

The TYC Dataset for Understanding Instance-Level Semantics and Motions of Cells in Microstructures

Christoph Reich

Tim Prangemeier

Heinz Koepl

Department of Electrical Engineering and Information Technology, Centre for Synthetic Biology,
Technische Universität Darmstadt

{christoph.reich, tim.prangemeier, heinz.koepl}@bcs.tu-darmstadt.de

https://christophreich1996.github.io/tyc_dataset

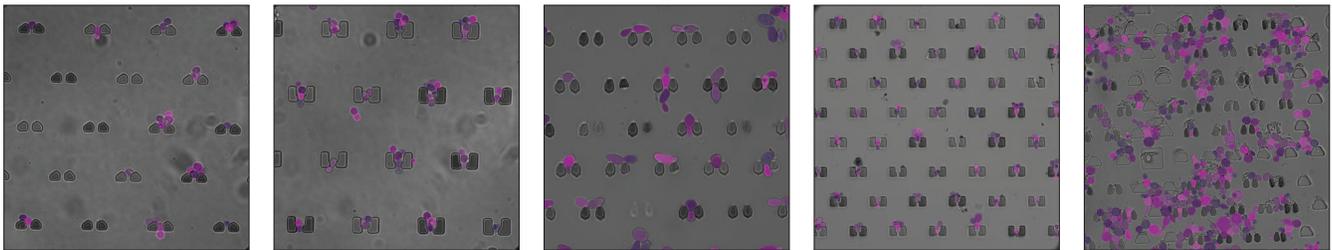


Figure 1. **Dataset teaser.** Samples of our trapped yeast cell dataset of high-resolution ($\geq 2048 \times 2048$) brightfield images overlaid with multi-class instance segmentation labels; shades of (dark) grey (■) indicate microstructures (traps) and shades of pink (■) indicate individual cell instances. Image brightness adjusted for better visualization. Best viewed in color; zoom in for details.

Abstract

Segmenting cells and tracking their motion over time is a common task in biomedical applications. However, predicting accurate instance-wise segmentation and cell motions from microscopy imagery remains a challenging task. Using microstructured environments for analyzing single cells in a constant flow of media adds additional complexity. While large-scale labeled microscopy datasets are available, we are not aware of any large-scale dataset, including both cells and microstructures. In this paper, we introduce the trapped yeast cell (TYC) dataset, a novel dataset for understanding instance-level semantics and motions of cells in microstructures. We release 105 dense annotated high-resolution brightfield microscopy images, including about 19k instance masks. We also release 261 curated video clips composed of 1293 high-resolution microscopy images to facilitate unsupervised understanding of cell motions and morphology. TYC offers ten times more instance annotations than the previously largest dataset, including cells and microstructures. Our effort also exceeds previous attempts in terms of microstructure variability, resolution, complexity, and capturing device (microscopy) variability. We facilitate a unified comparison on our novel dataset by introducing a standardized evaluation strategy. TYC and evaluation code are publicly available under CC BY 4.0 license.

1. Introduction

Detecting, segmenting, and tracking individual cells in microscopy images is fundamental for many biomedical applications [21, 51, 53, 80]. For example, accurate segmentation and tracking of individual cells are essential for analyzing the cellular processes of living cells in time-lapse fluorescence microscopy (TLFM) experiments [3, 66]. In general, computer vision-based single-cell analysis can aid the development of personalized medicine, early tumor detection, and the analysis of biological signal transduction, amongst others [32, 39, 65, 66, 79].

Deep neural networks have become the predominant workhorse of current state-of-the-art algorithms in the domain of computer vision [13, 38, 70, 81]. However, unleashing the full potential of deep neural networks requires vast amounts of data [7, 36, 87]. Subsequently, the widespread availability of large-scale public datasets, such as ImageNet [70], KITTI [25], Cityscapes [13], Ego4D [26], SA-1B [36], and LAION-5B [74], has been a major contributor to the recent success of deep neural network approaches. While large-scale (labeled) datasets in the domain of single-cell analysis are available, these datasets do not consider microstructured environments [8, 20, 62]. Significantly limiting the applicability of models trained without microstructures to analyze cells in microstructures due to the significant domain gap.

Dataset	# annotated pixels [10^7]	# cells	# traps	Resolution	Annotation type	# trap types
Bakker <i>et al.</i> [3]	–	1000	0	512×512	Cell outline	1
Reich <i>et al.</i> [67]	0.81	914	971	128×128	Instance segmentation	2
TYC dataset (ours)	46.39	14541	4405	$\geq 2048 \times 2048$	Instance segmentation	6

Table 1. **Dataset comparison.** High-level statistics of our TYC dataset (labeled set) and other segmentation dataset including cells and microstructures. Our TYC datasets exceeds existing efforts in all high-level statistics.

Microstructured environments are commonly employed to analyze populations of hundreds or thousands of cells individually [3, 14, 27, 33, 47, 58, 61, 64, 66, 76]. A constant flow of growth media hydrodynamically traps cells within the microstructures, enabling their analysis over time [3, 66]. Understanding the instance-level semantics and motions of cells in microstructured environments is particularly challenging due to the perceptual similarity of microstructures and cells (*cf.* Fig. 1). In this paper, we propose the trapped yeast cell dataset for understanding instance-level semantics and motions of trapped yeast cells. Our TYC dataset of high-resolution ($\geq 2048 \times 2048$) bright-field microscopy images includes both a labeled instance segmentation set and an unlabeled set of video clips. In total, we release 18946 instance masks, of cells and microstructures (*cf.* Fig. 1). Our labeled dataset exceeds all existing cells in microstructures datasets in terms of annotated pixels, number of instances (cells & traps), resolution, and microstructure variability (*cf.* Tab. 1). Our unlabeled set contains 261 curated high-resolution microscopy video clips for unsupervised understanding cell motions facilitating cell tracking.

In addition to the proposed dataset, we also present a standardized evaluation strategy to facilitate fair comparison of future work with our dataset. This may also serve to standardize the results of biological investigation, making them comparable between laboratories. In particular, we present an instance-level and semantic-level evaluation strategy for our labeled set. Our evaluation strategy considers both the downstream biological application and the raw segmentation performance for benchmarking.

To showcase the complexity of our TYC dataset, we report qualitative segmentation results of the Segment Anything Model (SAM), a recent foundation model for segmentation [36]. While SAM is able to provide fairly accurate instance masks for very simple cell and microstructure configurations, SAM fails to provide useful masks as scenes get more complex (*e.g.*, touching cells). This demonstrates the complexity of segmenting cells in microstructures.

2. Related Work

Many biological applications rely on microscopy image data for single-cell analysis [56]. Most applications require instance segmentation of cells [53, 54, 56]. Object tracking

is required to analyze cells through time [51]. The vast majority of state-of-the-art cell segmentation approaches rely on deep neural networks [57, 22]; prominent examples include, U-Net [69] StarDist [73], and Cellpose [60, 78]. Similarly, current cell tracking approaches also predominately rely on deep neural networks [4, 10, 31, 44, 46, 51, 80, 72].

The widespread use of deep learning in cell analytics tasks has motivated the construction of large and general datasets [57, 20]. Caicedo *et al.* [8] proposed the 2018 Data Science Bowl dataset for single-cell segmentation in fluorescence imagery, composed of 670 images with instance-wise annotations. Recently, Edlund *et al.* [20] released the large-scale microscopy image dataset LIVECell, for single-cell (instance) segmentation, with 5239 densely annotated phase-contrast images and eight cell types. The ISBI Cell Tracking Challenge [51, 80] offers multiple datasets for both cell segmentation and tracking. Anjum *et al.* [1] proposed CTMC with 86 annotated videos.

While significant efforts have been devoted to the development of cell analytics datasets, only very limited (annotated) data of cells in microstructures is available. Bakker *et al.* [3] released about 1000 cell outline annotations of yeast cells in microstructures. Recently, Reich *et al.* [67] proposed the first instance segmentation dataset of yeast cells in microstructured environments. These datasets, however, entail major limitations (*cf.* Tab. 1). Both datasets include only a limited number of annotated objects (cells and traps) and capture only a small number of different trap types. Additionally, the dataset by Reich *et al.* [67] only includes low-resolution crops of high-resolution brightfield microscopy images. The annotated set of our TYC dataset exceeds all previous datasets of cells in microstructures in terms of all high-level dataset statistics. Tab. 1 provides a comprehensive overview between existing datasets including cells and microstructures and our TYC dataset (labeled set).

To overcome the limited availability of annotated microscopy data, some approaches aim to synthesize microscopy images and annotations [37, 71]. While synthesized data can be effective for simple cell analytic tasks, these approaches are not able to synthesize complicated cell and microstructure scenes as well as vast and complex cell motions. Additionally, synthetic images commonly suffer from a domain gap w.r.t. real images, requiring approaches to address this domain gap when deep neural networks are pre-trained on synthetic data [23, 59, 82].

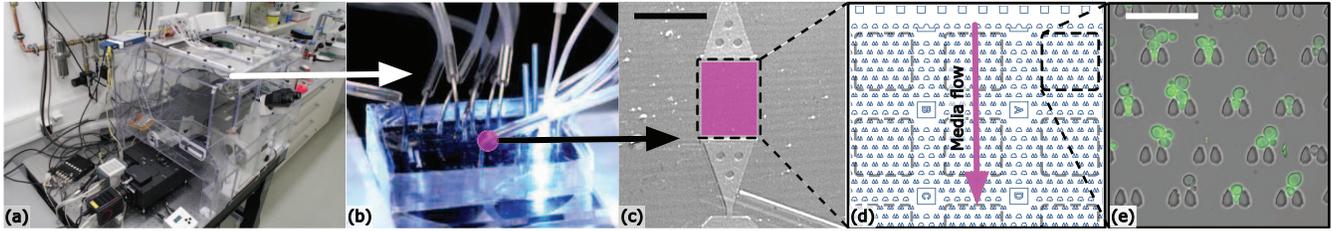


Figure 2. **Overview data acquisition process.** General TLFM experiment setup for single-cell fluorescence measurement. Microscope (from left to right), microfluidic chip (cell chamber within the purple dot), trap chamber (pink rectangle), trap microstructure design, traps with cells (cells in green, fluorescent overlay, right); black scale bar 1 mm white scale bar 10 μm .

3. Dataset

We made numerous design choices during data collection, curation, and annotation of our TYC dataset. This section reports these design decisions and provides dataset statistics. We selected 105 high-resolution ($\geq 2048 \times 2048$) brightfield images for our labeled set, from over a dozen TLFM experiments. The unlabelled video set contains 261 manually-selected video clips. The selection encompasses the entirety of the data distribution: it includes edge cases, various magnifications, and diverse focal positions. The key objects in the imagery are yeast cells (various strains) and trap microstructures. Six widespread trap geometries are included in the dataset (*cf.* Fig. 3), in order to ensure that models trained on our TYC dataset can generalize over a wide variety of microstructure geometries.

3.1. Data acquisition

We recorded the dataset with a computer-controlled microscope (Nikon Eclipse Ti, Fig. 2a). The yeast cells were cultured in a tightly controlled environment within a microfluidic chip (Fig. 2b) comprised of a microscopy cover slip and transparent Polydimethylsiloxane (PDMS). The trap chamber contains approximately 1000 trap microstructures (*cf.* Fig. 2c). Various trap geometries are available (*cf.* Fig. 3). The microstructures and a constant flow of yeast growth media hydrodynamically trap the cells, constraining them laterally in XY (*cf.* Fig. 2d & 2e). The axial constraint is provided by the coverslip and PDMS ceiling, the space between which is on the order of a cell diameter. This facilitates a continuously uniform focus of the cells. The entire chip is maintained at a temperature of 30 $^{\circ}\text{C}$ and together with the flow of yeast growth media this enables yeast to grow for prolonged periods and over multiple cell cycles. The cells bud daughter cells multiple times within the course of a typical experiment (budding approximately every 90min, experiments typically run for 15h), and these are flushed out by the continuous flow to avoid clogging of the chip.

We recorded time-lapse brightfield (transmitted light) and fluorescent channel imagery of the budding yeast cells every 10min. We used either NIS-Elements or $\mu\text{Manager}$

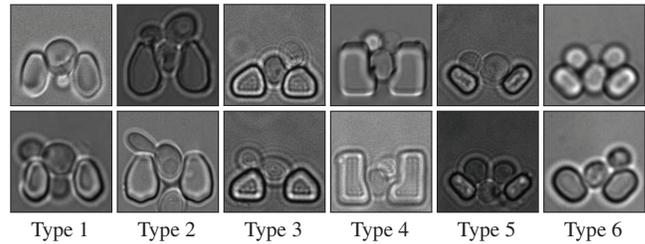


Figure 3. **Trap types.** Overview of trap microstructure geometries included in our TYC dataset, cropped from high-resolution microscopy images (*cf.* Fig. 1).

to control the microscope in different experiments [19]. Imagery from both 100 \times and 60 \times objectives, which are widespread in order to resolve the cells, are included. A CoolLED pE-100 and a Lumencor SpectraX light engine illuminated the respective channels and lighting conditions are varied throughout the dataset. Resolutions are 2048×2048 and 2304×2304 recorded with Hamamatsu cameras ORCA-Flash4.0 V2, V3, and with an ORCA-Fusion. Multiple lateral and axial positions were recorded sequentially at each timestep (Fig. 2). In addition to our own data, selected imagery from the dataset by Bakker *et al.* [3] are included in our TYC dataset (upscaled from 512×512 to 2048×2048 to match our data).

3.2. Biological application

Creating novel functionality from standardized parts is central to synthetic biology. Ideally, synthetic biological circuitry is rationally designed *in silico* to predetermined design specifications [5, 64], *e.g.*, to detect and kill cancer [75, 85], but also for a plethora of other applications and real-world problems. *Saccharomyces cerevisiae* (yeast) and *E. coli* are the model organisms of choice in synthetic biology [64], both in terms of development and characterization of novel biological circuitry. With TYC we focus on yeast.

TLFM plays a central role in the thorough characterization and standardization of biological investigation, as it is the only technique available to capture both the population heterogeneity and dynamics of synthetic circuitry on the single-cell level [39, 43, 45]. However, extracting the relevant information from within the imagery, *i.e.* measur-

ing fluorescence via cell segmentation, is currently a limiting factor and there is no standard for it or calibration set to compare vision model-based measurements upon. The TYC dataset will not only facilitate the development of new vision models that resolve the current bottleneck in extracting the measurement data reliably but also provide a benchmark for all of the future models to be calibrated against and contribute to the reproducibility of TLFM-based biological investigation and development.

3.3. Annotations and classes

We present a labeled set composed of 105 high-resolution brightfield microscopy images with instance-level annotations for cells and microstructures. An example of our annotated microscopy images is presented in Fig. 4.

Our dense annotations consist of non-overlapping pixel-wise instance maps with semantic class annotation. We annotated all images in-house to ensure consistent and high-quality annotations. Annotation time per image differs between about 30min for microscopy images with only a few cells (< 20 cell instance) and up to 2h for images with more than 300 cell instances (*cf.* Fig. 1 and 4).

We assume no overlap between instances since the employed microfluidic chips ensure a monolayer of cells. For ambiguous configurations (*e.g.*, touching and slightly overlapping cells), we label ambiguous regions to belong to the upper cell instance. Note due to accurate microfluidic chip production, ambiguous cases are a rarity.

Following prior work, our instance segmentation annotations include two semantic classes, cell and (microstructure) trap [67]. While instance segmentations of merely the cells suffice for most applications, we also include instance-level annotations of microstructures [56]. We motivate this design choice twofold. First, learning to distinguish between perceptually similar cell and trap instances might be enforced by explicitly learning to segment both cells and traps. Second, understanding the location of both cell and trap instance enables us to determine cells that are hydrodynamically trapped and those that are not trapped, which are likely washed out of the image and the whole chip [64, 67].

All pixels not labeled as a trap or cell instance are assumed to belong to the background. Since we only assume a single background class, our annotations can also be seen as panoptic segmentation labels [35]. In particular, panoptic segmentation distinguishes between *things* - countable objects like cells or traps and *stuff* - amorphous regions such as the background. We later use this property during evaluation and assume that cell and trap instances belong to the things category and the background to the stuff category.

During annotation, distinguishing between cell instances is well-defined by the cell membrane. Budding yeast cells, in which a daughter cell buds off the mother cell, represent a non-trivial edge case [18]. We label daughter cells as sepa-

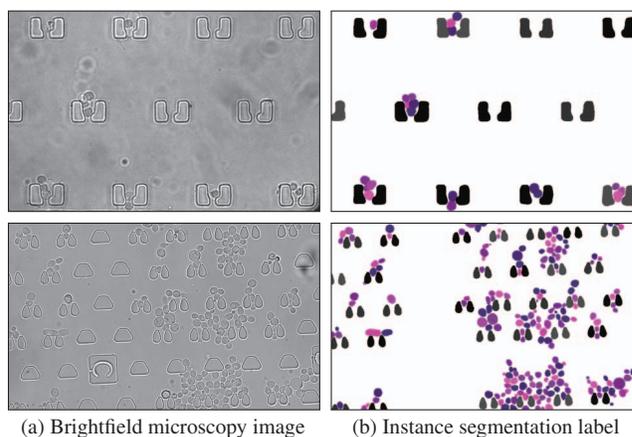


Figure 4. **Instance annotations.** Brightfield microscopy images including cells & microstructures on the left (a) and the corresponding instance segmentation labels on the right (b). Shades of grey (■ ■ ■) indicate microstructures (traps) and shades of pink (■ ■ ■) indicate individual cell. For visualization purposes, we adjusted the brightness of the microscopy image and show a crop of the full image. Best viewed in color; zoom in for details.

rate cell instances as soon as the connecting region between the mother and daughter becomes convex or thinner than the daughter cell’s diameter. The process of budding is shown in Fig. 16; additionally, Fig. 5 provides annotated examples of budding cells.

Fabrication errors can lead to damaged or fully broken microstructures. While we generally label a pair of traps as a single microstructure instance (*cf.* Fig. 4), broken traps lead to ambiguities. If a trap is damaged, but the general structure still follows the structure of an intact trap, we annotate the full trap pair (*cf.* Fig. 6 right example). In case the structure of a broken trap is vastly different from an intact trap, we refrain from labeling the broken trap (*cf.* Fig. 6 left example). In case a trap is broken and overlapped by a cell instance, we also refrain from labeling the broken trap (*cf.* Fig. 6 second left example). Note both previous cases lead to microstructure instances, including only a single trap, not a trap pair (standard case).

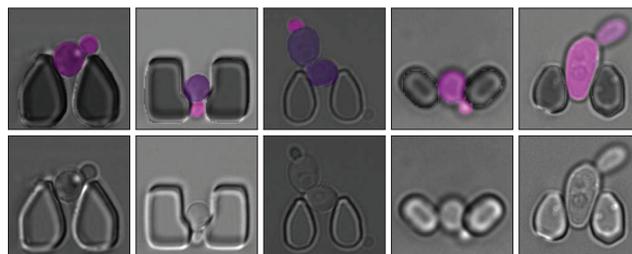


Figure 5. **Examples of budding yeast cells.** A daughter cells (small cells) bud from a trapped mother cells. The bottom row shows the raw image and the top row an overlay between the label and the raw image. Crops from full images are shown.

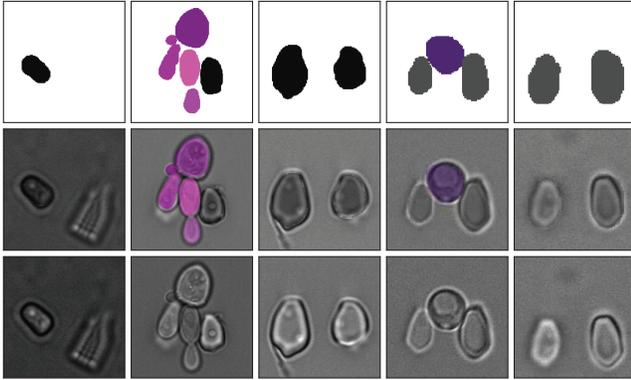


Figure 6. **Trap fabrication errors.** Examples of damaged and broken traps. The top row shows the instance segmentation label, the bottom row shows the brightfield image, and the middle row shows an overlay of both. Crops from full images are shown.

3.4. Statistical analysis

This subsection provides a statistical analysis of our labeled TYC dataset. We also compare our dataset to the dataset proposed by Reich *et al.* [67]. Note that a high-level comparison against related datasets is provided in Tab. 1.

Fig. 7 compares the distribution of semantic annotations. 6% of all annotated pixels belong to the cell class. Similarly, also 6% of all annotated pixels are labeled as traps. The dataset by Reich *et al.* [67] includes a higher pixel-wise density of cells (8%) and traps (16%). Notably, our dataset provides a more balanced semantic distribution than the dataset by Reich *et al.* [67].

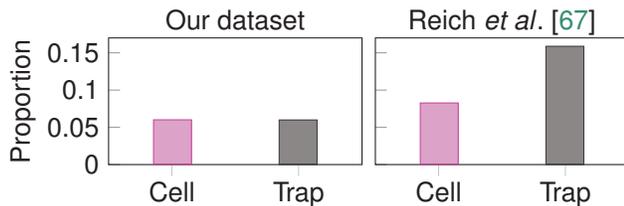


Figure 7. **Semantic distribution.** Portion of labeled pixels (y-axis) per semantic class (x-axis) for our TYC dataset and [67]. Note pixels not labeled as cell or trap belong to the background.

Our labeled dataset of brightfield microscopy images includes vastly different numbers of object instances. On average, a microscopy image of our labeled set includes 138 cells and 42 trap pairs. In the extreme, the maximum number of cells in an image is 469, and the maximum number of traps is 67. The minimum number of cells in a microscopy image is 6, and the minimum number of trap pair instances is 12. Fig. 8 showcases a histogram of object instances per image for both the cell and trap class. It can be observed that the range of cell instance counts is significantly larger than the range of the number of trap instances.

Fig. 9 showcases the histogram of object sizes in our labeled set and the dataset by Reich *et al.* [67]. We observe

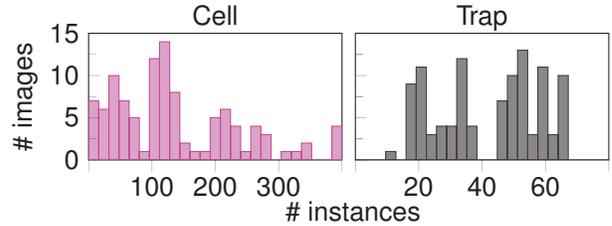


Figure 8. **Number of instance in images.** Histogram visualizing the frequency of number of object instances (cells & trap) in our labeled dataset. The left plot shows the cell class object counts (pink ■) and the right plot the trap class object counts (grey ■).

that while our labeled dataset includes more objects, our dataset also entails a broader distribution of cell and trap sizes. In particular, our labeled dataset includes more small cell instances than the dataset of Reich *et al.* [67]. While the dataset by Reich *et al.* [67] includes primary trap instances of the same shape, our dataset offers a wider range of different trap sizes.

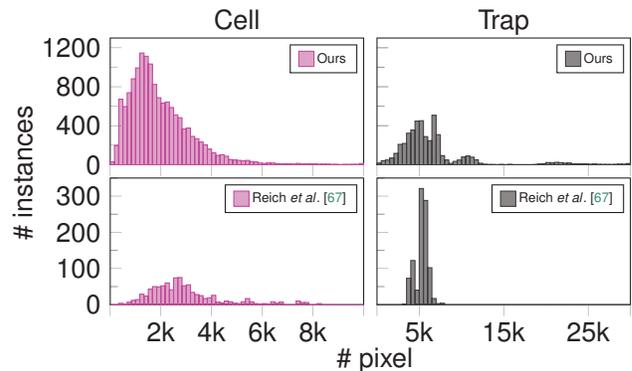


Figure 9. **Object size histogram.** Histogram showing the frequency of different object instance sizes (cells & trap) in our labeled dataset. The left plot shows the cell class object sizes (pink ■) and the right plot the trap class object sizes (grey ■).

Our TYC dataset includes six different microstructure trap geometries (Fig. 3). The frequency of these different trap types is presented in Fig. 10. Type 2 is routinely employed in our wet lab and is the most frequent in the dataset. Other traps serve more specialized roles.

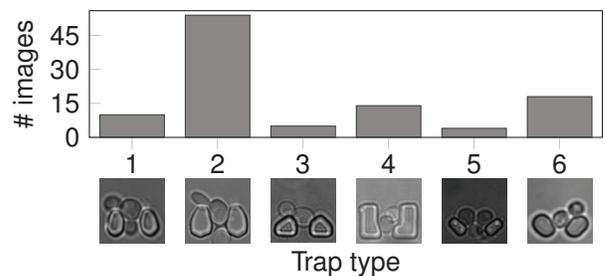


Figure 10. **Frequency of trap types.** Frequency of different trap types in our TYC dataset (labeled set).

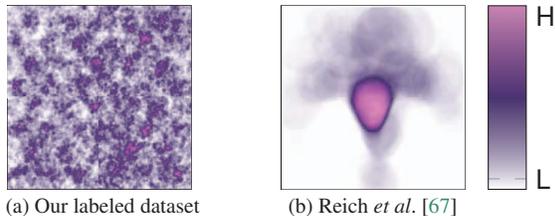


Figure 11. **Cell density map.** Cell location density map of (a) our dataset and (b) [67]. Areas in pink (■) indicate a high cell density (H). Regions in white indicate a low cell density (L).

The spatial distribution of cells in our labeled set is visualized in Fig. 11. We observe that our labeled set does not have a positional bias. In contrast, the dataset by Reich *et al.* [67] entails a strong positional bias, with cells primarily located in the center of the image. This bias is caused by the fact that the dataset by Reich *et al.* [67] utilizes crops of the full microscopy image, including only a single trap pair.

Similar to the cell positioning, the spatial distribution of traps also shows no predominant positional bias (*cf.* Fig. 12). While there is no general positional bias towards specific image regions, the regular grid-based positioning of the microstructured traps can be observed. Thus, understanding the general grid-based positioning of traps might enable accurate detection and segmentation of trap instances. In contrast to our dataset, the dataset by Reich *et al.* [67] entails a strong positional bias of trap positions. This is due to the fact that Reich *et al.* [67] utilizes trap-centered crops of the full microscopy images.

We include some challenging edge cases in our labeled dataset, in order to facilitate robust segmentation models. Examples of scenarios where additional objects are in the imagery include debris or contamination on the chip (*cf.* Fig. 13). In some TLFM experiments, microchips can get contaminated by debris. In total, our labeled set includes 11 microscopy images with such challenging edge cases. Debris adds additional complexity to the microscopy images since debris around cells perceptually affects the appearance of the cell borders.

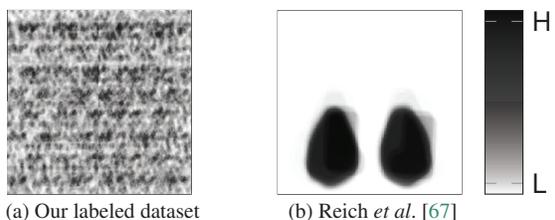


Figure 12. **Trap density map.** Trap location density map of (a) our dataset and (b) the dataset by Reich *et al.* [67]. Black (■) indicate regions where many traps are located (H). White areas showcase regions where only a few or no traps are located (L).

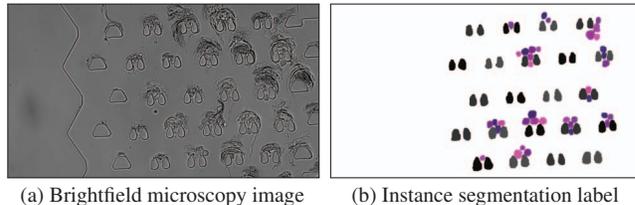


Figure 13. **Strong debris example.** Example of strong debris in the microfluidic chip. For visualization purpose we adjusted the brightness of the microscopy image and show a crop of the full image. Best viewed in color; zoom in for details.

3.5. Dataset split

We split our densely instance-wise annotated set into separate sets for training, validation, and testing. For testing, we provide two separate sets. The standard test set partly contains data from experiments that also contributed to the training and validation set. Note while data can be from the same experiment, we do not use the same specific position (*cf.* Fig. 2) for the separate sets. Additionally, we also provide an out-of-distribution (OOD) test set, which only includes images from separate experiments not present in the training, validation, and standard test set. A distribution shift occurs between experiments, due to variations, such as in microchip fabrication or lighting.

Our dataset split is described in Tab. 2. While we initially use a random split for the training, validation, and test set, we later curated all three sets to include a representative number of trap types, debris, and focal positions. For the OOD test set, we hand-pick specific samples.

Split	# images	# ann. pixels [10^7]	# cells	# traps
Training	81	35.53	12296	3448
Validation	8	3.69	950	310
Test	8	3.59	753	346
OOD Test	8	3.58	542	301

Table 2. **Dataset split.** Training, validation, test, and OOD test split of our labeled dataset.

3.6. Unlabeled video data

In addition to the labeled dataset, TYC also includes a large unlabeled dataset, including high-resolution TLFM clips. Motivated by the recent advances in unsupervised optical flow estimation [34, 50, 55, 68, 77, 83] and the use of optical flow for tracking [15, 17, 28, 41, 48, 84, 88], we aim to facilitate the unsupervised understanding of cell motions with our unlabeled dataset. Our unlabeled dataset might also facilitate general unsupervised representation learning [2, 9, 11, 24, 29]. In total, we provide 261 curated video clips, including 1293 high-resolution frames.

The clip length ranges from 3 frames to up to 11 frames with a Δt of 10min. Fig. 14 shows a histogram of the dif-

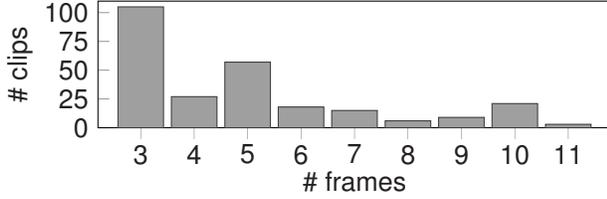


Figure 14. **Clip lengths histogram.** Frequency of clip lengths.

ferent clip lengths. While longer sequences are available, the majority of our clips are composed of three frames.

We curate our unlabeled dataset; in particular, we ensure no drift of the microfluidic chip. This is to ensure limited global motions, facilitating the learning of fine cell motions. Additionally, we only include clips where no significant amount of cells are washed in or out. While we have no significant amount of washed-out cells, our dataset still includes complex cases of vast cell motions and occasional washed-out cells. Examples of vast cell motions and washed-out cells are shown in Fig. 15.

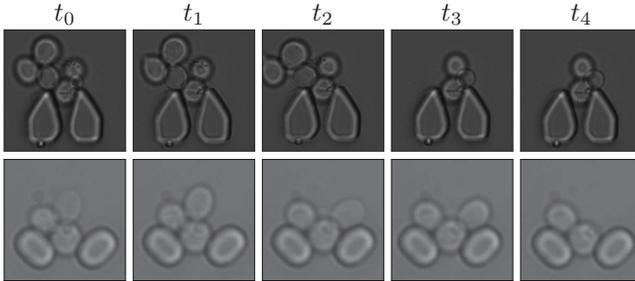


Figure 15. **Washed out cells.** Example of cells hydrodynamically washed out of a trap pair. In the upper clip, three cells are washed out. In the lower clip, only a single cell is washed away. Frames cropped from the full microscopy clip. Δt is 10min.

Similarly to the labeled dataset, our large unlabeled dataset also includes budding yeast cells. We provide a visualization of the temporal appearance of the budding process in Fig. 16. The process of budding is, in particular, challenging since small daughter cells appear out of the void due to the large Δt of 10min between frames.

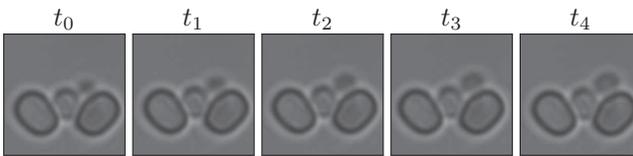


Figure 16. **Budding yeast cell over time.** Temporal sequence of our unlabeled set showcasing budding. A cell grows to the top of the trap. Frames cropped from the full video clip. Δt is 10min.

4. Evaluation

We propose a standardized evaluation approach for measuring the segmentation performance on our TYC dataset (labeled set). While our dataset offers instance-wise annotations and most biological applications require single-cell information, some biological applications might only rely on semantic-level information (*e.g.*, the fluorescence of all cells). To this end, we offer both semantic and instance-level evaluation approaches. Note our evaluation strategy is strongly inspired by the recent work from Reich *et al.* [67].

4.1. Semantic-level evaluation

For evaluating the semantic-level performance of segmentation approaches we utilize the cell class intersection-over-union (IoU). The cell class IoU is computed by:

$$\text{IoU}(p_c, g_c) = \frac{|p_c \cap g_c|}{|p_c \cup g_c|}, \quad (1)$$

where g_c is the ground truth set of pixels for the cell class and p_c denotes the predicted set of cell class pixels. The motivation for utilizing the cell class IoU is twofold. First, the cell class IoU is, in particular, significant for biomedical applications only requiring semantic predictions, *e.g.*, measuring the fluorescence of all cells. Second, as previous work (*e.g.*, [3] & [65]) reports this metric when evaluating on the dataset by Reich *et al.* [67] using the cell class IoU provides a point of comparison to existing work.

4.2. Instance-level evaluation

We propose to utilize the panoptic quality (PQ) for measuring the instance-level performance on our TYC dataset. In particular, we make use of the property that our instance segmentation labels can also be interpreted as panoptic annotations since we only have one background class and assume no instance overlap (*cf.* Sec. 3.3).

For calculating the PQ, we first match all predicted object masks of a semantic class with the ground truth masks. From this matching, three categories (per semantic class) emerge: True Positive (TP), False Positive (FP), and False Negative (FN) matches. We refer the reader to the work of Kirillov *et al.* [35] for details on the matching approach. After matching, the PQ is calculated per semantic class (cell, trap, & background) by:

$$\text{PQ} = \underbrace{\frac{\sum_{(p,g) \in \text{TP}} \text{IoU}(p,g)}{|\text{TP}|}}_{\text{Segmentation quality (SQ)}} \underbrace{\frac{|\text{TP}|}{|\text{TP}| + \frac{1}{2}|\text{FP}| + \frac{1}{2}|\text{FN}|}}_{\text{Recognition quality (RQ)}}, \quad (2)$$

where $\frac{1}{|\text{TP}|} \sum_{(p,g) \in \text{TP}} \text{IoU}(p,g)$ calculates the mean intersection-over-union of the matched mask predictions p and the label g . After computing the PQ for each semantic class, the full PQ is obtained by averaging over all classes.

The PQ is composed of the segmentation quality (SQ) and recognition quality RQ. This property allows us to separately analyze the segmentation performance and the performance of recognizing objects present in an image. We provide PyTorch [63] code for computing both metrics.

5. SAM Results

SAM (Segment Anything Model) is a recent foundation model for segmentation [6, 36]. Trained on 11M diverse images, SAM achieves remarkable zero-shot performance on new image distributions. The support for prompting [42] enables SAM to perform diverse segmentation tasks. SAM can be prompted with text, points, boxes, and dense masks. To overcome the ambiguity of segmenting an image, SAM generates multiple mask predictions alongside a score for each prediction. Note SAM only predicts instance masks and not semantic classes, limiting the application of our evaluation strategy. Thus, we provide qualitative results.

Follow-up work has evaluated SAMs zero-shot segmentation performance on biomedical imagery [12, 30, 49, 52]. While SAM achieves strong results on some biomedical tasks (*e.g.*, large-organ seg.) and even exceeds the state-of-the-art, SAM falls short to produce accurate results on other tasks (*e.g.*, pleural effusion seg.) [12, 49]. We utilize SAM to perform zero-shot segmentation of cells and traps.

We employ SAM with a ViT-H backbone [16, 40] to segment cells and traps. We prompt SAM with 256 points per side (default is 32) to segment the whole high-resolution microscopy image. To ensure no significant overlap between segmentation masks of individual objects we use a non-maximum suppression IoU threshold of 0.1 (default is 0.7). For all other parameters, we utilize the default.

Fig. 17 shows quantitative results of SAM on our TYC dataset. In simple cases, SAM yields fairly accurate predictions. In more complex cases (*e.g.*, debris) SAM fails to produce accurate segmentations. We especially observe that touching cells are often segmented as a single object. Additionally, SAM tends to segment background artifacts.

Fig. 18 presents both correct mask predictions and failure cases of SAM in more detail. If only a small number of cells are trapped in a single trap pair, SAM typically segments traps and cells correctly (*cf.* Fig. 18a). Note while we

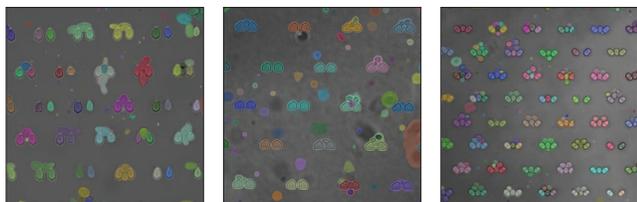
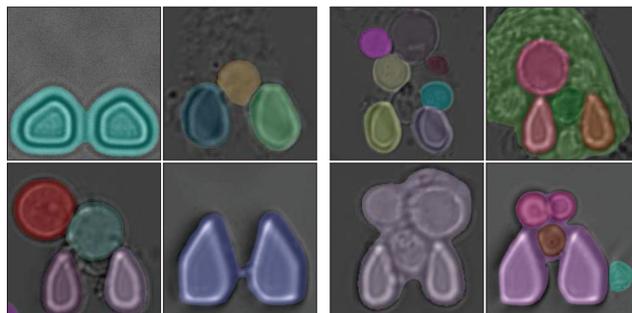


Figure 17. **SAM zero-shot results.** Qualitative segmentation results of SAM on our dataset. Colored masks indicate different instance predictions. Best viewed in color; zoom in for details.



(a) Good examples (b) Poor examples

Figure 18. **Good & poor SAM results.** (a) In some cases, SAM generates good mask predictions. (b) In other cases, SAM fails to predict accurate segmentation masks. Color coding as in Fig. 17.

define a pair of traps as a single instance, segmenting both traps separately is technically not wrong due to the ambiguity of zero-shot segmentation. If many cells are trapped in a trap pair, SAM often struggles to detect each cell instance (*cf.* Fig. 18b). If an image includes a lot of debris, SAM also struggles to predict correct segmentations.

6. Conclusion and Outlook

In this paper, we proposed a novel high-resolution dataset designed to facilitate the understanding of instance-level semantics and motions of yeast cells in microstructures. Our TYC dataset provides pixel-wise instance masks for segmenting yeast cells in microstructures — a widespread task in biological research. Alongside the labeled dataset, we also provide a large unlabeled dataset, including short microscopy video clips to facilitate the unsupervised understanding of cell motions and morphology. To ensure fair comparisons of future cell segmentation approaches, we propose a standardized strategy for evaluating segmentation performance on our dataset. Qualitative zero-shot results of the recent segmentation foundation model SAM demonstrate the complexity of our TYC dataset, as SAM often struggles to generate satisfactory mask predictions. Our effort aims to drive progress in the field of biomedical image analysis and facilitate the development of novel cell segmentation and tracking approaches, as well as make these comparable between laboratories.

While our dataset includes both image and video data, currently only image-level annotations are provided. However, some applications may require temporal labels. We are keen to extend our dataset in the future with video instance segmentation annotations [86], facilitating the development of unified segmentation and tracking approaches.

Acknowledgements. This work was supported by the Landesoffensive für wissenschaftliche Exzellenz as part of CompuGene (LOEWE). H.K. acknowledges support from the European Research Council (ERC) with the consolidator grant CONSYN (nr. 773196). C.R. acknowledges support by NEC Labs America, Inc.

References

- [1] Samreen Anjum and Danna Gurari. CTMC: Cell Tracking with Mitosis Detection Dataset Challenge. In *CVPRW*, pages 982–983, 2020. [2](#)
- [2] Mahmoud Assran, Quentin Duval, Ishan Misra, Piotr Bojanowski, Pascal Vincent, Michael Rabbat, Yann LeCun, and Nicolas Ballas. Self-Supervised Learning from Images with a Joint-Embedding Predictive Architecture. In *CVPR*, pages 15619–15629, 2023. [6](#)
- [3] Elco Bakker, Peter S Swain, and Matthew M Crane. Morphologically constrained and data informed cell segmentation of budding yeast. *Bioinformatics*, 34(1):88–96, 2018. [1](#), [2](#), [3](#), [7](#)
- [4] Tal Ben-Haim and Tammy Riklin Raviv. Graph Neural Network for Cell Tracking in Microscopy Videos. In *ECCV*, pages 610–626. Springer, 2022. [2](#)
- [5] Philip Bittihn, M Omar Din, Lev S Tsimring, and Jeff Hasty. Rational engineering of synthetic microbial systems: from single cells to consortia. *Current Opinion in Microbiology*, 45:92–99, 2018. [3](#)
- [6] Rishi Bommasani, Drew A Hudson, Ehsan Adeli, Russ Altman, Simran Arora, Sydney von Arx, Michael S Bernstein, Jeannette Bohg, Antoine Bosselut, Emma Brunskill, et al. On the Opportunities and Risks of Foundation Models. *arXiv preprint arXiv:2108.07258*, 2021. [8](#)
- [7] Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel Ziegler, Jeffrey Wu, Clemens Winter, Chris Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. Language Models are Few-Shot Learners. In *NeurIPS*, volume 33, pages 1877–1901, 2020. [1](#)
- [8] Juan C Caicedo, Allen Goodman, Kyle W Karhohs, Beth A Cimini, Jeanelle Ackerman, Marzieh Haghghi, CherKeng Heng, Tim Becker, Minh Doan, Claire McQuin, Mohammad Rohban, Shantanu Singh, and Anne E. Carpenter. Nucleus segmentation across imaging experiments: the 2018 Data Science Bowl. *Nature Methods*, 16(12):1247–1253, 2019. [1](#), [2](#)
- [9] Mathilde Caron, Hugo Touvron, Ishan Misra, Hervé Jégou, Julien Mairal, Piotr Bojanowski, and Armand Joulin. Emerging Properties in Self-Supervised Vision Transformers. In *ICCV*, pages 9650–9660, 2021. [6](#)
- [10] Yuan-Hsiang Chang, Hideo Yokota, Kuniya Abe, Chia-Tong Tang, and Ming-Dar Tasi. Automated Detection and Tracking of Cell Clusters in Time-Lapse Fluorescence Microscopy Images. *Journal of Medical and Biological Engineering*, 37:18–25, 2017. [2](#)
- [11] Ting Chen, Simon Kornblith, Kevin Swersky, Mohammad Norouzi, and Geoffrey E Hinton. Big Self-Supervised Models are Strong Semi-Supervised Learners. In *NeurIPS*, volume 33, pages 22243–22255, 2020. [6](#)
- [12] Dongjie Cheng, Ziyuan Qin, Zekun Jiang, Shaoting Zhang, Qicheng Lao, and Kang Li. SAM on Medical Images: A Comprehensive Study on Three Prompt Modes. *arXiv preprint arXiv:2305.00035*, 2023. [8](#)
- [13] Marius Cordts, Mohamed Omran, Sebastian Ramos, Timo Rehfeld, Markus Enzweiler, Rodrigo Benenson, Uwe Franke, Stefan Roth, and Bernt Schiele. The Cityscapes Dataset for Semantic Urban Scene Understanding. In *CVPR*, pages 3213–3223, 2016. [1](#)
- [14] Matthew M Crane, Ivan BN Clark, Elco Bakker, Stewart Smith, and Peter S Swain. A Microfluidic System for Studying Ageing and Dynamic Single-Cell Responses in Budding Yeast. *PLOS ONE*, 9(6):1–10, 2014. [2](#)
- [15] Douglas Decarlo and Dimitris Metaxas. Optical flow constraints on deformable models with applications to face tracking. *International Journal of Computer Vision*, 38(2):99–127, 2000. [6](#)
- [16] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. In *ICLR*, 2021. [8](#)
- [17] Bo Du, Shihan Cai, and Chen Wu. Object Tracking in Satellite Videos Based on a Multiframe Optical Flow Tracker. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 12(8):3043–3055, 2019. [6](#)
- [18] Andrea A Duina, Mary E Miller, and Jill B Keeney. Budding Yeast for Budding Geneticists: A Primer on the *Saccharomyces cerevisiae* Model System. *Genetics*, 197(1):33–48, 2014. [4](#)
- [19] Arthur Edelstein, Nenad Amodaj, Karl Hoover, Ron Vale, and Nico Stuurman. Computer Control of Microscopes using μ Manager. *Current Protocols in Molecular Biology*, 92(1):14–20, 2010. [3](#)
- [20] Christoffer Edlund, Timothy R Jackson, Nabeel Khalid, Nicola Bevan, Timothy Dale, Andreas Dengel, Sheraz Ahmed, Johan Trygg, and Rickard Sjögren. LIVECell—A large-scale dataset for label-free live cell segmentation. *Nature methods*, 18(9):1038–1045, 2021. [1](#), [2](#)
- [21] Neda Emami, Zahra Sedaei, and Reza Ferdousi. Computerized cell tracking: Current methods, tools and challenges. *Visual Informatics*, 5(1):1–13, 2021. [1](#)
- [22] Thorsten Falk, Dominic Mai, Robert Bensch, Özgün Çiçek, Ahmed Abdulkadir, Yassine Marrakchi, Anton Böhm, Jan Deubner, Zoe Jäckel, Katharina Seiwald, et al. U-Net: deep learning for cell counting, detection, and morphometry. *Nature Methods*, 16(1):67–70, 2019. [2](#)
- [23] Abolfazl Farahani, Sahar Voghoei, Khaled Rasheed, and Hamid R Arabnia. A brief review of domain adaptation. *Advances in Data Science and Information Engineering*, pages 877–894, 2021. [2](#)
- [24] Christoph Feichtenhofer, haoqi fan, Yanghao Li, and Kaiming He. Masked Autoencoders As Spatiotemporal Learners. In *NeurIPS*, volume 35, pages 35946–35958, 2022. [6](#)
- [25] Andreas Geiger, Philip Lenz, Christoph Stiller, and Raquel Urtasun. Vision meets robotics: The KITTI dataset. *The International Journal of Robotics Research*, 32(11):1231–1237, 2013. [1](#)

- [26] Kristen Grauman, Andrew Westbury, Eugene Byrne, Zachary Chavis, Antonino Furnari, Rohit Girdhar, Jackson Hamburger, Hao Jiang, Miao Liu, Xingyu Liu, et al. Ego4D: Around the World in 3,000 Hours of Egocentric Video. In *CVPR*, pages 18995–19012, 2022. 1
- [27] Alexander Grünberger, Wolfgang Wiechert, and Dietrich Kohlheyer. Single-cell microfluidics: opportunity for bio-process development. *Current Opinion in Biotechnology*, 29:15–23, 2014. 2
- [28] Dongmin Guo, Anne L Van de Ven, and Xiaobo Zhou. Red Blood Cell Tracking Using Optical Flow Methods. *IEEE Journal of Biomedical and Health Informatics*, 18(3):991–998, 2013. 6
- [29] Kaiming He, Xinlei Chen, Saining Xie, Yanghao Li, Piotr Dollár, and Ross Girshick. Masked Autoencoders Are Scalable Vision Learners. In *CVPR*, pages 16000–16009, 2022. 6
- [30] Yuhao Huang, Xin Yang, Lian Liu, Han Zhou, Ao Chang, Xinrui Zhou, Rusi Chen, Junxuan Yu, Jiongquan Chen, Chaoyu Chen, et al. Segment Anything Model for Medical Images? *arXiv preprint arXiv:2304.14660*, 2023. 8
- [31] Junbong Jang, Kwonmoo Lee, and Tae-Kyun Kim. Unsupervised Contour Tracking of Live Cells by Mechanical and Cycle Consistency Losses. In *CVPR*, pages 227–236, 2023. 2
- [32] Hannah Jeckel and Knut Drescher. Advances and opportunities in image analysis of bacterial cells and communities. *FEMS Microbiology Reviews*, 45(4):fuaa062, 2021. 1
- [33] Myeong Chan Jo, Wei Liu, Liang Gu, Weiwei Dang, and Lidong Qin. High-throughput analysis of yeast replicative aging using a microfluidic system. *Proceedings of the National Academy of Sciences*, 112(30):9364–9369, 2015. 2
- [34] Rico Jonschkowski, Austin Stone, Jonathan T Barron, Ariel Gordon, Kurt Konolige, and Anelia Angelova. What Matters in Unsupervised Optical Flow. In *ECCV*, pages 557–572, 2020. 6
- [35] Alexander Kirillov, Kaiming He, Ross Girshick, Carsten Rother, and Piotr Dollár. Panoptic Segmentation. In *CVPR*, pages 9404–9413, 2019. 4, 7
- [36] Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao, Spencer Whitehead, Alexander C Berg, Wan-Yen Lo, et al. Segment Anything. *arXiv preprint arXiv:2304.02643*, 2023. 1, 2, 8
- [37] Herbert T Kruitbosch, Yasmin Mzayek, Sara Omlor, Paolo Guerra, and Andreas Miliadis-Argeitis. A convolutional neural network for segmentation of yeast cells without manual training annotations. *Bioinformatics*, 38(5):1427–1433, 2022. 2
- [38] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep Learning. *Nature*, 521(7553):436–444, 2015. 1
- [39] Markus Leygeber, Dorina Lindemann, Christian Carsten Sachs, Eugen Kaganovitch, Wolfgang Wiechert, Katharina Nöh, and Dietrich Kohlheyer. Analyzing Microbial Population Heterogeneity—Expanding the Toolbox of Microfluidic Single-Cell Cultivations. *Journal of Molecular Biology*, 431(23):4569–4588, 2019. 1, 3
- [40] Yanghao Li, Hanzi Mao, Ross Girshick, and Kaiming He. Exploring Plain Vision Transformer Backbones for Object Detection. In *ECCV*, pages 280–296, 2022. 8
- [41] Kun Liu, Soeren S Lienkamp, Asako Shindo, John B Wallingford, Gerd Walz, and Olaf Ronneberger. Optical flow guided cell segmentation and tracking in developing tissue. In *ISBI*, pages 298–301, 2014. 6
- [42] Pengfei Liu, Weizhe Yuan, Jinlan Fu, Zhengbao Jiang, Hiroaki Hayashi, and Graham Neubig. Pre-train, Prompt, and Predict: A Systematic Survey of Prompting Methods in Natural Language Processing. *ACM Computing Surveys*, 55(9):1–35, 2023. 8
- [43] James CW Locke and Michael B Elowitz. Using Movies to Analyse Gene Circuit Dynamics in Single Cells. *Nature Reviews Microbiology*, 7(5):383–392, 2009. 3
- [44] Katharina Löffler and Ralf Mikut. Embed-Track—Simultaneous Cell Segmentation and Tracking Through Learning Offsets and Clustering Bandwidths. *IEEE Access*, 10:77147–77157, 2022. 2
- [45] Diane Longo and Jeff Hasty. Dynamics of single-cell gene expression. *Molecular Systems Biology*, 2(1):64, 2006. 3
- [46] Jean-Baptiste Lugagne, Haonan Lin, and Mary J Dunlop. DeLTA: Automated cell segmentation, tracking, and lineage reconstruction using deep learning. *PLoS Computational Biology*, 16(4):e1007673, 2020. 2
- [47] Tao Luo, Lei Fan, Rong Zhu, and Dong Sun. Microfluidic Single-Cell Manipulation and Analysis: Methods and Applications. *Micromachines*, 10(2):104, 2019. 2
- [48] Wenhan Luo, Junliang Xing, Anton Milan, Xiaoqin Zhang, Wei Liu, and Tae-Kyun Kim. Multiple Object Tracking: A Literature Review. *Artificial Intelligence*, 293:103448, 2021. 6
- [49] Jun Ma and Bo Wang. Segment anything in medical images. *arXiv preprint arXiv:2304.12306*, 2023. 8
- [50] Rémi Marsal, Florian Chabot, Angélique Loesch, and Hichem Sahbi. BrightFlow: Brightness-Change-Aware Unsupervised Learning of Optical Flow. In *WACV*, pages 2061–2070, 2023. 6
- [51] Martin Maška, Vladimír Ulman, Pablo Delgado-Rodríguez, Estibaliz Gómez-de Mariscal, Tereza Nečasová, Fidel A Guerrero Peña, Tsang Ing Ren, Elliot M Meyerowitz, Tim Scherr, Katharina Löffler, et al. The Cell Tracking Challenge: 10 years of objective benchmarking. *Nature Methods*, pages 1–11, 2023. 1, 2
- [52] Maciej A Mazurowski, Haoyu Dong, Hanxue Gu, Jichen Yang, Nicholas Konz, and Yixin Zhang. Segment Anything Model for Medical Image Analysis: an Experimental Study. *arXiv preprint arXiv:2304.10517*, 2023. 8
- [53] Erik Meijering. Cell Segmentation: 50 Years Down the Road. *IEEE Signal Processing Magazine*, 29(5):140–145, 2012. 1, 2
- [54] Erik Meijering, Anne E Carpenter, Hanchuan Peng, Fred A Hamprecht, and Jean-Christophe Olivo-Marin. Imagining the future of bioimage analysis. *Nature Biotechnology*, 34(12):1250–1255, 2016. 2
- [55] Simon Meister, Junhwa Hur, and Stefan Roth. UnFlow: Unsupervised Learning of Optical Flow with a Bidirectional Census Loss. In *AAAI*, volume 32, page 7251–7259, 2018. 6

- [56] Fatima Merchant and Kenneth Castleman. *Microscope Image Processing*. Academic Press, 2022. 2, 4
- [57] Erick Moen, Dylan Bannon, Takamasa Kudo, William Graf, Markus Covert, and David Van Valen. Deep learning for cellular image analysis. *Nature Methods*, 16(12):1233–1246, 2019. 2
- [58] Vigneswaran Narayanamurthy, Sairam Nagarajan, Fahmi Samsuri, TM Sridhar, et al. Microfluidic hydrodynamic trapping for single cell analysis: mechanisms, methods and applications. *Analytical Methods*, 9(25):3751–3772, 2017. 2
- [59] Sergey I Nikolenko. *Synthetic Data for Deep Learning*, volume 174. Springer, 2021. 2
- [60] Marius Pachitariu and Carsen Stringer. Cellpose 2.0: how to train your own model. *Nature Methods*, 19(12):1634–1641, 2022. 2
- [61] Long Pang, Jing Ding, Xi-Xian Liu, Haoyue Yuan, Yuxin Ge, Jianglin Fan, and Shih-Kang Fan. Microstructure-based techniques for single-cell manipulation and analysis. *TrAC Trends in Analytical Chemistry*, 129:115940, 2020. 2
- [62] Sapun Parekh and Schwendy Mischa. EVICAN Dataset, 2019. 1
- [63] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Kopf, Edward Yang, Zachary DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. PyTorch: An Imperative Style, High-Performance Deep Learning Library. In *NeurIPS*, volume 32, 2019. 8
- [64] Tim Prangemeier, François-Xavier Lehr, Rogier M Schoeman, and Heinz Koepl. Microfluidic platforms for the dynamic characterisation of synthetic circuitry. *Current Opinion in Biotechnology*, 63:167–176, 2020. 2, 3, 4
- [65] Tim Prangemeier, Christoph Reich, and Heinz Koepl. Attention-Based Transformers for Instance Segmentation of Cells in Microstructures. In *BIBM*, pages 700–707, 2020. 1, 7
- [66] Tim Prangemeier, Christian Wildner, André O Françani, Christoph Reich, and Heinz Koepl. Yeast cell segmentation in microstructured environments with deep learning. *Biosystems*, 211:104557, 2022. 1, 2
- [67] Christoph Reich, Tim Prangemeier, André O Françani, and Heinz Koepl. An Instance Segmentation Dataset of Yeast Cells in Microstructures. In *EMBC*, 2023. 2, 4, 5, 6, 7
- [68] Zhe Ren, Junchi Yan, Bingbing Ni, Bin Liu, Xiaokang Yang, and Hongyuan Zha. Unsupervised Deep Learning for Optical Flow Estimation. In *AAAI*, volume 31, pages 1495–1501, 2017. 6
- [69] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-Net: Convolutional Networks for Biomedical Image Segmentation. In *MICCAI*, pages 234–241, 2015. 2
- [70] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, et al. ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision*, 115(3):211–252, 2015. 1
- [71] Christian Carsten Sachs, Karina Ruzaeva, Johannes Seifarth, Wolfgang Wiechert, Benjamin Berkels, and Katharina Nöh. CellSium: versatile cell simulator for microcolony ground truth generation. *Bioinformatics advances*, 2(1):vbac053, 2022. 2
- [72] Tim Scherr, Katharina Löffler, Moritz Böhlend, and Ralf Mikut. Cell segmentation and tracking using CNN-based distance predictions and a graph-based matching strategy. *PLoS One*, 15(12):e0243219, 2020. 2
- [73] Uwe Schmidt, Martin Weigert, Coleman Broaddus, and Gene Myers. Cell Detection with Star-Convex Polygons. In *MICCAI*, pages 265–273, 2018. 2
- [74] Christoph Schuhmann, Romain Beaumont, Richard Vencu, Cade Gordon, Ross Wightman, Mehdi Cherti, Theo Coombes, Aarush Katta, Clayton Mullis, Mitchell Wortsman, Patrick Schramowski, Srivatsa Kundurthy, Katherine Crowson, Ludwig Schmidt, Robert Kaczmarczyk, and Jenia Jitsev. LAION-5B: An open large-scale dataset for training next generation image-text models. In *NeurIPS*, volume 35, pages 25278–25294, 2022. 1
- [75] Wen Si, Cheng Li, and Ping Wei. Synthetic immunology: T-cell engineering and adoptive immunotherapy. *Synthetic and Systems Biotechnology*, 3(3):179–185, 2018. 3
- [76] Nidhi Sinha, Haowen Yang, David Janse, Luc Hendriks, Ulfert Rand, Hansjörg Hauser, Mario Köster, Frans N van de Vosse, Tom FA de Greef, and Jurjen Tel. Microfluidic chip for precise trapping of single cells and temporal analysis of signaling dynamics. *Communications Engineering*, 1(1):18, 2022. 2
- [77] Austin Stone, Daniel Maurer, Alper Ayvaci, Anelia Angelova, and Rico Jonschkowski. SMURF: Self-Teaching Multi-Frame Unsupervised RAFT with Full-Image Warping. In *CVPR*, pages 3887–3896, 2021. 6
- [78] Carsen Stringer, Tim Wang, Michalis Michaelos, and Marius Pachitariu. Cellpose: a generalist algorithm for cellular segmentation. *Nature Methods*, 18(1):100–106, 2021. 2
- [79] Jing Sun, Attila Tárnok, and Xuantao Su. Deep Learning-Based Single-Cell Optical Image Studies. *Cytometry Part A*, 97(3):226–240, 2020. 1
- [80] Vladimír Ulman, Martin Maška, Klas EG Magnusson, Olaf Ronneberger, Carsten Haubold, Nathalie Harder, Pavel Matula, Petr Matula, David Svoboda, Miroslav Radojevic, et al. An objective comparison of cell-tracking algorithms. *Nature Methods*, 14(12):1141–1152, 2017. 1, 2
- [81] Athanasios Voulodimos, Nikolaos Doulamis, Anastasios Doulamis, Eftychios Protopapadakis, et al. Deep Learning for Computer Vision: A Brief Review. *Computational Intelligence and Neuroscience*, 2018, 2018. 1
- [82] Mei Wang and Weihong Deng. Deep Visual Domain Adaptation: A Survey. *Neurocomputing*, 312:135–153, 2018. 2
- [83] Yang Wang, Yi Yang, Zhenheng Yang, Liang Zhao, Peng Wang, and Wei Xu. Occlusion Aware Unsupervised Learning of Optical Flow. In *CVPR*, pages 4884–4893, 2018. 6
- [84] Mark Weber, Jun Xie, Maxwell D Collins, Yukun Zhu, Paul Voigtlaender, Hartwig Adam, Bradley Green, Andreas Geiger, Bastian Leibe, Daniel Cremers, et al. STEP: Segmenting and Tracking Every Pixel. In *NeurIPS Datasets and Benchmarks Track*, 2021. 6

- [85] Zhen Xie, Liliana Wroblewska, Laura Prochazka, Ron Weiss, and Yaakov Benenson. Multi-Input RNAi-Based Logic Circuit for Identification of Specific Cancer Cells. *Science*, 333:1307–1312, 2011. [3](#)
- [86] Linjie Yang, Yuchen Fan, and Ning Xu. Video Instance Segmentation. In *ICCV*, pages 5188–5197, 2019. [8](#)
- [87] Xiaohua Zhai, Alexander Kolesnikov, Neil Houlsby, and Lucas Beyer. Scaling Vision Transformers. In *CVPR*, pages 12104–12113, 2022. [1](#)
- [88] Jimuyang Zhang, Sanping Zhou, Xin Chang, Fangbin Wan, Jinjun Wang, Yang Wu, and Dong Huang. Multiple Object Tracking by Flowing and Fusing. *arXiv preprint arXiv:2001.11180*, 2020. [6](#)