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

TECHNISCHE UNIVERSITÄT DARMSTADT

Christoph Reich, Tim Prangemeier, and Heinz Koeppl
Department of Electrical Engineering & Information Technology, Centre for Synthetic Biology, Technische Universität Darmstadt

Introduction & Motivation

- Many biomedical applications require the segmentation of single cells in microscopy imagery

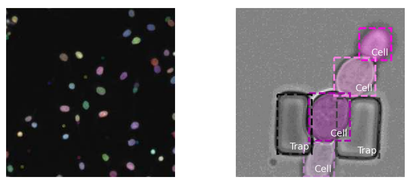


Fig. 1. Existing datasets (DSB2018 [1] left & Reich *et al.* [2] right).

Existing datasets only consider limited scenes (*e.g.*, no microstructures) or only offer limited annotations (*c.f.* Fig. 1)

Dataset Overview

Dataset	# ann. pix. [10^7]	# cells	# traps	Resolution	Annotation type	# trap types
Bakker *et al.* [3]	–	1000	0	512×512	cell outline	1
Reich *et al.* [2]	0.81	914	971	128×128	instance seg.	2
TYC dataset (ours)	46.39	14541	4405	≥2048×2048	instance seg.	6

Tab. 1. Dataset comparison of existing datasets including cells and traps *vs.* our TYC dataset.

We present the largest labeled dataset to include cells & microstructures

Data Acquisition

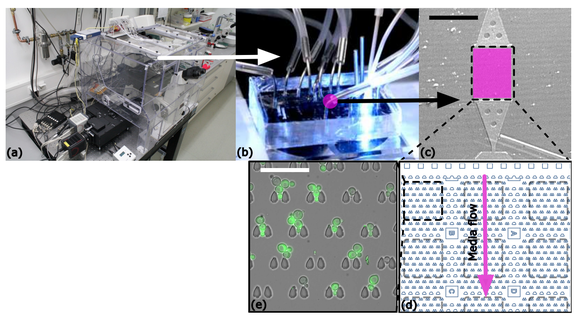


Fig. 2. Overview data acquisition process.

We utilize data from more than a dozen experiments

TYC Labeled Set

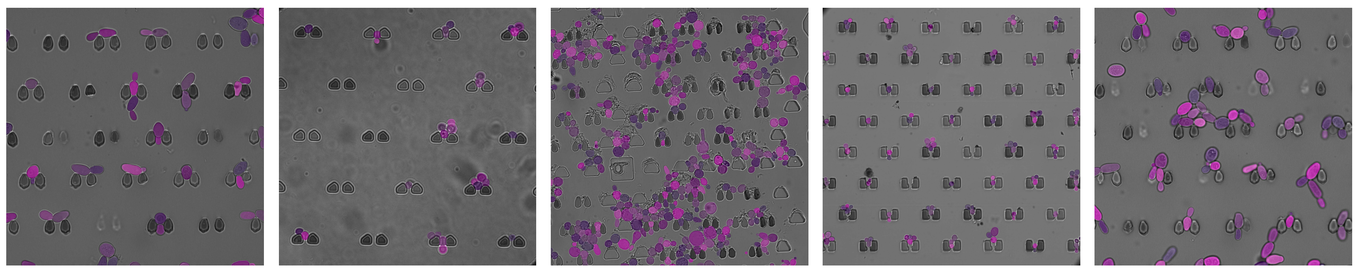


Fig. 3. Labeled sample of our TYC dataset; traps in grey (■■■) and cells in pink (■■■).

We release ∼19k instance masks for both cells and microstructures

TYC Unlabeled Set

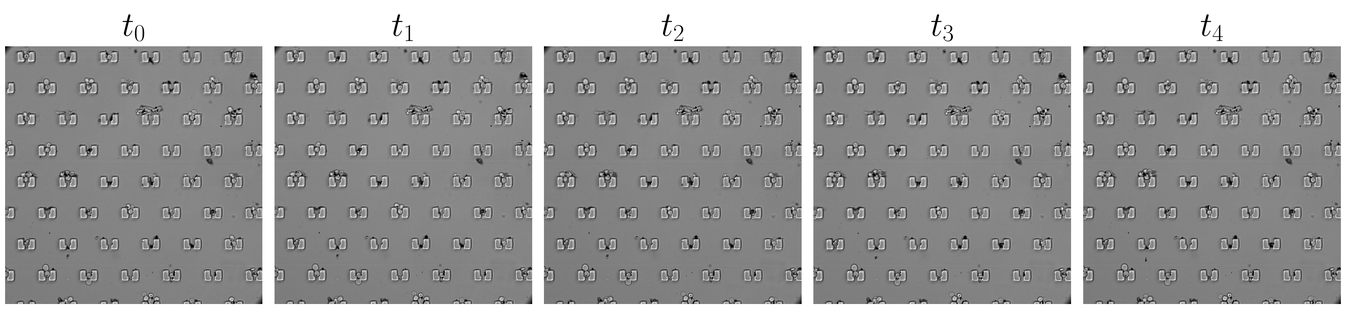


Fig. 4. Unlabeled brightfield microscopy image sequence.

We present 61 curated video clips to facilitate unsupervised learning

Evaluation

Semantic-level evaluation (cell class Intersection-over-Union)

$$\mathrm{IoU} = \frac{|p_c \cap g_c|}{|p_c \cup g_c|},$$

Instance-level evaluation (Panoptic Quality [4])

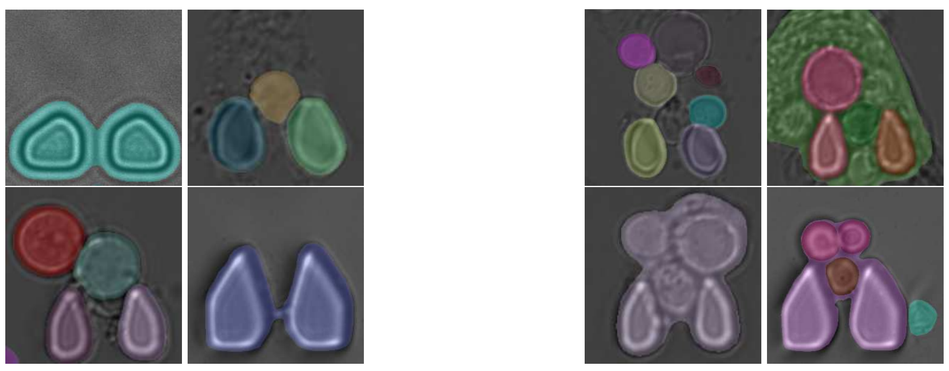
$$\mathrm{PQ} = \underbrace{\frac{\sum_{(p,g)\in \mathrm{TP}} \mathrm{IoU}(p,g)}{|\mathrm{TP}|}}_{\text{Segmentation quality (SQ)}} \underbrace{\frac{|\mathrm{TP}|}{|\mathrm{TP}| + \frac{1}{2}|\mathrm{FP}| + \frac{1}{2}|\mathrm{FN}|}}_{\text{Recognition quality (RQ)}},$$

Dataset Split

Split	# images	# ann. pix. [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

Tab. 2. Split of our labeled set.

Segment Anything Results



(a) Good examples **(b)** Poor examples

Fig. 5. Zero-shot results of SAM [5] on our TYC dataset.

SAM fails to produce accurate segmentations out of the box

References

[1] J. C. Caicedo, A. Goodman *et al.*, "Nucleus segmentation across imaging experiments: the 2018 Data Science Bowl," *Nature Methods*, vol. 16, no. 12, pp. 1247–1253, 2019.
[2] C. Reich, T. Prangemeier, A. O. Françani, and H. Koeppl, "An Instance Segmentation Dataset of Yeast Cells in Microstructures," in *EMBC*, 2023.
[3] E. Bakker, P. S. Swain, and M. M. Crane, "Morphologically constrained and data informed cell segmentation of budding yeast," *Bioinformatics*, vol. 34, no. 1, pp. 88–96, 2018.
[4] A. Kirillov, K. He, R. Girshick, C. Rother, and P. Dollár, "Panoptic Segmentation," in *CVPR*, 2019, pp. 9404–9413.
[5] A. Kirillov, E. Mintun, N. Ravi, H. Mao, C. Rolland, L. Gustafson, T. Xiao, S. Whitehead, A. C. Berg, W.-Y. Lo *et al.*, "Segment Anything," *arXiv preprint arXiv:2304.02643*, 2023.

Paper: `https://arxiv.org/abs/2308.12116`
Dataset: `https://tudatalib.ulb.tu-darmstadt.de/handle/tudatalib/3930`
Project Page: `https://christophreich1996.github.io/tyc_dataset/`
Code: `https://github.com/ChristophReich1996/TYC-Dataset`
Correspondence: `heinz.koeppl@bcs.tu-darmstadt.de`

Acknowledgement. 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.

ICCV23 PARIS