Aspects of the analysis of cell imagery: from shape to understanding

Li, C.

Chapter 6

Conclusions and Discussion

In this thesis, we have addressed five research questions regarding two applications: pollen classification and neutrophil tracking in the biomedical domain. The themes investigated include image classification, image segmentation, and object tracking. In this chapter, we summarize the main findings from our research. Additionally, we discuss the limitations of our methods and the possible solutions to address them. Lastly, we point out the trends for future work.

6.1 Main Findings

In each research chapter, we have proposed new strategies/approaches to answer the corresponding research questions. Here, we summarize these approaches and present the main findings inspired by experimental results.

(1) **Chapter 2** answers **RQ 1: Can the existing deep learning-based classification models work with images from morphologically similar pollen grains of related species and what is the performance of the different models? Chapter 2** reports on the first time that CNN models are applied to classify morphologically similar unacetolyzed pollen grains of two common genera and a species in the Urticaceae family. In this Chapter, the captured raw pollen data is a 3D stack that has 20 through-focus layers. In order to gain good performance from classification models, therefore, we select three projections of the raw 3D pollen images and treat them as 2D data as the input of the CNNs. In this manner, projections incorporate as many representative features as possible. Consequently, we compare the classification performances of three popular CNNs including VGG16, light-weighted MobileNet V1, and MobileNet V2. CNNs extract features from images automatically via the convolutional layer and these feature vectors are fed into fully connected layers for classification. We,

respectively, train the three mentioned models with random initialization parameters, i.e. model from scratch, as well as pre-trained parameters, i.e. pre-trained model, also known as transfer learning technique. Data augmentation and hard voting techniques are adopted so as to further improve the performance. We observe that for the pre-trained model, VGG16, MobileNet V1, and V2 achieve comparable results. For the model from scratch, VGG16 has around 10% accuracy higher than the MobileNet. It proves that VGG16 is more robust than the light-weighted MobileNet and the transfer learning technique has a more significant effect on the MobileNet. In addition, the newly trained VGG16 model based on the new aerobiological samples is further used in a case study, which proves the deep learning-based VGG16 works successfully on our complex pollen classification task.

(2) **Chapter 3** is extended work from **Chapter 2**. We intend to integrate more models not only from deep learning but also from traditional machine learning approaches and explore the deeper insights further from those approaches.

In **Chapter 3**, we first involve three extra deep learning models: a shallow AlexNet, a comparable VGG19, and a deeper ResNet50. Among the six deep learning-based models, we find that the classification performance does increase from shallow AlexNet to deeper ResNet50. The VGG16, VGG19, MobileNet V1, and V2 reach comparable performance within the performance range of AlexNet and ResNet50. Secondly, we construct traditional machine learning-based models. We explore and extract the handcrafted features from our pollen images and use different classifiers such as SVM, RF, MLP, Adaboost to classify pollen categories. In this process, feature selection and reduction methods are applied to remove the irrelevant and redundant features. The flat models in which only one classifier is used to classify all classes are constructed. Compared to flat models, a hierarchical strategy merges the classes which are more similar into subgroups and classifies these subgroups separately with selected classifiers at different stages. The experimental results imply the hierarchical-structured classification model outperforms the flat-structured model.

The research question **RQ 2: How does the performance of the traditional machine learning-based classification models compare to that of deep learning-based models?** is answered after exploring the performance of deep learning-based models and traditional machine learning-based models. We observe that all the investigated flat classification models yield lower performance compare to the deep learning-based models, even to the shallow AlexNet. The hierarchical model is able to accomplish a comparable performance as AlexNet after fine-tuning all hyper-parameters to optimal settings. However, it could not keep pace with other deeper deep learning-based models investigated. The possible reason is the handcrafted features are designed from expertise and experience instead of learning from the ground truth automatically. There is a huge chance that the extracted features cannot

thoroughly cover the representations of pollen images. Therefore, we conclude that deep learning-based models have powerful capacities and better performance on similar pollen classification tasks.

In addition, to answer **RQ 3: To what extent is it possible that both the traditional machine learning-based and the deep learning-based classification models perform well on a relatively small amount of data?**, we conduct experiments on two small-size datasets, which have around 500/1000- images. All the aforementioned classification models are implemented and compared on the two datasets, respectively. The experimental results show: with the decreasing number of images in the dataset, the classification accuracy decreases accordingly. However, the trend is the same in that deep learning-based models outperform all flat models. The hierarchical models can achieve comparable performance with a shallow AlexNet, but not with other deeper CNNs. Moreover, we observe the decreasing extent of different models. For the 1000- image dataset, the accuracy is reduced by 2-5% varies from different models, while a 4-8% decrease on the 500- image dataset. The decreasing accuracy is fluctuating more or less but follows the same trend.

According to the analysis above, we are able to guarantee the power of the deep learning-based models for handling our pollen classification task. To explore more details of deep learning models, an ablation study is conducted. Peeling the whole model from transfer learning, data augmentation, and hard voting techniques step by step, we discover that the transfer learning technique has a significant impact on the improvement of the classification accuracy with 12-13%. While data augmentation and hard voting improve the performance of 1-2% and 1-3% further, respectively. The ablation study reveals the reason why the deep learning model works well for pollen classification.

(3) In **Chapter 4**, we answer the fourth research question **RQ 4: To what extent is it possible to develop an automated algorithm that provide accurate support in the tracking of neutrophils from time-lapse sequences in the 2D spatial domain?** In order to address this question, we first project the 3D images containing neutrophils to 2D images for all time points of the time-lapse recording. Subsequently, we load the images in our processing pipeline to conduct cell tracking in 2D + T space. We divide a tracking task into three parts: cell segmentation, cell motion tracking, and trajectory linkage. We compare several 2D segmentation approaches including both rule-based (Watershed) and deep learning-based (U-Net-based) methods. We aim to find a segmentation model that can maximally benefit the cell tracking part. The ground truth data for segmentation are prepared manually. We provide a ground truth dataset of 240 images. One more U-Net model is applied to learn the cell movement from the ground truth annotation. This ground truth dataset with hundreds of labeled trajectories is annotated by biological experts by "Manual tracking". Last but not

least, we propose an extended Viterbi algorithm to link the trajectories. We add different heuristics to the basic Viterbi algorithm in order to deal with complex cell behavior such as cell merging and splitting. We find that our extended Viterbi algorithm achieves superior performance compared to other straightforward linkage methods.

(4) In **Chapter 5**, we expand our cell tracking study to 3D + T space so as to answer question **RQ 5: To what extent is it possible to develop an automated algorithm that provide accurate support in the tracking of neutrophils from time-lapse sequences in the 3D spatial domain?** We divide 3D cell tracking into three parts: 3D cell segmentation, feature extraction, and trajectory linkage. 3D segmentation is conducted using a 3D U-Net model. The 3D segmentation ground truth data is labeled manually. Next, instead of using a deep learning model to automatically learn the cell movement motion in **Chapter 4**, we design and manually extract the handcrafted cell features. Our proposed rule-based feature-weighted approach has three features which include cell distance, cell movement direction, and average cell movement distance. They are selected after observing the moving neutrophils and characterizing them. Subsequently, we link the trajectories with different linkage methods based on the calculated feature similarity between pairs of candidate cells frame by frame. In contrast to **Chapter 4**, using a rule-based method rather than a deep learning method is because deep learning models require a large amount of ground truth data. However, annotation in 3D + T space is not only laborious and time-consuming but also not effective without proper labeling tools. Nevertheless, to evaluate the performance, we still need ground truth data. We come up with the idea of adding an extra Z coordinate on the X, and Y-axis coordinates that have already been annotated in **Chapter 4** and correcting X, and Y coordinates accordingly. In this manner, we annotate 20 ground truth trajectories for our dataset. The experimental results show our specific pipeline designed for neutrophils does improve the tracking performance compared to other state-of-the-art methods.

6.2 Discussion

Our proposed methodologies and pipelines in the thesis have addressed five research questions and achieved promising results in terms of two applications. They, however, still have some limitations which can be discussed from the following perspectives.

6.2.1 Data Perspective

(1) In **Chapter 2** and **Chapter 3**, the 20-layer raw pollen image stack captured using a bright-field microscope, are through-focus-images. The expert has not adjusted the focal

level of each Z-plane and only scans through the Z-axis by automated procedures of the microscope itself. In this case, it can be possible that some representative feature information is lost. This is why we project the 3D data into three different projections to acquire more features and classify them as 2D images. The focused 3D raw pollen data collection will be considered using CLSM in further research, and it supposes to achieve a different classification perspective.

(2) In **Chapter 4** and **Chapter 5**, in fact, the expert captures neutrophils and macrophages in the tail of zebrafish larvae, aiming to track both of the two types of leukocytes. The raw 3D time-lapse data acquisition with a confocal microscope has been set to one-minute intervals over a 2 hr period of time. The longer the time interval, the longer distance the cell migrates, so it is easy to lose the track. All three channels of neutrophil, macrophage, and brightfield images from two groups, experimental and controlled groups, of zebrafish need to be taken at the same time within one minute. This means the number of scanning sections on the Z-axis is limited. This is why we originally obtain an 8-layer 3D image.

Moreover, neutrophils and macrophages have different morphology and movement pattern. It is impossible to design a tracking algorithm to fit both of them well. Furthermore, macrophages have a more irregular shape and they are difficult to be distinguished even by experienced experts, and thus, so far, also much less by algorithms. In contrast, neutrophils can be tracked easily in practice. This motivated us to focus on tracking neutrophils only.

Overall, with these limitations mentioned above, there is a lot of room to improve the methods. Not only by improving data quality, but also, working with macrophage tracking needs to be considered in future research.

(3) In **Chapter 4**, the ground truth preparation for 2D cell tracking trajectories, is accomplished in a straightforward manner, in which a cell candidate is manually tracked through the whole sequence directly. The expert does not take the merge and split events into account during the labeling process. This is why we use the linkage method to subsequently solve this problem. Because the deep learning model can not learn these modes from ground truth data.

In **Chapter 5**, due to the lack of effective and convenient labeling tools in 3D + T space, we use a rule-based method to do tracking. Ideally, deep learning models are the mainstream nowadays, and training a learning model from annotated ground truth data should achieve high tracking ability. There has been research after methods to produce 3D annotations [141][157] but these do not appear to be reproducible in our case. Therefore, designing our annotation tool is essential. It should be further considered for future developments in this area.

6.2.2 Hardware Perspective

Deep learning systems require powerful hardware because they have a large amount of data being processed. In this thesis, all the deep learning models that we used are trained on a dedicated server equipped with two NVidia GeForce GTX 2070 with 8 GB GPUs and a processor with 64 GB RAM in our group. With the two GPUs, we conducted pollen classification in **Chapter 2** and **3** successfully. In **Chapter 4**, We also trained the 2D cell segmentation and 2D cell motion tracking models successfully. However, for 3D cell segmentation in **Chapter 5**, the image size of $512 \times 512 \times 8$ is too big to use as the input and output of end-to-end training on our GPUs, even with the setting "batchsize=1". To solve this, we apply a strategy to divide the original image into small patches to train a deep learning model, after training the small patches are stitched to the original size again. Although the 3D cell segmentation can be done with this strategy, the tracking task requires 3D time-series data as the input and output of the model. This needs more GPU memory than the Setup we have been employing. Except for the difficulties of ground truth data labeling, this is another reason we use a rule-based method rather than a deep learning method to conduct a 3D cell tracking task in this thesis.

6.2.3 Algorithmic Perspective

Despite of the aforementioned limitations we have achieved promising results using our algorithms in terms of our application study, there is still room for improvement.

(1) Cell segmentation. Ideally, no matter with 2D or 3D cell segmentation, the over-segmentation and under-segmentation are the key problems that need to be prevented. The segmented results will directly affect the performance of the subsequent tracking part. However, not one approach from either rule-based or deep learning methods can perfectly solve it. In the research presented in this thesis, we compared several existing segmentation methods in order to find the method with the best performance with respect to our data. In this manner, the subsequent tracking will benefit from the good segmentation performance. So for the analysis of the neutrophils we focused on the tracking rather than designing our own segmentation algorithms. In the future work, it will be possible to take this into account.

(2) Cell tracking. Compared with some state-of-art methods, our pipelines for 2D cell tracking and 3D cell tracking improve the performance in the experiments from **Chapter 4** and **5**. But in fact, the tracking rate of algorithms still cannot accomplish sufficient accuracy compared to the manual annotation, especially in a long-range time-lapse sequence. In

particular, the cell merging and splitting events are not solved satisfactorily even with a lot of rules designed in our pipeline. This is an absolute challenge in many cell tracking tasks.

(3) Trajectory visualization. In the 2D tracking task in **Chapter 4**, we link the position of each tracked trajectory with colored lines and map it to the raw data. The trajectories can be shown through a 2D time-lapse movie easily. However, in **Chapter 5**, it is a much more complex process to visualize the tracked trajectories within a 3D + T space. The process could include but is not restricted to cell detection, cell coordinates calculation, and 3D space rotation. Thus, it could be a future visualization work that can be achieved together with 3D ground truth annotation as we mentioned previously.

(4) Comparison of 2D and 3D tracking results. In **Chapter 4** and **Chapter 5**, we demonstrate that our dedicated algorithms outperform other state-of-art methods. This is because all of the comparisons are based on the same standard, ground truth data. However, to date, it is hard to conclude whether the performance of 3D tracking is better than 2D tracking. In **Chapter 4**, we annotate 110 ground truth data while only 20 ground truth is labeled in **Chapter 5**. The overlapped 20 ground truth data can not realistically represent the performance of our two proposed pipelines due to the small size of sampling. Therefore, enlarging the number of ground truth data for 3D tracking is essential for future research.

(5) Macrophage tracking. All previously mentioned algorithms are designed for neutrophils. Neutrophils have the features of clear cell borders that can be easily distinguished while having a faster motion. Compared with neutrophils, macrophages move slower but are more irregular. It is even difficult for biologists to tell them apart. With these completely different characteristics, the proposed algorithms for not only segmentation but also tracking are definitely not suitable to use on the two types of cells simultaneously. For macrophages, segmentation should be the more challenging work compared with the tracking part. A slower movement and velocity could make it easier to trace these macrophages.

6.3 Future Research Directions

In the previous Chapters, we have presented many approaches to address the research questions regarding two applications. A wide variety of future work is also encouraged to advance them further. In this section, we briefly discuss the possible future research directions regarding the aforementioned themes.

(1) We have labeled the ground truth data for 2D cell tracking without taking the merging and splitting events into account. If we can consider all of these possible cell events during

ground truth labeling, the deep learning model used in the tracking part could learn more potential cell motion patterns so as to achieve a better performance in theory.

(2) At present, specifical focus on the difficulties of labeling of ground truth trajectories in 3D + T space is essential, and the design dedicated interactive tools very much needed. It will be very valuable to realize labeling of 3D time-series data in the cell tracking field. Suppose we already have a smart tool that can help label ground truth data, we can apply deep learning approaches to conduct 3D cell tracking tasks easily. Consequently, visualizing the cell trajectories in 3D + T space from different angles is also possible and it reduces the dependency on commercial software. A complete working route as mentioned above might be feasible in the future.

(3) Except for the neutrophils' tracking task explored in our thesis, tracking macrophages will be a more challenging task. Nevertheless, investigating the migration behaviors of macrophages is a necessary step to take in future research.