Received: 23 September 2022 | Revised: 31 March 2023 | Accepted: 2 April 2023

DOI: 10.1111/mmi.15064

RESOURCES

WILEY

Single-cell segmentation in bacterial biofilms with an optimized deep learning method enables tracking of cell lineages and measurements of growth rates

Eric Jelli [1,2,3] | Takuya Ohmura [2,4] | Niklas Netter [2,4] | Martin Abt [2] | Eva Jiménez-Siebert [2,3,4] | Konstantin Neuhaus [2,3,4] | Daniel K. H. Rode [2,3,4] | Carey D. Nadell [5] | Knut Drescher [2,3,4]

[1]Max Planck Institute for Neurobiology of Behavior – caesar, Bonn, Germany

[2]Max Planck Institute for Terrestrial Microbiology, Marburg, Germany

[3]Department of Physics, Philipps-Universität Marburg, Marburg, Germany

[4]Biozentrum, University of Basel, Basel, Switzerland

[5]Department of Biological Sciences, Dartmouth College, Hanover, USA

Correspondence
Knut Drescher, Biozentrum, University of Basel, Switzerland.
Email: knut.drescher@unibas.ch

Funding information
Bundesministerium für Bildung und Forschung, Grant/Award Number: TARGET-Biofilms; Deutsche Forschungsgemeinschaft, Grant/Award Number: DR 982/5-1 and DR 982/6-1; European Commission, Grant/Award Number: 716734 and 955910; Human Frontier Science Program, Grant/Award Number: LT000013/2019-C; Minna-James-Heineman-Stiftung; Schweizerischer Nationalfonds zur Förderung der Wissenschaftlichen Forschung, Grant/Award Number: 51NF40_180541 and TMCG-3_213801

Abstract

Bacteria often grow into matrix-encased three-dimensional (3D) biofilm communities, which can be imaged at cellular resolution using confocal microscopy. From these 3D images, measurements of single-cell properties with high spatiotemporal resolution are required to investigate cellular heterogeneity and dynamical processes inside biofilms. However, the required measurements rely on the automated segmentation of bacterial cells in 3D images, which is a technical challenge. To improve the accuracy of single-cell segmentation in 3D biofilms, we first evaluated recent classical and deep learning segmentation algorithms. We then extended StarDist, a state-of-the-art deep learning algorithm, by optimizing the post-processing for bacteria, which resulted in the most accurate segmentation results for biofilms among all investigated algorithms. To generate the large 3D training dataset required for deep learning, we developed an iterative process of automated segmentation followed by semi-manual correction, resulting in >18,000 annotated *Vibrio cholerae* cells in 3D images. We demonstrate that this large training dataset and the neural network with optimized post-processing yield accurate segmentation results for biofilms of different species and on biofilm images from different microscopes. Finally, we used the accurate single-cell segmentation results to track cell lineages in biofilms and to perform spatiotemporal measurements of single-cell growth rates during biofilm development.

KEYWORDS
3D segmentation, biofilm, deep learning, image analysis, image cytometry, *Vibrio cholerae*

Eric Jelli, Takuya Ohmura and Niklas Netter equal contribution.

1 | INTRODUCTION

Bacterial biofilms can be found in diverse environments and are highly abundant in the major microbial habitats on Earth (Costerton et al., 1978, 1987; Flemming & Wuertz, 2019). The cells inside these communities are bound together via a self-produced extracellular matrix, which can also protect the community from environmental stresses, phages, and predatory bacteria (Flemming & Wingender, 2010). As biofilms are comprised of metabolically active densely packed cells, resource gradients emerge, which leads to spatially distributed physiological heterogeneity inside biofilms (Ackermann, 2015; Jo et al., 2022; Evans et al., 2020; Stewart & Franklin, 2008). Additional physiological heterogeneity in biofilms can arise from stochastic gene expression at the single-cell level, and due to the presence of multiple genotypes. To investigate the mechanisms that contribute to or result from single-cell level heterogeneity in biofilms, robust methodologies for live single-cell analyses in situ in biofilms are required.

Advances in 3D fluorescence microscopy techniques and fluorescent reporter brightness have enabled the observation of individual cells in live 3D biofilms (Drescher et al., 2016; Hartmann et al., 2019; Stewart et al., 2013; Yan et al., 2016; Zhang et al., 2020). For densely packed biofilms, single-cell resolution is currently limited to regions of biofilms that are less than approximately 40 μm thick along the optical axis, mostly due to intense scattering of the fluorescence emission light by the cells in biofilms. Fluorescence-based single-cell imaging also requires excitation light intensities that may cause phototoxicity and interfere with biofilm physiology. Phototoxicity is particularly problematic in point scanning confocal microscopes so that more gentle 3D fluorescence microscopy techniques, such as multi-point scanning confocal and light sheet microscopy techniques, are better suited for time-lapse 3D imaging of live biofilms (Hartmann et al., 2019; Qin et al., 2020; Yordanov et al., 2021; Zhang et al., 2020).

Performing measurements on the single cells that can be observed in 3D microscopy images of biofilms requires the automated detection of each cell's boundary in the images, which is a process termed instance segmentation in computer vision research. Based on the instance segmentation results, image cytometry tools such as BiofilmQ (Hartmann et al., 2021) can then be used for quantifying single-cell parameters in biofilms. However, achieving instance segmentation with high accuracy has been a difficult challenge for 3D biofilm images because the cells are very closely spaced and the signal-to-background ratio is often low. Recent studies performing single-cell segmentation in biofilms still rely on sophisticated classical image analysis approaches based on filtering and thresholding (Hartmann et al., 2019; Qin et al., 2020). Such classical image analysis requires significant fine-tuning of parameters for different experimental settings, fluorescence labels, and cell shapes (Caicedo, Goodman, et al., 2019; Caicedo, Roth, et al., 2019), which limits their transferability between different labs and, therefore, their usefulness to the community. Furthermore, the segmentation accuracy of these classical segmentation approaches for biofilms has not been sufficient to perform measurements that rely on prolonged cell tracking, such as single-cell growth rate measurements. Data-driven segmentation approaches based on convolutional neural networks (CNNs) alleviate the problem of fine-tuning parameters of the analysis pipeline by 'learning' the segmentation task from annotated images (Çiçek et al., 2016; Stringer et al., 2021; Weigert et al., 2020; Wolny et al., 2020). In addition, CNN approaches often generate better segmentation performance compared with classical segmentation approaches (Laine et al., 2021; Moen et al., 2019; von Chamier et al., 2019). However, CNN segmentation methods come with a caveat: a suitable amount of training data needs to be generated.

Training data for CNNs for image segmentation consist of a raw image file and a file in which all cell outlines are correctly annotated, which is used as a learning target for the network. The manual annotation of single cells in 3D image stacks constitutes a serious barrier to the wide-spread implementation of deep learning workflows for cell segmentation, as manual labeling of 3D images is very time-consuming (Moen et al., 2019). Resources could be saved by only annotating the minimally required number of cells for the training of maximally accurate CNN models. Quantifications of minimally required annotations have been reported systematically for two-dimensional (2D) but not for 3D problems (Falk et al., 2019; Van Valen et al., 2016). As an alternative to human-annotated ground truth data, generation of synthetic training data was suggested (Hardo et al., 2022; Toma et al., 2022; Zhang et al., 2020). Nevertheless, there is always a gap between the real and synthetic data, which only can be bridged with elaborate randomization approaches (Tobin et al., 2017). To speed up annotation processes, iterative human-in-the-loop annotation schemes have been adapted for eukaryotic systems (Suga et al., 2021; Chen et al., 2020; Wang et al., 2021; Wolny et al., 2020).

Here, we develop a human-in-the-loop annotation framework to generate a dataset of more than 18,000 cell annotations in three-dimensional biofilms. Using this dataset, we determined the minimum number of 3D annotated cells for obtaining accurate deep learning segmentations. We also used this large training dataset to evaluate the performance of recent deep learning algorithms (Schmidt et al., 2018; Stringer et al., 2021; Weigert et al., 2020; Zhang et al., 2020, 2022) for single-cell segmentation in bacterial biofilms. For the well-performing StarDist algorithm, we developed a post-processing algorithm that optimizes the cell segmentation in biofilms, which outperforms the previous state-of-the-art segmentation methods for biofilms. By applying the trained segmentation algorithm to 3D images of biofilms of different species and microscopes, we revealed the robustness of our method. The accuracy of the optimized CNN-based segmentation method enabled us to perform spatiotemporally resolved measurements of single-cell growth rates in *Vibrio cholerae* biofilms and to accurately track cell lineages during biofilm growth.

2 | RESULTS

Following the standard approach of deep learning workflows for image cytometry (Jeckel & Drescher, 2021; LeCun et al., 2015),

we set up the image analysis pipeline for this study in three steps (Figure 1): We employed an annotation scheme to generate the training data used for training a CNN model (Figure 1a), as described in detail below. After training the CNN model, we applied the CNN model to volumetric images that were not used during the model training, to identify individual cells in 3D bacterial biofilm images (Figure 1b). Subsequently, the resulting instance segmentation can be used to extract single-cell parameters and whole-community parameters in an image cytometry analysis (Figure 1c). As a model system for improving single-cell image segmentation algorithms for 3D biofilm images, we used *V. cholerae* biofilms grown in microfluidic flow chambers. To evaluate the final segmentation algorithm, we also applied it to images of biofilms formed by other species, as described below.

2.1 | Generating 3D training data with an iterative semi-automated annotation scheme

Completely manual annotation of cells in 3D image datasets is a time-consuming process: In our dataset of bacterial cells in biofilms (*z*-spacing of 100 nm, *xy* pixel size of 63.2 × 63.2 nm^2) we found that the manual 3D annotation of a single cell takes 5 ± 2 min by an experienced person, and we noticed that 3D annotation of many cells is mentally difficult to endure due to the slow progress and sometimes ambiguous fluorescence signals from cells in close proximity to each other.

We, therefore, developed a workflow to accelerate our annotation procedure of bacterial cells in 3D images (Figure 2). For this, we first used a previously published segmentation pipeline (Hartmann et al., 2019), including an algorithmic improvement published recently (Jeckel et al., 2022), to perform a segmentation of a raw 3D biofilm image (Figure 2, step 1). This segmentation pipeline relies on classical image analysis methods (filtering, thresholding, watershed algorithm) and generates a good, but not perfect, instance segmentation result. This first segmentation result was inspected visually by a human and segmentation errors were manually corrected, using the napari software (Sofroniew et al., 2022). Based on this small 3D training dataset, a first CNN model was trained (Figure 2, step 2). As the CNN for this task, we chose StarDist (Weigert et al., 2020), an easy-to-train, state-of-the-art single-cell segmentation approach with a well-written code base, albeit with a restriction on segmenting star-convex shapes. In step 3 of our workflow for training data generation (Figure 2, step 3), we then used the first CNN model to obtain the instance segmentation prediction for a second raw 3D biofilm image. Due to the small amount of training data that the first CNN model was trained with, the segmentation result was poor, including many cases of over-segmentation errors. However, we observed that the cases of segmentation errors involved significant overlaps in the volume of the predicted objects, which enabled us to automatically identify the regions that require corrections. With a simple graphical user interface (Figure S1), a human annotator was thus able to iterate through the segmentation errors and perform an automated object fusion, an automated object splitting, or mark the object for a manual re-annotation. This automated error identification, followed by a human in the loop for error correction resulted in the corrected segmentation result for the second raw 3D biofilm image, which took approximately 8 h of annotation time (corresponding to 5 s per cell for 6077 cells, Figure S2). The corrected segmentation dataset was then added to the pool of training data, which was then used to train a new version of the network from scratch, that is, the second CNN model. This process was then repeated until *n* 3D raw biofilm images were correctly annotated, resulting in a pool of training datasets that was then used to create the n^{th} CNN model (Figure 2, step 3).

With this workflow for semi-automated annotation, we annotated all individual cells in *n* = 5 biofilm images (Figure S2), resulting

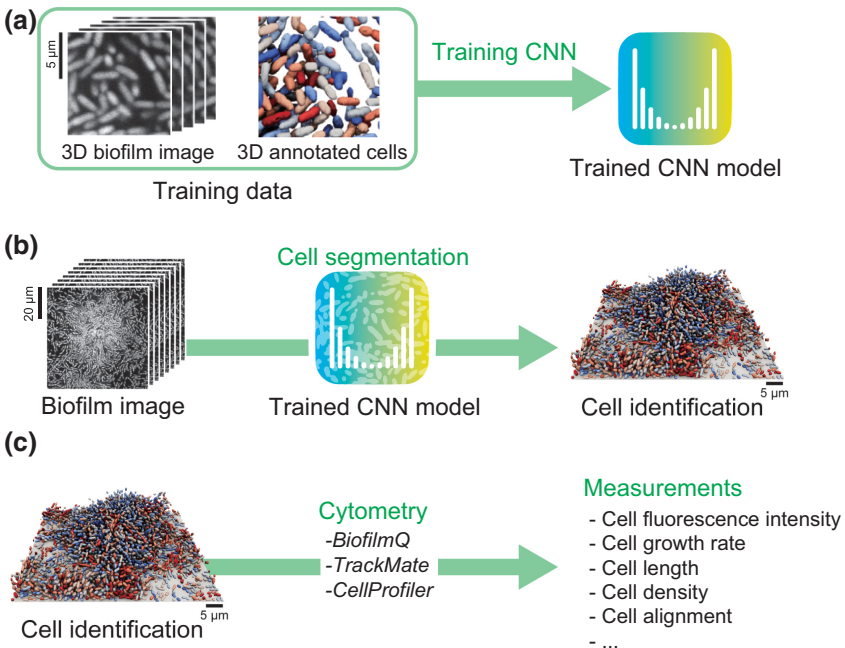

FIGURE 1 Deep-learning-based workflow for single-cell measurements in three-dimensional biofilms. (a) A convolutional neural network (CNN) model is trained with volumetric microscope images and a corresponding label image in which all cells are individually annotated. (b) The trained CNN model can be used to predict individual cell segmentations in previously unseen 3D images of biofilms. (c) Based on the single-cell segmentation, existing image cytometry tools, such as BiofilmQ, TrackMate, or CellProfiler, can be used to quantify single-cell parameters and BiofilmQ can also be used to quantify whole-biofilm parameters.

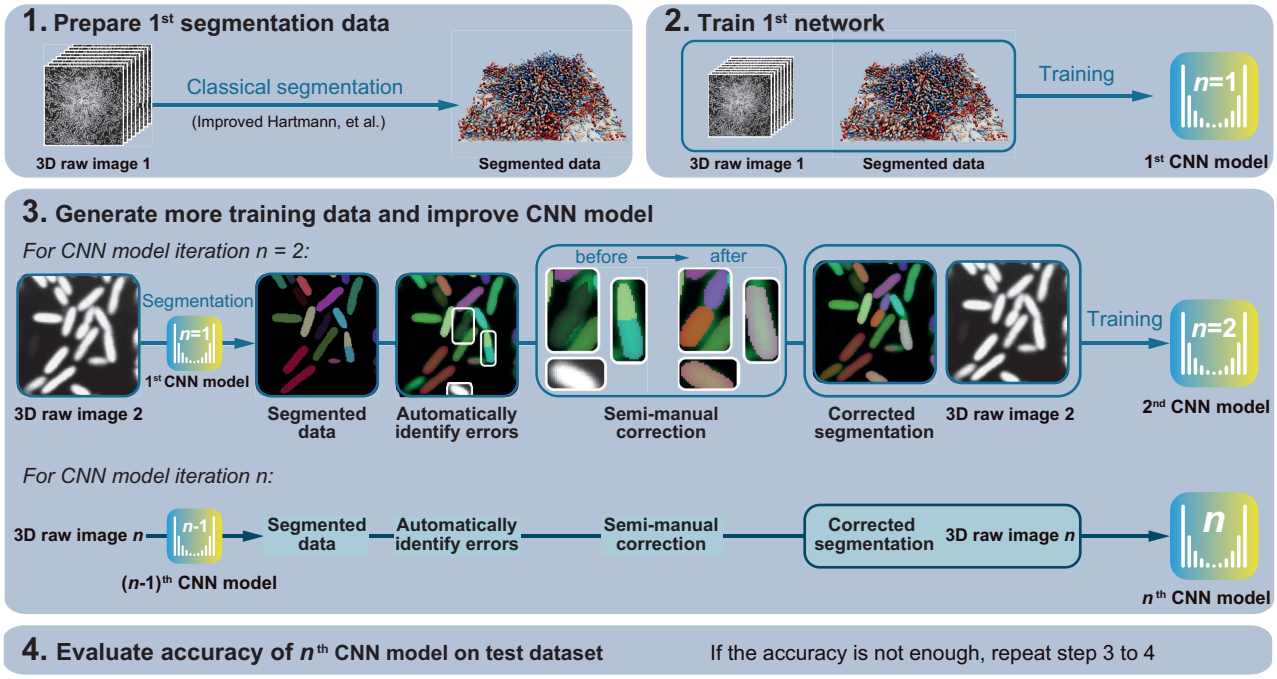

FIGURE 2 Workflow employed for obtaining convolutional neural network (CNN) training data of 3D biofilms with segmented cells. Step (1): Raw 3D microscopy images of biofilms are segmented with the "Improved Hartmann et al." method (see Experimental Procedures), which is a classical segmentation method based on filtering, as well as edge detection and watershed algorithms. The segmentation output was visually inspected by a human, and segmentation errors were corrected manually by re-annotation of cells. The resulting manually corrected segmentation data can be used to train the 1st CNN model. Step (2): The 1st CNN model is trained based on the raw image data and the 1st segmentation data obtained in Step 1. Step (3): Additional training data is generated by using a new, previously-unseen 3D biofilm image as input for the 1st CNN model, to generate a segmentation for this image. The resulting segmentation is semi-manually corrected for errors, using the tool described in Figure S1. The raw image and the corrected segmentation are then added to the pool of training data, which is then used to train the second version of the CNN model. This process of adding new training data is performed *n* times, ultimately resulting in *n* different pairs of 3D biofilm images and manually corrected segmentations, which are used as training data for the n^{th} CNN model. Step (4): After each iteration, the n^{th} CNN model is evaluated on a test dataset that was not included in the training data. The iteration process ends once the evaluation reaches saturation.

in annotations for 18,868 cells in 3D. This large training dataset is available for download as described in the Data and Code Availability section below. From this pool of manually corrected annotation data, four biofilms were used for training the final CNN model, and one biofilm was designated as the "test dataset", which was not used for training and instead only used for testing the accuracy of the segmentation result of the trained CNN models. This amount of training data saturated the average precision of the segmentation that can be obtained with the StarDist CNN (Figure 2, step 4).

To test the quality of the semi-manual iterative annotation process described in Figure 2, we generated a completely manual annotation for a subset of our dataset and compared the resulting two segmentations (Figure S3). Cell outlines in the fully-manual annotation were systematically drawn tighter around the cells (corresponding to smaller cells) compared with the semi-manual dataset (Figure S3a,c). By visual inspection, we found that most significant differences between the fully-manual and semi-manual annotations were due to over-segmented or poorly annotated cells in the fully-manual segmentation (Figure S3c–e). The fully manual segmentation result was, therefore, of lower quality than the semi-manual segmentation result.

2.2 | Increasing bacterial segmentation accuracy by optimizing the post-processing of the convolutional neural network output

Despite training the StarDist CNN with the manually corrected training dataset that is comprised of highly accurate cell shape annotations, the StarDist segmentation prediction on previously unseen 3D biofilm images contained a large number of cases of over-segmentation (Figure S5). Even though the StarDist outputs (the 3D segmentation label probability maps and the 3D distance maps) represented the cell shapes remarkably well (Figure 3a), StarDist's non-maximum-suppression post-processing step was not able to convert these outputs into accurate segmentations. These algorithmic problems ultimately stem from the fact that StarDist was designed for detecting star-convex shapes, and not for very elongated and curved cell shapes such as those of rod-shaped bacteria. To develop a solution, we first required a careful diagnosis of the StarDist segmentation problem for bacterial cells: For many cells in our *V. cholerae* biofilms, only very few voxels in the cell can be regarded as the center of a star-convex object while many more voxels of the cell are similarly or even more distant to the edge of the cell (Figure S6).

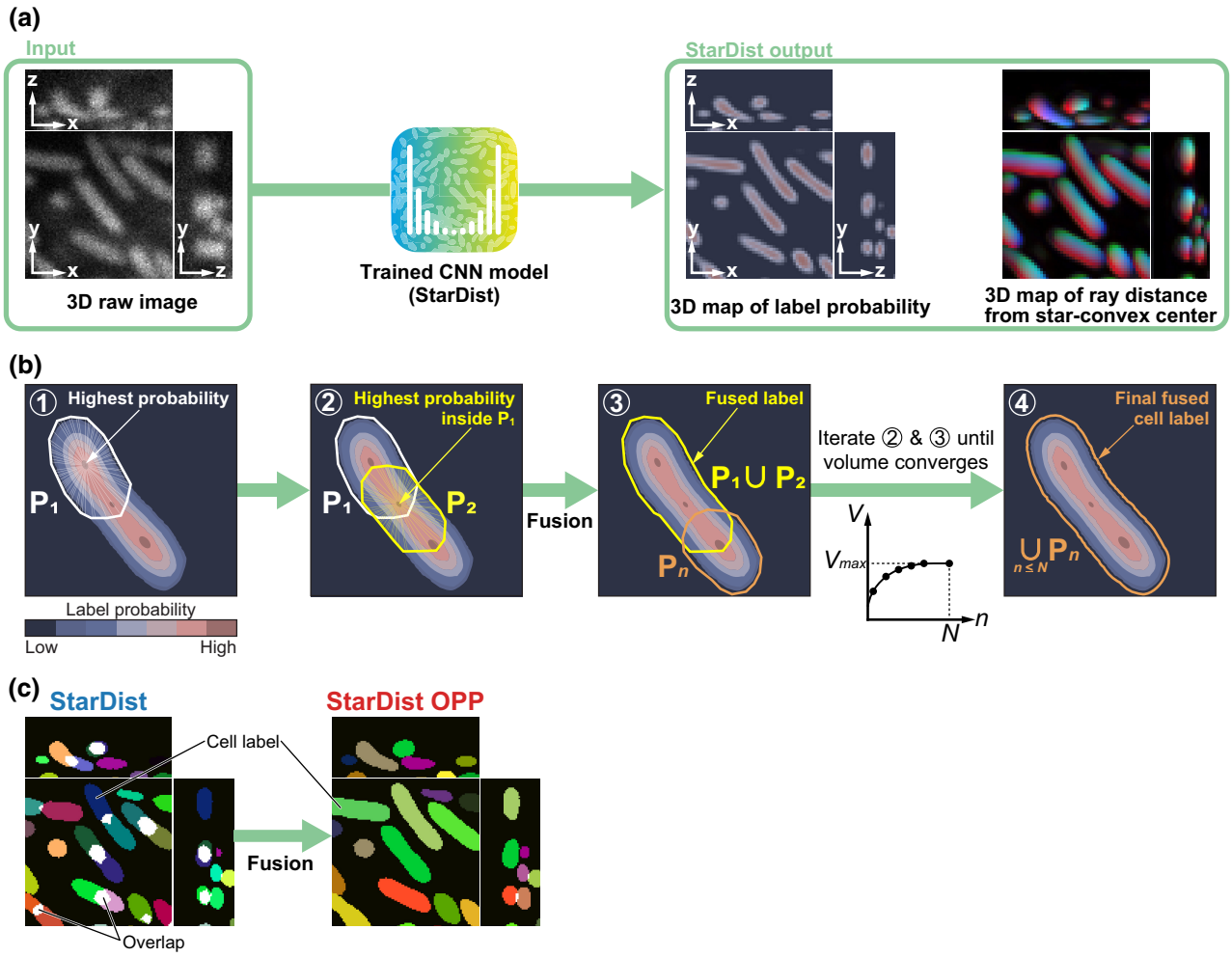

FIGURE 3 Algorithm for optimizing post-processing of the StarDist segmentation output avoids over-segmentation. (a) When the StarDist convolutional neural network (CNN) model is used to generate a single-cell segmentation of a 3D raw image, the output of StarDist is a 3D map of the "label probability", which quantifies the probability of each pixel to be inside a cell. In addition, StarDist generates multiple 3D distance maps, which describe the length of the StarDist rays spanning the star-convex polyhedra that approximate the cell shapes. Depicted are three distance maps on top of each other, using RGB values of the three distance map values. (b) Schematic illustration of the strategy underlying StarDist OPP: In Step 1, the pixel with the highest label probability is chosen and the corresponding polyhedron P_1 is constructed based on the distance map entries at that position, and P_1 is assumed to be the current cell shape. In Step 2, the polyhedron P_2 corresponding to the pixel within the current cell shape with the next highest probability is generated. In Step 3, cell shapes P_1 and P_2 are fused together to yield a new cell shape $P_1 \cup P_2$. Steps 2 and 3 are iteratively repeated until the volume V of the cell shape reaches V_{max} and no longer increases (see schematic graph), which yields the final fused cell segmentation label (Step 4). (c) Comparison of predictions by StarDist (left) and StarDist OPP (right) for the 3D region depicted in panel A. Cell labels in StarDist involve the overlap of regions with high label probability, which indicate segmentation problems. Colors indicate different cells, overlaps between different cell predictions are white, and the background is black.

Based on the distance to the object-edge, the StarDist neural network assigns a label probability to each voxel. Hence, for some bacterial cells, the correct star-convex center does not get assigned the highest label probability (Figure S6). Since StarDist's non-maximum-suppression algorithm reconstructs the segmented cell shape on the basis of the cell shape information in the voxel of highest label probability, incomplete shapes are reconstructed. In the next step, the voxels that should be part of the cell but were not included into the incomplete shape are picked up by the non-maximum-suppression algorithm to predict another incomplete cell shape. As a result, a single bacterial cell is often over-segmented into two or more objects.

To overcome the limitations of the StarDist non-maximum-suppression post-processing, we developed a new post-processing algorithm that reconstructs the bacterial cell shape based on multiple predicted shapes (Solovyev et al., 2021), instead of considering only the shape information in the voxel with highest assigned label probability (Figure 3b). Similar to the non-maximum-suppression algorithm, we start with the voxel of highest label probability and reconstruct the object shape associated with it. Within that object shape, we identify the pixel of the second highest probability and reconstruct the object shape associated with it. Then, these two shapes are fused and the resulting shape is iteratively extended

by the information associated with the pixel of highest probability that has not yet been taken into account. This iterative process stops once there are no pixels left within the current shape whose probability exceeds a predefined cutoff value, or once the volume of the current shape is not extended for a predefined number of iterations. After completion of this iterative shape-fusion process, a new object is generated starting at the pixel with the next highest remaining probability. With this optimized post-processing algorithm for StarDist, termed StarDist OPP, we obtain segmentation results with significantly less over-segmentation (Figures 3c and S5). Additionally, we noticed that StarDist overestimates the size of object shapes, and segmentation performance improves drastically when reducing the extents present in the 3D distance maps by one pixel. Finally, by testing StarDist OPP on synthetically generated training data, we ruled out any algorithmic limitations of StarDist OPP regarding curved, rod-like objects (Figures S8 and SI Text S1). We made StarDist OPP available as a python package as described in the Data and Code Availability section. A quantitative comparison of StarDist OPP with other segmentation methods is provided below.

2.3 | Performance comparison of StarDist OPP relative to alternative segmentation pipelines

We quantitatively evaluated the segmentation performance of StarDist OPP relative to the basic StarDist network, and different CNNs that were recently developed for bacterial cell segmentation: Cellpose (Stringer et al., 2021), a multi-class U-Net (Zhang et al., 2020), and BCM3D 2.0 (Zhang et al., 2022). All deep learning methods in this comparison were trained with the same ground truth dataset. In addition, we compared the state-of-the-art classical algorithms that are optimized for bacterial cell segmentation in *V. cholerae* biofilms by Hartmann et al., 2019, and an improved version of this algorithm based on a better watershed seed selection (Jeckel et al., 2022). For each of these segmentations, we computed the deviations from the ground truth. The performance of each segmentation result was quantified for different intersection-over-union (IoU) threshold values between 0 and 1. The IoU threshold determines how close the predicted shape and location of an object matches objects in the ground truth (Jeckel & Drescher, 2021; Weigert et al., 2020). Segmented objects were marked as true positive (TP) when the IoU with a single ground truth object exceeded the threshold; otherwise, they were marked as false positive (FP). Ground truth objects that could not be assigned to a predicted object above the chosen IoU threshold were marked as false negatives (FNs). Based on these three numbers, we calculated various metrics, such as the average precision, corresponding to the ratio TP/(TP + FN + FP) (Figure 4) (Taha & Hanbury, 2015).

Based on the average precision, StarDist OPP outperforms other segmentation methods for all IoU thresholds (Figure 4a). In comparison with the multi-class U-Net, StarDist OPP has more TPs and less FNs, while maintaining the same amount of FPs (Figure 4b–d). The improved performance of StarDist OPP is also visible when looking at the FNs and FPs in 3D renderings of biofilms (Figure 4e). In addition to the IoU-based performance metrics, we observed that StarDist OPP results in several qualitative segmentation improvements. This includes the observation that the distribution of cell aspect ratios (the ratio of cell length to width) in the training data is more closely matched by StarDist OPP compared with StarDist and the multi-class U-Net (Figure 4f). In addition, we observe that StarDist OPP has less cell–cell overlaps compared with StarDist. While small overlapping regions (volume $< 0.06\,\mu m^3$) generally arise due to ambiguity of the exact boundary separating neighboring cells, large overlapping regions arise due to systematic limitations in the post-processing. StarDist generates more of such large overlaps, which are visible as a peak at around $0.15\,\mu m^3$. These overlaps correspond to over-segmentation errors, which are identified here in an IoU-independent metric as follows: To quantify the relative abundance of these over-segmentation cases, we define overlaps with volumes higher than $0.06\,\mu m^3$ as over-segmentation errors and their total number as N_O. For each over-segmentation error, an additional cell was predicted, hence, we define the relative over-segmentation abundance as $N_O/(N_C\text{-}N_O)$, where N_C is the total number of predicted cells. The relative over-segmentation abundances are 14% and 3% for StarDist and StarDist OPP, respectively, indicating that StarDist generates more systematic errors.

Compared with both StarDist variants and the multi-class U-Net, the Cellpose CNN did not perform well (Figure 4a–d). This low performance likely results from the fact that Cellpose generates segmentation predictions for 2D image slices, which are then merged together for a 3D segmentation, whereas StarDist OPP predicts segmentation predictions directly for 3D images. Interestingly, the improved variant of the classical segmentation algorithm by Hartmann et al. (with manually fine-tuned segmentation parameters) performs only slightly worse than the StarDist methods and similarly good as the BCM3D 2.0 deep learning approach (Figure 4a). To visualize the difference in the performance of the improved Hartmann et al. algorithm and the StarDist variants, correctly detected cells, over-segmentation errors, as well as undetected cells are shown in biofilms side by side in Figure 4e (Ahrens et al., 2005). These renderings of the FNs and FPs in the biofilm reveal that the Hartmann et al. algorithm has a systematic segmentation problem towards the top of the biofilm, likely due to the reduction of fluorescence signal with increasing height in the biofilm.

As StarDist and the improved Hartmann et al. algorithm were involved in the semi-manual generation of the training and test datasets (Figure 2), we suspected that the dataset itself had an intrinsic bias that allowed these methods to perform better. Therefore, we validated the different segmentation methods against a ground truth dataset that was obtained by completely manual annotation (Figure S3). Compared with the fully manually annotated segmentation, StarDist OPP and StarDist again outperformed the Hartmann et al. results (Figure S3b). The curve of StarDist OPP in Figure S3b is on par with the curve of the semi-manual annotation, suggesting our trained StarDist OPP model used all the information in the training data. In conclusion, we have no indication that an intrinsic bias of our dataset is responsible for the good performance of StarDist OPP.

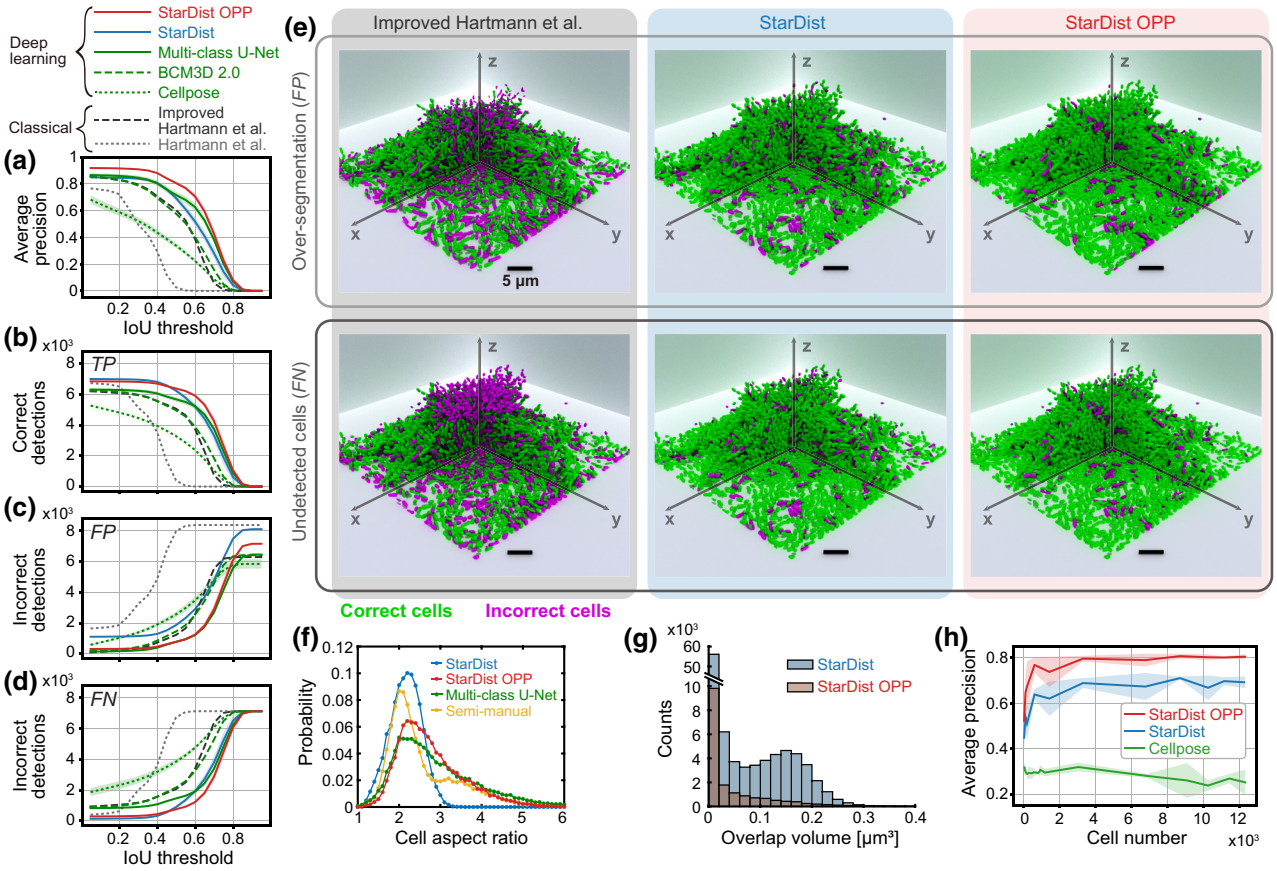

FIGURE 4 Performance evaluation of different methodologies for single-cell segmentation in 3D biofilms. (a) Average precisions of two classical segmentation approaches (Hartmann et al., Improved Hartmann et al.), four convolutional neural network (CNN) models (StarDist, multi-class U-Net, BCM3D 2.0, Cellpose), and a CNN model followed by the optimized post-processing strategy developed in this study (StarDist OPP). The average precision of a segmentation method is defined as the number of true positive detections divided by the sum of the true positive (TP), false positive (FP), and the false negative (FN) detections, that is, average precision = TP/(TP + FP + FN). The average precision is measured as a function of the intersection over union (IoU) between the segmentation result and the ground truth. (b) Number of true positive objects, corresponding to correctly identified cells, as a function of IoU. (c) Number of false positive objects, corresponding to incorrectly segmented (i.e., over-segmented) cells. (d) Number of false negative objects, corresponding to undetected cells. (e) 3D image renderings of biofilms, with colors indicating the cells that are segmented correctly (green) or incorrectly (magenta). The top row of images indicates over-segmentation (false positive objects) in magenta; the bottom row of images indicates undetected cells (false negative objects) in magenta. For illustrative purposes, a quarter of the 3D biofilm volume was removed to enable a view inside the biofilm core. (f) Probability distributions of cell aspect ratios (i.e., the ratio between cell length and cell width) in the training data (yellow data) and in the segmentations from StarDist (blue), StarDist OPP (red), and a multi-class U-Net (green). (g) Histograms of the volumes of overlaps between predicted cells for StarDist (blue) and StarDist OPP (red). (h) Dependence of average precision on the number of annotated cells in the training data.

To further examine the quality of the segmentation results generated by StarDist OPP without relying on the semi-manual or the fully manually annotated datasets, we evaluated 100 segmented cells selected randomly from the test dataset. These StarDist OPP segmentation predictions, together with the raw image, were presented to three human experts which had to decide between three options: (1) no human intervention is required, (2) human intervention might improve the result, and (3) the prediction requires human intervention. We introduced option (2) because for some cases of seemingly wrong predictions the annotators were unsure how to correct the prediction due to ambiguity in the raw imaging volume. On average, less than 20% of predictions required correction and the majority (~60%) of predictions required no intervention (Figure 5b), which is in agreement with the IoU-based average precision evaluation (Figure 4a). Furthermore, for our dataset, we observed that human annotators perceive segmentation predictions as qualitatively good (option (1)) that on average have an IoU of approximately 0.8 with the underlying semi-manual ground truth (Figure S9a, see Figure S7a for such cases). This is in agreement with the steep fall-off at IoU thresholds above 0.8 in the average precision curve for synthetic microscope data (Figure S8c). Both observations suggest that for our dataset, IoU values above 0.9 cannot be reached and are not required to be reached. This suggestion diminishes the significance of the steep fall-off of the average precision curve at around 0.8 (Figure 4a). By closer visual inspection of predictions with low IoU, we found that—despite the intensive human proofreading efforts—the semi-manually annotated labels still comprise some segmentation errors. In these cases, StarDist OPP often generates better segmentations than the ground truth (Figure S7b,c).

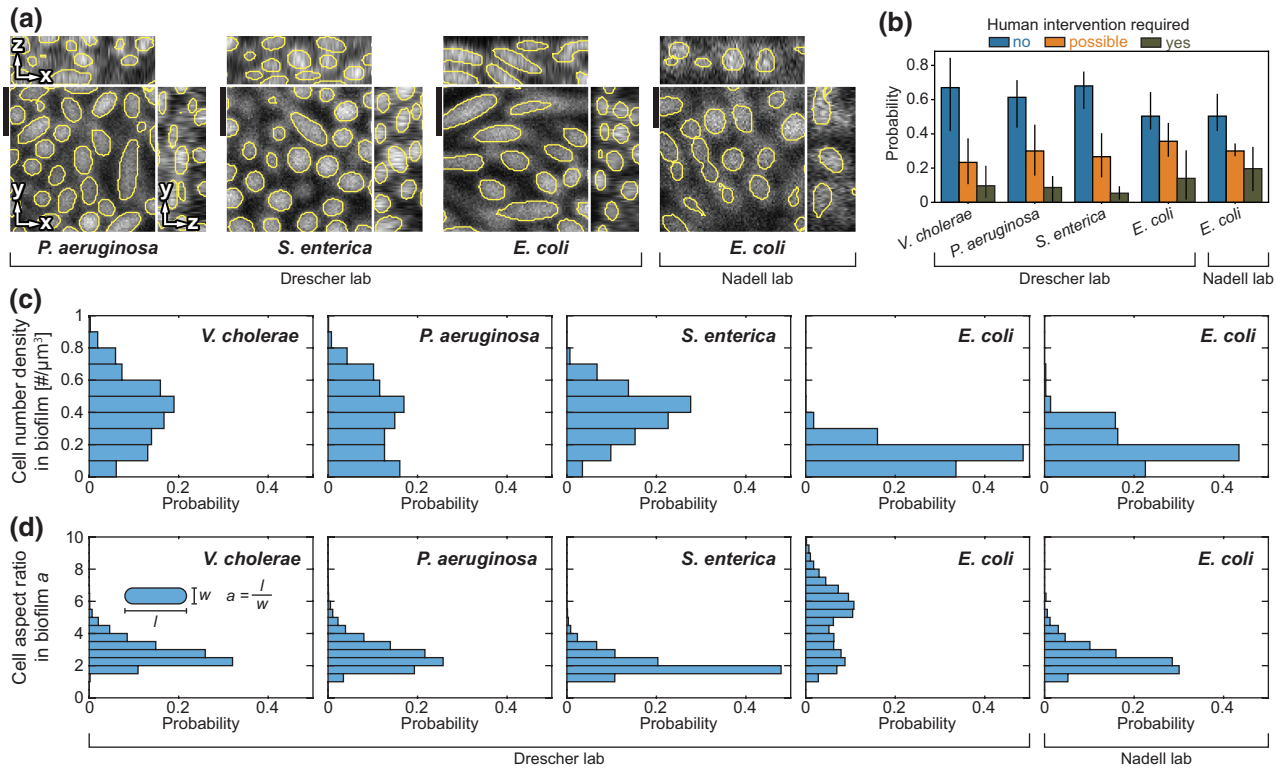

FIGURE 5 StarDist OPP segmentations on biofilms of different bacterial species. (a) Using our StarDist OPP model trained on *Vibrio cholerae* data, we generated segmentation labels (yellow lines) for cells in biofilms of four bacterial species from two different laboratories: From Knut Drescher's lab, images of *E. coli*, *P. aeruginosa*, and *S. enterica* are shown, and from Carey Nadell's lab, images of *E. coli* are shown. Each species is shown in a cropped 3D region with *xy*-, *xz*-, and *yz*-slices. The scale bar is 2 µm. (b) Validation results of segmented labels for biofilms of different bacterial species: 100 predicted cells are selected randomly from each species and are sorted manually by 3 human experts into 3 categories based on the required type of human intervention. For every segmented cell, the human annotators answered the question "Is human intervention required?" with one of 3 options "no", "possible", or "yes". Error bars are the standard deviation across 3 human experts. (c) Distribution of cell number density extracted from one biofilm of each species. (d) Distribution of cell aspect ratios extracted from one biofilm of each species.

2.4 | Identification of minimally required number of 3D annotated cells for training an accurate convolutional neural network

Given that our iterative semi-manual annotation scheme has provided more than 18,000 cell annotations in 3D, we sought to determine how many annotated cells are actually required for the training process of neural network approaches to yield accurate segmentation results in densely packed 3D microbial communities.

For this purpose, we trained replicate models ($n = 5$) of StarDist, StarDist OPP, and Cellpose networks on subsets of our training data with different amounts of annotated cells. We evaluated all networks against the same dataset that was not part of the training dataset and measured the average precision of the segmentation result (Figure 4h). Models trained on 100 annotated cells have an average precision (IoU threshold = 0.5) of approximately 0.5 for both StarDist and StarDist OPP. Increasing the number of annotated cells in the training data increases the average precision for both models, yet for more than 3000 annotated cells the average precision saturates, with a better performance by StarDist OPP (Figure 4h).

2.5 | StarDist OPP models trained on *Vibrio cholerae* dataset generate accurate single-cell segmentations for biofilms of other species and other imaging conditions

To test the transferability of StarDist OPP, we applied StarDist OPP models that were trained on our *V. cholerae* dataset to biofilm images of other bacterial species (*Escherichia coli*, *Pseudomonas aeruginosa*, and *Salmonella enterica*) from our lab, and to biofilm images (*E. coli*) from another research group that used a different confocal microscopy technique. Although we trained the StarDist OPP model with *V. cholerae* data, the segmentation predictions for different bacterial species worked remarkably well (Figure 5a). Due to lack of appropriate ground truth labels for the different bacterial species, we quantified the prediction quality with the same visual inspection approach by three human experts, as described above (Figure 5b). According to this validation procedure, segmentation performance on *P. aeruginosa* and *S. enterica* cells was comparable with the performance on *V. cholerae* cells. The segmentation performance on *E. coli* biofilms was similar for biofilms imaged in two different labs (Figure 5b), yet the segmentation performance on *E. coli* biofilms was slightly worse

compared with other species, likely due to the higher cell number density of the *E. coli* biofilms (Figure 5c). In summary, we conclude that StarDist OPP trained on our *V. cholerae* dataset can be transferred to single-cell segmentation tasks in biofilms of various species across research groups.

From the single-cell segmentation results for biofilms of different species, we extracted the single-cell level parameters "cell density" and "cell aspect ratio" (Figure 5c,d). These two parameters have been shown to account for the majority of differences in biofilm architecture between different species (Jeckel et al., 2022).

2.6 | Accurate 3D single-cell segmentation enables cell lineage tracking and spatiotemporal cell growth rate measurements in biofilms

Based on the accurate segmentation results provided by StarDist OPP, we tested whether the tracking accuracy is sufficient for cell lineage tracking. For this, we acquired 3D images of *V. cholerae* biofilm growth every 10 min for 19 h, starting from a cluster of just a few surface-attached cells. We segmented all cells in the 3D image of each time frame. Using BiofilmQ (Hartmann et al., 2021), we identified all cellular 3D centroid coordinates and transferred them to the TrackMate software (Ershov et al., 2022; Tinevez et al., 2017) to predict cell trajectories. When performing the tracking with segmentation results from StarDist OPP there are less tracking errors compared with the segmentation results from StarDist (Figure S10). StarDist OPP segmentations enabled us to obtain long cell tracks (Figure S10), and the cell trajectories during biofilm growth point radially outwards from the core of the biofilm (Figure 6a,b), which is consistent with results from lineage tracking in biofilms without single-cell resolution (Qin et al., 2020).

From the cell trajectories, we calculated the cell division rates, which we mapped to all spatiotemporal locations during biofilm development using BiofilmQ (Figure 6c, d). Figure 6c depicts the cell division rate at different times during biofilm development at different spatial positions in the biofilm. At the very early stage of biofilm development ($N_{cell} < 10$) the cell division rates are below 0.75 h^{-1}. For $N_{cell} > 10$, the majority of the biofilm remains at a division rate below 0.75 h^{-1}, yet the cells in the outer periphery of the biofilm (where the fresh medium flows by) have a division rate of up to 1.5 h^{-1}. For biofilms with $N_{cell} > 1000$, the cell division rate in the core of the biofilm approaches 0 h^{-1} (Figure 6c). The spatial profile of the cell division rate is radially symmetric, and cells with a similar distance from the surface have a similar division rate (Figure 6d). These spatiotemporal cell division rates match the expectation of nutrient- or oxygen-deprived slow-growing cells in the biofilm core, and faster-growing cells in the nutrient-rich biofilm periphery (Díaz-Pascual et al., 2021; Stewart & Franklin, 2008).

3 | DISCUSSION

Recent improvements in bacterial image cytometry tools have greatly advanced the analysis of single cells in microbiology research (Ducret et al., 2016; Hartmann et al., 2020, 2021; Mountcastle et al., 2021; Paintdakhi et al., 2016; Stirling et al., 2021), yet the key step in image cytometry for biofilms remains a challenge: the accurate segmentation of cells in 3D images of biofilms. In the analysis of time-lapse images of mother machines and other 2D systems, the most recent advances in segmentation relied on deep learning approaches (Hardo et al., 2022; Lugagne et al., 2020; O'Connor et al., 2022). These new methods enable the investigation of single-cell behavior over time, ultimately shedding light on the physiological heterogeneity in bacterial populations (Bakshi et al., 2021). As it is a longstanding paradigm that biofilms harbor bacterial cells of diverse physiological states, such analysis tools are required in biofilm research. However, a major barrier in applying new deep learning segmentation methods to 3D biofilm images has been the lack of sufficient training datasets, as the annotation of volumetric images is time-consuming and difficult to endure. With this study, we provide an extensive dataset of 18,868 cells annotated in 3D *V. cholerae* biofilm images to overcome this barrier.

The iterative semi-manual annotation scheme we presented in Figure 2 resulted in an approximately 60× faster annotation than fully-manual annotation (from 5 ± 2 min per cell for manual annotation to approximately 5 s per cell for semi-manual annotation). The semi-manual annotations differ from fully-manual annotations by resulting in slightly larger cells. Furthermore, visual inspection suggests that our semi-manual annotation scheme generates fewer errors than the fully-manual annotation. Yet despite our efforts to ensure a very high quality of the 3D annotation dataset we generated, there are still limitations to this dataset (Figure S7c). As our randomized visual inspection showed, opinions of expert human annotators differ strongly (Figure S9b), likely due to the optical resolution limitations and signal-to-background limitations, as described in the methods section on manual segmentation. These differences in the opinion of expert human annotators could lead to ambiguities during the manual curation step. To significantly improve the training dataset, these ambiguities have to be resolved by enhancing the information present in the microscope images, for example, by increasing the optical resolution, by additionally staining the cell membrane, or by random expression of two or more fluorescent proteins such that neighboring cells can be distinguished by their differing fluorescence properties.

Performing CNN-based segmentation of single cells in biofilms requires a suitable large amount of training data. Our data show that the segmentation performance of the trained CNNs did not substantially increase for more than 3000 annotated cells. This number is comparable with the 1500–3000 cells used per training run by Zhang et al., 2020. Systematic investigations of minimal amounts of training data for 2D systems report 10 to 1000 minimally required annotated cells and suggest that fewer cells are required for transfer learning (Falk et al., 2019; Van Valen et al., 2016). Unfortunately, the required number of annotated cells for 3D biofilms is significantly higher than for 2D systems. However, our evaluation of the performance of the *V. cholerae*-trained StarDist OPP model on biofilms of different species and from another lab suggests that our trained

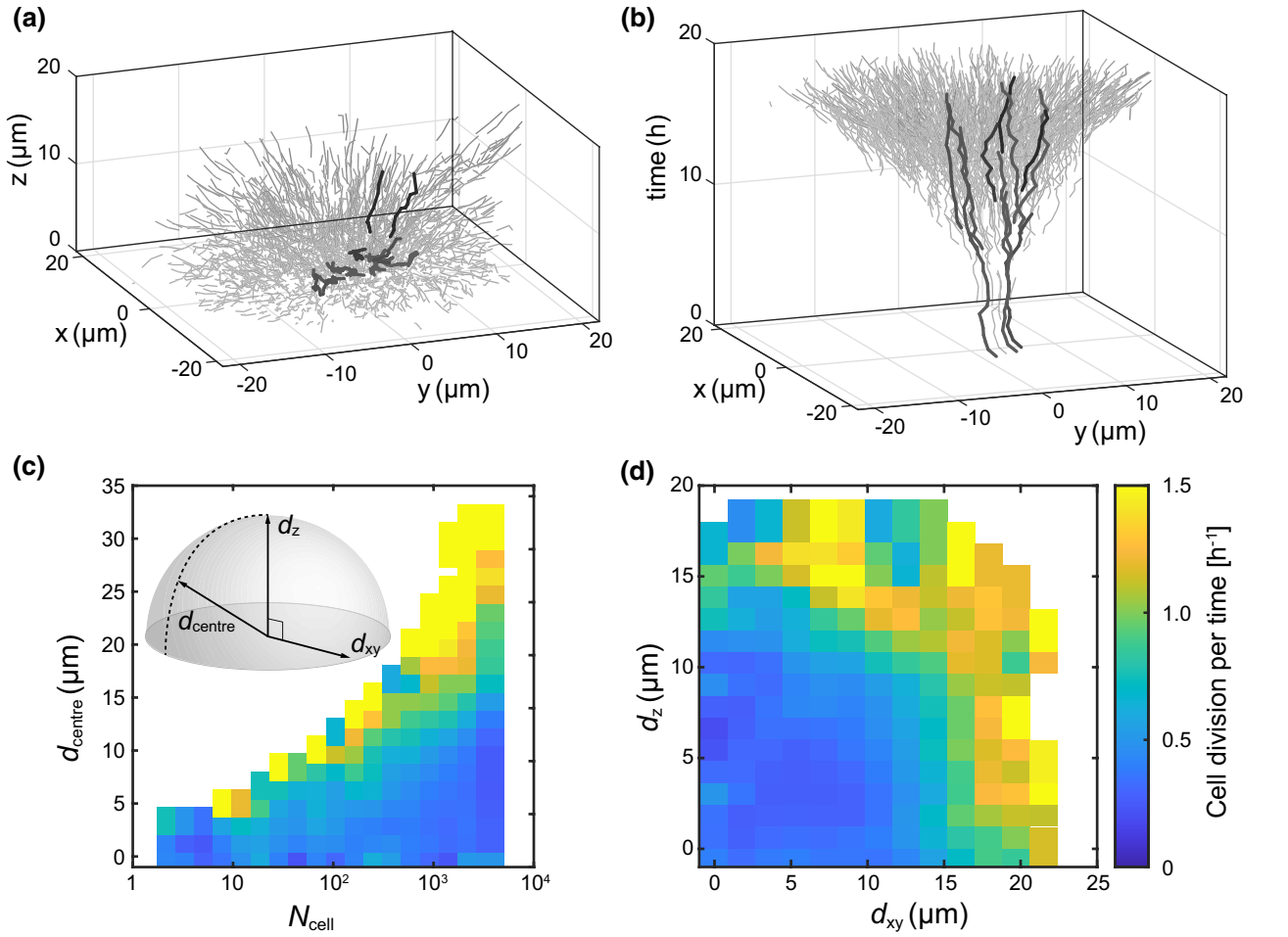

FIGURE 6 Based on accurate single-cell segmentation in biofilms, all individual cells can be tracked and cellular growth rates can be measured. (a, b) Tracked trajectories of cells during *Vibrio cholerae* biofilm formation are shown in light gray. A few example trajectories are indicated by dark gray and black lines. Panel (a) shows the trajectories in *xyz*-space. Panel (b) shows the trajectories in the *xy*-plane and time. (c) Spatiotemporal cell division rate during biofilm formation. The number of cells in the biofilm (N_{cell}) indicates the time of development. The *y*-axis shows the distance between the cell center and the center of the biofilm at the bottom of the biofilm (in the $z=0$ plane). Inset: Schematic figure illustrates the coordinates used in panels (c) and (d). (d) Spatial distribution of cell division events in a *Vibrio cholerae* biofilm that grew from 2074 cells to 5190 cells during the measurement interval. The horizontal axis shows the distance between the *xy*-projected cell centroid and the bottom center of the biofilm. The vertical axis shows the *z*-position of the cell centroid.

model can readily be applied by other research groups. In case the performance of the pre-trained StarDist OPP is not sufficient, our trained models should be improved iteratively by semi-manually curating imperfect predictions.

While some deep learning methods rely on classical algorithms, for example, connected components and watershed, to separate individual touching cells in an instance segmentation task (Zhang et al., 2020, 2022), deep learning methods also allow developing of new instance segmentation methods that have an intrinsic representation for individual objects (Cutler et al., 2022; Schmidt et al., 2018; Stringer et al., 2021; Weigert et al., 2020). In these methods, boundaries between two cells are not enhanced for better post-processing; rather the methods directly predict two individual cells. In this work, we used StarDist as it uses an intrinsic representation of individual 3D cells (Weigert et al., 2020), which is beneficial in densely populated microbial communities. StarDist uses a 3D U-Net backend (Çiçek et al., 2016) to predict object probabilities as well as distance maps (Figure 3a), which together contain the information to calculate single object segmentations with a sophisticated post-processing approach. StarDist was a major improvement over the previously available 3D U-Nets, which relied on accurate cell–cell boundary predictions for correct object separations.

As the standard post-processing of StarDist generated systematic over-segmentation errors on our dataset, we developed an optimized post-processing algorithm for bacterial cell shapes that iteratively fuses together proposed objects, which we termed StarDist OPP. One advantage of the StarDist CNN, which is used in StarDist OPP, is that its outputs enable the automatic identification of areas of ambiguous segmentation results, by searching for overlapping objects. Visual inspection of these areas and the results on synthetic microscope images (Figure S8b, SI Text S1) suggest that ambiguous and wrong segmentation results stem either from flawed training data or from

insufficient information in the microscope images to determine the exact boundaries of cells—these are also regions where expert human annotators cannot clearly decide where boundaries between two objects should be drawn. The training data contains similar errors due to the same limitations, as discussed above.

To circumvent the time-consuming annotation process a recent study trained their deep learning models with synthetic training data and achieved an average precision of ~50% at IoU = 0.5 on experimental images (Zhang et al., 2020). With the StarDist OPP, an average precision of ~80% at IoU = 0.5 is achieved for experimental images. Interestingly, M. Zhang et al., 2020 observed that their network performed less well on experimental test data compared with synthetic test data. This suggests that generating synthetic images that accurately mimic experimental data might require as much time as is needed for fine-tuning a classical segmentation algorithm, or for generating a manual or semi-manual annotation. Furthermore, even for an approach with synthetic training data, some manual annotation of experimental data is required to validate the segmentation. However, image segmentation efforts in other fields suggest that training on a combination of synthetic and experimental data could decrease the time spent on manual annotations while maintaining high segmentation performance (Cuéllar et al., 2022; Poucin et al., 2021).

In summary, we have presented an efficient strategy for obtaining the key ingredient for the application of CNNs to segment 3D bacterial biofilm images: large training datasets of 3D single-cell annotations. The 18,868 annotated 3D cells we generated are now available as a community resource for training neural networks. From this dataset, we identified the minimal number of cell annotations that are needed to obtain accurate segmentation results. Using our large training dataset, we also evaluated the performance of different methods for single-cell segmentation in biofilms, and we extended the StarDist CNN by optimizing the post-processing step for bacterial cell shapes, resulting in StarDist OPP, which generated the most accurate segmentation results across all algorithms. StarDist OPP models trained on our training dataset performed well on biofilms of different species across research groups. Finally, we demonstrated that the segmentation results provided by StarDist OPP are sufficient for accurate cell lineage tracking and single-cell growth rate measurements. The availability of our large training dataset, the StarDist OPP code and pre-trained models, in conjunction with easy-to-use CNN training and deployment tools like ZeroCostDL4Mic (von Chamier et al., 2021), will enable a more broad adoption of CNN-based segmentation in biofilm research.

4 | EXPERIMENTAL PROCEDURES

4.1 | Bacterial cultivation and biofilm growth

The strain used in this study was derived from the *V. cholerae* N16961 wild type, by introducing the *vpvC*^{W240R} allele conferring rugosity, and by introducing the Δ*crvA* (VCA1075) mutation that abolishes the typical comma-shape of *V. cholerae* to confer a straight cell shape. The strain also harbors a plasmid (pNUT542) carrying a gentamicin resistance and a P_{tac}-*sfgfp* expression system without *lacO*, resulting in the constitutive production of sfGFP. The resulting strain, which was used for all experiments in this study, is termed KDV613 (Hartmann et al., 2019). Biofilm experiments with *V. cholerae* were performed in M9 minimal medium, supplemented with 2 mM $MgSO_4$, 100 mM $CaCl_2$, MEM vitamins (Sigma), 0.5% (w/v) glucose and 15 mM triethanolamine (pH 7.1). Gentamicin (30 μg/mL) was also added to LB and M9 media, to maintain the plasmid pNUT542.

The *E. coli* strain used in this study is derived from the AR3110 wild type (Serra et al., 2013) with a point mutation that upregulates biofilm formation (Grantcharova et al., 2010). For imaging on a spinning disk confocal microscope, a derivative of this strain harboring a plasmid that leads to the P_{tac}-controlled production of sfGFP was used. For imaging on a point scanning confocal microscope, the AR3110 wild type with a chromosomal P_{tac}-controlled expression of mKate2 was used. The *S. enterica* strain used in this study is derived from the UMR1 wild type (Römling, Bian, et al., 1998) and also carries a point mutation that upregulates biofilm formation (Römling, Sierralta, et al., 1998), and also harbors a plasmid that leads to the P_{tac} controlled production of sfGFP. The *P. aeruginosa* strain used in this study (KDP63) is a derivative of the PAO1 wild type, harboring a plasmid producing the pX2 promoter-controlled production of the YPet fluorescent protein.

Biofilms of *V. cholerae*, *E. coli*, *P. aeruginosa*, and *S. enterica* were grown in microfluidic flow chambers (chamber dimensions: [width; height; length] = [500; 70; 7000] μm). Flow chambers were constructed from poly(dimethylsiloxane) bonded to glass coverslips using oxygen plasma. The microfluidic design included four independent channels on each coverslip. The manufacturing process for these microfluidic channels guarantees highly reproducible channel dimensions and surface properties.

For biofilm growth of *E. coli*, *S. enterica*, and *P. aeruginosa*, the culture conditions, media, and flow rates described in a previous study (Jeckel et al., 2022) were used. For biofilm growth of *V. cholerae*, each channel was inoculated with a culture of the *V. cholerae* strain, which was prepared as follows. Cultures were grown overnight at 28°C in liquid LB medium under shaking conditions, back-diluted 1:200 in LB medium in the morning, and grown to an optical density at 600 nm of 0.5. This culture was used for inoculating the flow channels. After inoculation of the channels, the cells were given 1 h to attach to the glass surface of the channel without flow, followed by the application of a flow of 100 μL min^{-1} for 45 s, with M9 medium, to wash away non-adherent cells and to remove LB growth medium from the channels. The flow rate was then set to 0.5 μL min^{-1}, to continuously supply fresh M9 medium into the channel. Flow rates were controlled using a high-precision syringe pump (Pico Plus, Harvard Apparatus).

4.2 | Microscopy and image acquisition

Most 3D biofilms used in this study (except the *E. coli* biofilm images obtained in Carey Nadell's lab, see description below) were

imaged using spinning disk confocal microscopy with a 50 μm pinhole disk (Yokogawa CSU) mounted on a Nikon Ti-E inverted microscope, using an Olympus 100× silicone oil immersion (refractive index = 1.406) objective with NA 1.35. By using this silicon oil objective, distortions at *z*-positions >10 μm into the biofilm are minimized. Fluorescence was excited using a 488 nm laser (Coherent OBIS), and the emission was imaged using an Andor iXon Ultra electron-multiplying charge-coupled device (EMCCD) camera with 90 ms exposure time per *z*-slice. The measured resolution of the optical system is *xy*-resolution of ~235 nm and *z*-resolution of ~710 nm (Yordanov et al., 2021). The theoretical limit of the resolution of our optical system was calculated as *xy*-resolution ~220.14 nm and *z*-resolution ~625 nm, using the Nikon Resolution Calculator (https://www.microscope.healthcare.nikon.com/microtools/resolution-calculator/). As is common when imaging into a thick specimen, the fluorescence intensity decreases with increasing depth into the specimen (Figure S4a). In our case, this is caused by scattering of the excitation light as it penetrates the biofilm, and further by scattering of the emission light as it exits the biofilm towards the objective. The Olympus 100× objective resulted in an image on the camera with 63.2 nm per *xy*-pixel, and we used 100 nm spacing in the axial *z*-direction for all experiments, except for experiments in which the cell lineages were tracked. In cell lineage tracking experiments we used a 400 nm spacing in the *z*-direction and a 30 ms exposure time per *z*-slice, to reduce photodamage and to enable biofilm growth that is unperturbed by excitation light exposure. Very-low-excitation light intensities were used for all experiments to reduce photodamage as much as possible. Focus drift was prevented using a Nikon hardware autofocus system. The hardware was controlled using μManager (Edelstein et al., 2014), and a feedback between image acquisition and live image analysis ensured that *z*-stacks were only imaged up to the *z*-height at which the biofilm ends, to further reduce light exposure of the biofilm. For the cell lineage tracking experiments, each biofilm was imaged every 10 min.

For the *E. coli* biofilm images obtained in Carey Nadell's lab, a Zeiss LSM 980 point-scanning confocal with a 40×/1.2NA water immersion objective was used. Pixel scaling was set to 52 × 52 nm and a *z*-spacing of 200 nm was used. A 594 nm excitation line was used to excite the mKate2 fluorescent protein. The pixel dwell time was 1.31 μs, and 4× line averaging was used. The spectral window for fluorescence emission collection was 604–682 nm and the detector gain was set to 600 V.

4.3 | Training data generation

Initially, a deconvolved volume of 716 × 476 × 100 voxels was segmented using a classical instance segmentation approach we termed "Improved Hartmann et al." This technique was based on a segmentation procedure introduced for single-cell segmentation in *V. cholerae* biofilms (Hartmann et al., 2019), which employed filters to reduce noise and to enhance the signal-to-background ratio, edge detection for semantic segmentation, and watershed for instance segmentation, followed by a few post-processing steps to remove detected objects that are not cells. This "Hartmann et al." procedure was improved recently (Jeckel et al., 2022) to result in what we term "Improved Hartmann et al." This resulted in a reduced oversegmentation by the watershed segmentation (Figure 4a).

We inspected the segmentation output from the Improved Hartmann et al. method for the 716 × 476 × 100 voxels for segmentation errors, and we manually corrected each visible error in 3D with napari (Sofroniew et al., 2022). This step took approximately one week for a skilled and trained human annotator. The volume was then split into training, validation, and testing sub-volumes and a first StarDist network was trained on the available training data.

This first trained StarDist network was then used to predict instance segmentations for additional microscope image data. During the manual inspection of the newly predicted segmentation volume, we realized that the segmentation errors of the StarDist network coincide with overlapping cell predictions. To facilitate the review process, we used the StarDist overlap labeling feature to mark all overlaps and created a graphical user interface in Matlab (Figure S1c), which iterates over all marked overlap positions. Instead of manually correcting the positions one-by-one, we manually classified each overlap into one of three categories (Figure S1d): Overlapping objects are separate cells ("split"), overlapping objects are parts of the same cell ("merge"), or the overlapping objects require manual correction ("manual"). As most of the objects were part of the first two classes ("split" or "merge"), the use of this graphical user interface significantly improved our turnover time for the manual correction of the segmentation result to approximately one day for a single biofilm, by a skilled and trained human. After the semi-manual correction of the segmentation prediction, we incorporated the image data into our training process and trained a new generation of StarDist model from scratch. This approach of iterative StarDist training and manual correction of the segmentation output was iteratively applied to obtain a manually corrected dataset of single-cell segmentations for five biofilms (Figure S2). One of these biofilm images was used only for testing purposes and never for training, whereas the other four biofilm images were used for CNN training (Figure S2).

After loading the 16-bit images, we converted the 16-bit integer values to double-precision numbers and rescaled their intensities such that the 1%-percentile was at 0.0 and the 99.8%-percentile was at 1.0. To augment the available training data, we generated additional online flips along the *x* and *y* image axes (probability = 0.5 each) and intensity transformations with factors drawn from a uniform distribution between [0.6, 2) and an additional intensity offset uniformly drawn from [−0.2, 0.2).

The visual perception of cell boundaries by the human annotator was facilitated by applying a 3D image deconvolution. For deconvolution, we used the software Huygens (Scientific Volume Imaging). The raw image and the deconvolved image are compared in Figure S4. While the cell signals in the raw image decrease for higher *z*-slices, the cell signals in the deconvolved image are relatively homogeneous (Figure S4a). The cell boundaries in the deconvolved image have a

better contrast than the ones in the raw image. Particularly in higher *z*-slices, cell signals that are visible in the raw image cannot be identified in the deconvolved image (Figure S4b,c). For the CNN training, however, we noticed that deconvolved input images degrade the segmentation performance. Therefore, deconvolved images were not used for training of the CNN, and all CNN trainings reported here were performed with raw microscopy images as input data.

4.4 | Artifacts and limitations of the semi-manually generated training dataset

When generating the training dataset, we made two deliberate choices that lead to the artifact of seemingly missing annotations for some regions with visible fluorescence signals. First, fluorescence signals that are visible only in one *z*-slice were not annotated. These signals stem from cells that are not attached to the biofilm and float freely in the microfluidic channel. Second, we chose that annotations should resemble the deconvolved cell shapes which are tighter than the visible fluorescence signal. Both choices explain seemingly missing annotations of visible fluorescence signals in 2D slices of our 3D images. If the fluorescence signal is not present in the previous and next *z*-slice, the signal stems from a floating cell. If the fluorescence signal is present also in a neighboring *z*-slice, there is also an annotation in this *z*-slice that is just drawn tighter around the fluorescence signal (Figure S4). We encountered the same phenomenon of recessed cell outlines also in our synthetic microscope images (Figure S8b).

While errors are present in our training dataset (Figures S3 and S7), during our visual inspections of the dataset, we did not encounter a generic class of segmentation errors other than over-segmentations and occasional under-segmentations. Therefore, we were not able to systematically investigate and correct other kinds of errors. The alternative, visual inspection and manual correction of each annotation would have been detrimental to our goal to speed-up generation of a training dataset with a semi-manual approach. Using visual inspection by three expert humans, we concluded that the majority of predictions based on the training dataset do not require human intervention, whereas for ~20% of predictions human annotators have problems to correct predictions (Figure 5b). Although we cannot rule out the possibility that the errors still contained in the dataset pose a significant limitation for our method, for our use cases, the method is sufficient.

We attribute the occasional under-segmentation errors to signal-to-background and optical resolution limitations. We did not systematically search for and correct under-segmentation errors as often correction of these errors was undecidable for human experts due to insufficient information in the raw image.

4.5 | Manual segmentation of cells in 3D biofilms

Since the training data generation depended on a StarDist prediction, we needed an additional image dataset with a completely manual annotation without a computational segmentation prediction. The complete manual annotation was performed using the napari interface (Sofroniew et al., 2022). The 3D volume that was annotated was a densely packed region of the raw image of "Biofilm 1" from Figure S2, with a volume of 257 ×257 ×129 voxels. The manual annotation of this region resulted in 1644 cells.

4.6 | Challenges in manual annotation and visual inspection of segmentations in 3D biofilms

During our efforts to visually inspect, manually annotate, or manually correct 3D single-cell segmentations in our imaging data we encountered several challenges that we describe below.

As the optical resolution in the *z*-direction is higher than the optical resolution in the *xy*-direction, the appearance of cell outlines compared with the underlying raw fluorescence signal differs depending on whether the *xy*-plane or the *xz*- or *yz*-planes are visualized. This feature of differing resolutions makes it more difficult to identify cell outlines in the *z*-direction compared with the *xy*-direction. This is especially true for horizontally aligned cells lying above each other. Here, the border between such cells can sometimes be inferred by the fact that a horizontally aligned object spanning more than five *z*-slices must be under-segmented as otherwise it would be too wide. Alternatively, it is sometimes possible to identify under-segmentation for horizontally aligned cells above each other by carefully observing the cell outlines in the *xy*-direction for several *z*-slices, and finding a *z*-slice where the cell extents are minimal in comparison with the extents in the previous and next *z*-slice.

The fall-off in fluorescence intensity with increasing *z*-depth into the biofilm results in lower signal-to-background ratio in higher *z*-slices (Figure S4a), which exacerbates the identification of individual cells in two distinct ways. First, separation of closely neighboring cells in higher *z*-slices gets more complicated. Second, cells in high *z*-slices with dim fluorescence might not be identified at all. It should be mentioned that this fall-off in fluorescence might influence the size of cell annotations. Although we observe that the decrease in mean cell brightness for higher *z*-depth in the biofilm correlates with a decrease in cell volumes (Figure S11), we cannot rule out a biological explanation for decreasing cell volumes.

The manual identification of object boundaries of cells with sizes close to the optical resolution limit is often difficult for annotators to endure for extended periods, as a decision has to be made based on insufficient information from the imaging data. The information in the imaging data is insufficient for this task, especially in areas deep into the tissue where cells are densely packed and where the signal-to-background ratio is diminished due to scattered light. From our manual accuracy quantification, we estimate that approximately 20% of cells fall into the category where human annotators cannot decide on how to improve an annotation (Figure 5b). Furthermore, there is biological uncertainty during cell division when it is visually difficult to determine with our cytoplasmic fluorescence signal whether the division septum is closed.

4.7 | Generation of synthetic training data

To generate synthetic training data, the cell-shaped objects were created by brute-force random placement: We started with an empty cube with an edge size of 256 pixels and a pixel size of 50 nm. Curved objects where generated by randomly sampling tori with a radius in the range of 24 to 36 pixels, a width in the range of 12 to 20 pixels, and an arc length in the range of 32 to 64 pixels. The ends of the torus were covered with semi-spheres with a radius according to the torus width. A total of 3500 objects were generated, randomly orientated, and placed inside the cube such that they do not overlap with previously sampled objects. If no placement could be found for a single object after 30 iterations of random orientation and placement, the object was discarded. After generation, only every second slice in the *z*-direction was used, resulting in a final pixel size of 100 nm in the *z*-direction and 50 nm in the *xy*-direction (Figure S8a).

We visually distorted the synthetic label data to mimic experimental microscope data (Figure S8b). The biological fluctuations in object signal intensity and gradient where enforced by randomly sampling the intensity from a Gaussian distribution (mean $\mu = 3000$, standard deviation $\sigma = 500$) and application of a Euclidian distance transformation. Then, the resulting volume was convolved with a synthetically created point-spread-function. Poisson noise (rate parameter $\lambda = 40$) was added to imitate the shot noise in a photon-counting detector, and Gaussian noise ($\mu = 500$, $\sigma = 10$) was added to account for intra-specimen scattering.

The synthetic point-spread-function (PSF) was created with the Fiji PSF Generator (version 1.1.1.2) (Griffa et al., 2010; Kirshner et al., 2013) using the Richards & Wolf 3D optical model with refractive index 1.405 in the immersion medium and accuracy computation "best" for a wavelength of 525 nm. An objective NA of 1.35 was used and the PSF was saved as a 32-bit image of $65 \times 256 \times 256$ pixels (ZYX) and a pixel size of 50 nm in the *xy*-direction and 100 nm in the *z*-direction.

4.8 | Training convolutional neural network models for cell segmentation

The 3D images in the available training dataset were divided into smaller sub-volumes ($64 \times 64 \times 128$ voxels) for the neural network training. The set of available sub-volumes was then divided into sub-volumes used for training, validation, and testing. We tested four CNN architectures for the final cell segmentation, StarDist (Weigert et al., 2020), Cellpose (Stringer et al., 2021), a multi-class U-Net (Zhang et al., 2020), and BCM3D 2.0 (Zhang et al., 2022).

The StarDist segmentation accuracy depends on the ray number to approximate the shape of the star-convex objects. By modeling all labels in the validation data as star-convex objects and comparing the model accuracy with respect to the input labels, we observed that 192 rays are sufficient to achieve a mean IoU value of 0.8 for the used training data in the anisotropic training data regime (Figure S12a). All StarDist models used in this study, therefore, used 192 rays (except for Figure S8, where 96 rays were used). The remaining StarDist parameters were based on default values: The U-Net backbone used two resolution levels. At each level, 2 convolutions with a $3 \times 3 \times 3$ voxel kernel size were used in the encoder and decoder, respectively. The filter size in the first convolution was set to 32. The filter size was doubled after each max-pooling step with a $2 \times 2 \times 2$ voxel field of view and correspondingly halved after each upsampling layer. The extra convolution layer in the network head contained 128 filters. The mini-batches contained 2 patches with $96 \times 96 \times 48$ voxels in *x*, *y*, *z*. The sampling algorithm enforced that 90% of all patches contain foreground voxels. The Adam optimizer was initialized with default parameters and a learning rate of 0.0003. A learning rate scheduler halved the learning rate if the loss did not change for 40 consecutive epochs. The mean absolute error was used for the distance loss and weighted in a 0.2 to 1 ratio with respect to the probability loss. The background regularization in the distance loss was set to 0.0001.

CellPose relies on a 2D training regime so that the sub-volumes ($128 \times 128 \times 64$ voxels) were sliced into 2D data along the *x*,*y*, and *z* axis respectively. As a result, the training time increased such that a full-scale hyperparameter scan on a single GPU was extremely time-consuming. Thus, an implementation with a data-parallelized training regime was implemented based on the MXNet support of Horovod (Chen et al., 2015; Sergeev & Del Balso, 2018). The training was performed on 8 nodes with 2 NVIDIA Quadro RTX 5000 each. After the hyperparameter optimization, the best parameter (learning rate = 0.00625, weight decay = 10^{-4}, momentum = 0.5) of a subset of the training data (5%) was verified in a two-week-long iterative training session on the CellPose reference implementation on a single GPU (Figure S12b). Since the single-GPU and multi-GPU implementations of CellPose training were identical with respect to the segmentation accuracy, we trained a multi-GPU implementation of CellPose with the determined hyperparameters on the complete training dataset and used it for our comparisons.

For the multi-class U-Net, we followed Zhang et al., 2020: The ground truth label images were transformed into semantic images where each pixel belongs to one of three classes: cell interior, cell boundary, or background. The CNN was trained to predict the correct class for each pixel. To predict single instances of cells based on the multi-class U-Net, first, a binary image was generated by thresholding on the predicted cell interior class. Then, by detecting connected components in the binary image, a label image was generated (Zhang et al., 2020).

For BCM3D 2.0, we implemented the code following the original paper (Zhang et al., 2022). In brief, two CARE networks (Weigert et al., 2018) were trained to predict the Euclidean distance transform of each cell and to predict boundaries between cells. The predicted Euclidean distance transform was used in the first step to generate a label map by thresholding and detecting connected components. In a second step, too big or too convex objects were detected and separated into smaller cells based on a watershed algorithm that used the information from the Euclidean distance transform and the predicted boundaries between cells.

4.9 | Quantification of segmentation accuracy without ground truth

Images of *E. coli*, *P. aeruginosa*, and *S. enterica* were acquired with a step size in the *z*-direction of 400 nm. To segment these volumes with our trained models, the volumes were interpolated with a cubic spline to match the 100 nm step size in the *z*-direction of the *V. cholerae* dataset. Images of *E. coli* biofilms acquired in Carey Nadell's lab with a point scanning confocal microscope were acquired with a step size in the *z*-direction of 200 nm. These images were also interpolated to a *z*-spacing of 100 nm with a cubic spline.

For each species, a biofilm image was predicted with our StarDist OPP model. From each of the resulting predictions, 100 predicted objects (i.e., cells) were randomly picked for evaluation by three human experts. Each annotator classified the same 100 predicted objects with a custom napari plugin (link to repository given in Data and Code Availability section) based on the question "Is human intervention required" into the categories "no," "possible," and "yes." The annotators knew the origin of the current labels with respect to species and lab. Each predicted object was presented with a context of the surrounding region extended by 18 pixels in all directions, the predicted object was highlighted, and all other objects in the field of view were visualized.

4.10 | Cell lineage tracking during biofilm development

For single-cell segmentation of 3D images in the biofilm growth time series, we employed a CNN-based content-aware image restoration step (Weigert et al., 2018) to boost the signal-to-noise ratio of the raw images. The image restoration network was used to boost the signal on 9 newly semi-manually annotated image/label pairs, which were then used to train a new StarDist network. At each time point, the 3D biofilm image was then segmented with the newly trained StarDist OPP. The training data was generated with 100 nm spacing between the stack levels to facilitate the semi-manual annotation, but the time series was captured with 400 nm spacing. Thus, a cubic interpolation was used to account for the missing layers. The cell centroid information was extracted from the volumetric segmentation and directly saved in XML format such that it can be used as a configuration for TrackMate 7.5.1 (Ershov et al., 2022; Tinevez et al., 2017) released as a Fiji 2.3.0/1.53f51 plugin (Schindelin et al., 2012). For the tracking process, the default settings for the LAP tracker (linking max distance = 0.948 µm with 15 pixels; track segment gap closing = 0.948 µm with 15 pixels and 5 frames; not allowing track segment splitting and merging) in TrackMate were used to follow the cells over time. In the tracking data, a new tracking ID was recognized as a cell division. The spatial cell division rate in Figure 6 was computed as the average of the number of cell divisions divided by the time between image frames (10 min). The tracks were analyzed with code based on *Numpy* (Harris et al., 2020) and *Pandas* (McKinney, 2010; The Pandas Development Team, 2023) and the tracking results were plotted with Matlab.

AUTHOR CONTRIBUTIONS

Eric Jelli: Writing—original draft; Writing—review & editing; Conceptualization; Investigation; Methodology; Validation; Visualization; Software; Data curation. **Takuya Ohmura**: Writing—original draft; Writing—review & editing; Investigation; Methodology; Validation; Visualization; Software; Data curation. **Niklas Netter**: Writing—original draft; Writing—review & editing; Investigation; Methodology; Validation; Visualization; Software; Data curation. **Martin Abt**: Data curation; Validation. **Eva Jiménez-Siebert**: Investigation; Methodology. **Konstantin Neuhaus**: Investigation; Methodology. **Daniel K. H. Rode**: Investigation. **Carey D. Nadell**: Investigation. **Knut Drescher**: Conceptualization; Investigation; Funding acquisition; Writing—original draft; Writing—review & editing; Supervision; Project administration.

ACKNOWLEDGMENTS

This work was funded by grants to K.D. from the European Union's Horizon 2020 research and innovation program through the European Research Council grant agreement No. 716734 and the Marie Skłodowska-Curie grant agreement No. 955910 (PHYMOT consortium), and through grants from the Deutsche Forschungsgemeinschaft (DR 982/5-1 and DR 982/6-1), Bundesministerium für Bildung und Forschung (TARGET-Biofilms), the Minna-James-Heineman-Stiftung, and the National Centre of Competence in Research AntiResist, funded by the Swiss National Science Foundation grant 51NF40_180541, and the Swiss National Science Foundation Consolidator Grant TMCG-3_213801. This work was also supported by a fellowship to T.O. from the Human Frontier Science Program (LT000013/2019-C). We thank the Max Planck Computing and Data Facility and the University of Basel computation facility sciCORE for GPU resources and support service. Open Access funding enabled and organized by Projekt DEAL.

CONFLICT OF INTEREST STATEMENT

The authors report no conflicts of interest.

DATA AVAILABILITY STATEMENT

The training dataset comprising raw images and 3D annotated cells in biofilms is available at Zenodo with DOI: 10.5281/zenodo.7704410. The same Zenodo record also contains raw data and segmentation predictions for biofilms of different species, the synthetic training data, and the trained StarDist OPP models. The code for the application, training, and evaluation of the used segmentation methods is available at https://github.com/erjel/iterative-biofilm-annotation. StarDist OPP is available as an independent Python package at https://github.com/gatoniel/merge-stardist-masks and StarDist OPP is also available as a napari plugin at https://github.com/gatoniel/napari-merge-stardist-masks. The napari plugin for the visual inspection and classification of segmentation performance is available at https://github.com/gatoniel/napari-validate-random-label-predictions.

ORCID

Eric Jelli https://orcid.org/0000-0001-6202-2449
Takuya Ohmura https://orcid.org/0000-0001-8022-2946

Niklas Netter https://orcid.org/0000-0001-5053-5554
Martin Abt https://orcid.org/0009-0001-6203-0702
Eva Jiménez-Siebert https://orcid.org/0000-0001-8168-0359
Konstantin Neuhaus https://orcid.org/0000-0002-9877-5765
Daniel K. H. Rode https://orcid.org/0000-0002-2950-9340
Carey D. Nadell https://orcid.org/0000-0003-1751-4895
Knut Drescher https://orcid.org/0000-0002-7340-2444

REFERENCES

Ackermann, M. (2015) A functional perspective on phenotypic heterogeneity in microorganisms. *Nature Reviews Microbiology*, *13*(8), 497–508. Available from: https://doi.org/10.1038/nrmicro3491

Ahrens, J., Geveci, B. & Law, C. (2005) *ParaView: an end-user tool for large data visualization*. Burlington, MA, USA: Elsevier.

Bakshi, S., Leoncini, E., Baker, C., Cañas-Duarte, S.J., Okumus, B. & Paulsson, J. (2021) Tracking bacterial lineages in complex and dynamic environments with applications for growth control and persistence. *Nature Microbiology*, *6*(6), 783–791. Available from: https://doi.org/10.1038/s41564-021-00900-4

Caicedo, J.C., Goodman, A., Karhohs, K.W., Cimini, B.A., Ackerman, J., Haghighi, M. et al. (2019) Nucleus segmentation across imaging experiments: the 2018 data science bowl. *Nature Methods*, *16*(12), 1247–1253. Available from: https://doi.org/10.1038/s41592-019-0612-7

Caicedo, J.C., Roth, J., Goodman, A., Becker, T., Karhohs, K.W., Broisin, M. et al. (2019) Evaluation of deep learning strategies for nucleus segmentation in fluorescence images. *Cytometry Part A*, *95*(9), 952–965. Available from: https://doi.org/10.1002/cyto.a.23863

Chen, J., Ding, L., Viana, M.P., Lee, H.W., Sluezwski, M.F., Morris, B. et al. (2020) The Allen cell and structure segmenter: a new open source toolkit for segmenting 3D intracellular structures in fluorescence microscopy images. *bioRxiv*. Available from: https://doi.org/10.1101/491035

Chen, T., Li, M., Li, Y., Lin, M., Wang, N., Wang, M. et al. (2015) MXNet: a flexible and efficient machine learning library for heterogeneous distributed systems. *arXiv*. Available from: https://doi.org/10.48550/arXiv.1512.01274

Çiçek, Ö., Abdulkadir, A., Lienkamp, S.S., Brox, T. & Ronneberger, O. (2016) 3D U-net: learning dense volumetric segmentation from sparse annotation. In: Ourselin, S., Joskowicz, L., Sabuncu, M.R., Unal, G. & Wells, W. (Eds.) *Medical image computing and computer-assisted intervention – MICCAI 2016*. Cham: Springer International Publishing, pp. 424–432. Lecture Notes in Computer Science. Available from: https://doi.org/10.1007/978-3-319-46723-8_49

Costerton, J.W., Cheng, K.J., Geesey, G.G., Ladd, T.I., Nickel, J.C., Dasgupta, M. et al. (1987) Bacterial biofilms in nature and disease. *Annual Review of Microbiology*, *41*(1), 435–464. Available from: https://doi.org/10.1146/annurev.mi.41.100187.002251

Costerton, J.W., Geesey, G.G. & Cheng, K.J. (1978) How bacteria stick. *Scientific American*, *238*(1), 86–95. Available from: https://doi.org/10.1038/scientificamerican0178-86

Cuéllar, S., Granados, P., Fabregas, E., Curé, M., Vargas, H., Dormido-Canto, S. et al. (2022) Deep learning exoplanets detection by combining real and synthetic data. *PLoS One*, *17*(5), e0268199. Available from: https://doi.org/10.1371/journal.pone.0268199

Cutler, K.J., Stringer, C., Lo, T.W., Rappez, L., Nicholas Stroustrup, S., Peterson, B. et al. (2022) Omnipose: a high-precision morphology-independent solution for bacterial cell segmentation. *Nature Methods*, *19*(11), 1438–1448. Available from: https://doi.org/10.1038/s41592-022-01639-4

Díaz-Pascual, F., Lempp, M., Nosho, K., Jeckel, H., Jo, J.K., Neuhaus, K. et al. (2021) 'Spatial alanine metabolism determines local growth dynamics of *Escherichia coli* colonies'. Edited by Karina B. Xavier, Gisela Storz, and Karina B. Xavier. *ELife*, *10*(November), e70794. Available from: https://doi.org/10.7554/eLife.70794

Drescher, K., Dunkel, J., Nadell, C.D., van Teeffelen, S., Grnja, I., Wingreen, N.S. et al. (2016) Architectural transitions in vibrio cholerae biofilms at single-cell resolution. *Proceedings of the National Academy of Sciences*, *113*(14), E2066–72. Available from: https://doi.org/10.1073/pnas.1601702113

Ducret, A., Quardokus, E.M. & Brun, Y.V. (2016) MicrobeJ, a tool for high throughput bacterial cell detection and quantitative analysis. *Nature Microbiology*, *1*(7), 1–7. Available from: https://doi.org/10.1038/nmicrobiol.2016.77

Edelstein, A.D., Tsuchida, M.A., Amodaj, N., Pinkard, H., Vale, R.D. & Stuurman, N. (2014) Advanced methods of microscope control using MManager software. *Journal of Biological Methods*, *1*(2), e10. Available from: https://doi.org/10.14440/jbm.2014.36

Ershov, D., Phan, M.-S., Pylvänäinen, J.W., Rigaud, S.U., Le Blanc, L., Charles-Orszag, A. et al. (2022) TrackMate 7: integrating state-of-the-art segmentation algorithms into tracking pipelines. *Nature Methods*, *19*(7), 829–832. Available from: https://doi.org/10.1038/s41592-022-01507-1

Evans, C.R., Kempes, C.P., Price-Whelan, A. & Dietrich, L.E.P. (2020) Metabolic heterogeneity and cross-feeding in bacterial multicellular systems. *Trends in Microbiology*, *28*(9), 732–743. Available from: https://doi.org/10.1016/j.tim.2020.03.008

Falk, T., Mai, D., Bensch, R., Çiçek, Ö., Abdulkadir, A., Marrakchi, Y. et al. (2019) U-net: deep learning for cell counting, detection, and morphometry. *Nature Methods*, *16*(1), 67–70. Available from: https://doi.org/10.1038/s41592-018-0261-2

Flemming, H.-C. & Wingender, J. (2010) The biofilm matrix. *Nature Reviews Microbiology*, *8*(9), 623–633. Available from: https://doi.org/10.1038/nrmicro2415

Flemming, H.-C. & Wuertz, S. (2019) Bacteria and archaea on earth and their abundance in biofilms. *Nature Reviews Microbiology*, *17*(4), 247–260. Available from: https://doi.org/10.1038/s41579-019-0158-9

Grantcharova, N., Peters, V., Monteiro, C., Zakikhany, K. & Römling, U. (2010) Bistable expression of CsgD in biofilm development of *Salmonella enterica* serovar typhimurium. *Journal of Bacteriology*, *192*(2), 456–466. Available from: https://doi.org/10.1128/JB.01826-08

Griffa, A., Garin, N. & Sage, D. (2010) Comparison of deconvolution software in 3D microscopy: a user point of view—part 1. *G.I.T. Imaging & Microscopy*, *12*(1), 43–45.

Hardo, G., Noka, M. & Bakshi, S. (2022) Synthetic micrographs of bacteria (SyMBac) allows accurate segmentation of bacterial cells using deep neural networks. *BMC Biology*, *20*(1), 263. Available from: https://doi.org/10.1186/s12915-022-01453-6

Harris, C.R., Jarrod Millman, K., van der Walt, S.J., Gommers, R., Virtanen, P., Cournapeau, D. et al. (2020) Array programming with NumPy. *Nature*, *585*(7825), 357–362. Available from: https://doi.org/10.1038/s41586-020-2649-2

Hartmann, R., Jeckel, H., Jelli, E., Singh, P.K., Vaidya, S., Bayer, M. et al. (2021) Quantitative image analysis of microbial communities with BiofilmQ. *Nature Microbiology*, *6*(2), 151–156. Available from: https://doi.org/10.1038/s41564-020-00817-4

Hartmann, R., Singh, P.K., Pearce, P., Mok, R., Song, B., Díaz-Pascual, F. et al. (2019) Emergence of three-dimensional order and structure in growing biofilms. *Nature Physics*, *15*(3), 251–256. Available from: https://doi.org/10.1038/s41567-018-0356-9

Hartmann, R., van Teeseling, M.C.F., Thanbichler, M. & Drescher, K. (2020) BacStalk: a comprehensive and interactive image analysis software tool for bacterial cell biology. *Molecular Microbiology*, *114*(1), 140–150. Available from: https://doi.org/10.1111/mmi.14501

Jeckel, H., Díaz-Pascual, F., Skinner, D.J., Song, B., Jiménez-Siebert, E., Strenger, K. et al. (2022) Shared biophysical mechanisms determine early biofilm architecture development across different bacterial species. *PLoS Biology*, *20*(10), e3001846. Available from: https://doi.org/10.1371/journal.pbio.3001846

Jeckel, H. & Drescher, K. (2021) Advances and opportunities in image analysis of bacterial cells and communities. *FEMS Microbiology Reviews*, *45*(4), fuaa062. Available from: https://doi.org/10.1093/femsre/fuaa062

Jo, J., Price-Whelan, A. & Dietrich, L.E.P. (2022) Gradients and consequences of heterogeneity in biofilms. *Nature Reviews Microbiology*, *20*(10), 593–607. Available from: https://doi.org/10.1038/s41579-022-00692-2

Kirshner, H., Aguet, F., Sage, D. & Unser, M. (2013) 3-D PSF fitting for fluorescence microscopy: implementation and localization application. *Journal of Microscopy*, *249*(1), 13–25. Available from: https://doi.org/10.1111/j.1365-2818.2012.03675.x

Laine, R.F., Arganda-Carreras, I., Henriques, R. & Jacquemet, G. (2021) Avoiding a replication crisis in deep-learning-based bioimage analysis. *Nature Methods*, *18*(10), 1136–1144. Available from: https://doi.org/10.1038/s41592-021-01284-3

LeCun, Y., Bengio, Y. & Hinton, G. (2015) Deep learning. *Nature*, *521*(7553), 436–444. Available from: https://doi.org/10.1038/nature14539

Lugagne, J.-B., Lin, H. & Dunlop, M.J. (2020) DeLTA: automated cell segmentation, tracking, and lineage reconstruction using deep learning. *PLoS Computational Biology*, *16*(4), e1007673. Available from: https://doi.org/10.1371/journal.pcbi.1007673

McKinney, W. (2010) Data structures for statistical computing in python. *Proceedings of the 9th Python in Science Conference*, *9*, 56–61. Available from: https://doi.org/10.25080/Majora-92bf1922-00a

Moen, E., Bannon, D., Kudo, T., Graf, W., Covert, M. & Van Valen, D. (2019) Deep learning for cellular image analysis. *Nature Methods*, *16*(12), 1233–1246. Available from: https://doi.org/10.1038/s41592-019-0403-1

Mountcastle, S.E., Vyas, N., Villapun, V.M., Cox, S.C., Jabbari, S., Sammons, R.L. et al. (2021) Biofilm viability checker: an open-source tool for automated biofilm viability analysis from confocal microscopy images. *NPJ Biofilms and Microbiomes*, *7*(1), 1–12. Available from: https://doi.org/10.1038/s41522-021-00214-7

O'Connor, O.M., Alnahhas, R.N., Lugagne, J.-B. & Dunlop, M.J. (2022) DeLTA 2.0: a deep learning pipeline for quantifying single-cell spatial and temporal dynamics. *PLoS Computational Biology*, *18*(1), e1009797. Available from: https://doi.org/10.1371/journal.pcbi.1009797

Paintdakhi, A., Parry, B., Campos, M., Irnov, I., Elf, J., Surovtsev, I. et al. (2016) Oufti: an integrated software package for high-accuracy, high-throughput quantitative microscopy analysis. *Molecular Microbiology*, *99*(4), 767–777. Available from: https://doi.org/10.1111/mmi.13264

Poucin, F., Kraus, A. & Simon, M. (2021) Boosting instance segmentation with synthetic data: a study to overcome the limits of real world data sets. In: *2021 IEEE/CVF international conference on computer vision workshops (ICCVW)*. New York City, NY, USA: IEEE, pp. 945–953. Available from: https://doi.org/10.1109/ICCVW54120.2021.00110

Qin, B., Fei, C., Bridges, A.A., Mashruwala, A.A., Stone, H.A., Wingreen, N.S. et al. (2020) Cell position fates and collective fountain flow in bacterial biofilms revealed by light-sheet microscopy. *Science*, *369*(6499), 71–77. Available from: https://doi.org/10.1126/science.abb8501

Römling, U., Bian, Z., Hammar, M., Sierralta, W.D. & Normark, S. (1998) Curli fibers are highly conserved between *Salmonella typhimurium* and *Escherichia coli* with respect to operon structure and regulation. *Journal of Bacteriology*, *180*(3), 722–731. Available from: https://doi.org/10.1128/JB.180.3.722-731.1998

Römling, U., Sierralta, W.D., Eriksson, K. & Normark, S. (1998) Multicellular and aggregative behaviour of *Salmonella typhimurium* strains is controlled by mutations in the AgfD promoter. *Molecular Microbiology*, *28*(2), 249–264. Available from: https://doi.org/10.1046/j.1365-2958.1998.00791.x

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T. et al. (2012) Fiji: an open-source platform for biological-image analysis. *Nature Methods*, *9*(7), 676–682. Available from: https://doi.org/10.1038/nmeth.2019

Schmidt, U., Weigert, M., Broaddus, C. & Myers, G. (2018) Cell detection with star-convex polygons. In: Frangi, A.F., Schnabel, J.A., Davatzikos, C., Alberola-López, C. & Fichtinger, G. (Eds.) *Medical image computing and computer assisted intervention – MICCAI 2018*. Cham: Springer International Publishing, pp. 265–273. Lecture Notes in Computer Science. Available from: https://doi.org/10.1007/978-3-030-00934-2_30

Sergeev, A. & Del Balso, M. (2018) Horovod: fast and easy distributed deep learning in TensorFlow. *arXiv*. Available from: https://doi.org/10.48550/arXiv.1802.05799

Serra, D.O., Richter, A.M., Klauck, G., Mika, F. & Hengge, R. (2013) Microanatomy at cellular resolution and spatial order of physiological differentiation in a bacterial biofilm. *MBio*, *4*(2), e00103-13. Available from: https://doi.org/10.1128/mBio.00103-13

Sofroniew, N., Lambert, T., Evans, K., Nunez-Iglesias, J., Bokota, G., Winston, P. et al. (2022) *Napari: a multi-dimensional image viewer for python*. Geneva, Switzerland: CERN. Available from: https://doi.org/10.5281/zenodo.7276432

Solovyev, R., Wang, W. & Gabruseva, T. (2021) Weighted boxes fusion: ensembling boxes from different object detection models. *Image and Vision Computing*, *107*, 104117. Available from: https://doi.org/10.1016/j.imavis.2021.104117

Stewart, E.J., Satorius, A.E., Younger, J.G. & Solomon, M.J. (2013) Role of environmental and antibiotic stress on staphylococcus epidermidis biofilm microstructure. *Langmuir*, *29*(23), 7017–7024. Available from: https://doi.org/10.1021/la401322k

Stewart, P.S. & Franklin, M.J. (2008) Physiological heterogeneity in biofilms. *Nature Reviews Microbiology*, *6*(3), 199–210. Available from: https://doi.org/10.1038/nrmicro1838

Stirling, D.R., Swain-Bowden, M.J., Lucas, A.M., Carpenter, A.E., Cimini, B.A. & Goodman, A. (2021) CellProfiler 4: improvements in speed, utility and usability. *BMC Bioinformatics*, *22*(1), 433. Available from: https://doi.org/10.1186/s12859-021-04344-9

Stringer, C., Wang, T., Michaelos, M. & Pachitariu, M. (2021) Cellpose: a generalist algorithm for cellular segmentation. *Nature Methods*, *18*(1), 100–106. Available from: https://doi.org/10.1038/s41592-020-01018-x

Suga, S., Nakamura, K., Humbel, B.M., Kawai, H. & Hirabayashi, Y. (2021) An interactive deep learning-based approach reveals mitochondrial cristae topologies. *bioRxiv*. Available from: https://doi.org/10.1101/2021.06.11.448083

Taha, A.A. & Hanbury, A. (2015) Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool. *BMC Medical Imaging*, *15*(1), 29. Available from: https://doi.org/10.1186/s12880-015-0068-x

The Pandas Development Team. (2023) *Pandas-Dev/Pandas: Pandas*. Geneva, Switzerland: CERN. Available from: https://doi.org/10.5281/zenodo.7658911

Tinevez, J.-Y., Perry, N., Schindelin, J., Hoopes, G.M., Reynolds, G.D., Laplantine, E. et al. (2017) TrackMate: an open and extensible platform for single-particle tracking. *Methods*, *Image Processing for Biologists*, *115*(February), 80–90. Available from: https://doi.org/10.1016/j.ymeth.2016.09.016

Tobin, J., Fong, R., Ray, A., Schneider, J., Zaremba, W. & Abbeel, P. (2017) Domain randomization for transferring deep neural networks from simulation to the real world. In: *2017 IEEE/RSJ international conference on intelligent robots and systems (IROS)*. New York City, NY, USA: IEEE, pp. 23–30. Available from: https://doi.org/10.1109/IROS.2017.8202133

Toma, T.T., Wu, Y., Wang, J., Srivastava, A., Gahlmann, A. & Acton, S.T. (2022) Realistic-shape bacterial biofilm simulator for deep learning-based 3D single-cell segmentation. In: *2022 IEEE 19th international symposium on biomedical imaging (ISBI)*. New York City, NY, USA: IEEE, pp. 1–5. Available from: https://doi.org/10.1109/ISBI52829.2022.9761655

Valen, V., David, A., Kudo, T., Lane, K.M., Macklin, D.N., Quach, N.T. et al. (2016) Deep learning automates the quantitative analysis of individual cells in live-cell imaging experiments. *PLoS Computational Biology*, *12*(11), e1005177. Available from: https://doi.org/10.1371/journal.pcbi.1005177

von Chamier, L., Laine, R.F. & Henriques, R. (2019) Artificial intelligence for microscopy: what you should know. *Biochemical Society Transactions*, *47*(4), 1029–1040. Available from: https://doi.org/10.1042/BST20180391

von Chamier, L., Laine, R.F., Jukkala, J., Spahn, C., Krentzel, D., Nehme, E. et al. (2021) Democratising deep learning for microscopy with ZeroCostDL4Mic. *Nature Communications*, *12*(1), 2276. Available from: https://doi.org/10.1038/s41467-021-22518-0

Wang, Y., Eddison, M., Fleishman, G., Weigert, M., Shengjin, X., Wang, T. et al. (2021) EASI-FISH for thick tissue defines lateral hypothalamus Spatio-molecular organization. *Cell*, *184*(26), 6361–6377.e24. Available from: https://doi.org/10.1016/j.cell.2021.11.024

Weigert, M., Schmidt, U., Boothe, T., Müller, A., Dibrov, A., Jain, A. et al. (2018) Content-aware image restoration: pushing the limits of fluorescence microscopy. *Nature Methods*, *15*(12), 1090–1097. Available from: https://doi.org/10.1038/s41592-018-0216-7

Weigert, M., Schmidt, U., Haase, R., Sugawara, K. & Myers, G. (2020) Star-convex Polyhedra for 3D object detection and segmentation in microscopy. In: *2020 IEEE winter conference on applications of computer vision (WACV)*. New York City, NY, USA: IEEE, pp. 3655–3662. Available from: https://doi.org/10.1109/WACV45572.2020.9093435

Wolny, A., Cerrone, L., Vijayan, A., Tofanelli, R., Barro, A.V., Louveaux, M. et al. (2020) 'Accurate and versatile 3D segmentation of plant tissues at cellular resolution'. Edited by Christian S Hardtke, Dominique C Bergmann, Dominique C Bergmann, and Moritz Graeff. *ELife*, *9*(July), e57613. Available from: https://doi.org/10.7554/eLife.57613

Yan, J., Sharo, A.G., Stone, H.A., Wingreen, N.S. & Bassler, B.L. (2016) Vibrio cholerae biofilm growth program and architecture revealed by single-cell live imaging. *Proceedings of the National Academy of Sciences*, *113*(36), E5337–E5343. Available from: https://doi.org/10.1073/pnas.1611494113

Yordanov, S., Neuhaus, K., Hartmann, R., Díaz-Pascual, F., Vidakovic, L., Singh, P.K. et al. (2021) Single-objective high-resolution confocal light sheet fluorescence microscopy for standard biological sample geometries. *Biomedical Optics Express*, *12*(6), 3372–3391. Available from: https://doi.org/10.1364/BOE.420788

Zhang, J., Wang, Y., Donarski, E.D., Toma, T.T., Miles, M.T., Acton, S.T. et al. (2022) BCM3D 2.0: accurate segmentation of single bacterial cells in dense biofilms using computationally generated intermediate image representations. *NPJ Biofilms and Microbiomes*, *8*(1), 1–13. Available from: https://doi.org/10.1038/s41522-022-00362-4

Zhang, M., Zhang, J., Wang, Y., Wang, J., Achimovich, A.M., Acton, S.T. et al. (2020) Non-invasive single-cell morphometry in living bacterial biofilms. *Nature Communications*, *11*(1), 6151. Available from: https://doi.org/10.1038/s41467-020-19866-8

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Jelli, E., Ohmura, T., Netter, N., Abt, M., Jiménez-Siebert, E., Neuhaus, K., Rode, D. K. H., Nadell, C. D. & Drescher, K. (2023). Single-cell segmentation in bacterial biofilms with an optimized deep learning method enables tracking of cell lineages and measurements of growth rates. *Molecular Microbiology*, *119*, 659–676. https://doi.org/10.1111/mmi.15064