Applying Deep Learning Techniques for Tracking Immune Cells

Submitted by

Marzieh Rahmani Moghadam

Bachelor of Electrical and Electronic Engineering

Master of Electronic Engineering

Submitted in fulfillment of the requirements for the degree of

Doctor of Philosophy

Department of Computer Science and Information Technology

School of Engineering and Mathematical Sciences

La Trobe University

Melbourne, Victoria, 3086

Australia

June 2020

Statement of Authorship

I clarify the thesis entitled “Applying Deep Learning Techniques for Tracking Immune Cells” submitted for the degree of Doctor of Philosophy is the result of my own work.

Except where reference is made in the text of the thesis, this contains no material published elsewhere or extracted in whole or in part from a thesis submitted for the award of any other degree or diploma. No other person's work has been used without due acknowledgment in the main text of the thesis. This thesis has not been submitted for the award of any degree or diploma in any other tertiary institution.

Marzieh Rahmani Moghadam

June 2020

Acknowledgements

Firstly, I would like to express my great appreciation to my principal supervisor, Professor Yi-Ping Phoebe Chen for her supervision during of my PhD studies. I am grateful for her valuable advice, encouragement, and supports that made this thesis possible. Also, my successes and experiences could not have been possible without such a great encouragement and guidance from my co-supervisor Associate Professor Zhen He.

I would like to sincerely acknowledge La Trobe University for providing a scholarship during my candidature. Also, I would like to extend my appreciation to the academic staff at Australian Regenerative Medicine Institute (ARMI), Monash University for providing biomedical imaging data, notably Professor Graham Lieschke who made this collaboration possible. I would also like to acknowledge the LIMS BioImaging Facility La Trobe Institute for Molecular Science for providing me with all the necessary bioimaging facilities.

Last, but certainly not least, I would like to thank my family for the support that they provide in my life. Also, I would like to express my appreciation to all my friends specially to Vahid Pazhakh and Zahra Mofrad for their valuable supports and encouragements during my PhD studies.

Table of Contents

List of Figures

List of Tables

Abstract

Recent advances in microscopy imaging techniques, capable of producing high resolution 3D time lapse images, have enabled biologists to observe biologic processes at unprecedented spatiotemporal resolution and at large scale. These advances range from development of optical properties of microscopes to improvements in fluorescent probes used for labelling different biologic structures such as immune cells. Leukocytes are cellular elements of the immune system that play crucial roles in defending organism against invading pathogens and studying their migratory pattern is a challenging task in cellular immunology.

Zebrafish as a vertebrate animal model have proven to be an optimal system for studying leukocyte biology, mostly due to the feasibility of intravital imaging of leukocyte behaviour within the live animal. As temporal and spatial analysis of these large imaging datasets by human operator is quite laborious, developing an automated cell tracking method is inevitable. However, the complex motion profile and amorphous nature of these cells make their detection and tracking on time lapse images extremely challenging.

Classical cell tracking techniques usually demand human intervention for pre-processing, feature selection, and parameter tuning. Additionally, large and complicated datasets could not be easily analysed by traditional techniques. Recently, with the development of deep learning as a powerful research area in the field of computer vision, advanced automated object tracking models have been introduced, which could learn robust features of objects with less or even no human interaction requirement.

This thesis focuses on effectively utilizing deep learning to tackle the common challenges involved in tracking highly amorphous and migratory leukocytes in a large-scale imaging dataset. The goal of this study is to develop more accurate automated deep learning-based cell tracking methods for analysing leukocyte trajectories in 2D and 3D fluorescence microscopy image sequences, minimising the requirement for human operator interaction.

We propose deep learning-based cell tracking methods (cell tracking by detection and cell tracking by linkage) with greater precision to calculate tracks of single and multiple cells by predicting the most probable matches from the adjacent frames within the image sequences. The proposed methods are trained through our generated real large training dataset, in which the robust spatial and temporal features of these highly mobile cells are extracted through 2D and 3D convolutional neural networks coupled with two additional learning channels. The evaluation of our deep learning-based cell tracking methods confirm their outstanding performance compared to other tracking methods.

To demonstrate the functionality of our method in the context of the biologic process, we implemented our techniques in analysing phagocytosis (engulfing large particles), as a major immune system defence mechanism. Also, at technical level, our method was employed to fix the problem of “fluorescence channel overlapping” as a common fluorescence microscopy imaging issue with cell detection and tracking. The experimental results of this study have proven the remarkable performance of our proposed cell tracking methods in rectifying common cell tracking problems compared to other state-of the-art methods in the field.

List of Publications

Moghadam, R. M. and Chen, Y. P. P. “Tracking Neutrophil Migration in Zebrafish Model Using Multi-Channel Feature Learning.” *IEEE Journal of Biomedical and Health Informatics*. (Accepted).

Moghadam, R. M. and Chen, Y. P. P. “Tracking Leukocytes in Intravital Time Lapse Images Using 3D Cell Association Learning Network.” *Artificial Intelligence in Medicine*. (Under Review).

Moghadam, R. M., Pazhakh, V., Lieschke, G. J., and Chen, Y. P. P. “Cell Tracking Repository for Biomedical Engineering Applications.” *Bioinformatics*. (Submitted).

Moghadam, R. M. and Chen, Y. P. P. “Applying Deep Learning-Based Cell Association Method on Biological Challenges.” *Gene*. (Submitted).

Chapter 1: Introduction

1.1 Motivation

The last couple of decades have witnessed the exponential growth of digital imaging techniques in medical applications. On the other hand, remarkable progress in developing *in vivo* models has made studying the biologic process in their natural physiologic environment easier than before. These advances have resulted in developing high-end intravital microscopy techniques capable of imaging live tissues using fluorescent probes, which allow researchers to monitor dynamic phenomena such as cell migration and provide the opportunity to produce high resolution multi-dimensional time lapse images. Despite all the advances in flexible biologic models and imaging techniques, the field still relies heavily on the visual interpretation of human operators to analyse complex imaging datasets. This gap between what humans and algorithms can differentiate is to some extent due to the greater complexity of biologic objects and their intrinsically dynamic behaviours. However, a visual examination of the large quantity of image data tends to be laborious and time consuming and can sometimes be subjective or biased by the expectations of the human operator performing the analysis.

To assist biologists in analysing multi-dimensional microscopy images, researchers in the field of computer vision have developed various algorithms aiming to enhance the accuracy of automated analysis of these complex data. The objective of these automated or semi-automated methods is to detect biologic objects such as cells in the 2D (x, y) or 3D (x, y, z) microscopy image sequences and calculate their trajectories. Often, the outlines of

the cells are also needed for the analysis. The task of producing these outputs is usually referred to as cell tracking. Once the tracks and the outlines have been computed, they can be analysed using algorithms that extract biologically interesting information, such as migration speeds, cell interactions, and categorising different cell populations.

Automated cell tracking algorithms can reduce the manual labour required and make the analysis more quantitative and reproducible. Furthermore, automated algorithms can be used to analyse subtle changes that are hard for a human to observe or quantify. A lot of effort has been put into the development of algorithms for automated cell tracking but handling massive datasets with tremendously divers features such as highly migratory and amorphous immune cells remain as challenging tasks. Thus, existing automated cell tracking methods are still considered error-prone for most microscopy datasets.

1.2 Aims

The main aim of this thesis is to use deep learning as an advanced computer vision technique to extract the robust spatial and temporal features of the microscopy image sequences to develop cell tracking methods with greater precision and minimum human operator interference. The project was conducted in close collaboration with cell biologists to ensure the identification of relevant problems could lead to the development of solutions that would work in real world scenarios. The research objectives of this thesis are as follows:

- Objective I: Construct a comprehensive data repository for learning methods.

Developing accurate deep learning-based cell tracking methods for studying leukocytes function in immune system is restricted by a lack of comprehensive ground truth datasets for training and validation. To address this limiting factor, a comprehensive leukocyte tracking data repository comprising a significant number of manually detected and tracked leukocytes was built.

- Objective II: Propose a learnable tracking method with a high level of accuracy.

Since the traditional cell tracking approaches cannot generate results with the expected accuracy and efficiency, it is necessary to apply new advanced computer vision approaches, such as deep learning, in the development of cell tracking models. A deep learning-based cell tracking method can extract the comprehensive features of complex cell behaviour from time lapse images, leading to improving the accuracy of cell detection and tracking. Therefore, a feature learning method is proposed to improve the 2D single cell detection and tracking accuracy by learning hidden characteristics of tracked cell in adjacent time points.

- Objective III: To effectively apply deep learning on accurate cell association.

The literature has proven that, despite the remarkable advances in cell detection, the cell tracking field still lacks powerful algorithms to accurately associate the multiple detected cells across time frames. The most challenging part of the cell association challenge is mostly related to the amorphous nature of cells, and their complicated motion profile through their migratory paths in 3D microscopy image sequences. Thus, the research objective is to propose a deep learning-based object

linkage method with high accuracy to address multi-cell tracking problems in relation to the number of tracks and track consistency.

- Objective IV: Apply our proposed cell tracking method to real world challenges.

  This research objective is to investigate the functionality of our proposed methods in the context of biological challenges such as analysing the phagocytosis procedure (i.e. engulfing large particles), as one of the major mechanisms by which leukocytes neutralise microbes, and also to fix the fluorescence channel overlapping issue, as one of the common fluorescence microscopy problems with cell detection and tracking.

1.3 Main Contributions

The main contributions of this work are described as follows:

- First, as part of this thesis, an extensive repository of immune cells imaging data has been prepared and properly labelled with a broad range of detection and tracking challenges. It is mostly used as training and evaluation dataset in deep learning-based methods. This repository has been constructed and labelled in close collaboration with a local Australian research group expert in leukocyte biology, and it is aimed to reflect the real challenges currently identified in the field.

- This thesis also makes a significant contribution to cellular immunology by reviewing classical cell tracking approaches and developing a novel learnable cell

detection and tracking model based on the proposed multi-channel feature learning (MCFL) algorithm. The MCFL model, inspired by deep learning, extracts the robust spatial and temporal features of 2D images of individual highly migratory leukocyte in adjacent time points within time lapse images. The evaluation of our learnable cell detection and tracking method confirms its outstanding performance (about 90% precision and 70% success accuracy) compared to other tracking methods.

- This thesis also proposes a novel deep learning-based cell linkage framework that contributes to immune cell identification and improves consistency in multi-cell tracking. The complex features of 3D fluorescent images of immune cells are learned through a 3D cell association learning network (3D CALN) using 3D convolutional neural networks. To perform multiple tracks, the trained 3D CALN finds the best paired matches between cells in two consecutive frames across the optimized samples. The evaluation results clearly demonstrate the superior performance of the proposed model with an optimized structure (about 95% accuracy in terms of tracking consistency).

- Lastly, this thesis provides practical contribution to the field of cellular immunology and biomedical imaging by adapting the proposed methods to real world challenges. The results demonstrate the functionality of our learnable tracking method for biological application and technical imaging problems suggested by our leukocyte biologist collaborators.

1.4 Thesis Outline

Chapter 2: This chapter provides a comprehensive literature review regarding cell detection and tracking techniques applied to common time lapse microscopy datasets of biological objects (whole cells or subcellular structures such as nucleus), along with different biomedical imaging techniques. This chapter also presents a brief review of the existing literature about deep learning approaches applied to cell detection and tracking. Finally, cell tracking evaluation methods are discussed in this chapter.

Chapter 3: This chapter describes the procedure for preparing high quality and large enough labelled training dataset from the raw microscopy images, as this is the critical component required for applying deep learning on cell tracking. The significant feature of our data repository is incorporating the most common cell tracking challenges in either the biomechanical or technical properties of a dataset.

Chapter 4: This chapter proposes a learnable cell detection and tracking framework based on the novel multi-channel feature learning model (MCFL), which is inspired by deep learning. The proposed model is designed using the convolutional neural network to learn the robust spatial and temporal features of neutrophils migration in time lapse 2D microscopy images. The cell tracking framework provide a set of candidates bounding boxes on the assigned cell through a sampling function based on the Particle Filter. Then it calculates the track by determining the most probable candidate cell in the next imaging frame. According to the evaluation results, the proposed framework has remarkable performance in cell tracking compared to other common state-of-the-art methods.

Chapter 5: This chapter presents a novel multi-cell association model that can accurately link multiple detected cells across time frames in a 4D dataset. The multi-cell association challenge in 3D microscopy images of immune cells is mostly related to the amorphous nature of cells, and their complicated motion profile in three dimensions and through their migratory paths. Therefore, a 3D cell association learning network (3D CALN) based on a deep learning approach is proposed to learn the cell complex shape and migration features such as morphological characteristics, rapid deformation and complex motion patterns. The comparison of our method to five other association methods has proved that our method has an outstanding performance in term of tracking consistency in multi-cell tracking, which makes it capable of handling the complex migratory patterns of numerous immune cells imaged in live animal models.

Chapter 6: This chapter focuses on the functionality of our proposed cell tracking method in the field of cellular immunology. Here, we have demonstrated how our tracking method can be applied to biomedical imaging datasets to identify and analyse biologic phenomena such as phagocytosis, as one of the major immune system mechanisms by which leukocytes neutralise microbes. On the other hand, this chapter also presents the capability of our proposed cell association model to fix the problem of fluorescence channel overlapping as one of the common fluorescence microscopy problems with cell detection and tracking methods.

Chapter 7: This chapter concludes this thesis and discusses the main challenges of this study, including some suggestions for similar studies in the field. Finally, future directions of this work are discussed at the end of the chapter.

Chapter 2: Literature Review

White blood cells (leukocytes) are the key cellular components of immune systems that play a crucial role through inflammation by defending organisms against invading microorganisms (microbes, pathogens) and also mediating wound healing [2]. Leukocytes function is highly dependent on their characteristics; most importantly, their exceptional migratory potent, which enables them to reach out to sites of infection in a short time [3]. As the mobility of leukocytes is so indispensable to their function, studying their migratory profile using different computational tracking methods is topical in the cellular immunology field [4]. Zebrafish infection models have been heavily in use to study phagocytic cells, neutrophils and macrophages, and responses to different microbial identities of pathogens. The major cellular inflammatory responses include migration toward site of infection or injury, phagocytosis, and secreting different anti-microbial materials. Traditionally, analysis of these anti-microbial activities of phagocytes imaged by complex microscopic techniques includes visually tracking of phagocytes in spatial (x, y, z in 3D imaging) and temporal (t) dimensions by a human expert (eye balling). Obviously, this type of manual analysis is time-consuming and laborious. Due to the ever-expanding size of biomedical datasets produced through large-scale imaging projects, the need for automated methods of image analysis is inevitable. Recent advances in computer vision and image processing techniques, along with developing of machine learning and deep learning techniques, have led to the development of automated cell tracking systems with the capability to morphologically dedifferentiate and track different cells. This chapter reviews the computer vision techniques adopted in medical image analysis, in particular,

for cell tracking. The review focuses on the challenges of immune cells, imaging techniques, and cell tracking in both classical and recent (deep learning) methods.

2.1 Introduction

Cells, the fundamental units of life, are the key actors in many biological processes. A significant part of our biologic knowledge has been acquired through observing biologic structure and processes at the microscopic level using different microscopes. Recent advances in imaging techniques have enabled biologists to observe biologic processes at unprecedented spatiotemporal resolution and at large scale. These advances range from the development of optical properties of microscopes to improvements in fluorescent probes used for labelling different biologic structures. Fluorescent proteins are a class of proteins that are excited by absorbing high-energy photons in a specific range and then releasing that energy by emitting a lower-energy visible photon in response [5]. Confocal microscopes are able to collect emission lights from different depths of the sample and form of a z-stack. Images taken from different z slices then could be reconstructed to form 3D images. Using this technology on live samples (intravital confocal imaging) allows us to study biological objects (e.g. leukocytes) in their physiologic state over time [6].

Leukocytes (white blood cells) are highly mobile cells that play a key role in defending organisms against invading microorganisms during infections. The function of these cells is highly associated with their migratory profile, which enables them to reach the site of infections in a short time. Studying the migratory profile of these cells as form of cellular immune response is a topical area in the field of inflammation research.

Zebrafish as a vertebrate model with a highly conserved genome and immune system similar to higher vertebrates has emerged in the biomedical research field. Their small size and optical transparency at the embryonic level have made zebrafish a favourable animal model particularly for *in vivo* visualising of biological processes including cellular inflammatory responses [7]. The zebrafish genome being fully sequenced along with feasibility of genetic manipulation have provided the opportunity to tag different cell types or biologic structure with different fluorescent proteins, which then could be easily visualised at high spatiotemporal resolution using state-of-the-art fluorescent microscopes [8].

Despite all of the remarkable advances in biologic models and microscopic techniques, with the ever-increasing size of imaging files and large-scale imaging datasets coupled with the possibility of imaging more complex biologic objects, automatic tracking techniques are still in their preliminary stages and image analysis tasks are still being performed manually, which is not only laborious and time consuming but also could be subject to bias and inconsistency.

In the field of computer vision, recent advances in digital computing technology have been taken on board by many automatic object tracking methods [9, 10] that aim to track and analyse an object over a sequence of complex three- (x, y, t) or four- (x, y, z, t) dimensional images. Despite all of these developments, when it comes to analysis of complex microscopic imaging datasets, the field still lacks powerful automated cell tracking algorithms capable of replacing human operators.

It is worth mentioning that the existing gap between what humans and algorithms can differentiate is partially due to more the complexity of biologic objects and their intrinsically dynamic status, which makes them highly amorphous structures compared to the usual macro objects conventionally tracked by these algorithms, such as cars or human. In other words, compared to traditionally often-solid tracked objects, biologic structures lack morphologically definitive features such as shape, size and angles, as they may interact with each other, split or merge or go through deformations such as general or polarised elongation, expansion in size or even shrinking [11], which may, in turn, also change their motion pattern over time [12].

The major aim in this field is to develop a powerful and fully automatic cell tracking algorithm that can analyse these highly dynamic and often-amorphous biological structures to provide the accurate results needed in biomedicine. Figure 2.1 illustrates the data acquisition procedure and our proposed automated tracking frameworks for leukocytes migrating in the zebrafish model.

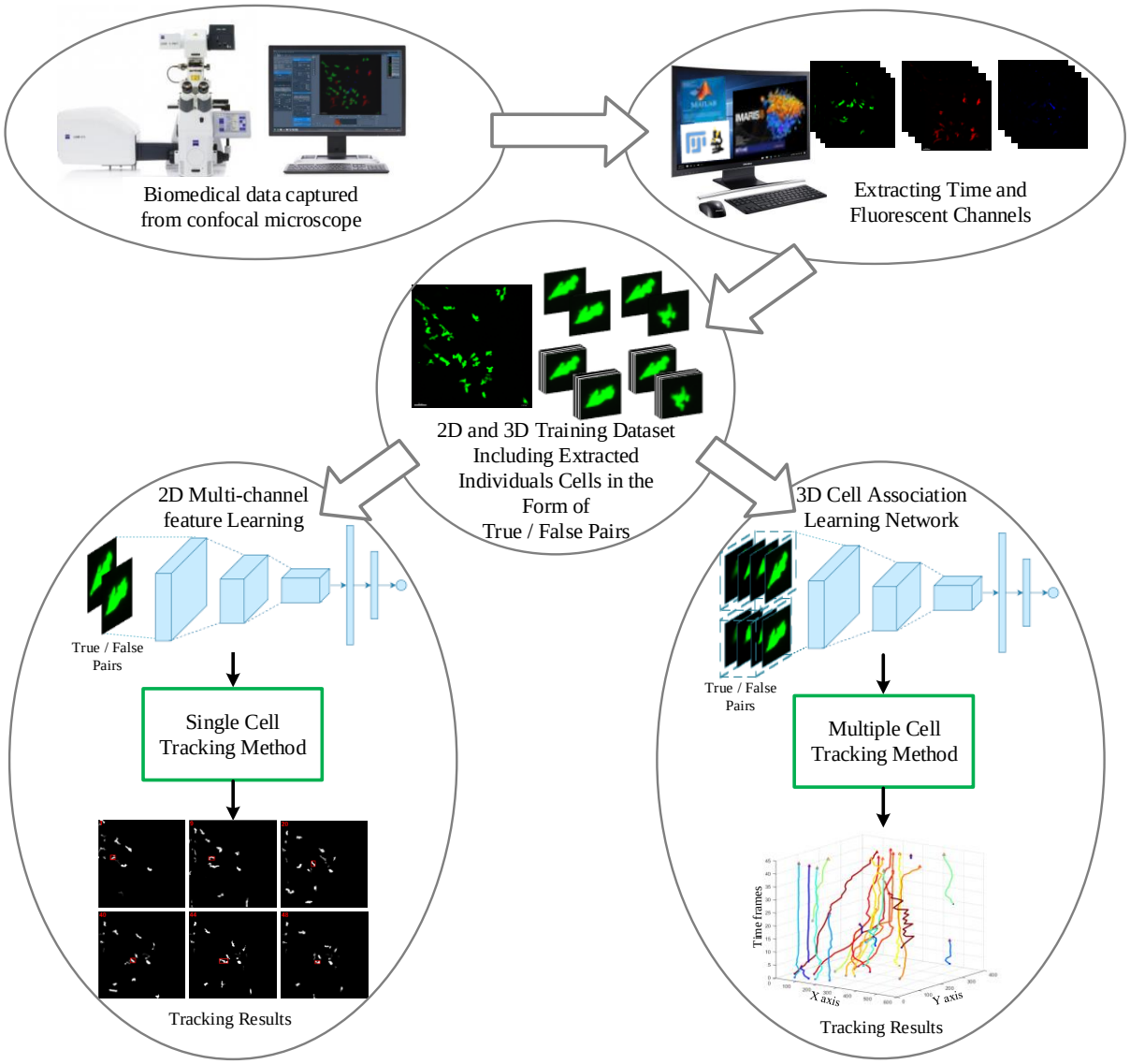


Figure 2.1: Overview of data acquisition procedure and proposed cell tracking frameworks.

This chapter provides a comprehensive review of the literature that led to the research questions addressed in this thesis. The literature review provides a thorough background on most appropriate tracking methods used in common time lapse microscopy images and presents arguments to support the research in this thesis. Firstly, cell characteristics, biologic systems and imaging techniques are briefly reviewed in sections 2.2, 2.3 and 2.4 respectively. Then the classical cell tracking approaches are introduced in section 2.5. In this section, we explain and compare the performance of different cell tracking methods in morphological cellular challenges such as deformation and complex motion. Next, we provide a complete comparison of recent tracking methods followed by a discussing of deep learning approaches applied in tracking in section 2.6. Lastly, the tracking evaluation parameters, conclusion and potential implications are discussed in sections 2.7 and 2.8, respectively.

2.2 Cell Types

One big difference between biologic particles, particularly cells, to other often solid objects traditionally tracked by tracking algorithms is the huge diversity and amorphous nature of biologic objects, which makes their tracking a super challenging task. Cells from different organisms or various tissues within the same organisms have their own unique characteristics and behaviour. For example, cell types such as fibroblasts or muscle stem cells, form clusters by adhering to each other and also to the substrate and are extremely hard to track, while others like haematopoietic stem cells are often non-adherent and show a Brownian motion pattern [13]. The other constraining factor is the density of cells, which

makes tracking even more challenging where there are too many cells in the field in close proximity of each other.

Phagocytes—including neutrophils and macrophages—are the main leukocytes (white blood cells) defending organisms against invader microorganisms as part of the innate immune response. To fulfil their patrolling job throughout the organism, these cells have evolved to migrate quickly to the site of inflammation and neutralise the foreign threat at site of invasion using their specialised microbicidal properties. So, their defensive mechanism, including releasing microbicidal granules or eating pathogenic particles, known as phagocytosis, is crucially dependant on their ability to relocate to the site of invasion. Despite the similarity between characteristics such as mobility and killing microbes, these two cell types are specialised in their own field and take different mechanisms to fulfil their duties. The main characteristics of these cell types, particularly neutrophils, is fast migration towards sites of inflammation and phagocytosing microbes.

2.3 Biologic systems

Biologic processes are usually studied either within a living organism (*in vivo*) or inside an experimentally conditioned environment in laboratory flasks (*in vitro*). Although *in vivo* experiments are the preferred type and generally produce more reliable results due to studying biologic processes in their natural physiologic environment, they have their natural constraints, such as cost, feasibility and ethical issues. In contrast, *in vitro* systems provide the opportunity to answer biologic questions in a cheaper, more controlled and ethically less challenging way, although quite often *in vitro* findings need to be confirmed in an *in vivo* set up [14].

2.4 Zebrafish Model

The zebrafish (*Danio rerio*) is a tropical teleost fish adopted as a vertebrate model organism to study different aspects of developmental biology [15]. The high conservation of developmental and physiological mechanisms between zebrafish and mammals (including humans), and the fundamental similarity of their genomes, underpin its applicability as a disease model [16]. The many attributes that support the use of zebrafish as an animal model include their optical clarity, rapid extra-uterine development, short breeding time, large clutches per mating, simplicity of maintenance and nurturing, availability of high density genetic maps, ease of mutagenesis, and the availability of many zebrafish models of monogenic human genetic diseases [16-20].

By expressing fluorescent proteins under the control of cell promoters (promoters are regulatory elements in the genome that control gene function), numerous transgenic zebrafish lines with specifically labelled cells and tissues have been generated, making feasible *in vivo* live imaging of different biologic processes and also tissue structures, in both transparent embryos and even adult fish [19, 21].

Transgenic zebrafish lines with fluorescent-labelled neutrophils and macrophages are examples of extant transgenic lines developed for this purpose (Figure 2.2). For this thesis, the zebrafish myeloperoxidase (*mpx*) [21, 22] promoters have been used.

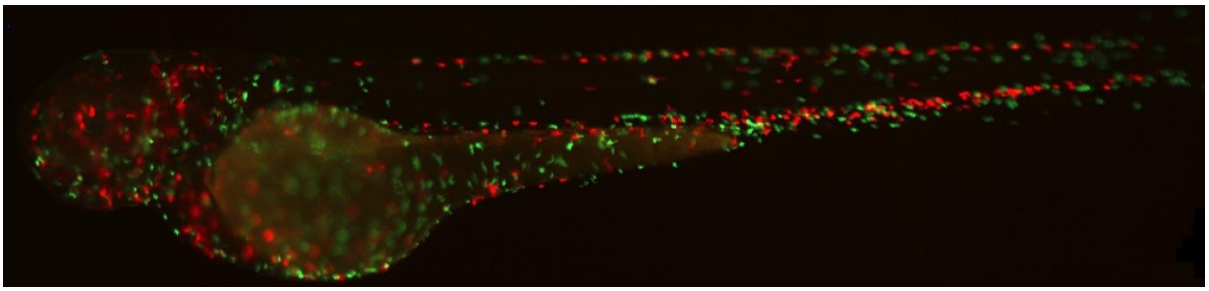


Figure 2.2: Transgenic zebrafish lines with fluorescent-labelled neutrophils and macrophages.

2.5 Medical Imaging Techniques

A wide range of microscopy techniques are available to be exploited to study different biological structures and processes, based on a range of factors including sample size, spatiotemporal resolution required to visualise the structure and process of interest, etc. In general, light microscopy techniques are divided into brightfield and fluorescence microscopy [23]. In brightfield microscopy, the sample is located between the objective lens and light source, and cells are visualised according to how they absorb, reflect or scatter passing light. As cells and majority of biologic samples are highly transparent and have minimum contrast to their background, phase shifting techniques (emphasising refractive index difference between cells and their surrounding environment) need to be applied before cells become visually differentiated from the background. The two most commonly used methods of phase shifting are “phase contrast” and “differential interference contrast (DIC)”, which make cells visible by providing a relatively darker background or pseudo–three-dimensional (3D) shaded effect on cells, respectively [24]. Brightfield microscopy is

mostly useful when cells are not labelled and causes minimum phototoxicity (light-induced cell damage).

In fluorescence microscopy, cells of biologic structures are labelled with fluorescent dyes (fluorophores) and then are visualised under florescent microscope. Fluorescent materials emit visible light (the emission wavelength) when they absorb a specific shorter wavelength light (the excitation wavelength), usually produced by specific lasers. In fluorescence microscope usually the same objective is used for both excitation and emission lights, so it is called epifluorescence [23].The shorter excitation light is the more energy it has, so emitting cells by a high energy laser may cause photobleaching (i.e. Fluorescent dye fading over time). So, high power exciting lights, such as UV, have higher phototoxicity and should be used with caution [25].

EGFP and mCherry, which label neutrophils and macrophages in Tg(mpeg1:mCherry/mpx:EGFP) [26] zebrafish embryos (Figure 2.2), have excitation and emission peaks of 488/509 nm and 587/610 nm, respectively [27].

2.6 Classical Tracking Methods

Cell tracking, video security surveillance, sports videos and autonomous-driving are among the traditional applications for visual tracking, also known as object tracking, in the field of computer vision [28, 29]. Simply, object tracking means automatically estimating objects' trajectory when they move in sequential video frames. Visual object tracking (VOT) and multiple object tracking (MOT) are two main branches of object tracking. Research in the field of VOT focus on tracking a single instance of an object (e.g. person,

animal, vehicle, etc.), usually annotated by a bounding box in the initial frame of a video. It focuses on the appearance and motion models that have to be learned on-line[30]. MOT, on the other hand, executes tracking of multiple instances of a single type of object while entering or leaving the field of view. MOT mainly improves the accuracy of association between detected objects in adjacent frames[31].

Applying object tracking, in the form of automated cell tracking, to analyse biomedical images including morphological assessments of different cell types, wound regeneration, metastatic malignancies and immune responses are among the topical fields in biological studies [32]. Due to the natural complexity of biological objects (both shape and motion), automated cell tracking is a quite challenging task compared to other tracking algorithms. Cell tracking challenges could be outlined into four categories (Table 2.1). The physical properties of biologic objects such as cell size or motion velocity, and also meaningful physiologic processes such as phagocytosis (i.e. eating other objects by immune cells) or cell divisions, are analysed by applying cell tracking methods.

Cell tracking methods have been exploited to address each of these challenges in a specific manner. In other words, cellular properties and motion profiles have been used for optimised tracking methods accordingly [33, 34]. Conventional classical cell tracking approaches are usually categorised in two classes: (1) cell tracking by model evaluation and (2) cell tracking by detection. The model evaluation and detection-based tracking methods utilise the temporal property of object of interest to enhance the accuracy of cell boundaries and tracks, respectively.

TABLE 2.1: CELL TRACKING CHALLENGES

Cell challenge	Definition
Cell deformation	Cells are non-rigid bodies, so they always change their shapes with time, such as elongation expansion, and shrinkage.
Cell complex behaviour	Cell migration entails complex motion with multiple modes.
Cell's living environment	There are many particles in the cytochylema, which contains dead cells, germs and other organic material.
Imaging dataset complexity	4D or 5D microscopy data including low resolution and massive images that sometimes contain many cells.

2.6.1 Cell Tracking by Model Evaluation

In model evolution algorithms [35-37], detection and tracking of each cell happens simultaneously by transmitting a mathematical model for each cell over sequential time points. In this scenario, the output data from each frame act as the basis for analysing the same object in the next frame. According to the mathematical formulation, model evolution algorithms are divided in two categories, explicit (aka parametric) or implicit (aka level set) methods. In the explicit model, the counter is expressed through a parametric curve or surface [35, 38, 39], while in the level set method [29], the contour implicitly is expressed as the zero-level of a higher-dimensional Lipschitz function [40, 41]. These methods are preferred for studying cell morphology in detail in samples imaged at a high spatiotemporal resolution.

2.6.1.1 Parametric Models

The active contour model, also known as snakes [39], works based on drawing an initial contour around the object, followed by sequential contour deformation until it overlaps with the object boundary. To perform tracking on the segmented object in this method, the formed contour will be used to initialize the model on the following frames. Although active contour-based techniques could potentially lead to better approximations of object morphology in a short time, their dependency on parametrization and low efficiency in cases of contour splitting and merging (absence of topological flexibility) are still considered as their major disadvantages [42, 43].

To improve cell tracking through better handling of topological operators such as cell divisions and also fine cell-cell contacts, Zimmer *et al.* [44] exploited the classic ‘‘snake” model through addition of repulsive forces between snakes. To overcome multi-cell segmentation and 3D tracking of fluorescent cells, Dufour *et al.* [45] suggested a deformable model based on 3D active meshes, in which the deformation problem is solved by expressing the surface in discrete domain. The drawbacks of this approach are the restrictive cost of computational overhead while tracking a large number of cells and the inability of the method to handle new cells entering the field of imaging.

2.6.1.2 Level Sets

In level sets-based approaches [37, 40, 46], multiple cells are tracked automatically based on their morphology and intensity profile. The cell edges are recognised as a closed curve according to the defined energy function. These energy functions are then minimised

according to the intensity coherency and spatial consistency of detected cells over sequential frames to perform tracking [40]. In these methods, the fronts are defined as the zero-level set of a scalar function defined on the image domain [47-49] and Eulerian equations are utilised to compute their evolution. The evolution in this method occurs naturally through splitting or merging the fronts, which leads to segmentation of disconnected objects originated from random initial fronts. This is how this method is capable of automated initialization, detection of incoming cells, and also handling of dividing cells. Noticeably, this type of segmentation-based object tracking does not implement the temporal consistency of the object, which may decrease the efficiency of tracking in case of dramatic spatial changes. In other level sets-based segmentation-tracking methods [50], post-processing analysis has been proposed, instead of frame-by-frame processing, aiming for better distinguishing between cell clusters and to form cell trajectories leading to more accurate cell segmentation, which obviously demands more computational power and memory.

TABLE 2.2: ADVANTAGES AND LIMITATIONS OF CELL TRACKING BY MODEL EVALUATION

Model Evaluation Methods	Advantages	Limitation
Parametric Models	• Better approximations of object morphology. • Combination of cell segmentation and tracking stages.	• Only tracks smooth cell motions. • Has difficulty in coping with changes in topology. • Highly dependent on the initial placement of the contour. • Lack of tracking multiple cells.
Level sets	• Naturally capture topological changes, such as cell division. • Automatic initialization. • Detects newly appeared cells.	• High computational cost. • Requires postprocessing steps. • Lack of temporal coherency. • Weak segmentation and tracking on poor contrast images.

2.6.2 Cell Tracking by Detection

This approach includes segmentation and tracking, initiated through segmentation of cells in single frames and then exploiting probabilistic frameworks to form the temporal associations between segmented cells in sequential frames of the video file, by either using a two-frame or multi-frame sliding window, or even for all frames at once.

Basically, segmentation includes differentiating meaningful objects (foreground) from the background and assigning each individual pixel into either of these categories. It is performed by comparing each pixel value to the segment pre-set threshold and labelling it with the difference by which it falls below or above the threshold [51, 52]. The simplicity of this approach makes it preferable over other more complicated segmentation methods such as applying predetermined cell density profiles [53], which obviously is more efficient when cell morphologies are more consistent.

Watershed transform is another fairly common image segmentation method, in which images are treated as a topographic relief. From mathematical morphology, a relief function represents edge evidence, usually computed through the morphological gradient. The function treats the images under process like a topographic map of a mountain landscape, in which object boundaries are defined as watershed lines. Thus, this relief function from the regional intensity minima divides the image into areas and delimiting contours. A variety of pre- and post-processing modifications, such as using markers or region merging, have been applied to make this segmentation approach more compatible for biologic object microscopy images [54, 55].

In this approach, object detection is then followed by object linkage [56]. Nearest neighbour, probabilistic models, and linear programming are among classical object linkage methods.

2.6.2.1 Nearest Neighbour Methods

The nearest neighbour approach simply links every segmented object in a given frame to the nearest object in the following frame [57]. In other words, a linking algorithm matches all particle coordinates frame by frame-by-form the final track [58]. When the object density is low in the field of imaging, the linking task is performed easily in an automated sequential way using separation-based criterion including nearest neighbour. In this approach, distances between all possible paired objects in the following frames are calculated, and then a pair with the shortest distance are considered as the actual pair and are linked together. By repeating this process for all detected objects, all trajectories in the field are formed, although in real world experimental conditions there are always some errors in tracks constructed by this method, due to the intrinsic complexity of biologic objects' motion profiles.

2.6.2.2. Probabilistic Models

Generally, probabilistic tracking methods [59-62] employ the Bayesian estimation concept to infer the true state of a dynamic system from noisy measurements of that system. In this approach, the "state" is defined by a mathematical vector, which collects and utilises all information including object position, shape, velocity, and intensity. The estimation is

composed of two steps, “state prediction” and “state update”, which employ dynamics and observation models, respectively. The dynamic model is based on a priori knowledge of state variation while the observation models is based on a priori knowledge of the measurements given a particular system state. In this approach, Kalman Filtering and sequential Monte Carlo schemes are commonly utilised for computational analysis of linear or Gaussian and non-linear or non-Gaussian cases, respectively. Ngoc S. *et al.* [62] proposed a tracking method that relies on the correlation of the image with a filter that changes in an adaptative manner to represent an object as it moves and deforms. To predict and update the objects position, this approach utilises a Kalman Filter and a cost function [62], which improves the reliability of prediction for the object position at time $t + l$ based on the knowledge of its past positions at t. This method has shown its functionality when applied to a sequence of 800 DIC microscopy images of a latex micro bead moving beside an axon. Kalman Filtering is an efficient way to deal with multi-target tracking as long as small number of objects are in the field [63-65], but the efficiency drops when the number of objects increases, mostly due to the recursive nature of this filter. This weakness has been partially improved by employing Particle Filtering, which engages multiple hypotheses [66-68] and was tested to track multiple hockey players[69], or several people in the ground and image planes simultaneously [70], although its application is limited as it requires careful tuning of several meta-parameters. Recently, the multi-Bernoulli filter, inspired by ant colonies, has been used to analyse cell behaviour dynamics in a pool of cells performing mitosis division and going through morphological changes [71].

The Bayesian framework is employed to complement the active contours method in probabilistic contour tracking algorithms [72, 73]. The main idea in this cell tracking approach is to predict the global motion and determinant features of the object in a

probabilistic way. Kalman Filter or related methods such as Particle Filters [74] are usually used for this aim, assuming that object deformations are minimal, a feature which is not common in dividing cells and biologic objects in general.

2.6.2.3 Linear Programming

Object association in linear-programming methods is performed based on a similarity score computed by combining a collection of features such as centroids, intensity and morphology [75-77]. Compared with Probabilistic-based tracking approaches, such as Kalman Filtering [64] or Particle Filtering [74], motion is not usually taken into account while optimizing the similarity score in linear programming. This is why dealing with long-term occlusions in this approach is challenging. In this regard, Jaqaman *et al.* [78] demonstrated how tasks such as frame-by-frame association of objects and also recognising newly appeared (i.e. cell division) and disappeared objects (i.e. merging), could be dealt with by solving a linear-assignment problem, although the long term occlusion problem still remains. Different software packages for tracking particles and cells such as uTrack [78], CellProfiler [79, 80] and TrackMate [81] have been developed based on the same approach.

TABLE 2.3: ADVANTAGES AND LIMITATIONS OF CELL TRACKING BY DETECTION

Tracking by Detection Methods	Advantages	Limitation
Nearest neighbour	• Low computational complexity. • Simple to implement.	• Easily lead to mismatches for highly mobile cells or in fields with a high cell density. • Hard to solve the problems of cell disappearance, splitting and merging.
Probabilistic Models	• Tracks moving and deforming objects. • Deals with partial occlusions. • Near real-time speeds.	• Lack of tracking highly deformable objects. • Lack of tracking highly mobile cells. • Lack of accurate tracking in high cell density field.
Linear programming	• Deals with complex cell behaviours such as disappearance, splitting and merging. • Deals with complex track interactions. • Not dependent on spatial features.	• Lack of considering long term occlusions. • Lack of tracking highly mobile cells.

2.7 Deep Learning and its Applications in Biological Object Tracking.

In the deep learning approach [82], several processing layers of computational models are utilised to learn the characteristic features of datasets through numerous magnitudes of abstraction. These approaches have proven to significantly improve the efficiency of sophisticated tasks in different fields including object detection [83], object tracking [28, 84, 85], visual object recognition [86], speech recognition [87], and even biological fields like drug discovery and genomics [88].

2.7.1 Deep Learning

"Deep learning" refers to a set of machine-learning techniques, specifically neural networks, which has significantly extended the range of problems that could be solved by computer vision. Compared with conventional machine learning, in which representations are manually designed by feature engineering, the learning process in deep learning can be supervised or unsupervised. In supervised learning (section 2.7.1.2), which is the most successful approach, an annotated dataset is provided for the algorithm to learn from, while in the unsupervised approach, known as self-organization, the algorithm extract features and patterns on unlabelled data for modelling of probability density.

Deep learning uses the backpropagation algorithm (section 2.7.1.3) to identify complex features in large data sets by optimising internal parameters used for computing the representation in each layer from the representation in the previous layer. While deep convolutional neural networks (section 2.7.1.4) have been mostly implemented in

processing images, video, speech and audio, recurrent networks have been mainly useful for sequential data such as text and speech.

In this section, a brief overview and recent history of deep learning is presented. More detailed historical background and an overview can be found in [82].

2.7.1.1 Main and Background

Defining some basic concepts are essential before describing the deep learning method [82]. To be able to detect and classify objects, a deep learning algorithm needs to identify characteristic features of the objects, and this task happens through employing representation learning methods. These representation learning methods are composed of several representation levels, formed through formulating non-linear modules [89]. Firstly, each module processes the raw input data to a slightly more abstract level through transforming the representation at a given level. Highly complex functions could be learned in this way. Then, through classification tasks, the most biased aspects of input are emphasised in a higher level of representation, while irrelevant variations are filtered out. For an image composed of arrays of pixels, features learnt as the first layer of representation usually are boundaries at a particular orientation or location, typically learned at the first layer of representation, while more specific features such as arrangements of edges, regardless of their location, are detected through the second layer of representation. The third layer incudes collection of patterns in first and second layers, which corresponds to the characteristics of similar objects. Subsequent layers formed by applying the same basics would employ the collection of represented features to perform the object detection task. Learning characteristic features of the data by employing a general-purpose learning

procedure, which is not designed by a human operator, is the striking difference between deep learning and other image analysis approaches.

2.7.1.2 Supervised Learning

Supervised learning, which is basically a training machine with a well annotated dataset, is the most common type of machine learning. For example, if the machine is to be trained to detect images of different objects like cars, houses, trees, etc., a large dataset containing these images needs to be collected and labelled as different categories (car, house, tree, etc.). While being trained, the machine is fed with images of one category and produces a vector of scores as output for that category. Ideally, the desired category should the highest score, which is unlikely to be achieved before training. An objective function is computed to measure the difference between the desired pattern of scores and output scores. This difference is then minimised by modifying the internal adjustable parameters, which are real numbers, usually called weights. Linear classifier is employed for practical applications of deep learning, which computes a weighted sum of the feature vector elements and classifies an input to a category when the weighted sum is above a threshold.

2.7.1.3 Backpropagation

The backpropagation algorithm [82] is the procedure to calculate the gradient of the loss function relative to each weight. Backpropagation is short for the backward propagation of errors as the error is calculated at the output and distributed backwards all over the network's layers. In other words, the gradient for the final layer of weights is computed

first, and conversely, the first layer's gradient of weights is calculated last. In this method, partial calculations of the gradient from one layer are reused to ascertain the gradient for the previous layer. Compared to the conventional method of calculating the gradient of each individual layer independently, this backwards error checking provides the opportunity to calculate the gradient at each layer more efficiently.

2.7.1.4 Convolutional Neural Network

The convolutional neural network (CNN) [90], as a class of deep neural networks, is designed to process datasets in the form of multiple layers (Figure 2.3). CNN have been designed to exploit local connections, shared weights, pooling, and the use of many layers as properties of natural signals to learn high level representations of image datasets. CNN is commonly formed in a series of stages, in which the first few stages are composed of "convolutional layers" and "pooling layers". By going through a convolutional layer, units are abstracted and organized in feature maps, in a way that each unit is correlated to local patches in the feature maps of the previous layer via a set of weights called a filter bank. The output of this local weighted amount is then passed through a non-linearity activation function such as a rectified linear unit (ReLU). Although all units within a feature map share the same filter bank, various feature maps in a layer employ different filter banks. This architecture has two immediate advantages. First, easy detection of characteristic local motifs is formed by highly correlated local groups of values in array data (e.g. images). Second, the local image statistics and other signals do not vary relative to location. So, units at different locations share the same weights and the same pattern could be detected in different parts of the array.

While the convolutional layer detects local conjunctions of features taken from the previous layer, similar features are merged into one by the pooling layer. Coarse-graining of each feature position could be used to improve the motif detection regardless of the slightly varying position of features within each motif. Commonly, the pooling unit computes the maximum of a local patch of units in one or a few feature maps. By taking input from patches that are shifted by more than one row or column, neighbouring pooling units reduce the dimension of the representation and subsequently create an invariance to small distortions. Usually, two or three stages of convolution, non-linearity and pooling are stacked. This structure is then followed by more convolutional and fully connected layers. Backpropagating gradient algorithms employed in CNN and regular deep networks are similar in basics and promote training of all weights in filter banks.

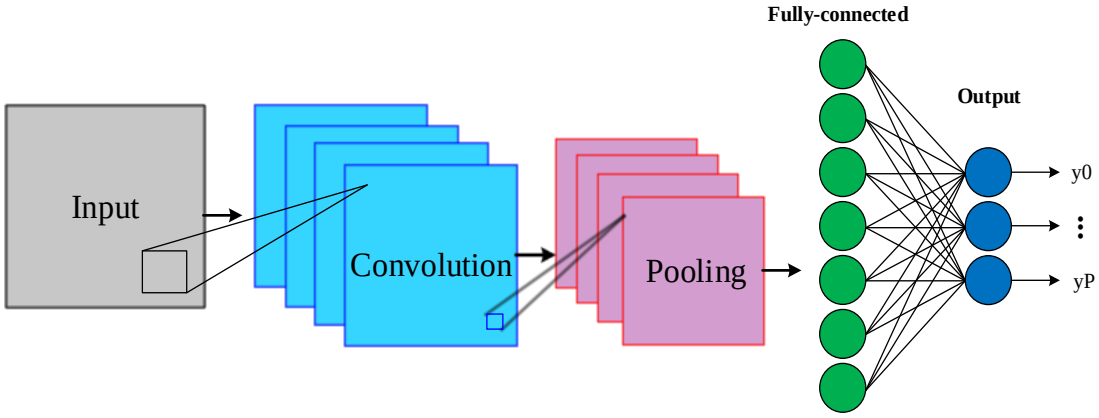


Figure 2.3: Framework of convolutional neural network including input layer, convolutional layers, pooling layers, fully-connected layer, and output layer ($y_0 \dots y_P$).

2.7.2. Deep Learning-Based Cell Tracking

Deep learning-based visual tracking is a topical field in computer vision and image processing, which has various applications in navigation, surveillance, robotics, traffic control and other related fields. Multiple deep learning-based trackers such as FCNT [91], SANet [92], TCNN [93] and MDNet [94] have proven their potential to remarkably improve the tracking performance. As an example, tracking performance on the OTB-100 [95] dataset is continuously improved by deep learning-based tracking approaches such as DeepSRDCF [96] and HDT [97].

Generally, deep learning techniques have been applied on medical [98-101] and biological sciences [102-105] at much lower rate compared to other natural images, mostly due to the intrinsic complexity and divergent features of biomedical images and also lack of large publicly available datasets. Generally, deep learning-based tracking methods are divided in two categories: tracking by object detection [106-111] and tracking by object linkage [112, 113].

2.7.2.1 Cell Tracking by Object Detection Using Deep Learning

This method views tracking as a real-time processing procedure. in this approach, location of the target object in the first frame is required by algorithm, then the object is tracked frame-to-frame, while the algorithm simultaneously updates the tracking model and rules. As tracking efficiency in this model is highly dependent on accurate detection of the object in every frame, deep learning is utilised to improve this task. In these methods, cell tracking problems are solved by feature learning [114-116] and employing a classifier to

differentiate cells or particles from the background. For example, to improve object detection and subsequently particle tracking in [106], convolutional neural network (CNN) [90, 116], has been utilised to learn the robust features of biologic objects in both 2D and 3D imaging dataset. Detected particles were then linked through the centre of mass method. Application of this approach is limited to samples with minimal particle dislocation between image frames[106]. In other example of deep learning application for biologic object detection [37], a CNN-based multi-task learning observation model was matched with an optimized model update strategy to be able to deal with the variation of the tracked cell over the tracking course. The constraint of this method is that its training is only limited to cell morphology and motion profile is not considered in training, thus fast dislocation often causes tracking being lost. In another study [112] convolutional and recurrent layers have been used to perform tracking according to temporal information including object appearance and motion. In this method, a pair of cropped objects are given to the network. The recurrent layers, acts as the tracker, is then updated with a single forward pass. This approach is also limited to small relocation distance of moving object. The other application of deep learning is improving tracking accuracy by detection of rare events such as mitosis [117-119]. As an example, mitosis events have been detected simultaneously in tracked stem cells by sharing the same features from convolutional layers through applying deep learning with convolutional structure and multi-output layers.[120]. Also, the object-tracking package DeepLabCut, which uses deep learning approaches for object tracking, has been employed for tracking flies and mice in behavioural studies in the field of neuroscience [109]. Recently, deep learning has been implemented in tracking software packages such as LEAP[111] and idtracker.ai[110]. However, these methods have been trained and tested on *in vitro* samples, not highly mobile cells in *in vivo* datasets.

Although methods of tracking by object detection using deep learning have demonstrated a promising performance in resolving false negative, false positive, and under-segmentation errors, the application of deep learning-based methods for analysis of intravital microscopy images containing complex motion patterns has been limited so far. Since cell motion profile is as important as spatial features for tracking, and by considering the potential and significant progress that deep structures have offered over previous methods, developing a deep learning-based framework to learn spatial and temporal features to improve the tracking accuracy of various cell types in *in vivo* imaging datasets can be the most appropriate framework. In this thesis, such a framework has been implemented for tracking migratory immune cells in sequence of 2D imaging dataset, which has been described in Chapter 4.

2.7.2.2 Cell Tracking by Object Linkage Using Deep Learning

Conventional tracking by object linkage approaches usually comprise two steps: (1) cell segmentation or detection and (2) cell linkage.

Image Segmentation could be defined as splitting an image file into multiple partition to identify meaningful objects or characteristics. This meaningful object in biological context is usually a cell or subcellular element such as nucleus. Deep learning has a great potential to be applied in segmentation task to differentiate the cell boundaries, background, and intracellular spaces by classifying the pixel level [84, 121-123]. U-Net[101, 124] and DeepCell [125] are two examples of early application of deep-learning for instance segmentation to perform single-cell analysis.

Moreover, deep learning-based object detection have also been adapted for instance segmentation. For example, Faster R-CNN [126] and Retinanet [127] work by predicting bounding boxes for all objects within an image. In this approach, redundant bounding-box predictions are excluded through non-maximum suppression. For general purpose datasets, Mask R-CNN[128], is among the most precise methods of instance segmentation, which predicts an object mask for each identified bounding box in image file.

The second step in object tracking, following object segmentation or detection, is called object linkage [56], which is basically identifying and linking the detected cell along image sequences to form a track. So far, most deep learning approaches on tracking by linkage have focused on object segmentation or detection, and classical object linkage approaches including nearest neighbour and linear programming are still adopted for the linking task, primarily due to a limited available training datasets. In other words, the 3D nature of biological object (cell or particle), make producing training dataset more difficult. For example, deep cell segmentation has been combined with a Viterbi algorithm tracker [129, 130] to enhance the cell detection and tracking accuracy in microscopy videos. In this algorithm, fully convolutional networks (FCN) [131], including convolutional and deconvolutional layers, have been utilised to execute pixel-wise segmentation. In this approach, cell tracks are represented as a Hidden Markov Model [132], within which individual cell locations and binary cell segmentations are represented as the hidden states and the observations, respectively. Then, the most likely state transitions between each frame is determined by the Viterbi algorithm. In other study [133], cell centroid detection is used for multi-cell tracking through applying U-Net [101] to calculate the spatiotemporal information of cells within and between frames and also for cell segmentation by dividing image pixels into cell boundaries, cell interiors and background. Simultaneous detection

and segmentation significantly improve the performance of this method in multi-cell tracking. ConvGRU, which extracts local features and inter-frame information [134], was incorporated into a stacked hourglass network by Payer *et al.* [135] for instance cell segmentation and tracking, which demonstrated acceptable performance even in segmentation of cells with very close membrane. To perform inter-frame object associations, these methods utilise overlap intersection-over-union (IOU).

More recently, the image processing tool DeLTA (Deep Learning for Time-lapse Analysis) [136] has been introduced, which uses two U-Net [101] deep learning models consecutively for cell segmentation and tracking, although employing two U-Net in single cell tracking is not computationally the most efficient configuration. To sum up, a review of the existing literature in the field demonstrates superior performance of deep network approaches over other methods in generating highly accurate tracking results. This encourages us to exploit the potential of deep network approaches in performing the cell association task, to better deal with the known problem of tracking consistency in multi cell association which has been described in Chapter 5.

TABLE 2.4: ADVANTAGES AND LIMITATIONS OF DEEP LEARNING-BASED CELL TRACKING

Deep Learning-Based Cell Tracking Methods	Advantages	Limitation
Deep learning on Tracking by Detection	• Highly accurate detection. • Able to resolve false positives, false negatives, and under-segmentation errors. • Low computational complexity.	• Hard to deal with tracking highly mobile cell. • Hard to deal with tracking multiple cells. • Mostly need first frame target location. • Lack of learning cell motion behaviour.
Deep learning on Tracking by linkage	• Accurate segmentation. • Deals with multi-call tracking. • Separate segmentation (detection) and tracking task. • Simple association models.	• High computational complexity. • Accuracy of tracking very dependent on accuracy of segmentation task. • Lack of learning cell motion behaviour.

2.8 Tracking Evaluation Metrics

For systematic assessment and comparing the performance of different algorithms in multiple and single object tracking, a range of metrics (Figure 2.4) have been introduced [137-139].

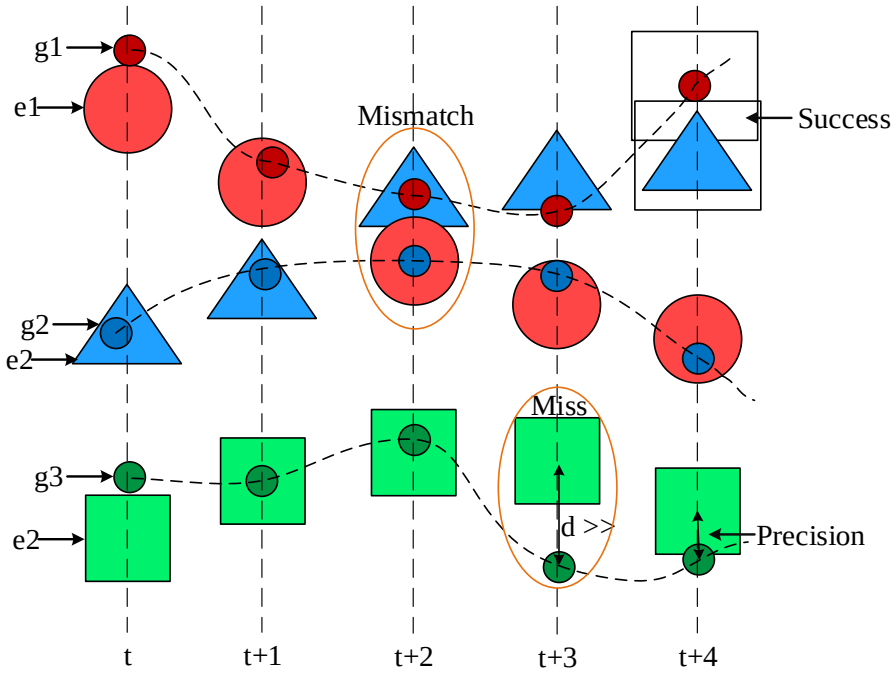


Figure 2.4: Tracking evaluation is shown for three estimated objects (e1-3) and their ground truths (g1-3) based on four parameters including success, precision, mismatch and miss. Success is based on the bounding boxes overlaps. The distance between centre locations is defined as precision. Mismatch is computed when a track selects the wrong object in its neighbourhood. An object is missed during tracking procedure when the distance between estimated and ground truth is higher than the threshold.

2.8.1 Evaluation Metrices for Single Object Tracking

Precision and success plots are the two evaluating factors proposed by [28] for performance analysis [28] of single object tracking methods. The precision plot is computed according to the centre location error, in such a way that, for any object, it is defined as the average Euclidean distance between the centre locations of the tracked objects and the manually annotated ground truths. Precision plot demonstrates the percentage of frames in which estimated location in tracking is within the given threshold distance of the ground truth. The second evaluation metric, success plot, is based on the bounding box overlap, defined as "overlap score", and computed as intersection of the estimated and ground truth regions over their union defining as $S = \frac{|r_g \cap r_e|}{|r_g \cup r_e|}$ where r_e and r_g are estimated and ground truth regions, respectively. The number of successful frames whose overlap S is larger than the given threshold is used to present the success plot for a sequence of image frames.

Conventionally, to calculate the average precision or success rate for tracking algorithms, they are run over a sequence of test images, beginning with the ground truth position in the first frame. One drawback of this evaluation method, which is called one-pass evaluation (OPE), is the varied sensitivity of different trackers to initialization at a different start frame. This issue is dealt with by perturbing the initialization either "temporally" (i.e., start at different frames) or "spatially" (i.e., start by different bounding boxes), which are called temporal robustness evaluation (TRE) and spatial robustness evaluation (SRE) [28], respectively.

2.8.2 Evaluation Metrices for Multiple Object Tracking

To precisely estimate object positions and perform the consistency of multiple object tracking, each object is assigned a unique track ID, which is preserved for the entire length of the image sequence, even on the occasions of temporary occlusion. Several evaluation metrices have been proposed for evaluation of multiple object tracking methods [137, 138, 140, 141], mainly focusing on precision and accuracy, called multiple objects tracking precision (MOTP) and multiple objects tracking accuracy (MOTA), respectively.

MOTP represents the power of the tracker in estimating object positions precisely and is computed by the total error in the estimated position for the matched object-hypothesis pairs throughout sequential frames, averaged by the total number of generated matches. Factors such as recognizing object configurations and consistency of tracking are not considered in the MOTP calculation. The MOTA, on the other hand, represents all object configuration errors made by the tracker over all frames including mismatches and misses [137]. Two evaluation metrices of identification precision (*IDP*) and identification recall (*IDR*) have been proposed for evaluating tracking consistency [138], in which the consistency error is calculated relative to the longest ground truth-hypothesis match tracks.

2.9 Conclusion

In this chapter, the features of different tracking methods in the existing literature have been reviewed. Improving cell tracking has a significant impact on refining computer-aided

diagnosis systems that assist biologists in investigating inflammatory responses against invading microorganisms through automated analysis of large-scale imaging datasets.

The common classical cell tracking techniques are categorized as “tracking by model evaluation” and “tracking by detection methods”. The advantages and limitations of common model evaluation and detection-based tracking algorithms are summarized in Tables 2.2 and 2.3, respectively. Most of the studies in model evaluation-based tracking focus on the accuracy of cell boundaries, which is very beneficial when analysing morphological changes in the cells, while most of the cell tracking by detection models concentrate on tracking accuracy.

Considering the importance of deep learning as a topical research area in the field of computer vision, the recent deep learning-based cell tracking models, cell tracking by-detection and by-linkage, have also been reviewed in this chapter. The advantages and limitations of each category are summarized is Table 2.4. The convolutional neural network (CNN) has become one of the most popular deep learning techniques in object detection and tracking, which is employed to learn the robust features of the objects. Most cell tracking by detection methods have adopted deep learning, aiming for more accurate detection and also to resolve false positives, false negatives and under-segmentation errors in the tracking stage. However, for live-cell microscopy image analysis, the impact of deep learning has been limited so far. Considering the complex motion profile of biologic objects, including cells and particularly migratory immune cells, and also the proven tracking potential of deep learning models, developing a deep learning-based framework to learn both spatial and temporal features of immune cells to improve the live cell tracking

accuracy seems the most appropriate framework. In this thesis, such a framework has been described in Chapter 4.

The majority of the studies in tracking by cell linkage focus on adopting deep learning in cell segmentation to differentiate cell boundaries and intracellular structures from the background. Despite the improved performance of these methods in accurate segmentation, their capacity in cell association is still dependent on simple classical approaches such as intersection over union or nearest neighbour, which often result in multiple broken trajectories. Hence, to improve the accuracy of cell tracking, the potential of deep network approaches in performing the cell association task is exploited to better deal with the known problem of tracking consistency in multi cell association which has been described in Chapter 5.

Finally, the most appropriate evaluation metrics that systematically assess and compare the performance of different algorithms for multiple and single object tracking have been reviewed.

Chapter 3: Data Acquisition

Tracking cells in multidimensional time-lapse fluorescence microscopy is a crucial task in the fields of computer vision and biomedical science. Studying leukocyte (a type of immune cell) and their migratory profile is one of the highly topical fields in inflammation research due to determining role of these cells during immune responses. Recent advances in microscopy techniques and in vivo animal models for studying inflammatory diseases have resulted in an obvious demand for developing better tracking algorithms enable of analysing complex migratory profile of these cells. Despite huge attempts in recent years to develop automated cell tracking systems based on advances in computer vision technology, the field still lacks a state-of-the-art software and method to overcome challenges arising from the intrinsic complexity of these cells in both their migration pattern and amorphous morphology. However, developing novel automated cell tracking methods capable of handling these challenges is highly restricted by shortage of comprehensive ground truth datasets for training and validation. Thus, in response to this demand, we created a data repository comprising a significant number of manually detected and tracked leukocytes and particles imaged in a zebrafish inflammation model. A variety of biomechanical or technical challenges in this dataset would provide a platform which could be used to develop tracking algorithms capable of handling these real-world challenges.

NOTE: The content of this chapter has been submitted to *the IEEE Transactions on Biomedical Engineering*.

Moghadam, R. M., Pazhakh, V., Lieschke, G. J., and Chen, Y. P. P. "Cell Tracking Repository for Biomedical Engineering Applications." *Bioinformatics*. (Submitted).

3.1. Introduction

Automated cell tracking is an important fields in biomedical engineering due to its various applications in the cell morphology assessments, cancer metastasis, wound healing, and immune response[142-147]. Migratory profile of leukocytes (white blood cells) as one of the highly mobile cell types is an essential element of understanding the immune response in inflammatory diseases. Optical transparency and feasibility of genetic manipulation along with a highly conserved immune system similar to higher vertebrates [148, 149], have made zebrafish embryos an excellent model organism to study immune responses and inflammation in vivo [7]. A common genetic manipulation of the zebrafish is labelling cells or biologic components with fluorescent reporter proteins. Fluorescent proteins are a class of proteins which are excited by absorbing high-energy photons in a specific range and then releasing that energy by emitting a lower-energy visible photon in response [5]. The fluorescent-labelled structures can then be imaged in multiple dimensions using different imaging techniques, including wide-field, confocal, multiphoton or light-sheet fluorescence microscopy [148, 150].

Intravital time lapse imaging of biologic processes provides a platform for visualising cellular morphology and behaviour over time in physiologic and pathologic condition within live organisms. Developing model organisms and technological advances in imaging platforms provide the chance for collecting multi-dimensional large-scale datasets containing huge amount of interpretable biologic information. Despite remarkable advances in biologic models and microscopy techniques, the field still lacks equally developed automated tracking system for efficient analysis of these datasets to avoid the extremely time-consuming process of manual analysis, which is often inefficient for quantitative analyses such as calculating cell volumes, boundaries, and velocities. Manual

analysis is also prone to being subjective or biased by human operator performing the analysis.

Along with the advances in digital computing technology and, several automated approaches on cell detection and tracking[151] and software packages such as Imaris (Bitplane), Volocity and FIJI [152] have been introduced in the field of computer vision and biomedical engineering aiming to solve challenges in automated analysis of biologic imaging dataset, including dynamic quantification of signalling pathways [153], cell motility studies [154], and measuring bacterial growth [155], and genetic[156-158]. However, detecting more complicated properties of biologic objects such as highly amorphous and elastic morphology, variable velocity and motion profiles, complex cell-cell interactions and natural phenomenon like cell death (disappearance), cell division (split) or merge in case of mitochondrial fusion are still challenging tasks for the current analytical capacity of available detecting and tracking systems.

The emergence of “deep learning” techniques, has significantly extended the range of problems that could be solved by computer vision through learning of sample characteristics by several levels of abstraction [159-162]. To be able to apply deep learning to cell tracking generally three major components are required, including: (1) constructing a labelled training dataset; (2) effective training of deep learning models on that dataset; and (3) testing trained models on new data. Thus, preparing high-quality training datasets is as important as developing the deep learning algorithm. The more diverse training datasets are, the higher chance those algorithms have to identify and track objects in actual analysis. This is the actual gap in the field of biomedical imaging that the proposed repository in this work is trying to fill.

We proposed an extended dataset of microscopy time laps images capturing leukocytes in the context of inflammation, with significant number of cells (~20,000 biologic objects including neutrophils, macrophages, and particles) being manually labelled and tracked by experts in the related field. Our aim is to provide a high-quality dataset to support development of deep learning algorithms capable of tracking leukocytes and possibly other highly amorphous and fast migrating cells in biomedical imaging fields.

Publicly available datasets for addressing cell and particle tracking challenges, such as [151, 163, 164], are useful tools for highlighting the biological objects properties to test different algorithms, however none of the already existing datasets includes imaging files of amorphous and highly mobile leukocytes taken through intravital 5D (x, y, z, f, t) confocal microscopy and at the size of our repository. The other unique feature of the presented dataset in this work is comprehensive labelling of cells in different imaging files with various characteristics, which is ideal for training deep learning-based tracking models.

3.2. Data Generation

Imaging data for this repository were adopted from a larger dataset originally created to study host-pathogen interaction in the zebrafish model [165] [166]. Zebrafish husbandry was undertaken according to standard protocols [167]. Three days post-fertilisation Tg(mpeg1:Gal4FF)×(UAS-E1b:Eco.NfsB-mCherry)×(mpx:EGFP) embryos, expressing Enhanced Green Fluorescent Protein (EGFP) in neutrophils [8] and mCherry in macrophages [168] were anesthetized by 0.168 mg/ml tricaine followed by microinjection of fungal spores or plastic beads (blue particles) into the tail muscle at the level of the dorsal

tip of the yolk extension to mimic local infection. Infected embryos then were mounted in low melting point agarose and site of infection was imaged by a Zeiss LSM 710 laser scanning microscope with a W Plan-Apochromat 20x, 1.0 NA objective at 28°C for minimum of three hours. ZEN software (2012, black edition 64 bit) was used for the acquisition. The size of each image is 512×512 pixels (2D images) and z-stacks were composed of 20–40 slices (31 ± 4) covering 35–130 μm of z-depth (a). The interval time between z-stacks was set to zero and imaging each z-stack took 33.24 ± 9.50 s (approximately 2 z-stacks per minute) to generate a 4D dataset (x, y, z, t), where x, y are length and width of the field of microscopy (2D images), z is the depth of imaging, and t dimension could be expressed as number of 3D frames. For this repository, the first 100 frames of selected imaging datasets were cropped and neutrophils (green cells), macrophages (red cells) and fungal spores (blue particles) were manually tracked to generate ground truth.

Confocal and multi-photon microscopes can also capture the intensity value i at a specific three-dimensional location (x, y, z). This intensity is related to the photons collected at the detector, which are in turn emitted by fluorescent substances in the sample, in response to excitation at specific frequencies f. As the observation is repeated in time t , the data become a 5-dimensional matrix $i(x, y, z, f, t)$. The 5D dataset was then converted to multiple 3D TIFF files through Imaris 7.4 (Bitplane). The fluorescence channels (red, green and blue) and time were separated by name in each 3D images. For some files, bright field channel is also provided. 2D dataset was generated by converting 3D stacks (slices) to maximum intensity projection (MIP). MIP is a method to represent highest intensity values along the Z-axis in a 3D image. Figure 3.1 shows schematic of the 3D surface

rendered, Z stack, and 2D image of leukocyte populations and microbial particles labelled with fluorescent tags.

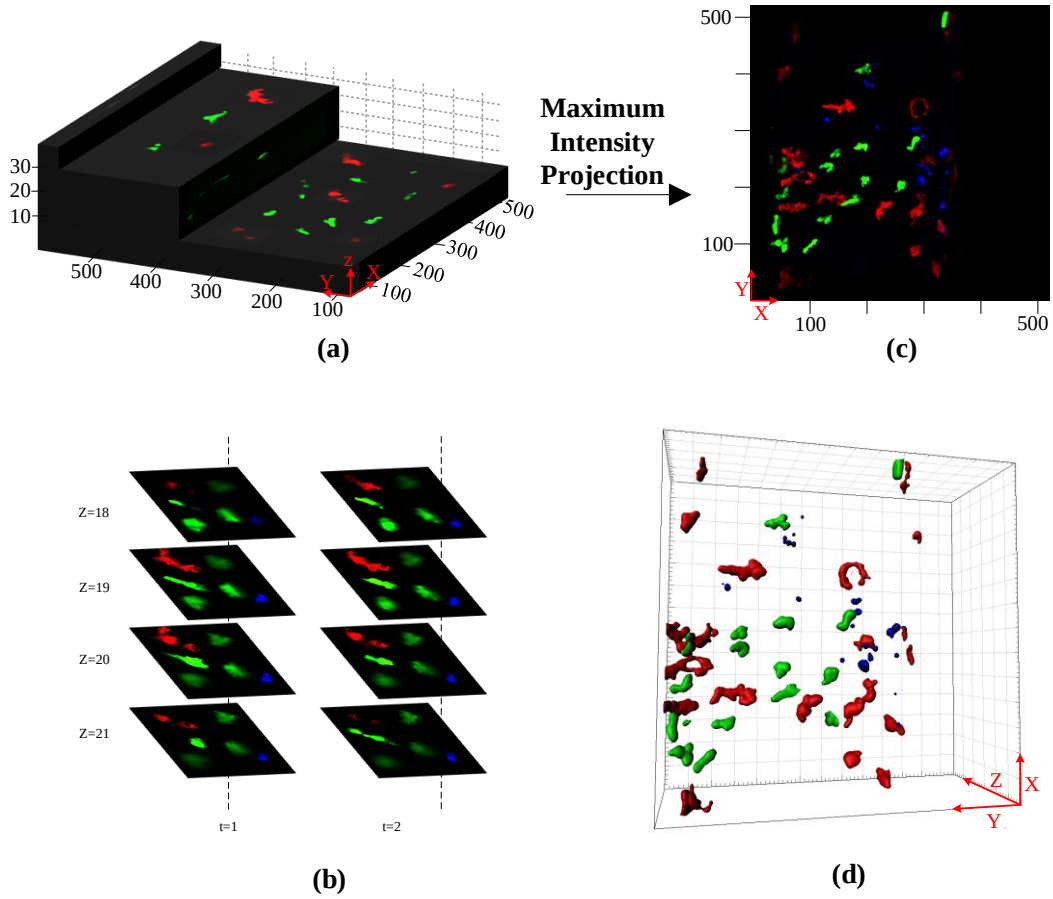


Figure 3.1: a) Zstack. b) Individual z slices from two time points. c)Maximum intensity projection (MIP) of a 3D (z-stacks) image of fluorescent leukocytes in zebrafish at one-time point. d)A schematic of the 3D surface rendered.

3.3. Data Annotation

We have generated a leukocyte tracking ground truth dataset, in which cells have been manually labelled and tracked in sequential microscopic images. The image acquisition pipeline of confocal microscopy data is depicted in (Figure 3.2). After acquisition, image sequences were extracted in 2D (x, y) and 3D (x, y, z) formats comprising different fluorescent channels. Finally, ground truth and training datasets were generated by labelling individual cell positions (centre location and bounding box) and tracking their dislocation through sequential images over time. The generated ground truths are mainly for neutrophils, but also some macrophages and microbial particles have been tracked according to the relevant challenges.

For labelling purpose, a MATLAB function was defined to annotate individual cell positions, selected and followed by an operator during time frames. Along with cell position, accurate bounding boxes were also assigned by masking the selected cell and applying thresholding technique [169] to specify cell boundary from background. The selected cells then were cropped and saved as images of individual cells, to generate a training dataset. Examples of labelled leukocytes for detection and tracking are shown in Figures 3.3 (a) and (b). In this approach, cells are detected by annotating the centroid position and bounding boxes while their tracks are produced by associating their centre position over time subsequently and labelled by track ID. A ground truth file was then generated through repeating the labelling procedure in an image sequence.

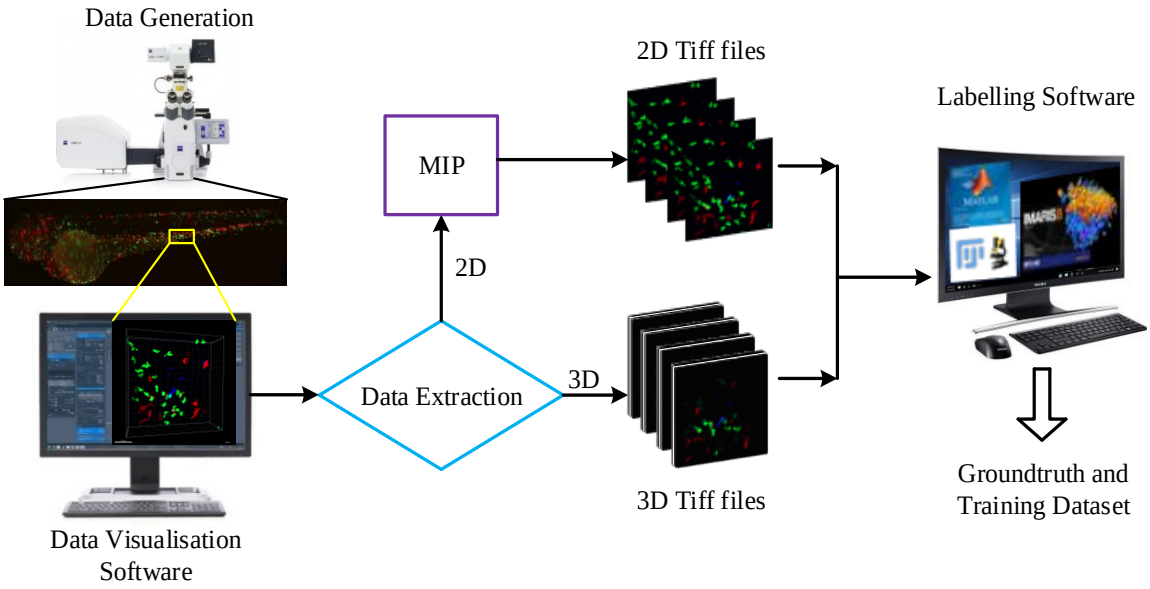
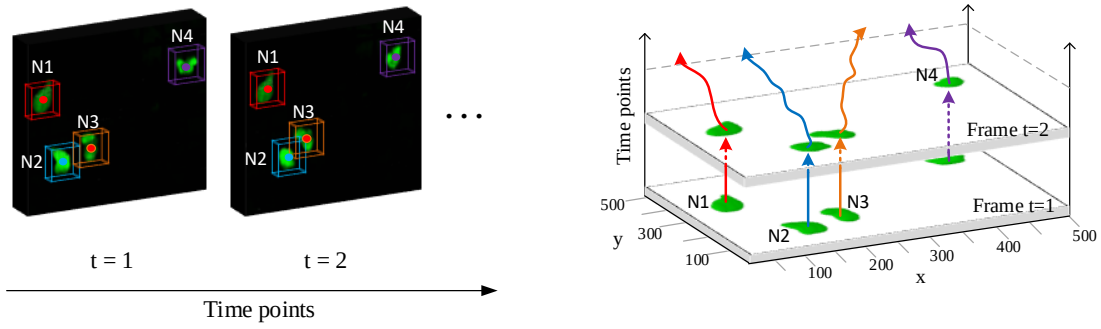


Figure 3.2: Fluorescent leukocytes and particles in the tailfin of a zebrafish larva are shown during microscopic scanning (three colours including green (neutrophils), red (microphages) and blue (particles)). The raw microscopic data can be displayed by a visualisation software (Zeiss). The 4D data is extracted to 2D and 3D images at different time points which are then separated to 3 florescent channels (green, red, blue). The cells are then labelled using some software on each individual fluorescent channel. Finally, the ground truth dataset is generated.



(a) (b)

Data ID	Track ID	Frame	Channel	Centre Location	Bounding Box
1	1	1	G	[253.32, 310.07, 2.42]	[238, 298, 0.5, 33, 23, 4]
1	1	2	G	[251.11, 313.36, 2.42]	[238, 299, 0.5, 26, 27, 4]
				• • •	
1	2	1	G	[125.06,227.53,3.56]	[114,210,1.5,22,37,5]
1	2	2	G	[120.55,229.51,3.99]	[105,215,0.5,29,30,6]
				• • •	
1	3	1	G	[92.29,208.49,7.03]	[79,194,3.5,24,28,6]
1	3	2	G	[98.14,205.65,5.97]	[88,192,2.5,22,25,7]
				• • •	
1	4	1	G	[65.38,275.13,11.47]	[54,259,8.5,22,39,6]
1	4	2	G	[65.72,276.52,11.45]	[54,261,8.5,22,39,6]
				• • •	

(c)

Figure 3.3: The labelling results on detection and tracking are shown in 3D images over time. (a) The labelled centres and boundaries are shown by dots and bounding boxes for four cells in two time points. (b) The labelled centres are then formed tracks by connecting them over time. (c) The ground truth is displayed for the example dataset (Data ID is 1). Track ID is 1 to 4 as this data has four cells (tracks). Number of frames are shown only for two time points while they have been continued to end of the tracks (depends on the track duration). Channels are represented by G (green), R (red), and B (blue). Centre location and bounding boxes are represented by three and six numbers which are the coordinates of the cell centre and boundary respectively. Finally, computed volume is shown in the last column for each individual cell.

Algorithm 1 represents the labelling procedure in both detection and tracking.

Algorithm 1: Data Labelling Procedure

Export microscopy data to image sequences on each fluorescent channel (G, R, B)

Data ID: $d_{id} = 1, \dots, 51$

Track ID: $t\_id = 1, .., n$

Number of frames: f $= 1, \dots, T$

Image sequence $I = I_1, \dots I_T$

for every d_id **do**

for every t_id **do**

for every f **do**

Display Image I_f

Select individual cell

Label the target cell as N_{t_id}

Compute centre and bounding box position

Save the d_id, t_id, f, centre, and bounding box in a table

end for

end for

end for

The tracking and detection are defined to serve as a ground truth for training and evaluating cell tracking algorithms. Measuring deviations from ground truth could be expressed as a metric that reflects the accuracy of different tracking methods [137]. A dataset composed of labelled individual cells could serve as a training dataset for deep

learning methods. Indeed, in light of recent applications of deep learning for object detection and tracking in highly variable scenarios, our dataset can provide the large number of labelled images required for the training of learning-based cell tracking models

Cell tracking challenges arise from many factors including complexity of biomechanical properties of leukocytes, such as being amorphous, having non-linear motion pattern and making indefinite contacts with their surrounding cells and tissues to fulfil their patrolling job in the organism. Apart from biologic complexity including high cell density, high velocity, high elasticity, cell aggregation, and highly cell-cell interactive there are also technical challenges related to in vivo microscopy artefacts such as overlap between emission spectra coming from different cell types labelled with various fluorescent proteins, photo toxicity of excitation lights, sample movement, and noise (Table 3.1).

To classify our dataset in terms of different detection and tracking challenges, we have categorised our dataset into 10 attributes shown in Table 3.1. For any given image sequence, the most outstanding attributes for automatic analysis have been indicated in Table 3.2. Green, blue, and red fluorescent channels are summarized to G, B, R.

TABLE 3.1: CELL TRACKING CHALLENGES IN TERMS OF BIOMECHANICAL AND TECHNICAL PROBLEMS

Attributes	Description
High cell density (HCD)	Higher number of cells present in the field of imaging passing by in close proximity of each other
High velocity (HV)	Cells migrating at velocity higher than average for zebrafish neutrophils (0.020 μm/s) [1], leading to non-overlapping dislocation between consecutive frames
High elasticity (HE)	High variability in cell shape, such as elongation and formation of protrusions
Cell aggregation (CA)	Accumulation of cells at infection site, causing cell overlapping each other in x,y or z dimension
Blurred signal (BS)	Out of focus sample, due to wide pine hole or thick sample
Moving sample (MS)	Unwanted physical movement of sample on microscope stage in x, y, or z directions
Channel mixing (CHM)	When excitation or emission wavelengths from different fluorescent sources in the same sample overlap and are not well separated
Highly cell-cell interactive (HCI)	When cells are in close vicinity and contact each other frequently
Noise (N)	Unwanted captured fluorescent signal usually caused by autofluorescence, and unoptimized instrument set up (imaging parameters)
Bright field (BF)	Bright field channel included in acquisition channels
Special features (SF)	Special features on imaging quality e.g. high magnification, poor red signal, etc.

TABLE 3.2: DISTRIBUTION OF CHALLENGES IN EACH IMAGE SEQUENCE FILE

Data ID	HCD	HV	HE	CA	BS	MS	CHM	HCI	N	BF	SF
01		•	•								High magnification
02	•	•	•								
03		•			•						
04		•		•							
05	•	•			•						
06	•		•					•			High magnification
07						•					
08			•	•							
09			•		•	•					
10				•		•					
11		•			•						
12		•				•		•			
13		•	•					•			
14		•		•		•					
15	•	•				•					
16	•	•	•	•		•					High magnification
17					•						
18					•	•					Poor red signal
19					•	•					
20				•		•		•			
21					•	•					
22			•			•	•GB				
23		•		•							
24			•	•		•					
25			•		•			•			High magnification
26					•	•					
27	•	•		•							
28	•				•						
29	•	•									
30	•	•				•					
31				•	•						
32				•							
33				•							
34	•										
35					•						
36					•	•					
37			•		•						
38					•						
39					•						
40					•						High magnification
41									•		
42						•					
43						•				•	
44					•	•					
45					•		•GR				
46			•			•	•GR				Saturated green signal
47									•		
48										•	
49							•GR/GB				Saturated all signals
50					•		•GR				
51	•	•	•			•					

Complexity of each image sequence has been indicated based on cell tracking challenges. A list of 51 image sequences has been identified in data ID. The outstanding challenges (attributes) of each data have been represented by dot in its specified column. Green, blue, and red fluorescent channels are summarized to G, B, and R in channel mixing attribute (CHM).

The attribute distribution in our dataset is shown in Figure 3.4 Some attributes occur more frequently than others, such as blurred signal (BS) and moving sample (MS).

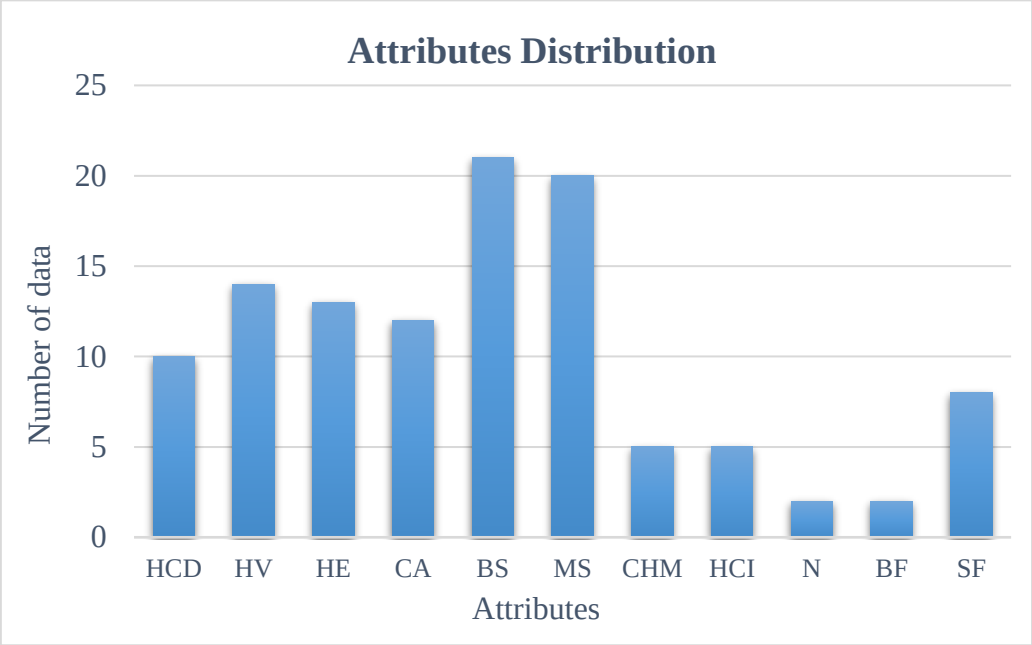


Figure 3.4: The distribution of biomechanical and technical challenges is represented on image sequences.

Figure 3.5 visually represents some of the challenges presented in our dataset. In order to maximize track duration, partially visible cells (weak emission signal) were also included in tracking, and only completely invisible cells were excluded.

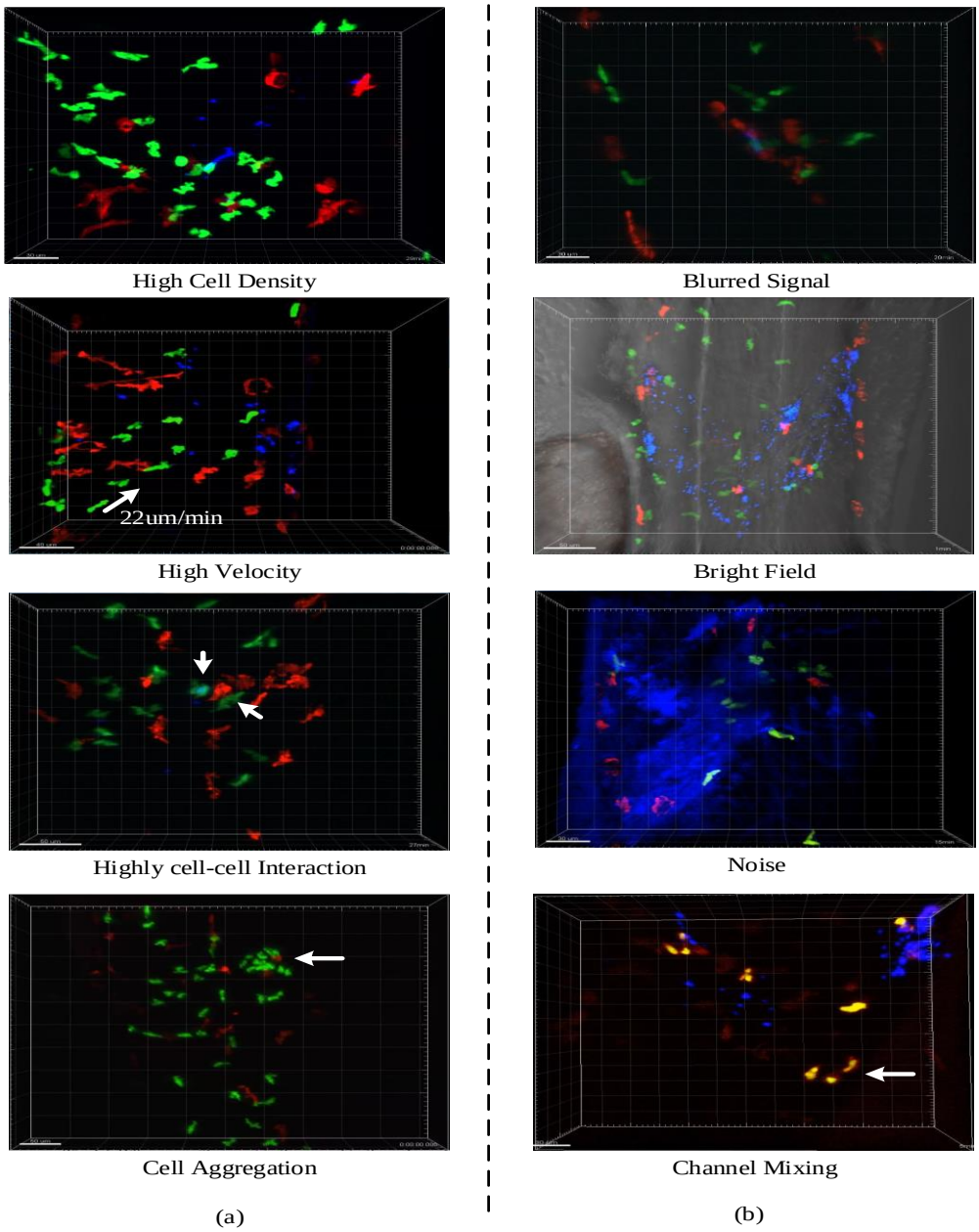


Figure 3.5: Visual demonstration of selected biomechanical and technical challenges on example datasets. a) The complexity of biomechanical properties included high cell density, high velocity highly cell-cell interaction and cell aggregation. b) Blurred signal, bright field, noise, and channel mixing are some example of technical challenges.

The dataset repository consists of 51 labelled videos occupying 94 GB of raw time laps images versions and 25,000 cropped and labelled images of individual biological objects, including neutrophils, macrophages and microbial particles, in both 2D and 3D. Table 3.3 provides the characterization of each 3D dataset including, data size and voxel size, image format, number of channels, and number of ground truth tracks for neutrophils, macrophages, and particle. Imaging files have also been categorized into three levels of easy, moderate, and hard based on their complexity. A characterization table is not provided for 2D dataset as only varying factors are the data size and the pixel size which will change to (x, y, t) and (x, y), respectively.

TABLE 3.3: DESCRIPTION OF EACH IMAGE SEQUENCE FILE AND THEIR GENERATED GROUND-TRUTH TRACKS

Data ID	Data Size (x, y, z, t)	Voxel Size (μm)	Format (bits)	No.Channels	No.Neutrophil Tracks	No.Macrophage Tracks	No.Particle Tracks	Complexity
01	512,512,29,100	0.642,0.642,2.94	16	3	5	5	-	hard
02	512,512,40,100	0.536,0.536,2.96	16	3	4	-	-	moderate
03	512,512,27,100	0.664,0.664,1.25	16	4	6	-	-	easy
04	512,512,45,100	0.524,0.524,2.70	16	3	5	-	-	easy
05	512,512,23,100	0.560,0.560,2.83	16	3	7	-	-	hard
06	512,512,32,100	0.507,0.507,2.38	16	3	1	-	-	easy
07	512,512,33,100	0.576,0.576,2.54	16	3	3	3	-	hard
08	512,512,30,100	0.594,0.594,2.45	16	3	3	-	-	hard
09	512,512,34,100	0.559,0.559,2.45	16	3	1	-	-	moderate
10	512,512,33,100	0.560,0.560,2.83	16	3	4	-	-	moderate
11	512,512,30,100	0.560,0.560,1.56	16	3	2	-	-	easy
12	512,512,33,100	0.648,0.648,1.74	16	3	7	4	-	easy
13	512,512,30,100	0.506,0.506,1.75	16	3	8	5	10	hard
14	512,512,33,100	0.573,0.573,1.95	16	3	12	-	-	easy
15	512,512,32,100	0.614,0.614,1.83	16	3	8	-	-	moderate
16	512,512,33,100	0.518,0.518,2.39	16	3	5	-	-	moderate
17	512,512,30,100	0.551,0.551,1.40	16	3	5	-	-	moderate
18	512,512,30,100	0.616,0.616,2.24	16	3	3	-	-	moderate
19	512,512,35,100	0.516,0.516,2.71	16	3	1	-	-	easy
20	512,512,35,100	0.576,0.576,1.91	16	3	5	3	-	moderate
21	512,512,29,100	0.595,0.595,1.35	16	3	4	-	-	easy
22	512,512,35,100	0.615,0.615,2.42	16	3	6	7	7	hard
23	512,512,33,100	0.573,0.573,2.17	16	3	5	-	-	easy
24	512,512,30,100	0.526,0.526,1.72	16	3	5	-	-	hard
25	512,512,30,100	0.471,0.471,1.78	16	3	3	4	5	hard
26	512,512,30,100	0.675,0.675,2.15	16	3	5	-	-	easy
27	512,512,33,100	0.573,0.573,2.17	16	3	5	-	-	moderate
28	512,512,33,100	0.586,0.586,2.12	16	3	3	-	-	easy
29	512,512,40,100	0.573,573,3.21	16	3	9	-	-	hard
30	512,512,33,100	0.764,0.764,2.39	16	3	4	-	-	hard
31	512,512,36,100	0.997,0.997,1.78	8	2	8	-	-	easy
32	512,512,33,100	0.613,0.613,2.30	16	3	4	-	-	moderate
33	512,512,33,100	0.649,0.649,1.87	16	3	5	-	-	moderate
34	512,512,33,100	0.708,0.708,2.10	16	3	5	-	-	moderate
35	512,512,40,100	0.582,0.582,2.42	8	3	3	-	-	easy
36	512,512,30,100	0.511,0.511,2.25	8	3	5	-	-	moderate
37	512,512,30,100	0.488,0.488,1.57	8	3	4	-	-	hard
38	512,512,12,100	0.554,0.554,2.25	8	3	6	-	-	easy
39	512,512,33,100	0.639,0.639,1.38	8	3	4	-	-	moderate
40	512,512,21,100	0.511,0.511,1.58	8	3	7	-	-	easy
41	512,512,21,100	0.511,0.511,1.82	8	3	3	-	-	easy
42	512,512,33,100	0.755,0.755,1.53	8	3	1	-	-	moderate
43	512,512,30,100	0.753,0.753,3.06	8	4	6	5	-	hard
44	512,512,29,100	0.639,0.639,1.77	8	3	2	-	-	hard
45	278,310,25,100	0.874,0.874,3.11	16	3	4	7	-	hard
46	512,512,22,100	0.583,0.583,2.77	16	3	3	4	7	hard
47	512,512,36,100	0.543,0.543,1.53	16	4	7	8	-	easy
48	512,512,40,100	0.680,0.680,1.58	16	4	4	3	-	easy
49	512,512,36,100	0.678,0.678,3.66	16	3	8	3	4	moderate
50	512,512,25,100	0.874,0.874,3.11	16	3	3	6	-	hard
51	512,512,35,100	0.507,0.507,3.08	16	3	23	-	-	moderate

Description of each image sequences including data size (x, y, z, t), voxel size, format, and number of channels (fluorescence and DIC). The number of generated ground-truth tracks for neutrophils, macrophages, and particle has been indicated for each data. The complexity of each image sequence is categorized to hard, moderate, and easy based on their cell tracking challenges.

3.4 Conclusion

In this chapter we present the data repository aimed to be used for improving novel automated cell tracking methods, more specifically deep learning-based tracking, in multi-dimensional microscopy image sequences. The significant feature of our dataset is the compilation of large enough number of labelled cells in different image sequences with a broad range of detection and tracking challenges to be used as training dataset for deep learning algorithms. Although our imaging dataset lacks diversity in term of cells from different tissues and organs, but it includes a comprehensive collection of phagocytic cell types participating in inflammatory response and their migratory tracks toward site of fungal infection. Thus, this data repository is mostly suitable for developing algorithms capable of tracking amorphous and high velocity migratory cells.

Chapter 4: Tracking Neutrophil Migration in Zebrafish Model Using Multi-Channel Feature Learning

Tracking cells over time is a crucial task in the fields of computer vision and biomedical science. Studying neutrophils (a type of immune cell) and their migratory profile is one of the highly topical fields in inflammation research due to determining role of these cells during immune responses. As neutrophils are generally of various shapes and motion, it remains challenging to track and describe their behaviours from multi-dimensional datasets captured by sophisticated confocal microscopes[170]. In this study, we propose a robust novel multi-channel feature learning (MCFL) model inspired by deep learning to extract the complex behaviour of neutrophils motion in time lapse images taken from zebrafish infection model[171]. In this model, the convolutional neural networks along with cell relocation distance and orientation channels learn the robust significant spatial and temporal features of an individual neutrophil taken from two consecutive time frames continuously. Additionally, we also proposed a new cell tracking framework to detect and track neutrophils in the original time-laps microscopy images, including sampling, observation, and visualisation functions. Our proposed cell tracking by multi-channel feature learning method has remarkable performance in rectifying common cell tracking problem compared to other state-of the-art methods.

NOTE: The content of this chapter has been submitted to *the IEEE Journal of Biomedical and Health Informatics*.

Moghadam, R. M. and Chen, Y. P. P. "Tracking Neutrophil Migration in Zebrafish Model Using Multi-Channel Feature Learning." *IEEE Journal of Biomedical and Health Informatics*. (Accepted).

4.1. Introduction

Visual tracking also known as object tracking, has been widely studied in the field of computer vision for a variety of applications including cell tracking, video security surveillance, sports videos, and autonomous driving. Object tracking involves automatic estimation of the object's trajectory when it moves in the video frames. Automated cell tracking as a type of object tracking is one of the fast-developing fields in biomedical imaging due to its various applications in the cell morphology assessments, cancer metastasis, wound healing, and immune response [32].

Leukocytes (white blood cells) are highly mobile cells and analysing their mobility profile by tracking techniques is an essential element of understanding the pathophysiology of inflammatory diseases. Inflammation is a critical process for multicellular animals during wound healing and also to keep their integrity against unicellular pathogenies microorganisms[2]. Optical transparency and a highly developed immune system comparable to higher vertebrates, have made zebrafish embryos an excellent model organism to study immune responses and inflammation *in vivo*[7]. The availability of extensive genetic resources and the feasibility of genetic manipulation have provided the opportunity to genetically label various cell types using different fluorescent tags and visually track them *in vivo* with high-resolution microscopy techniques [8]. Multiphoton and confocal microscopy are among the most common imaging techniques that can visualise biologic processes at high temporal and spatial resolution during inflammation [172].

Despite the advantages of the zebrafish model and availability of high-end microscopic techniques involved in producing highly informative intravital image, the segmentation and

analysis of information contained in large datasets is often quite challenging and makes the post-imaging analysis step a time-consuming process in this type of research. Due to the increasing size and complexity of the data produced by advancing *in vivo* imaging systems in biomedical fields, the need for the automated processing of these data sets seems inevitable.

Neutrophils are one of the major leukocytes which act as a cellular arm of the innate immune system[173]. One of the remarkable characteristics of neutrophils is their high mobility profile and their fast migration toward sites of inflammation throughout organisms[174]. These cells demonstrate a super-elongated morphology when migrating quickly (average speed of zebrafish neutrophil migration is 9.94 +- 0.36 μm/min [7]), while they form a rounded morphology at stationary phase.

The high mobility profile of neutrophils sometimes causes interruptions in cell tracking as the dislocation of cells between two consequential frames is so large that tracking algorithms may lose the track. Due to the patrolling role of these cells, they are constantly sensing their surrounding environment or other cells partially through physical polarized cell-cell contacts. This is why quite often they are too close to each other or form clusters and then separate in such a way that re-identifying cells after these interactions is a complicated task during cell tracking. The highly amoeboid and irregular mobility profile of neutrophils along with their dynamic shape, which is constantly changing due to the scavenger role of these cells, make tracking leukocytes over time a complicated task.

The level of complexity in biologic bio-particle tracking has forced some studies to limit their dataset to the least overlapping objects in sequential frames [175], or in other

case [7] performing a pre-analysis visual evaluation to exclude biologic objects with highly challenging parameters.

Detection-based cell tracking methods have been widely used in recent years [176]. Although most of these methods work through the real-time processing of datasets by exploiting feature extraction and tracking techniques, their potential for learning the characteristic features of shapeless cells is restricted. Feature discovery through training deep networks using large scale datasets followed by a transfer learning method has originated from deep learning[159], which is one of the most successful feature learning techniques[177]. In our proposed cell tracking method, we use convolutional neural networks (CNN) [178-180] to learn the robust features of neutrophils.

Based on CNN and tracking techniques, we propose a novel method to address the challenges (complex variations in shape and motion) of tracking neutrophils in zebrafish model. By combining CNN and two motion learning channels, both spatial and temporal features of neutrophils can be learned during training procedure. The proposed framework comprises of three tasks, including sampling, observation, and visualisation. It starts with a rectangular bounding box in the first frame of the image sequence which provides the target neutrophil position. Then, a set of candidates bounding boxes are produced by the sampling function based on particle filters in the following frames. Eventually, the observation function selects the most probable candidate from all possible cases and defines it as the final position of the cell in following frame. Afterwards, the visualisation task enhances the tracking performance by defining the threshold boundary (background or cell) on selected candidate to measure the cell area.

To best of our knowledge, this is the first work in which deep learning has been applied to learn both spatial and temporal features for cell detection and tracking.

The contributions of this work are summarized as follows:

1) We constructed a cell tracking dataset of neutrophil migration in the zebrafish model. The dataset is labelled with ground truth annotation prepared for both training and tracking aspects.
2) A new feature learning model (MCFL) is proposed to extract the features of shape and motion of associated cells in every pair of adjacent frames.
3) We proposed a novel cell tracking framework to address the cell tracking challenges. The framework entails sampling, observation, and visualization functions. With only one labelled cell in the beginning frame, the framework estimates the cell motions in successive frames.

This chapter is organized as follows. Related cell tracking methods and techniques are discussed in Section 4.2. A summary of data acquisition is given in section 4.3. Section 4.4 introduces the outline of our proposed method and framework. In section 4.5 the performance of our proposed cell tracking framework is evaluated and demonstrated. Finally, conclusions and suggestions for future work are presented in section 4.6.

4.2. Related Work

Different methodologies have been recommended for cell tracking in a way that each method has been optimised according to specific cell types and tracking profiles[33, 34]. Traditional tracking methods have been proposed to handle predictable appearance change with handcrafted features. For example, [181] proposed a tracking model to analyse the quantitative cancer cell cycle through spatial distribution, intensity, nuclei morphological appearance, and migration features, while [182] identified features on associated cells between frames within a contour similarity measurement. [183] acquired lineage trees and cell trajectories upon frame-by-frame association by proposing a global spatial-temporal data association method for the tree structure aiming to solve the maximum-a-posteriori problem. The algorithm proposed in [184] considers neutrophil's motions as tracks, forming adjoining time points by means of a keyhole model. This method links a neutrophil in two successive frames based on segmentation and movement direction to predict the most probable landing position of that target neutrophil in the next frame. [185] proposed the leukocyte's shape and size constrained tracking model based on active contour (snake) techniques and energy function. In this method, gradient vector flow is utilized to track high speed rolling leukocytes. In [40], the level-set based method, as a model-based contour evolution approach, identifies cells by associating cell (object) contours according to the initial contour in previous frame. In this model, cell boundaries are detected by minimising the energy function. In [71] the dynamic behaviour of cells is analysed through ant colony approach inspired multi-Bernoulli filter. The ant colony is used to model cell segmentation and the underlying cell motion in an efficient and adaptive way.

In recent years, several cell tracking methods have been developed through machine learning and deep learning approaches. In [186], a machine learning-based cell tracking

approach was proposed to optimize the tracking parameters of a given dataset with large scale feature learning. A level-set approach has been used in [187] to track cells in time-lapse fluorescence microscopy, in which each level-set function indicates a cell or nucleus as an object, in such a way that the evolution equation of each level-set is calculated by substituting the original weight from the energy function.

A machine learning-based pipeline is proposed in [188] to recognize single mRNA spots by determining local intensity maxima features and classifying them through supervised random forest classifier. To address the intrinsic and extrinsic cell appearance variation problem, a feature learning method is proposed based on deep learning and a classic principal component analysis model in [189]. In another study [119] , a fluorescent cell tracking algorithm is proposed based on hierarchical and CNN methods, in which correlation-filter tracking is performed based on the extracted features. [190] proposed a cell tracking model to improve the tracking generalisation appearance using CNN and multi-task learning techniques. In this method, tracking is performed as a real-time process in which the object is followed frame by frame when its location is assigned manually in the first imaging frame. The tracking model and rules are updated simultaneously as tracking proceeds. [191] predicts individual cell position and motion in successive frames using data-driven approach. In this method, cells between consecutive frames are associated based on association score to generate short trajectories (tracklets). The final trajectories are obtain by associating tracklets through cell motion field. [192] detects mitotic event in time-lapse phase contrast microscopy images by combining CNN and long short-term memory (LSTM). This approach detects mitotic sequences of variable lengths and also makes use of longer context information among frames.

Despite all the developments in cell tracking approaches, the field still lacks accurate automatic cell tracking model in which a shapeless cell with complex motion can be tracked consistently on time-lapse images.

4.3. Data Annotation

We built a training benchmark by manually cropping individual cells of any two consecutive frames which are then concatenated to form image pairs in size $60 \times 60 \times 2$. The pairs of images were then classified as true $(y = 1)$ when a particular cell in frames $t - 1$ and t form a pair of cells, and false $(y = 0)$ when there is a pair with different cells in two adjacent frames (Figure 4.1). The training benchmark consists of 55000 true and false samples (45% true and 55% false to bias the classifier). The source dataset was split into 60% for training, 22% validation, and 18% for testing. To increase the size of the dataset slightly shifting and rotating techniques were utilized.

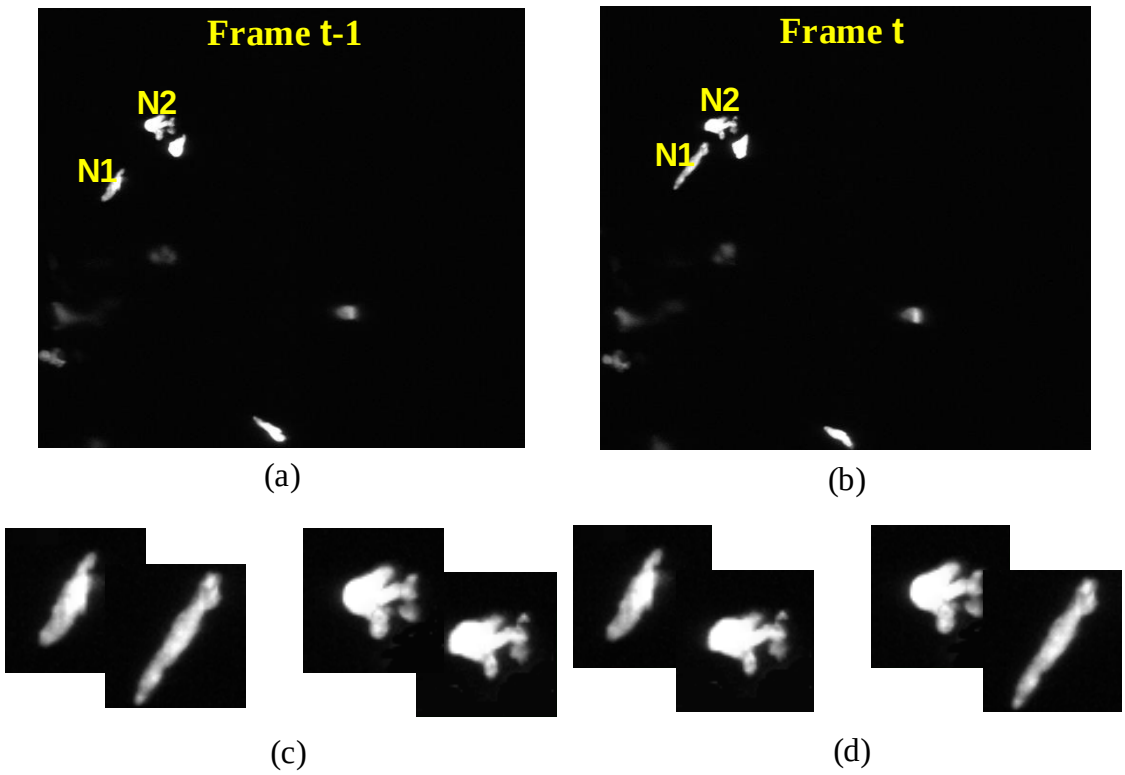


Figure 4.1: (a, b) Frame t-1 and t including marked neutrophils N1 and N2. (c) True samples consist of N1-N1 and N2-N2 pairs in two adjacent frames. (d) False samples consist of N1-N2 and N2-N1 pairs in two successive frames.

4.4. The Proposed Cell Tracking Method

Our proposed cell tracking framework tracks neutrophil's motion automatically in successive frames with only one assigned bounding box in the initial frame. The proposed framework shown in Figure 4.2 comprises three parts: first, sampling function generates a set of candidates bounding boxes in the following frame at the position of selected neutrophil in the first frame, which then will be updated according to predicted neutrophil position in current frame. Second, a new observation function which is the feedforward

network of pre-trained MCFL, determines the most probable candidate. Third, the visualisation function that applies some post-pressing strategies, enhances the accuracy of the tracking model.

Our proposed framework uses two datasets S_{train} and S_{track}. The training dataset is defined as:

$$S_{train} = \left\{ \left[(x_{t-1}, x_t), (d_{t-1,t}), (O_{t-1,t}) \right]_1, \ldots, \left[(x_{t-1}, x_t), (d_{t-1,t}), (O_{t-1,t}) \right]_n \right\}.$$

where (x_{t-1}, x_t), $(d_{t-1,t})$, (O_{t-1}, O_t) are the paired images, centre relocation distance (CRD), and orientation respectively, all of which are described in detail in next sections. The tracking dataset is $S_{track} = \{f_1, \ldots, f_t\}$ where f is an 2D image frame captured by the confocal microscope and t is the frame's number.

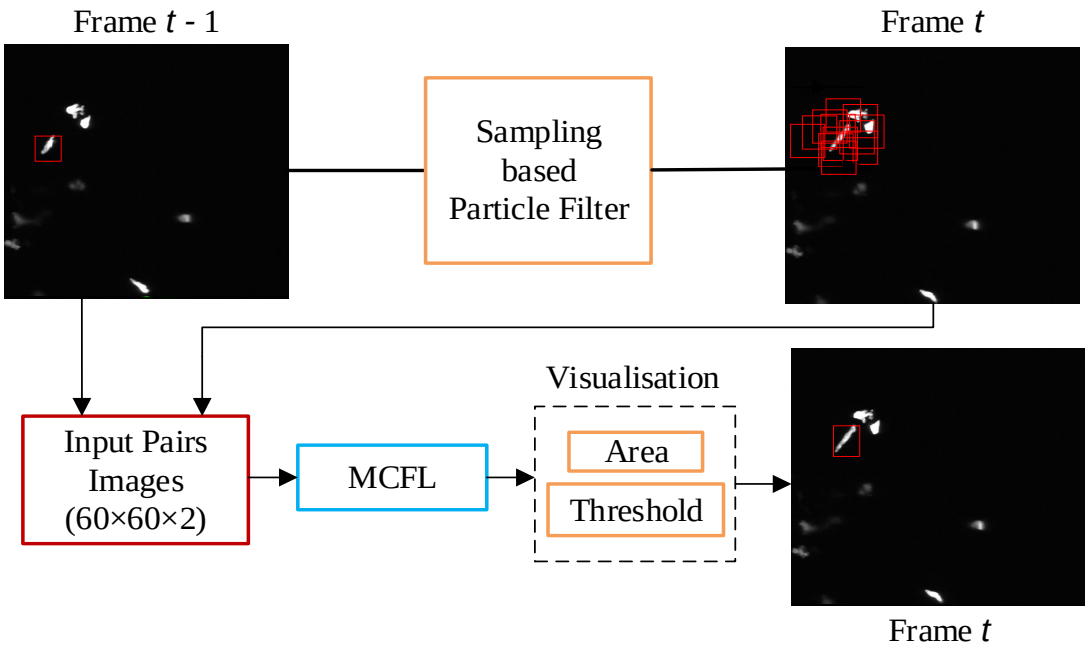


Figure 4.2: The framework of our proposed cell tracking framework representing sampling function, MCFL, and visualization function.

4.4.1. Sampling Function

The proposed cell tracking method exploits the particle filter as sampling function. The particle filter is a sequential Monte Carlo model that recursively estimates the hidden state variable of a dynamic system [193]. When the observation variable is defined as $S_t = [s_1, \dots, s_t]$ up to the tth frame, the model predicts the hidden state variable b_t by the estimation and update steps as follows:

$$v_t = \text{argmax}\, p(v_t \mid s_{t-1}) = \text{argmax} \int p(v_t \mid v_{t-1})\, p(v_{t-1} \mid s_{t-1})\, dv_{t-1} \quad (1)$$

$$p(v_t \mid s_t) = \frac{p(s_t \mid v_t) p(v_t \mid s_{t-1})}{p(s_t \mid s_{t-1})} \quad (2)$$

where $p(v_t \mid v_{t-1})$ and $p(s_t \mid v_t)$ are the dynamic and observation states, respectively.

In our proposed tracking model, the cell state variable is denoted by $V = \{v_1, \dots, v_N\}$ (N is the number of particles) which are representing the probable positions of cell in the next frame. More detail regarding the sampling function can be found in [194].

4.4.2. Observation Function

The novel observation function is proposed to predict the most associated cells in image sequences by learning the spatial and temporal features directly from the images in successive frames. This function consists of two parts: pre-training (MCFL) and tracking (feedforward network).

The proposed MCFL model (Figure 4.3) is composed three learning channels including

appearance embedding used for extracting spatial features and centre relocation distance (CRD) and orientation used for learning temporal features which are described in detail in 4.3.2.1, 4.3.2.2, and 4.3.2.3. The MCFL is trained by updating the kernel which is defined as $W = \{W_x, W_d, W_o\}$ where W_x, W_d and W_o are the parameters of the kernel in three aforementioned channels respectively. The concatenation layer of the network (Figure 4.3) is computed by concatenating the three learning channels which is defined as:

$$z = \sum_{i=1}^{3} (h_{W_i}(s_i)) \tag{3}$$

where $s = \{x, d, o\}$ is the input dataset and it maps to output z through function h. The output feature map is achieved by $z^L = h_{W^L}(z)$ where W^L is the learnable kernel of the last layer. The network is trained by minimizing the following cost function:

$$J = -\sum_{n}^{j=1} \left[y_j \,.\; log\left(F\left(z^L{}_j\right)\right) + \left(1 - y_j\right).\; log\left(1 - F\left(z^L{}_j\right)\right)\right], \qquad F\left(z^L{}_j\right) = \frac{1}{1 + e^{z^L{}_j + b_j}} \tag{4}$$

where y_i and $F(z^L{}_j)$ are the ground truth and predicted output for the jth input respectively, b_j is the bias, and n is the size of the training batch. In this study, we trained our model through 20 epochs on batch size of 500, and a learning rate of 0.01. Once the model is trained with a set of parameters W, it can be used in the tracking part as a feedforward network. In the tracking procedure, the observation function accepts the input of image pairs consisting the n repeated images of the target neutrophil in the first frame f_1 and a set of n candidates bounding boxes (V) generated by the sampling model in the following frame. Then the highest probable match between two adjacent frames is predicted through feedforward network continuously to form a track.

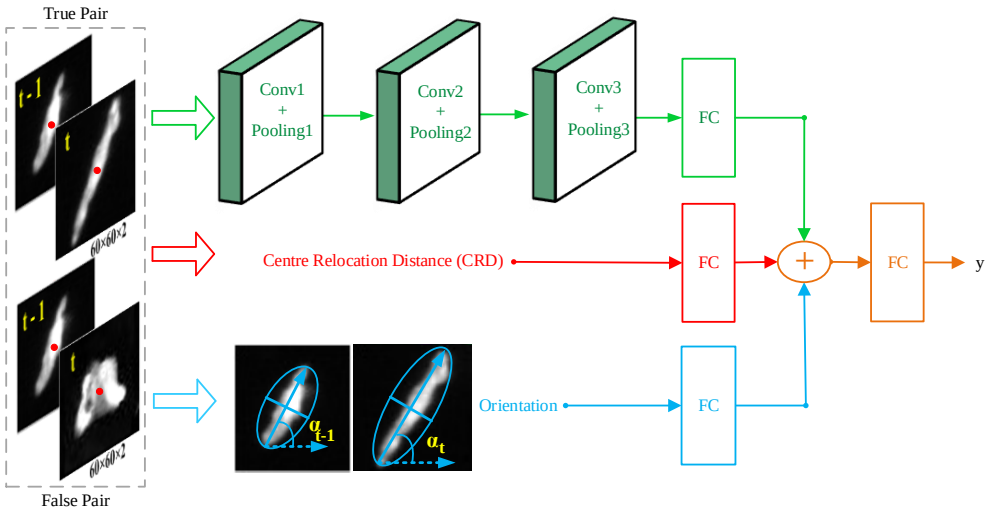


Figure 4.3: The structure of proposed MCFL model

4.4.2.1 Appearance Embedding

In the first learning channel the outstanding spatial features of the cropped paired images are extracted through the CNN architecture. CNN is applied in numerous deep learning networks to represent state-of-the-art performance. As the training sample images have small size, a CNN architecture with more layers is time-consuming with no improvement in the recognition result. To match the recognition performance and processing time, we designed our CNN model in three layers including convolutions which are paired with rectified linear unit (ReLU) and pooling, as illustrated in Figure 4.3.

The function of convolution layers is to apply various convolutional filters to extract the features of the input images. The activation function (ReLU) is exploited to distinct unspecific inhibition and specific excitation in the neural abstraction layers, which could be applied to many computer visions tasks. The pooling layer reduces the size of the input

and the number of parameters through down-sampling function.

Suppose that our CNN architecture has J layers and the weight and bias parameters are denoted by W_x and b_x respectively. The output nodes of each j convolution layer is defined as $a^j = f({W_X}^j a^{j-1} + {b_X}^j)$. The model starts by initializing the activation with $a^0 = X$ where $X = \{x_1, x_2\}$ is the input pair images, then the feature maps of layer $j-1$ are convolved with the learnable weight W_X^j. To train the network based on true and false classes, each single image must convolve with similar kernel and bias in each layer ($W_{x1}^j = W_{x2}^j$, ${b_{x1}}^j = {b_{x2}}^j$). The output feature map is formed through the ReLU activation function $f(\cdot)$. Finally, the size of the feature map is reduced along the pooling layer which is defined as $a^j = down(a^j)$ where $down(\cdot)$ is a sub-sampling function.

The input of the first learning channel is a pair of cropped single neutrophils in any two successive frames which are shown as true and false in Figure 4.3. The dimensions of each single image are 1 × M × N, where M and N are the image width and height. The two individual images are concatenated along the third dimension (M × N × 2) to form the input pair.

The CNN structure in our proposed model is: {input: $60 \times 60 \times 2$} - {conv 1: 30 @ 5×5 } - ReLU - {max-pooling 1:2×2} - {conv 2: 30 @ 5×5} - ReLU - {max-pooling 2: 2×2} - {conv 3: 50 @ 5×5} – ReLU- {max-pooling 3:2×2} - $\{FC_{image}\}$. Here, the size of the max-pooling layers is 2×2. Also, 30@ 5×5 represents the number of convolutional kernels (30) and their size (5×5) in each convolutional layer.

4.4.2.2 Centre Relocation Distance

Each neutrophil has a specific centre coordinate at (x_c, y_c). The probable movement of the neutrophil varies as its shape and motion changes, but it can be estimated by measuring the neutrophil's centre relocation when it moves from $t-1$ to t. Centres are denoted by a vector $C = [x_c(t-1), y_c(t-1), x_c(t), y_c(t)]$ where $\{x_c(t-1), y_c(t-1)\}$ and $\{x_c(t), y_c(t)\}$ represent the centre position of neutrophils at time $t-1$ and t respectively. Then the CRD of a specific neutrophil between two neighbouring frames is calculated as $d = \|C_t - C_{t-1}\|$. This feature is added as a second learning channel in MCFL model.

4.4.2.3 Orientation

By observing the movement of 10000 contiguous neutrophils in the neutrophil dataset, we assumed that the best way to predict the position of a neutrophil when it moves from time $t-1$ to time t is to follow its orientation in the previous frame. For this reason, in the third learning channel, the neutrophil's orientations in two adjacent frames is introduced to improve the training structure. Figure 4.4 illustrates the landing position of the neutrophil based on its orientation. Mathematically, the orientation is defined as:

$$Orientation = (\alpha_t - \alpha_{t-1}) \quad (5)$$

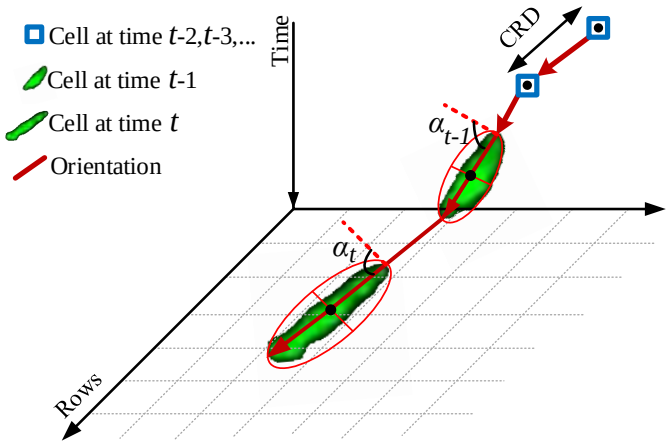
where α is the angle between the major axis of the specified ellipse and horizontal dotted line (x axis) for each single neutrophil (Figure 4.4). The orientation angle is in degree, ranging from -90° to 90°.The neutrophils in two neighboring frames are recognized as one when the orientation degree is close to zero. Cells are mostly circular when they do not have significant movement. In this case, the parent–child relationships are restricted to objects that are relatively close to their parents and follow the previous movements. Then

the orientation angle is defined between the neutrophil radius (in the movement direction) and x axis. To generate orientation ground truth for cropped single cells, we applied some image processing MATLAB functions on image mask.

4.4.3 Visualization Function

The aim of the visualization function is to enhance the performance of the track produced in previous steps through two post-processing steps: "Area" and "Threshold". Firstly, the

Figure 4.4: The position of neutrophil based on its orientation.



area of the predicted neutrophil in current frame is measured and compared to its area in previous frame. Since the area of cell does not change significantly between two sequential time points (two adjacent frames), a predicted neutrophil at time t whose area exceeds $\frac{3}{2}\times$ area in the previous frame is considered as two cells that are too close to distinguish as one cell or noise or outliers when fails $\frac{1}{3}\times$ area in the previous frame. In this situation, (joining two cells in current frame (t)), to predict the correct cell in frame $t+1$, the sampling function constructs a set of pair samples by concatenating the candidate bounding boxes in frame $t+1$ (time after joining) and frame $t-1$ (time before joining). In other

words, the image and position of target cell before joining is memorized and then concatenated to new candidates after joining situation even if two cells are joined for more than one time point. Figure 4.5 represents the accurate tracking result by considering the joining situations through visualization function. The second post-processing step is the definition of the intensity thresholds with double hysteresis threshold inspired by the Schmitt trigger: voxels below a lower threshold are classified as background, and those above a higher threshold are classified as neutrophils. The remaining voxels between these two levels are then classified as neutrophils if they are in contact with voxels above the high threshold, or as background otherwise. Both thresholds are automatically determined using Otsu's algorithm [195] (Figure 4.6).

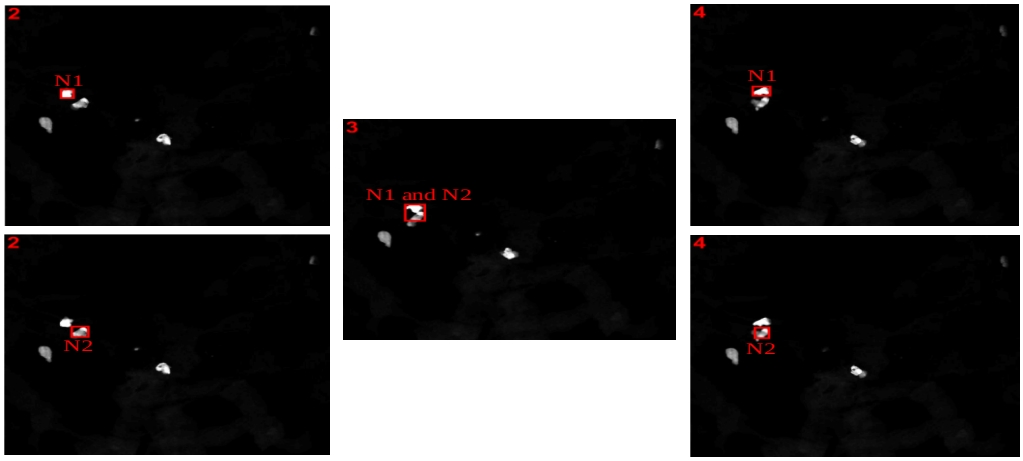
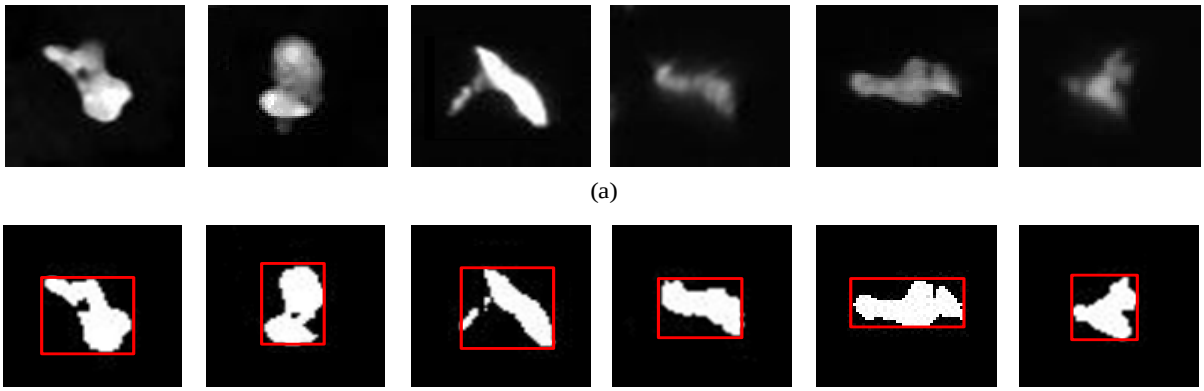


Figure 4.5: Tracking result for two neutrophils when they are jointed in frame3 and separated in frame4.



(a)

Figure 4.6: a) The predicted target image selected by our proposed cell tracking framework. b) binary image with accurate bounding boxes.

The complete cell tracking algorithm (Algorithm 1) is described as follows:

Algorithm 1 Multi channel feature learning

Step one: Training

Input: Training datasets ≔

$$S_{train} = \left\{ \left[(x_{t-1}, x_t), (d_{t-1,t}), (O_{t-1,t}) \right]_1, \dots, \left[(x_{t-1}, x_t), (d_{t-1,t}), (O_{t-1,t}) \right]_n \right\}$$

Aim: Minimise the cost function from (3) to optimize and update W.

Step two: Tracking

Input 1: Image sequences captured by microscope ≔

$$s_{track} = \{f_1, \dots, f_t\}.$$

Input 2: Set of candidates V= $v_1, \dots, v_N$. {N is the number of candidates}

Aim: Track the neutrophil which is assigned in first frame.

Output 1: The candidate image with highest probability.

Output 2: Accurate bounding box on the target by considering the neutrophil's area and threshold.

for i-th image from input

Sampling produce the candidate set V.

Observation function predicts the most probable candidate v_i, $i \in (1, N)$.

Visualisation function computes the selected neutrophil area $area$ and threshold (tr).

if $v_i > tr$, **then**

$v_i = 1$ (cell)

else

$v = 0$ (background)

end if

if $area(v_i) > \frac{3}{2} area(v_{i-1})$ **then**

$v_i = v_{i-1}$

end if

end for

4.5 Experiment and Results

This section represents the dataset, implementation details, performance evaluation, and comparison results.

4.5.1. Performance Evaluation

We validated the performance of our proposed tracking method in our dataset, represented on three video sequences which cover various challenging factors, such as deformation, fast motion, and collisions. Figures 4.7, 4.8, and 4.9 illustrate the qualitative tracking results of individual neutrophils with variations in size, shape, threshold, and density in our dataset. Our results clearly show a considerably improved performance in individual cell tracking.

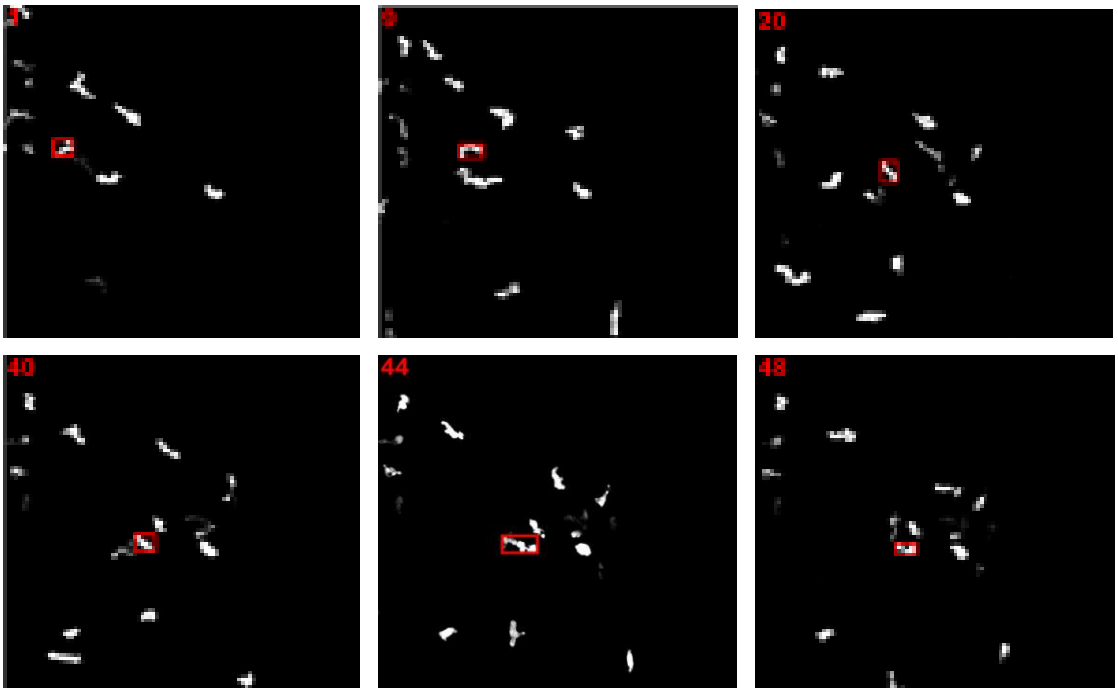


Figure 4.7: The tracking result of small neutrophil with shape variation and cramped in high density field.

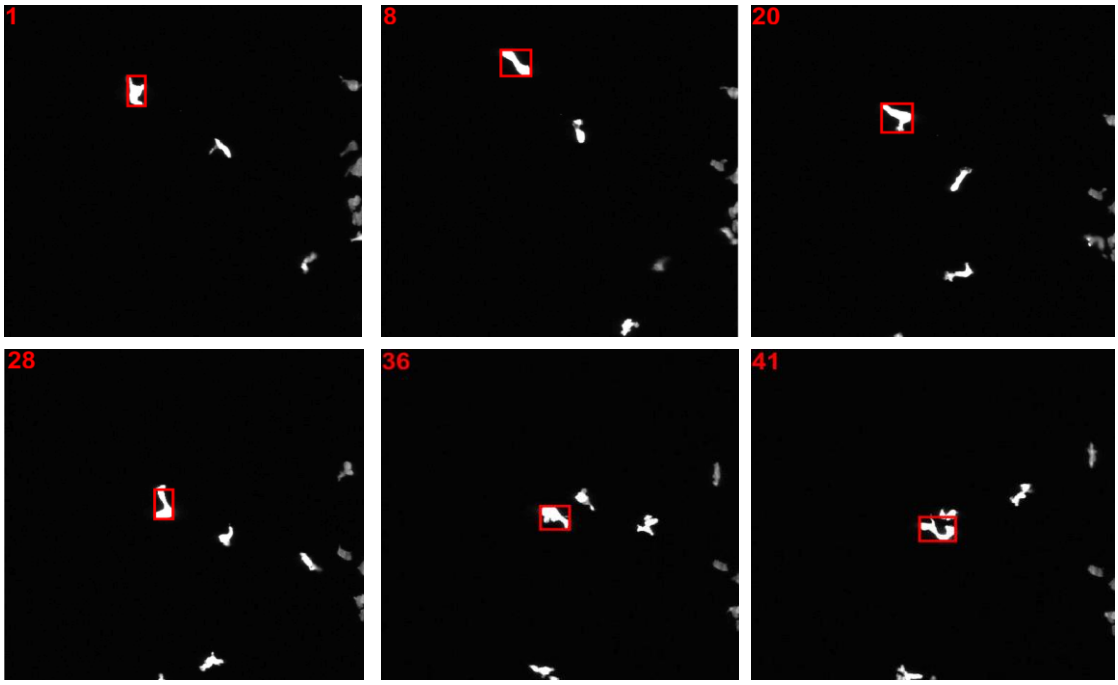


Figure 4.8: The tracking result of large neutrophil with shape variation in low density.

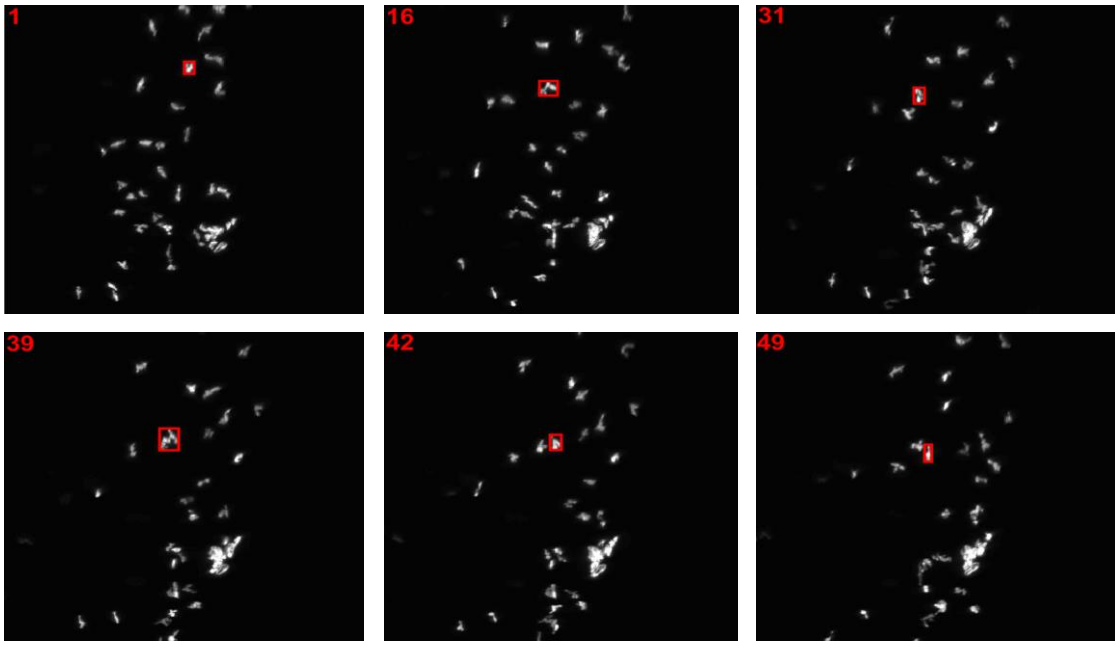


Figure 4.9: The tracking result of tiny and low threshold neutrophil with shape variation and cramped in high density field.

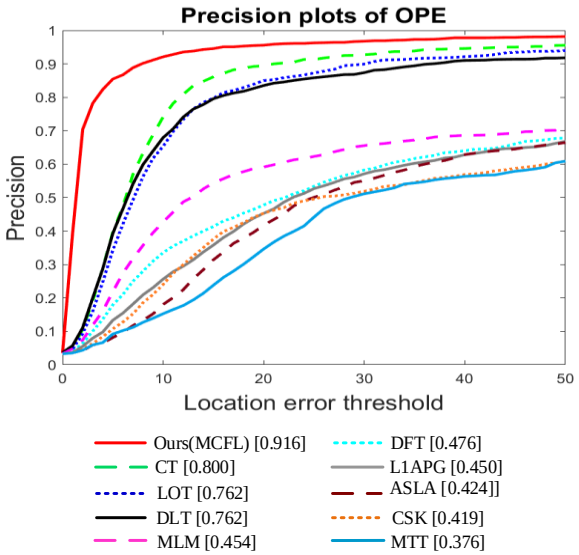
4.5.2. Robustness Estimation and Comparison with Other Methods

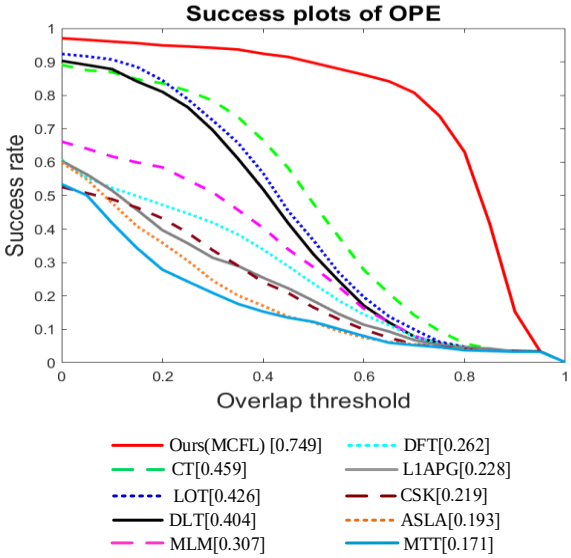
The evaluation methodology in [176] has been recently used in many visual tracking methods, and it generally comprises two different evaluation metrics: a success plot and a precision plot. The success plot measures the bounding box overlap between the predicted and ground truths, whereas the precision plot uses the Euclidean distance between the centre of the estimated targets and the ground truths. In this methodology, there are three commonly used evaluation methods: temporal robustness evaluation (TRE), start by different bounding boxes, which is based on a randomized starting frame, perturbing the initial bounding box randomly called spatial robustness evaluation (SRE), start by different bounding boxes, and one-pass evaluation (OPE) [176].

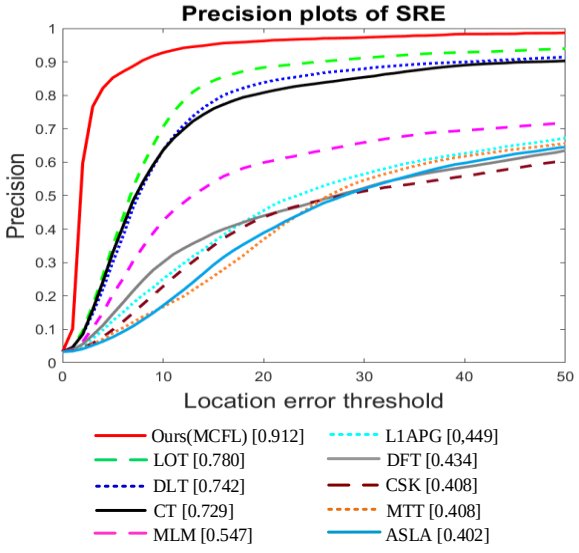
The precision and success plots at TRE, SRE, and OPE metrics are shown in Figure 4.10 for our proposed method compared with 9 other tracking methods which are tested in 10 randomly selected image sequences of our dataset (The characterization of each dataset including, data size and voxel size, image format, number of channels, and number of ground truth tracks for neutrophils, macrophages, and particle have been provided in Table 3.3, Chapter 3: Data Acquisition).

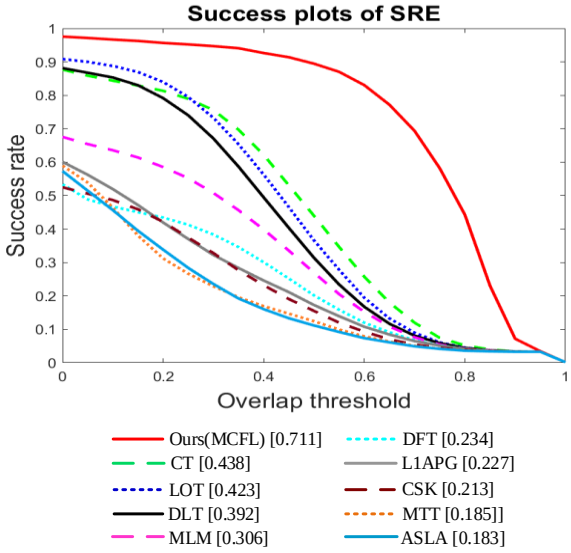
The original implementations of these trackers are available in [176] and prepared for evaluation metrices provided in online tracking benchmark [176]. These trackers view tracking as online processing procedure. In these approaches, location of the target object in the first frame is required by algorithm, then the object is tracked frame-by-frame, while the algorithm simultaneously updates the tracking model and rules.

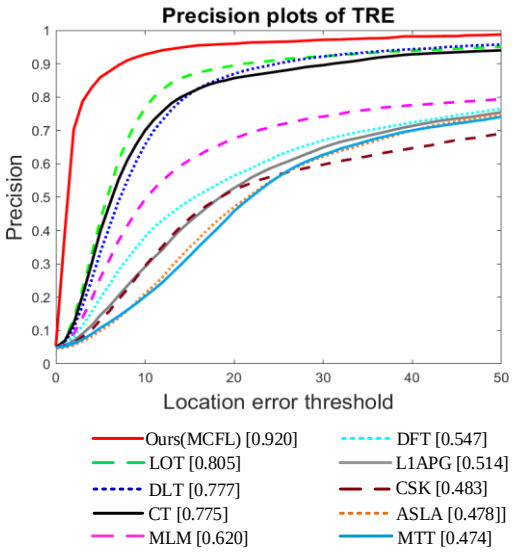
In terms of both success and precision evaluation metrics, our proposed model, performs remarkable results in comparison to other methods. In the precision plot of OPE, the precision score of our proposed MCFL is 0.916 and is ranked in first place. The next top four tracking methods are LOT [196] (0.800), DLT [116] (0.762), CT [197] (0.756), and MLT [12] (0.545). In the success plot of OPE, the area under curve (AUC) score of our MCFL model is 0.749, which is also ranked first. The next top four tracking methods are CT (0.459), LOT (0.426), DLT (0.404), and MLT (0.307). As shown in Figure 4.10, our proposed method performs the best result in both the TRE and SRE evaluation method in terms of precision and success metrics.


Precision plots of OPE
Precision
Location error threshold
Ours(MCFL) [0.916]
CT [0.800]
LOT [0.762]
DLT [0.762]
MLM [0.454]
DFT [0.476]
L1APG [0.450]
ASLA [0.424]]
CSK [0.419]
MTT [0.376]


Success plots of OPE
Success rate
Overlap threshold
Ours(MCFL) [0.749]
CT[0.459]
LOT[0.426]
DLT[0.404]
MLM[0.307]
DFT[0.262]
L1APG[0.228]
CSK[0.219]
ASLA[0.193]
MTT[0.171]


Precision plots of SRE
Precision
Location error threshold
Ours(MCFL) [0.912]
LOT [0.780]
DLT [0.742]
CT [0.729]
MLM [0.547]
L1APG [0,449]
DFT [0.434]
CSK [0.408]
MTT [0.408]
ASLA [0.402]


Success plots of SRE
Success rate
Overlap threshold
Ours(MCFL) [0.711]
CT [0.438]
LOT [0.423]
DLT [0.392]
MLM [0.306]
DFT [0.234]
L1APG [0.227]
CSK [0.213]
MTT [0.185]]
ASLA [0.183]


Precision plots of TRE
Precision
Location error threshold
Ours(MCFL) [0.920]
LOT [0.805]
DLT [0.777]
CT [0.775]
MLM [0.620]
DFT [0.547]
L1APG [0.514]
CSK [0.483]
ASLA [0.478]]
MTT [0.474]

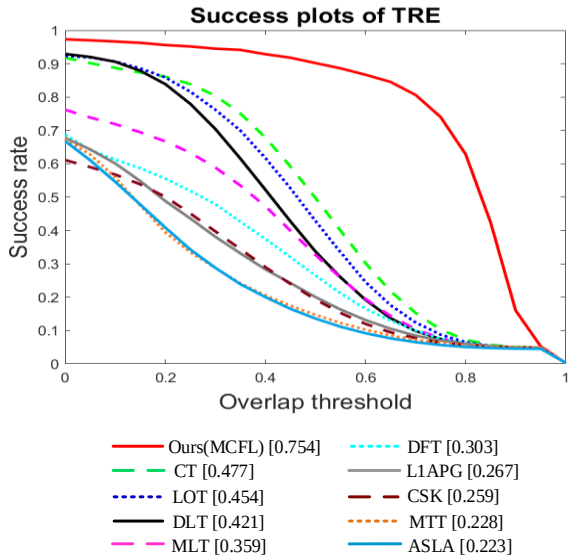

Success plots of TRE
Success rate
Overlap threshold
Ours(MCFL) [0.754]
CT [0.477]
LOT [0.454]
DLT [0.421]
MLT [0.359]
DFT [0.303]
L1APG [0.267]
CSK [0.259]
MTT [0.228]
ASLA [0.223]

Figure 4.10: The precision and success plots of OPE, TRE, and SRE for the 46 image sequences. The proposed tracking method (i.e., ours-MCFL in red) is ranked first among nine trackers including: CT [197], DLT [116], LOT [196], MLT[12], DFT [198], L1APG [199], CSK [200], MTT [201], and ASLA [202] in both the precision and success plots.

The performance of qualitative comparison results with other nine tracking methods on the same training and testing dataset is illustrated in Figure 4.11. The predicted targets are assigned by colour coded bounding boxes for each method. The tracking result verifies that our MCFL is a highly precise cell tracker by predicting the correct position of the target in all frames.

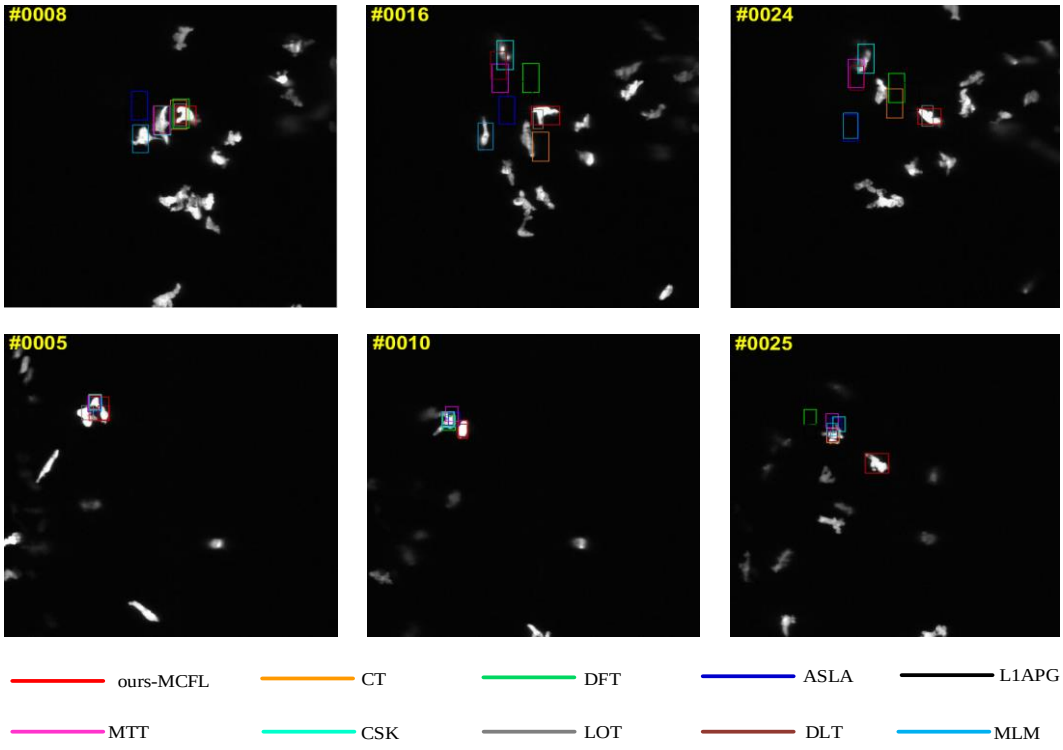


Figure 4.11: The qualitative results of the proposed tracking method (i.e., ours-MCFL) on two challenging tracking examples from our dataset.

4.5.3. Experiment Results on The Common Dataset

For further evaluation, our proposed MCFL was also validated on two different benchmarks: TB-100 [176] and the cell tracking dataset [190]. Four randomly selected examples of the qualitative comparison results are shown in Figure 4.12. The first two examples are the bright field cell tracking datasets from [190] and the other two example are the Bolt and Jogging dataset from the TB-100 dataset. This result indicates that our proposed method significantly outperforms the other tracking methods. To analyse our tracking method quantitively, we tested it on 5 cell image sequences and 5 videos of TB-100. The average centre location errors are summarized in Table 4.1, where the best and second-best results are highlighted in bold and underlined, respectively. Our proposed method, MCFL, achieved superior results on the cell tracking benchmark with the best score in dataset 1, 2, 3, and 5 and the second-best score in dataset 3. Furthermore, the proposed method has the best performance in the Jogging dataset and the second-best result in the Bolt dataset. While our tracker achieved superior results in both cell tracking benchmarks (zebrafish and the dataset in [190]), it did not performed as impressive with other various datasets, which is explainable by the fact that our tracker was only trained by cell images. Although the results in Car1, Car 24, and RedTeam are not the best or second-best, they are still in a good range.

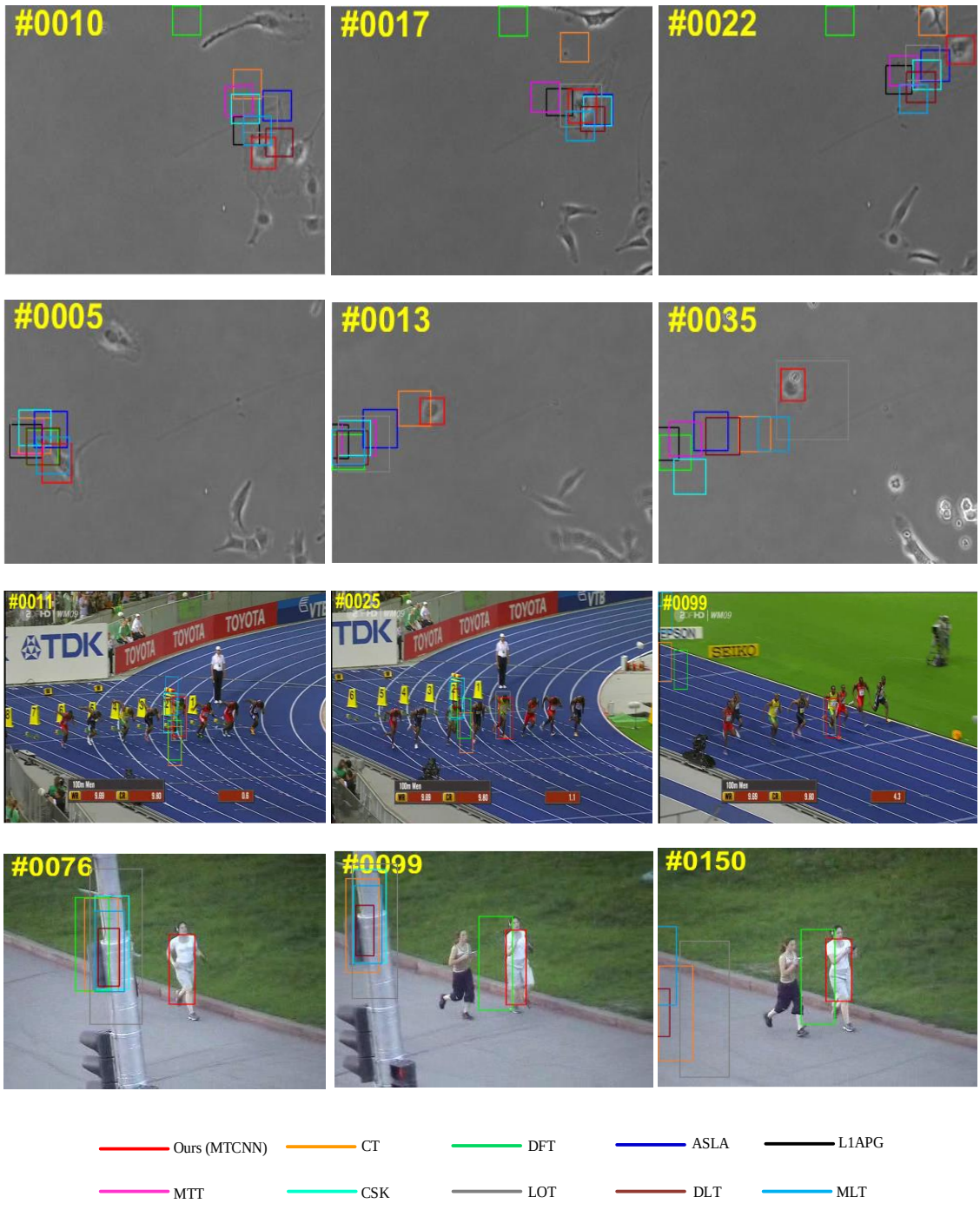


Figure 4.12: The qualitative results of the proposed tracking method (i.e., ours-MCFL) on two challenging sequences.

TABLE 4.1: THE AVERAGE CENTRE LOCATION ERROR ON TWO DIFFERENT DATASETS

Dataset	Ours (MCFL)	ASLA	CSK	CT	DFT	DLT	L1APG	LOT	MLM	MTT
1	**6.11**	32.86	20.18	56.13	50.35	6.99	30.68	18.53	15.93	29.89
2	**5.94**	56.82	64.41	15.62	35.81	36.57	57.97	49.00	31.24	33.48
3	**2.25**	17.56	14.74	3.87	27.85	3.32	3.36	6.22	7.34	3.22
4	3.62	7.58	44.41	5.03	31.85	**2.77**	14.72	40.48	7.55	28.05
5	**7.59**	35.15	46.04	18.09	42.28	20.95	27.31	35.94	16.02	33.51
Bolt	15.29	227.6	244.7	181.9	203.2	235.5	248.7	**9.71**	230.8	247.56
Car1	3.03	1.08	2.86	15.43	2.82	**1.07**	3.06	2.76	3.92	3.02
Car24	2.14	1.33	1.08	4.96	**0.85**	1.29	1.73	2.50	3.30	1.13
Jogging	**6.25**	44.92	43.99	51.29	25.31	43.94	36.73	43.92	46.69	47.62
RedTeam	4.17	1.79	1.80	5.26	1.61	**1.56**	2.31	3.53	46.50	1.62

4.6. Conclusion

We proposed a new cell tracking framework based on novel multi-channel feature learning model to track individual cells over 2D microscopic image sequences. Unlike previous works that extract spatial and temporal features manually or just employ new learning methods on spatial features such as classifying cells by their shape, in this work our proposed MCFL model learns the visual and motion features simultaneously from the time-lapse microscopy images. The CNN architecture is adapted as a spatial learning channel while the temporal features are learned through CRD and orientation learning channels. Sequential microscopic images of migrating neutrophils in zebrafish model used as the training and tracking datasets in this study. Extensive results on three datasets indicate the robustness of our proposed tracking method compared to the other 9 methods.

Although our method was designed to solve the difficulties of tracking individual neutrophils, there are still more challenging tasks in the field of biomedical imaging which need to be considered in our future works. First, we will attempt to expand our tracking method for multi-neutrophils tracking (currently it can only track a single cell). Second, as the source zebrafish dataset has 4 dimensions, we can develop our tracking algorithm to track neutrophils not only along x, y coordinates but also on z axis to increase the accuracy of the tracking performance when a neutrophil migrates through z axis. Third, the microscopy zebrafish dataset entails the second type of white blood cells called macrophage which can also be labelled through different fluorescent tags (in this study, we only export and track the neutrophils which are tagged with green fluorescent protein). We will attempt to extend the tracking method to track both types of white blood cells separately and predict their interactions.

Chapter 5: Tracking Leukocytes in Intravital Time Lapse Images Using 3D Cell Association Learning Network

Leukocytes are key cellular elements of the innate immune system in all vertebrates, which plays a crucial role in defending organisms against invading pathogens. Tracking these highly migratory and amorphous cells in *in vivo* models such as zebrafish embryos is a challenging task in cellular immunology. As temporal and special analysis of these imaging datasets by human operators is quite laborious, developing an automated cell tracking method is highly in demand. Despite the remarkable advances that have been made in cell detection, this field still lacks powerful algorithms to accurately associate the detected cell across time frames. The cell association challenge is mostly related to the amorphous nature of cells, and their complicated motion profile through their migratory paths. To tackle the 3D multi-cell association challenge, we propose a novel deep learning-based object linkage framework. For this aim, we trained the 3D cell association learning network (3D-CALN) with enough manually-labelled paired 3D images of a single fluorescent zebrafish's neutrophils from two consecutive frames. Our experiment results prove that deep learning is significantly applicable in cell linkage, particularly for tracking highly mobile and amorphous leukocytes. A comparison of our tracking accuracy with other available tracking algorithms shows that our approach performs well in relation to addressing cell tracking problems.

NOTE: The content of this chapter has been submitted to *the Artificial Intelligence in Medicine*.

Moghadam, R. M. and Chen, Y. P. P. "Tracking Leukocytes in Intravital Time Lapse Images Using 3D Cell Association Learning Network." *Artificial Intelligence in Medicine*. (Under Review).

5.1 Introduction

Leukocytes including neutrophils and macrophages are the first line of defence against invading pathogens [203]. To fulfil their duty in defending an organism against pathogens, these cells have evolved to migrate towards inflammation sites quickly and then neutralise the pathogens through different killing mechanisms including phagocytosis, releasing deadly granules and extracellular traps [204].

The emergence of antibiotic-resistant infections has shifted infection and immunology research towards finding alternative ways of controlling these infections, which is mostly achievable through a better understanding of the mechanisms by which the immune system defends the integrity of the biologic system against invading microorganisms [205]. Biologic processes could be studied within a living organism (*in vivo*) or within an experimentally conditioned environment in laboratory flasks (*in vitro*). While *in vitro* systems provide valuable information about the dynamics of cell behaviour, *in vivo* models have the advantage of studying biologic processes in their natural physiologic environment [14].

Despite their phytogenic distance from mammalians, the zebrafish has proven to be an excellent model for studying the dynamics of leukocyte behaviour *in vivo*, mainly due to its optical transparency and the feasibility of genetic manipulation [148], which provide the opportunity to capture meaningful information at the highest practical spatiotemporal resolution. The most common genetic manipulation of the zebrafish model is labelling cells or biologic components with fluorescent proteins, which then can be easily imaged using fluorescent microscopes [148, 150]. Fluorescent proteins are a class of proteins that are

excited by absorbing high-energy photons in a specific range and then releasing that energy by emitting a lower-energy visible photon in response [5].

The continuous refinement of fluorescent microscopy techniques including confocal microscopy [206], multiphoton microscopy [207] and light sheet microscopy [208] has remarkably improved the spatiotemporal resolution of imaging datasets at the cellular and subcellular levels, and at the same time, has made the imaging of more complex and dynamic biologic structures possible. As zebrafish *in vivo* studies are strongly associated with high-end microscopy techniques, this has led to the production of a huge amount of complex imaging datasets.

Despite the huge progress in development of flexible biologic models and imaging techniques, the field still lacks equally competitive automated image analysis algorithms, and human operators still do a better job when it comes to analysis of complex imaging datasets. This gap between what humans and algorithms can differentiate is to some extent due to the greater complexity of biologic objects and their intrinsically dynamic behaviours compared to the usual macro objects, such as cars or humans, which traditionally have been tracked by algorithms. In other words, compared to the aforementioned often-solid objects, bio particles lack morphologically definitive features such as shape, size and angles, as they may interact with each other, split or merge or go through deformations such as general or polarised elongation, expansion in size or even shrinking [11]. These morphological variations may, in turn, also change the motion pattern of biological objects over time [12]. So, proposing and developing a powerful tracking method to accurately analyse these complex cells is still essential.

With the advances in digital computing technology, many automatic object tracking methods [151] have been introduced in the field of computer vision that aim to track multiple bio particles (e.g. cells) over a sequence of complex three- (x, y, t) or four- (x, y, z, t) dimensional images. Conventional cell tracking approaches usually comprise two steps: (1) cell detection and (2) cell linkage.

During the detection phase, biologic objects and their locations are independently detected in each frame using diverse cell segmentation methods such as thresholding, active contours and deep neural network methods. Thresholding is one of the simplest and most commonly used methods for segmentation, which basically means defining a pre-set threshold value and comparing each image pixel to this threshold, in such a way that the pixels above and below the threshold will be labelled as foreground or background [169, 209, 210]. Active contour models, also known as snakes, are another detection method that characterise the boundaries of an object based on energy minimization [39]. Recently, with the development of deep learning [89], cell detection and segmentation methods have been optimised by learning the robust features of cells in biomedical images. Since 2015, several deep learning-based cell detection approaches have been proposed [211]. U-Net [101] and DeepCell [108] are two of the recent successful methods for instance segmentation, which work by classifying the pixel-level to the background, cell boundaries and cell interiors.

After object detection, the second step in object tracking is called object linkage [56], which is based on identifying and linking the detected cells along image sequences to form a track. Classical approaches include nearest neighbour, probabilistic models and linear programming.

Nearest neighbour is the most common and simplest approach. It links every detected cell to the nearest cell in the adjacent frame [57]. One example of this method is proposed in [212], which tracks human dendritic cell behaviour toward tumour cells based on a tightly interconnected system of immune cells. Additionally, detected objects in consecutive frames are linked according to the highest intersection over the union [213]. Although these approaches have low computational complexity, they may easily lead to mismatches for highly mobile cells or in fields with a high cell density.

The second cell-linkage approach relies on probabilistic models [59-62] such as Kalman Filter [214] and Particle Filter [215]. The Kalman Filter-based tracking searches the target cells in time lapse microscopy images by updating their state equation and minimizing their characteristic cost function to form a track [216]. In [217], the Kalman Filter is applied to a plant cell-tracking model to predict cell movement in a multicellular field. Then, the local graph matching method is applied to obtain cell pairs in the vicinity of the predicted position.

Linear programming is the third example of cell-linking approaches. It can determine complex cell behaviour such as merging, splitting and disappearing [78]. In this approach, the most likely pair of particles in consecutive frames, including newly appearing (birth) or disappearing particles (death), are identified by solving the linear-assignment problem [78]. Deep learning architecture has recently been used to solve the tracking problem, mostly focusing on tracking by detection [12, 106, 113] and classification [218].

Although these methods perform fairly well in detection-based object tracking on single cells, their performance is compromised when it comes to simultaneous multi-cell tracking,

mainly due to the highly dynamic state of cells, including changes in the number of cells due to merging and splitting or appearing and disappearing from the field, which makes them challenging to track over time. As result, it is important to develop accurate multiple cells tracking approaches based on deep learning techniques.

In this study, to tackle the common challenges of multi-cell tracking, we propose a deep learning-based cell linkage method for tracking zebrafish neutrophils in 3D fluorescence microscopy image sequences. The algorithm is first pre-trained through a large training set including cell pairs between subsequent frames. For this purpose, the robust features of moving cells are learned through 3D convolutional neural networks (3D-CNN). Then, the tracking procedure is initiated by predicting the most probable matches in the adjacent frames during image sequences. The proposed method conquers most of the challenges of multiple cell tracking, including rapid deformation, splitting and merging, as well as newly appearing and disappearing cells.

To outline our chapter, the proposed 3D cell association learning networks are presented in Section 5.2. Next, in Section 5.3, the learned multiple cells tracking framework is proposed including sampling and post-processing. In Section 5.4, the quantitative and qualitative experimental results are presented in addition to details on the construction of manually-labelled training and tracking datasets. Finally, the conclusion and potential implications are discussed in Section 5.5.

5.2 Proposed 3D Cell Association Learning Network

In this study, we propose the 3D cell association learning network (3D-CALN) as a training model that links identical cells (i.e. neutrophil) in successive frames. For this purpose, a dataset of true and false pairs was built by manually cropping and labelling individual cells in two neighbouring frames. In other words, when a pair is denoted as true, it is the same neutrophil in two time points, otherwise it is connected to a wrong cell in the next frame (false pair). The model proposed is based on a novel 3D-CNN to learn the robust features of complex cell behaviour.

To train the algorithm to recognise cell pairs, cropped images were passed through the 3D-CNN network to extract the morphological features. Figure 5.1 (a) shows the 3D-CNN architecture of the images of two individual cells in adjacent frames, namely the previous $(t-1)$ and current (t) frames. As the size of training images are small, a CNN architecture with more layers is time-consuming, with no improvement in the recognition results. To match the recognition performance and processing time, we designed our 3D-CNN model with three 3D-convolutions and max pooling, and two fully connected layers. In this process, two 3D single cell image (inputs) are convolved with a set of parametrized $3 \times 3 \times 3$ kernels in each 3D convolutional layer. All convolutional layers are paired with a rectified linear unit (ReLU) activation [219]. The max pooling layers perform a max operation for voxels in an 3D-window to reduce the input size by stride s=2. The two fully connected layers are similar to traditional neural networks, in which each perceptron unit is connected to inputs from all units of the previous layer and outputs to all units in the next layer [220].

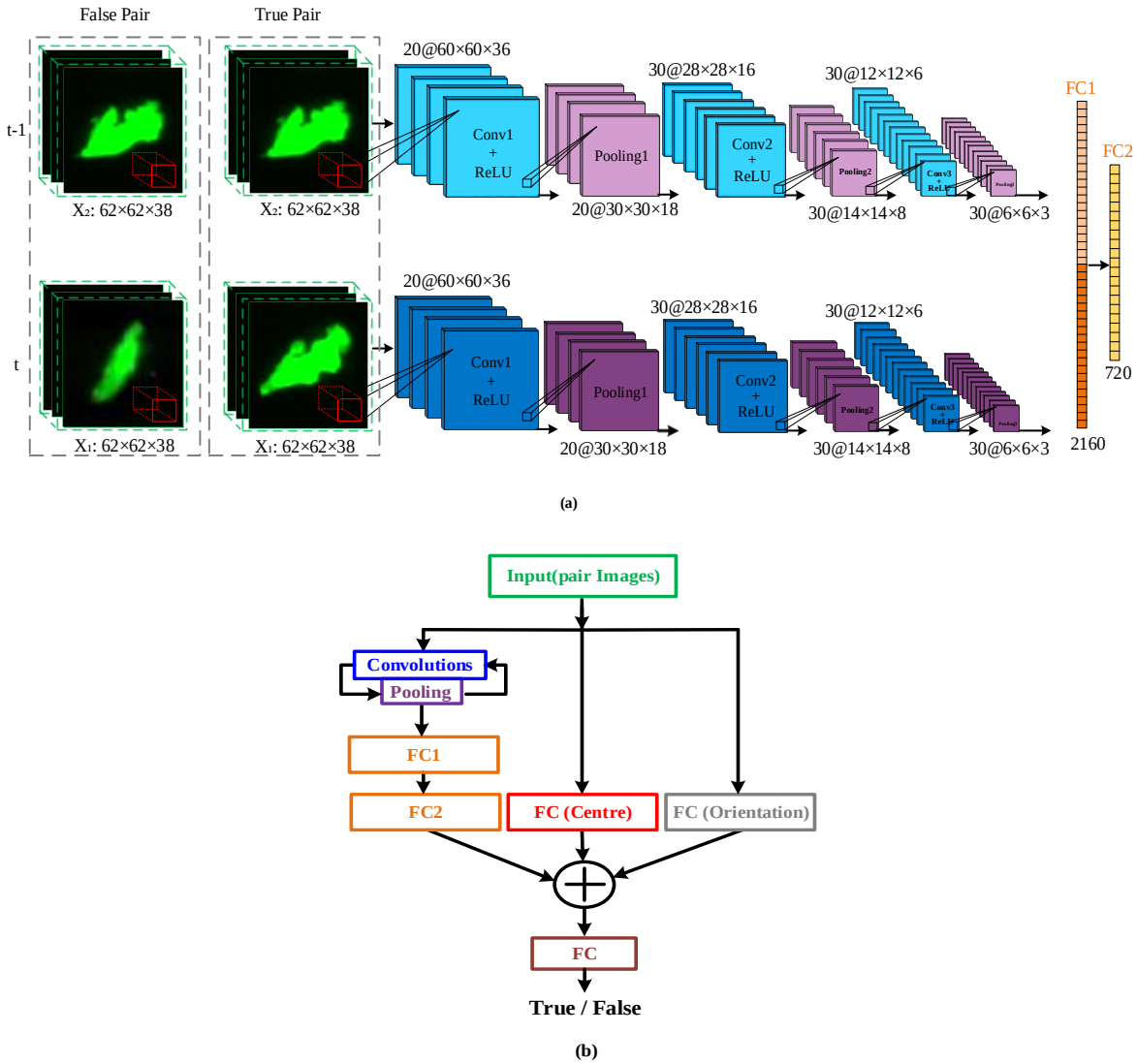


Figure 5.1: A cell association learning network is proposed for training the characteristics and behaviour of neighbouring cells during time lapse microscopy images. (a) Two individual cells (x_1, x_2) of consecutive frames $(t, t-1)$ fed to 3D-CNN (including three convolution-pooling layers and two fully connected layers). The lighter and darker cubes in each convolution and pooling neuron indicate the output feature map of one cell. The rectified linear unit (ReLU) activation function is applied on every convolutional layer. (b) Three parameters, including morphological characteristics, centre relocation distance (CRD) and cell motion direction are concatenated and then followed by a final fully connected layer and cost function to assign the true and false pairs.

To address the challenge of the amorphous nature of neutrophils (lack of consistent morphology in 3D image), the 3D-CNN network is joined to two new parameters, namely centre relocation distance (CRD) and cell motion direction. CRD is defined as the displacement of cell centres between adjacent frames. Cell motion direction is calculated based on an assumption taken from the keyhole algorithm [221], in which a moving object (e.g. neutrophil) will most probably follow its previous direction (time $t-1$ to time t).

Two new fully connected layers are defined as $FC_{centre}$and $FC_{Orientation}$which are used to train the network according to CRD and cell motion direction. The three input parameters are then concatenated and processed by a final fully connected layer followed by a cost function, which assigns a single similarity score between 0 and 1, indicating how likely it is that the two successive images belong to the same cell (Figure 5.1 (b)).

For more details on the training structure, the mathematical description is explained as follows. The 3D-CALN is trained with K input training samples as $I_{train} = \{(X^1, d^1, O^1), \dots, (X^K, d^K, O^K)\}$, in a way that each sample is composed of three elements: image, CRD and motion direction. The image element includes a pair of cells $X = \{x_1, x_2\}$ in two consecutive time points of size of $62 \times 62 \times 38 \times 2$. x_1 and x_2 are the individual cell images. CRD and motion direction are defined as below, respectively

$$d = \|C_t - C_{t-1}\| \tag{1}$$

$$O = \alpha_{t-1,t} \tag{2}$$

where C_t and C_{t-1} are the cell centre locations in neighbouring frames and $\alpha_{t-1,t}$ is an angle of cell movement respective to the x axis.

Suppose that 3D-CNN model has L layers and $\Theta_x^l = \{\theta_1 \dots \theta_J\}$ are the weights at layer l-th, so the output feature maps of layer l can be defined by

$$a^l = \phi(\Theta_x^l a^{l-1} + b_x^l) \tag{3}$$

which acts as input feature maps for the next layer. $\phi(\cdot)$ is the ReLU activation function and $a^0 = x$. So, as there are two inputs, each input is fed to one 3D-CNN including the convolution and pooling layers (upper and lower 3D-CNN in Figure 5.1). To find the similarity appearance score of single cells between two adjacent frames (true or false classes), these two inputs must convolve with similar kernels in each layer $\Theta_{x_1}^l = \Theta_{x_2}^l$. This process is followed to the last layer a^L, which is then fed to the two fully connected layers in each 3D-CNN. These two fully connected layers are concatenated to form one fully connected layer and then followed by another fully connected layer.

Next, the last layer of 3D-CNN concatenates to two fully connected layers associated with CRD and motion direction parameters. The output feature map is defined as:

$$A = \frac{1}{3}(a_X + a_C + a_o) \tag{4}$$

where $a_X = \left(a_{x_1}^L, a_{x_2}^L\right)$, $a_c = \phi(\Theta_C d + b_c)$, and $a_O = \phi(\Theta_O O + b_O)$. $\Theta_{\mathrm{C}}$and Θ_{O} are the CRD and motion direction learnable kernels.

The 3D-CALN is trained by minimizing the following loss function over all trajectories defined by:

$$\mathcal{L}(\Theta) = argmin \sum_j^J \left\| F\left(Z_j^L\right) - y_j \right\|^2 + \frac{\lambda}{2}\|\Theta\|^2 \tag{5}$$

where function $F(\cdot)$ denotes the predicted output, $Z^L = a^{L-1}.\Theta^{L-1}, l = 2, \dots, L)$, Θ represents the learnable weights at the last task, λ illustrates the weight decay parameters, and J is the number of neurons in the last feature map (last fully connected).

To minimize the loss function, networks weights and biases are adjusted through backpropagation. The level of adjustment is determined by computing the gradients of the cost function with respect to those parameters. The partial derivative of $\mathcal{L}(\Theta)$ is defined as below

$$\frac{\partial}{\partial \Theta_{i,j}^{l}} \mathcal{L}(\Theta) = \frac{\partial \mathcal{L}}{\partial Z_j^L} \frac{\partial Z_j^L}{\partial \Theta_{i,j}^{l}} \tag{6}$$

The error value of last layer (output) is defined as

$$\delta_j^L = \frac{\partial \mathcal{L}}{\partial Z_j^L} = F\left(Z_j^L\right) - y_j \tag{7}$$

The error value of the last fully connected layer is defined as

$$\frac{\partial Z_j^L}{\partial \Theta_{i,j}^{l}} = \frac{\partial Z_j^L}{\partial Z_j^{L-1}{}_j} \frac{\partial Z_j^{L-1}}{\partial \Theta_{i,j}^{l}} \tag{8}$$

$$\delta_j^{L-1} = \Theta^{L-1}.\delta_j^L \tag{9}$$

as: $Z_j^{L-1} = \Theta^{L-2}.A_j$ and $A_j = (a_X + a_c + a_s)_j$

$$\delta_j^{L-2} = \frac{\partial Z_j^{L-1}}{\partial \Theta_{i,j}^{l}} := (\delta_{c\,j}, \delta_{s\,j}, \delta_{x\,j}^{\;Lf}) \tag{10}$$

$$\delta_{c\,j} = \Theta^{L-2}.\delta_j^{L-1}.\frac{\partial \phi(\mathrm{z}_c)}{\Theta_c} \tag{11}$$

$$\delta_{s_j} = \Theta^{L-2}.\delta_j^{L-1}.\frac{\partial\phi(z_s)}{\Theta_s} \tag{12}$$

$\frac{\partial\phi(z)}{\Theta}$ is the derivative of ReLU activation function, which can be defined as:

$$\frac{\partial\phi(z)}{\Theta} = \begin{Bmatrix} 0 & z > 0 \\ 1 & z < 0 \end{Bmatrix} \tag{13}$$

Suppose $\delta_{x_j}^{Lf}$ and Θ_x^f represent the error value and weight of last fully connected of CNN, then they are defined as

$$\delta_{x_j}^{Lf} = \Theta^{L-2}.\delta_j^{L-1} \tag{14}$$

$$\delta_{x_j}^{Lf-1} = \Theta_x^f.\delta_{x_j}^{L} \tag{15}$$

The error values of the CNN's hidden layers are defined as follows

$$\delta_{x_j}^{l} = \Theta_x^l.\delta_{x_j}^{l+1}.\frac{\partial\phi(z^l)}{\Theta_x} \tag{16}$$

The gradient can be computed as

$$\frac{\partial\mathcal{L}}{\partial\Theta_{c_j}} = d.\delta_{c_j} \tag{17}$$

$$\frac{\partial\mathcal{L}}{\partial\Theta_{s_j}} = O.\delta_{s_j} \tag{18}$$

$$\frac{\partial\mathcal{L}}{\partial\Theta_{x_j}} = \sum_{l}^{L} a_{x_j}^{l}\delta_{x_j}^{l+1} \tag{19}$$

The model's weights then can be updated as follows

$$\Theta_c = \Theta_c - \varepsilon.\frac{\partial\mathcal{L}}{\partial\Theta_{c_j}} \tag{20}$$

$$\Theta_s = \Theta_s - \varepsilon.\frac{\partial\mathcal{L}}{\partial\Theta_{s_j}} \tag{21}$$

$$\Theta_x = \Theta_x - \varepsilon . \frac{\partial \mathcal{L}}{\partial \Theta_{x_j}} \tag{22}$$

The computed gradient and updated weights of all fully connected layers can be obtained similarly. The gradient and updated values of bias in each task are defined as

$$\frac{\partial \mathcal{L}}{\partial \mathrm{b}_j^l} = \frac{\partial \mathcal{L}}{\partial \mathrm{Z}_j^l} . \frac{\partial z_j^l}{\partial \mathrm{b}_j^l} \ , \qquad \frac{\partial z_j^l}{\partial \mathrm{b}_j^l} = 1 \quad \Rightarrow \quad \frac{\partial \mathcal{L}}{\partial \mathrm{b}_j^l} = \delta_j^l \tag{23}$$

$$b_j^l = b_j^l - \varepsilon . \delta_j^l \tag{24}$$

5.3 Learned Multi-Cell Tracking framework

Our proposed multi-cell tracking framework (Figure 5.2) is based on cell linkage during times in which the positions of the individual cells are pre-detected and pre-assigned by bounding boxes. Since the aim here was to precisely evaluate the accuracy of our proposed tracking model on cell linkage, we did not utilize any detection model and the tracking analyses are performed on the ground truth bounding boxes (the single cells in the input images of proposed framework on Figure 5.2 are assigned by ground truth bounding boxes). The aim of our proposed framework is denoted as an association function $\mathrm{h}: \mathrm{I}_{t-1} \rightarrow \mathrm{I}_t$ where I_t is a 3D image ($512 \times 512 \times 38$) at frame t. The proposed multi-cell tracking framework includes three tasks: sampling, 3D-CALN, and post-processing. A sampling function generates a set of possible pairs corresponding to all neighbouring cells in two time points by concatenating detected single cells, which are cropped and resized to $62 \times 62 \times 38$. Then, the optimized samples are fed to the trained 3D-CALN as an input batch to be predicted as true and false classes, where the true class indicates that the cell in the current frame should be associated with the cell in the previous frame and the false class

indicates that it should not. In this model, the final tracking result will be extracted from the most likely pairs after the post-processing task.

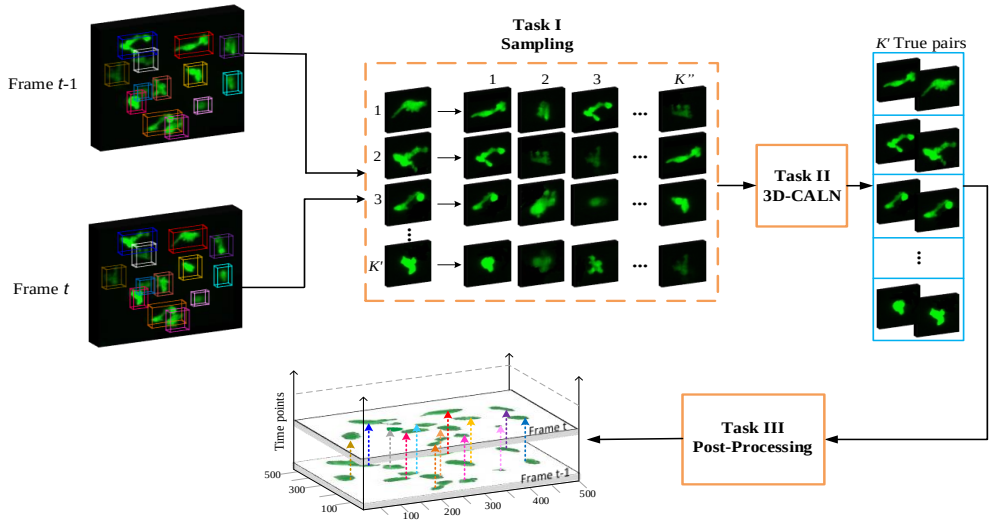


Figure 5.2: Proposed framework of multi-cell tracking method comprising three tasks: sampling, 3D-CALN and post-processing. Tracking is based on defining the highest probable matches in two adjacent frames. The detected K and K' number of neutrophils from two frames $t-1$ and t are used in the sampling function to generate a set of $K' \times K$ paired cells. The sampled set ($K' \times K''$) is then optimized by eliminating the pairs whose distance is significantly greater than the maximum cell relocation distance (K'' is the number of most probable cells in frame $t-1$, which can match to the cells in frame t). The true pairs are then determined by the trained 3D-CALN and finally after the post-processing task, they form tracks.

5.3.1 Task I (Sampling)

The sampling function matches the individual detected cells in two adjacent frames. Suppose all the detected cells at t-th frame in the sequence $t = 1, \dots, T$ denoted as

$$\mathbb{N}_t = \{N_t^1, \dots, N_t^K\} \in I_t \tag{25}$$

where $N_t^k = (X^k, d^k, O^k), k = 1, \dots, K$ represents each cell state vector including images, CRD and motion direction. The matching set between cells $\mathbb{N}_{t-1}$ and $\mathbb{N}_t$ are defined by

$$S_t = \{s_t^{1,1}, s_t^{1,2}, \dots, s_t^{K,K'}\} \tag{26}$$

where $s_t^{k,k'} = \{N_{t-1}^k || N_t^{k'}\}$ is a set of $M = K \times K'$ samples generated from concatenating all cells in the previous $t-1$ and current t frame (to simplify the notation of cells in the previous and current frames, we call them parents and children, respectively). K and K' indicate the total number of parents and children, respectively. Due to the abundance of samples, we restrict our selection to those samples in which the distance between parents and children cells is less than 3 times the cell diameter (this is much higher than the maximum cell relocation distance). Thus, our new optimized sampled set is defined as $S_t^{M'}$, $M' \subseteq M$.

5.3.2 Task II (3D-CALN)

A set of samples $S_t^{M'}$ fed into a trained 3D-CALN is summarized as $f : \mathbb{N}_{t-1} \rightarrow \mathbb{N}_t$ which is defined as cell association function in frame $t-1$ and frame t. 3D-CALN is trained based on the classifier to find the most related parent-child pairs (true). The output vector $P_t = f(S_t^{M'})$ can be interpreted as M' assignment probabilities in the true and false

classes. Thus, a set of K' (number of children) highest confidence matches $S_t^{K'}$ determined based on P_t which is defined as

$$S_t^{K'} := \left\{ \forall k', s_t^{k,k'} \in S_t^{M'} \mid p_t^{k,k'} = \max_{k'} P_t \right\}. \quad (27)$$

5.3.3 Task III (Post-Processing)

Collisions between cells have been detected as an important post-processing step. A collision can be distinguished by analysing cell volume, in the current frame using the following condition $V_{N^k} > \mu_N + \frac{5}{2}\sigma_N$, where V_{N^k} is the cell volume and μ_N and σ_N are the mean and standard deviation of the cells in frame t, respectively. In other words, if the volume of a cell exceeds the average volume plus $\frac{5}{2}$ standard deviation, then the cell is a form of two closed cells (merged). Since merge neutrophils do not move significantly, their characteristics (X, d, O) are replace by those at frame $t-1$ (before merging) until they split into two disjointed cells. Moreover, by measuring and comparing the cell, volumes in each frame, our proposed tracking algorithm can eliminate the over-segmented cells which are the result of a deficient detection method.

5.3.4 Cell Event Identification

In addition to one-to-one neutrophil mapping, the object events of cells include four types: newly appearing cells (birth), disappearing cells (death), split and merge, as shown in Figure 5.3. The challenges of birth, death, merge and split are considered in our tracking model through two conditions: condition I: when the number of children is less than the

parents ($K' < K$), there must be two assumptions: the cell at time $t-1$ does not appear in the current frame (death) or it is merged with another cell that has been explained in task III. Condition 2: when the number of children is more than parents ($K' > K$) we may have $b_t = K' - K$ newly appearing cells in frame t or the merged cells in the previous frame are separated in the current frame. In some cases, both assumptions in either condition 1 or 2 occur while $K' = K$.

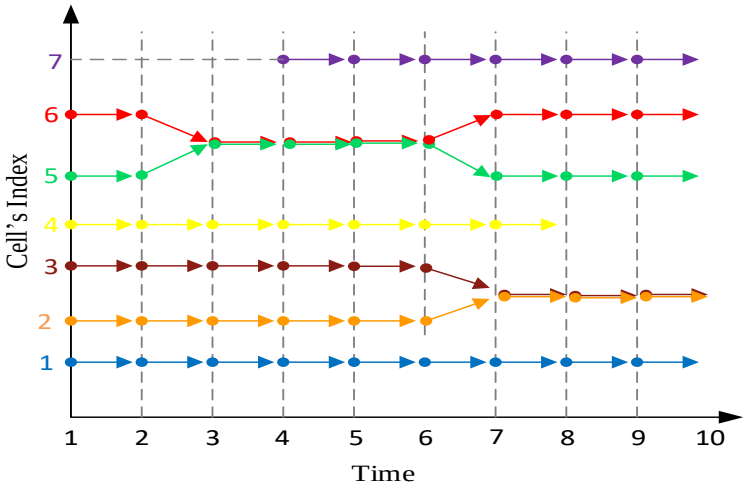


Figure 5.3: Cell events are displayed in 7 cells during 10 time points. Cell 1 is a normal cell that is moved constantly. Cell 2 and 3 are merged from time point 7. Cell 4 disappears from time point 8 (dead). Cell 5 and 6 are merged at time point 3 and they are then separated from time point 6. Cell 7 appears (born) at time 4.

Birth, death and split are defined automatically through 3D-CALN, while merge is specified in the post-processing task. As the number of predicted matches are determined according to the number of children, the death event is introduced when there is no child to be matched to the parents, which is defined as $\forall\, k = d, s_t^{k,k'} \notin S_t^{K'}$ (d is a death cell in the

current frame). However, in the birth event, there is a parent to match to a born cell, but it is classified as a false pair in the 3D-CALN model as

$$\forall k' = b\ , s_t^{k,k'} \in S_t^{K'}, p_t^{k,k'} \leq \gamma \tag{28}$$

where γ is an assigned threshold of acceptance for true pairs. Two or more cells are known as separated cells if the intersection of true matched cells is not empty (a parent has more than one child), which is illustrated as

$$\left\{\forall\ k, k', s_t^{k,k\prime} \in S_t^{K\prime}, \qquad \prod_k s_t^{k,k\prime} \neq \emptyset\right\}. \tag{29}$$

The completed proposed multi-cells framework is described in Algorithm 1.

Algorithm 1 learned Multi-Cell Tracking

Training:

$I_{train} = \{(X^1, d^1, O^1), \dots, (X^K, d^K, O^K)\}$: K input training set

Training the parameters Θ_X, Θ_C, Θ_O by minimizing the cost function.

Classifying cell pairs to true and false classes

Tracking:

$I_{track} = \{I_1, \dots, I_T\}$in the T sequence frames

for every I_t **do**

Defining a vector $\mathbb{N}_t = \{N_t^1, \dots, N_t^K\} \in I_t$ at frame t from detected cells.

Sampling M pairs by concatenating detected cells in two adjacent frames defined as $S_t = \{s_t^{1,1}, s_t^{1,2}, \dots, s_t^{K,K\prime}\}$.

Optimising samples to a set of M'samples defined as $S_t^{M\prime}$.

Computing the probabilities for all sampled pairs P_t through trained function.

Determining best matches through highest probability scores all $k = 1, \dots K$.

Event identification:

if $K' < K$ **then**

The pair is classified as death or merge

elseif $K' > K$**then**

The pair is classified as birth or split

if $p^{k,k\prime} < \gamma$ **then**

assigning as birth

else

assigning as split

end if

end if

Post-processing

if volume of $N^{k\prime}$ >mean +$\frac{5}{2}$ × std **then**

$N^{k\prime} \leftarrow N^k$

end if

end for

5.4 Experiments and Results

5.4.1 Training Setup

The source dataset used in the tracking experiments included 51 microscopy 3D image sequences (4D) ranging from 70 frames on average and containing an average of 12 moved neutrophils, which were individually cropped and resized to the size $62 \times 62 \times 38$ to match the input size of the training model (More details on the data acquisition can be found in Chapter 3: Data Acquisition). Eventually, we generated a set of 55,000 pairs by concatenating individual cropped cells in two successive frames, which are labelled by two-category including 45% true and 55% false to bias the classifier. In other words, a particular cell in frames $\mathrm{t}-1$ and t forms a pair of cells in the true class $(\mathrm{y}=1)$, while the false class $(\mathrm{y}=0)$ is a pair with different cells in two consecutive frames. To improve the denoising ability and robustness of the model, we also added some Gaussian noise to true samples. The source dataset was split into 60% for training, 22% for validation and 18% for testing. As neutrophils are of various shapes and motion profiles, our dataset could be considered as highly challenging compared to other existing visual tracking datasets.

5.4.2 Qualitative Validation of the Multi-Cell Tracking Framework on Our Dataset

We validated the performance of our proposed multi-cell tracking framework in our 4D test dataset. It is presented through two examples of a real dataset taken from neutrophils in live zebrafish, which covers various challenging factors, such as fast motion, cell merge and split (Figure 5.4). The results are shown for either high or low SNR examples (Figure 5.4 (a, b)). The colour coded bonding boxes display the neutrophils' positions and

identities. The figures are shown in 2D by applying the maximum intensity projection on the 3D image to be visually obvious.

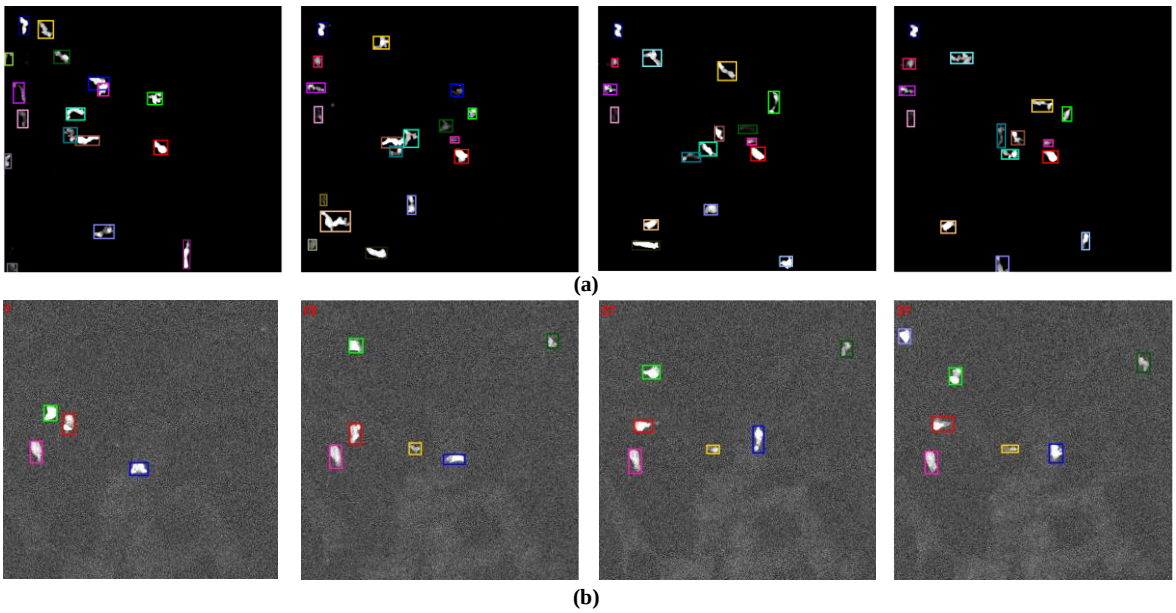


Figure 5.4: The tracking result of our proposed framework running on two 4D real data of fluorescent neutrophil migration on zebrafish larvae. The data are 3D image sequences including (a) maximum 23 neutrophils over 44 frames in high SNR (SNR=16dB) and (b) maximum 7 neutrophils over 31 frames in low SNR (SNR=8dB). The tracks are identified by colour-coded bounding boxes over the time. Newly appearing neutrophils (new tracks) are specified by new coloured bounding boxes.

The capability to track cells over the 4D dataset is the advantage of our proposed multi-cell tracking framework. Our results demonstrate a considerably improved performance in 3D cell tracking in which two cells collide or merge in 2D (x, y), but they are separated in 3D (x, y, z) (Figure 5.5). These results demonstrate three tracking examples that are identified by the colour-coded bounding boxes.

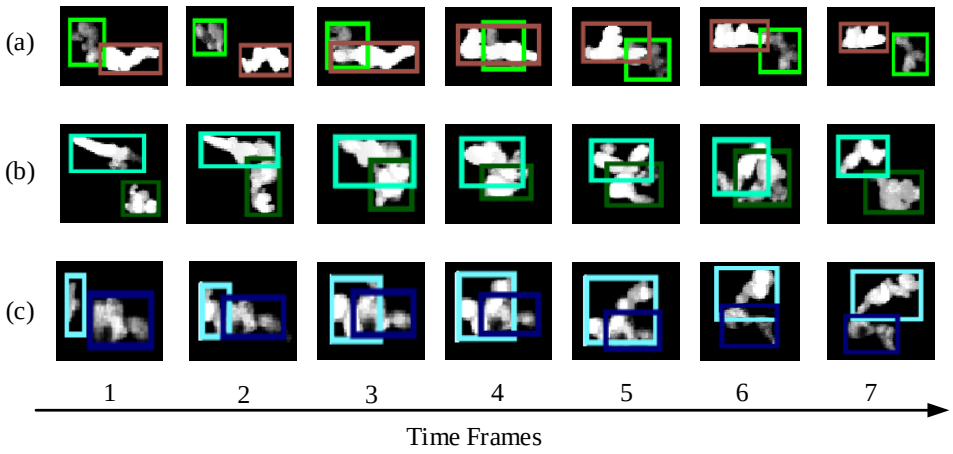


Figure 5.5: The advantages of 3D neutrophil tracking are shown in three situations over 7 frames in which cells are collided in x and y coordinates while being in different depth (z coordinate). (a) Two cells are moving toward each other. (b) Cells are moving in the same direction. (c) One new cell is appearing in the field of the microscopy.

The performance of our proposed tracking algorithm was qualitatively compared to five other tracking methods by running them on testing image sequences including maximum 7 neutrophils in 31 frames (Figure 5.6). In the IOU-Tracker [213], the threshold σ_{IOU} is set to 0.1 to maximize the tracking performance. The test results on LAP, Nearest Neighbour and Linear-Motion methods are derived from Trackmate software package [81] . We found that the most desired tracking results of these methods are achieved by adjusting the search radius to 15-20 pixel, which is relative to the size of the cells (the cells in our dataset are mostly in this range), and the maximum frame gap to 2. The comparison results demonstrated in Figure 5.6 indicate the superior performance of our proposed method in three situations. First, when two cells are too close to be separated as two dissociated cells, our method identifies them as merged cells and continues tracking them after separation. As shown in Figure 5.6 (i), IOU-Tracker [213], Nearest Neighbour, Linear-Motion and

Keyhole [221] lost the association, which causes the new track (on one of the collided cells) in the next time point to start with a new tracking ID (cells are identified as newly appearing ones), while IOU-Tracker and Keyhole track the wrong cell after the merge/split situation. Although LAP misses the cell positions at the merge point, it tracks the cells correctly by connecting the cell positions before merge and after separation. Second, our method illustrates the highest level of tracking consistency when cells move fast and asymmetrically. To display the speed of the cells' movement at each time point, we plotted the velocity on the ground truth tracks with colour codes (Figure 5.6 (a)). The speed of the moving cells increases, as shown by the change from darker to lighter colours. In the cases of IOU-Tracker, LAP, Linear Motion and Nearest Neighbour, the tracks are occasionally broken due to the cells' complex motion profile, but these trackers perform significantly better at a lower cell velocity and with more stable cell displacement (Figure 5.6 (iv)). Although Keyhole achieves a better performance when tracking fast migrating cells, the performance drops when the cell changes its direction sharply. Thus, these features distinguish our method from other multi-object tracking algorithms. In other words, the outstanding hallmark of our method is its capacity to track amorphous and complex motion features based on the trained 3D-CALN.

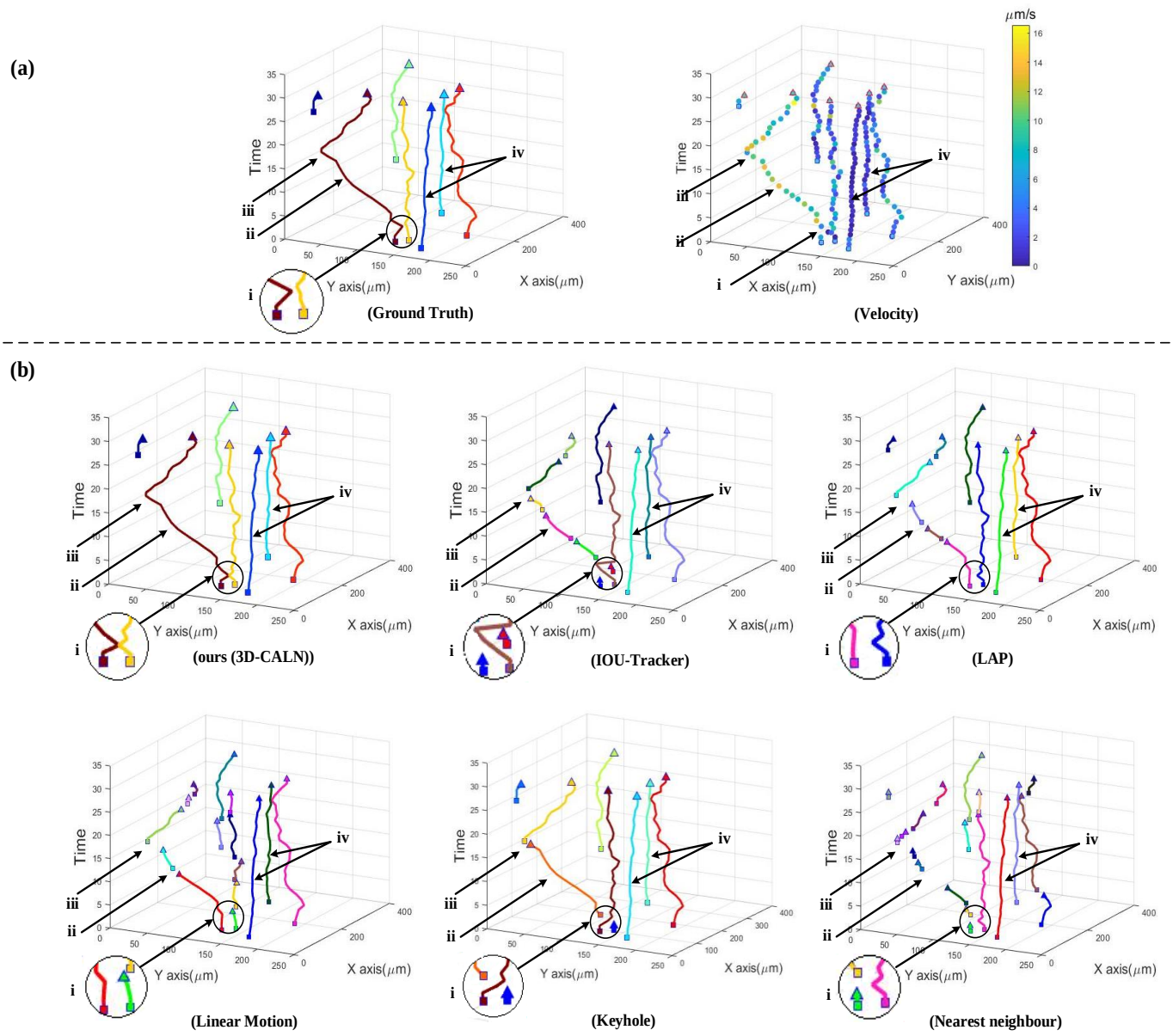


Figure 5.6: Tracking performance of the test real 4D imaging dataset including a maximum of 7 cells and 31 frames. (a) Ground truth of (X, Y, T) coordinates is shown as well as cell velocity which is depicted by color codes in the 3D plot in which the velocity increases as shown by the lighter colors. (b) Our proposed framework achieves remarkable results in regard to the most accurate tracking compared to the other 5 tracking algorithms. The tracking accuracy is studied in four situations (i, ii, iii, iv). Index i illustrates the merging point. Complex motion characteristics of tracks are shown in index ii and iii. Stable and low speed tracks are labelled by index iv.

For a better evaluation of the tracking results, our proposed method is also compared to the aforementioned five tracking methods on another challenging test dataset comprising 23 cells at 44 frames. The qualitative evaluation results presented in Figure 5.7 were analysed by TrackScheme, which is a track visualisation tool in the Trackmate software [81]. In this result, our proposed tracking method displayed a better performance compared to the other tracking methods with respect to less ID switching. The perfect tracking method must find 23 tracks, as there are 23 cells in this sample, so identifying more tracks proportionally indicates lost or mismatched tracks. According to this evaluation criterion, our method is ranked first with 27 tracks and LAP is the second top method with 30 tracks. The number of identified tracks (based on the number of track IDs) for Keyhole, Linear Motion, IOU-Tracker and Nearest Neighbour were 32, 38, 46 and 51, respectively.

In Figure 5.7 (a, b), red dots are used to illustrate the merge and split situations on our tracking result as well as LAP (of these five compared methods, only LAP performs merge/split situations). The merge event in Figure 5.7 (a) starts when a new cell appears at time point 16 and is simultaneously merged to one of the other tracked cells. In this case, our proposed method finds the merge/split points accurately by measuring the volume (which is explained in detail in section 3.3), while in LAP, the previous tracking points (before merge) are known as merge and they are shown correctly only at the point of split. Also, the cell disappears at frame 22 (two frames after the time of split), while it is still identified as a merge cell in LAP. For a better understanding, Figure 5.8 illustrates these two merge and split situations visually using TrackScheme tools.

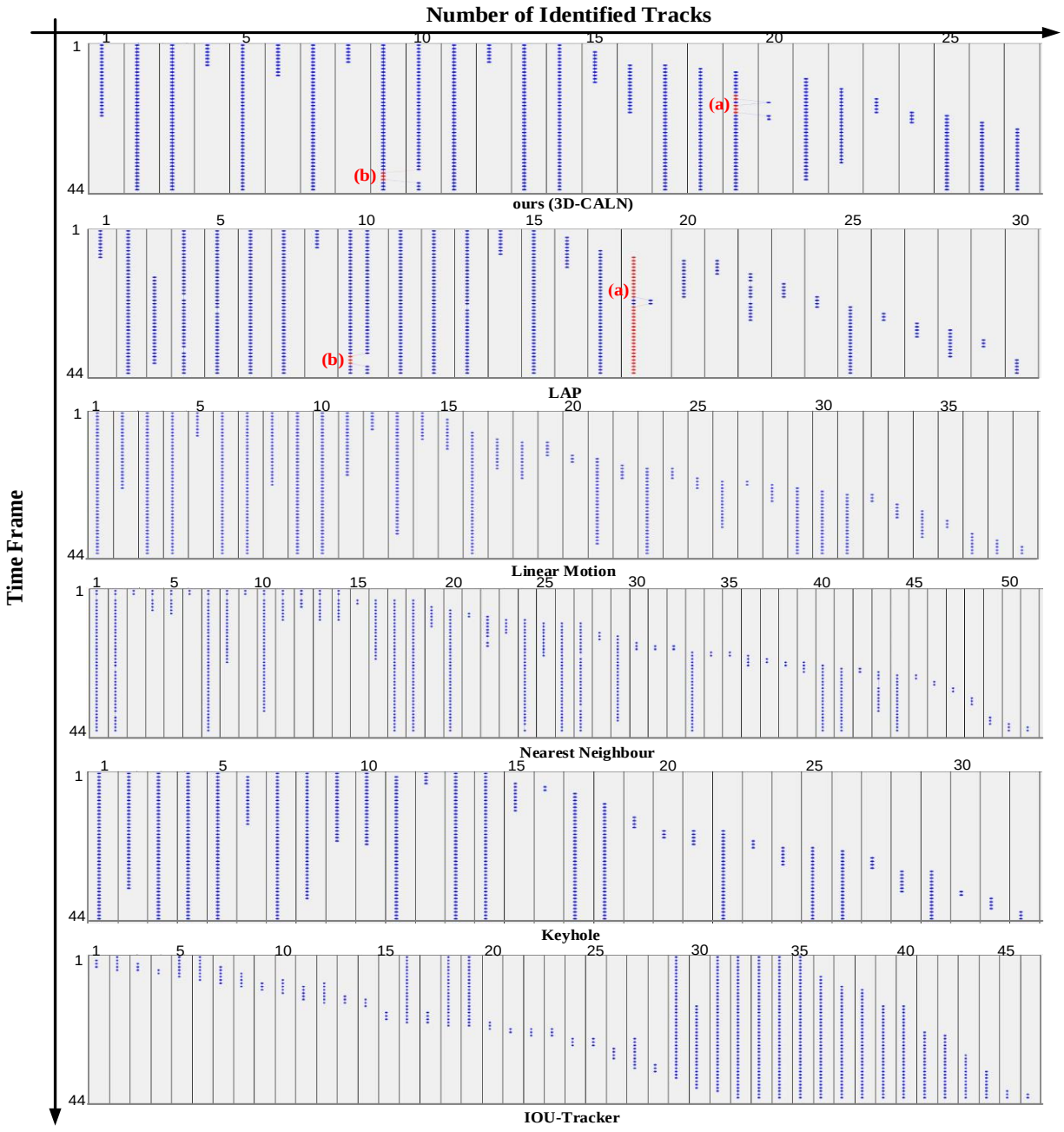


Figure 5.7: Tracking evaluation based on TrackScheme is shown in an example of 3D image sequences comprising 23 fluorescent neutrophils over 44 frames. The number of columns shows how many tracks are identified in each method. The tracking points for each track ID over time are also illustrated by dots in each column. As shown, our method with 27 track IDs achieves the best result compared to the other five tracking methods. The red dots (a) and (b) represent the merge and split statuses.

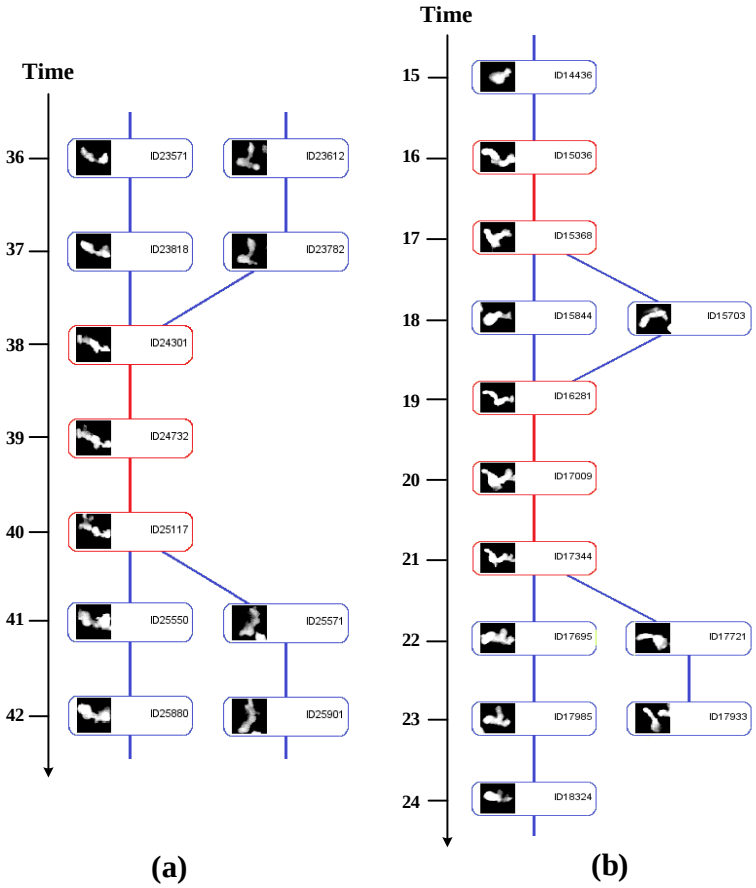


Figure 5.8: Two examples of merge and split situations are shown based on neutrophil images over time. (a) two neutrophils are merged at time 38 and split at 40. (b) a neutrophil appears at time 16 and is merged with a tracked neutrophil, while it is disjointing at time 18 and again merges at time 19. The tracked neutrophil continues as merged till time 21, when the two neutrophils are separated. The newborn cell that appears at time 16 leaves the imaging field at time 23.

5.4.3 Quantitative Validation of the Multi-Cell Tracking on our Dataset

The tracking accuracy can be tested by evaluating the track identity. We evaluated our tracking performance quantitatively based on ID precision (IDP) and ID recall (IDR) [138]. IDP and IDR are the ratios of the computed ground truths (detections) that are correctly identified. Table 5.1 and Table 5.2 summarise the IDP and IDR values of our proposed multi-cell tracking framework, which achieved superior results for both evaluation parameters IDP and IDR with an average of 96.97% and 95.75%, respectively. The next two top tracking results based on IDP and IDR were achieved by Keyhole (80.43%, 78.11%) and LAP (78.81%, 76.39%), respectively. Meanwhile, the merge and split accuracy was validated over 49 merge and 47 split situations, which are shown in Table 5.3. The average of the merge and split accuracy of our framework was 92.63%, which is significantly higher than LAP with 54.25%.

TABLE 5.1: COMPARISON OF THE SIX TRACKING METHODS BASED ON TRACK IDENTIFICATION PRECISION OVER TEN DIFFERENT TEST DATASETS

Data	ours(3D-CALN)	IOU-Tracker	Keyhole	Nearest Neighbor	LAP	Linear Motion
1	100	72.30	69.33	71.03	79.91	74.11
2	100	71.42	87.23	73.21	76.42	74.14
3	97.96	75.65	85.11	66.82	78.27	70.45
4	91.51	67.07	69.21	66.30	73.33	71.33
5	93.57	85.21	86.49	84.21	88.57	86.83
6	100	78.75	81.06	70.62	85.65	81.87
7	92.58	66.54	75.00	51.62	68.41	60.00
8	100	74.40	75.77	67.31	74.49	72.58
9	94.34	61.63	77.80	48.38	63.04	53.04
10	100	96.92	97.33	94.59	100	95.65
Avarage	96.97	74.99	80.43	69.41	78.81	74.00

TABLE 5.2: COMPARISON OF THE SIX TRACKING METHODS BASED ON TRACK IDENTIFICATION RECALL OVER TEN DIFFERENT TEST DATASETS

Data	ours(3D-CALN)	IOU-Tracker	Keyhole	Nearest Neighbor	LAP	Linear Motion
1	100	68.87	63.77	70.23	75.35	72.62
2	100	65.71	84.71	71.28	72.59	71.53
3	95.57	70.44	81.68	57.58	77.37	62.75
4	88.69	64.77	65.77	57.05	71.24	64.24
5	91.49	83.47	84.90	78.27	86.20	84.62
6	100	75.28	80.11	66.08	85.65	81.87
7	89.98	58.01	74.03	45.83	63.01	55.76
8	91.77	70.19	75.77	61.30	72.44	70.68
9	100	59.75	73.80	45.73	60.13	57.04
10	100	81.73	96.52	91.30	100	95.65
Avarage	95.75	69.8	78.11	64.46	76.39	71.68

TABLE 5.3: COMPARING OUR METHOD TO LAP IN TERMS OF MERGE AND SPLIT ACCURACY

	Merge	Split	Average
ours(3D-CALN)	95.91	89.36	92.63
LAP	65.95	42.55	54.25

5.5 Conclusion

In this study, we proposed a deep learning-based linkage method to track multiple cells in 4D florescence image. The proposed multi-cell tracking framework is designed based on two main steps. First, our proposed training model 3D-CALN extracts complex vision and motion features such as morphological characteristics, rapid deformation, complex motion patterns, merge and split. Then, the trained 3D-CALN finds the best matches in the current

frame across the optimized samples to perform multiple tracks. The proposed method was trained and then it tracked neutrophils in the 4D dataset that were already manually-labelled. We then evaluated our proposed tracking method on the real test dataset. The results demonstrate the superior performance of our approach compared to the other tracking methods. Although our tracking method has an outstanding performance in the consistent tracking of immune cells during their complex migratory pattern, it has several limitations similar to the other methods. For instance, when three or more cells are merged, the possibility of the track being lost or mismatched increases. In addition, so far, we have only tested our method on one frequency channel of our 4D dataset. In future work, we will extend our algorithm to simultaneously track cells in different channels. This will enable our algorithm to detect cell-cell interactions between different cell types imaged in different channels (5D imaging data sets (x, y, z, t and frequency channel). This potential can significantly impact the application of this method in analysing more complicated biologic processes such as phagocytosis [222] and intercellular communications [223].

Chapter 6: Applying Deep Learning-Based Cell Tracking Method on Biological Challenges

Cell tracking methods are designed to analyse cellular complex behaviour properties such as deformation, motion and velocity. These methods could also be used to detect and analyse specific biological procedures and problems through some modification. In this chapter, the functionality of our proposed deep learning-based 3D cell tracking model (chapter 5) is demonstrated firstly in the context of biological application by analysing phagocytosis procedure (engulfing large particles) as one of the major mechanisms by which leukocytes neutralise microbes. Secondly, our proposed method is employed in fixing the problem of fluorescence channel overlapping as one of the common fluorescence microscopy imaging problems on cell detection and tracking. The experimental results of this chapter provide proof of principle for effectiveness of our proposed cell tracking methods in solving real-world biological challenges.

NOTE: The content of this chapter has been submitted to *the Gene*.

Moghadam, R. M. and Chen, Y. P. P. "Applying Deep Learning-Based Cell Association Method on Biological Challenges." *Gene.* (Submitted).

6.1 Phagocytosis

The term phagocytosis refers to when a phagocytic immune cell, including neutrophils and macrophages, engulfs a $\geq 0.5\ \mu m$ particles into its cytoplasm [224]. Neutrophils and macrophages are the main effector cells of innate immune system that defend organisms against invader microorganisms through employing different mechanisms including fast migration towards the site of inflammation and phagocytosis. The phagocytosis rate is used as an indicator of the antimicrobial activity of phagocytes as, in the majority of cases, a phagocytosed (i.e. engulfed) microorganism will be neutralised by the intracellular killing mechanisms of these cells (Figure 6.1) [225]. Developing computational tools capable of detecting and analysing phagocytosis would have further applications in investigating other cellular biologic processes such as cell fusion and tissue remodelling[226].

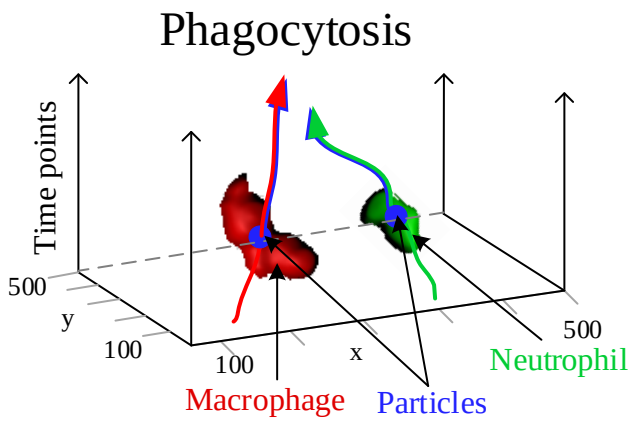


Figure 6.1: The schematic view of phagocytosing is displayed on both neutrophil and macrophage.

Figure 6.2 demonstrates four phagocytosis events selected from real time lapse fluorescence microscopy images taken from an infected zebrafish embryo. Each phagocytosing is highlighted by yellow boxes in 6 time points at frames 2, 23, 27, 47, 55 and 93.

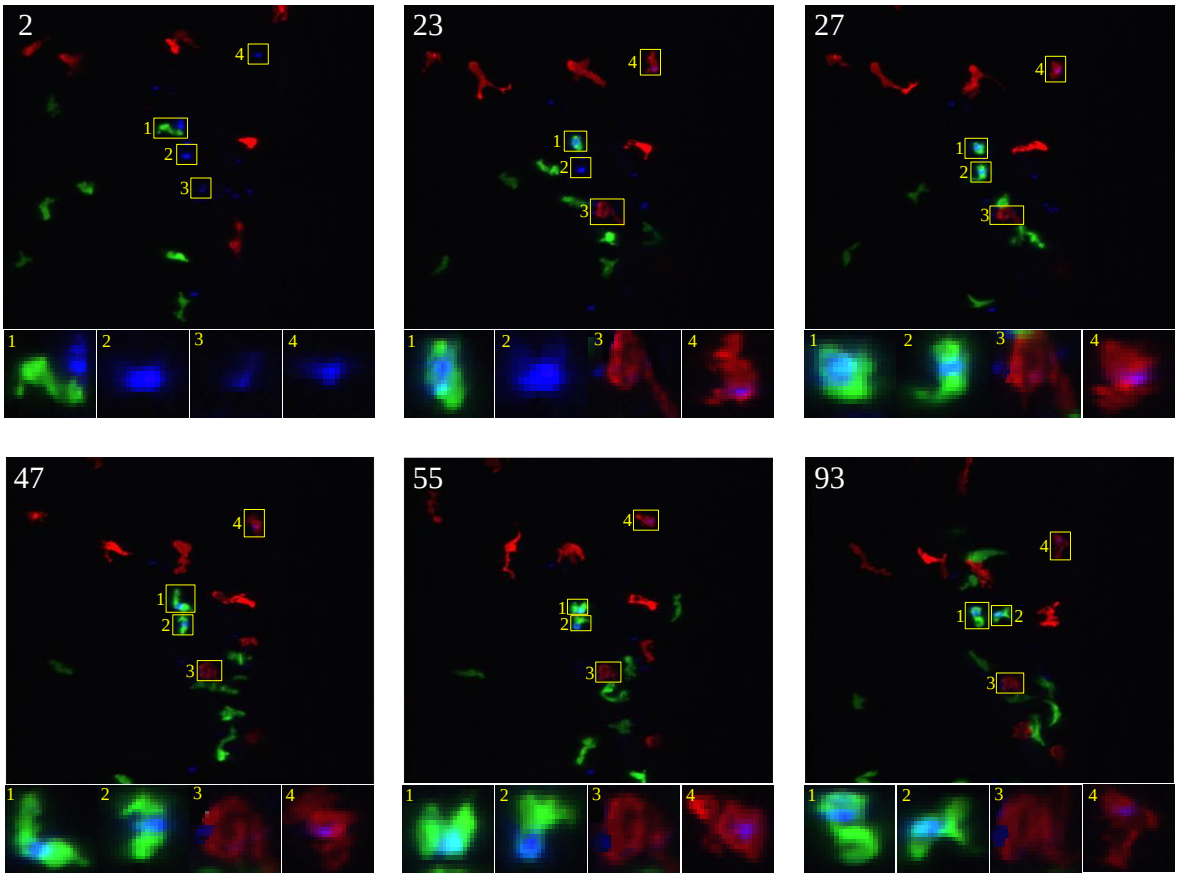


Figure 6.2: Demonstration of four phagocytosis event selected from a series of time lapse fluorescence microscopy images taken from an infected zebrafish embryo. In these images, neutrophils (green channel), macrophages (red channel) and particles (blue channel) are annotated in 4 positions in 6 different time points (frames 2, 23, 27, 47, 55 and 93). The cropped and magnified images of each yellow box have been presented at the bottom of each panel for better visualisation of phagocytosis procedure.

To analyse phagocytosis, our proposed multi-cell tracking model (chapter five) is adopted to track cells and particles from 5D input data (x, y, z, t, c) and report their positions (includes centre location and bounding boxes) during time frames. Then, the phagocytosis events are predicted by calculating the intersection over union on neutrophil-particle or microphage-particle bounding boxes in all frames. The intersection over union (IOU) is defined as follows:

$$IOU = \frac{|A \cup B| - |A \cap B|}{|A \cup B|} \quad (1)$$

where A and B are the bounding boxes of each predicted cell-particle pair at any time point. As the size of particles are significantly smaller than neutrophils and macrophages, a phagocytosis event has happened when the IOU of the neutrophil-particle or macrophage-particle is greater than 0.3. In other words, when the bounding boxes overlap of immune cells and particles is higher than 30%, it means the particle is engulfed by the immune cell.

To differentiate real phagocytosis events from transient overlaps between cells and particles, a post-processing step is introduced to eliminate the overlaps which last for less than two time points (i.e. the immune cells passed by in the vicinity of the particle, without actually phagocytosing it). Figure 6.3 displays the multi-cells tracking results of three fluorescent channels separately in Figure 6.3 (A) and then the phagocytosing results (Figure 6.3 (B)). The phagocytosis events are identified on the tracking results in two ways. First, the immune cells tracking results are illustrated and their overlap with particles are assigned by particle colour (blue in Figure 6.3 (B.i)). Second, the particles tracking results are generated and their interactions points with neutrophils and macrophage are highlighted by their specific colours, (green and red, respectively, in Figure 6.3 (B.ii)). Additionally, as particles do not move when located at extracellular matrix, any motion in particle tracking

points in x, y or z directions could be interpreted as the particle being phagocytosed by an immune cell in this scenario (Figure 6.3 (B.ii)).

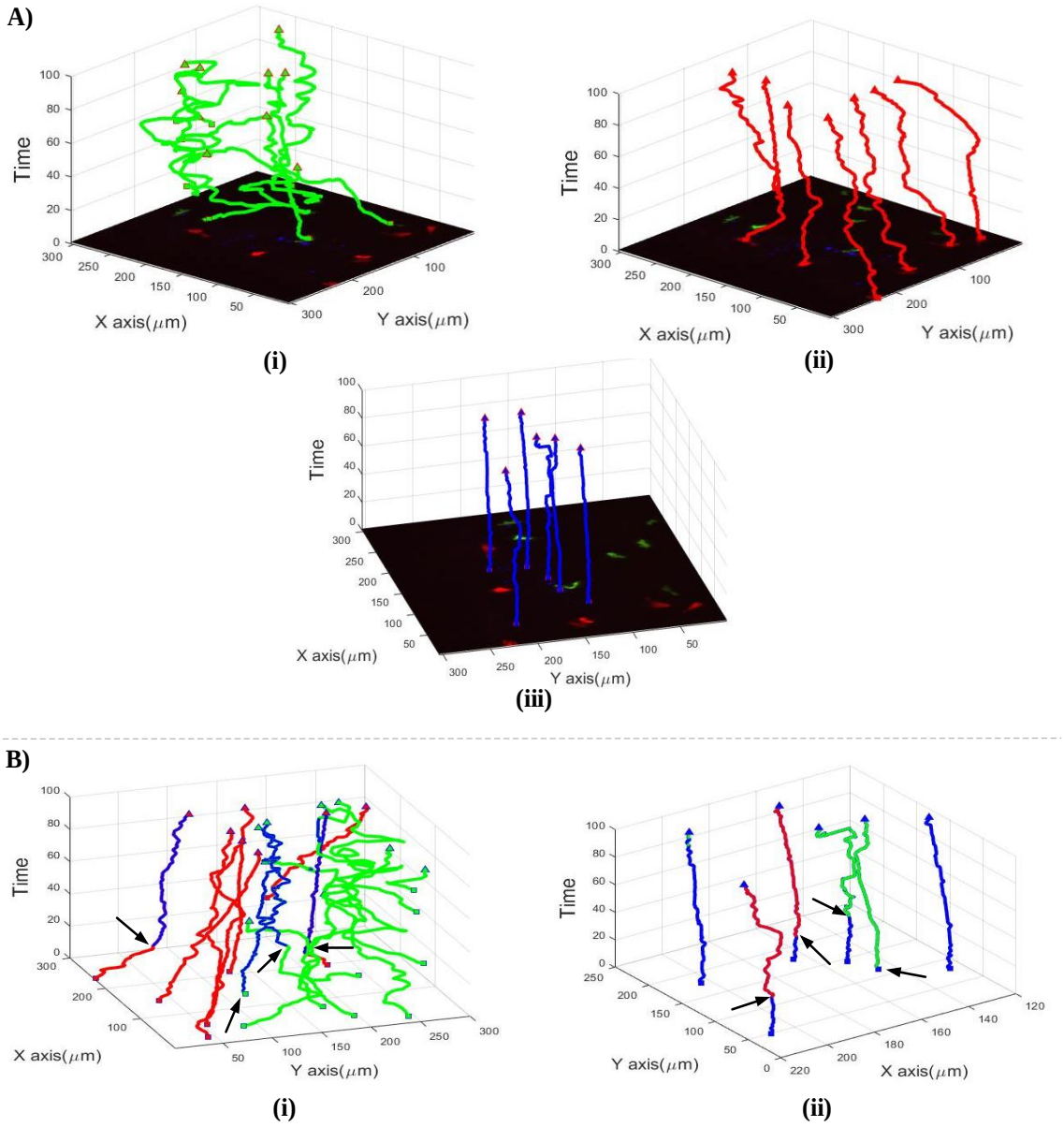


Figure 6.3: The phagocytosis events are spotted on tracking results using our proposed multi-cell tracking method on 5D (x, y, z, t, f) data. (A) shows the tracking results of three fluorescence channels separately including neutrophils (i), macrophage (ii) and particles (iii). (B) shows phagocytosis events based on neutrophils (green) and macrophages (red) tracks (i) and particle tracks (ii). The starting points of phagocytosis are assigned by arrows.

6.2 Fluorescence Channel Overlapping

The ongoing progress in fluorescent labelling methods, including genetic manipulation in transgenic animal models and various antibody staining approaches, along with recent advances in microscopy techniques, provide the opportunity to image complicated biologic structures labelled with several different fluorophores simultaneously. As some popular fluorophores spectrally overlap, precise quantitative analysis of emission light from multiple fluorescence channels is dependent on separation of collected signals from different fluorophores. Insufficient separation of emission signals from different channels makes quantitative analysis, particularly localization assessment, impossible.

At the moment a broad range of fluorescent proteins are available which offer a wide spectral distribution [227] with different excitation and emission peaks (Figure 6.4). By proper configuration of fluorescence filter sets, including excitation or emission filters and the dichroic mirrors, reliable separation of the signals from the most commonly used fluorophore combinations is achievable. However, the best spectral configurations are not always accessible in real world experiments, due to different properties of fluorescent proteins (pH stability, lifetime, etc.). The issue may also arise as result of different brightness of several fluorophores in the sample. All of the above-mentioned reasons could lead to the fluorophore cross talk problem, mainly presented as overlap between the emission spectra of two distinct fluorophores.

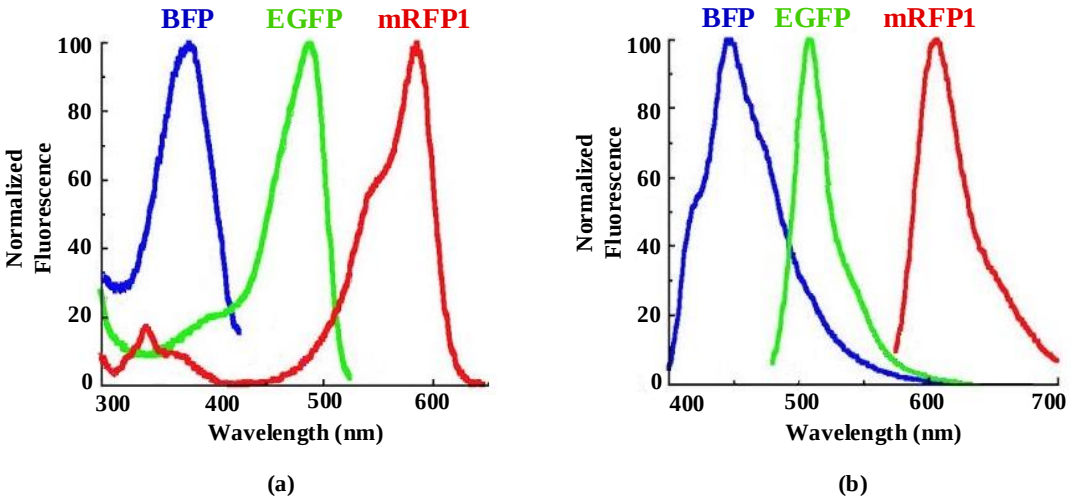


Figure 6.4: The excitation (a) and emission (b) spectra of blue (BFP), green (GFP) and red (mRFP1) fluorescent proteins [228]. BFPs are dim relative to other fluorescent proteins and tend to photobleach readily, with the excitation peak of 384 nm and emission at 448 nm [229]. EGFP is a bright, stable and well-established fluorescent marker, with excitation and emission peaks of 488-489 nm and 507-509 nm, respectively [230]. mRFP1 (excitation: 584 nm, emission: 607 nm) is less bright than other red fluorescent species; and although it matures rapidly but photobleaches more easily [231].

Generally, simultaneous excitation of multiple fluorophores would result in the fluorophore overlap issue. Figure 6.5 illustrates the confocal microscopy images of zebrafish embryos, in which immune cells (neutrophils and macrophages) are labelled with green and red fluorescence protein and microbial particles are labelled in blue. The fluorescence overlapping problem has been demonstrated between neutrophil – macrophage (green-red channels) in Figure 6.5 (A) and neutrophil – particle (green-blue channels) in Figure 6.5 (B). In both cases, the neutrophil is depicted not only in its correct channel (green) but also in other fluorescence channels (red or blue). In other words, the

emission signal from neutrophil fluorophore (EGFP) is also detected in two other channels, mostly due to being brighter than the other two fluorophores.

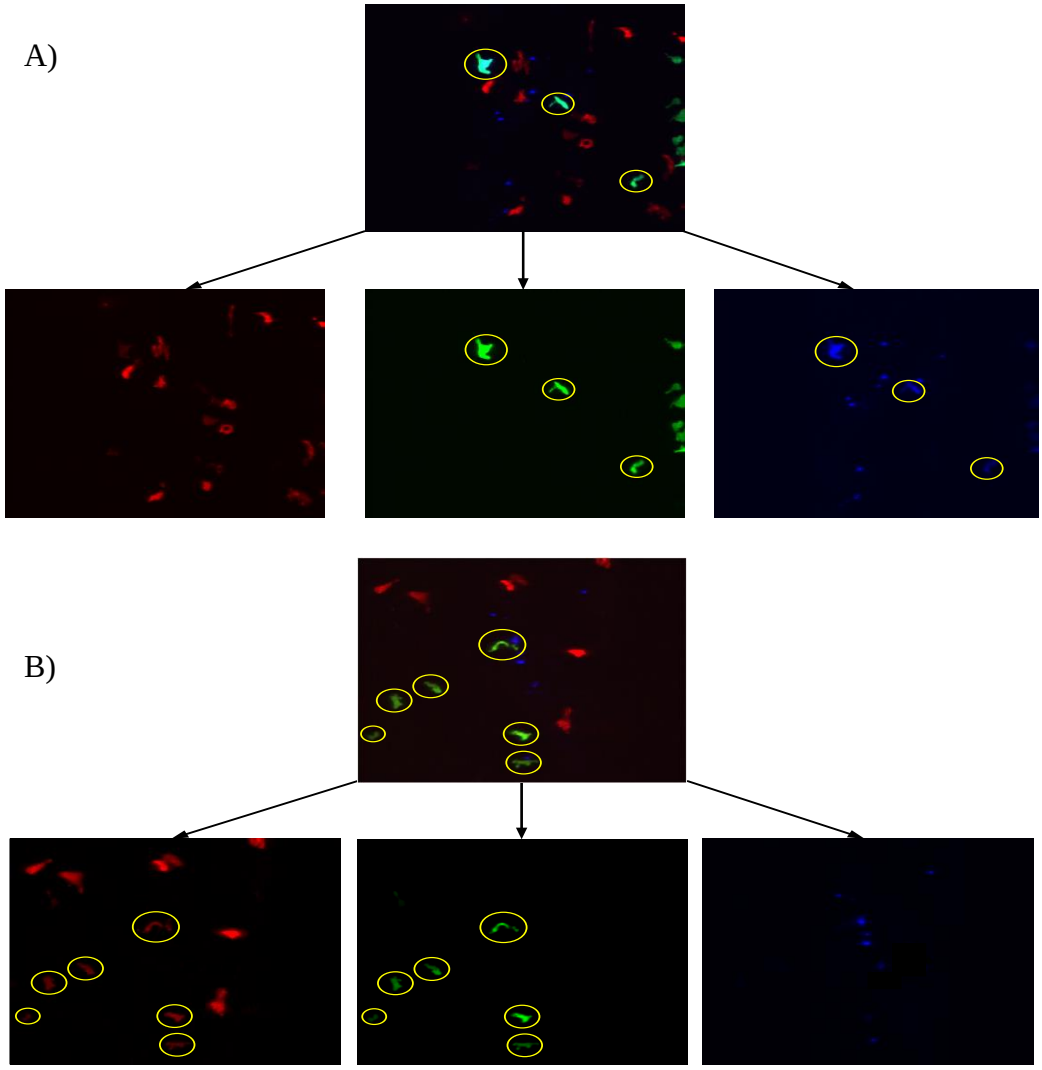


Figure 6.5: Fluorescent image of zebrafish immune cells (EGFP labelled neutrophils and mCherry-labelled macrophages) and microbial particles (labelled by blue fluorophore) taken by confocal microscopy. Images have been demonstrated as compound or individual fluorescent channels. (A) neutrophil – macrophage (green - red) overlapping channels. (B) neutrophil – particle (green - blue) overlapping channels. In both situations, the neutrophil signal (EGFP) has been detected in other channels (red or blue) due to the fluorescence overlapping problem.

The above example clearly demonstrates that separation is not always achievable by applying fluorescence filters in the microscope. Recently, linear unmixing [232] has been widely employed for analysis of spectral datasets and signal separation, aiming to overcome signal overlapping problems. Spectral imaging was originally used for spectral karyotyping [233], and then after being combined with linear unmixing, the technique was applied to immunohistochemistry [234], two-photon microscopy [235] and also confocal microscopy [236]. However, linear unmixing is a data-processing step and different microscopy brands have implemented it in their software in various ways.

Methods other than linear unmixing could also be implemented for spectrally dissociating fluorescence microscopy datasets, including simplistic approaches like unidirectional bleed-through of one fluorescence channel into the other. In this approach, by measuring the contribution of first channel into the second one, and performing subtractive compensation, the mixed signals are separated. This method has been utilised in primary versions of Fluorescence Resonance Energy Transfer (FRET) [237] and also in simultaneous fluorescent protein detection [238]. In other approaches such as Automatic Dye Separation [239], Automatic Component Extraction (ACE) and Blind Source Separation [240], in which spectral data are acquired through multiple detector. Applying this approach for spectral data separation is the dependent on existence of pure areas in dataset, with only a single fluorescent signal as reference spectra. Thus, spectral data separation by this approach in datasets with extremely overlapped fluorescence signals is not possible.

The problem of fluorescence channel mixing has significant impact on the accuracy of cell tracking results due to wrong cell detection in first place. In other words, the biological

objects labelled with a particular fluorescent tag will also appear in other fluorescence channels, where other cell types are supposed to be detected. This will eventually turn into inaccurate tracking results due to incorrect cell type, velocity, motion behaviour, deformation, etc.

In this chapter, we adapt our proposed cell association learning network (3D-CALN) introduced in chapter 5 to address the fluorescent channel mixing problem on cell tracking. To this end, the 3D-CALN predicts the most similar detected objects (due to overlapped fluorescent signal) between two different fluorescent channels. It follows this by post-processing steps to provide a resolved detected object with a particular fluorescence tag as output.

6.2.1 Method

To solve the overlapping fluorescence channels problem on cell tracking, we proposed a framework based on the3D cell association learning network (3D-CALN). First, 3D-CALN is trained on a large dataset composed of similar and non-similar cell pairs (Figure 6.6) from different fluorescence channels. The data is a set of manually cropped and labelled microscopy images of individual cells, which is explained in detail in chapter 3. To enhance the actuary of our training on similarity score, we expand our training dataset by pairing individual cell images with its distortion image, including motion blurred gaussian noise and low resolution. Then the feedforward network, followed by post-processing steps, fix the impact of the overlapped channels problem on cell detection and tracking. This happens through predicting the similarity score between detected individual cells from every two particular fluorescence channels and removing those cells that are

detected incorrectly due to fluorescent channel mixing (labelled by wrong protein tag). The output of the network is a set of detected cells that are allocated to their appropriate fluorescence channel and cell type.

The 3D-CALN learns the robust spatial and temporal features of cells by using two learning functions entails of appearance and regression (Figure 6.7). Due to complex motion profile and amorphous nature of immune cells, a pair of similar and non-similar cells feed to the learning network. To learn the cell appearance, the 3D convolutional neural network (CNN) with three hidden layers, two fully connected layers and one output layer, is applied while the regression function learns the location of cells in the image. The method is described in more detail in chapter 5.

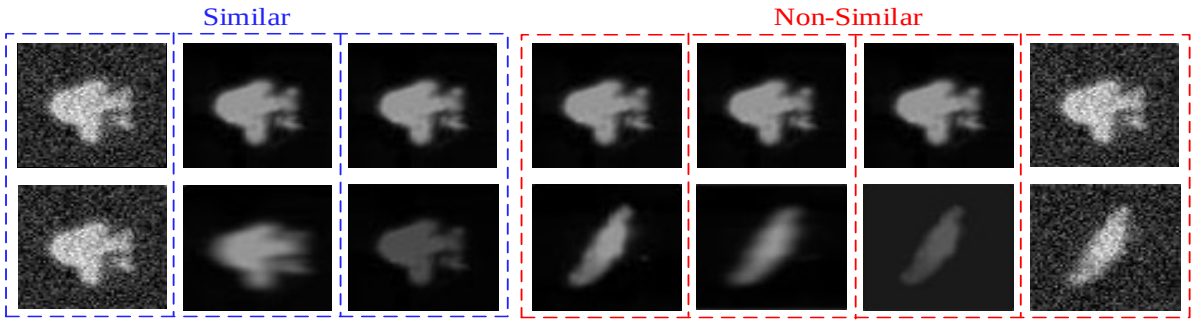


Figure 6.6: Similar and non-similar pairs of individual immune cells.

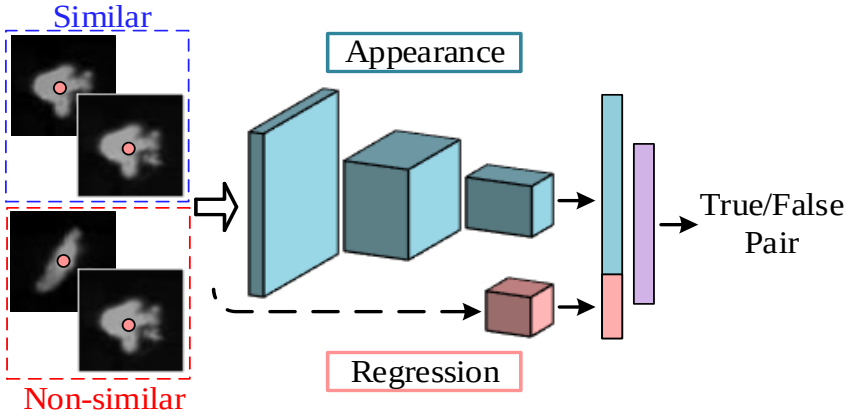


Figure 6.7: Applying 3D Cell Association Learning Network to learn the robust spatial and temporal features of fluorescent cells.

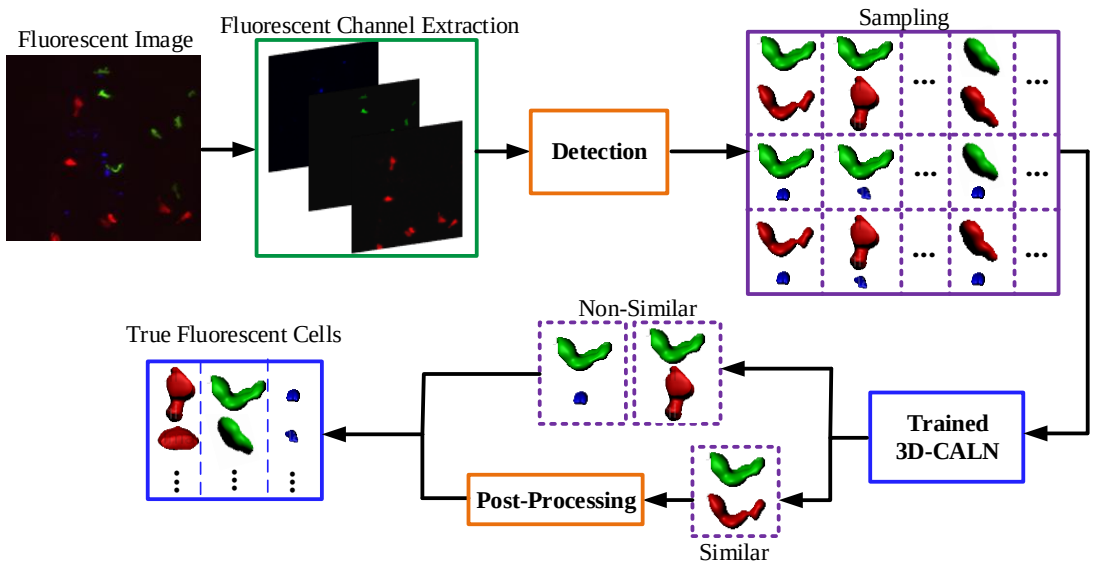


Figure 6.8: Proposed framework for addressing the problem of overlapping fluorescence channels.

Our proposed framework (Figure 6.8) entails five tasks including: channel extraction, detection, sampling, feedforward 3D-CALN and post processing. The procedure for addressing the fluorescent overlapping problem is as follows:

Firstly, raw microscopy data including N fluorescent channels is exported to N individual images $I = \{I_1, \dots, I_N\}$ (each one represents a specific fluorescent channel) using IMARIS and ImageJ software.

The detection algorithm predicts the location of individual cells on each channel by providing a set of target bounding boxes $B_i = D(I_i), i = 1, \dots, N$ where D is a detection model. We utilized the Faster R-CNN [126] structure as the detection model, which is composed of three modules. First, an input image is fed into the convolution layers to extract feature maps. Then, these maps are passed through a Region Proposal Network (RPN) to predict the object proposals. Finally, they are classified (in our work, we have

only one class of cells) and bounding boxes are predicted. Figure 6.9 illustrates the Faster R-CNN structure in our dataset. More detail on the Faster R-CNN can be found in [126].

.

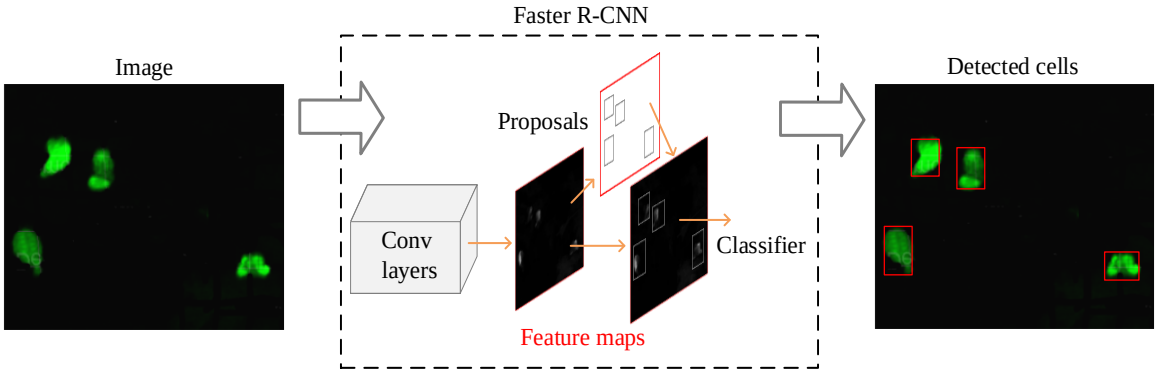


Figure 6.9: Structure of Faster R-CNN for detecting neutrophils displayed by bounding boxes in output image.

The detected targets are cropped and resized to form single cell images $\{x_{1_i}, \dots x_{K_i}\}$ where K_i is the total number of cells in channel i. Then, the sampling function generates a set of all possible paired cells between every two fluorescent channels, denoted as:

$$S = \bigcup_{\substack{i,j=1 \\ i \neq j}}^{N} \left\{ \left(x_{1_i}, x_{1_j}\right), \dots, \left(x_{K_i}, x_{1_j}\right), \left(x_{1_i}, x_{2_j}\right) \dots, \left(x_{K_i}, x_{K'_j}\right) \right\} \quad (2)$$

Thus, each sample $s_{k,k'} = \left(x_{k_i}, x_{k'_j}\right), i \neq j$ is identified by two parameters: cell index $k = 1, \dots, K,\ \ k' = 1, \dots, K'$, and channel index $i, j = 1, \dots, N$. Then the 3D-CALN trained on similar and non-similar classes predicts the most similar cell pairs S' (true class) among all input samples.

Finally, the post-processing function investigates the indexes of true samples, which are predicted by the trained 3D-CALN, and eliminates those indexes (detected cells) that appears incorrectly in other fluorescence channel. Thus, the highest likelihood samples (the most similar ones) are defined by:

$$H =: \{s_{k,k'} \mid \forall k, k', P(s_{k,k'}) > \gamma\} \tag{3}$$

Due to overlapping fluorescent signal, the number of detected cells in one fluorescent channel is greater than another. Therefore, the indexes of cells belonged to greater set and also identified in the highest similar set, are defined as overlapped cells. Then, removing these overlapped cells from a set of detected cells will rectify the cell detection that has resulted in accurate quantitative analysis and cell tracking.

6.2.2. Evaluation and Results

We validated our proposed framework on a real dataset (Figure 6.5 (A)) with a fluorescence channel mixing problem between green and red channels. The sample data shows 6 neutrophils, 12 macrophages and 7 particles in three green, red and blue fluorescence channels, respectively. However, neutrophils are detected in the red channel as macrophages due to overlapping between the fluorescence channels (there are only 6 correct macrophages). Figure 6.10 shows the samples similarity scores in terms of sample number, where samples with the highest likelihood are highlighted in red, representing samples with the fluorescent overlapping problem. Figure 6.11 displays the macrophage tracking results before applying our proposed solution to the overlapping fluorescence channels issue and after rectifying.

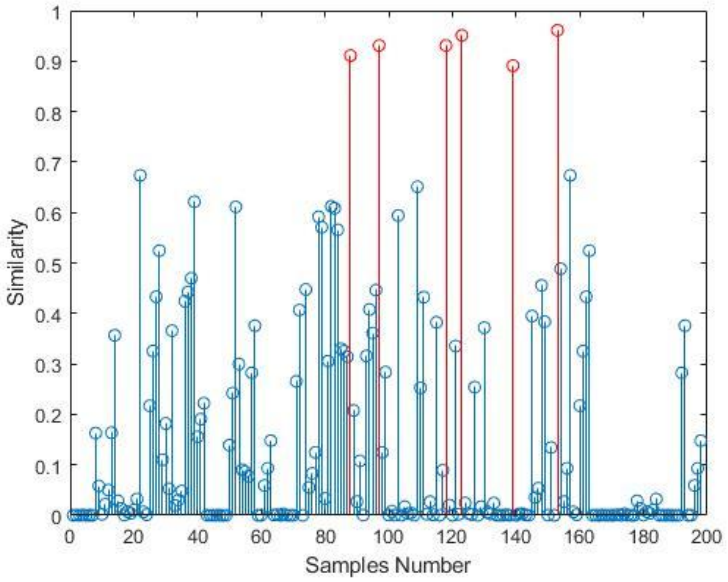
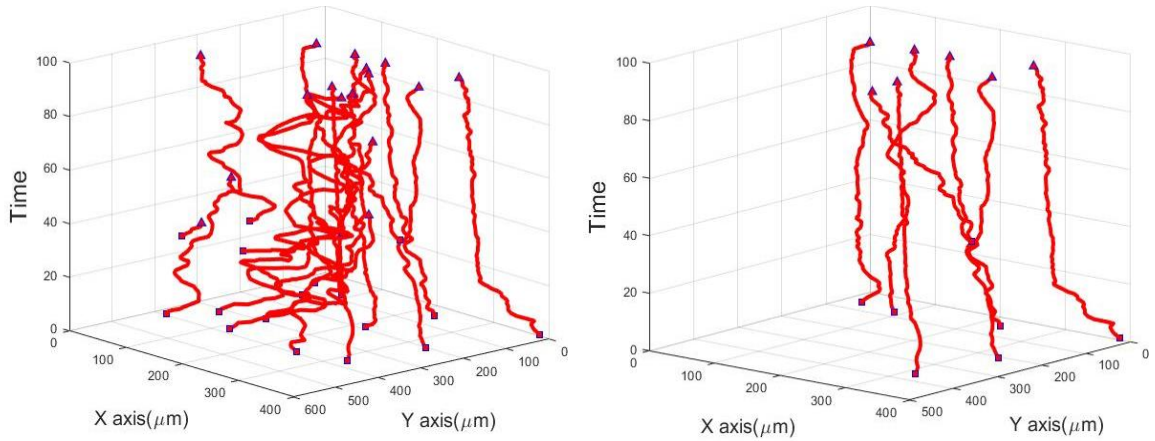


Figure 6.10: Similarity score in terms of sample number.



(a) (b)

Figure 6.11: Tracking macrophages with overlapping fluorescence channels (a) and after eliminating incorrect appeared fluorescent objects (b).

6.3 Conclusion

This chapter provides functional examples for application of our proposed cell tracking methods in solving real world biomedical imaging challenges. Firstly, our proposed 3D-CALN was applied to detect and quantify phagocytosis as a major immune cell mechanism against microbial particles, which is commonly employed as a measure for competency of innate immune responses in the field of cellular immunology [241].

Automated detection of phagocytosis, along with leukocyte cell tracking, provides valuable information about not only the migratory profile of these cells, but also their microbicidal activity through engulfing invading foreign particles.

Secondly, our proposed 3D-CALN has been adapted to address the problem of overlapping fluorescence channels on time lapse microscopy images, which negatively affects cell tracking accuracy. This approach has been designed in two parts: training 3D-CALN on similar and non-similar pairs of biological objects; and fluorescent channel correction on cell detection and tracking by feed forwarding the trained network followed by post-processing steps. The experimental results presented in this chapter demonstrate the clear separation of fluorescence channels in cell tracking results even in cases of highly overlapping fluorescence signals.

Chapter 7: Conclusion and Future Work

7.1 Conclusion

This thesis investigated the effectiveness of deep learning approaches for developing more accurate cell tracking, particularly when tracking highly mobile immune cells. Improving the accuracy of the cell tracking method has a significant impact on refining computer-aided diagnosis systems that assist biologists through automated analysis of large-scale imaging datasets revealing the inflammation process and immune system responses to invading microorganisms.

More recently, by taking advantage of deep learning as a powerful research area in the field of computer vision, advanced automated object tracking models have been introduced. The convolutional neural network (CNN) has become one of the most popular deep learning techniques in object detection and tracking, that works by learning the robust features of an objects. Although the CNN model has been widely employed for detection and tracking of common objects (humans, cars, etc.), its potential in identifying and tracking highly amorphous and migratory biologic particles within time lapse microscopy images has not been rigorously explored yet.

To do so, there are number of significant challenges which need to be addressed, including how to utilize power of deep learning to learn both spatial and temporal features of highly migratory and amorphous cells, which ultimately leads to improve the accuracy of the single cell detection and tracking. The other challenge is how to develop a deep

structure to associate 3D microscopy images of multiple cells within consecutive frames to improve tracking consistency. Finally, it is important to demonstrate how efficiently this method could be employed to biological applications and fix different real-world biomedical imaging problems. However, developing novel automated cell tracking methods capable of handling these challenges is highly restricted by shortage of comprehensive ground truth datasets for training and validation. The current thesis has been structured to address these points through applying the proposed methods on a real generated dataset and in close collaboration with local biologists specialized in the field of leukocyte biology.

The project started by constructing a data repository containing a substantial collection—about 25,000—of labelled immune cells extracted from multi-dimensional microscopy image sequences. The significant feature of our dataset is the compilation of large enough number of labelled cells in different image files covering a broad range of detection and tracking challenges to be used as a training and validation dataset for deep learning algorithms.

Then, the thesis focuses on single cell detection and tracking over 2D time lapse microscopy images. We proposed a multi-channel feature learning model inspired by deep learning to be trained by learning comprehensive spatial and temporal features of a single leukocyte (immune cell) within consecutive frames. Moreover, a cell tracking framework is introduced to first produces a set of candidates bounding boxes by the sampling function based on the Particle Filters in the following frames, then tracks cells by predicting the final position of the target cell in the following frame through the trained multi-channel feature learning network. Extensive comparison analysis indicated the robustness of our

method by running it parallel to other nine available algorithms on ten randomly selected image sequences of our dataset and also on two other different datasets, namely TB-100 [176] and an *in vitro* cell tracking dataset [190]. Our method proved to improve the precision and success accuracy by about 10% and 30% compared to the other nine methods.

We have also presented a novel multi-cell tracking framework based on 3D cell association learning network (a new deep structure of 3D CNN) to enhance tracking consistency with fewer broken tracks while tracking multiple highly migratory calls. The multi-cell tracking framework calculates multiple tracks through associating detected cells in adjacent frames. The capability to track cells over the 3D image sequence performed the advantage of our proposed multi-cell tracking framework on tracking cells on their complex motion in 3 dimensions. The proposed multi cell tracking framework improved the accuracy of tracking to as high as 95% (in terms of tracking consistency) compared to the other five available cell association algorithms.

Finally, as functional validation, we tested the capability and potential of our proposed cell tracking model for analysing a more specific biological process (phagocytosis). To examine phagocytosis (engulfing large particles by immune cells), as one of the major mechanisms by which leukocytes neutralise microbes, our proposed 3D multi-cell tracking framework was adapted to track the immune cells and particles simultaneously, and then their interactions were identified through calculating intersection over union.

Moreover, our model was applied to fix the problem of "fluorescent channel overlapping" as a common fluorescence microscopy imaging issue with cell detection and tracking. This was in two parts: training 3D-CALN on similar and non-similar pairs of

biological objects, and fluorescent channel correction on cell detection and tracking by feed forwarding the trained network followed by post-processing steps. The experimental results demonstrated the clear separation of fluorescence channels in cell tracking results even in cases of highly overlapping fluorescence signals.

7.2 Future Work

In this research project, a deep learning model was proposed to tackle the common challenges of tracking immune cells as highly shapeless and fast-moving cells. However, it is obvious that, by developing more advanced microscopy techniques (Ultramicroscope, Super-resolution microscopy, Light Sheet Fluorescence Microscopy, etc.), more sophisticated biologic processes will be visualised and the demand for computational analytical tools capable of analysing these complex patterns will be inevitable.

In future research, it would be interesting to study the potential of this model for analysing more sophisticated biologic patterns at the cellular and sub-cellular levels, including signalling and synapsing patterns between immune cells and other cell types, cell differentiation, tissue remodelling and regeneration.

References

[1] T. Nguyen, B. Payan, A. Zambrano, Y. Du, M. Bondesson, and C. Mohan, "Epigallocatechin-3-gallate suppresses neutrophil migration speed in a transgenic zebrafish model accompanied by reduced inflammatory mediators," *Journal of inflammation research*, vol. 12, p. 231, 2019.

[2] S. A. Eming, T. Krieg, and J. M. Davidson, "Inflammation in wound repair: Molecular and cellular mechanisms," *Journal of Investigative Dermatology*, vol. 127, no. 3, pp. 514-525, Mar 2007.

[3] K. M. Henry, C. A. Loynes, M. K. Whyte, and S. A. Renshaw, "Zebrafish as a model for the study of neutrophil biology," *Journal of Leukocyte Biology*, vol. 94, no. 4, pp. 633-42, Oct 2013.

[4] Q. Deng and A. Huttenlocher, "Leukocyte migration from a fish eye's view," *Journal of Cell Science*, vol. 125, no. Pt 17, pp. 3949-56, Sep 1 2012.

[5] G. J. Kremers, S. G. Gilbert, P. J. Cranfill, M. W. Davidson, and D. W. Piston, "Fluorescent proteins at a glance," *Journal of Cell Science*, vol. 124, no. 15, pp. 2676-2676, Aug 1 2011.

[6] V. Pinho, F. M. Coelho, G. B. Menezes, and D. C. Cara, "Intravital microscopy to study leukocyte recruitment in vivo," in *Light Microscopy*, pp. 81-90, 2011.

[7] J. R. Mathias, B. J. Perrin, T. X. Liu, J. Kanki, A. T. Look, and A. Huttenlocher, "Resolution of inflammation by retrograde chemotaxis of neutrophils in transgenic zebrafish," *Journal of Leukocyte Biology*, vol. 80, no. 6, pp. 1281-1288, Dec 2006.

[8] S. A. Renshaw, C. A. Loynes, D. M. I. Trushell, S. Elworthy, P. W. Ingham, and M. K. B. Whyte, "A transgenic zebrafish model of neutrophilic inflammation," *Blood*, vol. 108, no. 13, pp. 3976-3978, Dec 15 2006.

[9] P. Dendorfer, H. Rezatofighi, A. Milan, J. Shi, D. Cremers, I. Reid *et al*., "CVPR19 Tracking and Detection Challenge: How crowded can it get?," *arXiv preprint arXiv:1906.04567,* 2019.

[10] W. Luo, J. Xing, A. Milan, X. Zhang, W. Liu, X. Zhao, T.K. Kim, "Multiple object tracking: A literature review". *arXiv preprint arXiv:1409.7618,* 2014.

[11] F. Yang, M. A. Mackey, F. Ianzini, G. Gallardo, and M. Sonka, "Cell segmentation, tracking, and mitosis detection using temporal context," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 302-309, 2005.

[12] T. He, H. Mao, J. Guo, and Z. Yi, "Cell tracking using deep neural networks with multi-task learning," *Image and Vision Computing,* vol. 60, pp. 142-153, 2017.

[13] S. Lahooti, H. Yueh, and A. Neumann, "An image analysis strategy to study cell adhesion," *Colloids and Surfaces B: Biointerfaces*, vol. 3, no. 6, pp. 333-342, 1995.

[14] A. MacGowan, C. Rogers, and K. Bowker, "In vitro models, in vivo models, and pharmacokinetics: What can we learn from in vitro models?," *Clinical Infectious Diseases,* vol. 33, pp. S214-S220, Sep 15 2001.

[15] G. Streisinger, C. Walker, N. Dower, D. Knauber, and F. Singer, "Production of clones of homozygous diploid zebra fish (Brachydanio rerio)," *Nature*, vol. 291, no. 5813, pp.293-296, 1981.

[16] Y. Gibert, M. C. Trengove, and A. C. Ward, "Zebrafish as a genetic model in pre-clinical drug testing and screening," *Curr Med Chem,* vol. 20, no. 19, pp. 2458-66, 2013.

[17] C. Sullivan and C. H. Kim, "Zebrafish as a model for infectious disease and immune function," *Fish Shellfish Immunol,* Review vol. 25, no. 4, pp. 341-50, Oct 2008.

[18] E. E. Patton and L. I. Zon, "The art and design of genetic screens: zebrafish," *Nat Rev Genet,* vol. 2, no. 12, pp. 956-66, Dec 2001.

[19] D. M. Tobin, R. C. May, and R. T. Wheeler, "Zebrafish: a see-through host and a fluorescent toolbox to probe host-pathogen interaction," *PLoS Pathog,* Review vol. 8, no. 1, p. e1002349, Jan 2012.

[20] G. J. Lieschke and P. D. Currie, "Animal models of human disease: zebrafish swim into view," *Nat Rev Genet,* vol. 8, no. 5, pp. 353-67, May 2007.

[21] S. A. Renshaw, C. A. Loynes, D. M. Trushell, S. Elworthy, P. W. Ingham, and M. K. Whyte, "A transgenic zebrafish model of neutrophilic inflammation," *Blood,* vol. 108, no. 13, pp. 3976-8, Dec 15 2006.

[22] J. R. Mathias, B. J. Perrin, T. X. Liu, J. Kanki, A. T. Look, and A. Huttenlocher, "Resolution of inflammation by retrograde chemotaxis of neutrophils in transgenic zebrafish," *J Leukoc Biol,* vol. 80, no. 6, pp. 1281-8, Dec 2006.

[23] K. Thorn, "A quick guide to light microscopy in cell biology," *Mol Biol Cell,* vol. 27, no. 2, pp. 219-22, Jan 15 2016.

[24] G. Wassilew, H. Guski, E. F. Luschnikow, and W. M. Sakrebin, "Fundamentals of Morphometry and Its Use in Light and Electron-Microscopy," *Arkhiv Patologii,* vol. 39, no. 9, pp. 80-87, 1977.

[25] D. L. Coutu and T. Schroeder, "Probing cellular processes by long-term live imaging–historic problems and current solutions," *J Cell Sci,* vol. 126, no. 17, pp. 3805-3815, 2013.

[26] F. Ellett, L. Pase, J. W. Hayman, A. Andrianopoulos, and G. J. Lieschke, "mpeg1 promoter transgenes direct macrophage-lineage expression in zebrafish," *Blood,* vol. 117, no. 4, pp. e49-56, Jan 27 2011.

[27] G. Zhang, V. Gurtu, and S. R. Kain, "An enhanced green fluorescent protein allows sensitive detection of gene transfer in mammalian cells," *Biochem Biophys Res Commun*, vol. 227, no. 3, pp. 707-11, Oct 23 1996.

[28] Y. Wu, J. Lim, and M.-H. Yang, "Online object tracking: A benchmark," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2013, pp. 2411-2418.

[29] L. Leal-Taixé, A. Milan, I. Reid, S. Roth, and K. Schindler, "Motchallenge 2015: Towards a benchmark for multi-target tracking," *arXiv preprint arXiv:1504.01942*, 2015.

[30] A. W. Smeulders, D. M. Chu, R. Cucchiara, S. Calderara, A. Dehghan, and M. Shah, "Visual tracking: An experimental survey". *IEEE transactions on pattern analysis and machine intelligence*, vol. 36, no. 7, pp.1442-1468, 2013.

[31] W. Luo, J. Xing, A. Milan, X. Zhang, W. Liu, X. Zhao, T.K. Kim, "Multiple object tracking: A literature review". *arXiv preprint arXiv:1409.7618*, 2014.

[32] X. Trepat, Z. Z. Chen, and K. Jacobson, "Cell Migration," *Comprehensive Physiology*, vol. 2, no. 4, pp. 2369-2392, Oct 2012.

[33] E. Meijering, O. Dzyubachyk, I. Smal, and W. A. van Cappellen, "Tracking in cell and developmental biology," *Seminars in Cell & Developmental Biology*, vol. 20, no. 8, pp. 894-902, Oct 2009.

[34] Z. Y. Takeo Kanade, Ryoma Bise, "Cell image analysis: Algorithms, system and applications," *IEEE Workshop on Applications of Computer Vision (WACV '11)*, Kona, Hawaii, USA, pp. 374–381, January 2011.

[35] C. Zimmer, E. Labruyere, V. Meas-Yedid, N. Guillén, and J.-C. Olivo-Marin, "Segmentation and tracking of migrating cells in videomicroscopy with parametric

active contours: A tool for cell-based drug testing," *IEEE transactions on medical imaging,* vol. 21, no. 10, pp. 1212-1221, 2002.

[36] K. Li, E. D. Miller, M. Chen, T. Kanade, L. E. Weiss, and P. G. Campbell, "Cell population tracking and lineage construction with spatiotemporal context," *Medical image analysis,* vol. 12, no. 5, pp. 546-566, 2008.

[37] O. Dzyubachyk, W. A. Van Cappellen, J. Essers, W. J. Niessen, and E. Meijering, "Advanced level-set-based cell tracking in time-lapse fluorescence microscopy," *IEEE transactions on medical imaging,* vol. 29, no. 3, pp. 852-867, 2010.

[38] F. Leymarie and M. D. Levine, "Tracking deformable objects in the plane using an active contour model," *IEEE Transactions on Pattern Analysis & Machine Intelligence,* no. 6, pp. 617-634, 1993.

[39] M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: Active contour models," *International journal of computer vision,* vol. 1, no. 4, pp. 321-331, 1988.

[40] D. P. Mukherjee, N. Ray, and S. T. Acton, "Level set analysis for leukocyte detection and tracking," *IEEE Transactions on Image processing,* vol. 13, no. 4, pp. 562-572, 2004.

[41] J. A. Sethian, "Level set methods and fast marching methods," *Journal of Computing and Information Technology,* vol. 11, no. 1, pp. 1-2, 2003.

[42] D. Terzopoulos and D. Metaxas, "Dynamic 3D models with local and global deformations: Deformable superquadrics," in *Proceedings Third International Conference on Computer Vision*, pp. 606-615, 1990.

[43] N. Ray, S. T. Acton, and K. Ley, "Tracking leukocytes in vivo with shape and size constrained active contours," *IEEE transactions on medical imaging,* vol. 21, no. 10, pp. 1222-1235, 2002.

[44] C. Zimmer and J.-C. Olivo-Marin, "Coupled parametric active contours," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 27, no. 11, pp. 1838-1842, 2005.

[45] A. Dufour, R. Thibeaux, E. Labruyere, N. Guillen, and J.-C. Olivo-Marin, "3-D active meshes: fast discrete deformable models for cell tracking in 3-D time-lapse microscopy," *IEEE Transactions on Image Processing,* vol. 20, no. 7, pp. 1925-1937, 2010.

[46] A. Dufour, V. Shinin, S. Tajbakhsh, N. Guillén-Aghion, J.-C. Olivo-Marin, and C. Zimmer, "Segmenting and tracking fluorescent cells in dynamic 3-D microscopy with coupled active surfaces," *IEEE Transactions on Image Processing,* vol. 14, no. 9, pp. 1396-1410, 2005.

[47] R. Malladi, J. A. Sethian, and B. C. Vemuri, "Shape modeling with front propagation: A level set approach," *IEEE transactions on pattern analysis and machine intelligence,* vol. 17, no. 2, pp. 158-175, 1995.

[48] N. Paragios and R. Deriche, "Coupled geodesic active regions for image segmentation: A level set approach," in *European Conference on Computer Vision*, pp. 224-240, 2000.

[49] S. Osher and J. A. Sethian, "Fronts propagating with curvature-dependent speed: algorithms based on Hamilton-Jacobi formulations," *Journal of computational physics,* vol. 79, no. 1, pp. 12-49, 1988.

[50] D. Padfield, J. Rittscher, and B. Roysam, "Spatio-temporal cell segmentation and tracking for automated screening," in *5th IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, pp. 376-379, 2008.

[51] Q. Wu, F. Merchant, and K. Castleman, *Microscope image processing*. Academic Press, 2010.

[52] E. Meijering, I. Smal, O. Dzyubachyk, and J.-C. Olivo-Marin, "Time-lapse imaging," *Microscope Image Processing*, vol. 401, p. 440, 2008.

[53] N. N. Kachouie, P. Fieguth, J. Ramunas, and E. Jervis, "Probabilistic model-based cell tracking," *International Journal of Biomedical Imaging*, vol. 2006, 2006.

[54] C. Wählby, I. M. Sintorn, F. Erlandsson, G. Borgefors, and E. Bengtsson, "Combining intensity, edge and shape information for 2D and 3D segmentation of cell nuclei in tissue sections," *Journal of microscopy*, vol. 215, no. 1, pp. 67-76, 2004.

[55] X. Yang, H. Li, and X. Zhou, "Nuclei segmentation using marker-controlled watershed, tracking using mean-shift, and Kalman filter in time-lapse microscopy," *IEEE Transactions on Circuits and Systems*, vol. 53, no. 11, pp. 2405-2414, 2006.

[56] N. Chenouard, I. Smal, F. De Chaumont, M. Maška, I. F. Sbalzarini, Y. Gong et al., "Objective comparison of particle tracking methods," *Nature methods*, vol. 11, no. 3, p. 281, 2014.

[57] J. Dow, J. Lackie, and K. V. Crocket, "A simple microcomputer-based system for real-time analysis of cell behaviour," *Journal of cell science*, vol. 87, no. 1, pp. 171-182, 1987.

[58] Y. Bar-Shalom, T. E. Fortmann, and P. G. Cable, "Tracking and data association," *Journal of the Acoustical Society of America*, vol. 87, no. 918, 1990.

[59] A. Genovesio, T. Liedl, V. Emiliani, W. J. Parak, M. Coppey-Moisan, and J.-C. Olivo-Marin, "Multiple particle tracking in 3-D+ t microscopy: method and application to the tracking of endocytosed quantum dots," *IEEE Transactions on Image Processing*, vol. 15, no. 5, pp. 1062-1070, 2006.

[60] I. Smal, W. Niessen, and E. Meijering, "Bayesian tracking for fluorescence microscopic imaging," in *3rd IEEE International Symposium on Biomedical Imaging*, vol. 2006, pp. 550-553, 2006.

[61] W. J. Godinez, M. Lampe, S. Worz, B. Muller, R. Eils, and K. Rohr, "Tracking of virus particles in time-lapse fluorescence microscopy image sequences," in *4th IEEE International Symposium on Biomedical Imaging*, pp. 256-259, 2007.

[62] S. N. Ngoc, F. Briquet-Laugier, C. Boulin, and J.-C. Olivo, "Adaptive detection for tracking moving biological objects in video microscopy sequences," in *Proceedings of International Conference on Image Processing*, vol. 3, pp. 484-487,1997.

[63] J. Black, T. Ellis, and P. Rosin, "Multi view image surveillance and tracking," in *Workshop on Motion and Video Computing, Proceedings*, pp. 169-174, 2002.

[64] S. Iwase and H. Saito, "Parallel tracking of all soccer players by integrating detected positions in multiple view images," in *Proceedings of the 17th International Conference on Pattern Recognition, ICPR 2004*, 2004, vol. 4, pp. 751-754, 2004.

[65] M. Xu, J. Orwell, and G. Jones, "Tracking football players with multiple cameras," in *International Conference on Image Processing, ICIP'04*, vol. 5, pp. 2909-2912, 2004.

[66] J. Vermaak, A. Doucet, and P. Patrick, "Maintaining multi-modality through mixture tracking," in *Proceedings Ninth IEEE International Conference on Computer Vision*, p. 1110, 2003.

[67] J. Giebel, D. M. Gavrila, and C. Schnörr, "A bayesian framework for multi-cue 3d object tracking," in *European Conference on Computer Vision*, pp. 241-252, 2004.

[68] K. Smith, D. Gatica-Perez, and J.-M. Odobez, "Using particles to track varying numbers of interacting people," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, vol. 1, pp. 962-969, 2005.

[69] K. Okuma, A. Taleghani, N. De Freitas, J. J. Little, and D. G. Lowe, "A boosted particle filter: Multitarget detection and tracking," in *European conference on computer vision*, pp. 28-39, 2004.

[70] W. Du and J. Piater, "Multi-camera people tracking by collaborative particle filters and principal axis-based integration," in *Asian Conference on Computer Vision*, pp. 365-374, 2007.

[71] B. Xu, M. Lu, J. Cong, and B. Nener, "An Ant Colony Inspired Multi-Bernoulli Filter for Cell Tracking in Time-Lapse Microscopy Sequences", *IEEE journal of biomedical and health informatics*," 2019.

[72] H. Shen, G. Nelson, S. Kennedy, D. Nelson, J. Johnson, D. Spiller *et al*.,"Automatic tracking of biological cells and compartments using particle filters and active contours," *Chemometrics and Intelligent Laboratory Systems*, vol. 82, no. 1-2, pp. 276-282, 2006.

[73] Y. Rathi, N. Vaswani, A. Tannenbaum, and A. Yezzi, "Tracking deforming objects using particle filtering for geometric active contours," *IEEE transactions on pattern analysis and machine intelligence*, vol. 29, no. 8, p. 1470, 2007.

[74] M. D. Breitenstein, F. Reichlin, B. Leibe, E. Koller-Meier, and L. Van Gool, "Robust tracking-by-detection using a detector confidence particle filter," in *IEEE 12th International Conference on Computer Vision*, pp. 1515-1522, 2009.

[75] E. Meijering, O. Dzyubachyk, I. Smal, and W. A. van Cappellen, "Tracking in cell and developmental biology," in *Seminars in cell & developmental biology*, vol. 20, no. 8, pp. 894-902, 2009.

[76] H. Jiang, S. Fels, and J. J. Little, "A linear programming approach for multiple object tracking," in *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1-8, 2007.

[77] J. Berclaz, F. Fleuret, and P. Fua, "Multiple object tracking using flow linear programming," in *Twelfth IEEE international workshop on performance evaluation of tracking and surveillance*, pp. 1-8, 2009.

[78] K. Jaqaman, D. Loerke, M. Mettlen, H. Kuwata, S. Grinstein, S. L. Schmid *et al.*, "Robust single-particle tracking in live-cell time-lapse sequences," *Nature methods*, vol. 5, no. 8, p. 695, 2008.

[79] A. E. Carpenter, T. R. Jones, M. R. Lamprecht, C. Clarke, I. H. Kang, O. Friman *et al.*, "CellProfiler: image analysis software for identifying and quantifying cell phenotypes," *Genome biology*, vol. 7, no. 10, p. R100, 2006.

[80] C. McQuin, A. Goodman, V. Chernyshev, L. Kamentsky, B. A. Cimini, K. W. Karhohs *et al.*, "CellProfiler 3.0: Next-generation image processing for biology," *PLoS biology*, vol. 16, no. 7, p. e2005970, 2018.

[81] J. Y. Tinevez, N. Perry, J. Schindelin, G. M. Hoopes, G. D. Reynolds, E. Laplantine *et al.*, "TrackMate: An open and extensible platform for single-particle tracking", *Methods*, vol. 115, pp. 80-90, 2017.

[82] Y. LeCun, Y. Bengio, and G. J. n. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436-444, 2015.

[83] Z.-Q. Zhao, P. Zheng, S. T. Xu, and X. Wu, "Object detection with deep learning: A review". *IEEE transactions on neural networks and learning systems*, Vol. 30, no. 11, 3212-3232, 2019.

[84] E. Moen, D. Bannon, T. Kudo, W. Graf, M. Covert, and D. J. N. m. Van Valen, "Deep learning for cellular image analysis,", *Nature methods*, pp. 1-14, 2019.

[85] B. Babenko, M. H. Yang, and S. Belongie, "Robust object tracking with online multiple instance learning," *IEEE transactions on pattern analysis and machine intelligence*, vol. 33, no. 8, pp. 1619-1632, 2010.

[86] A. Eitel, J. T. Springenberg, L. Spinello, M. Riedmiller, and W. Burgard, "Multimodal deep learning for robust RGB-D object recognition," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 681-687, 2015.

[87] L. Deng, G. Hinton, and B. Kingsbury, "New types of deep neural network learning for speech recognition and related applications: An overview," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 8599-8603, 2013.

[88] E. Gawehn, J. A. Hiss, and G. Schneider, "Deep learning in drug discovery,", *Molecular informatics,* vol. 35, no. 1, pp. 3-14, 2016.

[89] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *nature,* vol. 521, no. 7553, p. 436, 2015.

[90] W. Rawat and Z. Wang, "Deep convolutional neural networks for image classification: A comprehensive review," *Neural computation,* vol. 29, no. 9, pp. 2352-2449, 2017.

[91] L. Wang, W. Ouyang, X. Wang, and H. Lu, "Visual tracking with fully convolutional networks," in *Proceedings of the IEEE international conference on computer vision*, pp. 3119-3127, 2015.

[92] H. Fan and H. Ling, "Sanet: Structure-aware network for visual tracking," in *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pp. 42-49, 2017.

[93] H. Nam, M. Baek, and B. Han, "Modeling and propagating cnns in a tree structure for visual tracking," *arXiv preprint arXiv,* 2016.

[94] H. Nam and B. Han, "Learning multi-domain convolutional neural networks for visual tracking," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 4293-4302, 2016.

[95] Y. Wu, J. Lim, M. H. Yang, and M. Intelligence, "Object tracking benchmark," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 37, no. 9, pp. 1834-1848, 2015.

[96] M. Danelljan, G. Hager, F. Shahbaz Khan, and M. Felsberg, "Learning spatially regularized correlation filters for visual tracking," in *Proceedings of the IEEE international conference on computer vision*, pp. 4310-4318, 2015.

[97] Y. Qi, S. Zhang, L. Qin, H. Yao, Q. Huang, J. Lim *et al.*, "Hedged deep tracking," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 4303-4311, 2016.

[98] N. Coudray, *P. S.* Ocampo, T. Sakellaropoulos, N. Narula, M. Snuderl, D. Fenyö *et al.*, "Classification and mutation prediction from non–small cell lung cancer histopathology images using deep learning," *Nature medicine,* vol. 24, no. 10, pp. 1559-1567, 2018.

[99] A. Esteva, B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau *et al.*, "Dermatologist-level classification of skin cancer with deep neural networks," *Nature,* vol. 542, no. 7639, pp. 115-118, 2017.

[100] S. Minaee, Y. Boykov, F. Porikli, A. Plaza, N. Kehtarnavaz, and D. J. a. p. a. Terzopoulos, "Image Segmentation Using Deep Learning: A Survey," *arXiv preprint arXiv,* 2020.

[101] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*, pp. 234-241, 2015.

[102] R. Poplin, P. C. Chang, D. Alexander, S. Schwartz, T. Colthurst, A. Ku *et al*., "A universal SNP and small-indel variant caller using deep neural networks," *Nature biotechnology,* vol. 36, no. 10, pp. 983-987, 2018.

[103] B. Alipanahi, A. Delong, M. T. Weirauch, and B. J. Frey, "Predicting the sequence specificities of DNA-and RNA-binding proteins by deep learning," *Nature biotechnology,* vol. 33, no. 8, pp. 831-838, 2015.

[104] C. Angermueller, T. Pärnamaa, L. Parts, and O. Stegle, "Deep learning for computational biology," *Molecular systems biology,* vol. 12, no. 7, 2016.

[105] R. Zhu, X. Tu, and J. Huang, "Using Deep Learning Based Natural Language Processing Techniques for Clinical Decision-Making with EHRs," in *Deep Learning Techniques for Biomedical and Health Informatics*, pp. 257-295, 2020.

[106] J. M. Newby, A. M. Schaefer, P. T. Lee, M. G. Forest, and S. K. Lai, "Convolutional neural networks automate detection for tracking of submicron-scale particles in 2D and 3D," *Proceedings of the National Academy of Sciences*, vol. 115, no. 36, pp. 9026-9031, 2018.

[107] H.-F. Tsai, J. Gajda, T. F. Sloan, A. Rares, and A. Q. Shen, "Usiigaci: Instance-aware cell tracking in stain-free phase contrast microscopy enabled by machine learning," *SoftwareX,* vol. 9, pp. 230-237, 2019.

[108] D. A. Van Valen, T. Kudo, K. M. Lane, D. N. Macklin, N. T. Quach, M. M. DeFelice *et al*., "Deep learning automates the quantitative analysis of individual cells in live-cell imaging experiments," *PLoS computational biology,* vol. 12, no. 11, p. e1005177, 2016.

[109] A. Mathis, P. Mamidanna, K. M. Cury, T. Abe, V. N. Murthy, M. W. Mathis *et al*., "DeepLabCut: markerless pose estimation of user-defined body parts with deep learning," *Nature neuroscience,1*546-1726, 2018.

[110] F. Romero-Ferrero, M. G. Bergomi, R. C. Hinz, F. J. Heras, and G. G. de Polavieja, "idtracker. ai: tracking all individuals in small or large collectives of unmarked animals," *Nature methods,* vol. 16, no. 2, p. 179, 2019.

[111] T. D. Pereira, *D. E.* Aldarondo, L. Willmore, M. Kislin, S. S. H. Wang, M. Murthy *et al.*, "Fast animal pose estimation using deep neural networks," *Nature methods,* vol. 16, no. 1, p. 117, 2019.

[112] D. Gordon, A. Farhadi, and D. Fox, "Re3: Re al-Time Recurrent Regression Networks for Visual Tracking of Generic Objects," *IEEE Robotics and Automation Letters,* vol. 3, no. 2, pp. 788-795, 2018.

[113] Y. Wang, H. Mao, and Z. Yi, "Stem cell motion-tracking by using deep neural networks with multi-output," *Neural Computing and Applications,* pp. 1-13, 2017.

[114] Z. Yi, "Foundations of implementing the competitive layer model by Lotka–Volterra recurrent neural networks," *IEEE Transactions on Neural Networks,* vol. 21, no. 3, pp. 494-507, 2010.

[115] Z. Yi, *Convergence analysis of recurrent neural networks*. Springer Science & Business Media, 2013.

[116] N. Wang and D.-Y. Yeung, "Learning a deep compact image representation for visual tracking," in *Advances in neural information processing systems,* pp. 809-817, 2013.

[117] D. C. Cireşan, A. Giusti, L. M. Gambardella, and J. Schmidhuber, "Mitosis detection in breast cancer histology images with deep neural networks," in *International Conference on Medical Image Computing and Computer-assisted Intervention,* pp. 411-418, 2013.

[118] W.-Z. Nie, W.-H. Li, A.-A. Liu, T. Hao, and Y.-T. Su, "3D convolutional networks-based mitotic event detection in time-lapse phase contrast microscopy image

sequences of stem cell populations," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 55-62, 2016.

[119] Y. Mao and Z. Yin, "A hierarchical convolutional neural network for mitosis detection in phase-contrast microscopy images," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 685-692, 2016.

[120] Y. Wang, H. Mao, Z. Yi, and Applications, "Stem cell motion-tracking by using deep neural networks with multi-output," *Neural Computing and Applications,* vol. 31, no. 8, pp. 3455-3467, 2019.

[121] J. Wang, J. D. MacKenzie, R. Ramachandran, and D. Z. Chen, "Neutrophils identification by deep learning and voronoi diagram of clusters," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 226-233, 2015.

[122] Y. Song, E. L. Tan, X. Jiang, J. Z. Cheng, D. Ni, S. Chen *et al.*, "Accurate cervical cell segmentation from overlapping clumps in pap smear images," *EEE transactions on medical imaging,* vol. 36, no. 1, pp. 288-300, 2016.

[123] H. Li, X. Zhao, A. Su, H. Zhang, J. Liu, and G. Gu, "Color space transformation and multi-class weighted loss for adhesive white blood cell segmentation," *IEEE Access,* vol. 8, pp. 24808-24818, 2020.

[124] T. Falk, D. Mai, R. Bensch, Ö. Çiçek, A. Abdulkadir, Y. Marrakchi *et al.*,"U-Net: deep learning for cell counting, detection, and morphometry," *Nature methods,* vol. 16, no. 1, pp. 67-70, 2019.

[125] D. A. Van Valen, T. Kudo, K. M. Lane, D. N. Macklin, N. T. Quach, M. M. DeFelice *et al.*, "Deep learning automates the quantitative analysis of individual cells in live-cell imaging experiments," *PLoS computational biology,* vol. 12, no. 11, 2016.

[126] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," in *Advances in neural information processing systems*, pp. 91-99, 2015.

[127] T. Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *Proceedings of the IEEE international conference on computer vision*, pp. 2980-2988, 2017.

[128] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask r-cnn," in *Proceedings of the IEEE international conference on computer vision*, pp. 2961-2969, 2017.

[129] G. D. Forney, "The viterbi algorithm," *Proceedings of the IEEE*, vol. 61, no. 3, pp. 268-278, 1973.

[130] D. E. Hernandez, S. W. Chen, E. E. Hunter, E. B. Steager, and V. Kumar, "Cell Tracking with Deep Learning and the Viterbi Algorithm," in *International Conference on Manipulation, Automation and Robotics at Small Scales (MARSS)*, pp. 1-6, 2018.

[131] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 3431-3440, 2015.

[132] L. Rabiner and B. Juang, "An introduction to hidden Markov models," *IEEE assp magazine*, vol. 3, no. 1, pp. 4-16, 1986.

[133] Z. Zhou, F. Wang, W. Xi, H. Chen, P. Gao, and C. He, "Joint Multi-frame Detection and Segmentation for Multi-cell Tracking," in *International Conference on Image and Graphics*, pp. 435-446, 2019.

[134] N. Ballas, L. Yao, C. Pal, and A. Courville, "Delving deeper into convolutional networks for learning video representations," *arXiv preprint arXiv*, 2015.

[135] C. Payer, D. Štern, T. Neff, H. Bischof, and M. Urschler, "Instance segmentation and tracking with cosine embeddings and recurrent hourglass networks," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 3-11, 2018.

[136] J.-B. Lugagne, H. Lin, and M. J. Dunlop, "DeLTA: Automated cell segmentation, tracking, and lineage reconstruction using deep learning," *PLoS computational biology,* vol. 16, no. 4, p. e1007673, 2020.

[137] K. Bernardin and R. Stiefelhagen, "Evaluating multiple object tracking performance: the CLEAR MOT metrics," *Journal on Image and Video Processing,* vol. 2008, p. 1, 2008.

[138] E. Ristani, F. Solera, R. Zou, R. Cucchiara, and C. Tomasi, "Performance measures and a data set for multi-target, multi-camera tracking," in *European Conference on Computer Vision*, pp. 17-35, 2016.

[139] P. Matula, M. Maška, D. V. Sorokin, P. Matula, C. Ortiz-de-Solórzano, and M. Kozubek, "Cell tracking accuracy measurement based on comparison of acyclic oriented graphs," *PloS one,* vol. 10, no. 12, p. e0144959, 2015.

[140] Y. Li, A. Dore, and J. Orwell, "Evaluating the performance of systems for tracking football players and ball," in *IEEE Conference on Advanced Video and Signal Based Surveillance,* pp. 632-637, 2005.

[141] A.-T. Nghiem, F. Bremond, M. Thonnat, and V. Valentin, "ETISEO, performance evaluation for video surveillance systems," in *IEEE Conference on Advanced Video and Signal Based Surveillance*, pp. 476-481, 2007.

[142] C. H. Huang, S. Sankaran, D. Racoceanu, S. Hariharan, and S. Ahmed, "Online 3-D tracking of suspension living cells imaged with phase-contrast microscopy,"

IEEE Transactions on Biomedical Engineering, vol. 59, no. 7, pp. 1924-33, Jul 2012.

[143] S. L. Charreyron, Q. Boehler, A. Danun, A. Mesot, M. Becker, B. J. Nelson, "A Magnetically Navigated Microcannula for Subretinal Injections," *IEEE Transactions on Biomedical Engineering,* pp. 1-1, 2020.

[144] X. P. Zhang, C. Leung, Z. Lu, N. Esfandiari, R. F. Casper, and Y. Sun, "Controlled aspiration and positioning of biological cells in a micropipette," *IEEE transactions on biomedical engineering,* vol. 59, no. 4, pp. 1032-1040, 2012.

[145] M. A. A. Dewan, M. O. Ahmad, and M. N. S. Swamy, "Tracking biological cells in time-lapse microscopy: An adaptive technique combining motion and topological features," *IEEE transactions on biomedical engineering,* vol. 58, no. 6, pp. 1637-1647, 2011.

[146] V. Khodadadi, E. Fatemizadeh, and S. K. Setarehdan, "Optimized kalman filter based on second momentum and triple rectangular for cell tracking on sequential microscopic images," in *22nd Iranian Conference on Biomedical Engineering (ICBME)*, pp. 251-256, 2015.

[147] H. N. Lim, M. Y. Mashor, N. Z. Supardi, and R. Hassan, "Color and morphological based techniques on white blood cells segmentation," in *2nd International Conference on Biomedical Engineering (ICoBE)*, pp. 1-5, 2015.

[148] D. Carradice and G. J. Lieschke, "Zebrafish in hematology: sushi or science?," *Blood*, vol. 111, no. 7, pp. 3331-42, Apr 1 2008.

[149] G. J. Lieschke and P. D. Currie, "Animal models of human disease: zebrafish swim into view," *Nature Reviews Genetics,* vol. 8, no. 5, p. 353, 2007.

[150] M. L. Suster, H. Kikuta, A. Urasaki, K. Asakawa, and K. Kawakami, "Transgenesis in zebrafish with the tol2 transposon system," *Methods Mol Biol*, vol. 561, pp. 41-63, 2009.

[151] V. Ulman *et al.*, "An objective comparison of cell-tracking algorithms," *Nature methods*, vol. 14, no. 12, p. 1141, 2017.

[152] J. Schindelin, I. Arganda-Carreras, E. Frise, V. Kaynig, M. Longair, T. Pietzsch *et al.*, "Fiji: an open-source platform for biological-image analysis," *Nature methods*, vol. 9, no. 7, p. 676, 2012.

[153] J. E. Purvis and G. Lahav, "Encoding and decoding cellular information through signaling dynamics," *Cell*, vol. 152, no. 5, pp. 945-956, 2013.

[154] J. C. Kimmel, A. Y. Chang, A. S. Brack, and W. F. Marshall, "Inferring cell state by quantitative motility analysis reveals a dynamic state system and broken detailed balance," *PLoS computational biology*, vol. 14, no. 1, p. e1005927, 2018.

[155] P. Wang, L. Robert, J. Pelletier, W. L. Dang, F. Taddei, A. P. Wright, A. Wang *et al.*, "Robust growth of Escherichia coli," *Current biology*, vol. 20, no. 12, pp. 1099-1103, 2010.

[156] Q. Chen, B. Chen, and C. Zhang, "Modeling conserved structure patterns for functional noncoding RNA," in *Intelligent Strategies for Pathway Mining*, pp. 151-173, 2014.

[157] J. K. H. Chiu and Y. P. P. Chen, "Efficient conversion of RNA pseudoknots to knot-free structures using a graphical model," *IEEE Transactions on Biomedical Engineering*, vol. 62, no. 5, pp. 1265-1271, 2014.

[158] F. Chen and Y. P. P. Chen, "Exploring cross-species-related miRNAs based on sequence and secondary structure," *IEEE transactions on biomedical engineering*, vol. 57, no. 7, pp. 1547-1553, 2010.

[159] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436-444, May 28 2015.

[160] C. Y. Lee and Y. P. P. Chen, "Prediction of drug adverse events using deep learning in pharmaceutical discovery," *Briefings in Bioinformatics*, 2020.

[161] Y. Gu, M. Shen, J. Yang, and G. Z. Yang, "Reliable Label-Efficient Learning for Biomedical Image Recognition," *IEEE Transactions on Biomedical Engineering*, vol. 66, no. 9, pp. 2423-2432, 2018.

[162] D. Lee, J. Yoo, S. Tak, and J. C. Ye, "Deep residual learning for accelerated MRI using magnitude and phase networks," *IEEE Transactions on Biomedical Engineering*, vol. 65, no. 9, pp. 1985-1995, 2018.

[163] M. Maška, V. Ulman, D. Svoboda, P. Matula, C. Ederra, *et al*., "A benchmark for comparison of cell tracking algorithms. *Bioinformatics"*, vol. *30, no.* 11, pp.1609-1617, 2014.

[164] N. Chenouard, I. Smal, F. De Chaumont, M. Maška, I. F. Sbalzarini, Y. Gong *et al*., "Objective comparison of particle tracking methods," *Nature methods*, vol. *11, no.* 3, pp. 281, 2014.

[165] V. Pazhakh, F. Ellett, B. A. Croker, J. A. O'Donnell, L. Pase, K. E. Schulze *et al.*, "beta-glucan-dependent shuttling of conidia from neutrophils to macrophages occurs during fungal infection establishment," *PLoS Biol*, vol. 17, no. 9, p. e3000113, Sep 2019.

[166] F. Ellett , V. Pazhakh, L. Pase, E. L. Benard, H. Weerasinghe, D. Azabdaftari *et al.*, "Macrophages protect Talaromyces marneffei conidia from myeloperoxidase-dependent neutrophil fungicidal activity during infection establishment in vivo," *PLoS pathogens*, vol. 14, no. 6, p. e1007063, 2018.

[167] C. Nüsslein-Volhard and R. Dahm, *Zebrafish : a practical approach*. Oxford: Oxford University Press, 2002.

[168] F. Ellett, L. Pase, J. W. Hayman, A. Andrianopoulos, and G. J. Lieschke, "mpeg1 promoter transgenes direct macrophage-lineage expression in zebrafish," *Blood, The Journal of the American Society of Hematology*, vol. 117, no. 4, pp. e49-e56, 2011.

[169] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE transactions on systems, man, and cybernetics*, vol. 9, no. 1, pp. 62-66, 1979.

[170] H. R. Manley, M. C. Keightley, and G. J. Lieschke, "The Neutrophil Nucleus: An Important Influence on Neutrophil Migration and Function," *Front Immunol*, vol. 9, p. 2867, 2018.

[171] V. Pazhakh, F. Ellett, B. A. Croker, J. A. O'Donnell, L. Pase, K. E. Schulze *et al.*, "beta-glucan-dependent shuttling of conidia from neutrophils to macrophages occurs during fungal infection establishment," *PLoS Biol*, vol. 17, no. 9, p. e3000113, Sep 2019.

[172] K. König, "Multiphoton microscopy in life sciences," *Journal of microscopy*, vol. 200, pp. 83–104, 2000.

[173] R. B. Nussenblatt, "Elements of the Immune System and Concepts of Intraocular Inflammatory Disease Pathogenesis," *Uveitis* pp. 1-36, 2010.

[174] J. H. Wilson and T. Hunt, *Molecular biology of the cell : a problems approach*. New York, London: Garland, 2002.

[175] J. Cui, S. T. Acton, and Z. L. Lin, "A Monte Carlo approach to rolling leukocyte tracking in vivo," *Medical Image Analysis*, vol. 10, no. 4, pp. 598-610, Aug 2006.

[176] Y. Wu, J. Lim, and M. H. Yang, "Online Object Tracking: A Benchmark," (in English), *IEEE Conference on Computer Vision and Pattern Recognition (Cvpr)*, pp. 2411-2418, 2013.

[177] O. K. Oyedotun and A. Khashman, "Deep learning in vision-based static hand gesture recognition," *Neural Computing & Applications,* vol. 28, no. 12, pp. 3941-3951, Dec 2017.

[178] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," *Communications of the Acm,* vol. 60, no. 6, pp. 84-90, Jun 2017.

[179] H. Nam and B. Han, "Learning Multi-Domain Convolutional Neural Networks for Visual Tracking," *IEEE Conference on Computer Vision and Pattern Recognition (Cvpr),* pp. 4293-4302, 2016.

[180] B. Dong, L. Shao, M. Da Costa, O. Bandmann, and A. F. Frangi, "Deep Learning for Automatic Cell Detection in Wide-Field Microscopy Zebrafish Images," *IEEE 12th International Symposium on Biomedical Imaging (ISBI),* pp. 772-776, 2015.

[181] F. H. Li, X. B. Zhou, J. W. Ma, and S. T. C. Wong, "Multiple Nuclei Tracking Using Integer Programming for Quantitative Cancer Cell Cycle Analysis," *IEEE Transactions on Medical Imaging,* vol. 29, no. 1, pp. 96-105, Jan 2010.

[182] Y. Y. Ren, B. L. Xu, J. Zhang, W. M. Zhang, and L. Xu, "A Generalized Data Association Approach for Cell Tracking in High-Density Population," *Fourth International Conference on Control, Automation and Information Sciences (Ccais),* pp. 502-507, 2015.

[183] R. Bise, Z. Z. Yin, and T. Kanade, "Reliable Cell Tracking by Global Data Association," *8th IEEE International Symposium on Biomedical Imaging,* pp. 1004-1010, 2011.

[184] K. M. Henry, L. Pase, C. F. Ramos-Lopez, G. J. Lieschke, S. A. Renshaw, and C. C. Reyes-Aldasoro, "PhagoSight: An Open-Source MATLAB (R) Package for the Analysis of Fluorescent Neutrophil and Macrophage Migration in a Zebrafish Model," *Plos One,* vol. 8, no. 8, Aug 30 2013.

[185] N. Ray, S. T. Acton, and K. Ley, "Tracking leukocytes in vivo with shape and size constrained active contours," *IEEE transactions on medical imaging,* vol. *21, no.* 10, pp. 1222-1235, 2002.

[186] X. H. Lou, M. Schiegg, and F. A. Hamprecht, "Active Structured Learning for Cell Tracking: Algorithm, Framework, and Usability," *IEEE Transactions on Medical Imaging,* vol. 33, no. 4, pp. 849-860, Apr 2014.

[187] O. Dzyubachyk, W. A. van Cappellen, J. Essers, W. J. Niessen, and E. Meijering, "Advanced Level-Set-Based Cell Tracking in Time-Lapse Fluorescence Microscopy," *IEEE Transactions on Medical Imaging,* vol. 29, no. 6, pp. 1331-1331, Jun 2010.

[188] A. C. Y. Wu and S. A. Rifkin, "Aro: a machine learning approach to identifying single molecules and estimating classification error in fluorescence microscopy images," *Bmc Bioinformatics,* vol. 16, Mar 27 2015.

[189] B. N. Zhong, S. Pan, C. Wang, T. Wang, J. Du, D. Chen, and L. Cao, "Robust Individual-Cell/Object Tracking via PCANet Deep Network in Biomedicine and Computer Vision," *Biomed Research International,* 2016.

[190] T. He, H. Mao, J. X. Guo, and Z. Yi, "Cell tracking using deep neural networks with multi-task learning," *Image and Vision Computing,* vol. 60, pp. 142-153, Apr 2017.

[191] J. Hayashida and R. Bise, "Cell Tracking with Deep Learning for Cell Detection and Motion Estimation in Low-Frame-Rate," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 397-405, 2019.

[192] Y.-T. Su, Y. Lu, M. Chen, and A. A. Liu, "Spatiotemporal joint mitosis detection using CNN-LSTM network in time-lapse phase contrast microscopy images," *IEEE Access,* vol. 5, pp. 18033-18041, 2017.

[193] M. S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking," *IEEE Transactions on Signal Processing,* vol. 50, no. 2, pp. 174-188, Feb 2002.

[194] Y. D. Wang N, "Learning a deep compact image representation for visual tracking," *Neural Information Processing Systems (NIPS),* pp. 809–817, 2013.

[195] N. Otsu, "Threshold Selection Method from Gray-Level Histograms," *IEEE Transactions on Systems Man and Cybernetics,* vol. 9, no. 1, pp. 62-66, 1979.

[196] S. Oron, A. Bar-Hillel, D. Levi, and S. Avidan, "Locally orderless tracking," *International Journal of Computer Vision,* vol. 111, no. 2, pp. 213-228, 2015.

[197] K. Zhang, L. Zhang, and M.-H. Yang, "Real-time compressive tracking," in *European conference on computer vision*, pp. 864-877, 2012.

[198] L. Sevilla-Lara and E. Learned-Miller, "Distribution fields for tracking," in *2012 IEEE Conference on computer vision and pattern recognition*, pp. 1910-1917, 2012.

[199] C. Bao, Y. Wu, H. Ling, and H. Ji, "Real time robust l1 tracker using accelerated proximal gradient approach," in *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1830-1837, 2012.

[200] J. F. Henriques, R. Caseiro, P. Martins, and J. Batista, "Exploiting the circulant structure of tracking-by-detection with kernels," in *European conference on computer vision*, pp. 702-715, 2012.

[201] T. Zhang, B. Ghanem, S. Liu, and N. Ahuja, "Robust visual tracking via multi-task sparse learning," in *2012 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2042-2049, 2012.

[202] X. Jia, H. Lu, and M. H. Yang, "Visual tracking via adaptive structural local sparse appearance model," in *IEEE Conference on computer vision and pattern recognition*, pp. 1822-1829, 2012.

[203] C. Rosales, N. Demaurex, C. A. Lowell, and E. Uribe-Querol, "Neutrophils: Their Role in Innate and Adaptive Immunity," *Journal of Immunology Research*, 2016.

[204] L. S. Chen, Y. Zhao, D. Lai, P. Zhang, Y. Yang, Y. Li *et al.*, "Neutrophil extracellular traps promote macrophage pyroptosis in sepsis," *Cell Death & Disease*, vol. 9, May 22 2018.

[205] R. Vivas, A. A. T. Barbosa, S. S. Dolabela, and S. Jain, "Multidrug-Resistant Bacteria and Alternative Methods to Control Them: An Overview," *Microbial Drug Resistance*, vol. 25, no. 6, pp. 890-908, Jul 1 2019.

[206] V. Rai and N. Dey, "The Basics of Confocal Microscopy," *Laser Scanning, Theory and Applications*, pp. 75-96, 2011.

[207] M. Yamada, L. L. Lin, and T. W. Prow, "Multiphoton Microscopy Applications in Biology," *Fluorescence Microscopy: Super-Resolution and Other Novel Techniques*, pp. 185-197, 2014.

[208] K. Chatterjee, F. W. Pratiwi, F. C. M. Wu, P. L. Chen, and B. C. Chen, "Recent Progress in Light Sheet Microscopy for Biological Applications," *Applied Spectroscopy*, vol. 72, no. 8, pp. 1137-1169, Aug 2018.

[209] K. Wu, D. Gauthier, and M. D. Levine, "Live cell image segmentation," *IEEE Transactions on biomedical engineering*, vol. 42, no. 1, pp. 1-12, 1995.

[210] Q. Wu, F. Merchant, and K. Castleman, *Microscope image processing*. Elsevier, 2010.

[211] E. Moen, D. Bannon, T. Kudo, W. Graf, M. Covert, and D. Van Valen, "Deep learning for cellular image analysis," *Nature methods*, p. 1, 2019.

[212] S. Parlato, A. De Ninno, R. Molfetta, E. Toschi, D. Salerno, A. Mencattini *et al.*, "3D Microfluidic model for evaluating immunotherapy efficacy by tracking dendritic cell behaviour toward tumor cells," *Scientific reports*, vol. 7, no. 1, p. 1093, 2017.

[213] E. Bochinski, V. Eiselein, and T. Sikora, "High-speed tracking-by-detection without using image information," in 1*4th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, pp. 1-6, 2017.

[214] R. E. Kalman, "A new approach to linear filtering and prediction problems," *Journal of basic Engineering*, vol. 82, no. 1, pp. 35-45, 1960.

[215] H. Shen, G. Nelson, S. Kennedy, D. Nelson, J. Johnson, D. Spiller *et al.*, "Automatic tracking of biological cells and compartments using particle filters and active contours," *Chemometrics and Intelligent Laboratory Systems*, vol. 82, no. 1-2, pp. 276-282, 2006.

[216] C. Tang, Y. Wang, and Y. Cui, "Tracking of Active Cells Based on Kalman Filter in Time Lapse of Image Sequences of Neuron Stem Cells," in *Proceedings of the International Conference on Bioinformatics & Computational Biology (BIOCOMP)*, p. 1, 2011.

[217] M. Liu, Y. He, Y. Wei, and P. Xiang, "Plant cell tracking using Kalman filter based local graph matching," *Image and Vision Computing*, vol. 60, pp. 154-161, 2017.

[218] S. Abousamra, S. Adar, N. Elia, and R. Shilkrot, "Localization and Tracking in 4D Fluorescence Microscopy Imagery," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 2290-2298, 2018.

[219] V. Nair and G. E. Hinton, "Rectified linear units improve restricted boltzmann machines," in *Proceedings of the 27th international conference on machine learning (ICML-10)*, pp. 807-814, 2010.

[220] K. Hornik, M. Stinchcombe, and H. White, "Multilayer feedforward networks are universal approximators," *Neural networks*, vol. 2, no. 5, pp. 359-366, 1989.

[221] C. C. Reyes-Alsdasoro, S. Akerman, and G. Tozer, "Measuring the velocity of fluorescently labelled red blood cells with a keyhole tracking algorithm," *Journal of microscopy*, vol. 229, no. 1, pp. 162-173, 2008.

[222] V. Depraetere, "“Eat me” signals of apoptotic bodies," *Nature cell biology*, vol. 2, no. 6, p. E104, 2000.

[223] S. Domhan, L. Ma, A. Tai, Z. Anaya, A. Beheshti, M. Zeier *et al.*, "Intercellular communication by exchange of cytoplasmic material via tunneling nano-tube like structures in primary human renal epithelial cells," *PloS one*, vol. 6, no. 6, p. e21283, 2011.

[224] S. Gordon, "Phagocytosis: An Immunobiologic Process," *Immunity*, vol. 44, no. 3, pp. 463-475, Mar 15 2016.

[225] C. Rosales and E. Uribe-Querol, "Phagocytosis: A Fundamental Process in Immunity," *Biomed Res Int*, vol. 2017, p. 9042851, 2017.

[226] J. J. Lim, S. Grinstein, and Z. Roth, "Diversity and Versatility of Phagocytosis: Roles in Innate Immunity, Tissue Remodeling, and Homeostasis," *Front Cell Infect Microbiol*, vol. 7, p. 191, 2017.

[227] N. C. Shaner, P. A. Steinbach, and R. Y. Tsien, "A guide to choosing fluorescent proteins," *Nature methods,* vol. 2, no. 12, pp. 905-909, 2005.

[228] J. Lippincott-Schwartz and G. H. Patterson, "Development and use of fluorescent protein markers in living cells," *Science,* vol. 300, no. 5616, pp. 87-91, 2003.

[229] B. P. Cormack, R. H. Valdivia, and S. Falkow, "FACS-optimized mutants of the green fluorescent protein (GFP)," *Gene,* vol. 173, no. 1, pp. 33-38, 1996.

[230] L. Wilson, P. T. Matsudaira, K. F. Sullivan, and S. A. Kay, *Green fluorescent proteins*. Academic Press, 1998.

[231] R. E. Campbell, O. Tour, A. E. Palmer, P. A. Steinbach, G. S. Baird, D. A. Zacharias *et al.*, "A monomeric red fluorescent protein," *Proceedings of the National Academy of Sciences,* vol. 99, no. 12, pp. 7877-7882, 2002.

[232] T. Zimmermann, "Spectral imaging and linear unmixing in light microscopy," in *Microscopy techniques*, pp. 245-265, 2005.

[233] E. Schröck S. Du Manoir, T. Veldman, B. Schoell, J. Wienberg, M. A. Ferguson-Smith *et al.*, "Multicolor spectral karyotyping of human chromosomes," *Science,* vol. 273, no. 5274, pp. 494-497, 1996.

[234] H. Tsurui, H. Nishimura, S. Hattori, S. Hirose, K. Okumura, and T. Shirai, "Seven-color fluorescence imaging of tissue samples based on Fourier spectroscopy and singular value decomposition," *Journal of Histochemistry & Cytochemistry,* vol. 48, no. 5, pp. 653-662, 2000.

[235] R. Lansford, G. Bearman, and S. E. Fraser, "Resolution of multiple green fluorescent protein color variants and dyes using two-photon microscopy and imaging spectroscopy," *Journal of biomedical optics,* vol. 6, no. 3, pp. 311-318, 2001.

[236] M. Dickinson, G. Bearman, S. Tille, R. Lansford, and S. Fraser, "Multi-spectral imaging and linear unmixing add a whole new dimension to laser scanning fluorescence microscopy," *Biotechniques,* vol. 31, no. 6, pp. 1272-1278, 2001.

[237] F. S. Wouters, P. J. Verveer, and P. I. Bastiaens, "Imaging biochemistry inside cells," *Trends in cell biology,* vol. 11, no. 5, pp. 203-211, 2001.

[238] T. Zimmermann and F. Siegert, "Simultaneous detection of two GFP spectral mutants during in vivo confocal microscopy of migrating Dictyostelium cells," *Biotechniques,* vol. 24, no. 3, pp. 458-461, 1998.

[239] F. Olschewski, "Living colors," *Imaging & Microscopy,* vol. 4, no. 2, pp. 22-24, 2002.

[240] R. A. Neher, M. Mitkovski, F. Kirchhoff, E. Neher, F. J. Theis, and A. Zeug, "Blind source separation techniques for the decomposition of multiply labeled fluorescence images," *Biophysical journal,* vol. 96, no. 9, pp. 3791-3800, 2009.

[241] C. A. Pfortmueller, C. Meisel, M. Fux, and J. C. Schefold, "Assessment of immune organ dysfunction in critical illness: utility of innate immune response markers," *Intensive Care Med Exp,* vol. 5, no. 1, p. 49, Oct 23 2017.