15th IAPR International Conference on Machine Vision Applications (MVA)
Nagoya University, Nagoya, Japan, May 8-12, 2017.

Computational single-cell classification using deep learning on bright-field and phase images

Nan Meng Hayden K.H. So Edmund Y. Lam
Department of Electrical and Electronic Engineering,
The University of Hong Kong, Pokfulam, Hong Kong
{nanmeng, hso, elam}@eee.hku.hk

Abstract

Automated cell classification is an important machine vision problem with significant benefits to biomedicine. We propose an efficient high-accuracy framework to classify cells based on bright-field and phase images using deep learning. With carefully designed network architecture and parameters, our network extracts features from single-cell images hierarchically and performs classification jointly. It can identify different types of cells without any human intervention and biological or hand-crafted features. Our experiments show that the system achieves a mean class accuracy of 96.5% on the single-cell images captured by an ultrafast time-stretch imager.

1 Introduction

Cell classification is of fundamental importance to personalized medicine, cancer diagnostics, and disease prevention, since the proportion of different types of cells provides important evidence for illness treatment. However, classifying different types of cells with high accuracy and speed is challenging, given the occlusion of other cells, apparent variation in shape and size, and effects from the external environment. Also, illumination variations, which cause the contrast between cell boundary and background to vary, have a detrimental effect on the accuracy of cell detection. Single-cell-based classification has its own additional challenges. Unlike cells in tissue, individual cells have similarly round shapes, making it a tough task to distinguish among different types.

To achieve the goal of identifying cells with high accuracy and speed from single-cell images, we propose to use deep learning methods. Convolutional neural networks (CNNs) have shown great potential in the field of image classification in recent years. One significant advantage of using CNN-based models is that the networks are able to learn distinguishing features automatically without human intervention. They are more versatile than hand-crafted features, leading to CNN generally outperforming traditional methods in image classification [1, 2] and object detection [3, 4].

Despite CNN's success, there are several challenges in using this framework to different applications such as cell images. First, CNNs require a large amount of data, since a large number of parameters need to be learned. Second, the intraclass variations are small since different types of cells have very similar appearances. Therefore, optimizing hyper-parameters of CNNs for cell images training data generally leads to overfitting and bias.

In this work, we overcome the challenges of CNNs for cell identification and propose some strategies to reduce the influence of these shortcomings. We outline a practical way to use CNN to learn useful features among different types of cells and then classify new cells automatically. This paper has the following contributions: 1) We propose a network based on CNNs for cell identification, which is shown to be effective on a single-cell dataset, and further outline details for how to construct the model with appropriate structures. 2) We propose a novel channel augmentation strategy via cascading several relevant images together. Our results show that this strategy contributes to extracting distinguishable features, further improving the classification accuracy.

2 Related Work

Various methods have been proposed for cell classification. A popular framework for these approaches is 1) feature extraction, 2) cell classification. An early work proposed by Perner et al. [5] introduced data mining techniques into image analysis. Relevant features are extracted among a large number of features found by image analysis. These relevant features are then inputted into a classifier. Later, Theera-Umpon [6] presented an automatic technique to extract four features from segmented nucleus and cytoplasm of bone marrow WBC and all four features are then classified by neural network. Foggia et al. [7] presented a heterogeneous set of features with Local Binary Patterns (LBP) to describe the mitotic cells and then used five classifiers to classify cells based on these features. The framework proposed by Nosaka and Fukui [8] combines a novel descriptor, i.e. Rotation Invariant Co-Occurrence among adjacent LBP (RIC-LBP), with a traditional multi-class SVM. The framework is invariant to local and global rotations of the cell image. All these approaches need manual features, which may suffer from intrinsic limitations related to human observations, and thus automatic classification with computer vision techniques is highly desirable.

Recently, deep learning approaches achieve great success especially after deep CNNs have been applied to classify images at a large scale [1]. One important reason is that deep CNNs have strong capacities to learn representational features not only including low and middle levels but also high-level vision ones. A straightforward application for cell classification using CNNs was proposed by Bayramoglu et al. [9] who improved classification accuracy by conducting different strategies for enhancing, augmenting and processing training data. The framework from Gao et al. [10] utilized three convolutional layers followed by two fully-connected layers to get better adaptability. Li et al. [11] explored extra cell images targeting on cross-

specimen analysis. They proposed deep convolutional architecture and evaluated on ICPR 2014 dataset. The main novelty pertains to the data augmentation step, which is performed by adding new specimens.

In this paper, we propose a CNN-based framework which receives single-cell images from the Quantitative Phase Imaging (QPI) system on FPGA and classifies them automatically. The QPI system [12, 13] records four single-angle images (bright-field images) of each cell with four fiber coupling angles and extracts phase image from these bright-field images [14, 15]. Our framework acquires all these five types of cell images from QPI imaging system. Experiments show that the framework is able to achieve high accuracy.

3 Single-cell Dataset and Preprocessing

3.1 Single-cell dataset

To evaluate the effectiveness of our CNN-based models, we test it on the single-cell image dataset obtained from the QPI system. For each cell, the system gets four bright-field images recorded from different angles. According to [15], the phase image $\phi(x, y)$ is calculated by

$$\phi(x,y) = Im\left[\mathcal{F}^{-1}\left\{\begin{array}{ll} C & ,\kappa_x = \kappa_y = 0 \\ \frac{\mathcal{F}\{G(x,y)\cdot \mathrm{FOV}\}}{2\pi\cdot(\kappa_x + i\kappa_y)} & ,otherwise \end{array}\right\}\right], \tag{1}$$

where $\mathcal{F}$ and $\mathcal{F}^{-1}$ correspond to forward and inverse Fourier Transform and $(\kappa_x + i\kappa_y)$ indicates Fourier spatial frequencies normalized as a linear ramp. C is an arbitrary integration constant, and FOV is the image field of view expressed in physical units. The complex phase shift $G(x, y)$ is calculated by

$$G(x,y) = \nabla\phi_x + i \cdot \nabla\phi_y, \tag{2}$$

where $(\nabla\phi_x, \nabla\phi_y)$ is the local phase shift obtained from four bright-field images. We have collected five types of cells: MCF7, OAC, OST, PBMC and THP1. The whole dataset includes 2500 images (500 images per type) and the size of each image is 305 × 305 × 5 which is obtained by combining 4 types of bright-field images and a phase image. Figure 1 shows a few typical examples of cells of various classes. Each column represents a specific type of cell while each row represents a specific type of bright-field image or phase image. Regardless what type the cells belong to, they have similarly round outline when flowing through the imaging system, making identification task challenging.

3.2 Channel augmentation

The cell images are first preprocessed by performing channel augmentation and intensity normalization. In particular, they are cascaded an additional channel of information and then each pixel in each channel is normalized to a number between 0 and 1.

In our imaging system, each cell is imaged with 4 different fiber coupling angles. Besides being used to compute the phase image, the 4 bright-field images by themselves may also provide additional information for classifying cell images. We thus also evaluate the effectiveness of employing them as additional imaging channels. Just as a color image has RGB channels, the cascaded image possesses high channel dimensions. Given that four types of bright-field images are collected from four different directions of cells, the cascaded images contain more angular information and the phase information benefit the network in learning better features. Channel augmentation aiming to enrich information of each individual image does not increase the size of the dataset used for training. However, it helps to train an effective model taking more information into consideration. Table 1 shows that compared with single channel or phase, images after channel augmentation processing contribute a 1% to 7% accuracy improvement, which reflects this augmentation strategy is effective for cell classification.

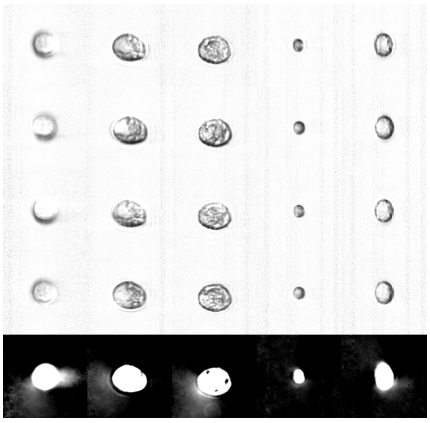

Figure 1: Examples of single-cell images dataset. Each column represents a specific type (From left to right: MCF7, OAC, OST, PBMC and THP1.) and each row exhibits a specific channel.

4 Network Design

We now focus on the design of the neural network and how to tune the structure to obtain a robust model to avoid overfitting. Traditional CNNs are hierarchical networks composed of two parts: feature extraction and a classifier. One of the most important layers in feature extraction is a convolutional layer, which leverages on three properties, namely sparse interactions, parameter sharing, and equivariant representations to extract transform invariant features to improve machine learning systems. The whole framework of our model is presented in Figure 2.

Considering that translation and rotation of cells may be present in images, we first utilize the properties of the traditional convolution layer to extract robust and invariant features for these variations. The channel augmentation process also introduces more angular information into these features as well. Second, a 2:1 downsampling max-pooling layer is used to make the output less redundant. After the pooling layer, there is a batch normalization layer. According to Ioffe and Szegedy's recent work [16], the batch normalization layer is effective to avoid overfitting. Experimental results presented in Figure 3 and Figure 4 provide strong

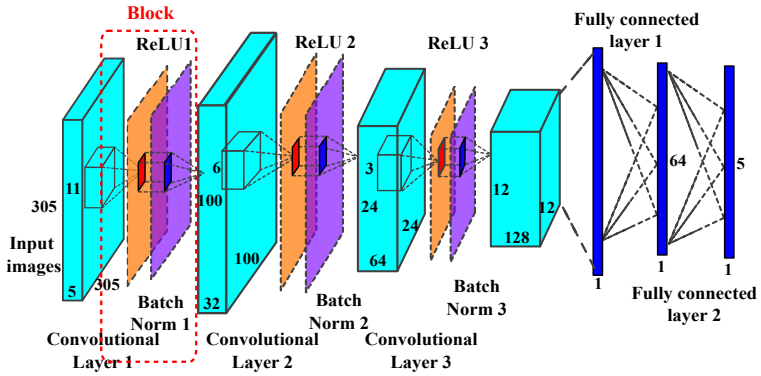

Figure 2: The structure of CNN-based model for cell identification.

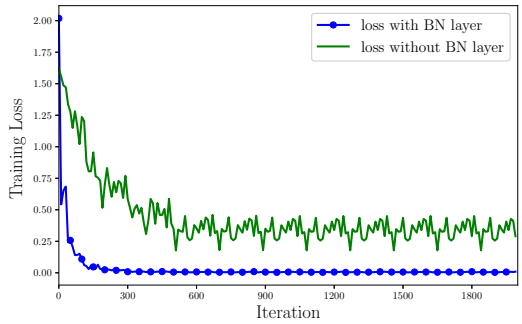

Figure 3: Training loss curves with and without the batch normalization layer.

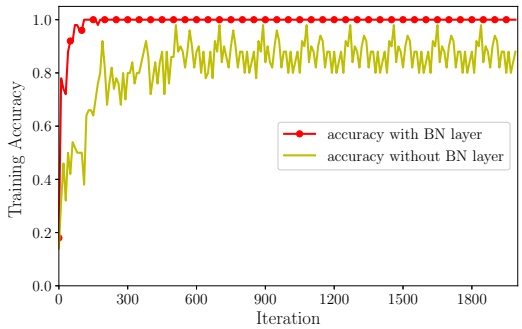

Figure 4: Training accuracy curves with and without the batch normalization layer.

evidence that with the batch normalization layer behind the active layer, the network can reach a better optimal point with lower loss and higher accuracy. We treat these three connected layers as a basic feature extraction unit, what we call a "block", from which we obtain more robust outputs. Figure 2 exhibits the detailed structure of blocks marked with the red dash line. Next, different blocks are cascaded hierarchically to extract high-level features. These abstract features are more distinguishable after three blocks of learning, and then they are fed into the two fully connected layers to classify cells.

5 Experiments

We use Caffe [17] to implement our CNN model. An Nvidia 980Ti graphics card is used to train the model

Table 1: Classification accuracy using our proposed method

Aspects	Average Accuracy	
	validation	test
channel 1	0.94	0.94
channel 2	0.96	0.95
channel 3	0.94	0.96
channel 4	0.94	0.92
channel 1-4	0.96	0.93
phase only	0.93	0.90
channel 1-4 & phase	**0.97**	**0.97**

with standard back-propagation, using SGD procedure. We randomly partition the cell images into three subsets, that is, 76% for training, 8% for validation, and 16% for test. This partition is utilized to all experiments. The network is trained using the minibatch stochastic gradient descent with a momentum factor of 0.9. Each iteration operates on a minibatch of 50 images that are sampled randomly from the training set. The network is trained for 12,000 iterations with a base learning rate of 0.01. The learning rate changes from 0.01 to 0.0001 after 500 iterations. After each round of the iterations in the training process, the layer parameters are updated based on the misclassification loss.

As mentioned, the bright-field images introduce angular information into the extracted features, since they are acquired from various directions. Therefore, a reasonable assumption is that by combining different types of bright-field images and phase images, the extracted features should be more distinguishable. To examine this assumption, we train the model several times with different inputs. First, four types of bright-field images are fed into the model respectively, and after several rounds of training, when the validation loss converges, we run the model using our test set. Next, all four types of bright-field images have been cascaded together in the third (channel) dimension and thus we obtain cascaded images with size $305 \times 305 \times 4$. We repeat the previous train-test procedure using the cascaded images. Results from Table 1 indicate that only integrating bright-field images provides little help to classification accuracy.

To further improve the performance of our network, we introduce the phase images which have been calculated from bright-field images. In order to demonstrate the usefulness of the phase images, an experimental comparative study has been conducted via training the

network with or without phase images.

Table 2 and Table 3 exhibit the precision and recall of cell classification results. They are calculated in a one-against-others manner. For example, the precision of THP1 is calculated by considering THP1 cells as positive instances and all the other cells as negative instances. Each row represents different training inputs and each column represents the precision or recall values of corresponding cell type. It is clear that our proposed classification framework is highly sensitive to each type of cells and with the help of channel augmentation it can achieve an overall best performance by combining bright-field images and phase images together.

Table 2: Precision of test

Aspects	Precision of each class **test**				
	THP1	OAC	MCF7	PBMC	OST
channel 1	0.8429	0.9551	**1.000**	0.8795	**1.000**
channel 2	0.9296	0.9872	0.9868	0.8588	0.9889
channel 3	0.9444	0.9753	**1.000**	0.9255	0.9753
channel 4	0.7838	0.9773	0.9880	0.8462	0.9870
channel 1-4	**1.000**	0.9859	0.9767	0.7907	0.9259
phase	0.8571	0.9167	0.9870	0.8642	0.8642
channel 1-4 &phase	**1.000**	**1.000**	0.9524	**0.9351**	0.9324

Table 3: Recall of test

Aspects	Recall of each class **test**				
	THP1	OAC	MCF7	PBMC	OST
channel 1	0.8551	**1.000**	0.9634	0.9012	0.9518
channel 2	0.8354	0.9872	**1.000**	**0.9359**	0.9889
channel 3	0.8947	0.9875	**1.000**	0.9355	**1.000**
channel 4	0.7945	**1.000**	**1.000**	0.8048	0.9870
channel 1-4	**0.9870**	0.9859	0.9882	0.9189	0.8065
phase	0.8571	0.8652	**1.000**	0.8642	0.9091
channel 1-4 &phase	0.9756	0.9770	**1.000**	0.9351	0.9324

6 Conclusion

In this paper, we developed a CNN-based model to classify single-cell images. In particular, we investigated the effect of channel augmentation strategies and we also compared the classification results with or without phase images and proved that channel augmentation was helpful to obtain better configuration of the network. Our results show that our CNN-based model is effective and can achieve high accuracy on single-cell image classification. This work was supported in part by the NSFC/RGC under Project N_HKU714/13, and GRF 17245716, and the Croucher Innovation Award.

References

[1] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems*, 2012, pp. 1097–1105.

[2] M. D. Zeiler and R. Fergus, "Visualizing and understanding convolutional networks," in *European Conference on Computer Vision*, September 2014, pp. 818–833.

[3] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *IEEE Conference on Computer Vision and Pattern Recognition*, June 2014, pp. 580–587.

[4] R. Girshick, "Fast R-CNN," in *IEEE International Conference on Computer Vision*, February 2015, pp. 1440–1448.

[5] P. Perner, H. Perner, and B. Müller, "Mining knowledge for HEp-2 cell image classification," *Artificial Intelligence in Medicine*, vol. 26, no. 1, pp. 161–173, October 2002.

[6] N. Theera-Umpon, "White blood cell segmentation and classification in microscopic bone marrow images," in *International Conference on Fuzzy Systems and Knowledge Discovery*, vol. 3614, August 2005, pp. 787–796.

[7] P. Foggia, G. Percannella, P. Soda, and M. Vento, "Early experiences in mitotic cells recognition on HEp-2 slides," in *IEEE International Symposium on Computer-based Medical Systems*, October 2010, pp. 38–43.

[8] R. Nosaka and K. Fukui, "HEp-2 cell classification using rotation invariant co-occurrence among local binary patterns," *Pattern Recognition*, vol. 47, no. 7, pp. 2428–2436, July 2014.

[9] N. Bayramoglu, J. Kannala, and J. Heikkilä, "Human epithelial type 2 cell classification with convolutional neural networks," in *IEEE International Conference on Bioinformatics and Bioengineering*, January 2016, pp. 1–6.

[10] Z. Gao, J. Zhang, L. Zhou, and L. Wang, "HEp-2 cell image classification with convolutional neural networks," in *IEEE Workshop on Pattern Recognition Techniques for Indirect Immunofluorescence Images*, December 2014, pp. 24–28.

[11] H. Li, J. Zhang, and W.-S. Zheng, "Deep CNNs for HEp-2 cells classification: A cross-specimen analysis," *arXiv:1604.05816*, April 2016.

[12] A. K. Lau *et al.*, "Interferometric time-stretch microscopy for ultrafast quantitative cellular and tissue imaging at 1 μm," *Journal of Biomedical Optics*, vol. 19, p. 076001, July 2014.

[13] J. Wu *et al.*, "Ultrafast laser-scanning time-stretch imaging at visible wavelengths," *Light: Science & Applications*, vol. 6, p. e16196, January 2017.

[14] T. T. Wong *et al.*, "Asymmetric-detection time-stretch optical microscopy (ATOM) for high-contrast and high-speed microfluidic cellular imaging," *Scientific Reports*, vol. 4, p. 3656, January 2014.

[15] J. Xie, X. Niu, A. K. Lau, K. K. Tsia, and H. K. So, "Accelerated cell imaging and classification on FPGAs for quantitative-phase asymmetric-detection time-stretch optical microscopy," in *IEEE International Conference on Field Programmable Technology*, December 2016, pp. 1–8.

[16] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," *arXiv:1502.03167*, March 2015.

[17] Y. Jia *et al.*, "Caffe: Convolutional architecture for fast feature embedding," in *ACM International Conference on Multimedia*, November 2014, pp. 675–678.