A Cell Tracking Method for Dynamic Analysis of Immune Cells Based on Deep Learning

Aya Watanabe, Kenji Fujimoto, Hironori Shigeta, Shigeto Seno, Yutaka Uchida, Masaru Ishii, and Hideo Matsuda*

***Abstract*—Since the dynamics of immune cells change about phenomena in a living body, it is very important to observe and analyze cell dynamics *in vivo* in real-time. For this purpose, it is necessary to extract the information for the analysis by accurately tracking individual cells. As a method for this, general object tracking algorithms based on CNN (Convolutional Neural Networks) have been actively studied in the field of computer vision. However, in cell tracking, there are a large number of cells in fluorescent images that are similar in color and shape. It is not easy to recognize individual cells once they are lost due to overlap with other cells. Thus it is difficult to generate a large amount of training data with correct tracking trajectories. To cope with the problem of insufficient training data of cell images, our method extends the data by image processing and by assigning pseudo-labels. Furthermore, to obtain information more suitable for dynamic analysis, we propose to apply the re-identification function based on Euclidean distance. We demonstrate the effectiveness of our method with application to time-lapse images of immune cells against multiple inflammatory stimulations.**

***Index Terms*—Cell tracking, bioimaging, deep learning, cell image analysis**

I. Introduction

It has been biologically and medically important to know the behavior of cells in a living body. For example, by observing the skin, we can see how the movement of immune cells such as granulocytes changes depending on conditions such as acute inflammation or allergy [1], and macrophages change according to the conditions of inflammatory stimuli [2]. Thus, immune cells in vivo have a lot of information that indicates the progress of diseases and the efficacy of drugs, and analysis of the behavior of immune cells is expected to contribute to the elucidation of the causes of diseases and the development of new therapeutic drugs.

In analyzing the behavior of cells, recent developments in bioimaging technology, such as two-photon excitation microscopy, have made it possible to observe living cells *in vivo* without damaging them. This has made it possible not only to capture microscopic cellular activity but also to capture dynamic information with a time axis that could not be captured in the past. This has led to more in-depth research. On the other hand, however, there is a problem that manually analyzing the behavior of cells one by one from a large number of moving images produced every day is physically and mentally burdensome. In addition, human analysis is susceptible to biases caused by the observer's subjectivity, such as mistaken identification of cells due to preconceived notions and erroneous tracking due to fatigue. Therefore, the development of an automatic and objective analysis method using a computer is an urgent issue.

To analyze cell dynamics, it is first necessary to extract the information for the analysis by accurately tracking individual cells. As a method for this, general object tracking algorithms have been actively studied in the field of computer vision. Those algorithms track the motion of the entire image rather than a specific object by mapping it to a region segmentation [3], and optical flow [4]–[6]. On the other hand, cell tracking algorithms track a cell by using the cellular motion of the entire image rather than a specific object [7]. The methods often utilize machine learning based on the results of learning the data with their ground truth.

Moreover tracking is hampered by the strong noise inherent in fluorescent images, and segmentation [8], which recognizes regions, cannot be performed in the same way. In optical flow, there is a concern about tracking accuracy in terms of pixel-by-pixel tracking. On the other hand, machine-learning-based tracking can perform robust tracking against noise by learning the general features of the object to be tracked. In recent years, machine learning has made remarkable progress because it can learn the general features of the object to be tracked and can perform noise-robust tracking.

However, in cell tracking, there are a large number of cells in fluorescent images that are similar in color and shape. In addition, the cells to be tracked are different from general objects in that they move with deformation, called migration, and it is very difficult to distinguish each cell by appearance. Therefore, it is not easy to generate training data with correct tracking trajectories. Furthermore, since the camera can only capture images at a low frame rate due to its phototoxicity in live fluorescence microscopy, it is difficult to recognize the target cells again once they are lost due to overlap with other cells.

In this study, we propose a more effective method of cell tracking for cell dynamics analysis by extending Tracktor [9], which is one of the object tracking methods based on convolutional neural networks (CNNs) among machine learning methods. Our method attempts to improve the tracking accuracy by generating a large amount of high-quality training data through data augmentation to address the annotation problem of cell images for training. In addition, to address the problem in low-frame-rate time-lapse images, we attempt to track target cells over a longer period by introducing reidentification based on Euclidean distance. We then demonstrate the effectiveness of this method in analyzing cellular dynamics by tracking the movement of immune cells in response to multiple inflammatory stimuli.

Manuscript received August 17, 2022; revised October 7, 2022; accepted November 11, 2022. This work was supported in part by JSPS KAKENHI Grant Numbers JP21K19827 Japan.

A. Watanabe, K. Fujimoto, H. Shigeta, S. Seno, Y. Uchida, M. Ishii, and H. Matsuda are with Osaka University, Suita, Osaka 565-0871 Japan.

*Correspondence: matsuda@ist.osaka-u.ac.jp(H.M.)

II. Cell Tracking

A. Cell Images

In this paper, time-lapse images of mouse skin, in which inflammation was induced by multiple inflammatory stimuli, were taken every minute by two-photon excitation microscopy. Two-photon excitation microscopy is a microscope that induces a two-photon absorption process by increasing the density of photons during excitation light illumination and observes fluorescence through an excited state. It is suitable for observation of the inside of living organisms [10]. And these data sets capture the movement of immune cells activated by either of two types of stimuli, LPS (lipopolysaccharide) stimulation or GM-CSF (granulocyte monocyte colony-stimulating factor) stimulation. There are two datasets of LPS stimulation (91 frames each) and 11 datasets of GM-CSF stimulation (2 datasets with 91 frames and 9 datasets with 61 frames). The cell image data used in the experiment consists of 512[pixels]×512[pixels] of 0.5μm per pixel in the x- and y-directions and 15 depth layers of 3.0μm per layer in the z-direction.

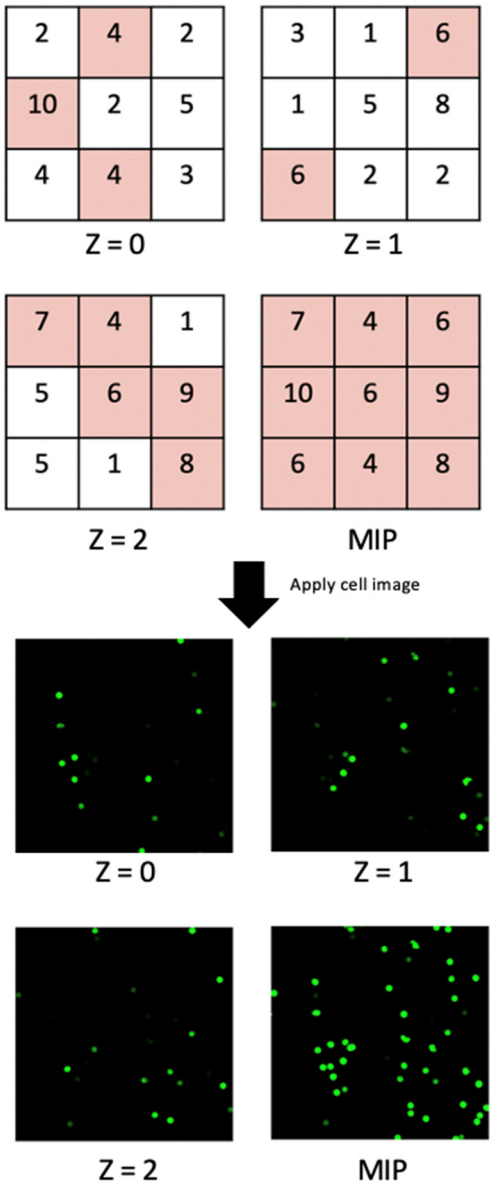

Fig. 1. Example of application of MIP to cell images.

We applied the maximum intensity projection (MIP) method to the images. The method is to project a three-dimensional structure onto a two-dimensional plane to facilitate recognition of the motion of an object of interest. The advantage of MIP is that the S/N ratio of the image is improved, and it is easier to detect cells and perform cell detection. Fig. 1 shows an example of MIP applied to cell images. The images consist of 15 slices on the z-axis and are projected by MIP, but in Fig. 1, only three images are extracted for convenience. In Fig. 1, immune cells are labeled green by EGFP (green fluorescent protein).

B. Tracktor

There are two types of CNN-based general object tracking methods: single-object tracking (SOT) [11, 12], which tracks one object at a time, and multi-object tracking (MOT) [13, 14], which tracks multiple objects simultaneously. While the SOT-based object tracking method predicts the position of an object in the next frame based on its initial position, the MOT-based tracking method uses an object detector, so there is no need to set initial values for the objects to be tracked, and it can respond to the appearance of new objects or the disappearance of objects. This makes it possible to respond to the appearance or disappearance of new objects. We propose a novel cell tracking method based on Tracktor [9], one of the MOT methods.

Tracktor consists of two steps: (i) detecting the position of an object in each frame separately and (ii) linking the corresponding detections over time to form a track. For each of these two steps, it is often necessary to build their training models. However, Tracktor utilizes the classifier and regressor in the object detector for tracking, allowing tracking to be performed without the need to build complex tracking models. The flow of the Tracktor algorithm applied to a cellular image is shown in Fig. 2. The performance of the method was demonstrated by applying it to object tracking using the MOT17 dataset [9].

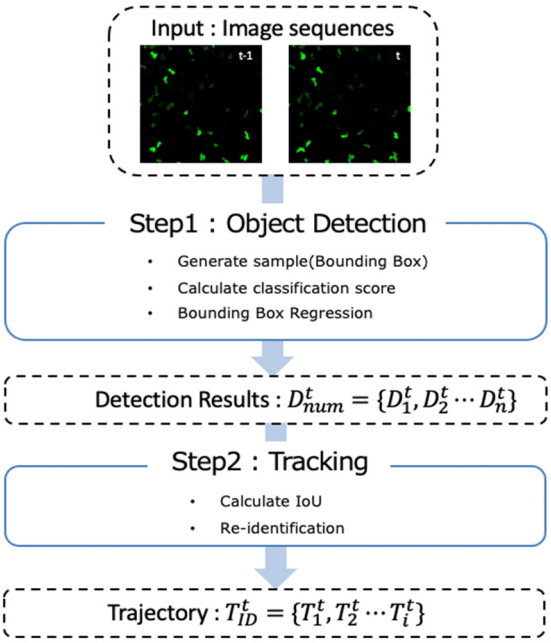

Fig. 2. A flow of tracking by Tracktor.

1) Object detection

The core part of Tracktor is an object detector that contains

a classifier and a regressor, which is based on Faster-RCNN [15]. Fig. 3 shows that the object detection step of Tracktor first takes a sequence of frame images as input and searches for cell *i*, the target of tracking, by sampling candidate rectangles around the position of the region where cell *i* was observed at time *t*-1 to estimate the position of cell *i* in the next frame $(B^1, B^2, \cdots, B^N)$. The resulting rectangle is called the sample and is used to estimate the position of the cell *i* to be tracked in the next frame. The resulting rectangles are called samples, and are denoted by (1).

$$B^n = \left(x_{left}, y_{top}, width, height\right) \tag{1}$$

Here, x_{left}, y_{top}, *width*, and *height* denote the x-coordinate of the upper left, y-coordinate of the upper left, width, and height of the rectangle, respectively.

Then, in the inner classifier part, among the samples $(B^1, B^2, \cdots, B^N)$ generated around the target, a positive score f^+ indicating foregrounds, and a negative score f^-, which indicates backgrounds. The sample candidate corresponding to the highest positive score is estimated as the optimal position in the next frame. Then, in the part of the regressor that encompasses the target, the location is estimated from multiple samples generated around the target, and since the rectangle that indicates the location of the target may not neatly surround the target, bounding box regression is applied to estimate the detailed location of the rectangle. Therefore, we apply the bounding box regression to estimate the detailed position of the rectangle.

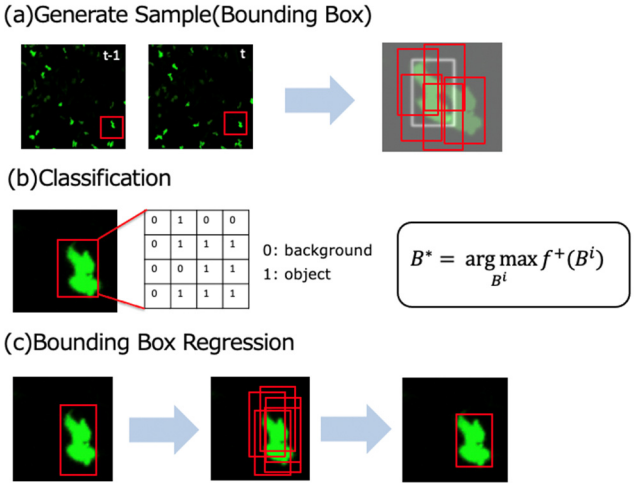

Fig. 3. A schema of object detection by image analysis. (a) Algorithm for generating sample(bounding box). (b) Algorithm for calculating object score. (c) Algorithm for regressing bounding box.

In applying Tracktor to cell tracking, there are a large number of cells with similar colors and shapes in cell images compared to general object images. Therefore, distinguishing the movement of individual cells from such images requires an enormous amount of time and effort. In other words, cell tracking has the problem that it is very difficult to label cells manually, and there is little training data. However, a sufficient amount of training data is essential to perform high-quality learning and accurate prediction. Therefore, we propose a method of repeating learning while increasing the amount of training data by data augmentation. This allows for high-quality learning and more accurate tracking.

2) Tracking

After cells are detected by the object detector, ID mapping is performed to determine whether they are the same object or not. At this time, the overlap rate in (2) is calculated as a criterion for determining whether or not they are the same object. Here r_a is the rectangle in the previous frame and r_b is the detected rectangle. If the overlap rate between the detected cell and all the cells predicted as the position of the next frame is less than 30%, the object is considered to have disappeared. If the overlap rate with all the cells detected in the previous frame in the detected object is less than 30%, the object is recognized as a new object.

$$\text{IoU} = \frac{r_a \cap r_b}{r_a \cup r_b} \tag{2}$$

However, a problem in applying Tracktor, which performs general object tracking, to cell tracking is the frame rate issue. Generally, when capturing the movement of cells, light is used to image internal biological information, but it is very difficult to capture cells *in vivo* at a high frame rate because there is concern that excessive irradiation may disrupt the internal functions of the tissues or the cells. Therefore, compared to general object images, which are usually captured with high frame rate video, the distance of the target cell traveled between frames is large, and if the target is temporarily lost due to overlap between cells or frame-out, the trajectory during tracking is interrupted as it is. In other words, it is difficult to continuously track the target cells over a long period, and it is impossible to obtain the information necessary for dynamic analysis, which continuously analyzes changes in the speed and direction of movement of the target. To cope with this issue, we adopt a re-identification function based on the Euclidean distance to re-recognize temporarily lost cells. This makes it possible to track individual cells over a long period.

C. Our Method

1) Data augmentation

In our method, only the green channel in each pixel of the input image is extracted as a preprocessing step to perform cell tracking independent of the background. This makes it easier to distinguish objects from the background and to detect cells. In addition, to increase the amount of missing training data, this method uses data augmentation by image processing and pseudo-labels. The data augmentation by image processing is a method of expanding the labeled training data, and generally includes image rotation, scaling, noise, and so on. In this study, image data were rotated by 90, 180, and 270 degrees, increasing the training data by a factor of four. Data augmentation with pseudo-labels is a method in which the prediction results with high confidence are used as training data. In this study, since long-term tracking is necessary for the analysis of cellular dynamics, we gave pseudo-labels to data which the predicted tracking is judged to be successful in more than 80% of all frames and we repeated the learning process by increasing the amount of training data.

2) Re-identification based on Euclidean distance

In our method, a reidentification function based on the Euclidean distance is introduced to track each targeted cell

for a longer period and obtain information suitable for dynamic analysis. The formula for calculating the Euclidean distance in this method is shown in (3). Where p_i represents the center position in the previous frame and q_i represents the center position of the detected cells. This makes it possible to re-identify and associate with IDs of cells that were temporarily undetected because they were partially hidden by other cells due to cell overlap, or because their trajectories were temporarily interrupted by frame-out.

$$D = \sqrt{\sum_{i=1}^{2}(q_i - p_i)^2} \tag{3}$$

III. Results

A. Evaluation Criteria

In this paper, we compare the tracking accuracy using a total of six indices [16] based on those in Table I, which are commonly used to evaluate performance in such two-class classification problems. Here, FP means false positive and FN means missed, while TP means successfully detected. In this experiment, "Successful detection (TP)" refers to the distance between a cell and one of the predicted trajectories in a given frame (here calculated based on IoU) that is less than a threshold value, which in this experiment is defined as 50%.

TABLE I: Evaluation Index of Binary(2-Class) Classification

		Correct trajectory (GT)	
		Positive	Negative
Predict trajectory	Positive	TP	FP
	Negative	FN	TN

Precision

Rate of correct trajectories among the predicted trajectories.

$$Precison = \frac{TP}{TP+FP} \tag{4}$$

Recall

Rate of correct predictions among the correct trajectories.

$$Recall = \frac{TP}{TP+FN} \tag{5}$$

IDSW (ID Switch)

The number of switched IDs for all data.

MT (Mostly Tracked)

Rate of mostly tracked trajectories, *i.e.,* ground-truth trajectories that are tracked as the same target at least 80% frames of their life span.

ML (Mostly Lost)

Rate of mostly lost trajectories, *i.e.*, ground-truth trajectories that are tracked at most 20% frames of their life span.

MOTA (Multi-Object Tracking Accuracy)

Rate of the correct prediction that was tracked correctly. Larger values indicate better results. g_t in (6) is the sum of the data of correct answers in frame *t*.

$$MOTA = \frac{FP+FN+IDSW}{g_t} \tag{6}$$

B. Datasets

1) MOT datasets

To test the usefulness of this method, we first used general object (person, car...) tracking datasets were used for the tracking. In this experiment, we used the MOT20 dataset. All of these datasets have different image sizes, each taken at 40-millisecond intervals.

2) Cell image datasets

In this paper, the cell image datasets are presented in Section II.A were used to verify the accuracy of our method. These data sets consist of 13 types (2 types of LPS stimuli and 11 types of GM-CSF stimuli) and each of the data were taken at 60-second intervals. In this paper, these 13 datasets were trained and tested by 5-fold cross-validation. The breakdown of these datasets is shown in Table II.

TABLE II: Breakdown of Cell Image Datasets for 5-Fold Cross-Validation

	Train	Test	
	Data name	Data name	The average number of cells per frame
Group1	01, 02, 03, 04, 06, 07, 08, 09, 11, 12, 13	05, 10	18.7
Group2	02, 03, 04, 05, 07, 08, 09, 10, 12, 13	01, 06, 11	33.3
Group3	01, 03, 04, 05, 06, 08, 09, 10, 11, 13	02, 07, 12	53.7
Group4	01, 02, 04, 05, 06, 07, 09, 10, 11, 12, 13	03, 08	43.2
Group5	01, 02, 03, 05, 06, 07, 08, 10, 11, 12	04, 09, 13	29.3

C. Cell Tracking Results

To compare the tracking accuracy of the method and to verify the usefulness of the method, tracking was performed here on a dataset of general object images as well as a cellular image dataset. Table III shows the results at this time, with higher values indicating better performance on the (↑) criterion and lower values indicating better performance on the (↓) criterion. The best results for each criterion are also shown in bold. Tables I and II show that our method improves the accuracy of cell tracking in 5 out of the total 6 indicators.

TABLE III: Evaluation of Cell Tracking Results on the Validation Set of MOT20 and Cell Images. The Best Results Are Shown in **bold**

	MOT20		Cell images	
	Tracktor	Our method	Tracktor	Our method
Precision↑	**87.9**	87.0	**93.1**	93.0
Recall↑	54.0	**56.5**	87.6	**91.4**
IDSW↓	10911	**10691**	457	**440**
MT↑	18.2	**20.9**	67.1	**77.9**
ML↓	19.5	**15.3**	9.18	**4.90**
MOTA↑	45.6	**47.1**	79.7	**83.0**

In particular, our method achieved better performance concerning the MT and ML criteria for both datasets. Thus, we confirmed that our method can continuously track cells for a longer time than existing Tracktor. Furthermore, Fig. 4

shows MOTA, which is commonly used to represent tracking accuracy, for each group of datasets. According to this figure, our method achieves better performance regardless of the number of cells in the image. Thus, we confirmed that our method can track more accurately than existing Tracktor even in areas crowded with cells.

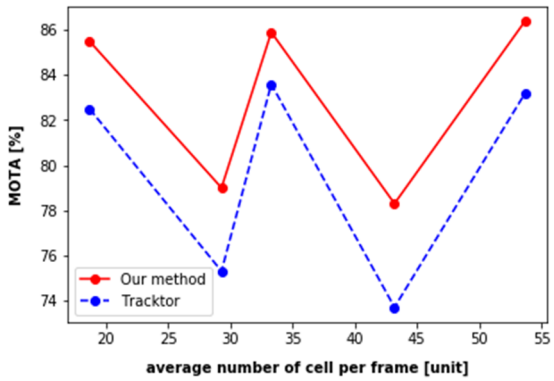

Fig. 4. MOTA for each group of datasets.

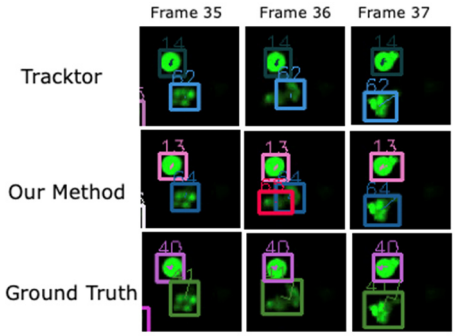

Fig. 5. Example results of FP.

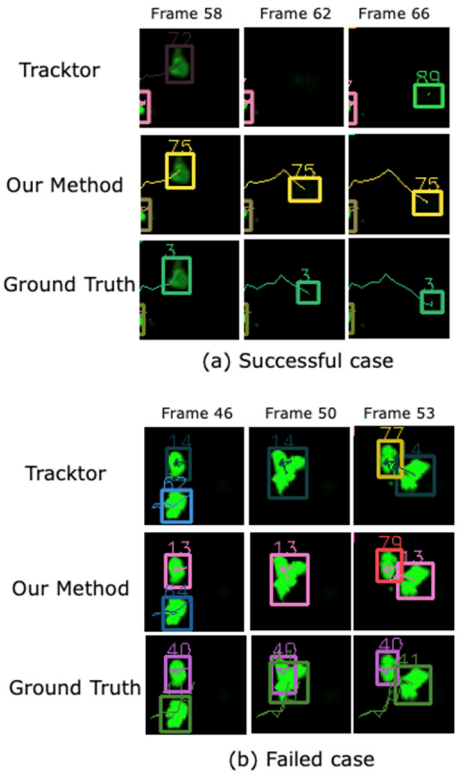

Fig. 6. Example results of our tracking method as shown in cell images.

However, in terms of the precision criterion, existing Tracktor achieved better performance. The reason for this is due to the increased FP. Fig. 5 shows an example of FPs. Frame 36 (middle) in Fig. 5, failed to capture cell migration and incorrectly detected cell deformation, and many FPs only temporarily captured cell deformation in this way. In cell tracking, it is more important to track the same target continuously than to track many targets accurately. Therefore, we consider continuous tracking more important than the reduction of the number of FPs.

Fig. 6 shows an example of the tracking results of our method. In Fig. 6 (a), MT is improved by reducing the number of missed cells in Frame 62 (middle) and associated IDSW. The improvement may be the effect of data augmentation of the training data.

However, In Fig. 6 (b), cells overlap each other in Frame 50 (middle) and another cell in Frame 53 (right). There are two possible reasons for this mistracking: the inability to detect individual cell overlaps and the failure to consider the motion vector from the previous frame of the target cell. Therefore, when objects overlap each other, as shown in Frame 50 in Fig. 6 (b), our method falsely detects them as a single cell. Furthermore, detection-based tracking, such as Tracktor, considers objects located close to the current position of the target cell in the next frame to be the same one. Therefore, in cases such as Fig. 6 (b), even though our method uses the reidentification function, the method may track the wrong cell due to similar location information.

IV. Conclusion and Future Work

In this paper, we propose an accurate cell tracking method for the analysis of cell dynamics. Our method tracks cells in images based on convolutional neural networks. We show that our method achieves better performance concerning continuous tracking of target cells, which is important for analyzing cell dynamics, compared to existing detection-based tracking methods.

However, the experimental results in Fig. 6 (b) show that our method misses the target. An intuitive way to solve this problem is to incorporate information about motion history during tracking, *i.e.,* to build a model during motion history to improve performance. To this end, we are considering incorporating the depth information lost during projection into the training in the detection step. By doing so, our method may achieve more accurate tracking.

The source code we developed is deposited at the GitHub repository (https://github.com/Hideo-Matsuda/LeukoTrack).

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

AW and HM conducted the research, analyzed the data, and wrote the paper; YU and MI provided imaging data with ground truth. KF, HS, and SS analyzed the data; all authors had approved the final version.

References

[1] D. Kreisel, R. G. Nava *et al.*, "In vivo two-photon imaging reveals monocyte-dependent neutrophil extravasation during pulmonary inflammation," *Proceedings of the National Academy of Sciences of the*

United States of America (PNAS), vol. 107, no. 42, pp. 18073-18078, 2010.
[2] S. McArdle, K. Buscher *et al.*, "Migratory and dancing macrophage subsets in atherosclerotic lesions," *Circulation Research*, vol. 125, no. 12, pp. 1038–1051, 2019.
[3] V. Paul, K. Michael *et al.*, "MOTS: Multi-object tracking and segmentation," in *Proc. The IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Long Beach, LA, June 16-20, 2019.
[4] B. Lucas and T. Kanade, "An iterative image registration technique with an application to stereo vision," in *Proc. International Joint Conference on Artificial Intelligence*, Vancouver, B.C., Canada, August 24-28, 1981, vol. 2, pp. 674–679.
[5] G. Franeback, "Two-frame motion estimation based on polynomial expansion," in *Proc. Scandinavian Conference on Image Analysis (SCIA)*, Halmstad, Sweden, June 29-July 2, 2003, pp. 363–370.
[6] H. Shigeta, S. Seno, S. Nishizawa *et al.*, "Analyzing Leukocyte migration trajectories by deformable image matching," in *Proc. 2019 IEEE 19th International Conference on Bioinformatics and Bioengineering (BIBE)*, pp. 94–98, Athens, Greece, October 28-30, 2019.
[7] E. Neda, S. Zahra, and F. Reza, "Computerized cell tracking: current methods, tools and challenges," *Visual Informatics*, vol. 5, no. 1, pp. 1-13, 2021.
[8] S. Carsen, W. Tim, M. Michalis, and P. Marius, "Cell-pose: A generalist algorithm for cellular segmentation," *Nature Methods*, vol. 18, pp. 100-106, 2021.
[9] P. Bergmann, T. Meinhard, and L. Leal-Taixe, "Tracking without Bells and Whistles," in *Proc. The IEEE/CVF International Conference on Computer Vision (ICCV)*, Seoul, Korea, Oct. 29-Nov. 2, 2019.
[10] M. Ishii, J. Egen *et al.*, "Sphingosine-1-phosphate mobilizes osteoclast precursors and regulates bone homeostasis," *Nature*, vol. 458, no. 7237, pp. 524–528, 2009.
[11] D. Martin, B. Goutam, S. Fahad, and F. Michael, "ATOM: Accurate Tracking by Overlap Maximization," in *Proc. The IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Long Beach, CA, June 16-20, 2019.
[12] Q. Wang, L. Zhang, L. Bertinetto, W. Hu, and P. Torr, "Fast online object tracking and segmentation: A unifying approach," in *Proc. The IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Long Beach, CA, June 16-20, 2019.
[13] W. Nicolai, B. Alex, and P. Dietrich, "Simple online and realtime tracking with a deep association metric," in *Proc. 2017 IEEE International Conference on Image Processing (ICIP)*, Beijing, China, September 17-20, 2017.
[14] F. Boukari and S. Makrogiannis, "Automated cell tracking using motion prediction-based matching and event handling," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 17, no. 3, pp. 959-971, 2018.
[15] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 6, pp. 1137-1149, 2017.
[16] J. H. Yoon, M. H. Yang, J. Lim, and K. J. Yoon, "Bayesian multi-object tracking using motion context from multiple objects," in *Proc. 2015 IEEE Winter Conference on Applications of Computer Vision*, Waikoloa Beach, HI, January 5-9, 2015, pp. 33-40.