MESENCHYMAL CELL MIGRATION MODES PREDICTION IN CONFOCAL MICROSCOPY IMAGES USING BAYESIAN DEEP LEARNING

Anindya Gupta[1]*, Veronica Larsson*[2]*, Nicolas Pielawski*[1]*, Damian Matuszewski*[1]
Staffan Strömblad[2]*, and Carolina Wählby*[1]

[1] Centre for Image Analysis, Dept. of Information Technology, Uppsala University, Sweden
[2] Department of Biosciences and Nutrition, Karolinska Institutet, Huddinge, Sweden

ABSTRACT

Cell migration is a complex and heterogeneous phenomenon consisting of many spatially and temporally regulated mechanisms. A complete understanding of cell migration phenomenon is indeed crucial to derive an adequate quantification for single cell-based analysis. However, the spatiotemporal behavior of dynamic cells is often complex and challenging for manual classification and analysis. Automatized algorithms are thus highly desired to facilitate mesenchymal migration sub-modalities prediction. Supervised deep learning techniques have shown promising outcomes in microscopy image analysis. However, their implication is confided by the amount of carefully annotated data. Weak supervision provides a simple, model-agnostic way to integrate the domain-expertise into a learning model. Additionally, bayesian predicitons can lead to a more informed decision, and the quality of prediction can be improved.

Index Terms— bayesian deep learning, convolution neural network, recurrent neural network, mesenchymal cell migration, long short-term memory, system microscopy

1. INTRODUCTION

The ability of the cancer cells to migrate constitutes a central aspect of cancer metastasis, which causes most cancer lethality. A complete understanding of cancer cell migration regarding environmental factors and drug treatment may provide us with clues on how to reduce the risk of metastasis. With an automated and quantitative approach to understanding cell migration process, we open up for large-scale systems microscopy experiments and drug screening. Typically, cells can adopt several substantially diverse migration modalities. Here, we focus on a specific type of movement called mesenchymal migration, where the cells adopt two distinct migration sub-modes: *discontinuous* (or random) and *continuous*, and can also switch between modes [1].

Quantitative analysis of migration sub-modalities is a critical aspect of cancer cell biology [2]. However, the study of cell migration yields an overabundance of experimental data that requires demanding processing and human-efforts. When the cell population becomes large, manual analysis becomes tedious, time-consuming, and error-prone. Also, the spatiotemporal behavior of dynamic cells such as motion heterogeneity, splitting, and persistence is often complicated and challenging for manual analysis. Automatized and integrated algorithms are thus highly desirable to predict different migration modes from the given image sequences. As cells are generally diverse and densely packed, developing efficient approaches for predicting migration modalities remain a challenging problem.

Aided by massive time-lapse microscopy imaging data, supervised deep learning can be successfully exploited for automated prediction of migration modalities. However, it strives for carefully curated large annotated training data [3]. With an increasing amount of time-lapse image sequences, annotating cell observations per frame is an expensive and time-consuming effort. The human annotators are often unable to annotate the migration modes with high confidence, which makes the ground-truth labels noisy. Also, an insufficient understanding of model outputs may provide suboptimal results [4, 5]. Therefore, the uncertainty quantification for the probabilistic interpretations of predictions is desirable.

During live-cell acquisitions, cells are exposed for a longer duration to capture the dynamic cellular responses over time. However, over-exposure can cause photo-toxicity, injuring live cells and possibly causing apoptosis, and thereby reducing the number of cells to be used in subsequent downstream experiments [1]. Predicting the migration mode with minimal imaging data will require less acquired data, eventually reducing the time efforts and data storage challenges. Also, live cells will be exposed to less photo-toxicity, increasing the number of viable cells available for downstream experiments. Recent studies have shown that the current cell morphology influences its future responses [6]. Therefore, it is interesting to explore whether CNNs can be used to predict migration modality based on the current cell morphology.

In this study, we compare existing CNN architectures on their potential to encode discriminative morphological representations for predicting the migration mode in a static cell

observation (or image). We employ bayesian CNNs for probabilistic prediction of mesenchymal cell migration modes from weakly annotated data.

2. IMAGE DATA

Image acq.: The high resolution images (1024×1024×3 px) of human non-small lung carcinoma (H1299) cells transfected with EGFP-paxillin (CMAC marker) and RubyRed-LifeAct (F-actin marker) are acquired using Nikon A1R confocal microscope with 60× objective. The cells were treated with 2.5 μg/ml and 10 μg/ml fibronectin (FN) concentrations to study the behavioral changes in the cancer cells. The images are acquired for 8–10 hr at 5 min intervals with a pixel resolution of 0.21 μm, resulting in 90-110 frames per time-lapse images. Altogether, the image dataset consists of 34 cells (2525 images) and 118 cells (6528 images) from 10 μg/ml and 2.5 μg/ml FN concentrations, respectively.

Annotation and data selection criteria: An expert biologist annotated each cell observation as either *discontinuous*, or *continuous* mode. The cell observations difficult to visually interpret were marked as with *unknown* label, and discarded from both training and testing phases. After filtering out of all unknown cells, we obtained 137 single cells (8648 images) to train and evaluate our method, covering 78, 54, and 5 cells labeled as to *continuous*, *discontinuous*, and *switching* mode, respectively. The *switching* cells consist images of both *continuous* and *discontinuous* modes. We randomly selected 102 (7048 images) and 31 (1600 images) cells for training and testing purposes, respectively.

Data augmentation: We augmented the 7048 training images by applying horizontal and vertical flipping, and multiple 90° rotations. The augmented images were further extended by including translation up to ±2 pixels from the centroid position in both *x*- and *y*- directions, resulting in 72 variations per image.

3. PROPOSED METHOD

preprocessing: Due to high variability in the cell sizes, the images were first downsampled to a constant and equal size (512×512×3 px) using Lanczos-3 kernel interpolation. For each annotated cell in each frame, we extracted 227×2227×3 px patch from the centroid to include sufficient contextual information as an input for the network model. The input images are then preprocessed by normalizing to zero mean and unit standard deviation.

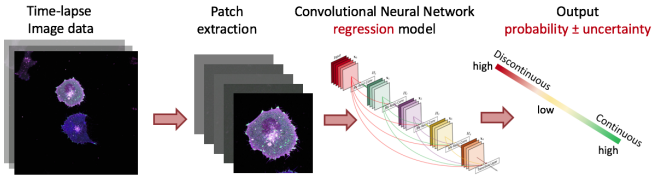

Fig. 1: The distribution of the predictions obtained by a classifier.

Weak supervision label assignment: Our annotations are noisy as manually characterizing the behavior of mesenchymal cell migration modes with a high confidence is challenging because of their inherent switching properties. Given the noisy behavior of the cells label, it is reasonable to describe the confidence regarding predictions instead of directly predicting their migration modalities. Motivated from [7], we thus introduce a weekly supervised learning based crtieria. Unlike the former approach, we focus on transforming the classification problem into a regression problem by genrating probablistic labels from the descrete frame-wise labels.

As we are provided with discrete frame-wise labels, i.e., $\{$*continous* and *discontinous*$\}$ for each single cell time-lapse sequence in the training phase, the goal is to predict probabilistic confindence for the training data, as well as for unseen testing data. To achieve that we linearly interpolate the frame-wise labels of each cell. Given a cell of length *n* and its labels of length *n'*, we linearly interpolate the labels in the continous range of 0 to 1 to assigning them to their corresponding *n'/n* frames. The range 0 - 0.4 corresponds to *continous* mode whereas the range 0.6-1 corresponds to *discontinous* mode; and both are linearly mapped for *n'/n* frames. The range between 0.41-0.59 corresponds to the region with less confident predictions, and also characterizing the conditions when cell starts switching migration modalities.

Bayesian Uncertainity: Uncertainties in Bayesian modeling are characterized as either aleatoric uncertainty capturing noise inherent in the observations or epistemic uncertainty accounting for model uncertainty [8]. Model uncertainty for a given image can be obtained by keeping the dropout mechanism switched on at test time and performing multiple predictions, which is an approximation to the posterior distribution based on Bernoulli approximated variational inference [4, 8].

We independently drop (with probability p_{drop}) the weights in all layers by drawing Monte Carlo (MC) sample from a Bernoulli distribution for each training sample x. The predictive uncertainity is estimated by:

$$\underbrace{\frac{1}{T}\sum_{t=1}^{T} diag(\hat{y}_t) - \hat{y}_t^{\otimes^2}}_{aleatoric} + \underbrace{\frac{1}{T}\sum_{t=1}^{T}(\hat{y}_t - \bar{y})^{\otimes^2}}_{epistemic}, \qquad (1)$$

where $\bar{y} = \sum_{t=1}^{T} \hat{y}_t/T$, $\hat{y}_t = Sigmoid\{f^{\theta_t}(x)\}$ and T refers to as the sampling rate. We fixed T= 50, as it was found in our case to be substantial for predictive mean estimation. The computation of MC samples for variational inference is fast and can also be applied to already trained networks.

3.1. Networks

We modified three existing networks, i.e., VGG16, Resnet 50, DenseNet, to our need for comparison purposes. We transformed each network model to be as a variational dropout network [9], i.e., each weight of a model has a dropout rate, for Bayesian approximation. This allows us to perform approximate but efficient Bayesian inference by straightforwardly using existing network models. Each convolutional layer was employed with $\mathcal{L}_2$ regularization to prevent overfitting, which is equivalent to putting a Gaussian prior on the network parameters, resulting in a maximum-a-posteriori (MAP) solution [4].

We employed the batch normalization operation specifically in the VGG16 network model. We replaced the fully-connected layers with global average polling operation, enabling us to use them with adaptive sizes of input. To obtain the regression output, we replaced softmax with sigmoid in the final activation layer. The dropout rate (p_{drop}) for the initial layer was fixed to 0.1, which grow exponentially at a rate of 0.05 for each subsequent layer. We found it to be a good compromise between getting a reasonable performance and uncertainty measures.

We trained the models in a 5-fold grouped stratified cross-validation scheme, ensuring that the images from the same cell are not represented in both training and testing sets. The number of images is larger for one migration mode than others, presenting a class imbalance during training. The stratified sampling also ensures that the relative class frequencies are approximately preserved in each fold to deal with class imbalance issues.

The training was performed for $29,000$ iterations with an initial learning rate of 0.01. The weights were initialized using Glorot normal distribution [**?**] and the biases were set to zeros. The weights were updated in a mini-batch of 4 samples using the ADAM optimizer. The network was optimized by minimizing the logarithmic hyperbolic cosine error as loss function and is defined as:

$$\mathcal{L}(y, \hat{y}) = \frac{1}{n} \sum_{i=1}^{n} \log(\cosh(y_i - \hat{y}_i)) \tag{2}$$

The network was implemented using Tensorflow backend in Keras, where the overall training took fifteen hours on a Titan X GPU.

4. EXPERIMENTS AND RESULTS

5. DISCUSSION AND CONCLUSION

The proposed model yields a substantial performance on the image dataset with different fibronectin concentration levels and thereby showing a great potential to be used as an automated method for large-scale live cells screening. However, whether the inclusion of temporal information improves the prediction performance or not is remained to be shown. Next, to reveal how and why CNN models can predict the migration modes, we will visualize the features of the cell images that were learned by the CNN models and contributed to their prediction: e.g., the protrusions and trailing edge. The proposed model yields the probability with an uncertainty that describes the confidence regarding predictions. If we manage to predict the migration behavior with just single frame then the phototoxity and staining can be overcome which is good in general.

Here, we demonstrate that a CNN can prospectively predict the migration sub-mode from a single image observation of a cell, and thereby overcoming the challenges of high photo-toxicity while minimizing the data acquisition efforts and data storage requirements. Furthermore, it addresses the inter- and intra-observer prediction variability by providing the model uncertainties while yielding high accuracy. For comprehensive and adequate quantitative analysis, this work focuses on harnessing the discriminative capabilities of deep learning based methods to perform real-time classification of two migration modes, and thereafter, to increase the understanding of underlying phenomenon that influences the cellular organization of each migration mode. Automation can help to reduce the noise induce by the multiple manual intervention. The study results showed the CNNs with automatically generated features can achieve desirable performance in cell migration mode prediction with just using a single cell observation.

This paper proposes an automatized approach for mesenchymal cell migration mode prediction. It discriminates candidates by encoding the morphological representations from the single cell observations (image patches), and also by encoding the spatiotemporal representation from image sequences. CNN and its variants offer several advantages in predicting cell migration modalities. The migration processes usually span several consecutive images, the migration modes can be predicted after considering both spatial and temporal information from adjacent image frames.

Given the spatiotemporal behvaior of imaging data, one preferred choice would have been exploiting recurrent neural networks (RNNs) to learn morphological and temporal representations. However, optimzing RNNs is complex due to the gradient vanishing and exploding problems, and computationally expensive and time-consuming training process over large time-lapse image sequences. Also, it is not trivial optimizing the additional hyperparameter, i.e., the ammount of temporal context needed for substantial perfromance gain.

6. ACKNOWLEDGMENTS

This work is partly funded by the Swedish Foundation for Strategic Research grant SB16-0046.

7. REFERENCES

[1] Hamdah Shafqat-Abbasi, Jacob M Kowalewski, et al., "An analysis toolbox to explore mesenchymal migration heterogeneity reveals adaptive switching between distinct modes," *Elife*, vol. 5, pp. e11384, 2016.

[2] Paola Masuzzo, Marleen Van Troys, Christophe Ampe, and Lennart Martens, "Taking aim at moving targets in computational cell migration," *Trends in cell biology*, vol. 26, no. 2, pp. 88–110, 2016.

[3] Anindya Gupta, Philip J Harrison, et al., "Deep learning in image cytometry: a review," *Cytometry Part A*, vol. 95, no. 4, pp. 366–380, 2019.

[4] Christian Leibig, Vaneeda Allken, et al., "Leveraging uncertainty information from deep neural networks for disease detection," *Scientific reports*, vol. 7, no. 1, pp. 17816, 2017.

[5] Zhi-Hua Zhou, "A brief introduction to weakly supervised learning," *National Science Review*, vol. 5, no. 1, pp. 44–53, 2017.

[6] Shori Nishimoto, Yuta Tokuoka, et al., "Predicting the future direction of cell movement with convolutional neural networks," *PloS one*, vol. 14, no. 9, pp. e0221245, 2019.

[7] Li Ding and Chenliang Xu, "Weakly-supervised action segmentation with iterative soft boundary assignment," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 6508–6516.

[8] Yarin Gal and Zoubin Ghahramani, "Dropout as a bayesian approximation: Representing model uncertainty in deep learning," in *international conference on machine learning*, 2016, pp. 1050–1059.

[9] Durk P Kingma, Tim Salimans, and Max Welling, "Variational dropout and the local reparameterization trick," in *Advances in Neural Information Processing Systems*, 2015, pp. 2575–2583.

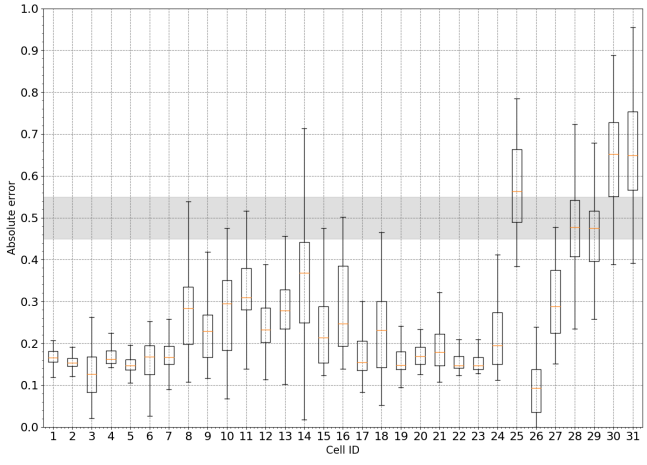

Fig. 2: The distribution of the predictions obtained by a classifier.

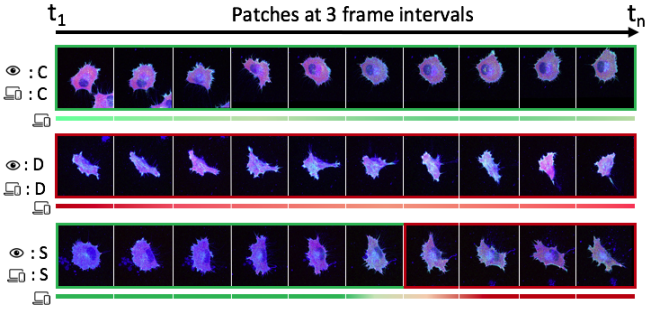

Fig. 3: 1a

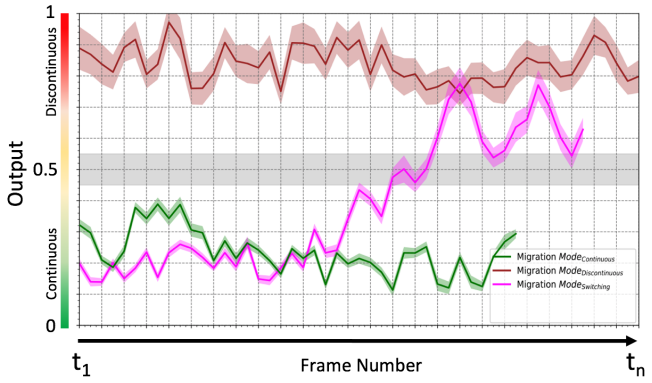

Fig. 4: 1a