Urine red blood cell classification based on Siamese Network

Xue Zhao\textsuperscript{1,a}, Jianhong Xiang\textsuperscript{1,b*} and Qingbo Ji\textsuperscript{1}

\textsuperscript{1}Department of information and communication engineering, Harbin Engineering University, Harbin, 150000, China
\textsuperscript{a}email: zxue12012021@163.com, \textsuperscript{b}email: xiangjianhong@hrbeu.edu.cn,
\textsuperscript{*}Corresponding author’s e-mail: xiangjianhong@hrbeu.edu.cn

Abstract. Urine examination is an important examination commonly used in medical in vitro examination. The morphological classification of red blood cells in urine plays an important role in the diagnosis of hematuria and renal diseases. In order to improve the accuracy and efficiency of urine red blood cell classification, a urinary red blood cell classification algorithm based on Siamese Network of dual-mode contrastive loss function was proposed. In the dual-mode contrastive loss function, the Cosine Similarity measure and a weight factor were added as the constraint of Euclidean distance in the Contrastive Loss function, which improves the similarity judgment ability of the Contrastive Loss function for the two input eigenvectors. The experimental results show that the Cosine Similarity measure has a positive effect on the Contrastive Loss function, and can effectively improve the accuracy of the whole classification model.

1. Introduction

With the advancement of science and technology, people's demand for a healthy life is becoming stronger and stronger. The Urinary sediment detection in the physical examination is to determine whether there are abnormal values in various indicators through the analysis and inspection of the formed components in the urine. When the amount of red blood cells in the urine exceeds a limit, it is a "hematuria" phenomenon, resulting in the patient developed nephritis and other diseases. When the amount of red blood cells in the urine exceeds a limit, the phenomenon of "hematuria" appears, causing the patient to develop nephritis and other diseases. The morphology of urine red blood cells can determine the source of hematuria or the cause of urinary tract disease. Renal hematuria is caused by damage to the filtration membrane in the body, which causes red blood cells to be squeezed and deformed when passing through the glomerular filtration membrane, leading to distortion of red blood cells in the urine; the morphology of urinary red blood cells of non-renal hematuria is the same as that of blood mature red blood cells.

The classification method of urinary erythrocyte image belongs to the fine-grained image classification research. Existing fine-grained image classification methods are mainly divided into strong supervised and weak supervised fine-grained classification algorithm. The strong supervised classification algorithm requires labeling not only the class labels of the data samples, but also the distinct feature areas. A Part-based R-CNN model was proposed, which is slower. A Mask-CNN model was proposed that preserves foreground features and removes background features. The high labeling requirements of strong supervised classification algorithm limit the development of image classification, so the weak supervised classification algorithm was born. Two-level Attention Models classification model was proposed. Hierarchical Bilinear Model was proposed to classify feature maps of different
convolution layers by using bilinear pooling techniques. Because the regional characteristics of urinary erythrocytes are difficult to extract, these algorithms are not suitable for the task of urinary erythrocyte classification.

In order to improve the classification accuracy and efficiency of urine red blood cell image, a dual-mode contrast loss function was proposed. Through adding the Cosine Similarity and a weighting factor as Euclidean distance constraints, the ability of the Contrastive Loss function to judge the similarity of two input feature vectors was improved. Finally, the classification accuracy of urine red blood cell image was promoted.

2. Related theories

Siamese Network is a network architecture built using two neural networks with the same network structure. The neural network can be a convolutional neural network CNN or a cyclic neural network RNN. Siamese refers to the sharing of weights between two branches \[11\]. The Siamese Network structure is shown in Figure 1. The Siamese Network takes two different data samples as input and transmits them to two neural networks respectively, and calculates the similarity of the high-level feature representations output for the two networks. When two neural networks have the same structure and share weights, it is called a narrow Siamese Network while a generalized Siamese Network can be composed of any two neural networks.

![Figure 1 General structure of Siamese Networks](image)

The core content of the Siamese Network is to learn the similarity relationship between two input samples through the same network model. For two input data sample images, using the Siamese Network to determine whether they are from the same class or different classes is a two-class problem.

As a dual-branch network, the Siamese Network is composed of a pair of data sample images. During training, the two branch networks each input a sample image, the parameter weights of each layer in the two branch networks are fully shared, and the two branches extract image features separately to obtain feature vectors.

The characteristics of the Siamese Network determine that it can realize small sample learning and prevent the interference of false samples. Therefore, it is often used in models that value the error tolerance rate, such as face recognition and fingerprint recognition. During training, the Siamese Network often uses the contrastive loss, which makes the distance between the feature vectors of two similar inputs as small as possible, and the distance between dissimilar inputs as large as possible. The contrastive loss function can be expressed as formula (1), and the Euclidean distance between two inputs is expressed as formula (2):

\[
L(W,(Y,X_1,X_2)) = (1-Y)\frac{1}{2}(D_w)^2 + Y\frac{1}{2}\max(0,m-D_w)^2
\]

\[
D_w (X_1,X_2) = \|X_1 - X_2\| = \left(\sum_{i=1}^{P}(X_{i1} - X_{i2})^2\right)^{\frac{1}{2}}
\]

Where: \(W\) is the weight of the Siamese Network model; \(Y\) is the label of the input data sample pair, \(Y=0\) indicates that the input data sample pair \(X_1\) and \(X_2\) belong to the same category, and \(Y=1\) indicates that \(X_1\) and \(X_2\) belong to different categories; \(D_w\) represents the \(L_2\) norm of \(X_1\) and \(X_2\) in the output feature
space, namely the Euclidean distance of the two feature vectors and the expression is as in formula (2); \( m \) represents a threshold called margin, which is the lower bound set by the contrastive loss for the Euclidean distance between different types of data samples.

3. Materials and Methods

3.1. Network model structure

The number of urine red blood cell data sets is small. In this paper, Siamese Network was selected as the network model of fine-grained image classification of urine red blood cells. In order to improve classification accuracy, a new dual-mode contrastive loss function was designed. Firstly, Cosine Similarity and a weight factor were introduced, and Euclidean distance was combined to complete the restriction of label similarity. The urine red blood cell classification model based on Siamese Network is shown in Figure 2.

![Figure 2: Urine red blood cell classification model based on Siamese Network](image)

Considering the small image size and the small number of the urine red blood cell data set, five convolution layers were selected as the feature extraction layer, and two fully connected layers were selected to achieve feature integration. In order to prevent the reduction of feature information, the max-pooling layer was used only once to extract the texture features in the input information. The network model structure design is shown in Figure 2.

The branch network of Siamese Network consists of seven learning layers, five convolution layers and two fully connected layers. A \( 3 \times 3 \) size convolution kernel was used in all convolutional layers. The output of the last fully connected layer of Siamese Network will not be transferred to the softmax layer. Instead, it is directly transferred to the loss function, which is used to calculate the Euclidean distance between the output characteristic graphs of the two branches and calculate the similarity of the two input samples.

3.2. Design of the dual-mode contrastive loss function

The contrastive loss function can be used to analyze the feature space of double input data samples in Siamese networks. It makes the distance between similar samples as small as possible, and the distance between dissimilar samples increases, as in equation (1).
When the label $Y=0$, $L = \frac{1}{2} (D_W)^2$, it means that the two samples are of the same class. It will adjust the parameters to minimize the distance between $X_1$ and $X_2$. If the Euclidean distance between $X_1$ and $X_2$ is large, it means that the current model parameters are not good enough, so the loss $L$ will be increased.

When the label $Y=1$, $L = \frac{1}{2} (\max(0, m - D_W))^2$, it means that the two samples are of different classes. It means that if the Euclidean distance between $X_1$ and $X_2$ is larger than the threshold $m$, do nothing, which greatly reduces the calculation amount of loss iteration; if the Euclidean distance between $X_1$ and $X_2$ is less than the threshold $m$, the distance between them will be increased to $m$.

European distance $D_W$ mainly represents the absolute difference between two vectors. It is usually used to analyze problems that need to distinguish the similarities and differences of vectors from the magnitude of the value itself, where Cosine similarity focuses on distinguishing the similarities and differences of vectors from the direction of the vector space, which is insensitive to the value. Therefore, combined with the Cosine similarity, a dual-mode contrastive loss function was proposed.

The Cosine similarity measure uses the cosine value of the angle between two inputs to express the similarity between them, as in equation (3):

$$\text{similarity} = \cos(\theta) = \frac{X_1 \cdot X_2}{\|X_1\| \|X_2\|} = \frac{\sum_{i=1}^{n} X_{1i} \times X_{2i}}{\sqrt{\sum_{i=1}^{n} (X_{1i})^2} \times \sqrt{\sum_{i=1}^{n} (X_{2i})^2}}$$

Here $X_1$ and $X_2$ respectively represent the components of the input vector $X_1$ and $X_2$. The value range of Cosine similarity is between [-1, 1]. If its value is "-1", the main directions of the two input vectors are considered to be opposite. If its value is "1", it means that the directions between the two vectors are the same. If its value is "0", it means that the two vectors are independent of each other. The closer the value of similarity is to "1", the more similar the two vectors are.

The Euclidean distance and the Cosine similarity were combined with a weight value $\alpha$. The smaller the Euclidean distance is, the more similar the two inputs are. The value range of the Cosine similarity measure is between [-1, 1], and the larger the Cosine value is, the more similar the two inputs are. Therefore, change the value range of the Cosine similarity to [0, 2] and take the reciprocal to get a decreasing function, as in equation (4).

$$\cos\_\text{dis} = \frac{1}{\text{similarity} + 1 + 10e^{-6}}$$

Where: $10e^{-6}$ is only to ensure that the denominator is not zero. Then the specific formula of similarity determination distance as in equation (5).

$$\text{dis} = D_W + \alpha \times \cos\_\text{dis}$$

Where: $D_W$ is shown in equation (2); $\alpha$ is a weight factor, which indicates the proportion of the Cosine similarity to the European distance.

The dual-mode contrastive loss function equation completed by using the above combination distance $\text{dis}$, as in equation (6).

$$L_m \left(W, (Y, X_1, X_2)\right) = \left(1 - Y\right) \frac{1}{2} (\text{dis})^2 + Y \frac{1}{2} \left(\max(0, m - \text{dis})\right)^2$$

Where $W$ is the weight of the Siamese Network model; $Y$ is the label of the input data sample pair, $Y=0$ means that the input data sample pair and belong to the same class, and $Y=1$ means that they belong to different classes; $m$ is a threshold margin, which is the lower bound of Euclidean distance between different types of data samples; $\text{dis}$ is the similarity judgment distance between the Cosine similarity and Euclidean distance.

4. Results & Discussion

4.1. Datasets
The urinary red blood cells (RBC) image data set used in this paper was provided by a company and labeled by professionals with medical knowledge. The sizes of urinary RBC images in the data set varies and it contains three sizes, 16×16, 20×20 and 24×24 respectively. According to the different morphology of urinary RBC, the dataset was divided into 9 types, which were normal RBC, rolls-like RBC, shadow RBC, crescent RBC, ring RBC, puckered RBC, G1 RBC, lateral RBC and erect RBC.

**Figure.3 Urine red blood cell dataset and labels**

### 4.2. The training process
The model designed in this paper was trained and tested on Windows 10 system with PyTorch deep learning framework. The environment configuration of the experiment is shown in Table 1. The number of iterations in the training process of the Siamese network was 1500 epochs, and each epoch means to traversal the data of the train set once.

| Software                  | Hardware                      |
|---------------------------|-------------------------------|
| Framework                 | PyTorch 1.8                   |
| Graphics acceleration     | CPU Inter(R) Core(TM) i9-10900X|
| library                   | CUDA 10.0                     |
| Deep learning library     | GPU Nvidia GTX 2080Ti         |
|                           | cuDnn 10.0                    |
|                           | Memory 32G DDR4               |

Test set contains 2344 images, and the number of types of red blood cells are: normal RBC (RBC-1) for 119, rolls-like RBC (RBC-2) 112, shadow RBC (RBC-3) 188, crescent RBC (RBC-4) for 76, ring RBC (RBC-5) for 157, puckered RBC (RBC-6) for 86, G1 RBC (RBC-7) for 65, the lateral RBC (RBC-8) for 1340 and erect RBC (RBC-9) for 195.

### 4.3. Experimental results and analysis
The weight factors $\alpha$ of different sizes were set to observe the influence of dual-mode contrastive loss function. The $\alpha$ values were set as 0, 0.04, 0.05, 0.1, 0.2, 0.3, 0.4. The Siamese Network training model using the dual-mode contrastive loss function is denoted as the SIAM_LOSS model. The number of training iterations was set as 1500 epochs, the model learning rate was set as 0.0001, and the Adam weight attenuation function was used for training.

The Table 2 shows the classification accuracy of each category. It can be intuitively seen from the table that when the value of weight factor $\alpha$ is close to 0.05, the overall classification accuracy of SIAM_LOSS model is the highest. As shown in Figure. 4, when the weight factor $\alpha = 0$, it means that the Cosine similarity has no effect on the Euclidian distance, in other words, the dual-mode contrastive
loss function is equal to the contrastive loss function. In this case, the dual-mode contrastive loss function is equal to the contrastive loss function, and the accuracy value is 89.93%. The accuracy rate reaches the maximum when $\alpha = 0.05$, and then decreased continuously. The Cosine similarity is less sensitive to numerical values. When the proportion of the Cosine similarity in the similarity judgment distance is increasing, the classification accuracy of the model decreases. When the influence factor $\alpha$ of the Cosine similarity in the similarity determination distance is within the range of $[0, 0.1]$, it plays an auxiliary role in the similarity determination distance, which can effectively improve the accuracy of the whole classification model.

| $\alpha$ | 0   | 0.04 | 0.05 | 0.1 | 0.2 | 0.3 | 0.4 |
|---------|-----|------|------|-----|-----|-----|-----|
| acc     | 0.8993 | 0.9117 | 0.9181 | 0.9053 | 0.9027 | 0.8879 | 0.8790 |

Figure 4 Influence of different $\alpha$ values on classification accuracy

Set the weight factor $\alpha$ in the dual-mode contrastive loss function as 0.05. For the convenience of expression, the Siamese Network model with the dual-mode comparison loss function is denoted as SIAM LOSS model in the following paper, and the Siamese Network model with the comparison loss function is denoted as SIAM model.

The Table 3 visually shows the precision of Siam_Loss model and Siam model in urine red blood cell (RBC) classification task. It can be seen from the table that the SIAM LOSS network model improves the classification sensitivity of RBC-1, RBC-2, RBC-5, RBC-7, RBC-8 and RBC-9. For RBC-6, RBC-3 and RBC-4, the precision is reduced, but it is still within the error range. At the same time, the precision of RBC-7 with lower sensitivity is improved by 10.25%, which is still to be improved.

| method   | RBC-1 | RBC-2 | RBC-3 | RBC-4 | RBC-5 | RBC-6 | RBC-7 | RBC-8 | RBC-9 |
|----------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| Siam     | 88.73 | 84.68 | 89.92 | 59.21 | 95.36 | 100   | 31.46 | 99.03 | 88.63 |
| Siam_Loss| 91.89 | 90.72 | 84.44 | 51.65 | 96.67 | 95    | 41.71 | 99.77 | 93.30 |

Table 4 Recall of Siam_Loss model and Siam model in urine red blood cell classification task (unit: %)

| method   | RBC-1 | RBC-2 | RBC-3 | RBC-4 | RBC-5 | RBC-6 | RBC-7 | RBC-8 | RBC-9 |
|----------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| Siam     | 50    | 83.93 | 64.09 | 59.21 | 91.71 | 70.65 | 86.15 | 99.25 | 100   |
| Siam_Loss| 57.14 | 78.57 | 80.85 | 61.84 | 92.99 | 88.37 | 87.69 | 98.73 | 100   |
The Table 4 shows the change of recall rate between SIAM_LOSS model and SIAM model in urine RBC classification task. As can be seen from the table, the SIAM_LOSS network model improves the classification recall rate of RBC-1, RBC-3, RBC-4, RBC-5, RBC-6, RBC-7 and RBC-9. Although the recall rate of RBC-2 and RBC-8 has been reduced, it is still within the error range. At the same time, the correct classification ability of RBC-1 is low. In general, the SIAM_LOSS network model has improved the ability to correctly classify urinary RBC samples.

The experimental results shows that the Accuracy of the SIAM_LOSS model reaches 91.81%, indicating that 91.81% of the total number of samples are correctly classified, where the Accuracy of the SIAM model reaches only 89.93%. Due to the similar appearance and shape of urine RBC, it is difficult to fully identify the detailed features, and it is difficult to take into account the detailed features of all categories, which is easy to cause confusion. In addition, due to the uneven distribution of urine red blood cell data set samples, some categories are not fully trained, which also leads to the unsatisfactory classification accuracy of some categories.

5. Conclusion
Aiming at the fine-grained classification task of urine RBC images, a classification method of urine erythrocyte using Siamese Network based on dual-mode contrastive loss function was proposed. Firstly, due to the small size and limited number of urine RBC images, the Siamese Network was proposed as the basic network for urine RBC classification in this paper. According to the size of the input data, the network model structure of 5 convolutional layers was designed. Secondly, by introducing the Cosine similarity and a weighting factor $\alpha$, and combining with Euclidean distance, a new dual-mode contrastive loss function was proposed. The experimental results show that when the weight factor $\alpha$ is within the range of $[0, 0.2]$, the Cosine similarity has a positive effect on the comparative loss function and can effectively improve the accuracy of the whole classification model. When the value of weight factor $\alpha$ is equal to 0.05, the total classification accuracy of SIAM_LOSS model is 7.12%, which is higher than that of SIAM model in the test set. The algorithm proposed is effective in the classification of urinary erythrocytes, which improves the sensitivity of some cell categories with less differentiation.

Acknowledgments
Of the many who provided us with constructive comments, we are specially grateful to Jianhong Xiang and Qingbo Ji, and the reviewers for their fruitful comments and corrections.

References
[1] Zhang, N., Donahue, J., Girshick, R., Darrell, T. (2014). Part-based R-CNNs for Fine-grained Category Detection. In: European Conference on Computer Vision.
[2] Wei, X. S., Xie, C. W., Wu, J. (2016) Mask-CNN: Localizing Parts and Selecting Descriptors for Fine-Grained Image Recognition.
[3] Xiao, T., Xu, Y., Yang, K., Zhang, J., Peng, Y. (2015) The application of two-level attention models in deep convolutional neural network for fine-grained image classification. In: 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR).
[4] Yu, C., Zhao, X., Zheng, Q., Zhang, P., You, X. (2018) Hierarchical Bilinear Pooling for Fine-Grained Visual Recognition. In: Computer Vision – ECCV 2018. Munich, Germany.
[5] Koch, G., Zemel, R., Salakhutdinov, R. (2015) July. Siamese neural networks for one-shot image recognition. In International Conference on Machine Learning (Vol. 2).
[6] Bikbov, B., Purcell, C. A., Levey, A. S., Smith, M., Murray, C. J. L. (2020). Global, regional, and national burden of chronic kidney disease, 1990–2017: a systematic analysis for the global burden of disease study 2017. The Lancet, 395(10225): 709-733.