Single-Cell Imaging in Bacterial Biofilms: Data Acquisition, Image Processing, and Quantitative Analysis

Yibo Wang

A Dissertation presented to the Graduate Faculty of the University of Virginia in Candidacy for the Degree of Doctor of Philosophy

Department of Chemistry

University of Virginia
August 2023

Andreas Gahlmann
David Cafiso
Ian Harrison
Linda Columbus
Cliff Stains
Scott Acton

Dedicated in loving memory

to

my grandmother, Guiqing Yang,

and to

my grandfather, Xisen Rong,

who brought me up and always believed in me

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude and appreciation to the individuals who have contributed to the completion of this Ph.D. thesis. Without their invaluable support, guidance and encouragement over the years, this achievement would not have been possible.

First and foremost, I would like to extend my gratitude to my supervisor, Dr. Andreas Gahlmann. I would like to share two brief anecdotes here. When I first joined the lab, I felt completely overwhelmed by the complexity of the lattice light sheet microscope. I was afraid of touching any optical components, thinking that a small bump could ruin the entire microscope. However, you trusted me and asked me to align the new 647nm laser, even though I was new to the lab. I gained so much confidence over the course, and for the first time, I believed that I belonged to and could thrive in the Gahlmann lab. After my second year, I wanted to pursue a concurrent master's degree in Data Science. I was worried that you would oppose to the idea because it might hinder my research progress. However, you were completely on board with the idea, and I still remember staying in your office to revise my program application until 7:30 pm on a Friday evening. Your insightful feedback and constructive criticism have consistently pushed me to strive for excellence. I am grateful for the opportunity to work under your guidance and for the invaluable knowledge and skills I have gained from our interactions.

I would also like to thank the members of my Ph.D. committee, Dr. David Cafiso, Dr. Linda Columbus, Dr. Ian Harrison, Dr. Cliff Stains, and Dr. Scott Acton. Your expertise, feedback, curiosity and constructive suggestions have significantly contributed to the refinement and enrichment of this thesis. I am sincerely grateful for the time and effort you dedicated to reviewing my work and providing invaluable insights over the years.

I would like to express my gratitude to my colleagues, coauthors and mentors: Dr. Mingxing Zhang, Dr. Ji Zhang and Dr. Alecia Achimovich. You were the ones that taught me step-by-step how to use the microscope, got me started in coding, as well as extensive microbiology techniques. Without you, I would not be able to contribute to the research projects immediately after joining the lab. Your intellectual discussion and willingness to help have made this journey more memorable. I also want to thank my collaborators in the Acton lab: Dr. Jie Wang, Tanjin Toma, and Dr. Scott Acton. You have been stellar collaborators and great problem solvers. Your instrumental role in tackling difficult image-processing problems has made our collaboration quite fruitful. I also want to thank my colleagues, Josh Prindle, Olivia de Cuba, Maddie Miles, Ting Yan, Deepika Parthasarrathy, and Eric Donarski. I am grateful for the countless intellectual debates, coffee breaks, lunches, sport events and social gatherings that provided a much-needed friendly and supportive environment crucial for my academic success. Your presence has not only made the experience more enjoyable but has also fostered a sense of belonging.

I would like to express my deepest gratitude to my family. My parents have been visionary and planted a seed early in my life to grow my passion for science. You have now finally witnessed it grow into something full and admirable. I appreciate your support in every decision I have made. You are, and have always been, my role models, and I would show the same level of positivity and respect to my children. I also want to thank my grandparents. Your constant support, patience, and encouragement made me who I am today. You have been my anchor throughout this demanding couple of years, especially through the years of pandemic that separated us.

Finally, I want to thank my loving wife and my best friend, Yoko Chan. Throughout our entire adulthood, we have traveled to so many different places and shared numerous joyful memories. You have been the witness to both my personal and professional growth from an inexperienced young student to a seasoned professional. You have played an integral role in every major decision I have made. It is because of your unwavering belief in me that I began my Ph.D. journey and will ultimately complete it despite any hardship and challenges posed by the global pandemic. There are more chapters to write in our lives, and I am looking forward to the endless possibilities we are about to embrace.

ABSTRACT

Bacterial biofilms are complex 3-dimentional (3D) structures with substantial spatial and temporal heterogeneity at the single-cell level. Simultaneous multi-cell tracking in 3D is thus critical for analyzing single-cell behaviors, such as motility and metabolism, as well as lineage tracing in biofilms. Due to phototoxicity and photobleaching concerns, fluorescence images are often subject to low signal-to-background ratios (SBRs). High cell density, and large relative cell movements from frame to frame add additional challenges for accurate segmentation and tracking of individual cells in living biofilms. To address these challenges, I used lattice light sheet microscopy (LLSM) to image 3D bacterial biofilms with high spatial and temporal resolution and without substantial light-induced degradation of the SBRs over time. I additionally incorporated a hermetically sealed microfluidic flow channel into the LLSM to sustain bacterial biofilm growth under precisely controllable physical and chemical conditions. To enable accurate cell segmentation in the acquired 3D movies, I trained convolutional neural networks (CNNs) to perform voxel classification and to translate 3D fluorescence images into 3D intermediate image representations that are more resistant to over- and under-segmentation errors. Using this approach, improved segmentation results are obtained even for low SBRs and/or high cell density biofilm images. In order to track individual cells, I further leveraged a separate machine learning algorithm to select cell features that facilitate linking corresponding cells between frames. We demonstrate the applications and limitations of our entire data processing pipeline by systematically evaluating tracking accuracy using both simulated and experimentally acquired time-lapse images. The combination of non-invasive imaging and machine-learning based computational image analysis pipeline provides new opportunities for investigating time-dependent phenomena in living bacterial biofilms with single-cell resolution.

Table of Contents

Chapter1: Introduction

1.1 Bacterial biofilms and host-pathogen interface

Bacterial biofilms are cohesive, multicellular microbial communities capable of adhering to biotic or abiotic surfaces[1, 2]. They constitute a major component of bacterial biomass on Earth, and thus have many direct impacts on public health, medical treatment and diagnosis as well as the food industry[3]. Undesired biofilm growth has been identified as a prominent problem in various industries. In healthcare and medical settings, most bacterial infection cases are associated with biofilms, which are known to develop resistance to antibiotic treatments[4]. The estimated global expenditure on wound care linked to biofilms exceeds two hundred billion dollars annually[5]. Therefore, alternative and more efficient biofilm control strategies are not only crucial for combating bacterial infections but also are an economic necessity.

Bacterial biofilms also play a positive, indispensable role in various industries, necessitating further research to harness their potentials. For example, bacterial biofilms play essential roles in wastewater management plans, where they either directly or indirectly aid the consumption of toxic materials in wastewater[6]. On the other hand, certain species of biofilms are capable of producing bioelectricity, encouraging scientists to research on the development of fuel cells powered by biofilms[7]. In relation to human health, commensal bacteria perform digestive functions and serve as a critical line of defense against pathogens.

It is paramount to closely investigate the biofilm cell biology and life cycle, whether the goal is to develop novel strategies to combat bacterial infections or to leverage beneficial microbial ecosystems for practical applications. It is believed that bacterial biofilm formation is a multi-step process. After bacteria interact with a surface, the production of extracellular polymeric substances (EPS) promotes cell-cell adhesion and ultimately allows biofilms to aggregate and mature[2]. However, the biofilm life cycle of different species is diverse over various stages and is yet to be fully understood. The understandings of biofilm formation were largely based on ensemble average data, and biofilm studies were typically performed without measuring single-cell statistics. However, with the recent development of advanced imaging modalities, researchers now are trying to understand bacterial biofilms as well as their spatial-temporal heterogeneity at the single-cell level. Recent studies have indicated that bacterial biofilms are known to promote phenotypic diversity (e.g. changes in growth rate or gene expression) among individual bacteria that enable the coordination of cellular behaviors and the development of heterogenous functionalities[1, 8-10]. However, conventional microbiological or biochemical experiments did not preserve the spatial-temporal information of individual cells, and therefore little was known about dynamic changes in motility, cell shapes, and gene expression profiles of individual cells in heterogeneous biofilms in response to different environmental cues and variable chemical signals. These mechanisms can enhance population survival in harsh environments and provide bacterial biofilms with emergent functional capabilities beyond those of planktonic bacterial cells. Therefore, rigorous investigations of single-cell behaviors are key to study not only for fundamental biology, such as biofilm life cycle, but also for practical applications associated with biofilms.

The host-pathogen interface, or host-microbe interface, represents the surface where pathogenic or commensal bacteria attach to and interact with host cells, making it physiologically relevant to study bacterial biofilms in human health. Bacterial biofilms are commonly observed to interact with host mucosal epithelia in the airway, gut, and genitourinary tract[11]. The host environment plays a critical role in the formation and regulation of bacterial biofilms in both health and disease, and thus is of high interest to researchers in various fields. However, little is known of how bacterial biofilm formation differs at host mucosal environments consisting of different structures, functions, and nutrient availability.

1.2 Advanced microscopy for imaging bacterial biofilms

Fluorescence microcopy has been a driving force for cell biology for many decades. The development of modern fluorescence microscopes has made it possible to acquire four-dimensional bacterial biofilm dynamics (x, y, z, and time). For instance, confocal microscopy and related imaging modalities like spinning disk confocal microscopy enable 3D time-lapse cellular and subcellular visualizations of live biological samples. In confocal microscopy, samples are scanned over a 2D or 3D raster, and a spatial pinhole (or equivalent) is typically used to block out-of-focus light, thereby enhancing contrast. Although confocal microscopy offers good contrast and spatial resolution, it cannot continuously resolve individual cells over a long period of time due to the use of high-intensity lasers and waste of photon budgets (i.e., a large amount of out-of-focus light is blocked and discarded)[12, 13].

To overcome the abovementioned problems, light sheet-based fluorescence imaging modalities have been recently developed, combining high resolution with fast imaging speed and low phototoxicity surpassing the capabilities of confocal microscopy[13-15]. Light sheet-based microscopy has gained popularity for non-invasive time-lapse imaging of live biological tissues and samples, including bacterial biofilms[16-18]. However, even light sheet microscopy has limitations. Depending on the objectives of the research aims, the cellular sizes may be too small or, the cellular density or tissue thickness may be too high to clearly resolve individual cells with diffraction-limited microscopy. Super-resolution imaging modalities, in contrast, such as structured illumination microscopy[19, 20], can theoretically improve the spatial resolution by two-fold, but such improvements come at the cost of decreased temporal resolution and increased concerns of photobleaching and phototoxicity[21, 22]. Hence, trade-offs between contrast, photobleaching, and spatial and temporal resolution need to be carefully considered when choosing an appropriate imaging modality for a specific research objective.

An additional challenge in the field is gaining access to the imaging of physiologically relevant biofilm and host-microbe interfaces. There have been approaches to target bacterial biofilm as well as the host-microbe interface with either *in vivo* or *ex vivo* microscopy. Each approach has its advantages and disadvantages. The *in vivo* approach allows for close investigation of biofilms with physiological relevancy, but it is not feasible to obtain images with single-cell resolution due to typically used low numerical aperture objective lenses as well as other constraints [23, 24]. On the other hand, *ex vivo* imaging of the tissue slices offer single-cell, high spatial resolution of the host-pathogen interface, but usually involves imaging fixed tissue slices[25]. While offering high spatial resolution, such approaches can only capture single snapshots in time and thus fail to encapsulate dynamic information on host-microbe interactions. Static culture experiments imaged using confocal microscopy on immortalized human cell lines or harvested living tissues from animal models can offer temporal information, but they do not accurately reproduce the complexity and dynamism of human host-microbe interfaces *in vivo*, where shear forces, chemical gradients, and environment changes due to fluid flow are present. Although it is unlikely that one imaging setup can address all the aforementioned limitations, it can be concluded that using a 3D imaging modality with both high spatial and temporal resolution to image live cells, coupled with microfluidic technology for precise control of the physical and chemical environment, is the best approach for imaging host-microbe interfaces.

In our lab, we use lattice light-sheet microscopy (LLSM), an advanced light sheet microscopy with high numerical aperture that combines diffraction-limited 3D spatial resolution (~250nm in xy and ~700 nm in z) with fast temporal resolution and low phototoxicity. LLSM is well-suited for imaging bacterial biofilms, capturing both rapid short-term events such as motility and cell dispersion, as well as long-term phenomena such as cell growth dynamics and cell divisions[26]. More importantly, LLSM can operate at illumination intensities that are below the levels of cellular phototoxicity. LLSM can reduce fluorophore photobleaching by 1-2 orders of magnitude compared to confocal microscopy, thus achieving better spatial temporal resolution and signal-to-background ratios over extended period of time. Furthermore, LLSM outperforms some super-resolution modalities in terms of data acquisition speed, while being able to still resolve individual cell boundaries in a dense biofilm. However, computational image analysis, such as image segmentation and cell tracking, becomes more challenging with LLSM due to small intercellular gaps between bacterial cells within densely packed biofilms and the diffraction-limited nature of the imaging system.

Therefore, I have developed machine learning and deep learning-based computational solutions to recognize, segment, classify and track individual bacterial cells in bacterial biofilms. Further details on these approaches will be provided in Chapter 3 and Chapter 4. In addition to software developments, this thesis also represents a first step towards addressing the challenges of imaging the host-pathogen interface by integrating microfluidic technology and best practices for running stable, long-term experiments on LLSM.

1.3 Deep learning for microscopy image processing

Modern microscopy modalities, such as confocal and light sheet microscopy, are able to generate a massive amount of raw, 3D time-lapse imaging data. However, processing these data, as well as doing quantitative, unbiased measurements, is not a trivial task. In this section, I introduce how deep learning techniques have revolutionized the field of microscopy image processing and briefly explain how and why deep learning techniques work.

Machine learning algorithms can learn from the data to predict desired outputs or discover patterns. Deep learning algorithms are normally considered a subset of machine learning algorithms that use artificial neural networks that mimic how a human brain functions. More formally, a machine learning algorithm can be viewed as a function, f, that maps an input, x, to an output, y, by updating model parameters, θ[27].

$$y = f(x; \theta)$$

Machine learning algorithms usually work on structured data, where the input, x, is typically a data frame or a matrix that contains m rows of data entry and n columns of features. A deep learning algorithm however typically works better for unstructured data, where features are difficult or impossible to extract effectively from highly complex data, such as texts, images, or videos.

There are a variety of tasks in the field of microscopy image processing that are well-suited for deep learning techniques, namely classification, image segmentation, object tracking and image reconstruction[28]. These tasks are usually fit into the supervised learning framework, which relies on labeled data (ground truth). Specifically, image classification has profound applications in the field of medical image analysis, where, for example, a deep learning classifier can help physicians classify cell types or cancer stages. It is one of the earlier applications of deep learning on microscopy imaging data, since training data are usually abundant and of low cost. In contrast, image reconstruction in microscopy data gained popularity only recently, when researchers realized deep learning techniques are able to map an image that has a low signal-to-noise ratio to an image of high signal-to-noise ratio by finding the underlying mapping functions that are too complicated to do mathematically[21, 29]. Image reconstruction is a method that is thought to computationally address the fundamental tradeoffs between imaging speed, image resolution and signal-to-noise ratio for any imaging modality. However, there are still concerns in the field, one of which is whether image reconstruction through the means of deep learning may produce any unexpected artifacts. Researchers must take extra caution on extracting data from the reconstructed images, and it is possible that certain intensity-based measurements can be biased or even error-prone.

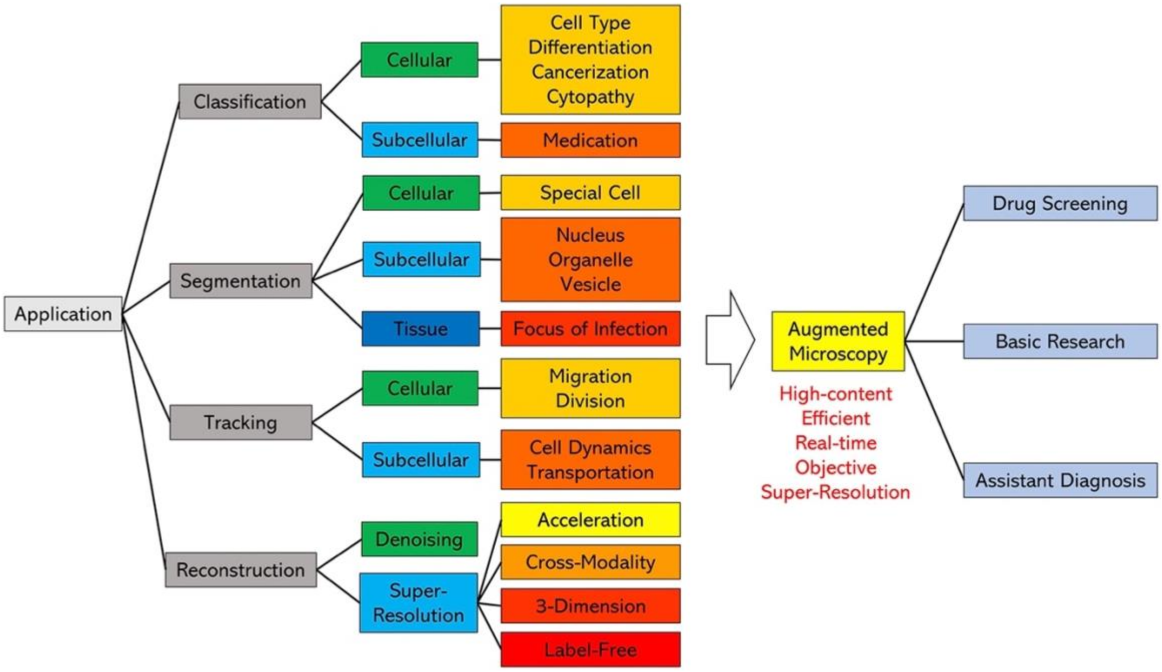


Figure 1.3.1. An overview of applications of deep learning in microscopy image analysis. Figure reproduced from Liu et al. 2021[28].

Image segmentation and tracking, the major topics of this thesis, have been widely investigated on different biological tissues and structures. Image segmentation is a method of dividing an image into subgroups or classes of image segments, typically of desired research target. Instance image segmentation additionally divides a subgroup or class into an individual object if there is more than one object within that subgroup. Instance segmentation requires not only the determination of the classes for each pixel in an image but also accurate determination of the boundaries of each object. Image segmentation can aid with image visualization; for instance, it is useful for automatically labeling cell compartmentalization and different tissues in an image. More importantly, the segmented image then provides single-object measurements that are crucial for biological investigations and statistical testing. Therefore, image segmentation is a crucial step for extracting meaningful biological information from a microscopy image. Additionally, tracking (or multi-object/multi-cell tracking) is a task that requires the detected object to be linked in time. It is especially useful to study single-cell motions and behaviors in a dynamic environment. In a biofilm study, a successful tracking experiment can provide time-dependent single-cell observables, which may shed light on cell fate or mechanisms of antibiotic resistance, and thus offer invaluable insights for the study of biofilm life cycle and emergent behaviors.

In conclusion, in this thesis, I present instrumentation as well as computational solutions for analyzing individual bacterial cells within bacterial biofilms using image-based assays and analyses. My contributions described here will open new avenues to investigate biofilm heterogeneity and collective behaviors.

Chapter 2: Bacterial biofilm imaging and microscopy

2.1 Lattice light sheet microscopy (LLSM) for 4D imaging

Fluorescence microscopy provides the best imaging capability for dynamic 3D live biological samples[30]. However, conventional wide-field and confocal fluorescence microscopy that are used routinely in most laboratories are not optimized for 4D (3D + time) long-term, high-temporal resolution imaging. A confocal microscope effectively scans the excitation beam and illuminates a section of the sample, but only the in-focus signals can be collected, and the out-of-focus signals will be blocked and thus will not contribute to image signals (Figure 2.1.1 Left Panel). Such a configuration results in less efficient use of the photon budget, and therefore further leads to photobleaching of the fluorophores and even phototoxicity to the cells over time. In Chapter 3 and 4, I will demonstrate quantitatively how signal-to-background ratio and imaging frequency have a profound impact on the capabilities for automatic image segmentation and tracking. In contrast, lattice light sheet microscopy illuminates a plane of the sample by using a thin light sheet[26]. All of the fluorescence signals are collected by another objective lens that is conjugated to the sample plane (**Figure 2.1.1** Right Panel). This approach not only increases the speed of the microscope, but also significantly reduces the photobleaching effects on fluorophores. For experimentally testing and comparing the photobleaching effect, a sample bacterial biofilm is imaged under both the LLSM and a spinning disk confocal microscope where similar image signal-to-noise ratio and quality were obtained. The normalized intensity of the image is plotted against the number of Z-stacks, and it is shown that the photobleaching effect, measured by the exponential decay constants after curve fitting, is approximately an order of magnitude less for the lattice light sheet microscope compared to a spinning disk confocal microscope (**Figure 2.1.2**).

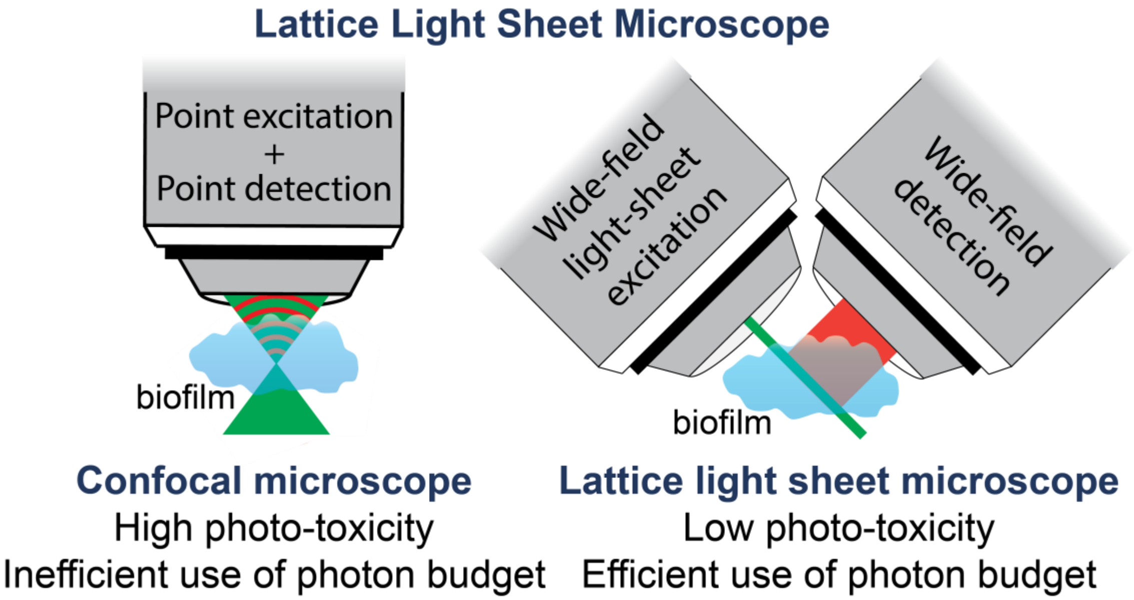


Figure 2.1.1. A schematic of how a lattice light sheet microscope compares to a confocal microscope for bacterial biofilm imaging.

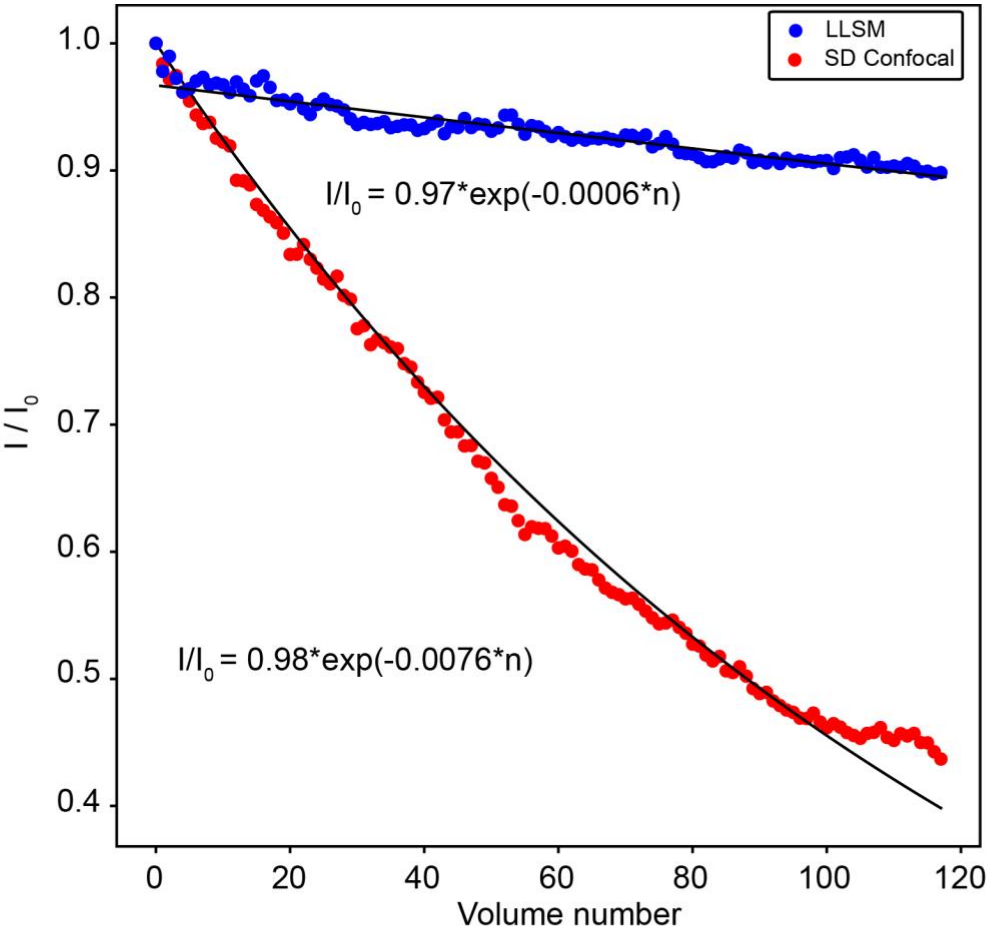


Figure 2.1.2. Quantitative comparison of photobleaching for LLSM and spinning disk confocal microscopy. Image volumes with similar voxel resolution (~100 nm) and initial signal-to-background ratios were recorded using GFP expressing *Shewanella oneidensis* MR-1 biofilms. Fluorescence intensity (normalized to the intensity of the first image volume I0) decreases as a function of acquired image volume. Data fitting using single-exponential decay functions show an order of magnitude decrease in the photobleaching rate for LLSM compared to confocal microscopy. Figure reproduced from Zhang et al. 2021[12].

Lattice light sheet microscopy has noticeable advantages over other fluorescence microscopes on 4D live-cell imaging over a long period of time, especially for imaging light-sensitive species (**Figure 2.1.3a**) and capturing highly dynamic cellular events (**Figure 2.1.3b**; **Figure 2.1.3c** for multi-channel imaging). LLSM makes imaging bacterial biofilm of different species possible, and it opens the door for more sophisticated image analysis as well as quantitative analysis of cells, which has not been previously achieved and reported in the literature on the single-cell level.

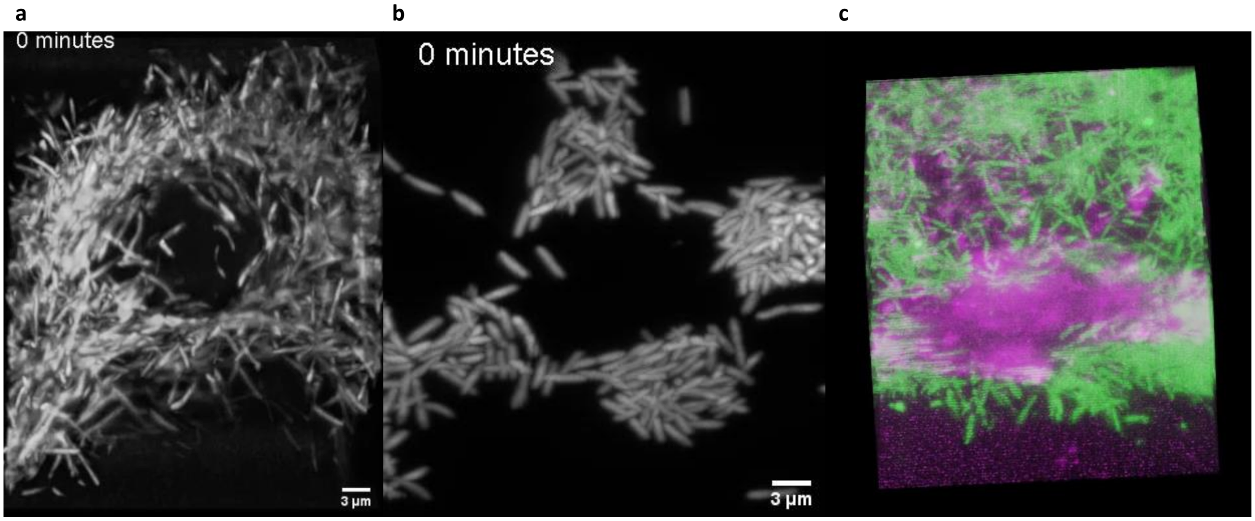


Figure 2.1.3. Example 4D biofilm images acquired by the lattice light sheet microscope. (a) tdTomato expressing *Myxococcus xanthus*. **(b)** GFP expressing *Shewanella oneidensis* **(c)** *M. xanthus* labeled with tdTomato (green) and the extracellular matrix labeled with Concanavalin A conjugated with Alexa fluor 647 (magenta).

2.2 LLSM-compatible microfluidic flow channel assembly

A microfluidic imaging system plays a crucial role in imaging bacterial biofilms due to its importance in facilitating precise and controlled experimentation, analysis, and observation at the microscale. This technology enables researchers to manipulate and image biofilms in a flow system that mimics a physiological environment. The Gahlmann lab previously developed a microfluidic flow channel that can be used for imaging live bacterial biofilms under the lattice light sheet microscope (the previous version of the channel assembly was published and provided in Zhang et al. 2021[12]). However, there are two major problems associated with the way the channel was previously assembled. The first is that the inlet and outlet were connected with PVC tubing sealed by heat shrinkable tubing on the junctions. However, the heat shrinkable tubing was observed to weaken over time allowing air bubbles to come in. In addition, they tend to fall off over time for long time-lapse imaging experiments. The second problem is that the LOCA-133 adhesive (My Polymers, Israel) tends to weaken over time in the heated basin, resulting in leakage from the channel. To tackle these problems, a slightly updated way to assemble the flow channel is described here. A thin layer of silicone is applied to the film gap at the bottom of the channel. This is observed to improve success rate to seal the channel and consistently maintain sealing of the channel in a heated basin for several days. Silicone tubing is then directly attached to the barbed inlet and outlet, which secures the junctions tightly and prevent air bubbles from coming in (**Figure 2.2.1**).

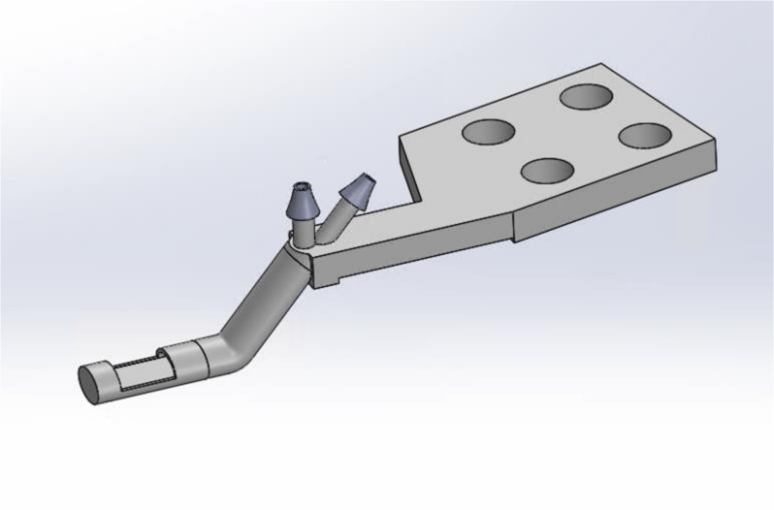


Figure 2.2.1. 3D model of 3D-printed microfluidic channel with improved barbed design of the inlet and outlet.

The channel inoculation protocol requires updating due to the new design of the channel and the use of tubing of larger diameters. Previously, in order to inoculate the channel with live bacterial cell cultures, a syringe was used to push cell culture through the tubing and the flow channel. Then, the flow rate was ramped up to 0.5 mL/hour for 20 minutes to flush away nonadherent cells. Next, the flow rate was reduced to 0.03 mL/hour for the duration of imaging. For a flow experiment, it is important to control the shear stress on bacterial biofilms. The shear stress τ can be calculated as following in a circular tubing:

$$\tau = \frac{4\eta Q}{r^3\pi}$$

In the equation, Q is the flow rate, η the dynamic viscosity and r the radial distance from the centerline of the channel. When the tubing size is within an order of magnitude to the size of the flow channel, it can be assumed that the adherent cells can also form bacterial biofilm on the tubing wall upstream to the channel, which makes biofilm imaging experiments difficult to control. It was hypothesized that bacterial biofilms on the tubing wall upstream to the channel reduced nutrient concentration and added cellular waste to the downstream medium. In practice, however, it is difficult to control the biofilm formation on tubing walls by optimizing shear stress, so it is favorable to bypass the upstream tubing when inoculating. As a result, in the updated version of the protocol, a needle is used to pierce the wall of the tubing right at the inlet, and the cell culture is directly injected into the inlet of the flow channel. This protocol guarantees that only pristine medium is supplied through the flow channel. It is observed that more robust and consistent biofilms were formed using the described protocol.

After inoculation, the channel was mounted on a piezo nanopositioning stage (Mad City Laboratories, NanoOP100HS) and immersed in the basin medium, where sucrose solution was used to match the refractive index of the growth media. The channel inlet port was connected to a syringe pump (Harvard Apparatus, Model 22), and the channel outlet port was connected to a waste container by using silicone tubing.

2.3 LLSM image acquisition and preprocessing

In order to image the microfluidic channel for stable, time-lapse image acquisition described above, there are additional challenges. The first challenge is that the sucrose solution in the basin evaporates, and thereby its refractive index is altered over time. This effect is less pronounced for room-temperature experiments, but poses a major problem for long-term image acquisition under high temperature, such as 30 °C or 37 °C commonly used for imaging bacterial biofilms. We previously added water to the basin at a constant rate to offset the evaporation, but this method proved to be inaccurate because the evaporation rate depends on many factors, such as the desired temperature in the basin, room temperature and humidity. Instead, I used a sucrose circulation system to always supply excessive amount of sucrose solution to the basin. A large amount of sucrose is premade in a tank and gets pumped into the basin at a constant rate driven by controlled air pressure. The basin inevitably overflows, and the overflowed liquid will be removed from a vacuum system built-in to the LLSM basin. Before the fresh sucrose solution reaches the basin, it will pass through a long heating pipe heated by the same hot water that is used to heat the basin. This guarantees that the fresh sucrose is heated at the desired temperature before it reaches the basin, and thus will not change the temperature of the microfluidic channel.

Another challenge is that we have observed sample drift for imaging 3D printed microfluidic channels. Sample drift not only moves the sample to the side of the field of view and ultimately moves the sample out of the field of view, but also blurs the image data since the sample is illumined by a less ideal and thicker part of the light sheet. The blurring effect results in poor or impossible downstream analyses such as single-cell segmentation. The drift is also problematic for automatic image acquisition and eventually cell tracking.

In order to alleviate this problem, I have adopted several strategies. I first glued a fluorescent bead pad to a 3D printed channel made from Clear Resin (Formlabs), shown in Zhang et al. 2021[12]. I estimated the drift by plotting the position of piezo micropositioning stage in z corresponding to the center position in the field of view of an arbitrarily chosen fluorescent bead against time. I fit the data with a linear equation, and the linear coefficient is the sample drift. With Clear Resin, the sample drift is estimated to be -0.14 µm/minute, or -8.4 µm/hour (**Figure 2.3.1a**). Compared to the size of the field of view, which is in the range of 40 µm to 60 µm, this is a significant amount of drift. In comparison, the original metal sample arm offers approximately an order of magnitude more stability, estimated to be -0.018 µm/minute (**Figure 2.3.1b**). Based on the direction of the drift and the specification of the resin material, I hypothesized that gravity pulls the channel down over time. Therefore, I used a machined metal scaffold that can be attached to the bottom of the channel giving extra support. However, this doesn't eliminate the drift (**Figure 2.3.1b**). I therefore tried a different material, Rigid 10k (Formlabs) to make the channel that is more rigid and stable. However, switching to a more rigid resin material is not observed to dampen the drift either (**Figure 2.3.1d**). A possible explanation of the phenomena is that all resin materials from Formlabs absorb a small amount of water and get swollen over time in the basin. By the time the thesis is completed, I haven't found any commercially available material that provides both great 3D printability and stability similar to stainless steel.

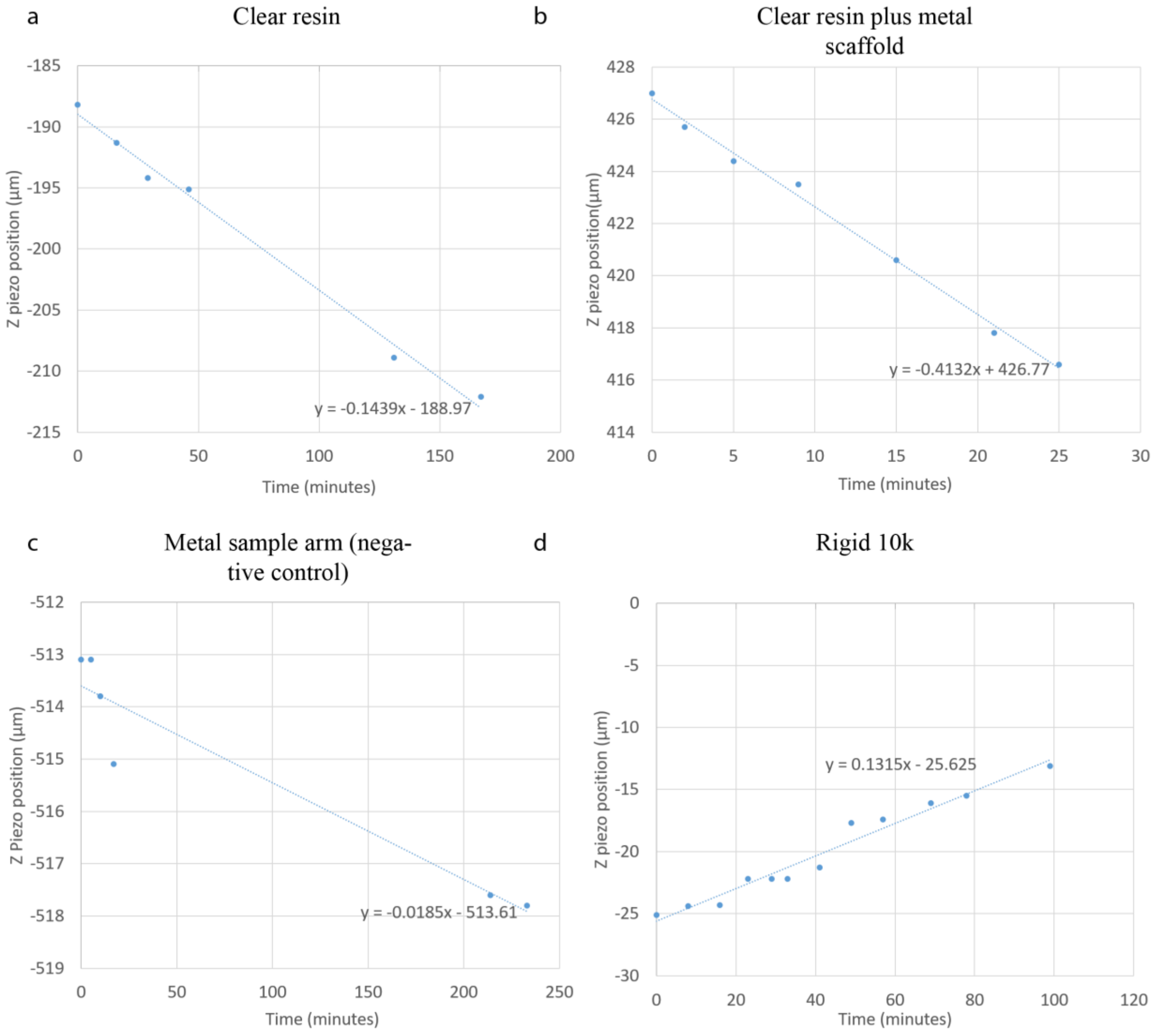


Figure 2.3.1. LLSM sample drift estimated by manually tracking fluorescent beads. (a) A fluorescent bead pad is glued to a 3D printed channel made from Clear Resin (Formlabs). The position of piezo micropositioning stage in z corresponding to the center position in the field of view of an arbitrarily chosen fluorescent bead is plotted against time. The dashed line represents the linear fit of the data, and the drift is estimated to be -0.14 µm/minute. **(b)** The 3D printed channel made from Clear Resin (Formlabs) is secured by attaching a metal scaffold on the bottom. The drift is estimated to be -0.41 µm/minute. **(c)** A fluorescent bead pad is glued to the original metal sample holder. The drift is estimated to be -0.019 µm/minute. **(d)** The 3D printed channel is made from Rigid 10k (Formlabs). The drift is estimated to be 0.13 µm/minute.

I therefore decided to develop a software solution to address the sample drift problem. Sample drift is a common problem for various microscopy techniques, and there are already software solutions available or built-in for commercial microscopes. However, for LLSM, no drift correction functionality is built in to the controlling software.

The pseudocode for real-time drift correction for LLSM is shown in **Figure 2.3.2**. The algorithm assumes that samples between frames move reasonably small, because the core of the algorithm is to use the cross-correlation functions to find the drift comparing two biofilm images alone. This is typically the case for most bacterial biofilms imaged.

Algorithm 1 Real-time specimen drift correction

```
Acquire a Zstack, $img0$
$baseline = img0$
for $n = 1, 2, \ldots$ do
    Acquire a Zstack, $imgN$
    if $n$ modulo by $updateBaseline$ is 0 then
        $baseline = imgN$
    end if
    Fourier transform $img0$ to $img0FT$
    Fourier transform $imgN$ to $imgNFT$
    Find the vector $[driftx, drifty, driftz]$ in unit of pixels that maximizes the cross correlation function between $img0FT$ and $imgNFT$
    Multiply $[driftx, drifty, driftz]$ by the microscope pixel sizes $[px, py, pz]$ to $[driftx, drifty, driftz]$ in metric units
    Transform $[driftx, drifty, driftz]$ in the microscope coordinates to $[driftxStage, driftyStage, driftzStage]$ in the stage coordinates
    Move the stage by $[-driftxStage, -driftyStage, -driftzStage]$
end for
```

Figure 2.3.2. Pseudocode for real-time drift correction for a Lattice light sheet microscope.

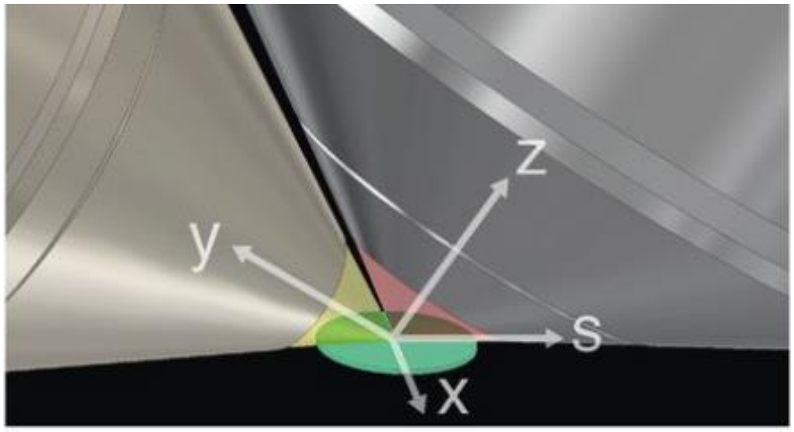


Figure 2.3.3. The x, y, and z directions are indicated for the microscope coordinates. The s-axis defines the direction the specimen moves from image plane to image plane. Figure reproduced from Chen et al. 2014[26].

The algorithm takes a baseline image, *img0*, and when a new image is taken, the two images are Fourier transformed, and then the cross-correlation function is used to find how much deviation there is to the new image from the baseline. A vector in pixel space is obtained and later transformed into metric units by multiplying the pixel sizes of the microscope. The drift vector is in the microscope coordinate, which is at an angle, θ, from the piezo micropositioning stage (**Figure 2.3.3**, also refer to the configuration in **Figure 2.1.1**). The drift in all three directions needs to be transformed using trigonometry as following:

$$driftY_{stage} \ = driftX_{microscpe}$$

$$driftZ_{stage} = driftZ_{microscpe} \times \cos(\theta) - \ driftY_{microscpe} \times \sin(\theta)$$

$$driftX_{stage} = driftY_{microscpe} \times \cos(\theta) + driftZ \times \sin(\theta)$$

Finally, the drift vectors will be communicated to the LLSM controlling software and used to direct the movement of the piezo micropositioning stage. I thank Andy Chiu from Sciotex here for implementing the LabVIEW interface that talks to the piezo micropositioning stage.

An image preprocessing pipeline for generating time-lapse 4D images is carried out in MATLAB and Fiji[31]. Raw 3D images are first background subtracted, then deskewed and deconvolved[26, 32]. The background subtraction is optional and is estimated by averaging intensity values of dark areas (devoid of cells) in the field of view. Deconvolution is performed using the Richardson-Lucy algorithm with 10 iterations using experimentally measured point spread functions (PSFs) as the deconvolution kernel. Deconvolution is only done for BCM3D 1.0, and not in BCM3D 2.0. The experimentally measured PSFs are obtained separately for each color channel using fluorescent beads (200 nm FluoSpheres®, Thermo Fisher) coated on a coverslip. A time sequence of 3D images is opened in Fiji as a hyperstack. A bounding box of a region of interest is manually selected, and then *batch_crop.ijm* is used to crop the bounding box in Fiji. Cropped individual 3D images are then saved to a specified folder by running hyperstack2se.ijm in Fiji. Fiji script *Correct_3D_drift.py* is then used to do additional steps of post-drift correction, since the piezo-micropositioner described previously has limited precision[33]. A rectangle ROI is manually selected and then used to compute the drift for better performance. Two separate rounds of post-drift correction might be needed on only *xy* and only *z* for optimal results. 3D images are rendered using the 3D Viewer plugin in Fiji or ChimeraX[34]. *renameForChimeraX.m* is used to rename default file names to be compatible with ChimeraX nomenclature. For rendering large 4D images, the data needs to be converted to a Chimera map (cmap) format. The rendered movie can therefore be displayed and saved using proper ChimeraX commands[34].

Chapter 3: 4D bacterial biofilm image segmentation using deep learning

3.1 Overview

Bacterial biofilm image segmentation refers to the process of separating the individual bacterial cells or clusters from the background and from each other within the biofilm image. It involves identifying and delineating the boundaries of the biofilm regions, enabling further analysis of the biofilm structure, composition, and growth patterns. 3D biofilm image segmentation was not fully solved in the literature given densely packed cells as well as small intercellular gaps between cells.

In this chapter, I present two different avenues for 3D biofilm image segmentation using deep neural networks trained from simulated data. We used physical image representations as well as their corresponding ground truth as training pairs for the *BCM3D 1.0* workflow (**Figure 3.1.1 left panel**). In this approach, we modeled physical image representations as closely to how real images are acquired by a lattice light sheet microscope, where representative emitter distributions, image noise and microscope point spread function (PSF) were carefully modeled (please see **Section 3.2.5** for more details). In contrast, *BCM3D 2.0* pipeline relies on both physical image representations and designed non-physical image representations as the training data, the purpose of which is to enable high accuracy of image segmentation using conventional image processing techniques further downstream (**Figure 3.1.1 right and bottom panel**). This approach is particularly effective when cell density is high and when image signal-to-background is low. Details of the two pathways as well as their performance comparisons will be discussed in the following sections.

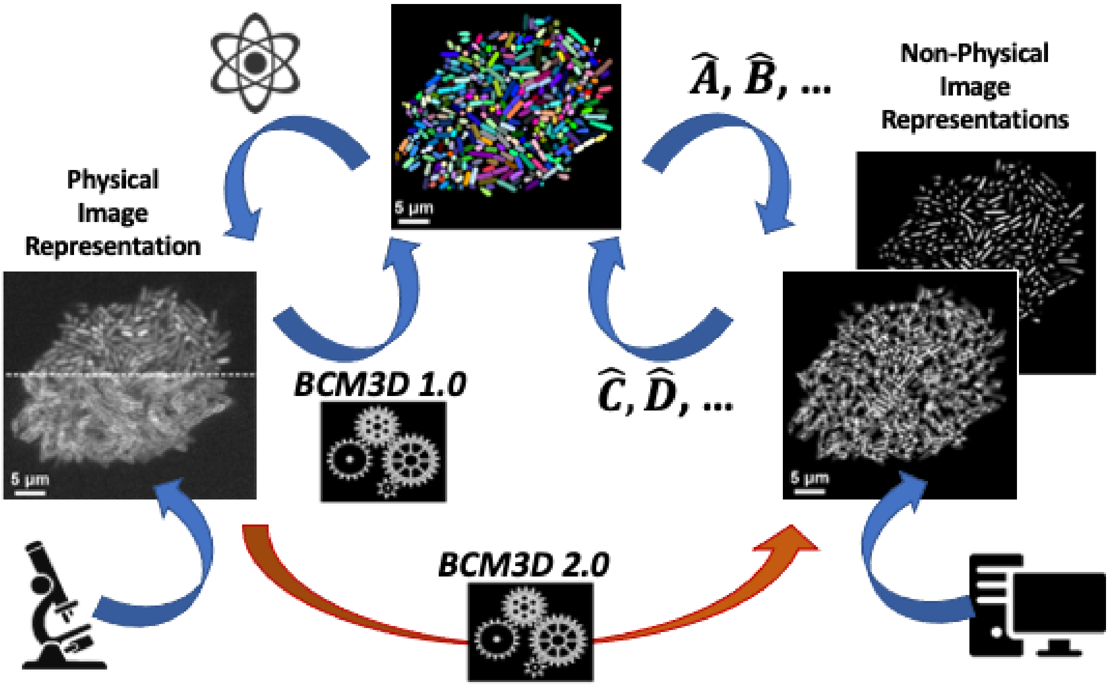


Figure 3.1.1. A schematic of how two versions of *BCM3D* (see details in the following sections) compare on 3D bacterial biofilm image segmentation.

3.2 BCM3D 1.0: pixel-wise classification using deep neural networks

3.2.1 Abstract

Fluorescence microscopy enables spatial and temporal measurements of live cells and cellular communities. However, this potential has not yet been fully realized for investigations of individual cell behaviors and phenotypic changes in dense, three-dimensional (3D) bacterial biofilms. Accurate cell detection and cellular shape measurement in densely packed biofilms are challenging because of the limited resolution and low signal to background ratios (SBRs) in fluorescence microscopy images. In this work, we present Bacterial Cell Morphometry 3D (*BCM3D*), an image analysis workflow that combines deep learning with mathematical image analysis to accurately segment and classify single bacterial cells in 3D fluorescence images. In

BCM3D, deep convolutional neural networks (CNNs) are trained using simulated biofilm images with experimentally realistic SBRs, cell densities, labeling methods, and cell shapes. We systematically evaluate the segmentation accuracy of *BCM3D* using both simulated and experimental images. Compared to state-of-the-art bacterial cell segmentation approaches, *BCM3D* consistently achieves higher segmentation accuracy and further enables automated morphometric cell classifications in multi-population biofilms.

3.2.2 Introduction

Biofilms are multicellular communities of microorganisms that grow on biotic or abiotic surfaces[35-38]. In addition to cellular biomass, biofilms also contain an extracellular matrix (ECM) which is composed of polysaccharides, DNA, and proteins. Individual cells in biofilms interact with other cells, the ECM, or with the substrate surface, and the sum total of these interactions provide bacterial biofilms with emergent functional capabilities beyond those of individual cells. For example, biofilms are orders of magnitude more tolerant towards physical, chemical, and biological stressors, including antibiotic treatments and immune system clearance[4, 35, 36, 39-41]. Understanding how such capabilities emerge from the coordination of individual cell behaviors requires imaging technologies capable of resolving and simultaneous tracking of individual bacterial cells in 3D biofilms.

Live cell-compatible imaging technologies, such as optical microscopy, can reveal the spatial and temporal context that affects cellular behaviors. However, conventional imaging modalities are not able to resolve individual cells within thick 3D biofilms over extended periods of time. For example, the diffraction-limited lateral *x,y*-resolution (~230 nm) of a confocal fluorescence microscope is barely sufficient to resolve bacterial cells positioned next to each other on flat glass coverslips. Even worse, the diffraction-limited axial *z*-resolution (570 nm) is comparable to the size of a single bacterial cell, so that densely-packed cells become unresolvable in the axial *z*-dimension[42, 43]. Notable exceptions include loose biofilms (low cell density), spherical cell shapes[44, 45], and mutant *Vibrio cholera* biofilms, in which cell-cell spacing is increased through the overproduction of ECM materials[1, 8, 9]. While single-cell resolved images have been obtained in such special situations, conventional optical microscopy modalities are not generally capable to accurately resolve and quantitatively track individual cells in dense 3D biofilms.

While super-resolution derivatives of confocal microscopy, known as Image Scanning Microscopy[46], can improve spatial resolution, a perhaps more important limitation for long-term live-cell imaging is photodamage to the specimen (phototoxicity) and to the fluorophores used for labeling (photobleaching)[32, 47, 48]. In confocal microscopy-based approaches, undesired out-of-focus fluorescence emission is filtered out by confocal pinholes to yield optically-sectioned images with high contrast, i.e. high signal-to-background ratios (SBRs). However, repeated illumination of out-of-focus regions during laser scanning and high light intensities at the focal volume result in rapid photobleaching of fluorophores and unacceptable phototoxicity for light sensitive specimens[32, 47, 48]. In fact, confocal fluorescence microscopy (as well as its super-resolution derivatives) uses illumination light intensities that are two to three orders of magnitude higher than the light intensities under which life has evolved[48]. The high rates of phototoxicity and photobleaching make confocal-based microscopy unsuitable for high frame-rate time-lapse imaging of living specimens over many hours and days[1, 8, 47, 49, 50].

In recent years, light sheet-based fluorescence excitation and imaging approaches have been developed to overcome the drawbacks of confocal microscopy. Among these, lattice light sheet microscopy (LLSM)[26, 32] and field synthesis variants thereof[51], axially-swept light sheet microscopy (ASLM)[52, 53], dual-view light sheet microscopy[17, 54], and single-objective oblique plane light sheet microscopy[15, 55-59] now combine excellent 3D spatial resolution with fast temporal resolution and low phototoxicity at levels that cannot be matched by confocal microscopy. Specifically, light sheet-based microscopy approaches can operate at illumination intensities that are below the levels of cellular phototoxicity, even for notoriously light sensitive specimens, and reduce fluorophore photobleaching by 20-50 times compared to confocal microscopy, while maintaining comparable spatial resolution and contrast/SBR[26, 56].

An additional challenge in high-resolution biofilm imaging is data quantification. Even if sufficient resolution and high SBRs can be achieved to visually discern, i.e. qualitatively resolve individual cells, robust computational algorithms are still needed for automated cell segmentation and quantitative cell tracking. Towards this goal, image processing approaches based on the watershed technique and intensity thresholding have been developed over the years for single-cell segmentation in bacterial biofilms[1, 8, 9, 50]. The broad applicability of watershed- and threshold-based image processing algorithms is however limited, because these algorithms require manual optimization of many user-selected parameters. Even with optimal parameters, watershed- and threshold-based image processing methods often produce sub-optimal segmentation results, especially when cell densities are high, when SBRs are low, and when cellular fluorescence intensities are not uniform across the cytosol or the cell surface. To overcome the drawbacks of traditional mathematical image processing approaches, automated solutions based on supervised training of deep convolutional neural networks (CNNs) have been used in recent years with great success for a wide range of problems in biomedical image analysis[60].

Here, we present Bacterial Cell Morphometry 3D (*BCM3D*)[48], a generally applicable workflow for single-cell segmentation and shape determination in high-resolution 3D images of bacterial biofilms. *BCM3D* uses CNNs, *in silico*-trained with computationally simulated biofilm images, in combination with mathematical image analysis to achieve accurate single cell segmentation in 3D. The CNNs employed in *BCM3D* are based on the 3D U-Net architecture and training strategy, which has achieved excellent performance in biomedical data analysis benchmark tests[60]. The mathematical image analysis modules of *BCM3D* enable post-processing of the CNN results to further improve the segmentation accuracy. We establish that experimental bacterial biofilms images, acquired by lattice light sheet microscopy, can be successfully segmented using CNNs trained with computationally simulated biofilm images, for which the ground-truth voxel-level annotation maps are known accurately and precisely. By systematically evaluating the performance of *BCM3D* for a range of SBRs, cell densities, and cell shapes, we find that voxel-level segmentation accuracies of >80%, as well as cell counting accuracies of >90%, can be robustly achieved. *BCM3D* consistently outperforms previously reported image segmentation approaches that rely exclusively on conventional image processing approaches. *BCM3D* also achieves higher segmentation accuracy on experimental 3D biofilm data than *Cellpose*[61], a state-of-the-art, CNN-based, generalist algorithm for cell segmentation and the algorithm used by Hartmann *et al.*[8], a specialized algorithm designed for bacterial cell segmentation based on traditional mathematical image processing. We expect that *BCM3D*, and CNN-based single-cell segmentation approaches in general, combined with non-invasive light sheet-based fluorescence microscopy will enable accurate cell tracking over time in dense 3D biofilms. This capability will launch a new era for bacterial biofilm research, in which the emergent properties of microbial populations can be studied in terms of the fully-resolved behavioral phenotypes of individual cells.

3.2.3 Results

Cell segmentation by thresholding CNN confidence maps

CNNs have been shown to perform well on pixel-level classification tasks for both 2D and 3D data[62, 63]. Bacterial biofilms, however, present a unique challenge in this context. The cell shapes to be segmented are densely packed and barely resolvable even with the highest resolution optical microscopes. Additionally, living biofilms in fluorescence microscopes can only be imaged with low laser intensities to ameliorate phototoxicity and photobleaching concerns. Unfortunately, low intensity fluorescence excitation also reduces the SBR in the acquired images. So far, it remains unclear to what extent single-cell segmentation approaches can accurately identify and delineate cell shapes in bacterial biofilm images obtained under low intensity illumination conditions. To address this question, we implemented an *in-silico* CNN training strategy and systematically evaluated its voxel-level classification (cell morphometry) and cell counting accuracies using simulated biofilm images with cell densities and SBRs similar to those encountered in experimental data (see Methods).

We compared two commonly used cell labeling approaches, namely genetic labeling through the expression of cell-internal fluorescent proteins and staining of the cell membranes using fluorescent dyes (**Figure 3.2.1**). For both labeling approaches, voxel-level segmentation and cell counting accuracies, obtained by thresholding CNN confidence maps (see Methods), depend mostly on cell density, whereas the SBR plays a less important role (**Figure 3.2.2a-c** and **3.2.2d-f**). For cell-internal labeling, SBRs of >1.7 and cell densities of <60% consistently produce voxel-level classification accuracies of >80% and cell counting accuracies of >95%. On the other hand, SBRs of <1.7 and cell densities of >60% lead to lower segmentation accuracies. While lower segmentation accuracies are expected for higher cell densities and lower SBRs, the sharp dropoffs observed here may indicate a fundamental performance limitation of the CNNs employed. Still, the voxel-level classification and cell counting accuracies consistently surpass previous approaches for bacterial cell segmentation for commonly encountered cell densities and SBRs. Specifically, the cell counting accuracies obtained by Hartmann *et al.*[8], *Seg3D*[64], and Yan *et al.*[9] quickly drop to zero as a function of increasing IoU matching threshold (a quantitative measure of cell shape similarity relative to the ground truth, see Methods), indicating that cell shapes are not accurately estimated by conventional image processing approaches (**Figure 3.2.2g-i**). We also evaluated the segmentation accuracy of *Cellpose*, a recently developed, CNN-based cellular segmentation algorithm[61]. The segmentation accuracy of *Cellpose* is comparable or superior to the best-performing conventional image processing approaches – a considerable achievement given that *Cellpose* was trained primarily on images of eukaryotic cells. However, being a pre-trained generalist model, the segmentation accuracy of *Cellpose* is lower than the accuracy achieved by the specialist *in silico*-trained CNNs of *BCM3D*, which were trained specifically for 3D bacterial biofilm segmentation. Overall, the cell counting accuracies obtained by *BCM3D* are higher than other methods and remain higher even for IoU matching thresholds larger than 0.5, indicating that cell shapes are more accurately estimated by the *in silico*-trained CNNs.

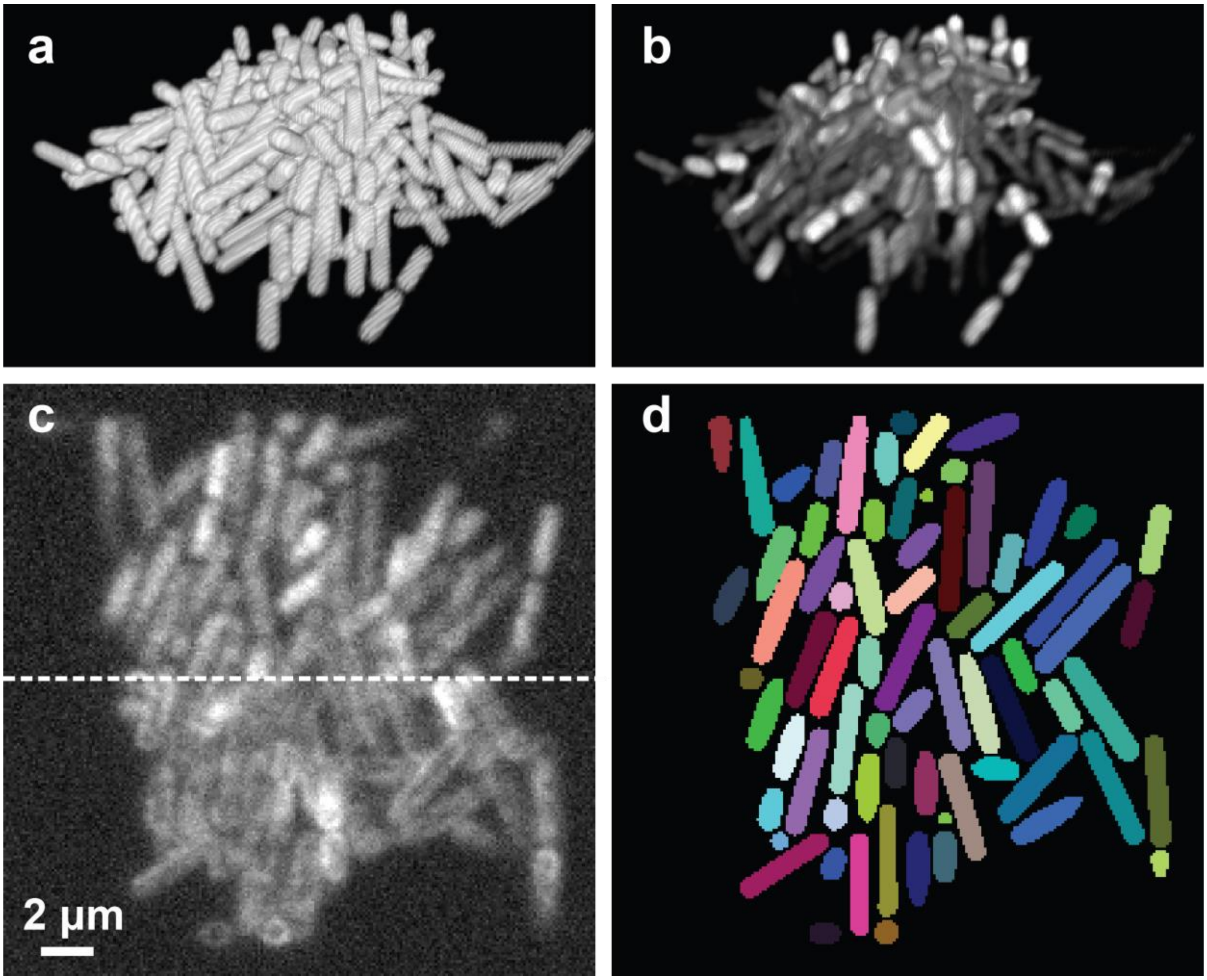


Figure 3.2.1. Simulation of fluorescent biofilms images and annotation maps used for CNN training. (a) Representative cell arrangements obtained by CellModeller. Due to the stochastic nature of biofilm growth, different cell arrangements are obtained in each new simulation. However, cell density is reproducible for each new simulated biofilm (typically $N = 10$ different biofilm simulations are used for CNN training, see Methods) **(b)** Simulated 3D fluorescence image based on the cell arrangements in **a**. **(c)** XY slice through the 3D simulated fluorescence image in **b** (upper panel shows cells expressing cytosolic fluorescent proteins, lower panel shows cells stained with membrane-intercalating dyes). **(d)** Ground truth cell arrangements giving rise to the image shown in **c**. Voxels are displayed as black (background), or in different colors (indicating different cells).

The accuracies of single-cell shape estimation and cell counting are predominantly affected by cell density. The variation is more prominent for membrane-stained cells, because *inter*-cellular fluorescence intensity minima are less pronounced when cell membranes are labeled and cells physically contact each other (red arrow in **Figures 3.2.2c** and **3.2.2f**). By contrast, intracellular fluorophores produce the highest intensities at the cell center, so that the gaps between cells are more readily resolvable. Also noteworthy is the sharp drop-off in segmentation accuracies for SBRs of <1.7 for all cases. In such low SBR regimes, fluorescence signals of the cells become too difficult to be distinguished from the background. As a result, the CNNs falsely identify random noisy patterns in the background as cells. Additionally, thresholding of the CNN confidence maps often yields connected voxel clusters that contain multiple bacterial cells. False identification and incomplete delineation of cells cause the pronounced decrease in segmentation accuracy for SBRs of <1.7.

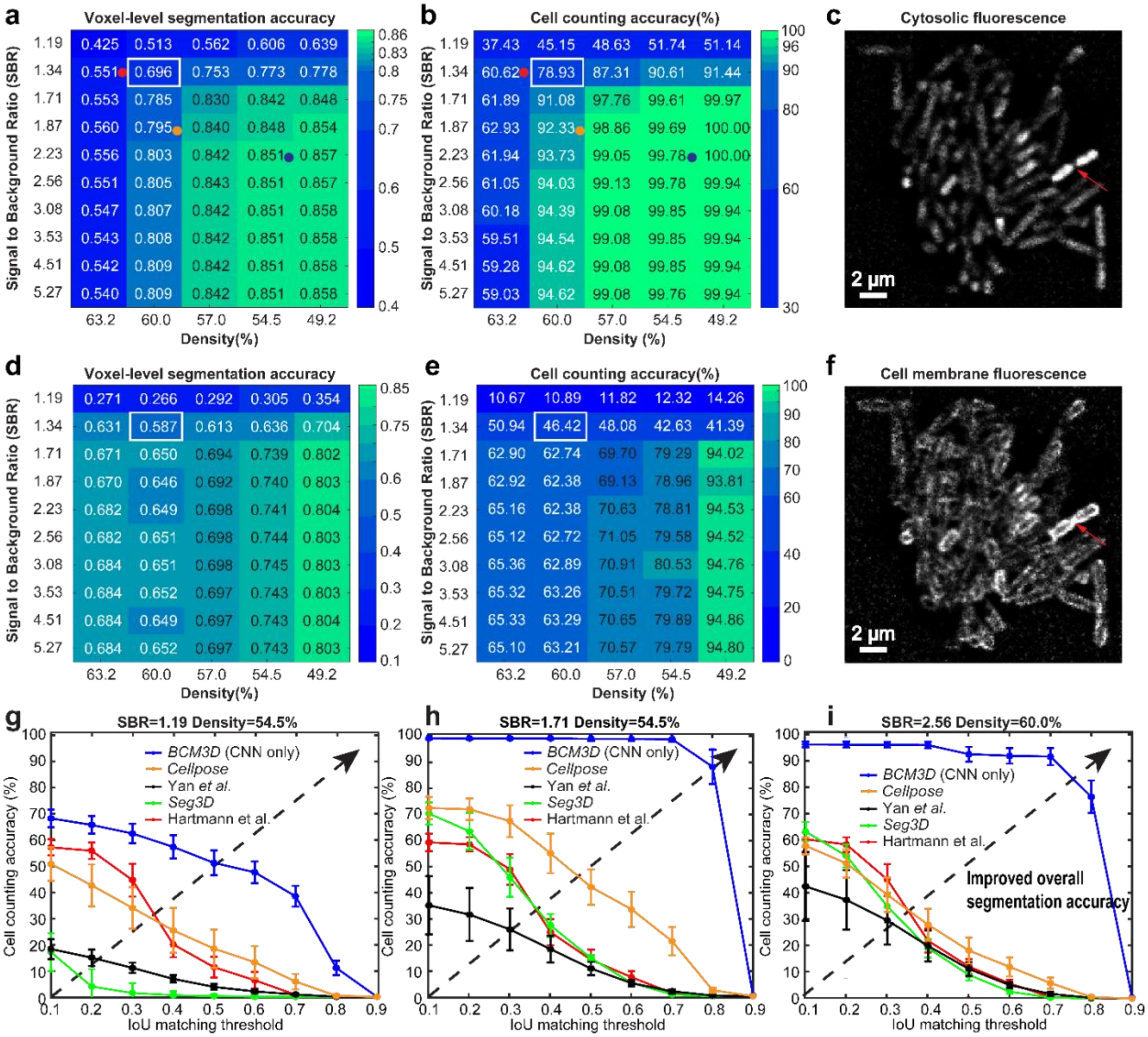


Figure 3.2.2. Performance of *BCM3D* using *in silico*-trained CNNs only on previously unseen simulated biofilm images. (a) The voxel-level segmentation accuracy quantifies whether each voxel has been assigned to the correct class (‘cell interior’, ‘cell boundary’, or ‘background’). Solid circles represent the maximum local density and average SBRs encountered in experimental datasets (red, orange and blue: *E. coli* expressing GFP). **(b)** The cell counting accuracy (using an IoU matching threshold of 0.5 for each segmented object, see Methods) averaged over N = 10 replicate datasets for cells labeled with cytosolic fluorophores. **(c)** Example image of cells labeled with cytosolic fluorophores (Cell density = 60.0%, SBR = 1.34, indicated by white rectangle in panels a and b. Similar images were generated N = 10 times with different cell arrangements). **(d)** Voxel-level segmentation accuracy and **(e)** cell counting accuracy averaged over N = 10 replicate datasets for cells labeled with membrane-localized fluorophores. **(f)** Example image of cells labeled with membrane-localized fluorophores (Cell density = 60.0%, SBR = 1.34, indicated by white rectangles in panels d and e. Similar images were generated N = 10 times with different cell arrangements). The red arrows indicate a close cell-to-cell contact point. **(g), (h) and (i)** Comparison of segmentation accuracies achieved by conventional segmentation approaches

(Hartmann *et al.*, *Seg3D*, Yan *et al.*), *Cellpose*, and *BCM3D* (only using *in silico*-trained CNNs). Three simulated datasets (cytosolic fluorophores) with different SBRs and cell densities are shown. Segmentation accuracy is parameterized in terms of cell counting accuracy (*y* axis) and IoU matching threshold (*x* axis, a measure of cell shape estimation accuracy). Each data point is the average of $N = 10$ independent biofilm images. Data are presented as mean values ± one standard deviation indicated by error bars. Curves approaching the upper right-hand corner indicate higher overall segmentation accuracy, as indicated by the dashed arrows.

Post-processing of CNN confidence maps

To better identify individual cells in low SBR and high cell density datasets, we developed a graph-based post-processing module (see Methods) that takes advantage of the fact that bacterial cell shapes are highly conserved for a given species. Briefly, we transformed the CNN 'cell interior' confidence maps into 3D point cloud data that trace out the central axes of individual cells. This transformation was achieved by medial axis extraction using size-constrained inscribed spheres[65]. Single-cell axes are then identified as linearly clustered data points by *LCuts* – a graph-based data clustering method designed to detect linearly oriented groups of points[66]. The so-identified single-cell axes are then mapped back onto the original segmentation volumes to obtain estimates of the 3D positions, shapes, and orientations of the now separated cells.

Post-processing with *LCuts* takes advantage of *a priori* knowledge about expected bacterial cell sizes by removing erroneously segmented volumes that are significantly smaller than the expected value and by splitting incompletely segmented volumes representing fused cells. Improvements in cell counting accuracy of up to 15% and 36% are observed for cells labeled with cytosolic fluorophores (**Figure 3.2.3a-c**) and membrane-localized fluorophores (**Figure 3.2.3d-f**), respectively. The more substantial improvement for membrane-stained cells is due to fact that CNNs trained on membrane-stained cells are more prone to erroneously identifying speckled background noise as fluorescence signals in low SBR images. In addition, membrane-intercalating fluorophores of two adjacent cells are in close proximity, making it difficult to resolve fluorescence signals from two separate cells due to spatial signal overlap (see the red arrow, **Figure 3.2.2c** and **3.2.2f**). *LCuts* thus provides an important benefit in improving the cell counting accuracy to an extent not achieved by currently available thresholding- or watershed-based post-processing algorithms.

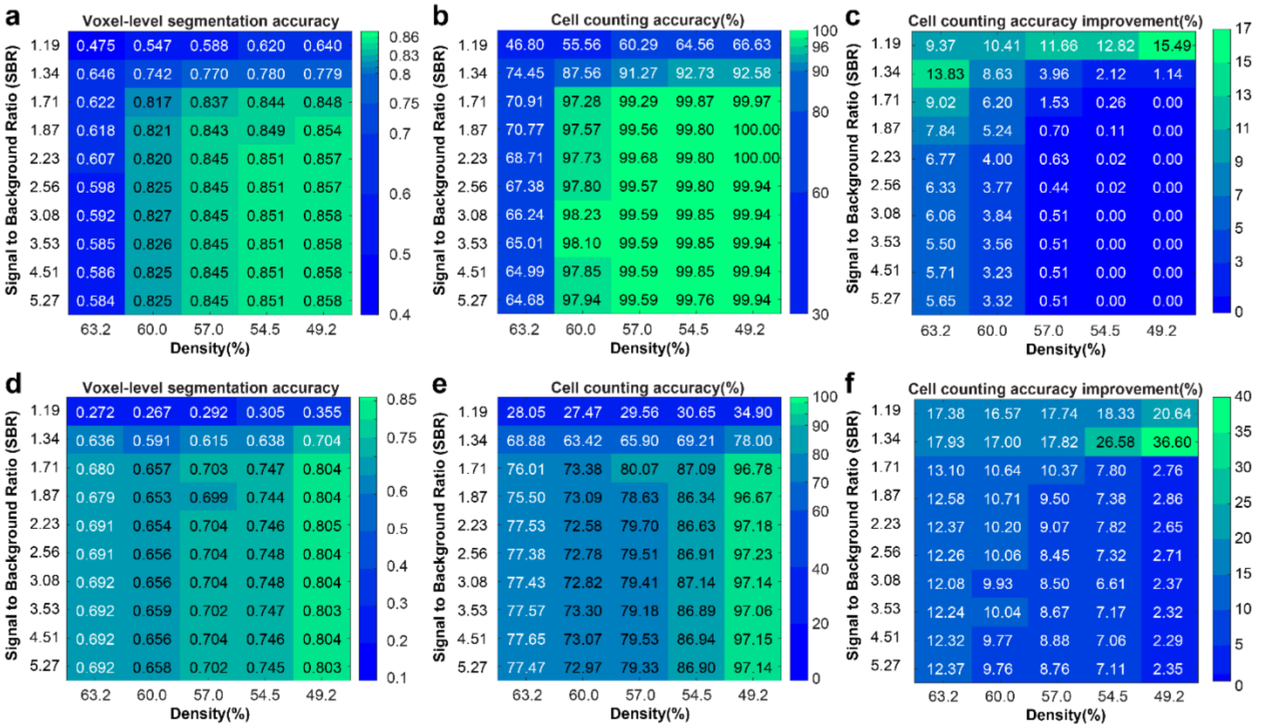


Figure 3.2.3. Performance of *BCM3D* (*in silico*-trained CNNs *and* additional post-processing by *LCuts*) on previously unseen simulated data. (a) Voxel-level segmentation accuracy and **(b)** cell counting accuracy (using a IoU matching threshold of 0.5 for each segmented object) averaged over N = 10 replicate datasets for cells labeled with cytosolic fluorophores. **(c)** Improvement relative to *silico*-trained convolutional neural networks without post-processing. **(d)** Voxel-level segmentation accuracy and **(e)** cell counting accuracy averaged over N = 10 replicate datasets for cells labeled with membrane-localized fluorophores. **(f)** Improvements relative to *silico*-trained convolutional neural networks without post-processing.

Segmentation of experimental biofilm images

To test the performance of *BCM3D* on experimentally acquired biofilm images, we acquired time-lapse images of GFP expressing *E. coli* biofilms every thirty minutes for ten hours (see Methods). We then manually annotated one 2D slice in the 3D images at the t = 300, 360, and 600-minutes time points (see Methods). When referenced to these manual segmentation results, the *LCuts*-processed CNN outputs consistently achieved better cell counting accuracies than conventional segmentation methods (**Figure 3.2.4**). Initially, *Cellpose* and the Hartmann *et al*. algorithm outperform the *in silico*-trained CNNs on two out of three of the test images (*t* = 360 and 600 min), for which our *in silico*-trained CNNs struggle with undersegmentation problems. However, mathematical post-processing of the CNN outputs by *LCuts* corrects some of these errors, so that the integrated *BCM3D* workflow achieves improved results compared to *Cellpose* and Hartmann *et al*. at each of the indicated time points. Visual inspection of the segmentation results is also informative. *Cellpose* accurately segments individual cells in low density regions, but suffers from oversegmentation errors in high density biofilm regions. The Hartmann *et al*. algorithm provides reasonable estimates of cell positions in low- and high-density biofilm regions, but again struggles with cell shape estimation. On the other hand, the integrated *BCM3D* workflow (CNN + *LCuts*) produces biologically reasonable cell shapes regardless of cell density (**Figure 3.2.4**).

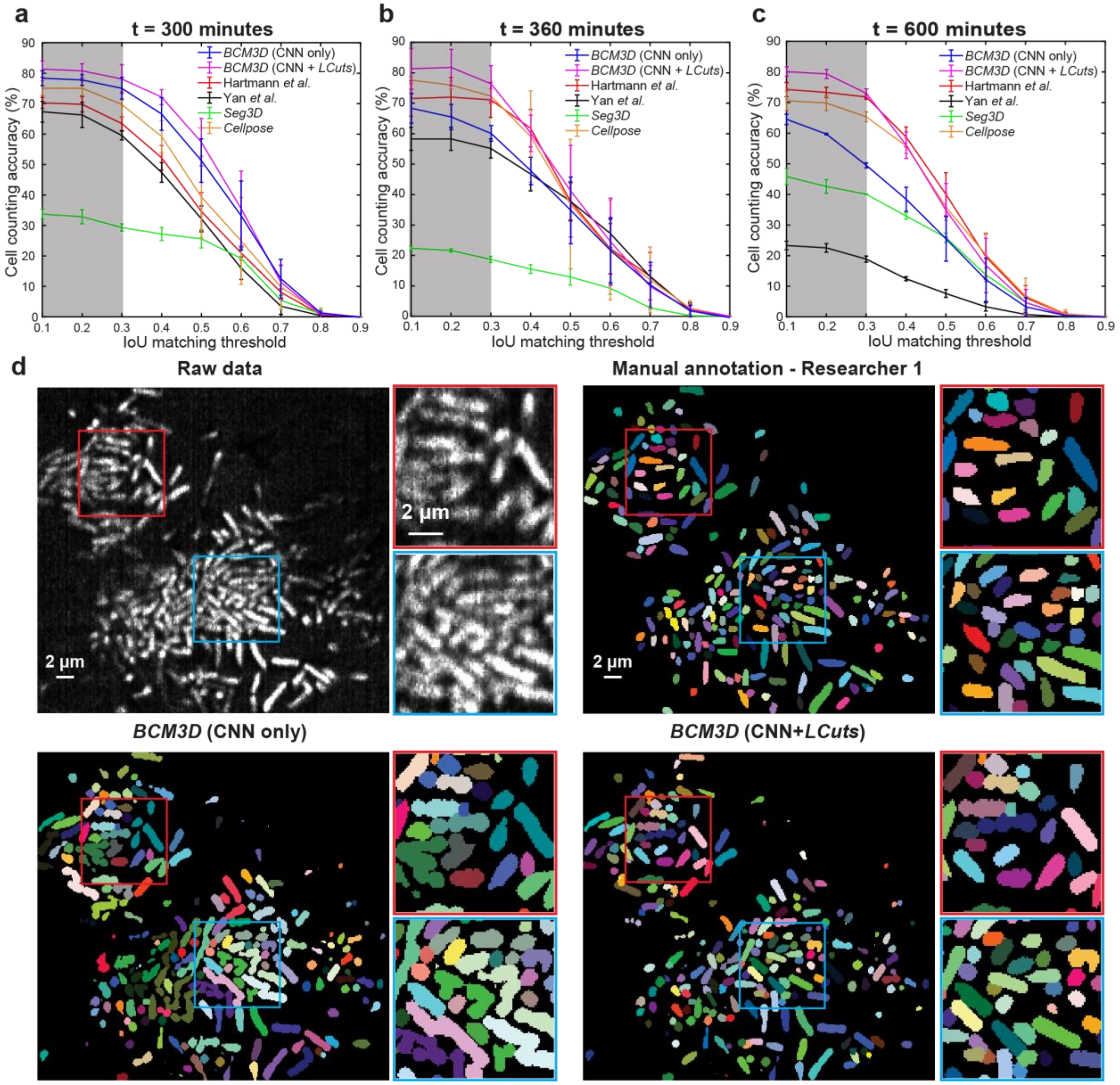


Figure 3.2.4. **Comparison of segmentation accuracies achieved by conventional segmentation approaches (Hartmann *et al.*, *Seg3D*, Yan *et al.*), *Cellpose*, and *BCM3D*.** The estimated SBRs are 2.2, 1.8, and 1.3, respectively. The estimated cell densities are 54.8%, 59.0%, and 64.6%, respectively. **(a-c)** Three experimental *E. coli* datasets (cytosolic expression of GFP) acquired at different time points after inoculation of cells. Segmentation accuracy is parameterized in terms of cell counting accuracy (*y* axis) and IoU matching threshold (*x* axis). Each data point is the average of the cell counting accuracies calculated using annotation maps traced by N = 3 different researchers. Data are presented as mean values ± one standard deviation indicated by error bars. Curves approaching the upper right-hand corner indicate higher overall segmentation accuracy. **(d)** Comparison of segmentation results achieved at the t = 600 minutes time point by manual annotation, and by *BCM3D* using *in silico*-trained CNNs only and after further refinement

of CNN outputs using *LCuts*. Similar results were also obtained at the t = 300 and t = 360 minute time points.

We attribute the more rapid drop-off of the cell counting accuracy as a function of increasing IoU matching threshold in **Figure 3.2.4** to the following factors. First, human annotation of experimentally acquired biofilm images differs from the ground truth segmentation masks that are available for simulated data. The shape mismatches between algorithm segmented and manually annotated cell shapes lead to a global lowering of voxel-level segmentation accuracy and thus a more rapid drop-off of the cell counting accuracy as a function of increasing IoU matching threshold. Because bacterial cell shapes are not accurately captured by manual annotation, cell counting accuracies referenced to manual annotations should be compared only at low IoU matching thresholds (0.1-0.3, shaded grey in **Figure 3.2.4a-c**), as also pointed out previously[67]. We also note that bacterial cells in experimental images appear motion-blurred if they are only partially immobilized and therefore wiggle during image acquisition. Furthermore, optical aberrations and scattering effects were not included in training data simulations, which may decrease the performance of the CNNs on experimental data. Still, at IoU matching threshold < 0.3, the cell counting accuracy of *BCM3D* remains above 75% while also producing biologically reasonable cell shapes. Thus, the bacterial cell segmentation results of *BCM3D* represent a substantial improvement over other approach (Figure 3.2.4).

To demonstrate that *BCM3D* can achieve similarly high segmentation accuracies for membrane-stained cells in different cellular arrangements, we analyzed a small patch of a *M. xanthus* biofilm, which was stained with the membrane intercalating dye FM4-64 (**Figure 3.2.5a**). In contrast to *E. coli* biofilms, the submerged *M. xanthus* biofilm imaged here features cells in a mesh-like arrangement with close cell-to-cell contacts, which presents a unique challenge for 3D single-cell segmentation. To obtain reference data for 3D segmentation accuracy determination, we manually annotated each *xy*, *xz*, and *yz* slice of an entire 3D image stack (**Figure 5b**). When referenced to this 3D manual segmentation result, *BCM3D* (**Figure 3.2.5c**) produced cell counting accuracies above 70% at low (0.1-0.3) IoU matching thresholds, whereas segmentation results obtained by conventional image processing (Hartmann *et al.*) and by generalist CNN-processing (*Cellpose*) produced cell counting accuracies <50% in the same IoU matching threshold region (**Figure 3.2.5d**). We note however that neither *Cellpose* nor the Hartmann *et al.* algorithms were specifically optimized/designed for segmenting membrane-stained cells. Indeed, the performance of *Cellpose* on this type of biofilm architecture is inferior to the results achieved using the *in silico*-trained CNNs of *BCM3D* alone (without using *LCuts* post-processing). One reason might be that the pre-trained, generalist *Cellpose* model has not been trained sufficiently on long, thin, and highly interlaced rod-shaped cells, such as those contained in a *M. xanthus* biofilm.

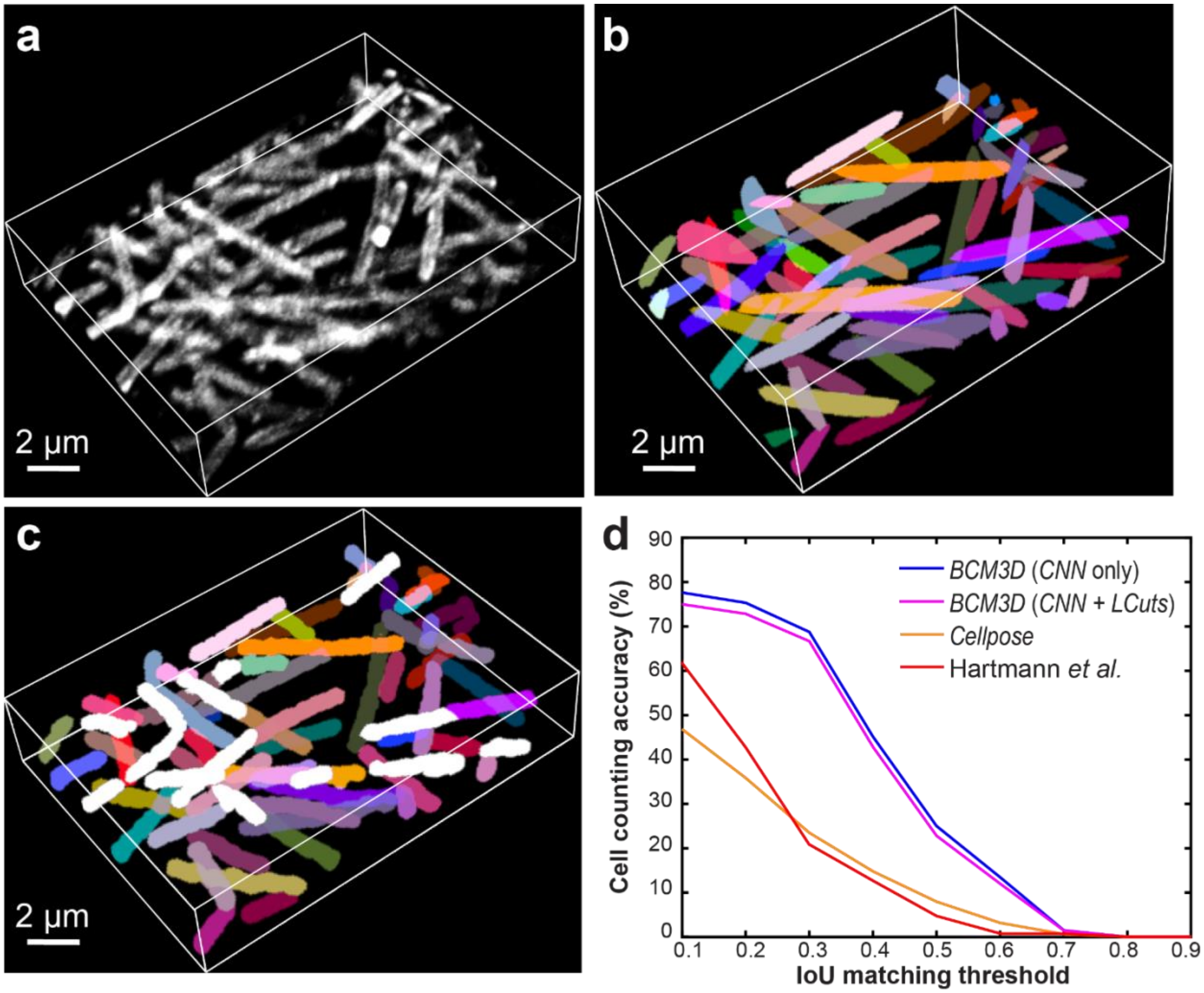


Figure 3.2.5. **3D Segmentation accuracy evaluation using *M. xanthus* biofilm images (cell density = 36.2%, and SBR = 1.58) using *in silico*-trained CNN processing. (a)** Maximum intensity projection of a 3D *M. xanthus* fluorescence image. Cells were labeled with membrane-intercalating dye, FM4-64. Similar images were obtained at *N* = 120 different time points. **(b)** Maximum intensity projection of the manually obtained 3D segmentation result. **(c)** Maximum intensity projection of a CNN-based 3D segmentation result after *LCuts* post-processing. Cells that can be matched with the GT are displayed in the same colors as GT or otherwise colored in white. **(d)** Segmentation accuracy of compared algorithms parameterized in terms of cell counting accuracy (*y* axis) and IoU matching threshold (*x* axis).

Morphological separation of mixed cell populations

Given the improved segmentation results obtained using *BCM3D*, we reasoned that the same CNNs may have additional capacity to assign segmented objects to different cell types based on subtle morphological differences in the acquired images. Differences in the imaged cell morphologies arise due to physical differences in cell shapes (e.g. spherical vs. rod-shaped cells) or due to differences in the fluorescent labeling protocols (e.g. intracellular vs. cell membrane labeling), because fluorescence microscopes simply measure the spatial distributions of fluorophores in the sample. The ability to separate different cell morphologies is important for the study of multispecies biofilms, where interspecies cooperation and competition dictate population-level outcomes[37, 68-75]. Separation of differentially labeled cells is also important for the study of gene activation in response to cell-to-cell signaling[76]. Expression of cytosolic fluorescent proteins by transcriptional reporter strains is a widely-used technique to visualize activation of a specific gene or genetic pathway in living cells. Such genetic labeling approaches can be complemented by chemical labeling approaches, e.g. using membrane intercalating chemical dyes that help visualize cells non-specifically or environmentally-sensitive membrane dyes that provide physiological information, including membrane composition[77, 78], membrane organization and integrity[79-81], and membrane potential[82, 83]. Chemical and genetic labeling approaches are traditionally implemented in two different color channels. However, there are important drawbacks to using multiple colors. First and foremost, the amount of excitation light delivered is increased by the necessity to excite differently colored fluorophores, raising phototoxicity and photobleaching concerns. Second, it takes N times as along to acquire N-color images (unless different color channels can be acquired simultaneously), making it challenging to achieve high temporal sampling in time-lapse acquisition. For these reasons, methods that extract complementary physiological information from a single-color image are preferable.

We evaluated the ability of *BCM3D* to automatically segment and identify rod-shaped and spherical bacterial cells consistent with shapes of *E. coli* and *S. aureus* in simulated images. To segment cells in two-population biofilms, we trained CNNs that classify pixels into five different classes: ‘background’, ‘cell interior of population 1’, ‘cell boundary of population 1’, ‘cell interior of population 2’ and ‘cell boundary of population 2’. Thresholding the CNNs confidence maps can achieve cell counting accuracies larger than 90% for both cell types independent of their population fractions (**Figure 3.2.7a**). Post-processing of this result using *LCuts* improved the cell counting accuracy by less than 0.5% on average, indicating that under-segmented cell clusters are not prevalent in this dataset.

We next evaluated the ability of *BCM3D* to automatically segment and separate membrane-stained cells that express cytosolic fluorescent proteins from those that do not. Again, the cell counting accuracy is consistently above 80% for all tested mixing ratios (**Figure 3.2.7b**). Finally, we applied *BCM3D* to experimentally acquired biofilm images of two different *E. coli* strains. Both strains were stained by the membrane intercalating dye FM4-64, but the second strain additionally expressed GFP (**Figure 3.2.6**). The cells were homogeneously mixed prior to mounting to randomize the spatial distribution of different cell types in the biofilm (see Materials and Methods). Multiple 2D slices from the 3D image stack were manually annotated and compared with the results obtained by *BCM3D*. Consistent with the single-species experimental data, a cell counting accuracy of 50% is achieved for each cell type at a 0.5 IoU matching threshold and, at lower IoU matching thresholds, the counting accuracies increased to 60% to 70%, (**Figure 3.2.7cd**). Thus, using appropriately trained CNNs in *BCM3D* enables automated and accurate cell type assignments based on subtle differences in cell morphologies in mixed population biofilms – a capability not available using conventional image processing methods.

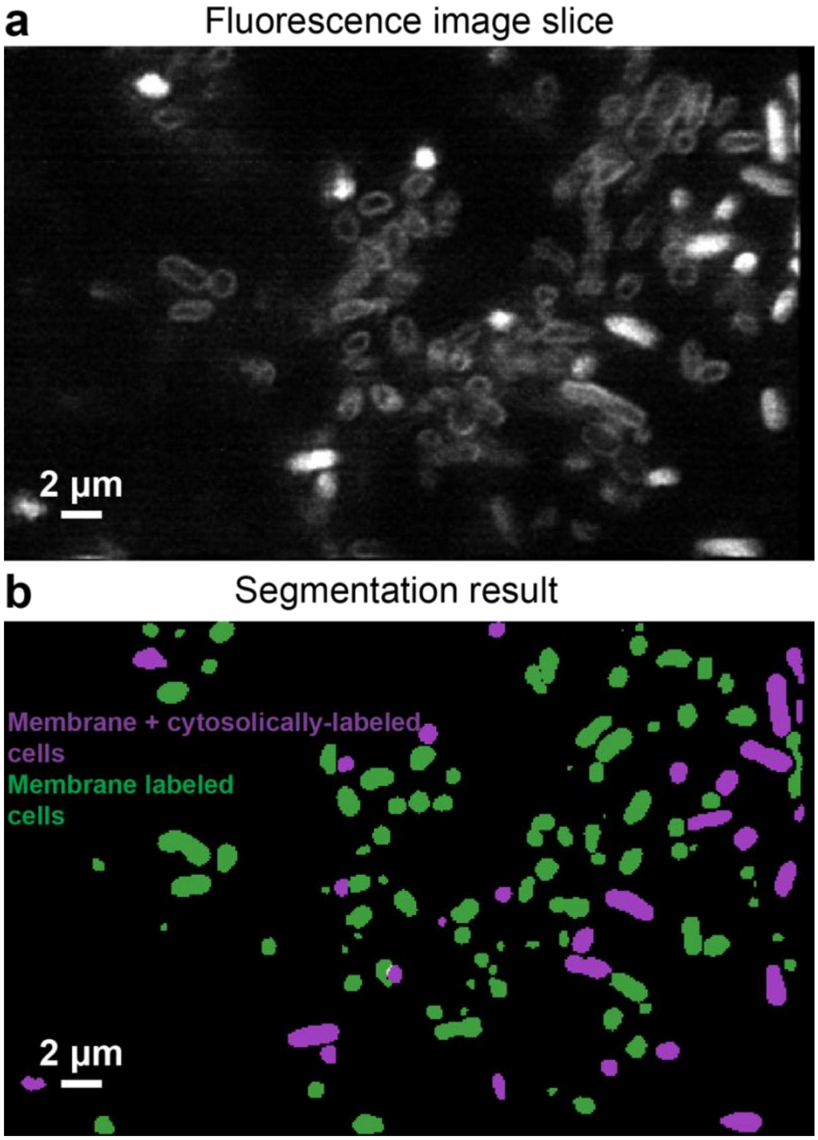


Figure 3.2.6. Segmentation of experimental, mixed-population biofilms containing membrane-stained cells and membrane-stained cells that additionally express an intracellular fluorescent protein. (**a**) Experimental 2D slice of a mixed *E. coli* population containing membrane-stained cells and membrane-stained cells that additionally express an intracellular fluorescent protein. The mixing ratio at the time of inoculation was 50:50. All cells were labeled by the FM4-64 membrane-intercalating dye. **(b)** *BCM3D* segmentation result corresponding to the image shown in (a). Membrane-stained cells are displayed in green, and cells that were both membrane-stained and cytosolically-labeled are displayed in magenta. Similar results were also obtained in $N = 3$ independent experiments.

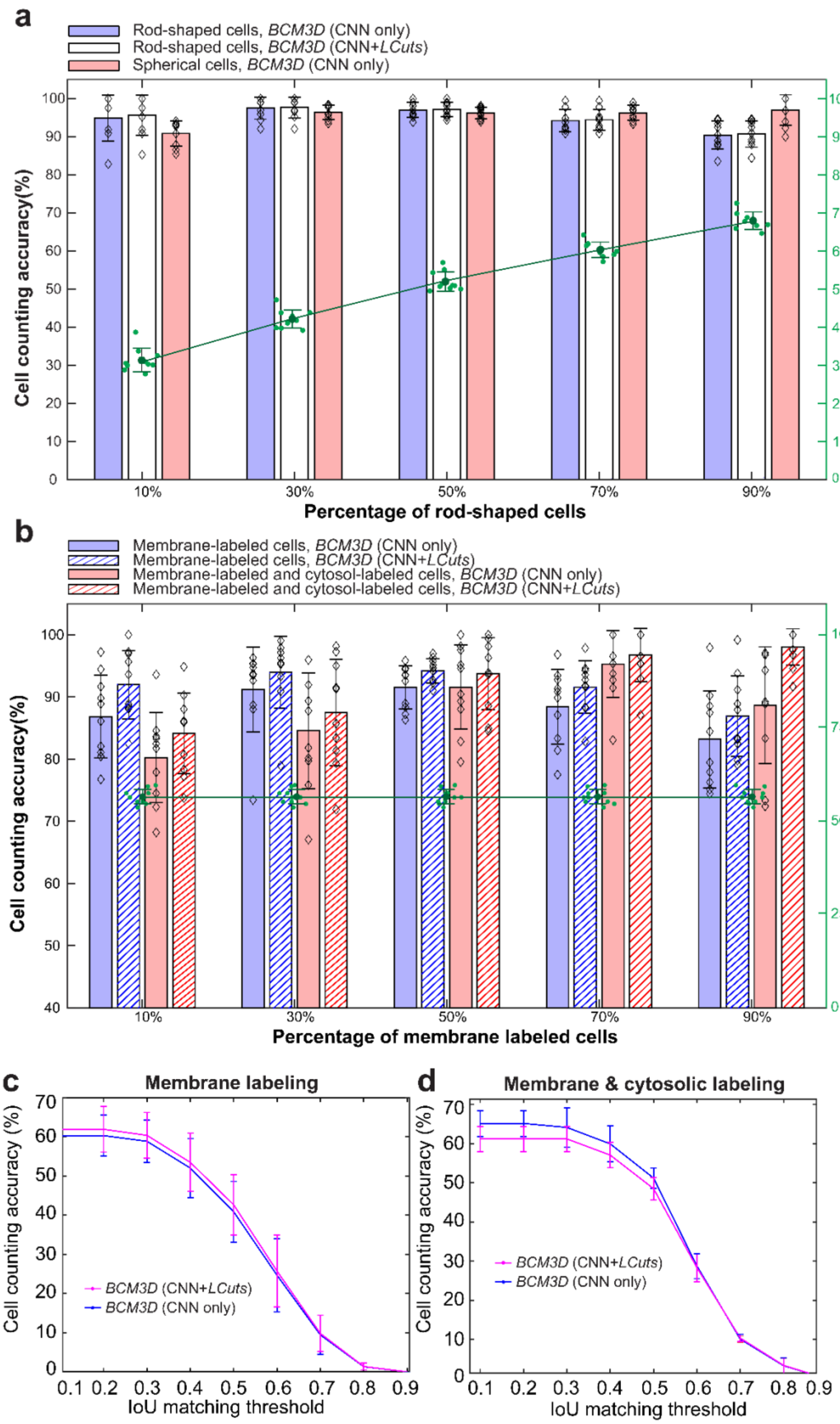

a
Rod-shaped cells, *BCM3D* (CNN only)
Rod-shaped cells, *BCM3D* (CNN+*LCuts*)
Spherical cells, *BCM3D* (CNN only)
Cell counting accuracy(%)
Density (%)
0
10
20
30
40
50
60
70
80
90
100
10%
30%
50%
70%
90%
Percentage of rod-shaped cells
b
Membrane-labeled cells, *BCM3D* (CNN only)
Membrane-labeled cells, *BCM3D* (CNN+*LCuts*)
Membrane-labeled and cytosol-labeled cells, *BCM3D* (CNN only)
Membrane-labeled and cytosol-labeled cells, *BCM3D* (CNN+*LCuts*)
Cell counting accuracy(%)
Density (%)
40
50
60
70
80
90
100
0
25
50
75
100
10%
30%
50%
70%
90%
Percentage of membrane labeled cells
c
Membrane labeling
Cell counting accuracy (%)
BCM3D (CNN+*LCuts*)
BCM3D (CNN only)
0
10
20
30
40
50
60
70
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
IoU matching threshold
d
Membrane & cytosolic labeling
Cell counting accuracy (%)
BCM3D (CNN+*LCuts*)
BCM3D (CNN only)
0
10
20
30
40
50
60
70
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
IoU matching threshold

Figure 3.2.7. Performance of *BCM3D* on mixed-population biofilm images. (a) Cell counting accuracy of *BCM3D* on simulated images containing different ratios of rod-shaped and spherical cells. Black diamonds represent the counting accuracy for *N* = 10 independently simulated datasets. Green dots represent the cell density for each independent dataset. Error bars represent ± one standard deviation. **(b)** Cell counting accuracy of *BCM3D* on simulated images with different ratios of membrane-labeled, and membrane-labeled and interior fluorescent protein expressing cells. Black diamonds represent the counting accuracy for *N* = 10 independently simulated datasets. Green dots represent the cell density for *N* = 10 independent datasets. Error bars represent ± one standard deviation. **(c** and **d)** Cell counting accuracy of *BCM3D* on experimental images of (c) membrane-labeled, and (d) membrane-labeled and interior fluorescent protein expressing *E. coli* cells (mixing ratio ~ 1:1). Each data point is the average of the cell counting accuracies calculated using annotation maps traced by three different researchers (*N* = 3). Data are presented as mean values ± one standard deviation indicated by error bars.

3.2.4 Discussion

CNNs have been successful applied to many different problems in biological image analysis, but their ability to segment individual cells in 3D and time-lapse 3D bacterial biofilm images has not yet been fully explored. Here, we demonstrated a CNN-based image analysis workflow, termed *BCM3D*, for single-cell segmentation and shape classification (morphometry) in 3D images of bacterial biofilms. In this work, we applied *BCM3D* to 3D images acquired by lattice light sheet microscopy. However, *BCM3D* readily generalizes to 3D images acquired by confocal microscopy or advanced super-resolution microscopy modalities, provided that realistic image formation models are used to simulate the training datasets. The use of simulated training data is a major advantage of *BCM3D*, because it overcomes inconsistencies inherent in manual dataset annotation (**Figure S5** and **S6**) and thus solves the problem of obtaining sufficient amounts of accurately annotated 3D image data. The ability to use simulated training data provides needed flexibility not only in terms of the microscope platform used for imaging, but also in terms of the bacterial cell shapes that are to be segmented.

We systematically investigated the advantages and limitations of *BCM3D* by evaluating both voxel- and cell-level segmentation accuracies using simulated and experimental datasets of different cell densities and SBRs. *BCM3D* enabled accurate segmentation of individual cells in crowded environments and automatic assignments of individual cells to specific cell populations for most of the tested parameter space. Such capabilities are not readily available when using previously established segmentation methods that rely exclusively on conventional image and signal processing algorithms.

While *BCM3D* surpasses the performance of previous approaches, we stress that further improvements are possible and, for long-term, high frame-rate time-lapse imaging experiments, absolutely needed. Our systematic analysis revealed that high cell density and low SBR datasets are particularly challenging for the CNNs used in this work. Future work will therefore focus on increasing the contrast and resolution in bacterial biofilm images. While, the use of optical super-resolution modalities can provide higher spatial resolution, such resolution improvements often come at a cost of reduced image contrast and faster photobleaching/phototoxicity. Software solutions that can process images with limited resolution and low SBRs will therefore play a tremendously important role in biological imaging. *BCM3D* is a general workflow that integrates computational simulation of training data, *in silico*-training of CNNs for a *specific* task or a *specific* cell type, and mathematical post-processing of the CNN outputs. Incorporating different training strategies and different CNNs, such as the generalist CNN used in *Cellpose*[61], into the *BCM3D* workflow will enable automated cross-validation of segmentation results when a ground truth or manual annotation map is not available. Furthermore, CNN-based image processing modules developed for contrast enhancement and denoising have also surpassed the performance of conventional methods based on mathematical signal processing[21, 84-86]. Incorporating these tools into the *BCM3D* workflow promises to further improve the single-cell segmentation accuracies. We anticipate that the ability to accurately identify and delineate individual cells in dense 3D biofilms will enable accurate cell tracking over long periods of time. Detailed measurements of behavioral single-cell phenotypes in larger bacterial communities will help determine how macroscopic biofilm properties, such as its mechanical cohesion/adhesion and its biochemical metabolism, emerge from the collective actions of *individual* bacteria.

3.2.5 Methods

Lattice Light Sheet Imaging of Bacterial Biofilms

Fluorescence images of bacterial biofilms were acquired on a home-built lattice light sheet microscope (LLSM). LLSM enables specimen illumination with a thin light sheet derived from 2D optical lattice[26, 87]. Here, a continuous illumination light sheet was produced by a time-averaged (dithered), square lattice pattern[26], and the illumination intensity at the sample was <1 W/cm^2. The submicrometer thickness of the excitation light sheet is maintained over long propagation distances (~30 µm), which enables optical sectioning, and thus high resolution, high contrast imaging of 3D specimens comparable to confocal microscopy. However, fluorophore excitation by a 2D light sheet reduces phototoxicity, because each excitation photon has multiple opportunities to be absorbed by fluorophores in the excitation plane and produce in-focus fluorescence. Widefield fluorescence images corresponding to each illuminated specimen plane are recorded on a sCMOS detector (Hamamatsu ORCA Flash v2). In this work, 3D biofilm images were acquired by translating the specimen through the light sheet in 200 nm steps using a piezo nanopositioning stage (Physik Instrumente, P-621.1CD). The data acquisition program is written in LabVIEW 2013 (National Instruments).

Ampicillin resistant *E.coli* K12, constitutively expressing GFP[88], were cultured at 37 degrees overnight in LB medium with 100 µg/ml ampicillin. Overnight cultures were diluted 100 times into the same culture medium, grown to an optical density at 600 nm (OD600) of 0.6 – 1.0, and then diluted by an additional factor of 10. Round glass coverslips with the diameter of 5 mm were put into a 24-well plate (Falcon) and 400 µL of cell culture was added to the well. Cells were allowed to settle to the bottom of the well and adhere to the coverslip for 1 hour. The round coverslips were then mounted onto a sample holder and placed into the LLSM sample-basin filled with M9 medium. GFP fluorescence was excited using 488 nm light sheet excitation. Biofilm growth was imaged at room temperature every 30 min for a total of 20 time points. At each time point, a single 3D image stack contained 400 images, each acquired with a 15 ms exposure time to avoid motion blur.

M. xanthus strain LS3908 expressing tdTomato under the control of the IPTG-inducible promoter[89] and DK1622 (WT) were cultured in the nutrient rich CYE media at 30 degrees Celsius until it reached an OD600 of 0.6 - 1.0. Media was supplemented with 1 mM IPTG for tdTomato expressing cells. Chitosan (Thermo Fisher)-coated 5 mm round glass coverslips were prepared by incubating coverslips with 1% (w/v) chitosan (1.5 % glacial acetic acid (v/v)) at room temperature for 1 hour. Coverslips were then rinsed with water and placed into a 24-well plate (Falcon) with 350-400 µL of undiluted cell culture. WT cells were stained directly in the 24 well plate with 5 ng/ml FM4-64 (Thermo Fisher) dye. Cells were allowed to settle and adhere to the coverslip for 2 hours. After the settling period, the coverslip was gently rinsed with CYE media to flush away unattached cells. The rinsed coverslip was then mounted onto a sample holder and placed into the LLSM sample-basin filled with MC7 starvation buffer. tdTomato and FM 4-64 fluorescence was excited using 561 nm light sheet excitation. The 3D image stack contained 400 2D images. Each 2D slice was acquired with an exposure time of 30 ms.

For mixed population biofilm imaging, ampicillin resistant *E.coli* K12, constitutively expressing GFP[88], and ampicillin resistant *E.coli* K12, expressing mScarlet (pBAD vector, arabinose induce) were cultured separately at 37 degrees overnight in LB medium with 100 μg/ml ampicillin. Overnight cultures were diluted 100 times into the same culture medium, grown to an optical density at 600 nm (OD600) of 0.6 – 1.0, and then diluted to an OD of 0.1. After dilution, the two strains were mixed together. Round glass coverslips with the diameter of 5 mm were put into a 24-well plate (Falcon) and 500 μL of cell culture was added to the well. Cells were allowed to settle to the bottom of the well and adhere to the coverslip for 1 hour. The cell culture medium was then removed and replaced by 500 uL M9 medium containing 0.2% (w/v) arabinose. The co-culture was incubated at 30 degrees overnight. 10 mins before imaging, the co-culture was stained with 5 ng/ml FM4-64 (Thermo Fisher) dye. 3D image stacks of 20 planes with 5 ms exposure time per frame were acquired using 488 nm excitation.

Raw Data Processing

Raw 3D images were background subtracted and then deskewed and deconvolved[26, 32]. The background was estimated by averaging intensity values of dark areas (devoid of cells) in the field of view. Deconvolution was performed using the Richardson-Lucy algorithm with 10 iterations using experimentally measured point spread functions (PSFs) as the deconvolution kernel. The experimentally measured PSFs were obtained separately for each color channel using fluorescent beads (200 nm FluoSpheres®, Thermo Fisher) coated on a coverslip[90]. 3D images were rendered using the 3D Viewer plugin in Fiji[31] or ChimeraX[34].

Generation of simulated biofilm images

To generate data for training of CNNs, we computationally simulated fluorescence images of 3D biofilms, for which spatial arrangements among individual cells are known precisely and accurately. Growth and division of individual rod-shaped cells in a population were simulated using CellModeller, an individual-based computational model of biofilm growth (**Figure 3.21a**)[91]. In individual-based biofilm growth models, cells are the basic modeling units. Each cell is characterized by a set of parameters, including its 3D position, volume, and spatial orientation. All the cells in the simulated biofilm are then allowed to evolve in time according to predefined biological, chemical, and mechanical rules. For example, cells grow at a defined rate and then divide after reaching a certain volume threshold. Cellular collisions that are due to cell growth are alleviated by imposing a minimum distance criterion between cells at each time point. For our simulations, we chose cell diameter and cell length (*d*, *l*) parameters consistent with a given bacterial species, namely (1 µm, 3 µm) for *E. coli*[92], (0.7 µm, 6 µm) for *M. xanthus*[93], and (1 µm, 1 µm) for spherically symmetric *S. aureus*[94].While the cell volume can be readily adjusted in CellModeller, the cellular volume density, which is determined by the intercellular spacing, is not directly adjustable. We therefore adjusted the cellular volume density after each simulation by scaling the cellular positions (cell centroids) and thus the intercellular distances by a constant factor, while leaving cell sizes, shapes, and orientations unchanged. This post-processing procedure enabled simulation of the exact same 3D cell arrangements at adjustable cell volume densities.

We fluorescently labeled simulated cell volumes and surfaces according to two commonly used labeling strategies in fluorescence microscopy. To simulate expression of intracellular fluorescent proteins, the fluorescence emitters were placed at random positions within the cell volume. To simulate membrane staining, the fluorescence emitters were placed at random positions on the cell surface. Each cell contained between 500 - 1000 fluorophores to simulate expression level variations between cells, which is often observed in experimental images. Once the fluorophore spatial distributions were determined, a 3D fluorescence image (**Figure 3.2.1b**) was computationally generated. Each fluorophore was treated as an isotropic point emitter, so that it would produce a diffraction-limited point-spread-function (PSF) on the detector. Experimentally measured 3D PSF shapes (see Raw Data Processing) were used as the convolution kernel. Next, the fluorescence signal intensity was scaled by multiplying the image by a constant factor and then a constant background intensity was added to the image at ~200 photons per pixel, as measured in experimental data. This procedure enabled independent adjustments of the fluorescence signal and background to obtain signal-to-background ratios (SBRs) consistent with experimental data. In a final step, we introduced Poisson-distributed counting noise, based on the summed background and signal intensities, as well as Gaussian-distributed camera read-out noise (experimentally calibrated for our detector at 3.04 photons per pixel on average)[95]. This resulting image data (**Figure 3.2.1c**) was then processed in the same manner as experimental data (see Raw Data Processing). In contrast to experimental data, generation of the corresponding voxel-level annotation maps is fast and error free, because the underlying ground truth cell arrangements are known *a priori* (**Figure 3.2.1d**).

To mimic imaging of reporter gene expression in a subset of cells, we simulated biofilm images, in which all cells were stained at the cell surface (e.g. with a membrane intercalating fluorescent dye) and a subset of cells additionally contained intracellular fluorophores (e.g. through the expression of an intracellular fluorescent protein) (**Figure 3.2.8ab**). The mixing ratios between membrane-labelled, and membrane and interior labelled cells were 10:90, 30:70, 50:50, 70:30 and 90:10. Ten different cell arrangements containing ~300 cells were simulated for each ratio. To train the CNNs (see next section), six datasets were used, all with a 50:50 mixing ratio.

To mimic imaging of cells with different morphologies, we simulated biofilms containing spherical and rod-shaped cells **(Figure 3.2.8cd)**. Cell arrangements were first simulated using rod shaped cells and then a fraction of rod-shaped cells is replaced with spherical cells. The size of the rod-shaped cells is that of *E. coli* (~3 × 1 µm, length by diameter). The size of the spherical cells is that of *S. aureus* (~1 µm in diameter)[96]. Both cell types were labelled by intracellular fluorophores, as described above. The mixing ratios between rod-shaped and spherical cells were 10:90, 30:70, 50:50, 70:30 and 90:10. Ten different cell arrangements containing ~300 cells were simulated for each ratio. To train the CNNs (see next section), we picked one image from each mixing ratio for a total of five images.

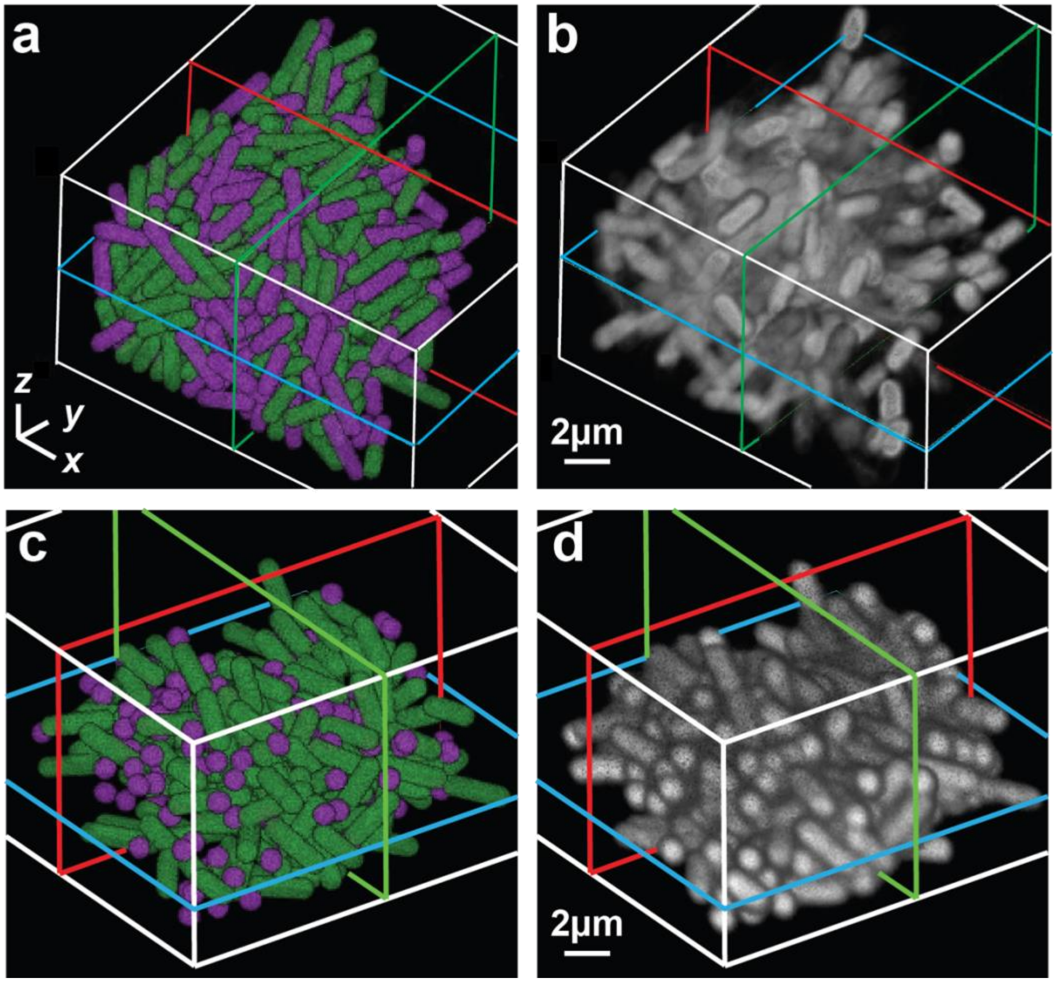


Figure 3.2.8. Simulation of mixed labeling and mixed cell shape biofilms. Cell arrangements (green indicates membrane labeled cells, magenta indicates membrane labeled cells that simultaneously express interior fluorescence protein). (**b**) Simulated fluorescence image based on the cell arrangements in (a) as displayed by the volume viewer plugin of Fiji[97]. *N* = 10 independent images with similar cell densities, but different cell arrangements, where analyzed for each data point in **Figure 6**. (**c**) Cell arrangements (green indicates rod-shaped cells, magenta indicates spherical shaped cells). (**d**) Simulated fluorescence image based on the cell arrangements in (c) as displayed by the volume viewer plugin of Fiji[97]. *N* = 10 independent images with similar cell densities, but different cell arrangements, where analyzed for each data point in **Figure 3.2.6**.

Training the convolutional neural networks

We trained 3D U-Net CNNs for voxel-level classification tasks[98] within the NiftyNet platform[99] (network architecture depth 4, convolution kernel size 3, ReLU activation function, 32 initial feature maps, and random dropout of 0.5 during training). To achieve robust performance, we trained these networks using five to ten simulated biofilm images with randomly selected cell densities and signal-to-background ratios (see Generation of simulated biofilm images). The same raw data processing steps used for experimental data (see Raw Data Processing) were also applied to simulated data. 3D deconvolved simulated data and their corresponding voxel-level annotations were used to train the CNNs. Each image used for training contained ~9 million voxels. We trained CNNs by classifying each voxel as ‘background’, ‘cell interior’ or as ‘cell boundary’ based on the underlying cell arrangements. For mixed-species biofilms, two additional classes, ‘cell interior’ and ‘cell boundary’ of the second species, were used. This type of annotation scheme has been shown to increase separation of bacterial cells in 2D[100]. For data augmentation, we applied NiftyNet’s built-in scaling, rotation, and elastic deformation functions. Instead of the original cross-entropy loss function combined with uniform sampling, we used the Dice loss function and ‘balanced sampler’, so that every label has the same probability of occurrence in training. All networks were trained for 2000 to 3600 iterations with a learning rate of 0.0001. Using these parameters, it took approximately 24 hours to train the CNNs on a NVIDIA Tesla V100 GPU with 16 GB memory.

Thresholding of CNN-produced confidence maps

Voxel-level classification by CNNs generates different confidence maps (one confidence map for each annotation class). The confidence values range between 0 and 1 and represent the confidence of assigning individual voxels to a given class. After thresholding the ‘cell interior’ confidence map to obtain a binary image, connected voxel clusters can be isolated and identified as single cell objects using 3D connected component labeling[101]. A conservative size-exclusion filter was applied: small objects with a volume ~10 times less than the expected cell size were considered background noise and filtered out using an area open operator[101]. Since the cell-interior volumes do not contain the cell boundaries, we dilated each object by 1-2 voxels to increase the cell volumes using standard morphological dilation[101]. The threshold value to segment individual cell objects based on the 'cell interior' confidence map was determined by plotting the overall voxel-level segmentation accuracy , quantified as the Intersection-over-Union value (IoU value, aka Jaccard index[102]) versus the confidence value thresholds. Optimal voxel-level segmentation accuracies were consistently obtained using confidence thresholds between 0.88 and 0.94. Throughout this work, we use 0.94 for cells labeled with intracellular fluorophores and 0.88 for cells labeled with membrane-localized fluorophores.

Post-processing of U-Net result using a refined *LCuts* algorithm

Thresholding of the 'cell interior' confidence map produces a binary segmentation result (background = 0, cell interior =1), where groups of connected, non-zero voxels identify individual cells in most cases. However, when cells are touching, they are often not segmented as individuals, but remain part of the same voxel cluster (undersegmentation). On the other hand, a single cell may be erroneously split into smaller sub-cellular objects (oversegmentation). Finally, in datasets with low SBR, connected voxel clusters may be detected that do not correspond to cells and thus produce false positive objects. To address these errors and improve the segmentation accuracy further, we included additional mathematical image analysis steps to post-process the CNN results and reduce undersegmentation and oversegmentation errors.

Step 1. False positive objects are identified by evaluating the coefficient of variation[103, 104] for each connected voxel cluster i:

$$CV_i = \frac{\sigma_i}{\mu_i} \quad (1)$$

where σ_i and μ_i denote the standard deviation and the mean of the intensity taken over all voxels contained in connected voxel cluster i. If the coefficient of variation is larger than ρ, then the current object will be classified as a false positive object and removed from the confidence map by setting all its voxels to zero. The removed objects will then no longer be counted when evaluating the cell counting accuracy. The value of ρ is selected based on the coefficient of variation of the background. For the datasets analyzed here, this sample coefficient of variation was determined to be $\rho = 1.1$. After CV-filtering, objects smaller than 25% of the expected bacterial cell size are also removed by setting its voxels to zero. The remaining connected voxel clusters are then considered for further processing.

Step 2. To identify and delineate individual cells in the connected voxel clusters identified in the previous step, we implemented medial axis extraction using the method of inscribed spheres[65], with the constraint that the sphere radii do not exceed the expected diameter of a single bacterial cell (e.g. $d = 0.8$ µm). The set of N inscribed spheres are tangent to the object's surface and parameterized by $(x_i, y_i, z_i; r_i<d/2)$ for $i = 1, \ldots, N$. Determination of the $(x_i, y_i, z_i; r_i)$ coordinates is achieved using the Euclidean distance transform of the objects' boundary[105], so that the points with coordinates (x_i, y_i, z_i) reliably trace out the central cell axes of individual bacterial cells.

Step 3. To separate different linear segments after cell axis extraction, we used a refined version of the linear cuts (*LCuts*) algorithm[66]. *LCuts* is a graph-based data clustering method designed to detect linearly oriented groups of points with certain properties. The fundamental elements of a weighted mathematical graph are nodes, edges, and edge weights. Here, the points with coordinates (x_i, y_i, z_i) represent the graph nodes. Edges are the connections among nodes. Edges are assigned weights, for example, to reflect the confidence that two nodes belong to the same group. *LCuts* achieves grouping by assigning weights to edges in the fully connected graph to reflect the similarity between two nodes. The features of each node include its location and direction, where the location of each node is simply its Cartesian coordinates. The direction of each node is found by first determining its 5-hop neighborhood, removing nodes at large relative angles, and evaluating the major direction of the outlier removed neighborhood.

The algorithm to separate the nodes into different groups is a recursive graph cutting method[66]. Graph cuts (*e.g.* nCut[106]) disconnect the edges between two groups of nodes when the combined weights of these edges are minimized. The weights, between node *i* and node *j*, are calculated as follows:

$$w_{ij} = w_D \cdot w_T \qquad (2)$$

where

$$w_D = \begin{cases} e^{-D_{ij}^2/\sigma_D^2} & \text{if } D_{ij}^2 \leq r \\ 0 & \text{if } D_{ij}^2 > r \end{cases} \qquad (3)$$

$$w_T = e^{-\left(\cos(\theta_{ij})-1\right)^2/\sigma_T^2} \qquad (4)$$

w_D weighs the distance between two nodes and w_T weighs difference between node directions. D_{ij} is the Euclidean distance between node *i* and node *j*, and r is set to eliminate edges between two far away nodes. θ_{ij} is the relative angle between the directions of nodes *i* and *j*. σ_D and σ_T are adjustable parameters that control the rate of exponential decay. *LCuts* continues to separate groups of nodes until each group satisfies a stopping criterion. The stopping criterion is biologically inspired based on the expected length *L* of a single bacterial cell and a group's linearity after each recursion. *LCuts* yields linearly oriented groups of points that trace out the central axes of individual cells. Importantly, cell separation is achieved without having to specify the number of cells in the biofilm in advance. Furthermore, to limit the need for optimization of postprocessing routines, the four adjustable parameters used in *LCuts*, namely cell diameter d, the cell length L, and the decay parameters σ_D and σ_T are chosen based on *a priori* knowledge about the bacterial cells under investigation. We found that the performance of *LCuts* is not sensitive to the particular values of d, L, σ_D and σ_T as long as they are consistent with the imaged bacterial cell sizes and shapes. Identification of single cells provided by *LCuts* alleviates under-segmentation errors of the CNN-based segmentation.

Step 4. The final output of linear clustering can provide length, location and orientation of each cell. Based on these linear clusters, the cellular architecture of the biofilms can be reconstructed by placing geometrical models of cells in space. For fast computation, spherocylinders are used as the geometrical model using a radius consistent with the known sizes of bacterial cells. To further refine the cell surfaces to better align with the CNN-segmented volumes, we enclosed the inscribed spheres found in Step 2 in a convex hull.

Performance Evaluation

We quantified segmentation accuracy both at the cell-level (object counting) and at the voxel-level (cell shape estimation). To quantify the cell-level segmentation accuracy, we designated segmented objects as true positive (*TP*) if their voxel overlap with the ground truth or the manual annotation resulted in an IoU value larger than a particular IoU matching threshold. This criterion ensures one-to-one matching. A threshold of 0.5 is typically chosen when reporting single cell counting accuracy values [61, 67]. We follow this convention here. If the segmented cell object could not be matched to a ground truth/manually annotated cell volume, then it was counted as a false positive (*FP*) and the IoU value of that segmented object was set to zero. If a ground truth/manually annotated cell volume was not identified in the image, then it was counted as false negative (*FN*). The cell (object) counting accuracy was then defined as *TP*/(*TP*+*FP*+*FN*). The average IoU value over all segmented objects in the image quantifies the voxel-level segmentation accuracy, i.e. the accuracy of cell shape estimation.

To evaluate the accuracy of cell segmentation on experimental data, three researchers separately traced the cell contours on experimental 2D slices by using freehand selections in Fiji ROI Manger[97]. Because human annotation is very time consuming (about 50 hours for a complete 3D dataset containing ~300 cells in a 22 x 32 x 12 um^3 volume), one to three single 2D slices were selected for each dataset. One exception is the 3D *M. xanthus*, for which the cell outlines in all available x, y and z slices were traced manually. For straight, rod-shaped cells, the centroids of the resulting 2D cell contours all fall within the cell interior volume. To group together the contours belonging to the same cells, the centroid of each contour was projected along the x, y and z dimension. If the projected centroid was enclosed by any other contour in a different slice, then the centroid of that contour was projected onto the plane of the initial contour. Two contours were labeled as related if they contained each other's projected centroids. This process is repeated for all possible contour pairs and their relationship is recorded in an adjacency matrix. Next, related contours were assigned to individual cells. To separate incorrectly grouped contours, we additionally identified clusters of centroids using the DBSCAN point clustering algorithm[107]. In a final step, we manually removed incorrectly traced contours. Cells are reconstructed by creating convex hulls with the grouped contours. This procedure determined the approximate positions, shapes, and orientations of individual cells in the 3D biofilm.

To estimate the SBRs of both simulated and experimental images, we manually selected and determined the intensities of approximately ten 'signal' and ten 'background' regions in the images. We computed the SBR as the mean signal intensity divided by the mean background intensity. To estimate the local density of a biofilm cluster of simulated images, we sliced the considered several 3D tiles within the biofilm of size 64 by 64 by 8 voxels. We then estimated the local density as total cell volume contained in each tile divided by the tile volume. We calculated the mean density of the 10 densest tiles to define the 'local density' metric reported for each dataset in the paper. To estimate the cell density in an experimentally acquired biofilm image, the same calculations were performed on either 3D manual annotations (if available) or binary masks obtained by CNN-processing.

3.2.6 Acknowledgements

Chapter 3.2 was adapted from Zhang et al. 2020[108]. This work was supported in part by the US National Institute of General Medical Sciences Grant No. 1R01GM139002 (A.G and S.T.A.) and by a Jeffress Memorial Trust Award in Interdisciplinary Research (to A.G and S.T.A.). We thank Lotte Søgaard-Andersen, Larry Shimkets, Huiwang Ai and Ingmar Riedel-Kruse for providing bacterial strains used in this work. We thank Karsten Siller, computational research consultant at the University of Virginia, for help with utilizing the Universities super-computing resources. We thank Knut Dresher and Eric Jelli for providing the segmentation results using the Hartmann *et al.* algorithm.

3.3 BCM3D 2.0: improved segmentation performances by image-to-image translation using deep neural networks

3.3.1 Abstract

Accurate detection and segmentation of single cells in three-dimensional (3D) fluorescence time-lapse images is essential for observing individual cell behaviors in large bacterial communities called biofilms. Recent progress in machine-learning-based image analysis is providing this capability with ever increasing accuracy. Leveraging the capabilities of deep convolutional neural networks (CNNs), we recently developed bacterial cell morphometry in 3D (*BCM3D*), an integrated image analysis pipeline that combines deep learning with conventional image analysis to detect and segment single biofilm-dwelling cells in 3D fluorescence images. While the first release of *BCM3D* (*BCM3D 1.0*) achieved state-of-the-art 3D bacterial cell segmentation accuracies, low signal-to-background ratios (SBRs) and images of very dense biofilms remained challenging. Here, we present *BCM3D 2.0* to address this challenge. *BCM3D 2.0* is entirely complementary to the approach utilized in *BCM3D 1.0*. Instead of training CNNs to perform voxel classification, we trained CNNs to translate 3D fluorescence images into intermediate 3D image representations that are, when combined appropriately, more amenable to conventional mathematical image processing than a single experimental image. Using this approach, improved segmentation results are obtained even for very low SBRs and/or high cell density biofilm images. The improved cell segmentation accuracies in turn enable improved accuracies of tracking individual cells through 3D space and time. This capability opens the door to investigating time-dependent phenomena in bacterial biofilms at the cellular level.

3.3.2 Introduction

Most terrestrial bacteria live in extended 3-dimensional tissue-like communities, named biofilms. As multicellular communities, bacteria can successfully colonize various biotic and abiotic surfaces. Biofilm-dwelling bacteria interact intimately not only with each other and the surface they reside on, but also with a self-produced extracellular matrix (ECM) that consists of proteins, DNA, and polysaccharides[37, 109, 110]. The sum total of these interactions helps biofilms develop emergent capabilities beyond those of isolated cells[37, 109, 111, 112]. Most notably, biofilms are more tolerant towards physical, chemical, and biological stressors[4, 112, 113]. Understanding how such capabilities emerge from the cooperative or antagonistic behaviors among individual cells requires live-cell compatible imaging technologies that are capable of resolving and tracking single cells within dense 3D biofilms.

Recently developed light sheet-based fluorescence imaging modalities combine high resolution with fast imaging speed and low phototoxicity at levels that cannot be matched by confocal microscopy[13-15]. Light sheet-based microscopy modalities are therefore increasingly used for non-invasive time-lapse imaging of eukaryotic cells and tissues[114-116] as well as bacterial biofilms[16-18]. Depending on the type of biofilm, the cell density may however be too high to clearly resolve the gaps between cells with diffraction-limited microscopy. Super-resolution imaging modalities, such as structured illumination microscopy[19, 20], improve the spatial resolution, but experimental improvements in spatial resolution come at the cost of decreased temporal resolution and increased light exposure to the specimen, which again raises photobleaching and phototoxicity concerns[21, 22]. An additional challenge arises for cell tracking studies. Tracking motile cells may require high frame rate imaging to achieve sufficient temporal resolution. Higher frame rates need to be accompanied by a proportional decrease in excitation laser intensities to mitigate photobleaching and phototoxicity. The decreased excitation laser intensities then result in lower signal-to-background ratios (SBRs) and signal-to-noise ratios (SNRs) in the individual images. The inherent trade-offs between spatial and temporal resolution, SBR/SNR, and photobleaching and phototoxicity is driving the continued development of new and improved image processing approaches that extract ever increasing amounts of useful information from the available experimental images.

Image processing pipelines based on supervised training of deep convolutional neural networks (CNNs) have been shown to outperform conventional image processing approaches for a variety of tasks in biomedical image analysis[62, 117]. For 3D biofilm image segmentation, we have recently developed Bacterial Cell Morphometry 3D (*BCM3D 1.0*), which achieved state-of-the-art performance for bacterial cell counting and cell shape estimation[108]. *BCM3D 1.0* does not rely on manually annotated training data, but instead combines *in silico*-trained CNNs for voxel classification with graph-theoretical linear clustering (mLCuts[118]) to post-process the thresholded CNNs outputs (i.e. the confidence maps for voxel-level classification). Using this approach, *BCM3D 1.0* automatically identifies individual cells in 3D images of 3D bacterial biofilms, reports their 3D shape and orientation, and classifies cell types with different morphologies. However, processing images with low SBRs and high cell densities remains challenging. Specifically, over- and under-segmentation errors increase in frequency for low SBR and high cell density images.

Cellpose[119] ,*StarDist*[120] and the work by Scherr *et al.*,[121] are CNN-based approaches that create intermediate image representations for better segmentation. We reasoned that solving an image-to-image translation task may prove to be a more robust strategy for handling extreme imaging conditions than the voxel classification approach implemented in *BCM3D 1.0* or, at least, yield complementary segmentation results to *BCM3D 1.0*. Two different intermediate image representations are generally employed. The first representation is used to locate objects and the second representation is used to highlight the boundaries of objects. In previous work[119-121], the CNN-predicted Euclidean distance to the nearest background pixel/voxel or the CNN-predicted object/background probability map was used to locate objects. Generation of boundary representations vary more widely: *StarDist* and *Cellpose* use star-convex polygons and spatial gradients separately to give complete boundaries, which can be used for object shape estimation. Scherr *et al*. instead enhance boundary regions that are close to other objects to prevent them from merging. Inspired by these approaches, we expanded the *BCM3D* workflow with a complementary CNN-based processing pipeline that translates the raw 3D fluorescence images into two distinct intermediate image representations that, in combination, are more amenable to conventional mathematical image processing, namely seeded watershed[122] and Otsu thresholding[123]. For object localization, we adapted the approach used by *StarDist*[67] and Scherr et al.[121]. For boundary information, however, we generated a new intermediate image representation that provides a complete 3D boundary of an object and additionally highlights whether the boundary is near other objects. We establish that, when combined and processed appropriately, these intermediate image representations provide biofilm segmentation results with higher accuracy than *BCM3D 1.0*. Importantly and in contrast to *BCM3D 1.0*, generation of intermediate image representations does not require image deconvolution as a pre-processing step. Deconvolution can lead to noise amplification[124] that then leads to false positive object segmentation with physiologically unreasonable shapes. We show that, using intermediate image representations, experimentally acquired biofilm images can be successfully segmented using CNNs trained with computationally simulated biofilm images – a feature shared with *BCM3D 1.0* that provides the flexibility to segment a wide variety of different cell shapes[108]. The segmentation performance of this new approach, which we term *BCM3D 2.0*, is superior to *Omnipose* and *Cellpose 2.0*, two recently-developed, state-of-the-art, CNN-based cell segmentation approaches, especially for dense biofilms imaged at low SBRs. The improvements in segmentation accuracy of *BCM3D 2.0* enables accurate multi-cell tracking, which is demonstrated using 3D simulated and experimental time-lapse biofilm images.

3.3.3 Results

Cell segmentation using intermediate image representations

High cell density and low SBR datasets are encountered often in biofilm research, especially when cells touch each other and biofilms extend farther into the vertical (*z*-) dimension, so that light scattering becomes a pronounced background contribution[12]. We therefore sought to specifically improve bacterial cell segmentation accuracy for high cell density and low SBR biofilm images. Our previous approach (*BCM3D 1.0*) relied on deconvolution as a preprocessing step to sharpen the image and to increase the SBR. However, deconvolution can introduce artifacts into an image, such as ringing[125], and noise amplification[126], and thereby introduce errors into the segmentation results. The segmentation pipeline of *BCM3D 2.0*, in contrast, works on the raw image data directly without the need for deconvolution.

We compared two commonly used cell labeling approaches, namely cell interior labeling through expression of cytosolic fluorescent proteins and cell membrane staining with membrane-embedded fluorescent dyes. For cell interior labeling (**Figure 3.3.1ab**), *BCM3D 2.0* consistently produces cell counting accuracies of >95% for SBRs > 1.3 and cell densities < 65%. A clear drop-off in cell counting accuracy is observed for SBRs of 1.19 but cell counting accuracies of >70%

are still achieved even for high cell densities of 65%. Importantly, the performance of *BCM3D 2.0* on low SBR datasets represents a substantial improvement (>20%) over the performance of *BCM3D 1.0*. Membrane staining (**Figure 3.3.1cd**) produces even more challenging images for segmentation, due to the less pronounced fluorescence intensity minima between cells (red arrow in **Figure 3.3.1bd**). We again observe a drop in cell counting accuracy for SBRs of 1.19. This drop-off is however much less pronounced than for the previous results obtained with *BCM3D 1.0*, and represents an even larger (>29%) improvement over *BCM3D 1.0* for such extremely low SBR datasets. Visual inspection of slices through the image volumes (**Figure 3.3.1bd**) reveals that even for SBR = 1.3, the cell bodies are difficult to distinguish for expert human annotators, especially for membrane-stained cells. Despite the low contrast in the SBR = 1.3 datasets, *BCM3D 2.0* is still able to achieve >90% cell counting accuracies, which, depending on cell density, represents a 6-26% increase over the performance of *BCM3D 1.0*.

To determine the improvement in cell shape estimation, we evaluated the cell counting accuracies as a function of IoU matching threshold for a SBR of 1.3 and a cell density of 62% (the IoU matching threshold is a quantitative measure of cell shape similarity relative to the ground truth). The cell counting accuracies obtained by *BCM3D 2.0* are consistently higher than *BCM3D 1.0* (CNN + *LCuts*) and substantially higher than *Omnipose* and *Cellpose 2*, especially for IoU matching thresholds larger than 0.5, indicating that cell shapes in this high density, low SBR dataset are most accurately estimated by *BCM3D 2.0* (**Figure 3.3.1ef**). A similar trend is observed for single-cell segmentation accuracy and single-cell boundary F1 score[127] – two additional metrics for segmentation accuracy (**Table 3.3.1**). We note that we trained *Omnipose, Cellpose 2,* and *BCM3D 2.0* using the same simulated training data for this comparison. Consistent with previous findings[128], *Cellpose*, the precursor algorithm to *Omnipose*, did not produce physiologically-reasonable cell shapes (**Figure 3.3.2**). Taken together, these results establish that more robust cell segmentation can be achieved using the *BCM3D 2*.0 image processing pipeline, which uses CNNs to generate two distinct intermediate image representations for subsequent mathematical processing.

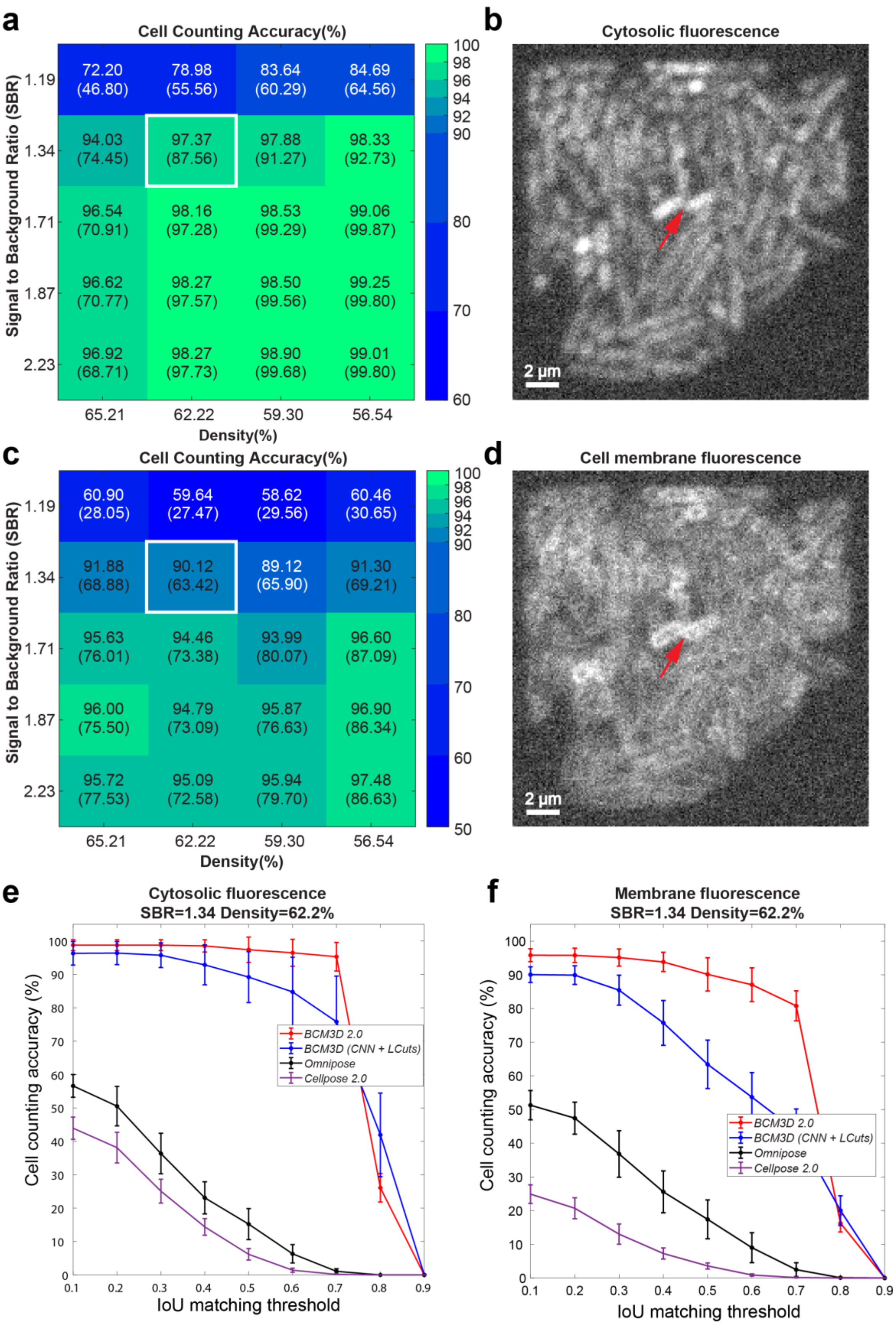

a
Cell Counting Accuracy(%)
Signal to Background Ratio (SBR)
1.19
72.20 (46.80)
78.98 (55.56)
83.64 (60.29)
84.69 (64.56)
1.34
94.03 (74.45)
97.37 (87.56)
97.88 (91.27)
98.33 (92.73)
1.71
96.54 (70.91)
98.16 (97.28)
98.53 (99.29)
99.06 (99.87)
1.87
96.62 (70.77)
98.27 (97.57)
98.50 (99.56)
99.25 (99.80)
2.23
96.92 (68.71)
98.27 (97.73)
98.90 (99.68)
99.01 (99.80)
65.21
62.22
59.30
56.54
Density(%)
100
98
96
94
92
90
80
70
60
b
Cytosolic fluorescence
2 µm
c
Cell Counting Accuracy(%)
Signal to Background Ratio (SBR)
1.19
60.90 (28.05)
59.64 (27.47)
58.62 (29.56)
60.46 (30.65)
1.34
91.88 (68.88)
90.12 (63.42)
89.12 (65.90)
91.30 (69.21)
1.71
95.63 (76.01)
94.46 (73.38)
93.99 (80.07)
96.60 (87.09)
1.87
96.00 (75.50)
94.79 (73.09)
95.87 (76.63)
96.90 (86.34)
2.23
95.72 (77.53)
95.09 (72.58)
95.94 (79.70)
97.48 (86.63)
65.21
62.22
59.30
56.54
Density(%)
100
98
96
94
92
90
80
70
60
50
d
Cell membrane fluorescence
2 µm
e
Cytosolic fluorescence
SBR=1.34 Density=62.2%
Cell counting accuracy (%)
100
90
80
70
60
50
40
30
20
10
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
IoU matching threshold
BCM3D 2.0
BCM3D (CNN + LCuts)
Omnipose
Cellpose 2.0
f
Membrane fluorescence
SBR=1.34 Density=62.2%
Cell counting accuracy (%)
100
90
80
70
60
50
40
30
20
10
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
IoU matching threshold
BCM3D 2.0
BCM3D (CNN + LCuts)
Omnipose
Cellpose 2.0

Figure 3.3.1. Performance of *BCM3D 2.0* on previously unseen simulated biofilm images. **(a)** Cell counting accuracy (using an Intersection-over-Union (IoU) matching threshold of 0.5 for each segmented object) averaged over *N*=10 replicate datasets for cells labeled with cytosolic fluorophores. IoU is well-documented metric[67, 108, 129] quantifying the amount of overlap between predicted cell and actual cell volumes. **(b)** Example image of cells labeled with cytosolic fluorophores (Cell density = 62.2%, SBR = 1.34, indicated by white rectangle in panel a). **(c)** Cell counting accuracy (using an IoU matching threshold of 0.5 for each segmented object) averaged over *N*=10 replicate datasets for cells labeled with membrane-localized fluorophores. **(d)** Example image of cells labeled with membrane-localized fluorophores (Cell density = 62.2%, SBR = 1.34, indicated by white rectangles in panel c). The red arrow indicates a close cell-to-cell contact. **(e and f)** Comparison of segmentation accuracies achieved by *BCM3D 1.0 (*CNN + *LCuts)*, *Omnipose* trained from scratch, *Cellpose 2* fine-tuned and *BCM3D 2.0* for cytosolic and membrane labeling, respectively (SBR = 1.34, cell density = 62.2%). Segmentation accuracy is parameterized in terms of cell counting accuracy (*y* axis) and IoU matching threshold (*x* axis). Each data point is the average of *N*=10 independent biofilm images. Data are presented as mean values ± one standard deviation.

Table 3.3.1. Quantitative comparison of single cell level segmentation accuracy between *BCM3D 1.0* and *BCM3D 2.0*.

	Cytosolic labeling		Membrane labeling	
	SSA*	SBF1**	SSA*	SBF1**
BCM3D 1.0 (CNN + *LCuts)*	**0.796 ± 0.021**	0.983 ± 0.008	0.756 ± 0.009	0.961 ± 0.007
BCM3D 2.0	**0.791 ± 0.004**	**0.995 ± 0.001**	**0.773 ± 0.005**	**0.988 ± 0.002**
Omnipose	0.599 ± 0.011	0.868 ± 0.017	0.615 ± 0.021	0.824 ± 0.025
Cellpose 2	0.566 ± 0.016	0.825 ± 0.016	0.572 ± 0.008	0.820 ± 0.009

*SSA and **SBF1 estimate how accurately the shape of a matched object compare with it of the corresponding ground truth. Here, the IoU threshold is 0.5 and the distance error tolerance for SBF1 is √3 voxels. (See Methods for details). Data are presented as mean values ± one standard deviation, the best performance (if different within error) is marked in bold.

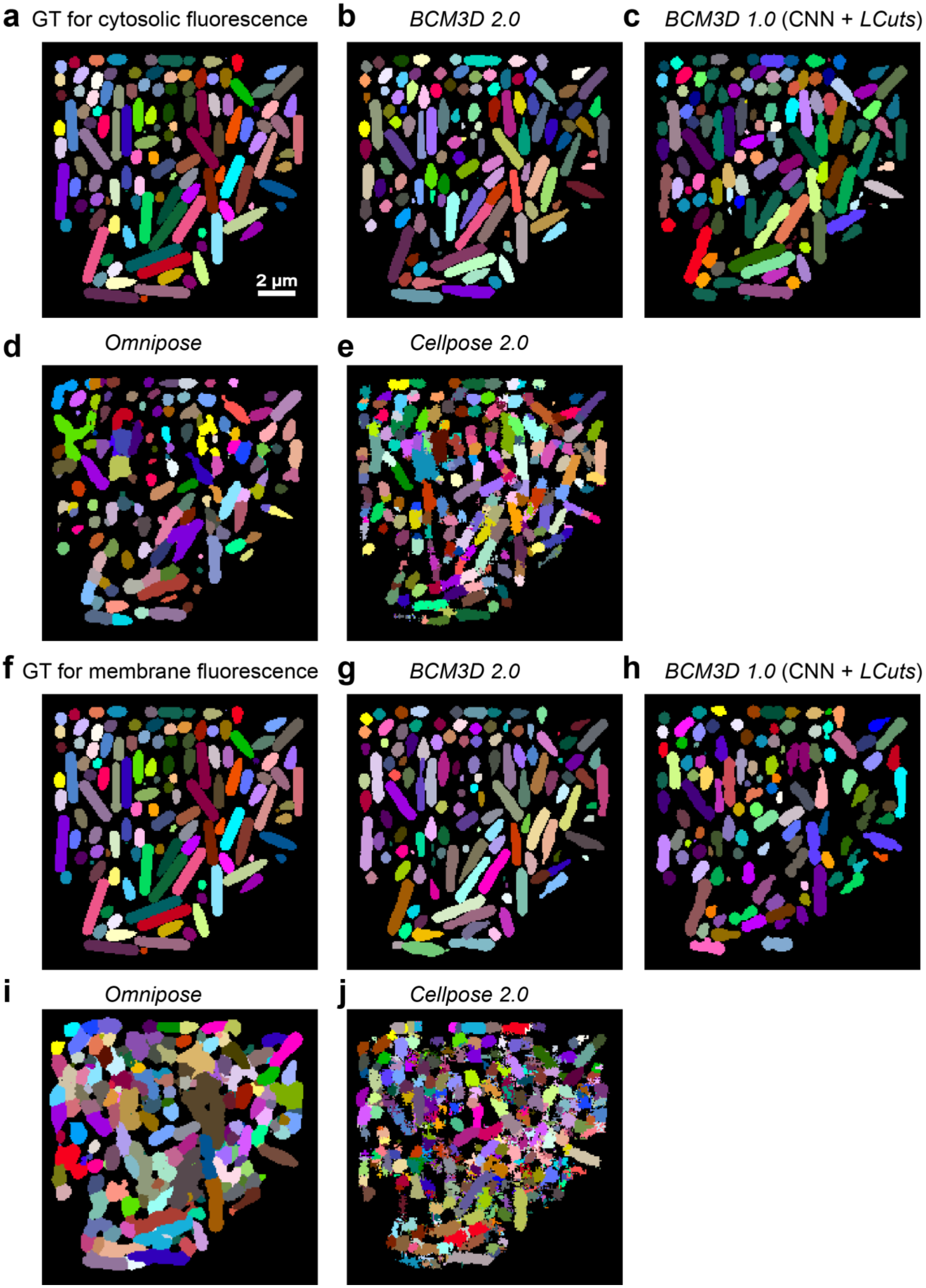

a GT for cytosolic fluorescence
b *BCM3D 2.0*
c *BCM3D 1.0* (CNN + *LCuts*)
2 µm
d *Omnipose*
e *Cellpose 2.0*
f GT for membrane fluorescence
g *BCM3D 2.0*
h *BCM3D 1.0* (CNN + *LCuts*)
i *Omnipose*
j *Cellpose 2.0*

Figure 3.3.2. Visualization of segmentation results for simulated datasets used for Figure 3.3.1e and f. (a, f) 2D cross section through the ground truth cell volumes used to generate the simulated images in **Figure 3.3.1b** and **d**. (b-e) Segmentation results produced by *BCM3D 2.0*, *BCM3D 1.0* (CNN+*LCuts*), *Omnipose*, and *Cellpose 2.0* respectively for cytosol-labeled biofilms. (g-j) Segmentation results produced by *BCM3D 2.0*, *BCM3D 1.0* (CNN+*LCuts*), *Omnipose*, and *Cellpose 2.0* respectively for membrane-labeled biofilms.

Segmentation of experimentally obtained biofilm images

To test the performance of *BCM3D 2.0* on experimental data, we first tested *BCM3D 2.0* on a previously published *E.coli* biofilm image, for which manual annotation masks are available [108]. For this dataset, which features a relatively low cell density, *BCM3D 2.0* performed on par with *BCM3D 1.0* (CNN + LCuts), but again outperformed *Omnipose* and *Cellpose 2* (**Figure 3.3.3)**. (We only considered recently developed 3D instance segmentation approaches for these comparisons. In previous work[108], we established that *BCM3D 1.0* (CNN + *LCuts*) outperformed both *Cellpose* and the segmentation algorithm developed by Hartmann *et al.*) To further demonstrate the versatility of *BCM3D 2.0*, we acquired images of a thick *S. oneidensis* biofilm expressing GFP that has an order of magnitude more cells. Visual inspection of the segmentation results obtained by applying *BCM3D 2.0*, showed physiologically reasonable cell shapes for a majority of segmented objects (**Figure 3.3.4**). For this biofilm image, which has a good SBR, *BCM3D 2.0* outperforms all other approaches, including *Omnipose*, *BCM3D 1.0* (CNN + *LCuts*) and *Cellpose 2.0* (**Figure 3.3.5)**.

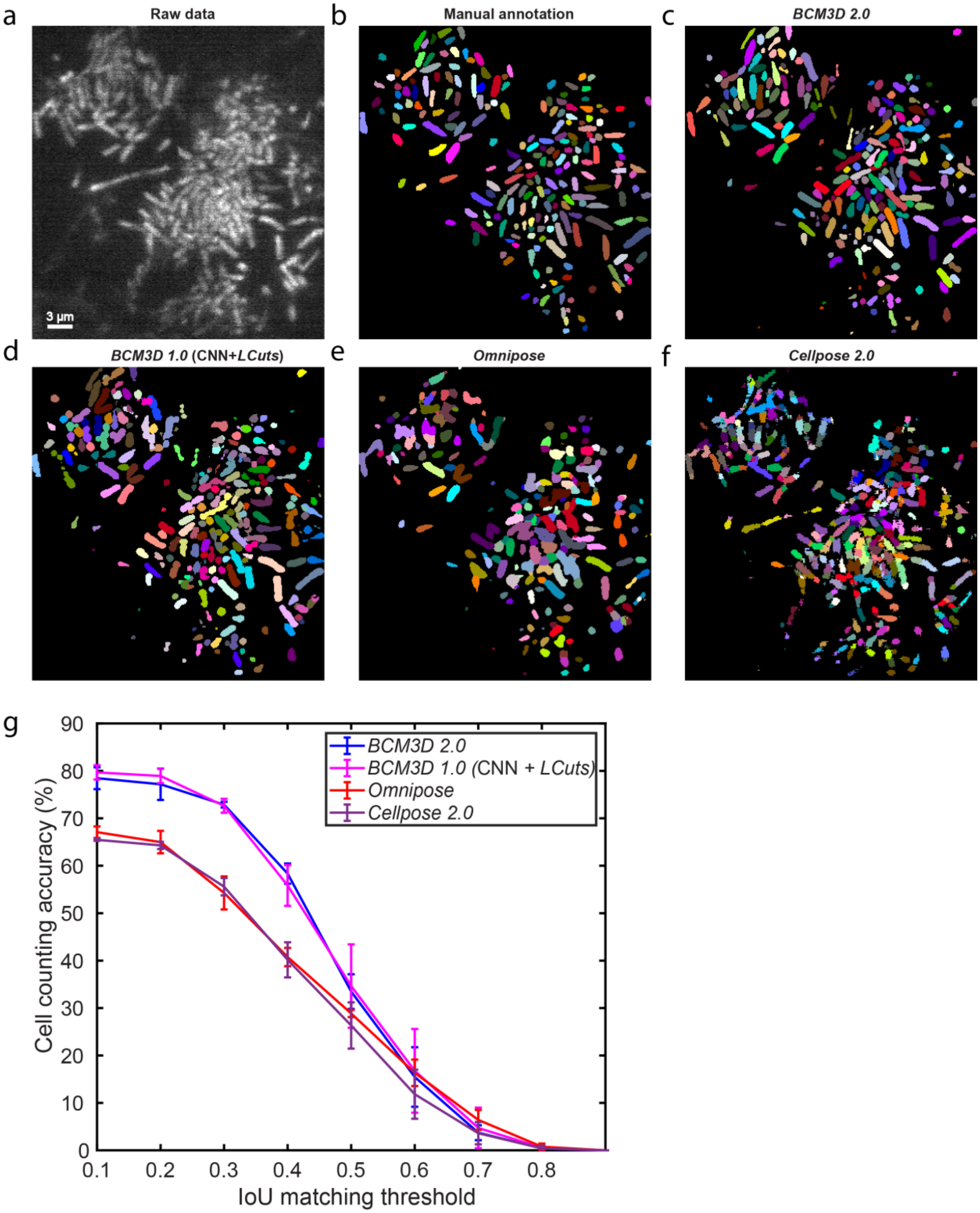


Figure 3.3.3. Comparison of segmentation results on Zhang et al. 2020 $t = 600$ min data.[108] Noted the images have been rotated 90 degrees and flipped horizontally in this figure. (a) Same 2D cross section shown in Zhang et al. 2020 **Figure 4** t = 600 min. (b) Manual annotation result.

(c) Segmentation result produced by *BCM3D 2.0*. (d) Segmentation result produced by *BCM3D 1.0 CNN + LCuts*. (e) Segmentation result produced by *Omnipose* trained from scratch for up to 800 epochs using the same simulated training dataset that *BCM3D 2.0* was trained on. (f) Segmentation result produced by *Cellpose 2.0* fine-tuned for up to 250 epochs using the randomly selected 2D slices of the same simulated training dataset that *BCM3D 2.0* was trained on. (g) Segmentation accuracy is parameterized in terms of cell counting accuracy (y-axis) and IoU matching threshold (x-axis). Each data point is the average of the cell counting accuracies calculated using annotation maps traced by $N = 3$ researchers. Data are presented as mean values ± one standard deviation indicated by error bars.

The above quantitative comparisons were made with reference to manual annotation results made in selected 2D slices. Manual annotation of 3D biofilm images is however too time consuming for thousands of cells distributed in 3D space[108]. We therefore chose to additionally assess the segmentation accuracy using representative morphological observables that are available after segmentation, namely object volume, object solidity (volume fraction of the object as compared to the smallest convex polygon that encloses it), major axis length, longer minor axis length, and the ratio of the two minor axes lengths (longer minor axis divided by the shorter one). We performed principal component analysis (PCA) using these morphological observables and project each segmented object onto a plane spanned by the first two principal components. For simulated data (for which the ground truth is known) this approach shows a distribution for which the correctly segmented objects are concentrated near the origin, whereas the incorrectly segmented objects are predominantly located at the periphery (**Figure 3.3.4a** inset).

We applied the same PCA approach to experimental segmentation results obtained for a S. oneidensis biofilm containing ~3000 cells. Similar to simulated data, most of the segmentation objects cluster near the origin of the two principal component axes (**Figure 3.3.4a**). However, several segmented objects are asymmetrically scattered around the periphery of the distribution. Inspecting the 3D shapes of a few manually selected objects revealed that, consistent with simulated data, physiologically reasonable cell shapes cluster near the center of the distribution, while oddly shaped objects predominantly localize at the periphery. To automatically separate oddly shaped objects from the physiologically reasonable, rod-shape shaped objects, we trained a 3D CNN (independent of BCM3D) with manually validated segmentation objects (obtained from experimental data, see methods). The trained network efficiently (accuracy of 97% on the validation set) separates rod-shaped objects (~82% of total) from oddly shaped objects (~18% of total). This classification enables the display of both subpopulations separately even though they are completely intermixed in 3D space (**Figure 3.3.4bc**). In contrast, Omnipose and Cellpose 2.0 have 40% and 5% rod-shaped objects respectively identified by the shape classifier on this dataset.

We next compared the distributions of solidity and minor axis ratio between rod-shaped and oddly shaped populations. Rod-shaped objects are characterized high values of solidity and minor axis ratio (**Figure 3.3.4de**). In contrast, solidity and minor axis ratio for oddly shaped objects take on values less than one and thus show a much broader distribution (**Figure 3.3.4de** insets). These results show that, when using *BCM3D 2.0*, ~82% of cells are segmented with physiologically reasonable cell shapes. The remaining 18% of cells can then be subjected to further processing to identify and correct the remaining segmentation errors[108, 118, 127] and/or be subjected to further scrutiny to determine whether they are due to aberrant cell shapes exhibited by sick or intoxicated cells[128].

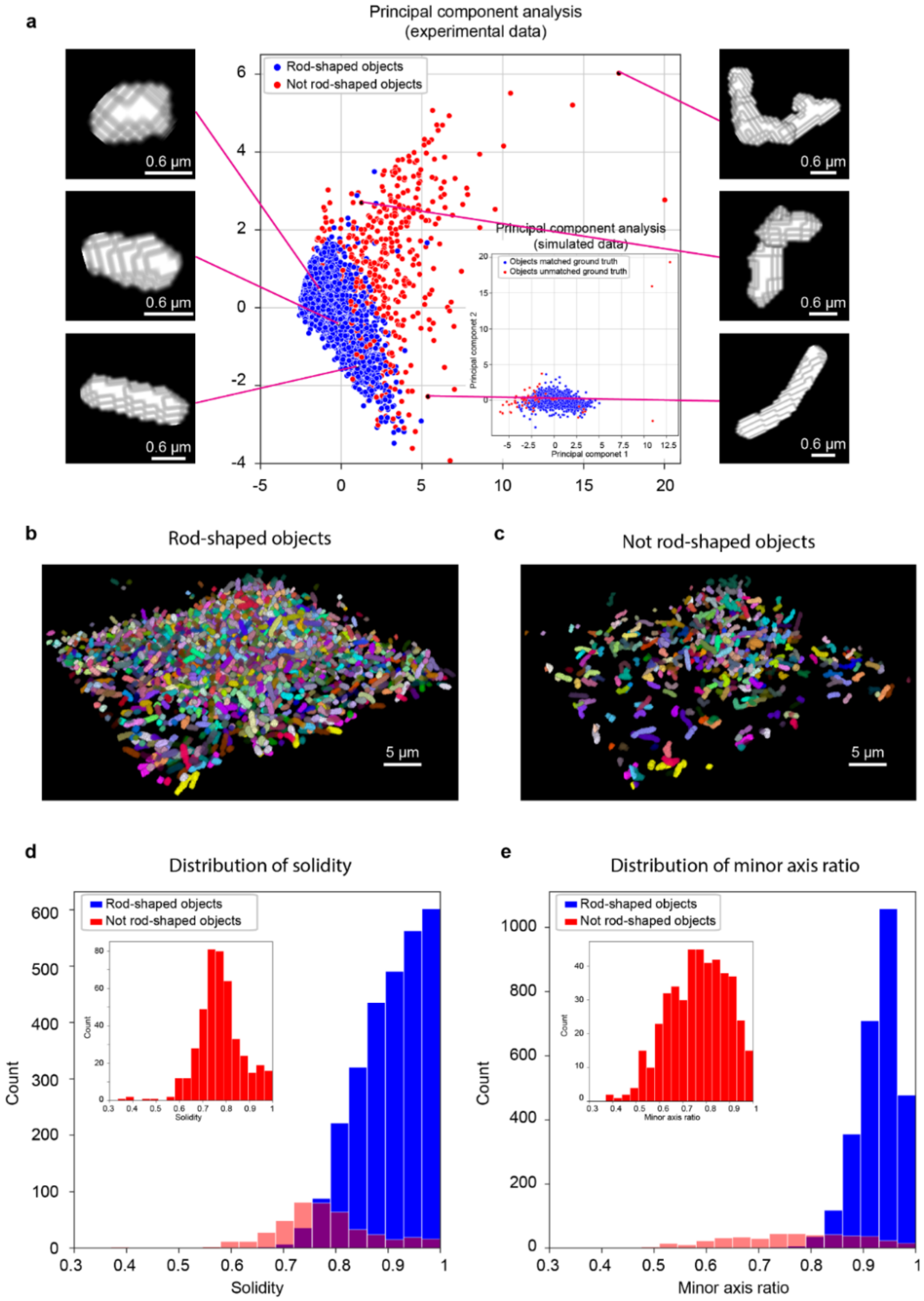


Figure 3.3.4. Performance of *BCM3D 2.0* on experimental biofilm images. (a) Principal component analysis of the segmentation objects (obtained from experimentally acquired images) that were classified by a pre-trained 3D CNN as either physiologically reasonable rod-shaped cells or oddly shaped (not-rod shaped) cells. Three examples cell shapes of each class are shown to the

right and left, respectively. Inset: same analysis on simulated biofilm images. (b) 3D segmentation results for objects classified as physiologically reasonable rod-shaped cells. (c) 3D segmentation results for objects classified as oddly shaped. (d) and (e) Comparison of the solidity and minor axis ratio distributions of rod-shaped and oddly shaped objects.

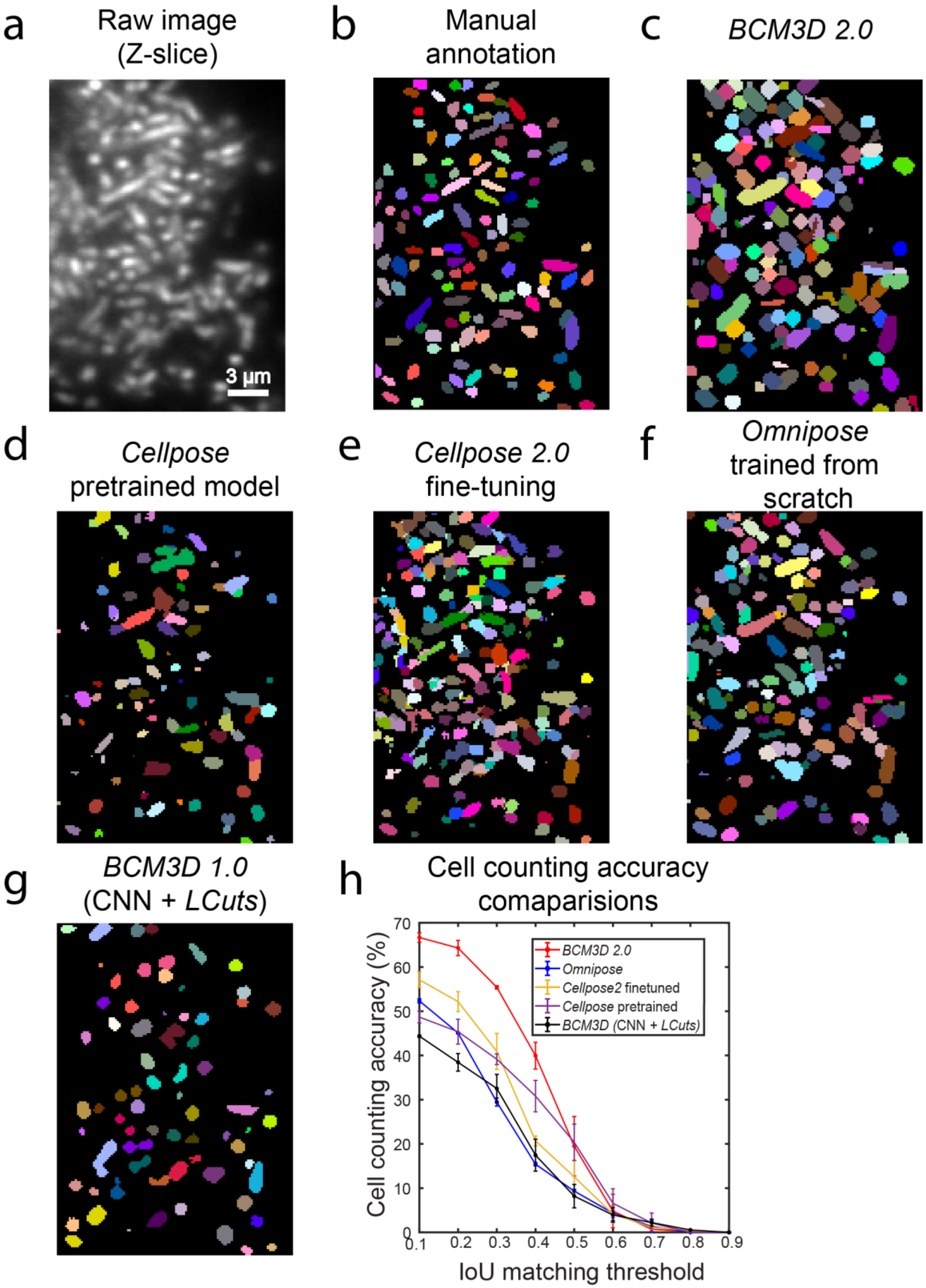

a
Raw image
(Z-slice)
3 µm
b
Manual
annotation
c
BCM3D 2.0
d
Cellpose
pretrained model
e
Cellpose 2.0
fine-tuning
f
Omnipose
trained from
scratch
g
BCM3D 1.0
(CNN + *LCuts*)
h
Cell counting accuracy
comaparisions
BCM3D 2.0
Omnipose
Cellpose2 finetuned
Cellpose pretrained
BCM3D (CNN + *LCuts*)
Cell counting accuracy (%)
0
10
20
30
40
50
60
70
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
IoU matching threshold

Figure 3.3.5. Performance comparisons on *S. oneidensis* dataset in Figure 3.3.4. (a) Small 2D cross section selected from the large 3D biofilm shown in **Figure 3**. (b) Manual annotation result. (c) Segmentation result produced by *BCM3D 2.0*. (d) Segmentation result produced by the pretrained *Cellpose* model. (e) Segmentation result produced by *Cellpose2* fine-tuned using the simulated training data from this work. Cellpose2 can be used for quickly prototyping new specialist models from a generalist model with minimal new training data. We note that *Cellpose2* only uses 2D data for training and fine-tuning, so thirty different 2D slices in all three dimensions of the 3D training data were selected to fine-tune the network. (f) Segmentation result produced by *Omnipose* trained from scratch for up to 800 epochs using the same simulated training dataset that *BCM3D 2.0* was trained on. Validation loss was observed to stop decreasing after approximately 600 epochs. (g) Segmentation result produced by *BCM3D 1.0 CNN + LCuts* (h) Segmentation accuracy is parameterized in terms of cell counting accuracy (y-axis) and IoU matching threshold (x-axis). Each data point is the average of the cell counting accuracies calculated using annotation maps traced by $N = 2$ researchers. Data are presented as mean values ± one standard deviation indicated by error bars.

3.3.4 Discussion

We expanded the *BCM3D* workflow with a complementary CNN-based processing pipeline, named *BCM3D 2.0*, which transfers raw 3D fluorescence images to intermediate image representations that are more amenable to conventional mathematical image processing (specifically, seeded watershed and single- and multi-level Otsu thresholding). Using the *BCM3D 2.0* image processing pipeline, unprecedented segmentation results are obtained, especially for challenging datasets characterized by low SBRs and high cell densities. *BCM3D 2.0* consistently achieves better segmentation accuracy than *Cellpose 2.0* and *Omnipose*, as well as our predecessor algorithm, *BCM3D 1.0*, which represented the previous state-of-the-art for 3D cell segmentation in bacterial biofilms.

We used the segmentation results provided by *BCM3D 2.0* as the input to a nearest neighbor tracking algorithm to explore the possibility of simultaneous multi-cell tracking in 3D biofilms. We found that accurate, automated multi-cell tracking in 3D time-lapse movies is possible with a nearest neighbor tracking algorithm, if the relative cell movement (RM) between consecutive frames is small. Depending on the type of biofilm and the bacterial species, small RM values can be achieved using moderate time resolutions of 1-5 minutes. However, for the motile *S. oneidensis* cells imaged here, a time resolution of 5 minutes was insufficient for automated nearest-neighbor cell tracking in dense biofilm regions. Tracking accuracy is reduced especially if cells undergo large and unpredictable displacements within the biofilm, and when cells associate or dissociate to and from the biofilm. Even so, single-cell observables, such as growth rates and cell division times can still be extracted based on manual tracking establishing that such information is in fact contained in movies acquired with a time resolution of 5 minutes. Because manual cell tracking is not feasible for biofilms containing thousands of cells, future work will have to focus on extracting this information in an automated manner, efficiently and accurately. Machine-learning based solutions will likely prove to be useful in this context.

A clear experimental solution would be to image biofilms at high time resolutions. However, every fluorescence imaging modality is subject to trade-offs between the achievable spatial and temporal resolution, image contrast (SBR), and phototoxicity and photodamage. If reducing the total radiation dose delivered to the cells is an experimental necessity, light sheet-based microscopy approaches offer substantial advantages over confocal microscopy[12].

While *BCM3D 2.0* is capable of segmenting biofilm datasets of lower SBR than previous methods, further modifications to the image processing pipeline may be needed to enable the tracking of extremely light sensitive or highly motile bacterial species. Additional modifications could be made to further improve segmentation accuracy for datasets with even lower SBRs than those successfully segmented here. On the other hand, more sophisticated tracking algorithms could be employed that consider additional features beyond the Euclidian distances between objects. Recently developed deep learning-based cell trackers for both 2D and 3D data[130, 131] are primarily designed for mammalian cells with unique cell shapes. These approaches utilize additional similarity features that inform cell linking across different frames. To what extent such approaches would improve tracking of bacterial cells that have more homogeneous cell shapes remains to be explored. Further benefits may also be gained by utilizing punctate cell labeling schemes[17] or adaptive microscopy approaches, in which higher illumination intensity frames are interspersed with lower illumination intensity frames and the segmentation results in lower SBR frames are informed by the more accurate results obtained in the higher SBR frames.

In summary, the ability to accurately identify and track individual cells in dense 3D biofilms over long periods of time will require the combination of non-invasive fluorescence microscopy approaches for long-term time-lapse imaging and sophisticated image analysis and multi-object tracking tools that provide robust results even for datasets with limited spatial and temporal resolution, and SBR. Here, we have presented an image processing pipeline that enables improved segmentation of dense biofilm-dwelling cells based on 3D fluorescence images of low SBR. The feasibility of simultaneous, automated multi-cell tracking using a simple nearest neighbor tracking algorithm was demonstrated on high time resolution datasets, while manual tracking was possible on lower time resolution datasets. The tools developed here can thus be leveraged to improve our understanding of how coordinated behaviors among biofilm-dwelling cells eventually produce in the macroscopic properties of bacterial biofilms.

3.3.5 Methods

Lattice Light Sheet Microscope Imaging of Bacterial Biofilms

Fluorescence images of bacterial biofilms were acquired on a home-built lattice light sheet microscope (LLSM). LLSM enables specimen illumination with a thin light sheet derived from a 2D optical lattice[26, 87]; here, an intensity uniform light sheet was produced by dithering a square lattice. The average illumination intensity across the light sheet was less than 1 W/cm^2. The sub-micrometer thickness of the light sheet is maintained over a propagation distance of ~30 µm to achieve high resolution, high contrast imaging of 3D specimens comparable to confocal microscopy but with lower concomitant photobleaching and phototoxicity. Widefield fluorescence images of illuminated planes in the specimen are recorded on a sCMOS detector (Hamamatsu ORCA Flash v2). 3D biofilm images were acquired by translating the specimen through the light sheet in 200 nm step sizes using a piezo nano-positioning stage (Mad City Labs, NanoOP100HS). The data acquisition program is written in LabVIEW 2013 (National Instruments).

Kanamycin resistant *S. oneidensis* MR-1, constitutively expressing GFP, were cultured at 30 °C overnight in LB medium with 50 µg/ml Kanamycin. Overnight cultures were diluted 100 times into the same culture medium, grown to an optical density at 600 nm (OD600) of 0.4 – 1.0, and then diluted to OD600 ~ 0.05 using M9 media with 0.05% (W/V) casamino acids. Poly-l-lysine coated round glass coverslips with the diameter of 5 mm were put into a 24-well plate (Falcon) and 400 µl of diluted cell culture was added to the well. Cells were allowed to settle to the bottom of the well and adhere to the coverslip for 1 hour. After the settling period, the coverslip was gently rinsed with M9 media to flush away unattached cells. Then 400 µl of M9 media (0.05% casamino acids) were added to ensure immersion of the coverslips. The well plate was set in a 30 °C chamber for 72-96 hours to allow dense biofilms to develop. Media were exchanged every 24 hours. Before imaging, the coverslip was rinsed again with fresh M9 media. The rinsed coverslip was then mounted onto a sample holder and placed into the LLSM sample-basin filled with M9 media. GFP was excited using 488 nm light sheet excitation. 3D biofilm stacks were acquired by translating the specimen through the light sheet in 200 nm or 235 nm steps. Each 2D slice was acquired with an exposure time of 5 ms or 10 ms.

Samples for time-lapse images were prepared by the same procedures, except imaging was started after either 24-hour or 48-hour cell attachment period, and the imaging experiment was carried out in LM medium (0.02% (W/V) yeast extract, 0.01% (W/V) peptone, 10 mM HEPES (pH 7.4), 10 mM $NaHCO_3$) with a lactate concentration of 0.5 mM.[132]

Raw Data Processing

Raw 3D stacks were deskewed and rotated as described previously[90], but the deconvolution step was omitted. If necessary, background subtraction can be applied to reduce background signal. 3D images were rendered using the 3D Viewer plugin in Fiji[97] or ChimeraX[133]. Sample drift over the course of a time-lapse imaging experiment was corrected by Correct 3D Drift[33], a Fiji plug-in that performs registration by phase correlation, a computationally efficient method to determine translational shifts between images at two different time points.

Generation of simulated biofilm images

Data for CNNs training was computationally generated as described previously[108]. Briefly, CellModeller[91], an individual-based computational model of biofilm growth, was used to simulate growth and division of individual rod-shaped cells in a population (**Figure 3.3.6a**). A minimum distance criterion between cells is imposed at each time point to alleviate cellular collisions that are due to cell growth. We chose cell diameter and cell length (d, l) parameters consistent with the bacterial species under investigation, namely (1 μm, 3 μm) for *E. coli*[92], and (0.6 μm, 2 μm) for *S. oneidensis*[134]. Training data should closely represent the experimental data to ensure optimal segmentation results. Unrepresentative cell diameter and cell length parameters can result in over- or under-segmentation errors and the predictions of non-physiological cell shapes (**Figure 3.3.7**). 3D fluorescence intensity images (**Figure 3.3.6b**) were generated by convolving randomly positioned fluorophores in the cytoplasm or the membranes of simulated cells (**Figure 3.3.6cd**) with experimentally measured point spread functions (PSFs), and then adding experimentally measured background and noise (Poisson detection noise, based on the summed background and signal intensities, as well as Gaussian read noise, experimentally calibrated for our detector at 3.04 photons per pixel on average)[95].

The fluorescence signal intensity in the simulated images was adjusted to match the signal to background ratios (SBRs) of experimentally acquired data. To estimate the SBRs of both simulated and experimental images, we manually selected 10 ‘signal’ and 10 ‘background’ regions in the images using the Oval tool in Fiji and calculated their means respectively. A ‘signal’ region is defined to be any region that contains only pixels within a cell (foreground) and a ‘background’ region contains only pixels outside cells, but in regions that are close to the cells. These regions are judged by the researchers rather than by any computer algorithm to ensure accuracy. The SBR was then calculated by dividing the mean signal intensity by the mean background intensity. Consistent with our previous results (Zhang *et al, Nature Communications*, 2020), the SBR is a good metric quantifying “difficulty to segment” in simulated data, which has homogeneous cells densities and the exact same biofilm architectures (i.e. cell positions). For heterogeneous cell densities in experimental biofilms, the SBR can vary considerably through space. We therefore quantify the SBR in experimental images locally at regions of highest cell density, but we note that this metric should only be used qualitatively and we refrain from making any direct, quantitative comparison of segmentation performance between biofilms of *different* architectures. To estimate the local density of a biofilm, the image was partitioned into several 3D tiles 64 by 64 by 8 voxels in size and the total cell volume contained in each tile was divided by the tile volume. The reported cell density was computed as the average of the 10 densest tiles for each dataset.

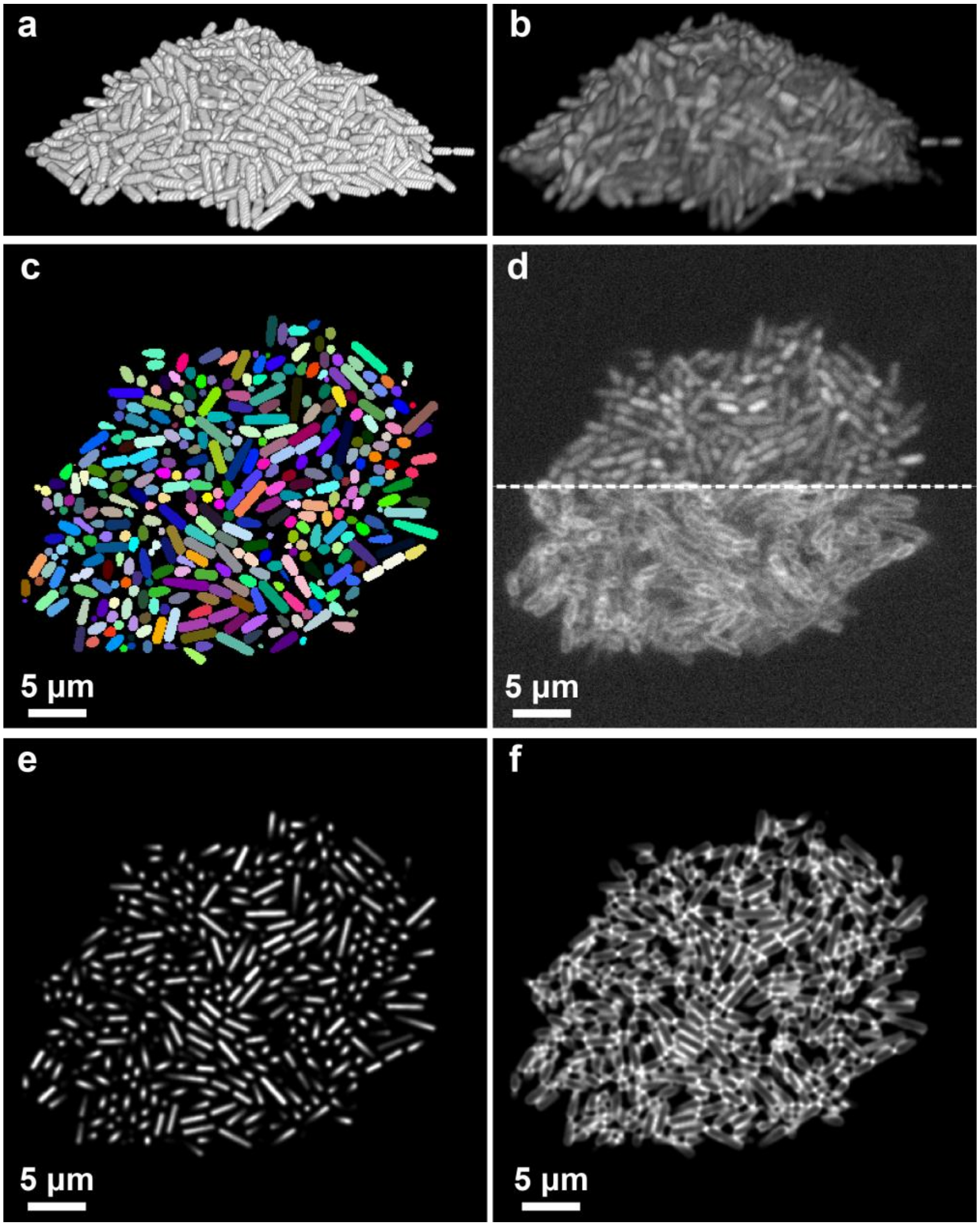


Figure 3.3.6. Simulation of fluorescent biofilms images and intermediate image representations. (a) Cell arrangements obtained by CellModeller. (b) Simulated 3D fluorescence image based on the cell arrangements in a. (c) Ground truth information of a 2D slice. Different cells are shown in different colors and intercellular spaces (background voxels) are displayed in black. (d) 2D slices of the simulated fluorescence image corresponding to the ground truth shown in c. The upper panel shows cells containing cytosolic fluorophores, the lower panel shows cells with fluorescently stained membranes. (e and f) Intermediate image representations generated from the ground truth information shown in c. See text for details.

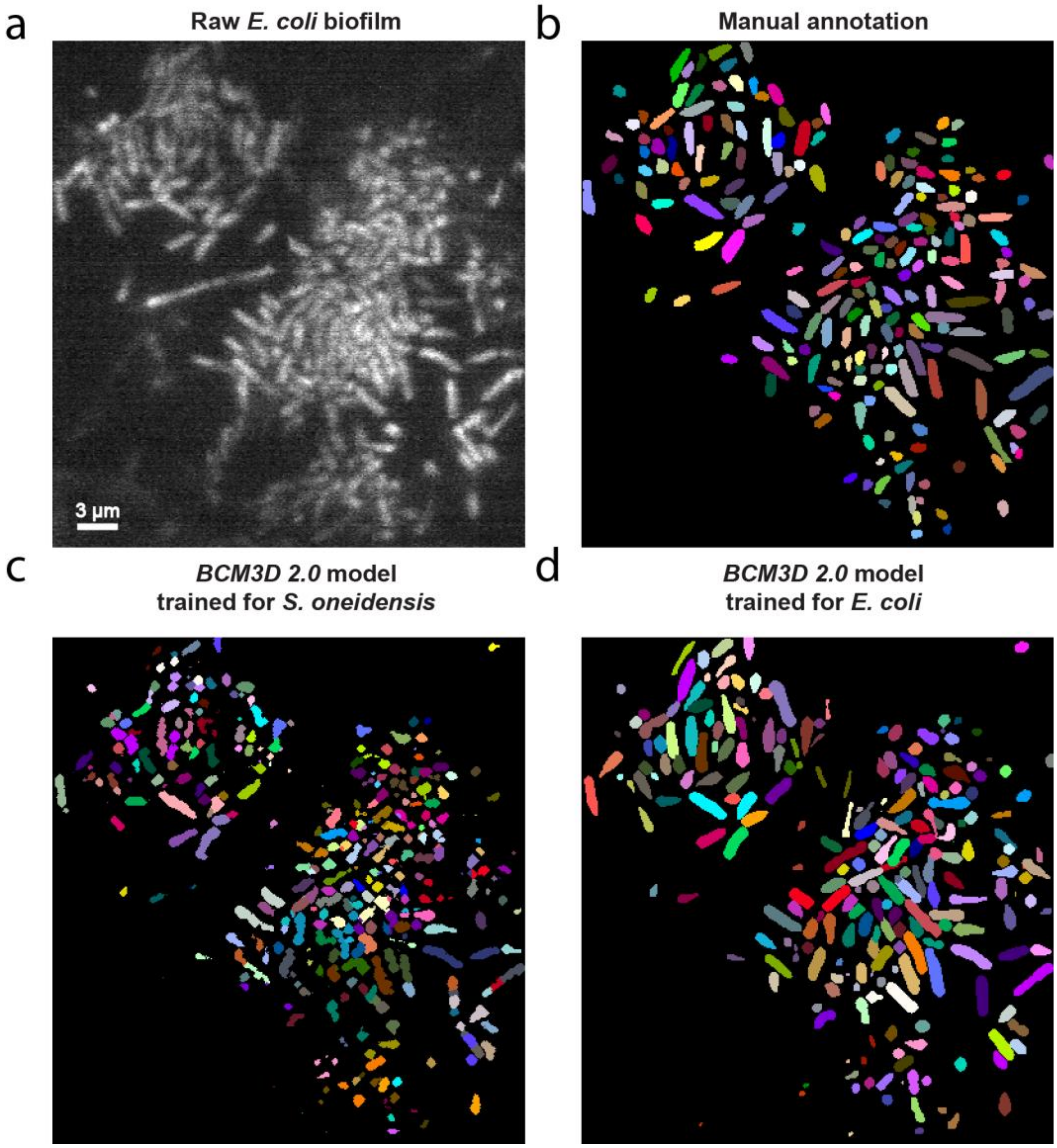


Figure 3.3.7. Segmentation results on Zhang et al. 2020 *t* = 600 min data using different training data simulation parameters.[108] (a) Same 2D cross section shown in Zhang et al. 2020 **Figure 4** t = 600 min. Noted the images have been rotated 90 degrees and flipped horizontally in this figure. (b) Manual annotation result. (c) Segmentation result produced by *BCM3D 2.0* using cell diameter and cell length (d, l) parameters (0.6 µm, 2 µm) in simulation. (d) Segmentation result produced by *BCM3D 2.0* using cell diameter and cell length (d, l) parameters (1 µm, 3µm) in simulation.

Generation of intermediate image representations

To generate ‘distance to nearest cell exterior’ images (**Figure 3.3.6e**, **Figure 3.3.8**) from ground truth data, the Euclidean distance of each voxel inside a cell to the nearest voxel not belonging to that cell was calculated. The so-obtained distances were then normalized to the maximum value of that cell (**Figure 3.3.8c**). In order to obtain a steeper gradient in distance values, the distance values were additionally raised to the third power (**Figure 3.3.8d**), so that the resulting images show highly peaked intensity near the cell center. In a final step, the ‘distance to nearest cell exterior’ images were smoothed by Gaussian blurring (kernel size = 5 voxels in each dimension) (**Figure 3.3.8e**).

To help distinguish touching cells, we calculated a second image representation, the ‘proximity enhanced cell boundary’ image (**Figure 3.3.8**). First, we subtracted the normalized distances to the nearest voxel not belonging to this cell (**Figure 3.3.8c**) from the binary map (**Figure 3.3.8f**). Second, we calculated the inverse of the Euclidean distance of each voxel inside a cell to the nearest voxel belonging to another cell, an intermediate image representation that has been proven useful to prevent objects merging in 2D[121] (**Figure 3.3.8g**). These two intermediate images were then multiplied together (**Figure 3.3.8h**) and small holes in the resulting images (**Figure 3.3.8h inset**) were filled using grayscale closing (**Figure 3.3.8i inset**). The resulting intermediate images provides a complete boundary of an object but also highlights whether the boundary is in close proximity to any other objects. Compared to previous methods that only provide a complete boundary or only provide boundary areas that are close to any other objects, this new intermediate image representation provides a more informative boundary representation. In a final step, the ‘proximity enhanced cell boundary’ images were smoothed by Gaussian blurring (kernel size = 5 voxels in each dimension) (**Figure 3.3.8i**).

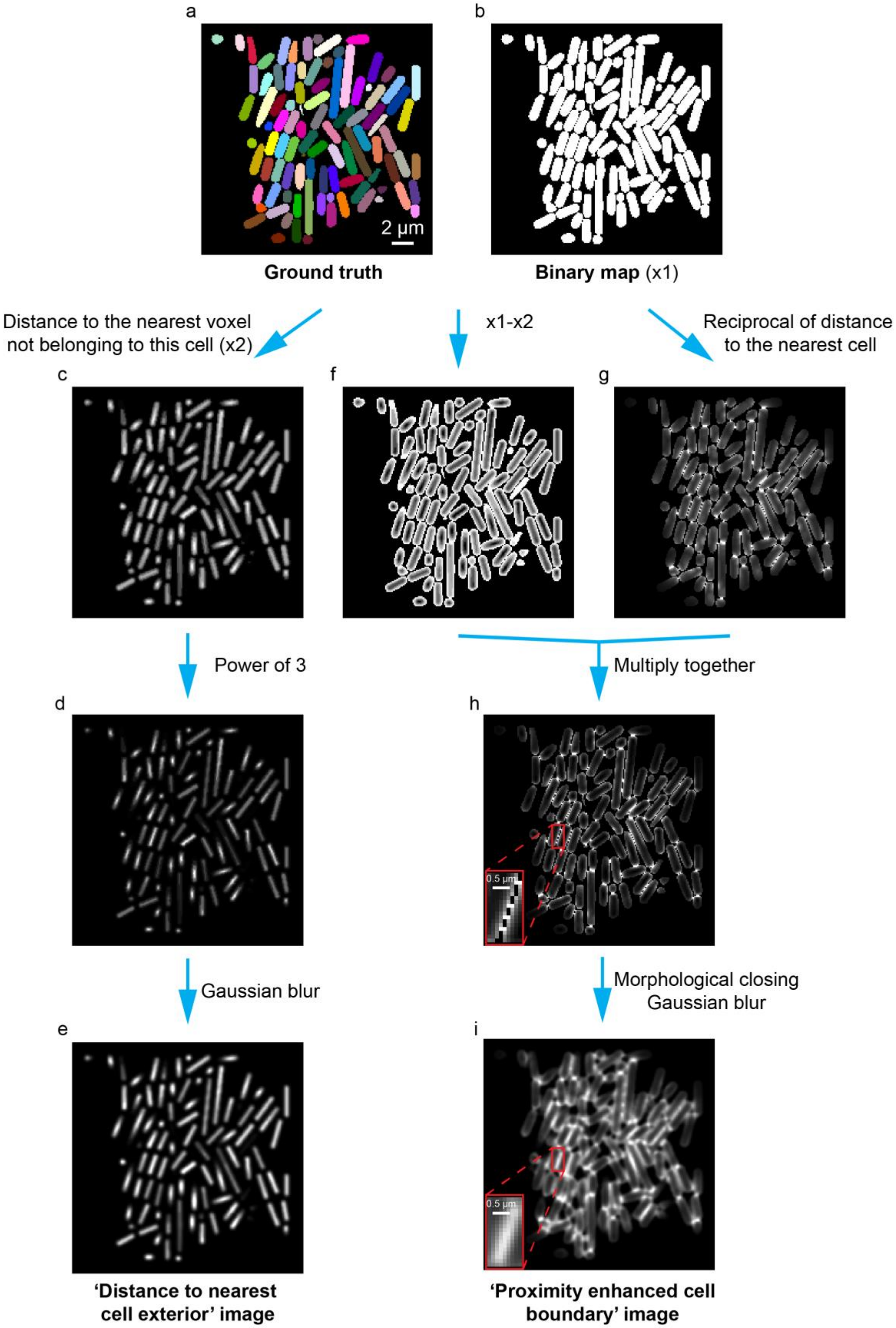

a
b
2 µm
Ground truth
Binary map (x1)
Distance to the nearest voxel not belonging to this cell (x2)
x1-x2
Reciprocal of distance to the nearest cell
c
f
g
Power of 3
Multiply together
d
h
0.5 µm
Gaussian blur
Morphological closing
Gaussian blur
e
i
0.5 µm
'Distance to nearest cell exterior' image
'Proximity enhanced cell boundary' image

Figure 3.3.8. Schematic of generating intermediate image representations. (a) Ground truth cell positions. (b) Binary maps based on the ground truth. (c) Images of distance to the nearest voxel not belonging to this cell. (d) A steeper gradient in distance values is obtained by raising each voxel value in panel c to the third power. (e) Smooth images to get 'distance to nearest cell exterior' image representation. (f) Obtain cell boundary by subtracting (c) from (b). (g) Highlight boundary areas that are close to other cells by calculating reciprocal of distance to the nearest cell. (h) Multiply (f) and (g). The inset shows small holes between two cells' boundary. (i) Small holes (inset) are removed in the 'proximity enhanced cell boundary' image by morphological closing and Gaussian blurring.

Training the convolutional neural network

To generate the above-mentioned intermediate image representations from experimental data, we trained 3D U-Net-based CNNs with residual blocks using the CSBDeep Python package[21]. Residual blocks allow the model to internally predict the residual with regard to inputs for each layer during training. This strategy provides better performance, because solvers are more efficient in solving residual functions than unreferenced functions, and it helps alleviate vanishing or exploding gradients problems for deep neural networks[135]. We employed a network architecture depth of 2, a convolution kernel size of 3, 32 initial feature maps, and a linear activation function in the last layer. Increasing U-Net depth or numbers of initial features didn't produce superior results for our test cases (**Figure 3.3.9**). To achieve robust performance, we trained this network using ten to twenty simulated biofilm images with randomly selected cell densities and signal-to-background ratio. To ensure the broad applicability of these networks, half of these images were biofilms containing cells expressing cytosolic fluorescence and the other half were biofilms containing membrane-stained cells (see **Figure 3.3.6d**). The loss function was taken as the mean absolute error (MAE) between the generated and the target images. The networks were trained for 100 epochs with 100 parameter update steps per epoch and an initial learning rate 0.0004 (**Figure 3.3.10**). The learning rate is reduced by half if the validation loss is not decreasing over 10 epochs. We selected the best weights based on performances on validation set for all the processing steps described below. Using these parameters, it took approximately 1 hour to train the CNNs on a NVIDIA Tesla V100 GPU with 32 GB memory.

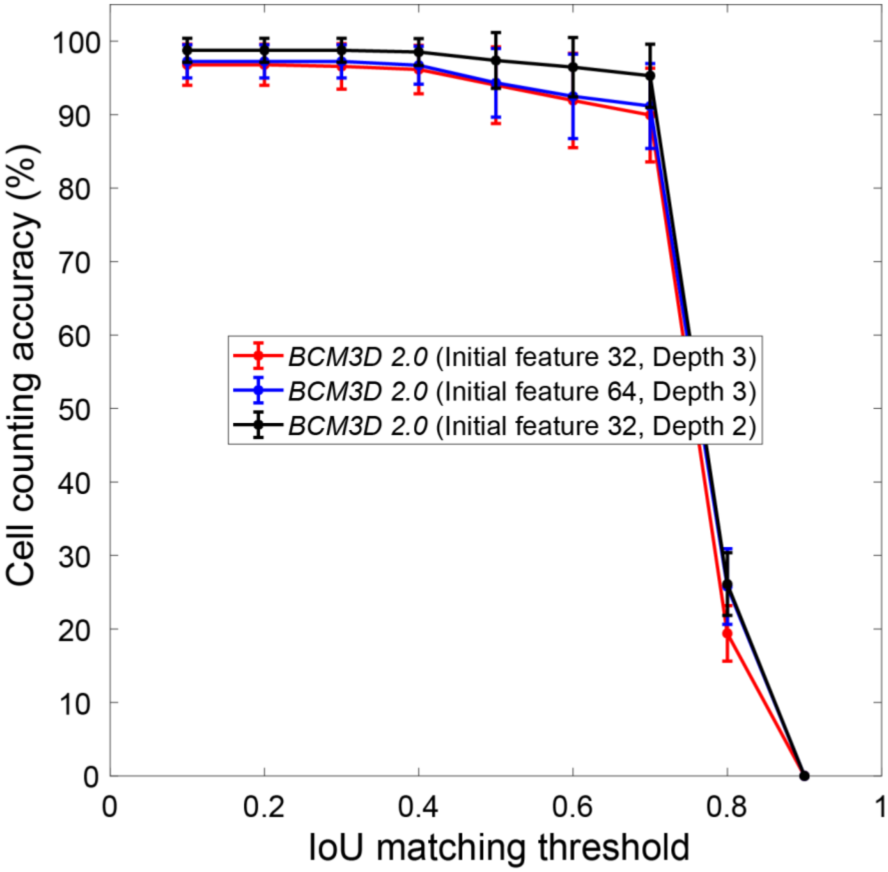


Figure 3.3.9. Comparison of segmentation accuracies achieved by BCM3D 2.0 using different hyperparameter combinations. The networks were trained with U-Net backbone depth 3, 32 initial features, U-Net depth 3, 64 initial features, and U-Net depth 2, 32 initial features respectively. Segmentation accuracy is parameterized in terms of cell counting accuracy (y axis) and IoU matching threshold (x axis). Each data point is the average of N=10 independent biofilm images. Data are presented as mean values ± one standard deviation.

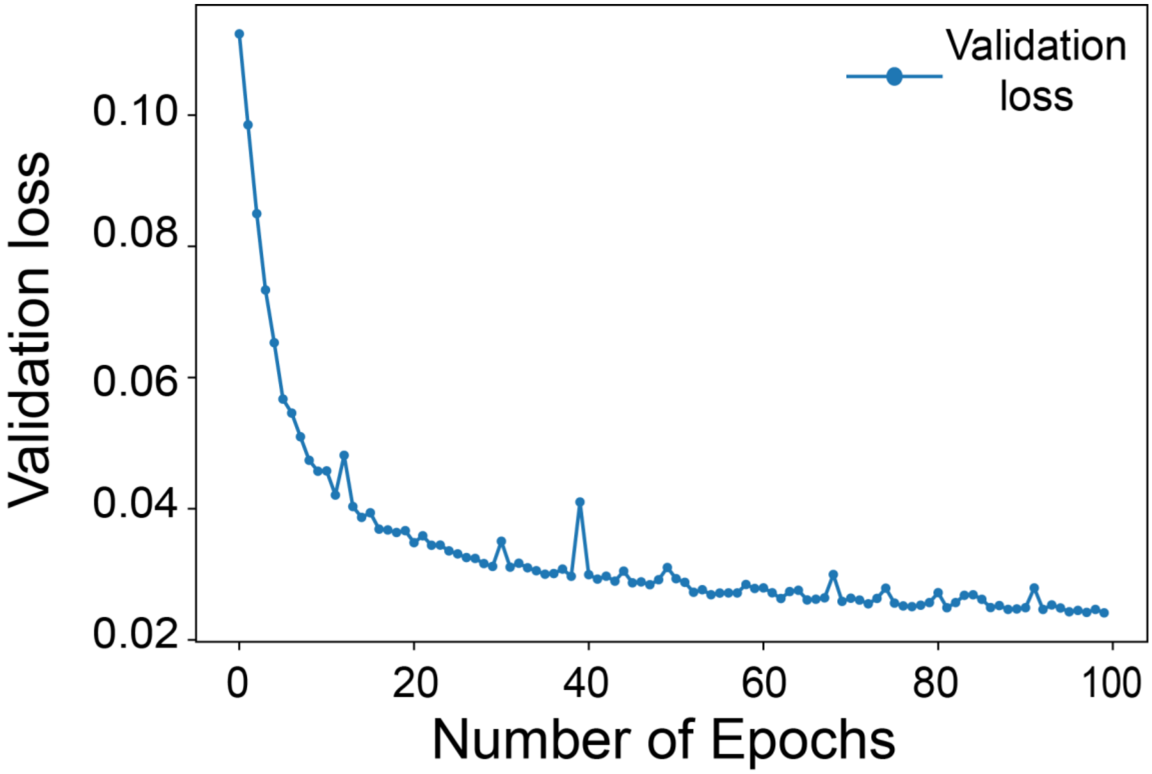


Figure 3.3.10. Validation loss over number of epochs in training. A U-Net network architecture depth of 2, a convolution kernel size of 3, 32 initial feature maps was trained for 100 epochs. Twenty simulated biofilm images with randomly selected cell densities and signal-to-background ratio were used to train the model.

To obtain instance segmentation results from intermediate image representations predicted by trained CNNs, we applied single- and multilevel Otsu thresholding[123], [136], and seeded watershed[122] (scikit-image Python library[137], **Figure 3.3.11** and **3.3.12**).

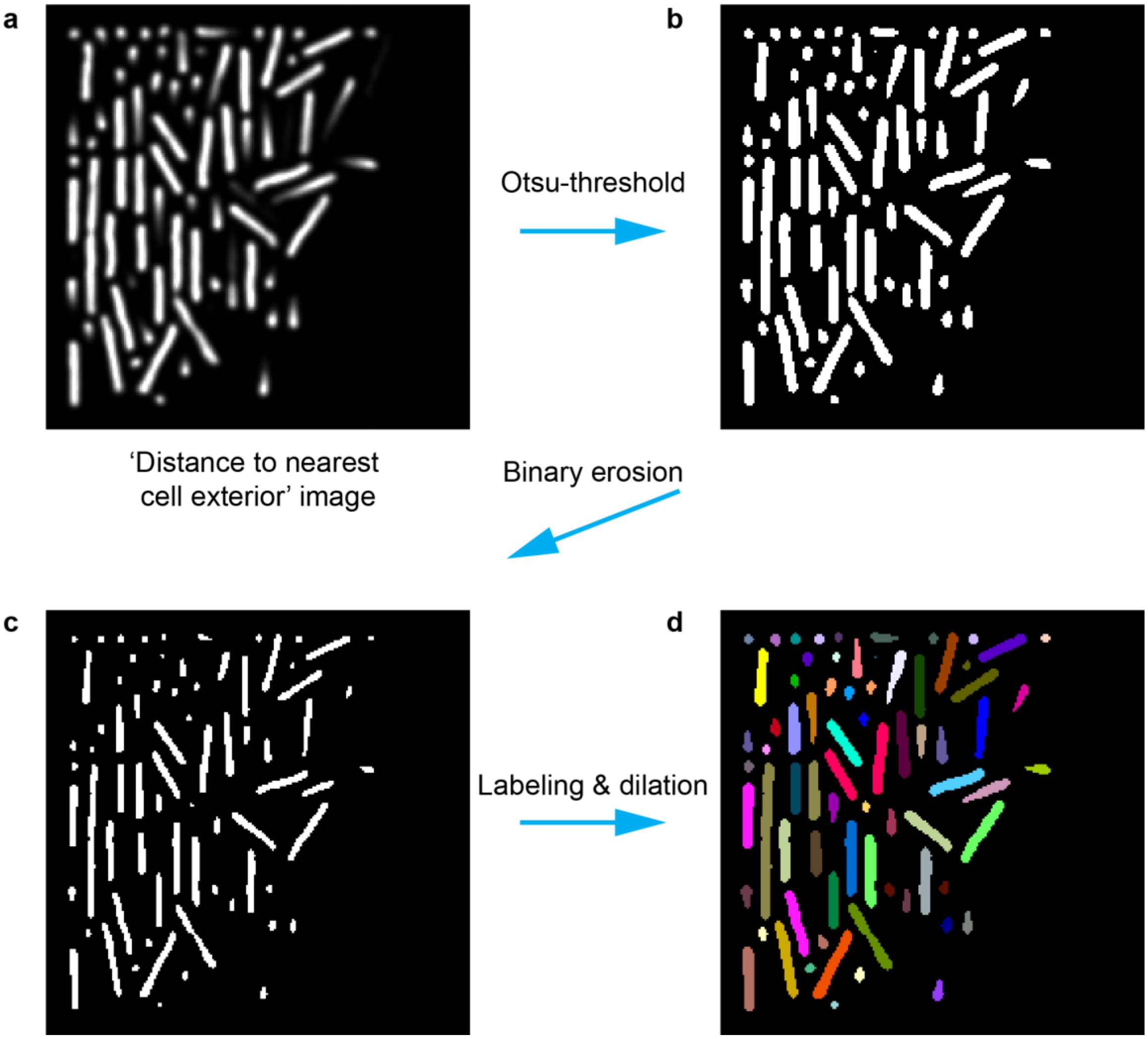


Figure 3.3.11. Post-processing step 1: Thresholding of CNN-produced 'distance to nearest cell exterior' image. (a) CNN-produced 'distance to the nearest cell exterior' image. (b) Apply Otsu-threshold to obtain binary images. (c) To split clusters that are only connected by one or two voxels, the boundary voxels of each object were set to zero (binary erosion). (d) Identify individual cell objects by labeling connected regions and then add back boundary voxels that were erased in panel c.

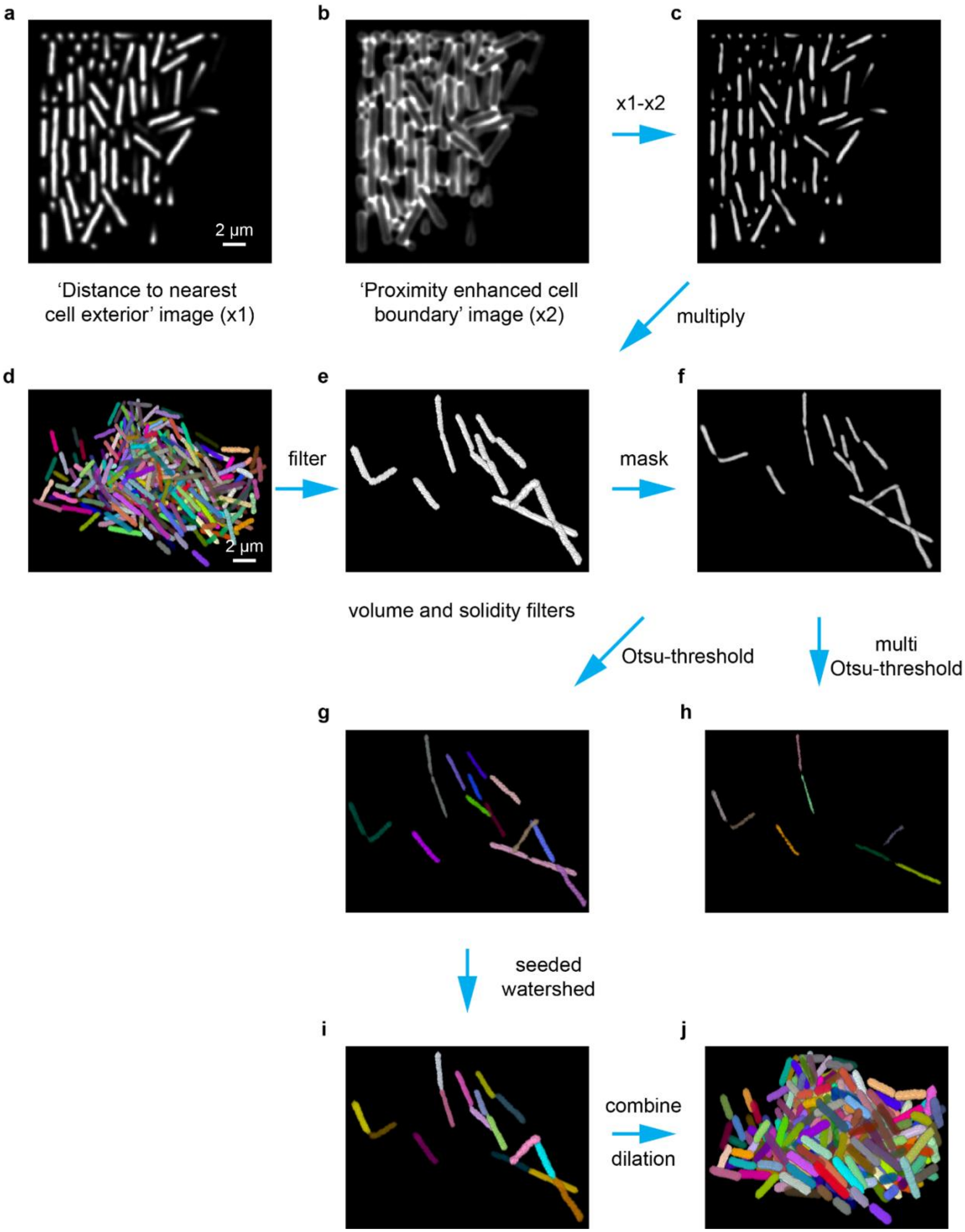

a
b
c
2 µm
x1-x2
'Distance to nearest cell exterior' image (x1)
'Proximity enhanced cell boundary' image (x2)
multiply
d
e
f
2 µm
filter
mask
volume and solidity filters
Otsu-threshold
multi Otsu-threshold
g
h
seeded watershed
i
j
combine
dilation

Figure 3.3.12. Post-processing step 2: Combining of intermediate image representations. (a) CNN-produced 'distance to nearest cell exterior' image. (b) CNN-produced 'proximity enhanced cell boundary' image. (c) Generate difference map by subtracting (b) from (a) and set negative values to zero. (d) Segmentation results from the process describe in **Figure S3.3.11**. (e) Identify objects that need further processing using volume and solidity filters (see Methods). Identified objects are outlined in a binary mask. (f) Mask image in panel c by multiplying it with image in panel e. (g) To split under-segmented clusters, apply seeded-watershed to image in panel f. Seeds are obtained by applying Otsu-thresholding to the image in panel f. (h) If there still are under-segmented clusters in the image in panel g, additional seeds are obtained by applying multi-level Otsu-thresholding the image in panel f. (i) Combine segmented objects from images in panels g and h. (j) Combine (i) and (d) to get final segmentation results.

To test whether segmentation objects have physiologically reasonable cell shapes, we separately trained a 3D CNNs based classification model using tensorflow 2.0. We adapted a network architecture from Zunair et.al.,[138]; mainly includes three 3D convolutional layers, one global average pooling layer and a sigmoid activation function in the last layer. To achieve robust performance, we trained this network using 733 manually confirmed segmentation objects from experimental data (411 reasonable shaped objects, 322 oddly shaped objects). Training data were augmented by rotation and flip. The loss function was taken as the binary cross entropy between the model output and the corresponding target value. The networks were trained for 100 epochs with a batch size of 5 and an initial learning rate 0.0002. The learning rate is reduced by a half if the validation loss is not decreasing over 15 epochs. Using these parameters, it took approximately 17 mins to train the CNNs on a NVIDIA Tesla V100 GPU with 32 GB memory.

Performance evaluation

Segmentation accuracy was quantified as cell counting accuracy and cell shape estimation accuracy. The cell counting accuracy (CA) was calculated as previously described[108]:

$$CA = \frac{TP}{TP + FP + FN}$$

where, TP is the number of true positives, FP is the number of false positives, and FN is the number of false negatives. Cell shape estimation is evaluated by two separate measures. Single-cell segmentation accuracy (SSA) takes the mean Intersection-over-Union (IoU) value (aka the Jaccard index[139]) over segments that have a matching ground truth/manual annotation object:

$$SSA = \frac{1}{N_{match}} \sum_{i}^{N_{match}} \frac{|Seg_i \cap GT_i|}{|Seg_i \cup GT_i|}$$

where, $|Seg_i \cap GT_i|$ is volume of overlap between the predicted object and the ground truth object, and $|Seg_i \cup GT_i|$ is the volume enclosed by both the predicted object and the ground-truth object. We note that the SSA metric can take on high values even if the shape of a segmented object does not accurately represent the shape of the corresponding ground truth object. For example, a predicted round object with a diameter of 20 covered by a ground truth square object with a length of 20 gives a 0.8 IoU value, which could be interpreted as good performance. From a biological perspective however, this would signify a substantial inaccuracy in shape estimation. To measure differences in cell shape in a more discriminating way, we additionally computed a single-cell boundary F1 score (SBF1)[127]. The SBF1 of the abovementioned square vs circular object example is 0.67. The SBF1 score is computed as

$$SBF1 = \frac{1}{N_{match}} \sum_{i}^{N_{match}} \frac{2 \cdot prcecision_i \cdot recall_i}{prcecision_i + recall_i}$$

where precision is the ratio of matching boundary points in a matched segmentation object to the total points of its boundary. Similarly, recall is the ratio of the matching boundary points to the total points of ground truth boundary. According to the definition of boundary F1 score[140], a distance error tolerance is used to decide whether a point on the predicted boundary has a match on the ground truth boundary. For our 3D data, we use $\sqrt{3}$ voxels.

3.3.6 Acknowledgements

Chapter 3.3 was adapted from Zhang & Wang et al. 2022[141].This work was supported in part by the US National Institute of General Medical Sciences Grant 1R01GM139002 (A.G.) and by a grant from the Trans University Microbiome Initiative (TUMI) at the University of Virginia.

Chapter 4: Simultaneous multi-cell tracking and lineage tracing in bacterial biofilms

4.1 Overview

Multi-object (multi-cell) tracking in biology involves the detection, identification, and tracking of multiple objects, such as cells, organisms, or particles, over time in biological systems. It plays a crucial role in various research areas, including cell biology, neuroscience, microbiology, and ecology. By accurately tracking and analyzing the movements and interactions of multiple objects, researchers can gain valuable insights into dynamic biological processes and phenomena.

Multi-object tracking in biology often involves the integration of various technologies, including imaging systems, computer vision algorithms, and computational modeling. High-resolution microscopy such as LLSM allows researchers to capture time-lapse images of biological systems, providing the data necessary for tracking analysis. Computer vision algorithms and machine learning techniques are employed to detect and track objects of interest, distinguishing them from the background and other objects in the field of view. Computational modeling and data analysis techniques help interpret the tracked object trajectories, quantify their movements, and extract relevant features or parameters for further analysis.

One of the primary goals of multi-object tracking in bacterial biofilms is to understand the behavior and characteristics of individual cells within a complex, heterogeneous structure. This includes studying cell migration, cell-cell interactions, and lineage tracing as well as other processes that occur within living bacterial biofilms. By tracking multiple cells simultaneously, researchers can uncover patterns, trajectories, and correlations that provide insights into underlying biofilm formation mechanisms.

4.2 Simultaneous multi-cell tracking and lineage tracing using a nearest neighbor approach

4.2.1 Tracking accuracy is linearly correlated with segmentation accuracy

A nearest neighbor tracking algorithm is a simple yet effective tracking algorithm that is based on minimizing the distance travelled globally by all cells. A nearest neighbor approach falls into the tracking-by-detection paradigm, in which tracking relies on detection of the objects first. For such method, it is important to understand the relationship between tracking accuracy and detection(segmentation) accuracy. Intuitively, tracking accuracy should depend on segmentation accuracy. More specifically, it is hypothesized that the number of edge operations in the *AOGM* metric is linearly correlated with the number of missing objects or falsely detected objects. The edge deletion (*ED*), addition (*EA*) terms are closely related to oversegmentation and undersegmentation respectively: an oversegmentation mistake is likely to result in an edge deletion, while an undersegmentation mistake will add one or more edge additions. To test such a hypothesis, I tracked individual cells in simulated biofilm images with different cell densities and SBRs. If only the Euclidean center-of-mass distance between two cells is considered in the tracking algorithm, then tracking accuracy improves linearly with cell counting accuracy (**Figure 4.2.1**). By including additional terms such as cell morphology, it is hypothesized that improved tracking accuracies can be obtained even if segmentation errors occur frequently or low frame acquisition rates are employed.

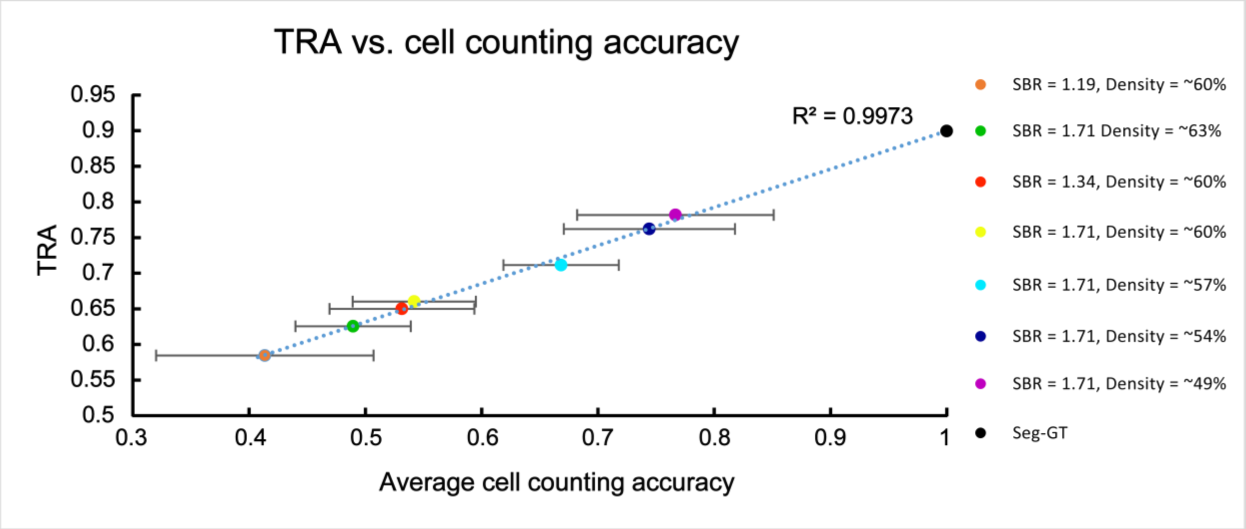


Figure 4.2.1. Tracking accuracy (*TRA*) over cell counting accuracy on simulated biofilms.

4.2.2 Accurate BCM3D 2.0 segmentation enables multi-cell tracking in biofilms.

Simultaneous multi-cell tracking and lineage tracing is critical for analyzing single-cell behaviors in bacterial biofilms. We asked whether the cell segmentation performance of *BCM3D 2.0* was sufficient to enable accurate tracking of individual cells in biofilms. To address this question, we employed a tracking-by-detection approach using simulated biofilm images of different SBRs (**Figure 4.2.2a**). We evaluated tracking accuracy as a function of SBR using the widely used *TRA* metrics based on Acyclic Oriented Graph Matching (*AOGM*)[142]. In acyclic oriented graphs, cells in different time frame are represented as vertices and linkages between cells from frame-to-frame are represented as edges. When the cells (vertices) are placed at their actual spatial coordinates, then the cell linkages (edges) represent the branches of a spatially resolved lineage tree (**Figure 4.2.2b**). The *TRA* metrics quantify the minimum number of elementary graph operations that are needed to transform an estimated graph into a ground truth graph. *TRA_edge* considers three edge operations only, while *TRA_full* considers all six graph operations[142].

To link the same cells across two different time points, we used a nearest neighbor algorithm[143]. When using spatial distance as the sole metric for cell linking, the *AOGM* tracking accuracy has a positive correlation with SBR (**Figure 4.2.2c**), which highlights the importance of accurate cell segmentation in multi-object tracking-by-detection[60]. *BCM3D 2.0* enables a tracking accuracy that is similar to the ground truth tracking accuracy (same nearest neighbor tracking algorithm applied to the ground truth segmentation masks) for SBRs of 1.65 and higher. We note that, given the high cell density in this test dataset, the ground truth tracking accuracy does not reach the optimum (100%) even with error-free segmentation. This is due to inherent limitations in how parent-daughter relationships are assigned. At SBR's less than 1.65, tracking accuracy decreases rapidly due to the lack of consistent segmentation results. The importance of accurate segmentation is clearly evident from the linear dependence of *TRA* as a function of cell counting accuracy (**Figure 4.2.2d**).

Another key factor for simultaneous multi-object tracking is the time resolution[60]. The relative movement (RM) of objects from frame to frame is therefore a useful metric to quantify the level of difficulty for cell tracking. The relative movement ($RM_{i,j}$) in time frame *i*, for a given cell *j* is defined as the ratio between the distance of cell *j* to itself between frame i and *i*+1 and the distance of cell *j* in frame *i* to its closest neighbor at frame *i* + 1. The <*RM*> metric is then the average $RM_{i,j}$ of all cells for each frame[144]. A dataset with <*RM*> values of 1 or more means that any tracking method that considers only distance (and distance related features) is likely to fail, whereas a dataset with a <*RM*> value of less than 0.5 is considered challenging[144]. For the simulated biofilm images here, *RM*~0.2, which indicates that the time resolution may be good enough for single cell tracking using a nearest neighbor algorithm. Indeed, under these conditions, many cells can be tracked for several generations (**Figure 4.2.2b**). However, even at *RM*~0.2, some cell division events are missed, so that a few branches of the lineage tree are not successfully traced. Even so, the subset of correctly detected cell division events allows for the estimation single-cell doubling cycles in the biofilm (**Figure 4.2.2e**). To quantify these trends, we tested how time resolution affects tracking accuracy. When the time resolution is decreased by a factor of two and three, the *TRA_edge* metrics decrease from 91% to 87% and 81%, respectively. The percentage of the parent-daughter misassignment error, quantified as the edge-correction (EC) error over the number of total errors, increases from 1.4% to 3.6 and to 5.2 % (**Figure 4.2.2f**). Taken together, these results show that segmentation based multi-object tracking accuracy is highly dependent on segmentation accuracy (which depends on image SBR and cell density[108]), as well as time resolution. It is critical to consider these parameters, when single-cell resolved observables, such as cell trajectories, single cell volume increases, and single-cell doubling times, need to be measured.

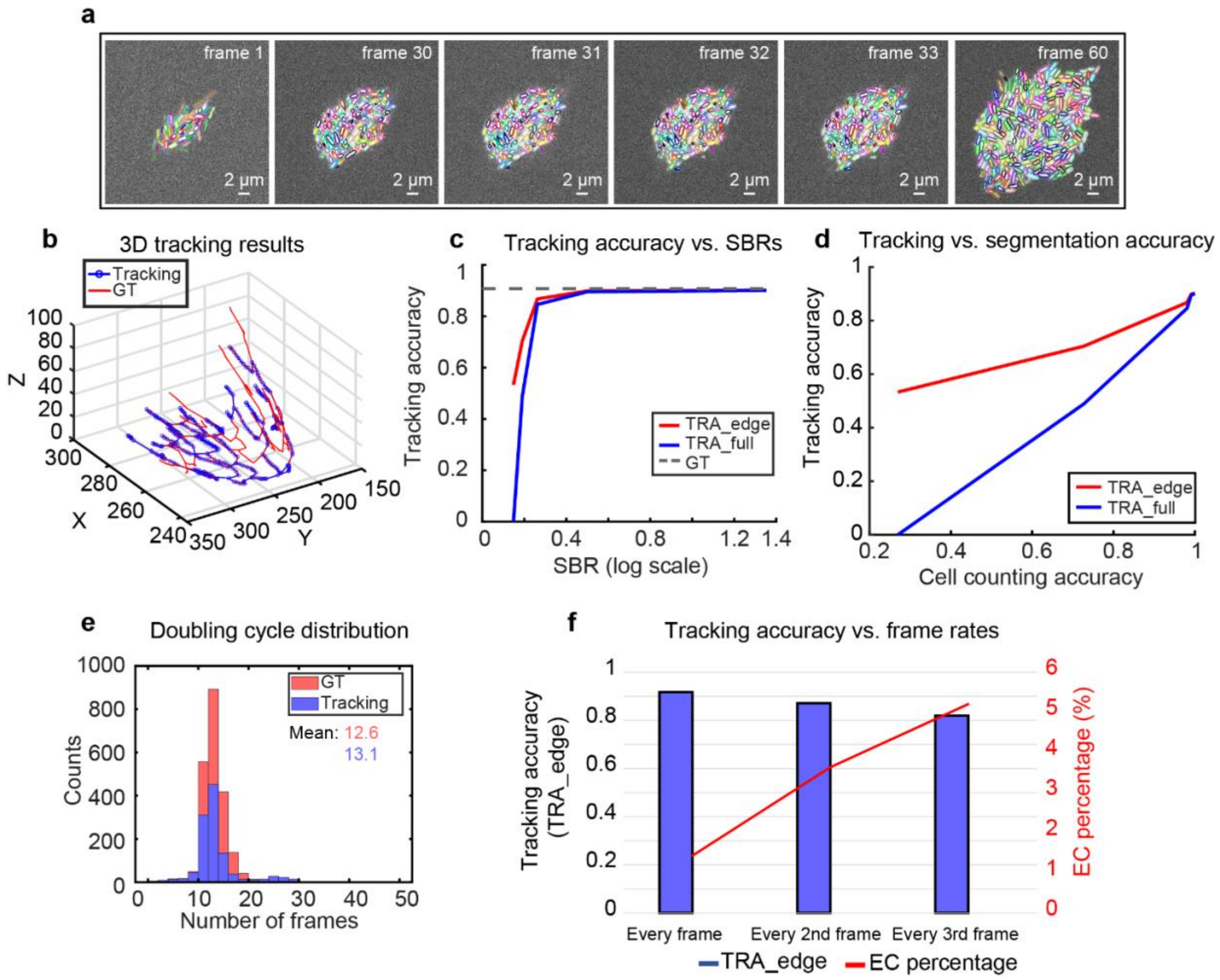


Figure 4.2.2. Multi-cell tracking in simulated biofilms. (a) Simulated fluorescence time-lapse images of growing *E. coli*-like biofilm. The SBRs of these images are ~1.65. Contours are color-coded based on segmentation and tracking results. (b) An example of 3D tracking and lineage tracing for simulated biofilm images. For clarity, spatial trajectories and lineages originating from only a single ancestor cell is displayed. The estimated graph is shown in blue and the corresponding ground truth graph is shown in red. The entire biofilm contains over sixty graphs of this type. (c) Two AOGM metrics calculated as *TRA_edge* and *TRA_full* are plotted against image SBR. The grey dashed line indicates tracking of the GT segmentation using the same tracking algorithm. (d) Same data as in panel c plotted as a function of cell counting accuracy at IoU = 0.5, a segmentation accuracy metric that increases for increasing SBR in the raw images[108]. (e) Doubling cycle distribution of simulated data and corresponding tracking results. A completed cell cycle is defined as a track in which the parent cell is able to split twice. This threshold results in a lower count numbers of estimated cell division, but does not alter the shape of the distribution. (f) *TRA_edge* (left axis) and edge correction (EC) percentage (right axis) for different temporal sampling rates. EC percentage indicates how many parent-daughter relationships are misassigned based on the tracking results.

4.2.3 Multi-cell tracking in the initial phase of *S. oneidensis* biofilm

Cell segmentation and subsequent multi-cell tracking in experimentally acquired 3D images presents additional challenges that were not modeled in the computationally simulated data. These challenges include optical aberrations in the imaging system, broader cell shape distributions in experimental biofilms, cell motility, and association and dissociation dynamics of individual cells to and from the biofilm. To determine whether the *BCM3D 2.0* segmentation results enable improved multi-cell tracking using a nearest neighbor algorithm, we manually traced a subset of ancestor cells over the course of a 15-minute 3D biofilm movie acquired with a time resolution of 30 seconds (**Figure 4.2.2ab**). Manual determination of cell-to-cell correspondences in consecutive image volumes generated 583 cell-cell and 3 parent-daughter linkages. Taking this manual annotation as the reference graph, the *RM* metric was determined to be ~0.2 and the *TRA_edge* metric was determined to be 93.5%. Steadily increasing single cell volumes for four selected cells allowed us to measure growth rates of 7.4×10^{-3}, 3.8×10^{-3}, 3.4×10^{-3}, and 0.6×10^{-3} μm^3/min (**Figure 4.2.2c**). Cell division events are also readily detected by the algorithm as a sudden decrease in cell volume. In two of the four selected cases, cell division led to the dispersal of the daughter cell. We found a high number of cell dispersion events resulting in the termination of trajectories, most often right after cell division (**Figure 4.2.2c**).

Although *BCM3D 2.0* in combination with high-frame rate imaging enables accurate cell tracking, it may not be feasible to maintain high-frame rate volumetric imaging for extended periods of time due to phototoxicity and photobleaching concerns. To further test the limits of nearest neighbor tracking, we tracked *S. oneidensis* biofilm growth for five hours at a time with a resolution of 5 minutes (**Figure 4.2.2de**). During this time period, the number of cells increases from ~300 to ~1400 cells. The relative cell motion in this dataset, estimated by the distances between manually tracked cell centroids, is 0.5 ± 0.2 µm (mean ± standard deviation, $N = 5$). The average spacing of the biofilm, as calculated by the average distance of each cell to the nearest neighbor, changes over time. The average spacing for the first frame, estimated by the average distance to the nearest neighbor for each cell, is 1.2 ± 0.6 µm (mean ± standard deviation, $N \sim 300$), and is 1.1 ± 0.2 µm (mean ± standard deviation, $N \sim 1400$) for the last frame. We manually traced a subset of founder cells over the course of the experiment, generating 262 cell-cell and 17 parent-daughter linkages. For this manually selected subset, the *RM* metric was ~0.4 and *TRA_edge* metric was determined to be 80.0%. While the nearest neighbor tracking algorithm is capable of making overall accurate cell-cell linkages for a few consecutive frames, automated nearest-neighbor tracking of the same cells for long time periods and correctly detecting all cell-division events is not readily possible (**Figure 4.2.2f**). It is, however, possible for human annotators to track individual cells from the segmentation results under such imaging conditions. Single-cell growth rates and single-cell division times can then be readily extracted (**Figure 4.2.2f, Figure 4.2.3,**). The measured growth rates are in excellent agreement with the values obtained with high time resolution imaging (**Figure 4.2.2c**). A small number of segmentation errors can be detected by manual tracking, as indicated by the boxes in **Figure 4.2.2f**, but these errors don’t preclude estimations of single-cell observables. These results indicate that quantitate information about single-cell behaviors is contained even in low time resolution 3D movies of bacterial biofilms. Future work will need to focus on extracting the information as accurately as, but faster than, a human annotator.

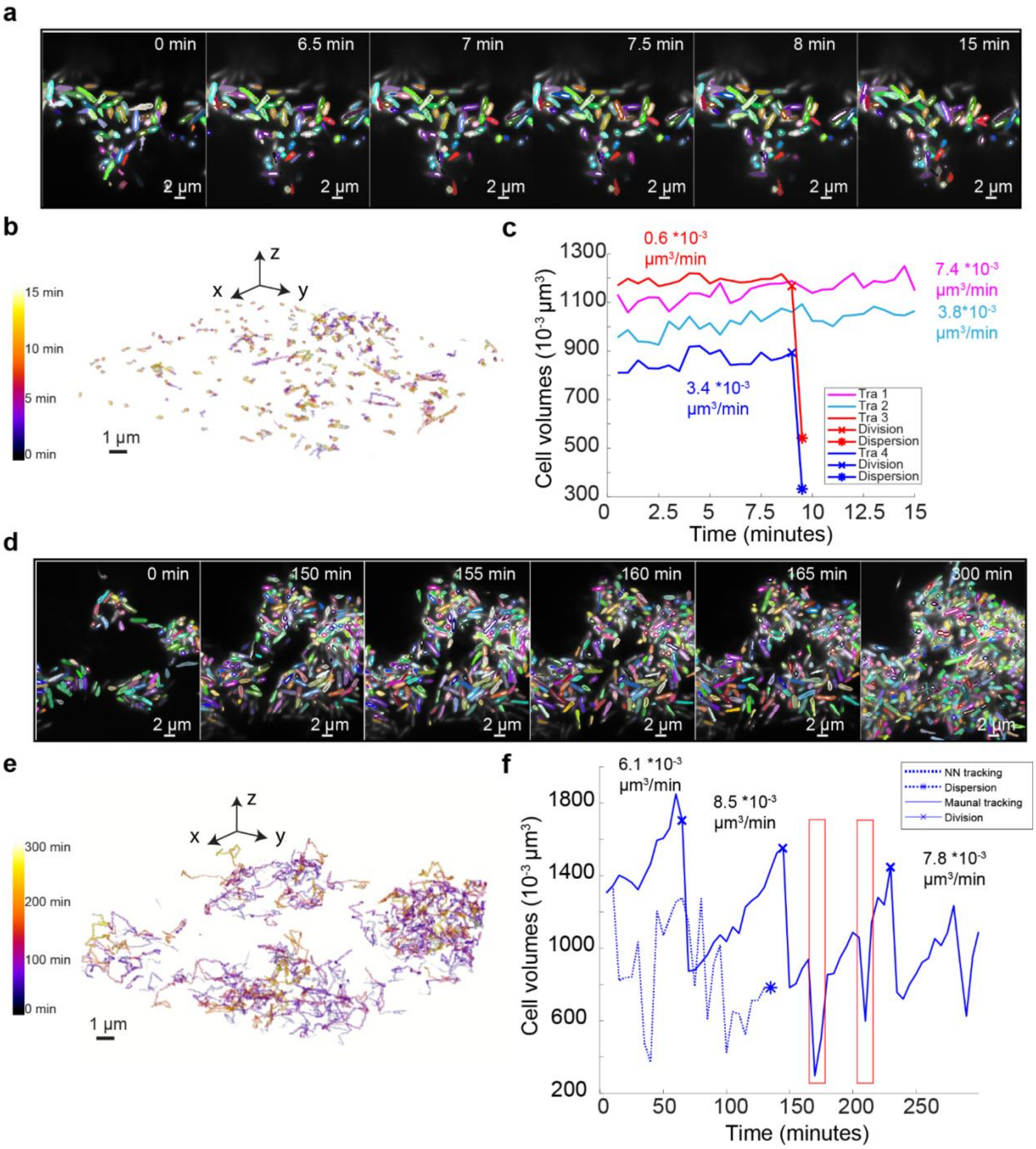


Figure 4.2.2. (a) Experimentally acquired fluorescence time-lapse images of a growing *S. oneidensis* biofilm with overlaid single-cell segmentation contours. Images were acquired every 30 seconds for 15 minutes. Corresponding cells in different frames are displayed in the same color. (b) Individual cell trajectories in the biofilm shown in panel a. Cells move very little during the short 15-minute imaging time. (c) Cell volumes over time for four example cells. Cell division and dispersion events are indicated for each trajectory. Single-cell growth rate was measured by calculating the slope of linearly fitted line for each curve (d) Experimentally acquired fluorescence

time-lapse images of a growing *S. oneidensis* biofilm with overlaid single-cell segmentation contours. Images were acquired every five minutes for five hours. Corresponding cells in different frames are displayed in the same color. (e) Individual cell trajectories in the biofilm shown in panel a. Cell displacements are more pronounced over the 5-hour imaging time. (f) Evolution of cell volumes over time for a single selected cell. The solid line represents a manually annotated trajectory, and the dashed line represents the nearest neighbor tracking trajectory of the same initial cell. Cell division and dispersion events are indicated on each trajectory. The single-cell growth rate was measured by calculating the slope of the trajectory segment between consecutive cell division events. The red arrow and the boxes indicate periodic underestimation of the cell volume due to oversegmentation.

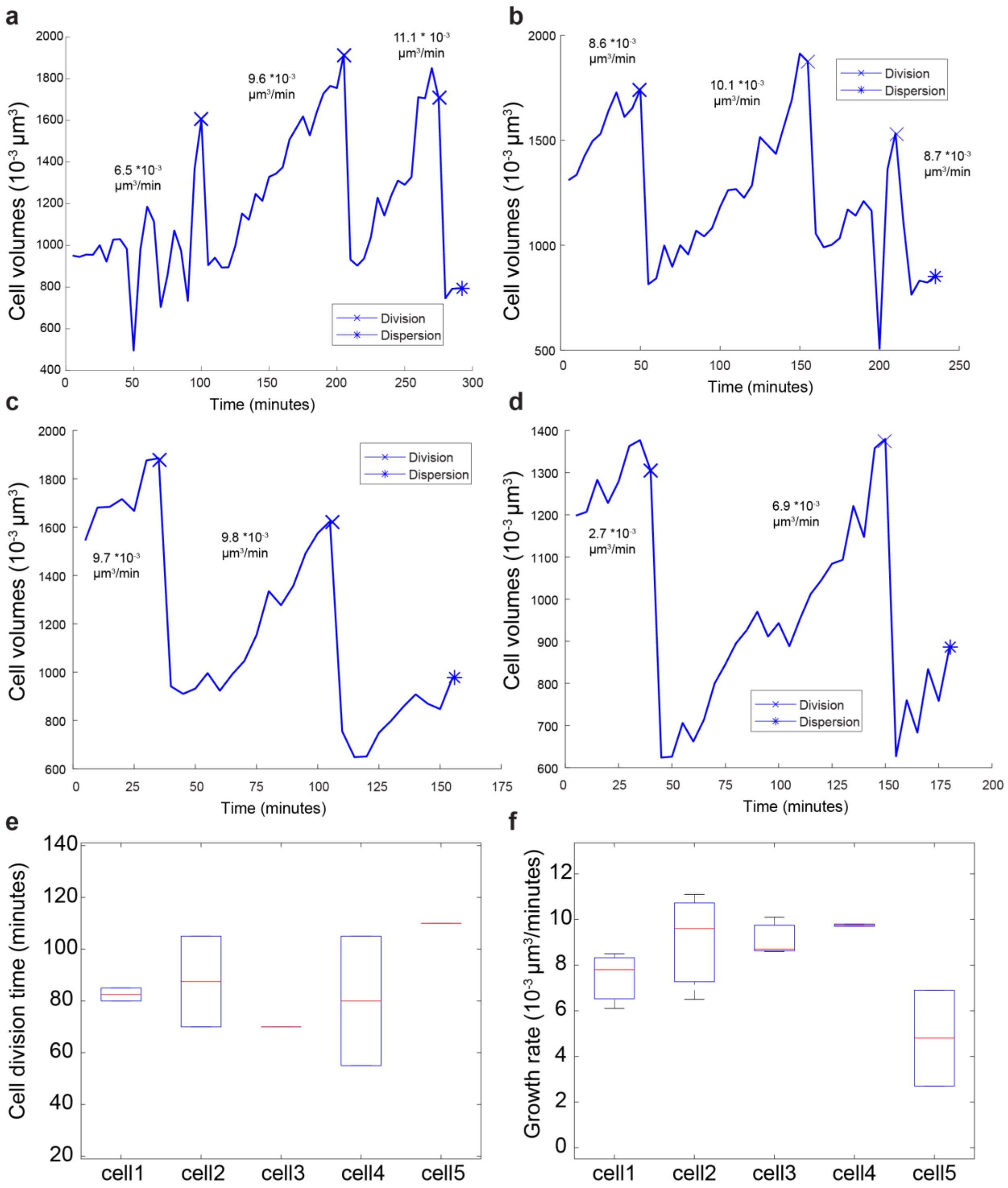


Figure 4.2.3. (a-f) Cell volumes over time for four additional cells. The solid line represents a manually annotated trajectory where cell division and dispersion events are detected manually and indicated on the trajectories. Single-cell growth rate was measured by calculating the slope of a line that connects both the beginning and end points for each trajectory. (e) Box plot of cell division

time for the selected fives cell (four in this figure plus one in Figure 4.2.2). (f) Box plot of single-cell growth rate for the selected five cells.

4.2.4 New time-dependent single-cell observables for biofilm studies

We have demonstrated that single-cell trajectories and lineage information are contained in the time-lapse data in the previous section. Although it currently remains challenging to automatically obtain complete, error-free trajectories as well as lineage information over several generations, there is quantitative information about single-cell behaviors that can answer biological questions and can be accurately measured even with the nearest neighbor tracking results. Time-dependent features such as cellular movement and growth rate don't require complete trajectories over hundreds of frames and therefore can be extracted with acceptable accuracy from tracking just a few frames.

Such single-cell observables open a door for studying biofilm heterogeneity on the single-cell level. To demonstrate this, I hypothesized that the cells on the surface have different cellular behaviors compared to interior cells because they are in a different physical and chemical environment with different levels of nutrient concentration, oxygen concentration, and shear rate, etc.

In order to test whether there is a spatial heterogeneity in the biofilm, I analyzed the *S. oneidensis* biofilm discussed in the previous section, which represents an early stage of biofilm development, including initial surface attachment of hundreds of cells as well as two to three generations of cell divisions. I divided the biofilm into the surface cells and the interior cells based on their locations in the biofilms, and measured features of individual cells. (**Figure 4.2.4ab**). However, I didn't observe any statistically different distributions in terms of either cellular movement or cellular growth (**Figure 4.2.4cd**). This is also consistent with the manually-traced single-cell trajectory data in which randomly chosen cells grow and divide at similar rates (**Figure 4.2.3ef**).

For future experiments, the Gahlmann lab plans to carry out similar experiments in a more physically relevant environment for much longer periods of time (e.g., a dual-channel experiment to study the biofilm morphology at the host-pathogen interface). It is interesting to see whether the surface cells and the interior cells behave differently and at what point during the course of biofilm development does this heterogeneity occur.

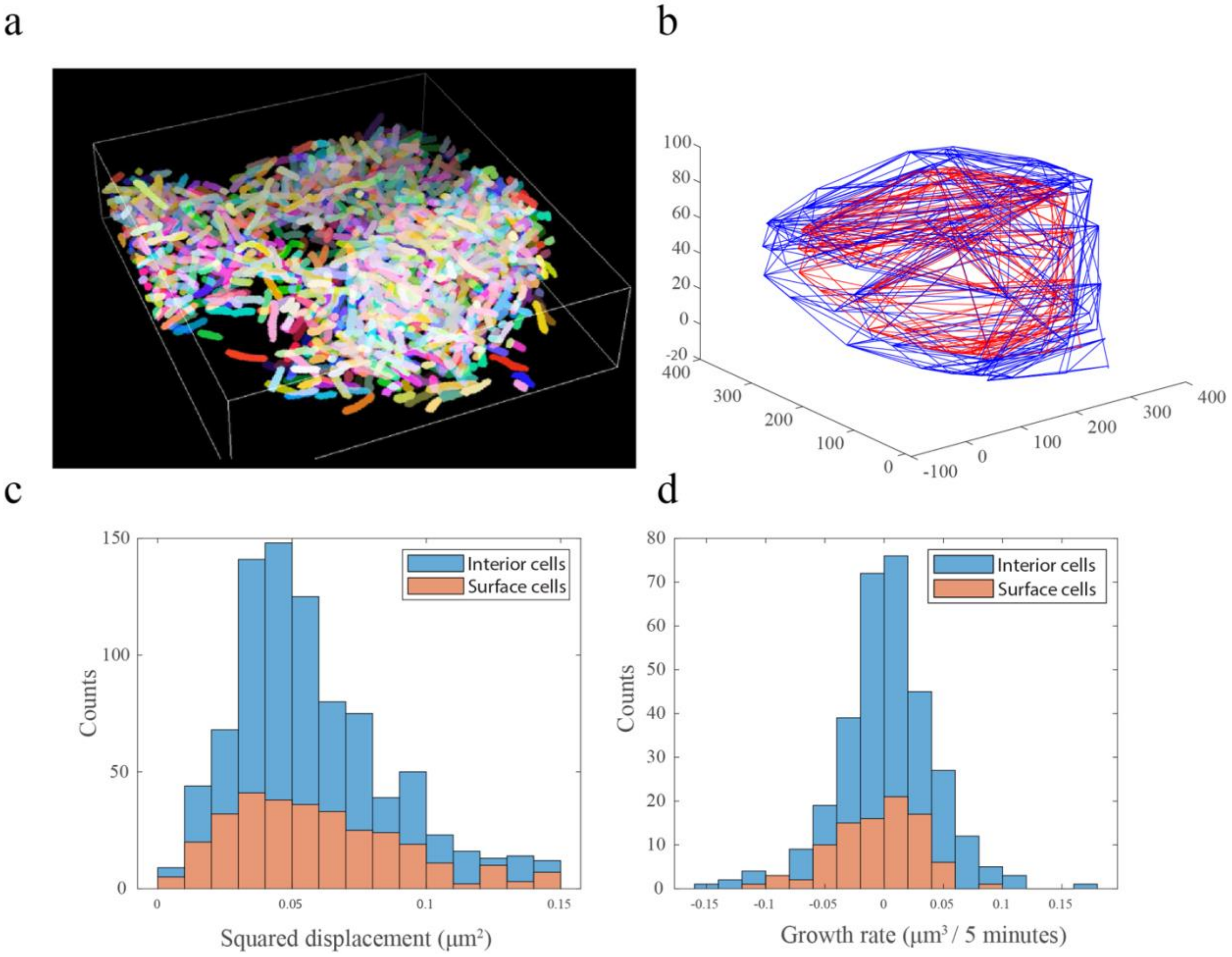


Figure 4.2.4. Time-dependent single-cell observables in a biofilm. (a) 3D rendering of the last panel in **Figure 4.2.2d** at t = 300 mins. (b) Convex hulls are used to differentiate biofilm surfaces and interiors. Surface cells are cells that have centroids inside the blue convex hull but outside the red convex hull, while interior cells are cells that have centroids inside the red convex hull. (c) Squared displacement distributions of the cells on the biofilm surface or interior. (d) Growth rate distributions of the cells on the biofilm surface or interior.

4.2.5 Methods

Tracking

Simpletracker in MATLAB was used to build tracking graphs and spatially resolved lineage trees[143]. *Simpletracker* implements the Hungarian algorithm and nearest neighbor trackers for particle tracking that links particles between frames in 2D or 3D. We used 1 µm and 1.5 µm as the maximum distance threshold for cell linking for simulated and experimental data, respectively. We used the nearest neighbor algorithm to associate the centroids of segmented objects in subsequent frames, such that the closer pairs of centroids are linked first. In order to determine a cell division event, a distance threshold of 1 µm and 1.5 µm for simulated and experimental data, respectively, a cell volume threshold of 1.5 (parent cell should be 1.5 times larger than the daughter cell), and a cell length threshold of 1.5 (parent cell should be 1.5 times longer than the daughter cell), were used to determine parent-daughter relationships between cell pairs on consecutive frames.

Tracking evaluation

To quantify tracking accuracy, we used the acyclic oriented graph metric (*AOGM*)[142]. The *AOGM* value is calculated as the weighted sum of the number of graph operations required to convert the estimated graph to the ground truth graph, i.e.:

$$AOGM = w_{NS}NS + w_{FN}FN + w_{FP}FP + w_{ED}ED + w_{EA}EA + w_{EC}EC$$

The tracking accuracy can then be computed using a normalized *AOGM* value, where $AOGM_0$ is the number of operations to build the ground truth graph from an empty graph:

$$TRA = 1 - min(AOGM, AOGM_0)/AOGM_0$$

There are three types of graph operations that are associated with detection errors: the number of false negatives (*FN*), the number of false positives (*FP*), and the number of missed splits (*NS*: *m* reference cells ($m > 1$) are assigned to a single segmented cell). There are also three types of graph operations that are associated with object linking: edge deletion (ED), addition (*EA*), and alteration of the semantics of an edge (*EC*: The semantics of an edge can either represent the same cells over time or represent a parent-daughter relationship). To focus on object matching over time (i.e. the association performance of the algorithm), we used an equally weighted sum of the lowest number of graph operations on edges only (*TRA_edge*). To give a more comprehensive view, we used an equally weighted sum of the number of graph operations on all six operations (*TRA_full*).

To estimate tracking accuracy for experimental data, we manually traced a small subset (n = 25) ancestor cells over time based on *BCM3D 2.0* segmentation masks. Two researchers performed tracking independently, manually determining parent-daughter relationships within the lineages originating from the ancestor cells. This lineage information was then used to compute *TRA_edge*.

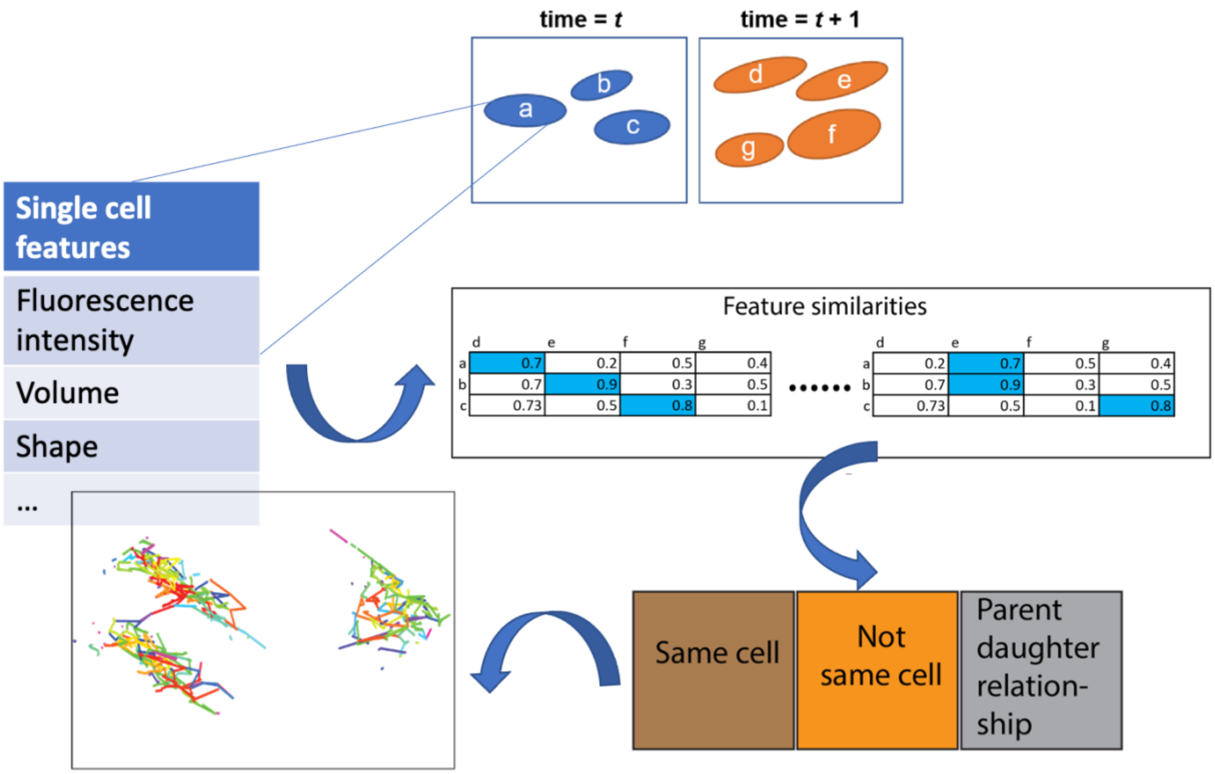
4.3 Simultaneous multi-cell tracking and lineage tracing using machine learning

4.3.1 Results

The nearest neighbor approach shows promising results in tracking bacterial biofilms at an early stage for both simulated images and real experimental images. However, the limitations are evident, such that it fails for high Relative Movement (*RM*) datasets, which are common for many biofilm species using imaging protocols that are optimized for image quality over time. It is therefore necessary to seek for a more robust tracking algorithm that can work with high *RM* datasets. Initially, our lab tried to engineer a cost function that takes into consideration multiple cellular parameters, such as cell shapes and sizes to match cells between frames, which fit seamlessly into the Hungarian algorithm framework. However, the number of parameters in the cost function adds up too quickly to optimize, and it is observed that such optimization needs to be done on every single new dataset.

A machining learning-based approach is superior because it resembles much more to how a human annotator makes tracking decisions of linking bacterial cells. Humans are not only good at judging distances, but can also intuitively evaluate and rank the similarities based on shapes, fluorescence intensities and volume, etc., between candidate cells, thereby making much more reliable decisions when the biofilms are dense and the cells have significant movement between frames. In order to mimic how a human annotator make a series of decisions on distance, shapes, and sizes, I therefore developed a machine leaning-based approach to explicitly consider all the single-cell features that a human may consider when making informed decisions on tracking. ML-based models are much better than nearest-neighbor or Hungarian tracking algorithms because it is completely data-driven, and therefore there is no need to determine feature importance beforehand. We also argue that a machine learning approach may have the potential to outperform human annotators because a human relies on feelings rather than accurate quantitative measurements of the cells. The machine learning approach can also readily detect cell dispersion events and parent-daughter relationships if enough representative, high-quality data is provided.

More specifically, the machine learning framework for tracking bacterial cells is illustrated in **Figure 4.3.1**. Single-cell parameters (features) from the segmentation masks are collected for each frame. Pairwise feature similarities scores are calculated and reorganized into a design matrix. The design matrix is then fed into a classical machine learning algorithm to make multiclass classification, in which there are three classes: ‘Same cell’, ‘Not same cell’, and ‘Parent-daughter relationship’ respectively (**Figure 4.3.2**). I have tested several algorithms including logistic regression, XGboost[145] and the Random Forest algorithm[146]. The Random Forest algorithm is determined to consistently yield better results (comparisons not shown). Finally, the classification results can be easily transformed to trajectories as well as the format that complies with the literature[142]. More details of this approach can be found in the Methods section.



	d	e	f	g
a	0.7	0.2	0.5	0.4
b	0.7	0.9	0.3	0.5
c	0.73	0.5	0.8	0.1

	d	e	f	g
a	0.2	0.7	0.5	0.4
b	0.7	0.9	0.3	0.5
c	0.73	0.5	0.1	0.8

Figure 4.3.1. A schematic of simultaneous multi-cell tracking and lineage tracing using machine learning.

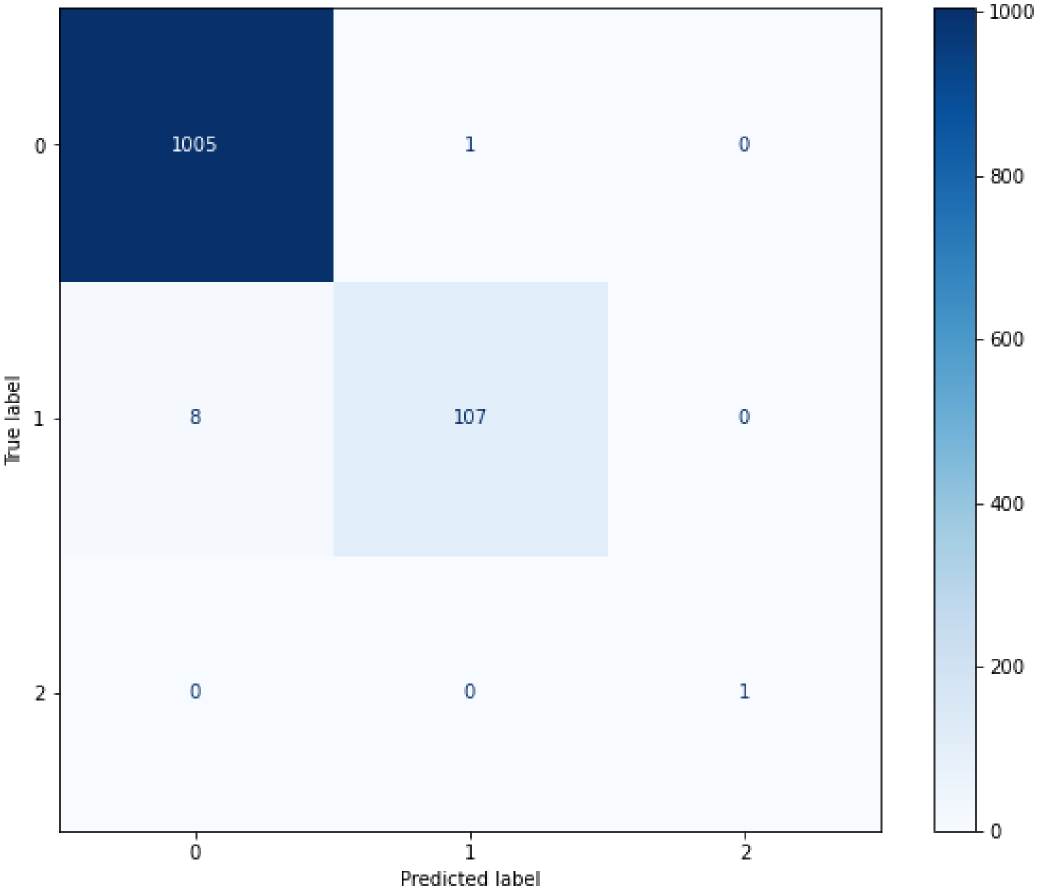


Figure 4.3.2. Confusion matrix of the test subset. ‘0’ class denotes a different cell pair between two consecutive frames, ‘1’ class denotes the same cell pair between two consecutive frames, and ‘2’ class denotes parent-daughter relationship between two consecutive frames

For the simulated biofilm images tested in Zhang et al. 2021 (*RM* ~ 0.2), vast majority of the trajectories as well as lineage information can be correctly traced for several generations (**Figure 4.3.3a**). At *RM* ~ 0.2, almost all cell division events are captured. As a result, the detected cell division events allow for accurate estimation of single-cell doubling cycles of the entire population in the biofilm (**Figure 4.3.3b**).

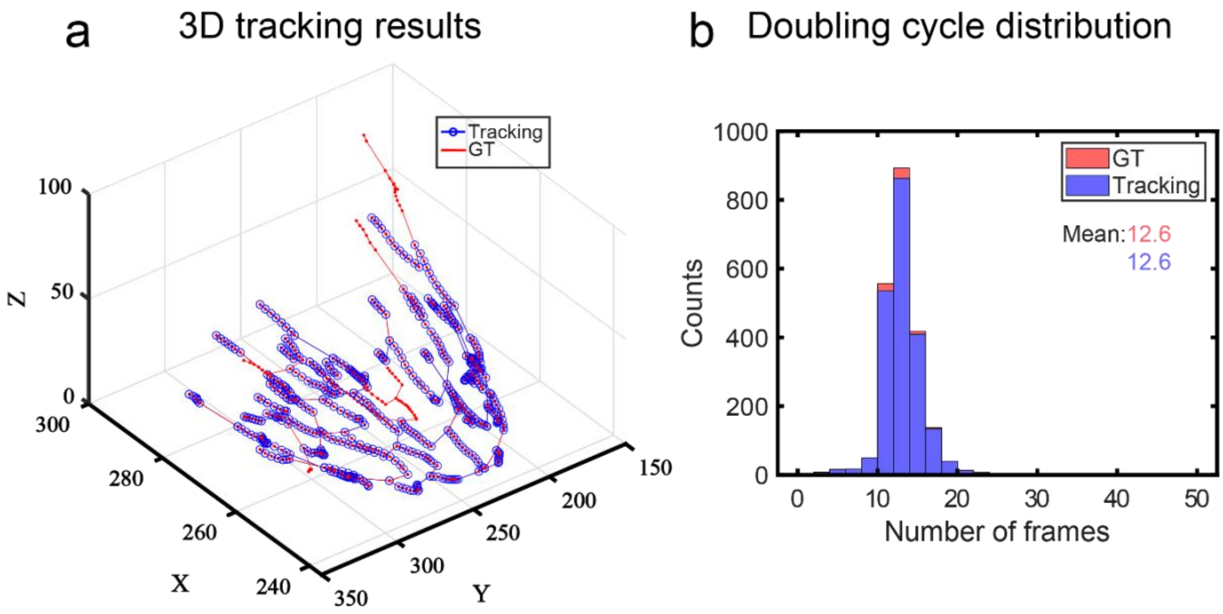


Figure 4.3.3. (a) An example of 3D tracking and lineage tracing for simulated biofilm images using machine learning. For clarity, spatial trajectories and lineages originating from only a single ancestor cell is displayed. The estimated graph is shown in blue and the corresponding ground truth graph is shown in red. The entire biofilm contains over sixty graphs of this type. (b) Doubling cycle distribution of simulated data and corresponding tracking results. A completed cell cycle is defined as a track in which the parent cell is able to split twice. This threshold results in a lower count numbers of estimated cell division, but does not alter the shape of the distribution.

4.3.2 Discussion

Machine learning-based tracking algorithms, which utilize single-cell parameters to compute similarity scores, are more robust and less prone to the accumulation of tracking errors than conventional tracking approaches. A segmentation result that produces a linkage with low pair-wise similarity score can be flagged for requiring increased scrutiny and possible re-evaluation of human-annotated data. Critical events such as cell divisions and cell dispersions are naturally detectable by the design of the machine learning approaches, and thus can be quantitatively assessed by the researchers.

Although we argue that the machine learning-based algorithm closely mimics how a human annotator makes the linkages, there is still more sophisticated information that human annotators use that our approaches cannot leverage. For example, human annotators can go back and forth in time to not only check pairwise similarities between frames, but also confirm the trajectories over multiple frames when making decisions. Human annotators can intuitively leverage time-dependent information that is contained in the data. However, the current version of the method does not consider any time-dependent features, which likely will further improve tracking accuracy. Although the incorporation of time-dependent features might not be as straightforward, it is still plausible to add features such as velocity, change in shape, etc.

Another extension that I have not implemented is the use of deep learning models to extract deep single-cell features that can be used to calculate pairwise similarity scores. The reason why I didn’t start using deep learning is because that all bacterial cells are small and their shapes are of simple geometries carrying little information to extract deep features. As a result, the use of deep learning may greatly reduce the speed for a tracking algorithm. However, this assumption may not hold if we want to incorporate time-dependent features. The use of deep learning models such as Long short-term memory (LSTM) or Transformers are well-suited to extract time-dependent information, similar to what a human annotator would do, and are expected to outperform classical models that consider only single time frames[147, 148].

Upon successful experimental validation and the inclusion of time-dependent features, we will make the developed machine learning-based multi-cell tracking algorithms and any associated documentation and test data freely available to the research community.

4.3.3 Methods

Feature extraction from segmentation masks

Single-cell features are extracted from each frame using *regionprops* function in Python[137]. For this chapter, only 'area', 'centroid', 'extent', 'major_axis_length', 'max_intensity', 'mean_intensity' 'min_intensity', 'minor_axis_length', 'solidity', 'weighted_moments_normalized' were considered for comparing pairwise feature similarity score. For detailed explanations of all supported features, please refer to *scikit-image*[137]. It is possible that more sophisticated features used in the literature could potentially improve tracking performances, but it is beyond the scope of this chapter.

Calculate pairwise feature similarity using vectorization

Pairwise feature similarity score, *Sim(p,q)* was defined as the Euclidian distance between each object between consecutive frames. Where p denotes an object at any frame t_x and q denotes an object at t_{x+1}, and n is the number of dimensions for each feature.

$$Sim(p,q) = \sqrt{\sum_{i=n}^{n} (q_i - p_i)^2}$$

The naïve approach to calculate pairwise feature similarities require nested loops, which has complexity of O(P*Q), where P is the number of cells detected at frame t_x and Q is the number of cells at t_{x+1}. This approach typically takes hours if not days for tracking large biofilms. Vectorization is a much more efficient way to calculate this in Python. There are overall two steps associated with vectorization in order to calculate pairwise feature similarities. The first step is to expand two matrices with shape (P, N) and (Q, N) with broadcasting to produce an expanded array with shape (P, Q, N), where N denotes the dimension of the features. The second step is to reduce the added dimension using an aggregate operation, creating the desire matrix with shape (P, Q). A similar approach can also be used to calculate and evaluate intersection over union (IoU) between ground truth and segmentation results.

Generate Ground truth

Ground truth tracks were generated by two researchers linking objects from BCM3D 2.0 segmentation masks. The pairwise class was thus defined as '0' or 'not same cell', '1' or 'same cell', and '2' or 'have parent-daughter relationship'. The 'not same cell' class was also limited by a 'close_n' parameter, in which only the closest n objects (n = 10 in this chapter) were selected to be included in the training data to reduce computational time and alleviate the class imbalance problem.

Random forest classifier

The data frame was scaled to have a zero mean and unit variance. Then the data was randomly split into 80/20 for training and validation. A random forest[146] classifier with maximum depth of 5, and a balanced class weight was used to train the model. Feature importance in a random forest algorithm can be measured by mean decrease in impurity (MDI). MDI measures how many times a given feature is used to split a node within a decision tree (**Figure 4.3.2**). Therefore, a higher MDI for a given feature indicates that it is more important to form the trained random forest decision trees. The error bars (standard deviations) of MDI showcase the variability of feature importance among decision trees.

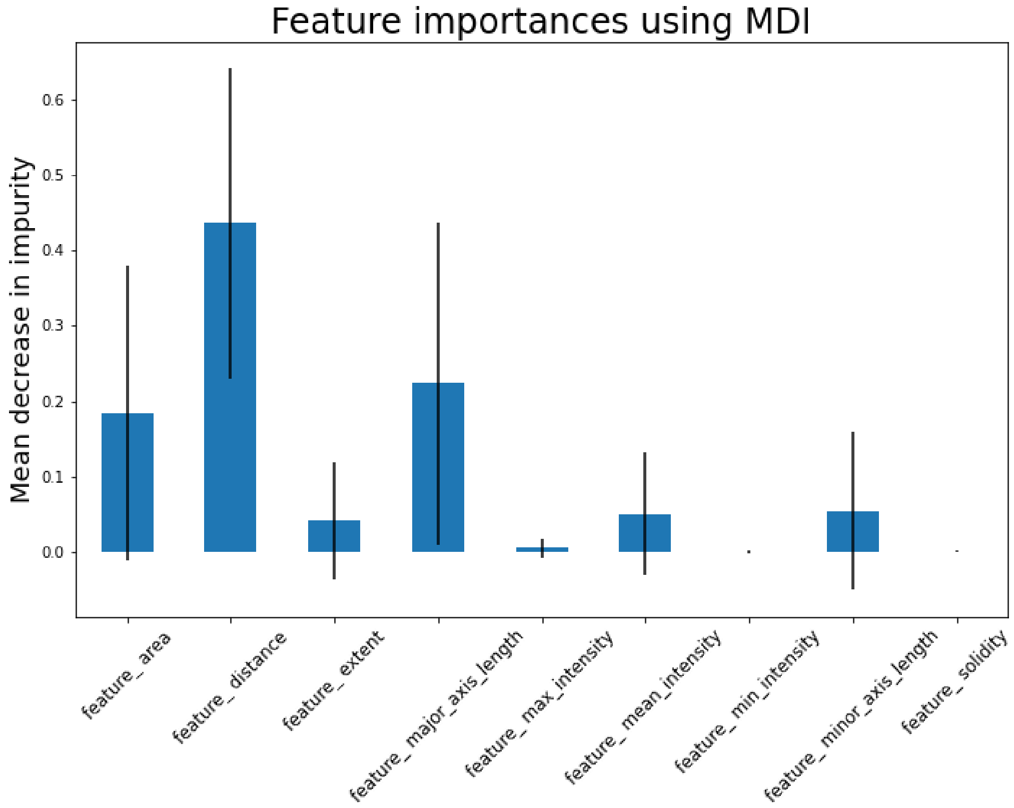


Figure 4.3.2. An example of feature selection process using mean decrease in impurity (MDI) and ranking of feature importance. Feature importance of the trained model is measured in mean decrease in impurity. The error bar indicates plus or minus one standard deviation.

Chapter 5: Conclusion and future directions

5.1 Conclusion

Bacterial biofilms have evolved to form complex heterogeneous structures, but there is still much to learn about the life cycle in various species, as well as their relationships to different biotic or abiotic surfaces. Bacterial biofilms have both positive and negative impacts on our daily lives, making them a subject of great interest for researchers across various fields. This thesis aims to provide a detailed, comprehensive guide on instrumentation as well as computational solutions for investigating bacterial biofilms at the single-cell level in a physiologically relevant environment.

Due to the dynamic nature of bacterial biofilms, the lattice light sheet microscope (LLSM) offers the best balance between spatial and temporal resolution for 3D imaging, enabling single-cell resolution for hours or even days. While LLSM provides superior imaging capability for bacterial biofilms over other commercially available microscopes, implementing a microfluidic channel is not as straightforward due to constricted space in the microscope basin. This thesis outlines best practices for using microfluidic channels as well as associated challenges and corresponding solutions. Two of the major challenges discussed are maintaining a steady refractive index in the basin and significant sample drift caused by the 3D printable resin. The thesis introduces computational solutions as well as LLSM implementation and demonstration of the code for real-time sample drift correction.

LLSM can generate terabytes of time-lapse data, but extracting quantitative information at the single-cell level is not a trivial task. Image segmentation, specifically instance segmentation, poses a key challenge in accurately and efficiently obtaining information of single cells, such as cell size, morphology, and locations. Deep learning approaches have shown to outperform conventional image processing techniques. Our research group was among the first to adopt 3D convolutional neural networks (CNN), named *BCM3D 1.0*, to successfully segment biofilm image data. We demonstrated that training a CNN with carefully simulated data is not only a viable approach but also outperforms the one trained with manually annotated data for 3D biofilm segmentation, which was considered the gold standard in image segmentation. Once segmentation masks are obtained, we can get single-cell level information and perform morphology classification, enabling more sophisticated experiments that involve different species and cell morphologies without the need for multiple fluorophores or laser lines.

Building upon the concept of *BCM3D 1.0*, *BCM3D 2.0* represents an improved version with a carefully designed workflow to enhance segmentation accuracy, even for low signal-to-background images and dense microbial communities. *BCM3D 2.0* uses the concept of image translation in deep learning and is integrated with downstream image analysis pipelines to precisely find the cell boundaries and segment individual cells. Being able to segment individual cells accurately is an important requirement to link the same cell over time and track its trajectories and behaviors. This task is called multi-object tracking (multi-cell tracking) in the field of image processing and computer vision.

While there are multiple tracking algorithms available in the field of biology research, they are primarily designed for tracking eukaryotic cells, which are distinct from one another and also normally larger than the diffraction limit, resulting in clear cellular boundaries in the images. As a result, they tend to fail for dense biofilm images. In order to track individual cells in biofilms, we first attempted to use a most basic tracking algorithm assuming minimal cell movement between frames with sufficient temporal resolution. However, the vanilla solution for tracking bacterial biofilms is shown to be linearly correlated with segmentation accuracy, and only performed well for high-temporal resolution data with slow cellular motion. In contrast, a machine learning-based tracker leverages cellular information beyond just distance, leading to more accurate linking even for low-temporal resolution data. Such a tracker can also effectively trace cell trajectories based on cellular features contained in the training data without relying on any hardcoded thresholds.

Individual trajectories of single bacterial cells as well as lineage information open the door to novel experiments on bacterial biofilms. The new information enhances our understanding for biofilm formation of each species and how they interact with different surfaces. It therefore can be useful for informing a biofilm formation model and calibrating agent-based simulations. Preliminary results already indicate significant differences from different species from the hypothesized mechanism of a generic biofilm formation model. The tools developed in this thesis can also enhance our understanding of how coordinated behaviors among biofilm-dwelling contribute to the macroscopic properties of bacterial biofilms. The tracking results can shed light on persister cell formation and potentially improve our understanding of antibiotic resistance attributed to bacterial biofilms[149].

In summary, the ability to accurately identify and track individual cells in dense 3D biofilms over long periods of time requires the combination of non-invasive fluorescence microscopy approaches for long-term time-lapse imaging and sophisticated image analysis and multi-object tracking tools that yield robust results. Each component discussed in the thesis is essential for obtaining optimal results. Any improvements in imaging, segmentation or tracking will have a positive impact on all other aspects of biofilm research (**Figure 5.1**).

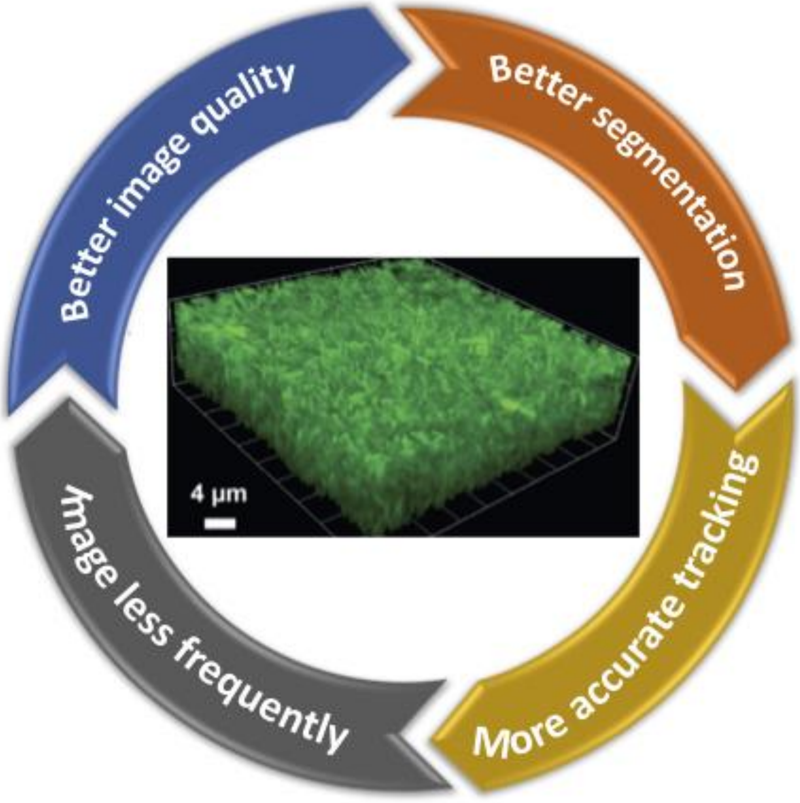


Figure 5.1. Illustrative summary for bacterial biofilm imaging and image processing.

5.2 Future directions

As the field of artificial intelligence advances rapidly, unprecedented opportunities arise for researchers in biology and biochemistry to leverage these advancements and investigate biological systems that were previously considered too complex to probe. As I finish writing this thesis, new candidates of deep learning-based models have emerged, holding the potential to yield improved biofilm image segmentation and tracking results. For instance, diffusion models have recently exhibited promising outcomes in modeling natural images that can perform various image processing tasks. Diffusion models are a class of generative models that approximate the distribution of real images from a simple parametric distribution, and it is shown that they can be repurposed for image segmentation with high efficiency[150].

Moreover, in the context of multi-object tracking, deep neural networks may outperform classical machine learning algorithms, which were tested in this thesis and rely on human-specified features. Specifically, a transformer model, renowned for its success across diverse fields, can be employed not only to extract simple cellular features but also to capture time-dependent features or "fingerprints" of a cell[148]. This approach has the potential to identify cells and discriminate their neighboring cells in 3D and time, mimicking how a human annotator traces the movement of a cell in a crowded biofilm environment over time. However, a deep learning approach inevitably demands a large amount of high-quality training data, and the current methods of obtaining training data, relying on simple biofilm simulation procedures or human annotation, may not suffice in terms of quantity and quality. Two avenues can be pursued to address this challenge: firstly, enhancing the current simulation procedures by calibrating them using real biofilm formation data, and secondly, adopting the popular human-in-the-loop approach, which can significantly accelerate human annotation.

Furthermore, the innovation of microscopes and microfluidic systems is an ongoing process. Improvements in the stability of microfluidic systems will enhance image quality over time, thereby improving both segmentation and tracking accuracy. While automated long-term imaging of bacterial biofilms has been achieved, the precision of the stage remains a limiting factor. Employing a more stable piezo nanopositioning system that offers a wider range of movement will greatly aid in data acquisition. Additionally, incorporating fiducials such as fluorescent beads into the workflow may increase the accuracy of drift correction algorithms.

Lastly, the Gahlmann lab aims to study bacterial biofilms not only on simple surfaces but also on sophisticated engineered surfaces and eukaryotic host tissues, such as human organoid-derived epithelia. The current imaging system and image processing modules can be adapted and applied to investigate the host-pathogen interface with high spatial and temporal resolution. It is possible to miniaturize a dual-channel tissue chip designed to support intestinal segment-specific cell differentiation and epithelial tissue functionality[151]. Making these chips optically accessible and compatible with lattice light sheet microscopy would enable the study of gut-colonizing bacteria, such as *Shigella flexneri*, and their survival strategies in different intestinal niches prior to causing acute gastrointestinal disease by infecting the colonic epithelia.

In conclusion, by harnessing the potential of artificial intelligence, advancements in microscopy and microfluidic systems, researchers in the field of bacterial biofilm will gain in-depth understanding of such complex and heterogeneous biological systems and their interactions in various contexts.

REFERENCES

1. Drescher, K., Dunkel, J., Nadell, C.D., van Teeffelen, S., Grnja, I. et al. Architectural transitions in Vibrio cholerae biofilms at single-cell resolution. *Proceedings of the National Academy of Sciences of the United States of America* **113**, E2066-E2072 (2016).

2. Flemming, H.-C. & Wingender, J. The biofilm matrix. *Nature reviews microbiology* **8**, 623 (2010).

3. Huttenhower, C., Gevers, D., Knight, R., Abubucker, S., Badger, J.H. et al. Structure, function and diversity of the healthy human microbiome. *nature* **486**, 207 (2012).

4. Yan, J. & Bassler, B.L. Surviving as a community: antibiotic tolerance and persistence in bacterial biofilms. *Cell host & microbe* **26**, 15-21 (2019).

5. Cámara, M., Green, W., MacPhee, C.E., Rakowska, P.D., Raval, R. et al. Economic significance of biofilms: a multidisciplinary and cross-sectoral challenge. *npj Biofilms and Microbiomes* **8**, 42 (2022).

6. Sehar, S. & Naz, I. Role of the biofilms in wastewater treatment. *Microbial biofilms-importance and applications*, 121-144 (2016).

7. Liu, T., Yu, Y.Y., Deng, X.P., Ng, C.K., Cao, B. et al. Enhanced Shewanella biofilm promotes bioelectricity generation. *Biotechnology and Bioengineering* **112**, 2051-2059 (2015).

8. Hartmann, R., Singh, P.K., Pearce, P., Mok, R., Song, B. et al. Emergence of three-dimensional order and structure in growing biofilms. *Nature Physics* **15**, 251 (2019).

9. Yan, J., Sharo, A.G., Stone, H.A., Wingreen, N.S. & Bassler, B.L. Vibrio cholerae biofilm growth program and architecture revealed by single-cell live imaging. *Proceedings of the National Academy of Sciences* **113**, E5337-E5343 (2016).

10. Stewart, P.S. & Franklin, M.J. Physiological heterogeneity in biofilms. *Nature Reviews Microbiology* **6**, 199-210 (2008).

11. Hall-Stoodley, L. & McCoy, K.S. Biofilm aggregates and the host airway-microbial interface. *Frontiers in Cellular and Infection Microbiology* **12**, 969326 (2022).

12. Zhang, J., Zhang, M., Wang, Y., Donarski, E. & Gahlmann, A. Optically Accessible Microfluidic Flow Channels for Noninvasive High-Resolution Biofilm Imaging Using Lattice Light Sheet Microscopy. *The Journal of Physical Chemistry B* (2021).

13. Power, R.M. & Huisken, J. A guide to light-sheet fluorescence microscopy for multiscale imaging. *Nature methods* **14**, 360-373 (2017).

14. Stelzer, E.H. Light-sheet fluorescence microscopy for quantitative biology. *Nature methods* **12**, 23-26 (2015).

15. Voleti, V., Patel, K.B., Li, W., Campos, C.P., Bharadwaj, S. et al. Real-time volumetric microscopy of in vivo dynamics and large-scale samples with SCAPE 2.0. *Nature methods* **16**, 1054-1062 (2019).

16. Yordanov, S., Neuhaus, K., Hartmann, R., Díaz-Pascual, F., Vidakovic, L. et al. Single-objective high-resolution confocal light sheet fluorescence microscopy for standard biological sample geometries. *Biomedical Optics Express* **12**, 3372-3391 (2021).

17. Qin, B., Fei, C., Bridges, A.A., Mashruwala, A.A., Stone, H.A. et al. Cell position fates and collective fountain flow in bacterial biofilms revealed by light-sheet microscopy. *Science* (2020).

18. Parthasarathy, R. Monitoring microbial communities using light sheet fluorescence microscopy. *Current opinion in microbiology* **43**, 31-37 (2018).

19. Heintzmann, R. & Huser, T. Super-Resolution Structured Illumination Microscopy. *Chemical reviews* **117**, 13890-13908 (2017).

20. Gustafsson, M.G., Shao, L., Carlton, P.M., Wang, C.R., Golubovskaya, I.N. et al. Three-dimensional resolution doubling in wide-field fluorescence microscopy by structured illumination. *Biophysical journal* **94**, 4957-4970 (2008).

21. Weigert, M., Schmidt, U., Boothe, T., Müller, A., Dibrov, A. et al. Content-aware image restoration: pushing the limits of fluorescence microscopy. *Nature methods* **15**, 1090 (2018).

22. Chen, J., Sasaki, H., Lai, H., Su, Y., Liu, J. et al. Three-dimensional residual channel attention networks denoise and sharpen fluorescence microscopy image volumes. *Nature methods*, 1-10 (2021).

23. Lim, Y., Shiver, A.L., Khariton, M., Lane, K.M., Ng, K.M. et al. Mechanically resolved imaging of bacteria using expansion microscopy. *PLoS biology* **17**, e3000268 (2019).

24. Koyama, L.A.J., Aranda-Díaz, A., Su, Y.-H., Balachandra, S., Martin, J.L. et al. Bellymount enables longitudinal, intravital imaging of abdominal organs and the gut microbiota in adult Drosophila. *PLoS Biology* **18**, e3000567 (2020).

25. Oliveira, R.A., Ng, K.M., Correia, M.B., Cabral, V., Shi, H. et al. Klebsiella michiganensis transmission enhances resistance to Enterobacteriaceae gut invasion by nutrition competition. *Nature microbiology* **5**, 630-641 (2020).

26. Chen, B.-C., Legant, W.R., Wang, K., Shao, L., Milkie, D.E. et al. Lattice light-sheet microscopy: imaging molecules to embryos at high spatiotemporal resolution. *Science* **346** (2014).

27. Goodfellow, I., Bengio, Y. & Courville, A. Deep learning. (MIT press, 2016).

28. Liu, Z., Jin, L., Chen, J., Fang, Q., Ablameyko, S. et al. A survey on applications of deep learning in microscopy image analysis. *Computers in Biology and Medicine* **134**, 104523 (2021).

29. Qiao, C., Li, D., Guo, Y., Liu, C., Jiang, T. et al. Evaluation and development of deep neural networks for image super-resolution in optical microscopy. *Nature methods* **18**, 194-202 (2021).

30. Lichtman, J.W. & Conchello, J.-A. Fluorescence microscopy. *Nature methods* **2**, 910-919 (2005).

31. Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M. et al. Fiji: an open-source platform for biological-image analysis. *Nature methods* **9**, 676-682 (2012).

32. Li, D., Shao, L., Chen, B.C., Zhang, X., Zhang, M. et al. Extended-resolution structured illumination imaging of endocytic and cytoskeletal dynamics. *Science* **349**, aab3500 (2015).

33. Parslow, A., Cardona, A. & Bryson-Richardson, R.J. Sample drift correction following 4D confocal time-lapse imaging. *Journal of visualized experiments: JoVE* (2014).

34. Goddard, T.D., Huang, C.C., Meng, E.C., Pettersen, E.F., Couch, G.S. et al. UCSF ChimeraX: Meeting modern challenges in visualization and analysis. *Protein Science* **27**, 14-25 (2018).

35. Hall-Stoodley, L., Costerton, J.W. & Stoodley, P. Bacterial biofilms: from the natural environment to infectious diseases. *Nature reviews microbiology* **2**, 95 (2004).

36. O'Toole, G., Kaplan, H.B. & Kolter, R. Biofilm formation as microbial development. *Annual Reviews in Microbiology* **54**, 49-79 (2000).

37. Nadell, C.D., Drescher, K. & Foster, K.R. Spatial structure, cooperation and competition in biofilms. *Nature Reviews Microbiology* **14**, 589 (2016).

38. Persat, A., Inclan, Y.F., Engel, J.N., Stone, H.A. & Gitai, Z. Type IV pili mechanochemically regulate virulence factors in Pseudomonas aeruginosa. *Proc Natl Acad Sci U S A* **112**, 7563-7568 (2015).

39. Zhang, Q., Lambert, G., Liao, D., Kim, H., Robin, K. et al. Acceleration of Emergence of Bacterial Antibiotic Resistance in Connected Microenvironments. *Science* **333**, 1764-1767 (2011).

40. Donlan, R.M. Biofilms: microbial life on surfaces. *Emerg Infect Dis* **8**, 881-890 (2002).

41. Kovach, K., Davis-Fields, M., Irie, Y., Jain, K., Doorwar, S. et al. Evolutionary adaptations of biofilms infecting cystic fibrosis lungs promote mechanical toughness by adjusting polysaccharide production. *NPJ Biofilms Microbiomes* **3**, 1 (2017).

42. Berk, V., Fong, J.C., Dempsey, G.T., Develioglu, O.N., Zhuang, X. et al. Molecular architecture and assembly principles of *Vibrio cholerae* biofilms. *Science* **337**, 236-239 (2012).

43. Vidakovic, L., Singh, P.K., Hartmann, R., Nadell, C.D. & Drescher, K. Dynamic biofilm architecture confers individual and collective mechanisms of viral protection. *Nature microbiology* **3**, 26 (2018).

44. Stewart, E.J., Ganesan, M., Younger, J.G. & Solomon, M.J. Artificial biofilms establish the role of matrix interactions in staphylococcal biofilm assembly and disassembly. *Scientific reports* **5**, 13081 (2015).

45. Stewart, E.J., Satorius, A.E., Younger, J.G. & Solomon, M.J. Role of environmental and antibiotic stress on Staphylococcus epidermidis biofilm microstructure. *Langmuir* **29**, 7017-7024 (2013).

46. Gregor, I. & Enderlein, J. Image scanning microscopy. *Current opinion in chemical biology* **51**, 74-83 (2019).

47. Waldchen, S., Lehmann, J., Klein, T., van de Linde, S. & Sauer, M. Light-induced cell damage in live-cell super-resolution microscopy. *Scientific reports* **5**, 15348 (2015).

48. Chen, B.C., Legant, W.R., Wang, K., Shao, L., Milkie, D.E. et al. Lattice light-sheet microscopy: imaging molecules to embryos at high spatiotemporal resolution. *Science* **346**, 1257998 (2014).

49. Nadell, C.D., Ricaurte, D., Yan, J., Drescher, K. & Bassler, B.L. Flow environment and matrix structure interact to determine spatial competition in Pseudomonas aeruginosa biofilms. *eLife* **6** (2017).

50. Diaz-Pascual, F., Hartmann, R., Lempp, M., Vidakovic, L., Song, B. et al. Breakdown of Vibrio cholerae biofilm architecture induced by antibiotics disrupts community barrier function. *Nat Microbiol* (2019).

51. Chang, B.J., Kittisopikul, M., Dean, K.M., Roudot, P., Welf, E.S. et al. Universal light-sheet generation with field synthesis. *Nature methods* **16**, 235-238 (2019).

52. Dean, K.M., Roudot, P., Welf, E.S., Danuser, G. & Fiolka, R. Deconvolution-free Subcellular Imaging with Axially Swept Light Sheet Microscopy. *Biophys J* **108**, 2807-2815 (2015).

53. Chakraborty, T., Driscoll, M.K., Jeffery, E., Murphy, M.M., Roudot, P. et al. Light-sheet microscopy of cleared tissues with isotropic, subcellular resolution. *Nature methods* **16**, 1109-1113 (2019).

54. Wu, Y., Wawrzusin, P., Senseney, J., Fischer, R.S., Christensen, R. et al. Spatially isotropic four-dimensional imaging with dual-view plane illumination microscopy. *Nature biotechnology* **31**, 1032-1038 (2013).

55. Dunsby, C. Optically sectioned imaging by oblique plane microscopy. *Optics express* **16**, 20306-20316 (2008).

56. Voleti, V., Patel, K.B., Li, W., Perez Campos, C., Bharadwaj, S. et al. Real-time volumetric microscopy of in vivo dynamics and large-scale samples with SCAPE 2.0. *Nature methods* **16**, 1054-1062 (2019).

57. Flemming, H.C. & Wuertz, S. Bacteria and archaea on Earth and their abundance in biofilms. *Nat Rev Microbiol* **17**, 247-260 (2019).

58. York, A. in https://andrewgyork.github.io/high_na_single_objective_lightsheet/ (Github; 2020).

59. Bouchard, M.B., Voleti, V., Mendes, C.S., Lacefield, C., Grueber, W.B. et al. Swept confocally-aligned planar excitation (SCAPE) microscopy for high-speed volumetric imaging of behaving organisms. *Nature photonics* **9**, 113 (2015).

60. Ulman, V., Maska, M., Magnusson, K.E.G., Ronneberger, O., Haubold, C. et al. An objective comparison of cell-tracking algorithms. *Nature methods* **14**, 1141-1152 (2017).

61. Stringer, C., Wang, T., Michaelos, M. & Pachitariu, M. Cellpose: a generalist algorithm for cellular segmentation. *bioRxiv*, 2020.2002.2002.931238 (2020).

62. Moen, E., Bannon, D., Kudo, T., Graf, W., Covert, M. et al. Deep learning for cellular image analysis. *Nature methods*, 1 (2019).

63. Caicedo, J.C., Roth, J., Goodman, A., Becker, T., Karhohs, K.W. et al. Evaluation of Deep Learning Strategies for Nucleus Segmentation in Fluorescence Images. *Cytometry Part A* **95**, 952-965 (2019).

64. Reyer, M.A., McLean, E.L., Chennakesavalu, S. & Fei, J. An Automated Image Analysis Method for Segmenting Fluorescent Bacteria in Three Dimensions. *Biochemistry* **57**, 209-215 (2018).

65. Choi, H.I., Choi, S.W. & Moon, H.P. Mathematical theory of medial axis transform. *Pac J Math* **181**, 57-88 (1997).

66. Wang, J., Batabyal, T., Zhang, M.X., Zhang, J., Aziz, A. et al. LCuts: Linear Clustering of Bacteria Using Recursive Graph Cuts. *2019 IEEE International Conference on Image Processing (ICIP)*, 1575-1579 (2019).

67. Weigert, M., Schmidt, U., Haase, R., Sugawara, K. & Myers, G. in 2020 IEEE Winter Conference on Applications of Computer Vision (WACV) 3655-3662 (2020).

68. Liu, J., Prindle, A., Humphries, J., Gabalda-Sagarra, M., Asally, M. et al. Metabolic co-dependence gives rise to collective oscillations within biofilms. *Nature* **523**, 550-554 (2015).

69. Prindle, A., Liu, J., Asally, M., Ly, S., Garcia-Ojalvo, J. et al. Ion channels enable electrical communication in bacterial communities. *Nature* **527**, 59-63 (2015).

70. Humphries, J., Xiong, L., Liu, J., Prindle, A., Yuan, F. et al. Species-Independent Attraction to Biofilms through Electrical Signaling. *Cell* **168**, 200-209 e212 (2017).

71. Liu, J., Martinez-Corral, R., Prindle, A., Lee, D.-y.D., Larkin, J. et al. Coupling between distant biofilms and emergence of nutrient time-sharing. *Science* **356**, 638-642 (2017).

72. Mitri, S. & Foster, K.R. The genotypic view of social interactions in microbial communities. *Annu Rev Genet* **47**, 247-273 (2013).

73. Drescher, K., Nadell, C.D., Stone, H.A., Wingreen, N.S. & Bassler, B.L. Solutions to the public goods dilemma in bacterial biofilms. *Curr Biol* **24**, 50-55 (2014).

74. Persat, A., Nadell, C.D., Kim, M.K., Ingremeau, F., Siryaporn, A. et al. The mechanical world of bacteria. *Cell* **161**, 988-997 (2015).

75. Papenfort, K. & Bassler, B.L. Quorum sensing signal-response systems in Gram-negative bacteria. *Nat Rev Microbiol* **14**, 576-588 (2016).

76. Kroos, L. Highly signal-responsive gene regulatory network governing Myxococcus development. *Trends in Genetics* **33**, 3-15 (2017).

77. Moon, S., Yan, R., Kenny, S.J., Shyu, Y., Xiang, L. et al. Spectrally resolved, functional super-resolution microscopy reveals nanoscale compositional heterogeneity in live-cell membranes. *Journal of the American Chemical Society* **139**, 10944-10947 (2017).

78. Bramkamp, M. & Lopez, D. Exploring the existence of lipid rafts in bacteria. *Microbiol. Mol. Biol. Rev.* **79**, 81-100 (2015).

79. Zou, S.B., Hersch, S.J., Roy, H., Wiggers, J.B., Leung, A.S. et al. Loss of elongation factor P disrupts bacterial outer membrane integrity. *Journal of bacteriology* **194**, 413-425 (2012).

80. Gonelimali, F.D., Lin, J., Miao, W., Xuan, J., Charles, F. et al. Antimicrobial properties and mechanism of action of some plant extracts against food pathogens and spoilage microorganisms. *Frontiers in microbiology* **9**, 1639 (2018).

81. Parasassi, T., De Stasio, G., d'Ubaldo, A. & Gratton, E. Phase fluctuation in phospholipid membranes revealed by Laurdan fluorescence. *Biophysical journal* **57**, 1179-1186 (1990).

82. Strahl, H. & Hamoen, L.W. Membrane potential is important for bacterial cell division. *Proceedings of the National Academy of Sciences* **107**, 12281-12286 (2010).

83. Prindle, A., Liu, J., Asally, M., Ly, S., Garcia-Ojalvo, J. et al. Ion channels enable electrical communication in bacterial communities. *Nature* **527**, 59-63 (2015).

84. Krull, A., Vicar, T. & Jug, F. Probabilistic Noise2Void: Unsupervised content-aware denoising. *arXiv preprint arXiv:1906.00651* (2019).

85. Krull, A., Buchholz, T.-O. & Jug, F. in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition 2129-2137 (2019).

86. Buchholz, T.-O., Prakash, M., Krull, A. & Jug, F. DenoiSeg: Joint Denoising and Segmentation. *arXiv preprint arXiv:2005.02987* (2020).

87. Valm, A.M., Cohen, S., Legant, W.R., Melunis, J., Hershberg, U. et al. Applying systems-level spectral imaging and analysis to reveal the organelle interactome. *Nature* **546**, 162 (2017).

88. Glass, D.S. & Riedel-Kruse, I.H. A synthetic bacterial cell-cell adhesion toolbox for programming multicellular morphologies and patterns. *Cell* **174**, 649-658. e616 (2018).

89. Cotter, C.R., Schuttler, H.B., Igoshin, O.A. & Shimkets, L.J. Data-driven modeling reveals cell behaviors controlling self-organization during Myxococcus xanthus development. *Proc Natl Acad Sci U S A* **114**, E4592-E4601 (2017).

90. Gao, R., Asano, S.M., Upadhyayula, S., Pisarev, I., Milkie, D.E. et al. Cortical column and whole-brain imaging with molecular contrast and nanoscale resolution. *Science* **363**, eaau8302 (2019).

91. Rudge, T.J., Steiner, P.J., Phillips, A. & Haseloff, J. Computational modeling of synthetic microbial biofilms. *ACS Synthetic Biology* **1**, 345-352 (2012).

92. Reshes, G., Vanounou, S., Fishov, I. & Feingold, M. Cell shape dynamics in Escherichia coli. *Biophysical journal* **94**, 251-264 (2008).

93. Starruß, J., Peruani, F., Jakovljevic, V., Søgaard-Andersen, L., Deutsch, A. et al. Pattern-formation mechanisms in motility mutants of Myxococcus xanthus. *Interface focus* **2**, 774-785 (2012).

94. Harris, L.G., Foster, S. & Richards, R.G. An introduction to Staphylococcus aureus, and techniques for identifying and quantifying S. aureus adhesins in relation to adhesion to biomaterials: review. *Eur Cell Mater* **4**, 39-60 (2002).

95. Yan, T., Richardson, C.J., Zhang, M. & Gahlmann, A. Computational correction of spatially variant optical aberrations in 3D single-molecule localization microscopy. *Optics express* **27**, 12582-12599 (2019).

96. Muller, F.D., Schink, C.W., Hoiczyk, E., Cserti, E. & Higgs, P.I. Spore formation in Myxococcus xanthus is tied to cytoskeleton functions and polysaccharide spore coat deposition. *Mol Microbiol* **83**, 486-505 (2012).

97. Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M. et al. Fiji: an open-source platform for biological-image analysis. *Nature methods* **9**, 676-682 (2012).

98. Çiçek, Ö., Abdulkadir, A., Lienkamp, S.S., Brox, T. & Ronneberger, O. in International conference on medical image computing and computer-assisted intervention 424-432 (Springer, 2016).

99. Gibson, E., Li, W., Sudre, C., Fidon, L., Shakir, D.I. et al. NiftyNet: a deep-learning platform for medical imaging. *Computer Methods and Programs in Biomedicine* **158**, 113-122 (2018).

100. Van Valen, D.A., Kudo, T., Lane, K.M., Macklin, D.N., Quach, N.T. et al. Deep Learning Automates the Quantitative Analysis of Individual Cells in Live-Cell Imaging Experiments. *PLoS computational biology* **12**, e1005177 (2016).

101. Acton, S.T. Fast Algorithms for Area Morphology. *Digital Signal Processing* **11**, 187-203 (2001).

102. Jaccard, P. The distribution of the flora in the alpine zone. *New Phytologist* **11**, 37-50 (1912).

103. Yu, Y. & Acton, S.T. Speckle reducing anisotropic diffusion. *Trans. Img. Proc.* **11**, 1260-1270 (2002).

104. Tabassum, N., Vaccari, A. & Acton, S. Speckle removal and change preservation by distance-driven anisotropic diffusion of synthetic aperture radar temporal stacks. *Digital Signal Processing* **74**, 43-55 (2018).

105. Yan, T.Q. & Zhou, C.X. A Continuous Skeletonization Method Based on Distance Transform. *Comm Com Inf Sc* **304**, 251-258 (2012).

106. Shi, J.B. & Malik, J. Normalized cuts and image segmentation. *Ieee T Pattern Anal* **22**, 888-905 (2000).

107. Ester, M., Kriegel, H.-P., Sander, J. & Xu, X. in Proceedings of the Second International Conference on Knowledge Discovery and Data Mining 226-231 (AAAI Press, Portland, Oregon; 1996).

108. Zhang, M., Zhang, J., Wang, Y., Wang, J., Achimovich, A.M. et al. Non-invasive single-cell morphometry in living bacterial biofilms. *Nature communications* **11**, 1-13 (2020).

109. Flemming, H.C., Wingender, J., Szewzyk, U., Steinberg, P., Rice, S.A. et al. Biofilms: an emergent form of bacterial life. *Nat Rev Microbiol* **14**, 563-575 (2016).

110. Jayathilake, P.G., Gupta, P., Li, B., Madsen, C., Oyebamiji, O. et al. A mechanistic Individual-based Model of microbial communities. *PloS one* **12**, e0181965 (2017).

111. Persat, A., Nadell, C.D., Kim, M.K., Ingremeau, F., Siryaporn, A. et al. The mechanical world of bacteria. *Cell* **161**, 988-997 (2015).

112. De la Fuente-Núñez, C., Reffuveille, F., Fernández, L. & Hancock, R.E. Bacterial biofilm development as a multicellular adaptation: antibiotic resistance and new therapeutic strategies. *Current opinion in microbiology* **16**, 580-589 (2013).

113. Frost, I., Smith, W.P., Mitri, S., San Millan, A., Davit, Y. et al. Cooperation, competition and antibiotic resistance in bacterial colonies. *The ISME journal* **12**, 1582-1593 (2018).

114. Mimori-Kiyosue, Y. Imaging mitotic processes in three dimensions with lattice light-sheet microscopy. *Chromosome Research*, 1-14 (2021).

115. Rosenberg, J., Cao, G., Borja-Prieto, F. & Huang, J. Lattice light-sheet microscopy multi-dimensional analyses (LaMDA) of T-cell receptor dynamics predict T-cell signaling states. *Cell systems* **10**, 433-444. e435 (2020).

116. Manley, H.R., Potter, D.L., Heddleston, J.M., Chew, T.L., Keightley, M.C. et al. Frontline Science: Dynamic cellular and subcellular features of migrating leukocytes revealed by in vivo lattice lightsheet microscopy. *Journal of leukocyte biology* **108**, 455-468 (2020).

117. Jeckel, H. & Drescher, K. Advances and opportunities in image analysis of bacterial cells and communities. *FEMS microbiology reviews* (2020).

118. Wang, J., Batabyal, T., Zhang, M., Zhang, J., Aziz, A. et al. in 2019 IEEE International Conference on Image Processing (ICIP) 1575-1579 (IEEE, 2019).

119. Stringer, C., Wang, T., Michaelos, M. & Pachitariu, M. Cellpose: a generalist algorithm for cellular segmentation. *Nature methods* **18**, 100-106 (2021).

120. Weigert, M., Schmidt, U., Haase, R., Sugawara, K. & Myers, G. in Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision 3666-3673 (2020).

121. Scherr, T., Löffler, K., Böhland, M. & Mikut, R. Cell segmentation and tracking using CNN-based distance predictions and a graph-based matching strategy. *PloS one* **15**, e0243219 (2020).

122. Meyer, F. & Beucher, S. Morphological segmentation. *Journal of visual communication and image representation* **1**, 21-46 (1990).

123. Otsu, N. A threshold selection method from gray-level histograms. *IEEE transactions on systems, man, and cybernetics* **9**, 62-66 (1979).

124. Laasmaa, M., Vendelin, M. & Peterson, P. Application of regularized Richardson–Lucy algorithm for deconvolution of confocal microscopy images. *Journal of microscopy* **243**, 124-140 (2011).

125. Liu, R. & Jia, J. in 2008 15th IEEE International Conference on Image Processing 505-508 (IEEE, 2008).

126. Starck, J.-L., Pantin, E. & Murtagh, F. Deconvolution in astronomy: A review. *Publications of the Astronomical Society of the Pacific* **114**, 1051 (2002).

127. Wang, J., Zhang, M., Zhang, J., Wang, Y., Gahlmann, A. et al. Graph-theoretic post-processing of segmentation with application to dense biofilms. *IEEE Transactions on Image Processing* **30**, 8580-8594 (2021).

128. Cutler, K.J., Stringer, C., Wiggins, P.A. & Mougous, J.D. Omnipose: a high-precision morphology-independent solution for bacterial cell segmentation. *bioRxiv* (2021).

129. Laine, R.F., Arganda-Carreras, I., Henriques, R. & Jacquemet, G. Avoiding a replication crisis in deep-learning-based bioimage analysis. *Nature methods* **18**, 1136-1144 (2021).

130. Magnusson, K.E., Jalden, J., Gilbert, P.M. & Blau, H.M. Global linking of cell tracks using the Viterbi algorithm. *IEEE Trans Med Imaging* **34**, 911-929 (2015).

131. Moen, E., Borba, E., Miller, G., Schwartz, M., Bannon, D. et al. Accurate cell tracking and lineage construction in live-cell imaging experiments with deep learning. *bioRxiv*, 803205 (2019).

132. Thormann, K.M., Saville, R.M., Shukla, S., Pelletier, D.A. & Spormann, A.M. Initial phases of biofilm formation in Shewanella oneidensis MR-1. *Journal of bacteriology* **186**, 8096-8104 (2004).

133. Pettersen, E.F., Goddard, T.D., Huang, C.C., Meng, E.C., Couch, G.S. et al. UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Science* **30**, 70-82 (2021).

134. Venkateswaran, K., Moser, D.P., Dollhopf, M.E., Lies, D.P., Saffarini, D.A. et al. Polyphasic taxonomy of the genus Shewanella and description of Shewanella oneidensis sp. nov. *International Journal of Systematic and Evolutionary Microbiology* **49**, 705-724 (1999).

135. He, K., Zhang, X., Ren, S. & Sun, J. in Proceedings of the IEEE conference on computer vision and pattern recognition 770-778 (2016).

136. Liao, P.-S., Chen, T.-S. & Chung, P.-C. A fast algorithm for multilevel thresholding. *J. Inf. Sci. Eng.* **17**, 713-727 (2001).

137. Van der Walt, S., Schönberger, J.L., Nunez-Iglesias, J., Boulogne, F., Warner, J.D. et al. scikit-image: image processing in Python. *PeerJ* **2**, e453 (2014).

138. Zunair, H., Rahman, A., Mohammed, N. & Cohen, J.P. in International Workshop on PRedictive Intelligence In MEdicine 156-168 (Springer, 2020).

139. Jaccard, P. The distribution of the flora in the alpine zone. 1. *New phytologist* **11**, 37-50 (1912).

140. Csurka, G., Larlus, D., Perronnin, F. & Meylan, F. in Bmvc, Vol. 27 10.5244 (2013).

141. Zhang, J., Wang, Y., Donarski, E.D., Toma, T.T., Miles, M.T. et al. BCM3D 2.0: accurate segmentation of single bacterial cells in dense biofilms using computationally generated intermediate image representations. *npj Biofilms and Microbiomes* **8**, 99 (2022).

142. Matula, P., Maska, M., Sorokin, D.V., Matula, P., Ortiz-de-Solorzano, C. et al. Cell Tracking Accuracy Measurement Based on Comparison of Acyclic Oriented Graphs. *PloS one* **10**, e0144959 (2015).

143. Tinevez, J. SIMPLETRACKER a simple particle tracking algorithm that can deal with gaps. *Jean-Yves Tinevez* (2012).

144. Wen, C., Miura, T., Voleti, V., Yamaguchi, K., Tsutsumi, M. et al. 3DeeCellTracker, a deep learning-based pipeline for segmenting and tracking cells in 3D time lapse images. *eLife* **10**, e59187 (2021).

145. Chen, T. & Guestrin, C. in Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining 785-794 (2016).

146. Ho, T.K. in Proceedings of 3rd international conference on document analysis and recognition, Vol. 1 278-282 (IEEE, 1995).

147. Hochreiter, S. & Schmidhuber, J. Long short-term memory. *Neural computation* **9**, 1735-1780 (1997).

148. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L. et al. Attention is all you need. *Advances in neural information processing systems* **30** (2017).

149. Lewis, K. Persister cells and the riddle of biofilm survival. *Biochemistry (Moscow)* **70**, 267-274 (2005).

150. Baranchuk, D., Rubachev, I., Voynov, A., Khrulkov, V. & Babenko, A. Label-efficient semantic segmentation with diffusion models. *arXiv preprint arXiv:2112.03126* (2021).

151. Kim, H.J., Huh, D., Hamilton, G. & Ingber, D.E. Human gut-on-a-chip inhabited by microbial flora that experiences intestinal peristalsis-like motions and flow. *Lab on a chip* **12**, 2165-2174 (2012).