Estimating Z-position of Motile Cells for Robotic Cell Manipulation

Jiaqi Chen, Jiaqi Wang, Zhuoran Zhang

Abstract— **Obtaining the position of the manipulated cell and manipulation tool is an essential step in robotic cell manipulation. While existing methods are capable of locating the x-y position of cells and manipulation tools within the focal plane, it remains a challenging task to locate the z-position of the manipulated cell, especially when the cell is motile and changing its appearance in microscopic images. Here we propose a new strategy for estimating z-positions of motile cells. Taking advantage of the shallow depth of field of an optical microscope, we transform the z-position solving problem into a multi-class classification problem. Different from the existing depth of focus and depth of defocus methods, our strategy takes a single image of a cell as input and classifies it into different originating z-positions. The multi-class classification problem is then solved by a deep learning classification approach. Using motile sperm as an example, the proposed strategy achieved a Top-1 accuracy of 77.3% and Top-2 accuracy of 96.1%. The proposed strategy provides a new approach for estimating z-positions of motile cells from a single monocular microscopic image, thus paving the road for 3D robotic cell manipulation.**

Keywords: Robotic cell manipulation, Automation at micro-scale, Robot vision.

I. INTRODUCTION

The synergy between robotics and cell biology has become increasingly strong over the past decades. Robotic systems have been developed for patterning [1], grasping [2], and aspirating and injecting single cells [3][4]. In robotic cell manipulation, it is essential to obtain the positions of the manipulated object (i.e., cell) and the manipulation tool (e.g., a glass micropipette). The task discussed in this paper is obtaining the position of a motile sperm, which is an essential step for clinical infertility treatment (e.g., injecting a sperm into an egg cell) [4]. Since the position information of sperm cells is in three dimensions, it is necessary to accurately position the x, y, and z-positions of the sperm cell head in the in-focus state of the microscope field of view for further microscopic manipulation of sucking by the robot. Due to its motile nature, a sperm often swims in and out of the microscope focal plane. This property results in different visual appearances of the sperm head in out-of-focus images than in focused images [5]. Even with mature technology for the positioning of sperm cells in the x and y-axis directions in the in-focus state [6], the positioning of the z-axis is still limited by many factors, including the small range of in-focus, which is always on a micron scale, caused by the small depth of field of the microscope lens. Due to such limitation, obtaining the z-position of sperm cells in the in-focus state requires high precision for practical operations.

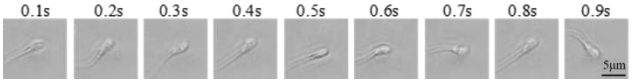

Fig. 1. During movement, motile sperm change their appearance within the focal plane. Scale bar, 5 μm.

Previous studies on analyzing z-position of sperm cells in the in-focus state mainly covered two methods, namely depth from focus [7] and depth from defocus [8]. The general principle of depth from focus algorithm is to obtain a set of focus pictures by photographing objects at different depths to form a focus stack and calculate the characteristic values of each image in the stack, such as the pixel gray level, the corresponding edge value [9], etc. The image with the best performance for the indicator is identified by the algorithm as the in-focus state. For autofocusing algorithms like depth from focus, the ideal output is always defined as having a maximum value at the location of the best-focused image and decreasing with increasing defocus [10]. However, this method is not suitable for imaging sperm cells. Since sperm is active and the movement is unpredictable, the time required for the camera to collect a focus stack and the bandwidth limitation of the camera itself is difficult to meet the focus changes during sperm movement.

For the other method, depth from defocus, it is necessary to establish a one-to-one correspondence between the focus and the depth by measuring a certain feature of the image, such as the blur of the observed object as the feedback information of the depth [10][11]. Since sperm swimming is a three-dimensional movement, the movement of sperm in the z-axis direction will affect the observation of the blur degree of sperm cells in the microscopic field of view, which will cause errors in the establishment of the corresponding relationship between defocus degree and cell depth.

This paper presents a strategy that is capable of simultaneous completion the positioning and prediction of the distance of sperm cells to the in-focus state based on deep learning. We use a DNN classification model to learn and classify the morphology of sperm cells at different z positions, so as to predict the relative position of sperm with the in-focus state as a reference based on the image of sperm under the microscope in real-time operation. Compared with conventional focus measure-based methods, with the use of this classification strategy, quantitative results show that it can be more accurate to locate the sperm.

The authors are with School of Science and Engineering, The Chinese University of Hong Kong, Shenzhen, 517182, Guangdong Province, China (e-mail: 119010017@cuhk.edu.cn, wangjiaqi@cuhk.edu.cn). Corresponding author: Zhuoran Zhang (zhangzhuoran@cuhk.edu.cn).

This work is supported by the University Development Fund of CUHKSZ (UDF01002141), and Guangdong Basic and Applied Basic Research Foundation (2021A1515110023).

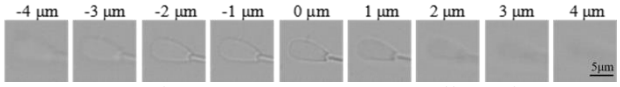

Fig. 2. Images of a sperm head captured at different focal planes (z-positions). Scale bar, 5 µm.

II. Experimental Setup

A. System Setup

The robotic cell manipulation system is built around a standard inverted microscope (Eclipse Ti2, Nikon). A CMOS camera (Basler MED ace 23 MP 164 color, Basler) is connected to the microscope to capture images at a bright field. The objective in the experiment was set to 40× for sperm observation, which is identical to the magnification used in clinical sperm manipulation tasks. All cell slides used in the study were located on a motorized microscope slide stage (ProScan III, Prior). During the automated procedure of image acquisition, the movements including x, y, and z-axis translation motions of sperm slides were achieved by the motorized Prior stage.

B. Preparation of Sperm Samples

All sperm samples in the experiment were collected from bovine semen. Holding treatment was preprocessed on collected semen samples to reduce cell motility for observation. The semen sample was aspirated at a volume of 5 mL using a pipette, then mixed with the diluent (PBS, Aladdin, China) in a volume ratio of 1:1. For each imaging experiment, 5 µL of sperm cell sample was added to the slide for subsequent image acquisition.

III. Methods

A. Problem Formulation

To obtain the z-position of sperm, the depth from focus methods need to obtain a stack of images from different focal planes, then calculate the focus measure score for each image in the stack and select the image with the highest focus measure score. This trial-and-error-based searching strategy is time-consuming and cannot satisfy the speed requirement in locating the z-position of a motile sperm. In contrast, the depth from defocus method can achieve z-position estimation in real time because it takes a single image as input. In terms of accuracy, both methods fail to address the changes in sperm appearance during sperm movement. For instance, sperm movement introduces noise in the focus measure score for the depth from focus method, and it is infeasible to establish a look-up table for each sperm for the depth from defocus method. Due to the motile nature of sperm, it is highly desired to estimate z-position of motile sperm in real-time (using a single image) with high accuracy.

Taking advantage of both methods, this paper proposes a novel strategy for sperm z-position estimation. The depth from focus method is capable of distinguishing images that are in-focus and out-of-focus but is slow in speed, whereas the depth from defocus method is fast in speed but does not distinguish in-focus and out-of-focus images. This paper uses a single image as input while distinguishing the in-focus status of the image. This naturally changes the problem of estimating z-position into a classification problem: in-focus and out-of-focus. Considering the limited depth of field of optical microscopes, the out-of-focus images could further be classified into images from different focal planes. Thus, the z-position estimation problem becomes a multi-class classification problem: i.e., given a single image as input, classify the image into different focal planes and predict its z-position (focal plane). The multi-class classification problem can then be solved by deep learning approaches.

The proposed strategy also takes advantage of the limited depth of field of optical microscopes. Objects within the depth of field are simultaneously in focus, while the limited depth of field gives a natural discrete "label" of the originating focal plane of each image. This also determines the resolution of the proposed method for estimating z-position. The depth of field of a microscope follows:

$$d = \frac{\lambda \cdot n}{NA^2} + \frac{n}{M \cdot NA} e$$

where d represents the depth of field, λ is the wavelength of illuminating light, n is the refractive index of the medium (typically 1.000 for air or 1.515 for immersion oil) between the coverslip and the objective front lens element, and NA is the objective numerical aperture. The variable e is the smallest distance that can be resolved by a detector that is placed in the image plane of the microscope objective, whose lateral magnification is M.

For a 40× objective with a numerical aperture of 0.65, the depth of field is 1.0 µm, thus, the resolution for z-position estimation is 1.0 µm. Using an objective with a higher numerical aperture could further increase the resolution.

B. Transforming Z-position Estimation into a DNN-based Classification Approach

Through the deep learning model, the problem of selecting images in which sperm is in-focus is transformed into the problem of classifying the morphology of sperm in different z-positions [see Fig. 3]. For a given sperm, a stack of 2N+1 images is formed by moving the slide stage up and down a fixed distance D to change the z-position for N times respectively, with the focused state as the starting position. Then, the in-focus image is set as label 0. Under the same principle, each obtained image above the focal plane is set as label 1, label 2, and so on, and each image below the focal plane is set as label N, label N+1, and so on.

The classification model adopted in this work is ResNet

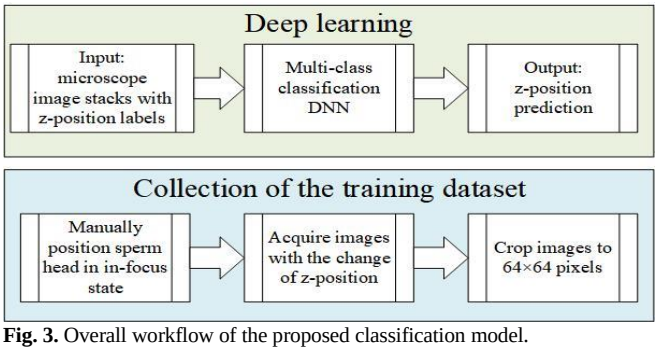

Fig. 3. Overall workflow of the proposed classification model.

Table 1. Performance of z-position predicted by different DNN-based models

DNN Models	Top-1 Accuracy (%)	Top-2 Accuracy (%)	Model Size	Training Time (s)	Inference Time (ms)
AlexNet	65.1	91.1	58.4 M	283	21
GoogleNet	76.3	95.7	41.4 M	612	32
ResNet34	77.3	96.1	85.3 M	564	29
ResNet50	76.12	96.34	94.4 M	743	32
ResNet101	76.5	96.7	170.7 M	1201	41

34 [12], with the classification procedure shown in Fig. 4. ResNet34 is a BasicBlock-based convolutional neural network with 34 layers. BasicBlock consists of two weight layers, utilizing a residual mapping strategy to map input feature x to H(x)-x as the feature map to output. According to the crucial position of depth of a neural network in learning, compared with other DNN models, ResNet can solve the problem of gradient explosion and gradient disappearance caused by network deepening, and successfully achieve deeper network depth to obtain more information. As shown by our results in Section IV.A, ResNet34 provides a balanced trade-off between classification accuracy and inference speed. With the help of this classification strategy, after taking an image of the current sperm, the model can feed back the category to which the sperm image belongs. As the per analysis of the depth of field, the selection criteria for in-focus images of sperm are within ±1 μm of the microscope focal plane.

To train the model, a dataset containing 900 images of sperm was collected. The dataset was constructed by 100 image stacks acquired from 100 sperm cells, with 90 stacks used for training and validating and 10 stacks for testing. For each image stack, it consisted of 9 images ranging in z-positions from -4 μm to +4 μm with a spacing of 1 μm relative to the microscope focal plane. To collect the dataset, the robotic system automatically controlled the motion of the z-axis of the motorized microscope stage, while grabbing and saving images simultaneously to acquire data set samples. The chosen sperm cells were initially put on the Prior stage, and were manually located at the in-focus state under the microscopic field of view. Acquisition of the sperm was then performed with the change of 1 μm on z-positions of sperm until 4 μm above and below its focal plane respectively, but fixed x and y-positions. As shown in Fig. 5 (a) and (b), the images out of the 4 μm range (above or below the focal plane) become blurry and it is hard for a human to distinguish the out-of-focus status from the image. After image processing with the principle of choosing the center of the sperm head as the center of the image, a stack of images in the size of 64×64 pixels for one sperm was recorded (see Fig. 2). Since each chosen sperm cell was detected in a continuous frame of images, the difference of visual appearances of a sperm head in out-of-focus states could be recorded in succession from its in-focus state.

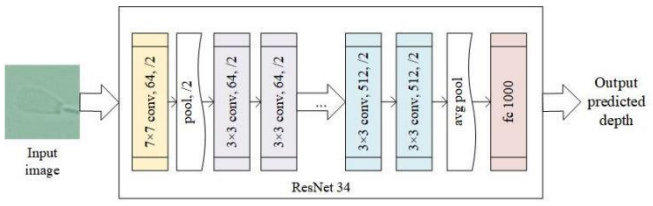

Fig. 4. Illustration of ResNet 34 model of taking a single microscopic image as input and predicting the z-position of the sperm.

IV. Results And Discussions

A. Comparison of Performance of Different DNN Models

Aimed to find a deep learning model with a good trade-off of accuracy, speed, and size, other state-of-the-art DNN classification models were compared, including ResNet50, ResNet101, GoogleNet [13], and AlexNet [14]. The prediction results on the same dataset are summarized in Table 1. Top-1 and Top-2 accuracy are used to evaluate the performance of DNN models. These models achieved the Top-1 accuracy with the range from 65.1% to 77.3%, and a Top-2 accuracy ranging from 91.1% to 96.7%. Comparing the model size of each algorithm, GoogleNet takes the smallest occupied size. While for the speed of training and testing, AlexNet is the fastest. Among these five models, ResNet 34 achieved the highest Top-1 accuracy of 77.3%, with the second fastest inference time of 29 ms. Hence, ResNet 34 is finally chosen as the DNN-based classification method to estimate z-position of sperm.

B. Comparison of the Proposed DNN Method versus Conventional Focus Measure-based Methods

Besides adopting DNN models for prediction, we also examined the performance of conventional focus measure-based methods, for example, Entropy [15] and Tenengrad [16]. For an immotile sperm, the performance based on a dataset derived from images of sperm heads at different z-positions can be seen in Fig. 6 (a), (b). The results imply that both methods are able to capture the relationship between the z-positions of sperm and the corresponding microscopic images, and they both reach the local maximum (i.e., peak) when matching with the sharpest in-focus sperm head images (z = 0 μm). However, when applying the focus measure-based methods for motile sperm, even capturing the sperm images within the same in-focus focal plane, both algorithms show a noisy focus measure curve. Within the same focal plane, the intrinsic sperm movement (e.g., sperm rotation along the head axis) changes the appearance of the sperm in the image, thus adding noise to the focus measure

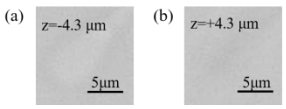

Fig. 5. Microscopic images of sperm out of the range of -4 μm to 4 μm of the in-focus plane. Scale bar, 5 μm.

curves and making the focus measure-based methods inapplicable to estimating z-positions of motile sperm.

In contrast, for motile sperm with changing appearance, the proposed DNN classification method achieved a Top-1 accuracy of 77.3% and a Top-2 accuracy of 96.1% for estimating sperm z-positions [see the confusion matrix in Fig. 8]. This is mainly because the training dataset has included images of sperm with different appearances, and during training, the DNN model classifies images of different moving sperm into their actual z-position. The movement-induced appearance change has been learned and incorporated into the model training process. Considering the in-focus state of the sperm head image is chosen within the range of ±1 μm relative to the microscope focal plane, the results confirm us that DNN-based classification method is a valid and efficient strategy for selecting the in-focus image of sperm heads, as well as predicting the corresponding z-positions.

V. Conclusion

In this study, we propose a strategy to transform the z-positioning problem of sperm cells into a deep learning-based classification problem, which can both predict the current position of sperm and address the difficulty of in-focus positioning. For comparing the performance of selecting sperm in-focus state, we exemplify different DNN models, as well as two traditional methods. Their performance is not unimodal within the selected range of z-positions and is limited to immotile sperm, which is contrary to the actual demand. Due to those limitations, we turned our solution to deep learning. Through the separate training of different DNN-based models, ResNet34 is examined to achieve a better balance in accuracy, speed, and model size. Using this model, we can estimate the sperm's z-position of sperms in robotic cell manipulation.

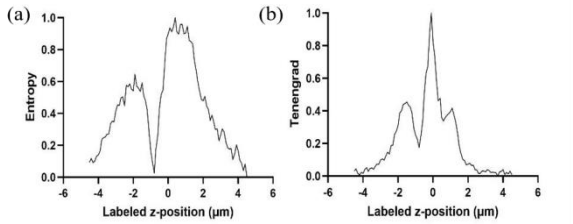

Fig. 6. Performance of focus-measure based algorithms for estimating the z-position of immotile sperm. For immotile sperm, both Entropy and Tenengrad algorithms could locate the correct in-focus z-position (0 μm).

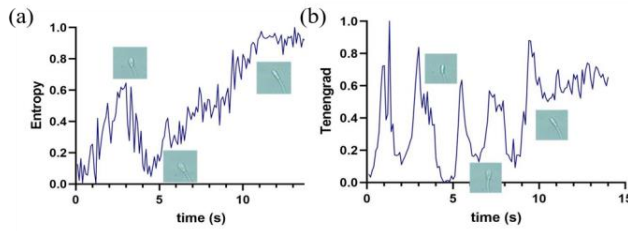

Fig. 7. For motile sperm, conventional focus-measure based methods failed to estimate their z-positions. Within the same in-focus focal plane, the intrinsic sperm movement changed sperm appearance thus adding noise into the focus measure.

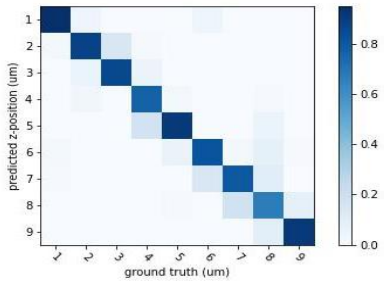

Fig. 8. Confusion matrix map of ResNet34.

References

[1] Kim, K., Cho, J., Pyo, J., Kang, S., & Kim, J. (2017). Dynamic object recognition using precise location detection and ANN for robot manipulator. 2017 International Conference on Control, Artificial Intelligence, Robotics & Optimization (ICCAIRO).

[2] Liu, X., Kim, K., Zhang, Y., & Sun, Y. (2009). Nanonewton force sensing and control in microrobotic cell manipulation. The International Journal of Robotics Research, 28(8), 1065–1076.

[3] Ladjal, H., Hanus, J.-L., & Ferreira, A. (2013). Micro-to-nano biomechanical modeling for assisted biological cell injection. IEEE Transactions on Bio-Medical Engineering, 60(9), 2461–2471.

[4] Fukui, W., Kaneko, M., Kawahara, T., Yamanishi, Y., & Arai, F. (2012). Geometrically-constrained cell manipulation for high speed and fine positioning. Journal of the Robotics Society of Japan, 30(6), 655–661.

[5] Liu, G., Shi, H., Zhang, H., et al. (2022). Fast noninvasive morphometric characterization of free human sperms using deep learning. Microscopy and Microanalysis: The Official Journal of Microscopy Society of America, Microbeam Analysis Society, Microscopical Society of Canada, 1–13.

[6] Sun, Y., Duthaler, S., & Nelson, B. J. (2004). Autofocusing in computer microscopy: selecting the optimal focus algorithm. Microscopy Research and Technique, 65(3), 139–149.

[7] Grossmann, P. (1987). Depth from focus. Pattern Recognition Letters, 5(1), 63–69.

[8] Subbarao, M., & Surya, G. (1994). Depth from defocus: A spatial domain approach. International Journal of Computer Vision, 13(3), 271–294.

[9] Valdecasas, A. G., Marshall, D., Becerra, J. M., & Terrero, J. J. (2001). On the extended depth of focus algorithms for bright field microscopy. Micron (Oxford, England: 1993), 32(6), 559–569.

[10] Sun, Y., Duthaler, S., & Nelson, B. J. (2005). Autofocusing algorithm selection in computer microscopy. 2005 IEEE/RSJ International Conference on Intelligent Robots and Systems.

[11] Zhou, C., Lin, S., & Nayar, S. K. (2011). Coded aperture pairs for depth from defocus and defocus deblurring. International Journal of Computer Vision, 93(1), 53–72.

[12] Wu, S., Zhong, S., & Liu, Y. (2018). Deep residual learning for image steganalysis. Multimedia Tools and Applications, 77(9), 10437–10453.

[13] Szegedy, C., Liu, W., Jia, Y., et al. (2015). Going deeper with convolutions. 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR).

[14] Iandola, F. N., Han, S., Moskewicz, M. W., et al. (2016). SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5MB model size. In arXiv [cs.CV].

[15] Huang, W., & Jing, Z. (2007). Evaluation of focus measures in multi-focus image fusion. Pattern Recognition Letters, 28(4), 493–500.

[16] Pech-Pacheco, J. L., Cristobal, G., Chamorro-Martinez, J., & Fernandez-Valdivia, J. (2002). Diatom autofocusing in brightfield microscopy: a comparative study. Proceedings 15th International Conference on Pattern Recognition. ICPR-2000.