in the spatial dimension $W \times H$, where the first c element of z is calculated as:

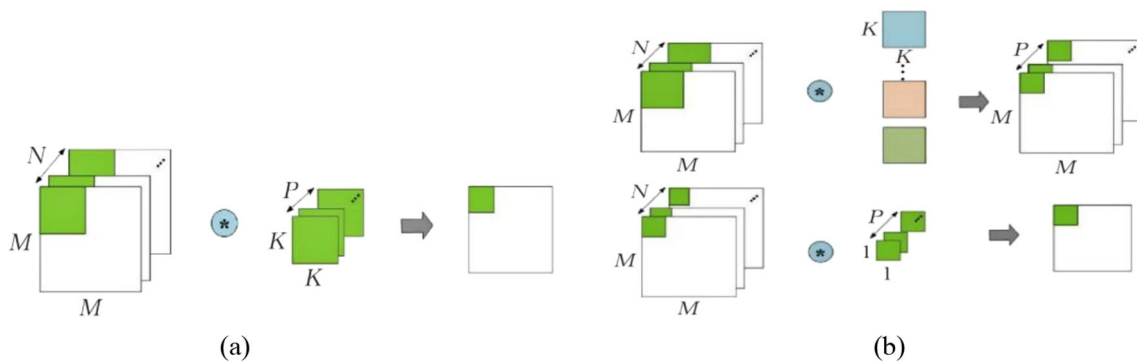


Fig. 2 Comparison of different convolution model.(a) standard convolution;(b) deep separable convolution

$$z_c = F_{sq}(u_c) \frac{1}{W \times H} \sum_{i=1}^W \sum_{j=1}^H u_c(i, j) \tag{6}$$

Step 2 (Excitation) In order to make full use of the channel aggregation information of the previous stage, the dependence of each channel information is obtained. The excitation of each channel is controlled by learning the activation of specific samples for each channel through a channel-dependent filtering mechanism. The feature mapping U is then re-weighted to generate the output of the SE block, which can then be entered directly into subsequent layers. The excitation operation consists of two fully connected layers with two active layer operations. The specific formula is as follows:

$$s = F_{ex}(z, W) = \sigma(W_2 \sigma(W_1 z)) \tag{7}$$

where δ and σ are the activation functions ReLU and Sigmoid, respectively, the dimension reduction layer is $W_1 \in R^{C \times C}$, the dimension reduction ratio is r (set to 16) and the ascending dimension layer is $W_2 \in R^{C \times C}$. The final output is obtained by re-adjusting the transformed output U with activation, which is written as follows

$$x_c = F_{scale}(u_c, s_c) = u_c \cdot s_c \tag{8}$$

where $X = [x_1, x_2, \dots, x_C]$ and $F_{scale}(u_c, s_c)$ refer to the corresponding channel product between the feature map $u_c \in R^{W \times H}$ and the scalar s .

For the partial missing and redundant feature of multi-view cell images in urine microscope image, the SE module seeks to calculate the weight of the output convolution channel, which makes the network more efficient by emphasizing the useless features between important cell features and suppression background channels. The SE module can be used as a direct replacement for raw blocks of any deep architecture, seamlessly integrated into any CNN model. In this paper, the SE block is embedded in the convolution network multiple times to extract the salient features of Erythrocyte and Leukocyte, and the multi-view cell recognition rate is improved. In addition, although the SE module increases the number of network parameters, it does not affect the calculation speed too much.

Multi-view cell identification network structure

This paper proposes a multi-view cell recognition network for cell recognition of urine microscopic images. The network structure is shown in Fig. 3. There are 17 convolution layers in the network: 2 two-dimensional convolutional layers (Conv2D) and 15 deep separable convolutional layers (SeparableConv2D). Each convolution layer is followed by a Batch Normalization function and a ReLU activation function. Among them, two Conv2D layers are set with 8 sets of 3x3 convolution kernels. In addition, five short-cut connection modules are set in the network, and each shortcut connection module consists of three deep separable convolutional layers. It is composed of a shortcut connection to speed up

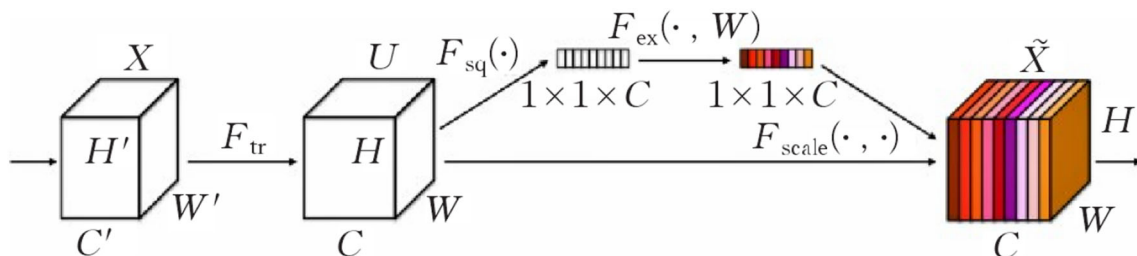


Fig. 3 Squeeze-and-Excitation Modules

network convergence while preventing the gradient disappearing. Each shortcut connection module is followed by a SE module that enhances the network’s generalization capabilities by enabling the network to perform dynamic channel feature recalibration. After the last convolutional layer, a Spatial Pyramid Pooling (SPP) layer is introduced to remove the restriction on the fixed size of the network. The SPP layer pools the features and produces a fixed-length output. Give the full connection layer which can avoid cropping the network at the very beginning. It can not only input information of any size, but also improve the accuracy and reduce the overall training time. The last layer of the fully connected layer uses the Softmax as activation function. In the multi-classification problem, k kinds of possibilities can be predicted (k is the number of kinds of sample classes). it is assumed that the input feature $isx^{(i)} \in R^{n+1}$, The sample is labeled as $y^{(i)}$, That is to say, the training set can be formed as $S = \{(x^{(1)}, y^{(1)}), (x^{(2)}, y^{(2)}), \dots, (x^{(m)}, y^{(m)})\}$. Function $h_{\theta}(x)$ and cost function $J(\theta)$ can be written as follows,

$$h_{\theta}(x^{(i)}) = \begin{bmatrix} P(y^{(i)} = 1|x^{(i)}; \theta) \\ P(y^{(i)} = 2|x^{(i)}; \theta) \\ \vdots \\ P(y^{(i)} = k|x^{(i)}; \theta) \end{bmatrix} = \frac{1}{\sum_{j=1}^k e^{\theta_j^T x^{(i)}}} \begin{bmatrix} e^{\theta_1^T x^{(i)}} \\ e^{\theta_2^T x^{(i)}} \\ \vdots \\ e^{\theta_k^T x^{(i)}} \end{bmatrix} \quad (9)$$

where $\theta_1, \theta_2, \dots, \theta_k \in R^{n+1}$ are parameters of our proposed model, $\frac{1}{\sum_{j=1}^k e^{\theta_j^T x^{(i)}}}$ is the normalization of the probability distribution, so that the sum of all the probabilities is 1.

$$J(\theta) = -\frac{1}{m} \left(\sum_{i=1}^m \sum_{j=1}^k P\{y^{(i)} = j\} \ln \frac{e^{\theta_j^T x^{(i)}}}{\sum_{l=1}^k e^{\theta_l^T x^{(i)}}} \right) \quad (10)$$

Network training

Deep convolutional neural network training is to optimize the network through the back propagation algorithm (BP) to get the optimal parameters. At present, the Adam algorithm is the mainstream method in practical applications. It has excellent convergence and its adaptive learning rate is more advantageous than the fixed or exponential decay learning rate of the stochastic gradient descent algorithm (SGD). However, the generalization performance is worse than the SGD method with Momentum in some data sets, and it is easy to converge to a less than ideal minimum. If the SGD method with the momentum is adopted, the gradient updating inflexibility problem will occur. In order to improve the generalization ability of the network and reduce the network training time, this paper uses an adaptive gradient descent with weight

attenuation (AdamW) to improve the generalization performance of the network. It is assumed that samples $\{x^{(1)}, x^{(2)}, \dots, x^{(m)}, \}$ are randomly selected from the training set $y^{(i)}$, where $y^{(i)}$ is the actual value corresponding to the sample $x^{(i)}$. Therefore, we firstly calculate the average gradient of m samples:

$$g_t = \frac{1}{m} \nabla_{\theta_{t-1}} \sum_i L(f_t(x^{(i)}; \theta_{t-1}), y^{(i)}) + w \theta_{t-1} \quad (11)$$

The first moment estimate and the second moment estimate of the gradient are written on basis of eq. (11)

$$m_t = \beta_1 m_{t-1} + (1-\beta_1) g_t \quad (12)$$

$$v_t = \beta_2 v_{t-1} + (1-\beta_2) g_t^2 \quad (13)$$

The update equation for the model parameters can be expressed as

$$\Delta \theta_t = -\eta \frac{m_t}{\sqrt{v_t} + \epsilon} + \omega \theta_{t-1} \quad (14)$$

where the deviation corrections of the first moment and the second moment are $m_t = \frac{m_t}{1-\beta_1}$ and $v_t = \frac{v_t}{1-\beta_2}$, respectively; learning rate is η ; the exponential decay rates of the first moment and the second moment estimate are $\beta_1 = 0.9$ and $\beta_2 = 0.999$; weight decay is $\omega = \omega_{norm} \sqrt{\frac{b}{B \cdot T}}$, where b is the batch size, the total number of training samples per batch is B , and the number of training is T , $\omega_{norm} = 0.005$.

Experimental results and analysis

Comparison algorithm and feature analysis

In this paper, the urine sediment image taken by the NIKON microscope under a 100x objective lens has an image size of 64×480. The target composition region in the 1550 images are selected for training. All the images are marked by the graduate students of the Department of Urology in Chongqing medical university and it was confirmed by pathologists. Since there are many experimental data, we use grouping strategy to make independent experiments. All data are divided into 5 groups: A, B, C, D and E. This paper proposes a multi-view residual network cell recognition algorithm for urine microscopic images, its hardware and software platform: CPU: Intel (R) Core (TM) i7–8700 CPU @ 3.20GH; GPU: NVIDIA GeForce GTX 1070Ti; operation System: ubuntu 16.04LTS; deep learning framework: TensorFlow. In order to verify the effectiveness of the proposed cell recognition detection algorithm, the current optimal algorithm is used as

a comparison algorithm, which is DenseNet network,^[15] SSD network^[16] and ResNet network.^[17] For the classification results, the accuracy (ACC), sensitivity (SENS), specificity (SPEC) and the area under ROC curve were used to evaluate the classification results.

The traditional method for identifying urine sediment image is to divide the input image into small patches containing only individual components. The segmented small region is the target image that we need to extract the features. The used features are about shape size, gray value information, texture information, etc. of the component. Finally, according to the extracted features, they are input to the recognition classifier for urine sediment image recognition, where the feature vector is used as a node, and constructing a conditional probability table to form a network of discriminant classifiers to obtain the result of urine sediment image recognition. This paper is a deep learning-based recognition algorithm, which uses the residual network to extract the deep features of the microscopic image of the urine, and then the deep network classifier for identification.

Urinary microscopic image target recognition was performed due to the advanced deep features used in this paper. In order to analyze the validity of this paper, we also use traditional machine learning algorithms for comparison. In conclusion, the selected traditional machine learning method is named as low-level feature cascade (LLFC). If we want to identify different kinds of regions in machine learning, we first need to determine the most representative features of these objects, which can not only represent this kind of things, but also be as different as possible from other things. This is the pre-processing step of the identification work. In the process of pre-processing, we first determine which features can represent the particularity of a certain class, and can also be converted into representative parameter types. We often cannot determine the category of a thing by a certain feature, so we need to collect enough features to perform subsequent classification work, and these multiple features can be combined into one feature vector to represent. The selection of features is particularly important because the recognition of objects can only have a good performance when the correct feature information is selected. The image recognition requires some related properties of the observed image as features, such as brightness, texture and target edge contour shape; the red and white cells in the urine sediment image have obvious characteristics, including tubular cells, epithelium, Crystallization is not easily distinguished and identified. The so-called focus on these three particles has been studied. In order to ensure the recognition effect on tubular cells, epithelium and crystallization, some attributes such as the original aspect ratio of the pixel in the target area are gradually added

during the actual experiment. A total of 33 features can be seen. The feature vector of the feature composition is sufficient for identifying tubular cells, epithelium, and crystallization.

Parameter setup

The designed network in this paper consists of 5 convolutional layers and 6 pooling layers and 3 fully connected layers; the first convolutional layer convolution kernel size is 11×11 with steps 4 and contains 96 convolution kernels; The convolutional layer convolution kernel has a size of 5×5 with step 1 contains 256 convolution kernels; the third convolutional layer convolution kernel has a size of 3×3 with step 1 contains 384 convolution kernels; The convolutional layer convolution kernel has a size of 3×3 with step 1 contains 256 convolution kernels; the 5th convolutional layer convolution kernel has a size of 3×3 with step 1 contains 256 convolution kernels; The pooling layer pooling window size is 3×3 , the step is 2, and the maximum pooling model is adopted; the second pooling layer size is 3×3 , the step size is 2, and the maximum pooling mode is adopted; The pooling layer pooling window size is 3×3 , the step size is 2, and the maximum pooling mode is adopted; the fourth pooling layer pooling window size is 3×3 , the step size is 1, and the maximum pooling mode is adopted; The pooling layer pooling window size is 3×3 , the step size is 2, and the average pooling mode is adopted; the sixth pooling layer pooling window size is 3×3 , the step size is 2, and the average pooling mode is adopted; The feature map of the activation function ReLU is output as a multi-scale feature. The feature maps selected by the second convolutional layer, the fourth convolutional layer, and the fifth convolutional layer are respectively pooled, and the feature fusion is performed through a fully connected layer, and finally the Softmax is input layers to classify and identify the cell feature. The convolutional layer learning rate is set to 0.001, and the fully connected layer learning rate is 0.01, attenuation is 10 for 10 times per iteration. The attenuation weight and the momentum are set from 0.001 to 0.8, and the discard rate of the fully connected layer is 0.5.

Qualitative and quantitative analysis

Figure 4 shows the fine-tuning process for the identification of urine microscopic image cells on the self-built dataset using the multi-view residual model proposed in this paper (using four-fold cross-validation, 100 rounds per iteration), Fig. 4 The fine-tuning iterative process with the original image in the data subset as the verification data is shown separately. As the number of iterations

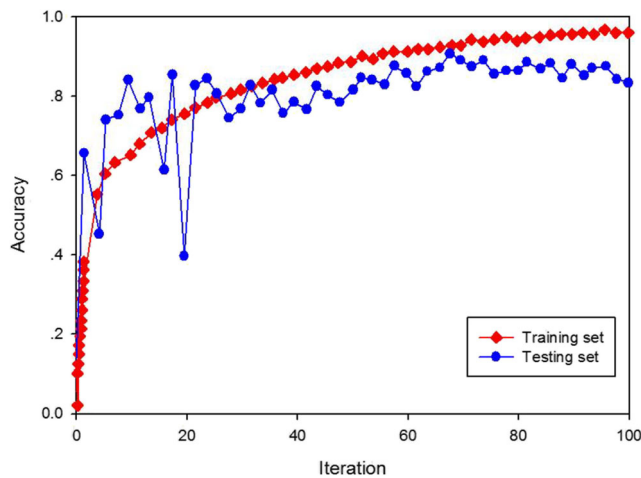


Fig. 4 Training process of improved multi-view residual network (four-fold cross validation)

increases, the accuracy of the training set gradually increases. The accuracy of the verification set will oscillate over a large range when the model begins to iterate, and as the iteration progresses, the amplitude of the oscillation decreases. When the four-fold cross-validation experiment is carried out, the maximum accuracy for the improved model is 0.907, 0.911, 0.913 and 0.905 after the completion of the 78th, 72nd, 76th, and 82nd rounds. With subset C as the verification set, the improved residual model with an accuracy of 0.9133 is used as the final semantic recognition model. According to the experimental results, the improved residual model can better complete the detection and identification of cell regions. Table 1 shows the comparison of recognition accuracy under different data sets. It can be seen that the recognition accuracy of the deep learning algorithm is much larger than the traditional artificial feature method. In the deep learning algorithms, our proposed algorithm in this paper is also optimal. Table 2 is the performance comparison for different methods. It is indicated that our proposed algorithm is robust to microscopic urine image, especially on the simple background image.

It can be seen from Fig. 5 that there are a large number of impurities with different shapes and colors in the urine sediment

Table 1 Comparison of recognition accuracy under different data sets

Dataset	SDD	DenseNet	ResNet	LLFC	Proposed
A	74.11%	85.47%	87.22%	81.12%	91.26%
B	61.11%	66.93%	71.09%	70.15%	78.35%
C	69.12%	78.22%	81.25%	66.89%	88.28%
D	68.28%	85.28%	88.20%	75.38%	89.37%
E	51.91%	52.90%	59.75%	58.96%	59.78%

Table 2 Performance comparison for different methods

Algorithms	ACC (%)	SENS (%)	SPEC (%)	AUC (%)
SDD	95.12	97.36	92.18	97.2
DenseNet	96.29	98.10	93.9	96.5
ResNet	93.08	93.85	91.15	96.2
LLFC	92.06	92.13	84.56	94.1
Proposed	97.15	98.73	95.85	97.7

images. When we use the multi-view residual network to identify the components of the image, the qualitative results show that the model can detect the obvious cells. It can identify some impurities with a larger area than the cell image. However, the trained samples in this paper are marked as background for some interference targets. This shows that the algorithm has a higher accuracy. If the number of sample are increased and cover all the interference targets, the recognition result will be much higher. When using DenseNet network for component recognition, the algorithm finds as many targets as possible, but it has poor effect on some adhesion targets. The main reason is that the algorithm is only a single-view learning, and can not learn from the characteristic information of multiple angles. When the impurities in the image are close to the color of the cell, the SSD algorithm will recognize the impurity as a cell. Moreover, due to the uneven color inside the cell, a misclassification occurs inside the cell during recognition.

Urine sediment testing is a routine testing program in hospitals. By combining the deep model and constructing the composition detection model with multi-view residual features, this paper has the following advantages: (1) high recognition accuracy, combined with multi-view characteristics, can better detect cells from urine images; (2) The results can be quantitatively analyzed. The traditional algorithm can only qualitatively judge the quality of the segmentation from the final segmentation image, but the component detection and recognition model can quantitatively analyze the recognition result according to the precision accuracy, sensitivity and specificity. Based on these characteristics, our proposed algorithm is more suitable for use on the urine sediment automatic recognition system, which has great application prospects.

Conclusion

Urine sediment testing is a routine testing program in hospitals. Traditional testing methods rely on medical personnel to operate microscopes to manually identify the components contained in urine sediment. This manual method consumes a lot of manpower, so it is necessary to develop a more automated urine sediment detection instrumentation, and the identification of urine sediment in the automatic test of urine

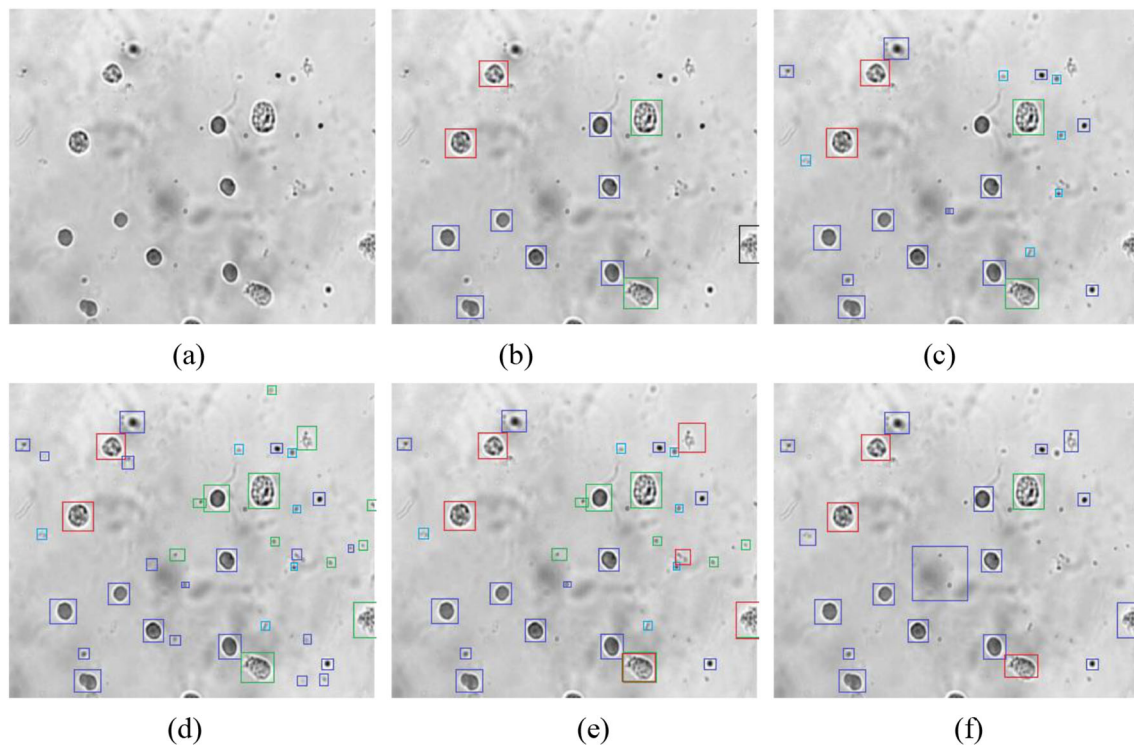


Fig. 5 Qualitative analysis of different algorithms; **a** Original image; **b** SDD; **c** Proposed; **d** DenseNet; **e** ResNet; **f** LLFC

sediment is a difficult problem. How to detect images more accurately is a key step for automatic identification of urine sediment. By combining the deep model and constructing the composition detection model with multi-view residual features, a multi-view urine cell recognition method based on multi-View deep residual learning is proposed to solve some existing problems, such as multi-view cell gray change and cell information loss in the natural state. Firstly, the convolutional network is designed to extract the urine sediment features from different perspectives based on the residual network, and depth-wise separable convolution is introduced to reduce the network parameters. Secondly, Squeeze-and-Excitation block is embedded to learn feature weights, using feature re-calibration to improve network representation, and the robustness of the network is enhanced by adding spatial pyramid pooling. Finally, for further optimizing the recognition results, the Adam with weight decay optimization method is used to accelerate the convergence of the network model. Experiments on self-built urine microscopic image data-set show that our proposed method has state-of-the-art classification accuracy and reduces network computing time.

Acknowledgements The fund is from the Affiliated Hospital of Jining Medical University, Clinical significance of urine protein detection in pregnant women during different pregnancy and correlatin with Physiological indicators during pregnancy (No. MP-2016-011).

Compliance with ethical standards

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

Human and animal rights 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 paper.

References

1. Zhou, X., Xiao, X., Ma, C., A study of automatic recognition and counting system of urine-sediment visual components. *International Conference on Biomedical Engineering & Informatics*. IEEE, 2010. 256–268.
2. Li, C. Y., Fang, B., Wang, Y., et al. Automatic Detecting And Recognition Of Casts In Urine Sediment Images. *International Conference on Wavelet Analysis & Pattern Recognition*. IEEE, 2009:036–045.
3. Mei-Li, S., Rui, Z., Urine Sediment Recognition Method Based on SVM and AdaBoost. *International Conference on Computational Intelligence & Software Engineering*. IEEE, 2009:1286–1296.
4. Fu, C., Xia, S. R., and Zhang, Z. C., [The study of SVM-based recognition of particles in urine sediment]. *Chinese Journal of Medical Instrumentation* 32(6):409–412, 2008.
5. Zaman, Z., Fogazzi, G. B., Garigali, G. et al., Urine sediment analysis: Analytical and diagnostic performance of sediMAX? — A new automated microscopy image-based urine sediment analyser. *Clin. Chim. Acta* 411(3–4):140–154, 2010.

6. Luo, H., Ma, S., Wu, D. et al., Mumford-Shah Segmentation for Microscopic Image of the Urinary Sediment. 2007 1st International Conference on Bioinformatics and Biomedical Engineering. IEEE: 112–118, 2007.
7. Canlong, Z., Yanping, T., Qiang, W. et al., A more effective algorithm of automatic recognition urinary sediment. *Computer Engineering & Applications* 46(3):232–235, 2010.
8. Yong, Y., and Ping, L., Segmentation of urine sediment image based on improved Canny operator. *Computer Engineering and Applications* 25(8):253–266, 2010.
9. Zhou, Y., and Zhou, H., Automatic Classification and Recognition of Particles in Urinary Sediment Images. *Lecture Notes in Electrical Engineering* 107(226):1071–1078, 2012.
10. Shen, M. L., and Chen, D. R., Study on urinary sediments classification and identification techniques. *Proceedings of SPIE - The International Society for Optical Engineering* 2006:25–30.
11. Yixiong, L., Zhihong, T., Meng, Y. et al., Object detection based on deep learning for urine sediment examination. *Biocybernetics and Biomedical Engineering* 38(3):661–670, 2018.
12. Yu, H., Jing, W., Iriya, R. et al., Phenotypic antimicrobial susceptibility testing with deep learning video microscopy. *Anal. Chem.* 12(7):1128, 2018.
13. Valenzuela, R., Momingstar, W. A., and Makker, S. P., The renal pathology of chediak-higashi disease: Usefulness of the urinary sediment as a confirmatory diagnostic test. *Hum. Pathol.* 8(2): 230–232, 1977.
14. Yixiong, L., Rui, K., Chunyan, L. et al., An End-to-End System for Automatic Urinary Particle Recognition with Convolutional Neural Network. *J. Med. Syst.* 42(9):165, 2018.
15. Kang, R., Liang, Y., Lian, C. et al., CNN-Based Automatic Urinary Particles Recognition. 34(32):129–141, 2018.
16. Hans, C., Merchant, F. A., Shah S, K., Decision fusion for urine particle classification in multispectral images. the Seventh Indian Conference on Computer Vision, Graphics and Image Processing, 2010. 419–426.
17. Brock, D. A., and Hundley, J. M., Identifying Calcium Oxalate Crystals in Urine. *Lab. Med.* 26(11):733–735, 1995.
18. Delanghe, J., New screening diagnostic techniques in urinalysis. *Acta Clin. Belg.* 62(3):155–161, 2007.
19. Targ, S., Almeida, D., Lyman, K., Resnet in Resnet: generalizing residual architectures. arXiv preprint arXiv:1603.08029, 2016.
20. Sünderhauf N, Shirazi S, Dayoub F, et al. On the performance of convention features for place recognition. *Intelligent Robots and Systems (IROS), 2015 IEEE/RSJ International Conference on. IEEE, 2015: 4297–4304.*
21. Redmon, J., Divvala, S., Girshick, R. et al., You only look once: Unified, real-time object detection. *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.:*779–788, 2016.
22. Ren, S., He, K., Girshick, R. et al., Faster r-cnn: Towards real-time object detection with region proposal networks. *Adv. Neural Inf. Proces. Syst.:*91–99, 2015.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.