SAT: Segment and Track Anything for Microscopy

Nabeel Khalid[1,5][a], Mohammadmahdi Koochali[1][b], Khola Naseem[1,5][c], Maria Caroprese[6][d], Gillian Lovell[3][e], Daniel A Porto[7][f], Johan Trygg[2,4][g], Andreas Dengel[1,5][h], and Sheraz Ahmed[1][i]

[1]*German Research Center for Artificial Intelligence (DFKI) GmbH, 67663 Kaiserslautern, Germany*

[2]*Sartorius Corporate Research, Umeå, Sweden*

[3]*Sartorius BioAnalytics, Royston, United Kingdom*

[4]*Computational Life Science Cluster (CLiC), Umeå University, Umeå, Sweden*

[5]*RPTU Kaiserslautern–Landau, 67663 Kaiserslautern, Germany*

[6]*Sartorius Digital Solutions, Royston, United Kingdom*

[7]*Sartorius BioAnalytics, Ann Arbor, United States*

{*nabeel.khalid, mohammadmahdi.koochali, andreas.dengel, sheraz.ahmed*}*@dfki.de*

,{maria.caroprese, gillian.lovell, daniel.porto, johan.trygg}@sartorius.com

Keywords: Biomedical, Healthcare, Deep learning, Cell Segmentation, Cell tracking, Segment Anything, Track Anything, Microscopy

Abstract: Integrating cell segmentation with tracking is essential for a detailed and dynamic understanding of cellular behavior. This combination enhances the study and quantification of cell morphology, movement, and interactions, providing valuable insights into various biological processes and diseases. Traditional methods require full masks or bounding boxes for each cell, which is labor-intensive and expensive. To address this challenge, SAT: Segment and Track Anything for Microscopy is presented. This method requires only a few point annotations per cell in the first frame of each sequence and then automatically performs cell segmentation and tracking for all subsequent frames. This approach reduces annotation time and effort, making it practical for large-scale studies. The method was evaluated on two diverse datasets, achieving over 80% Multiple Object Tracking Accuracy (MOTA), demonstrating its robustness and effectiveness in various cell tracking scenarios.

1 INTRODUCTION

Cell tracking is essential in biology and medicine, offering insights into cellular behavior and responses to stimuli (Newman et al., 2011). In cancer research, cell tracking aids in studying tumor growth, metastasis, and the efficacy of anti-cancer drugs, while in stem cell research, it helps observe differentiation and regenerative potential (Aramini et al., 2022). This technique is vital in drug development for assessing drug impact and efficacy and in immunology for understanding immune cell interactions and responses (Yazdi and Khotanlou, 2024). Accurate cell segmentation is essential for tracking, providing data to monitor cell movement and behavior over time (Chou et al., 2023). Without precise segmentation, tracking algorithms may misidentify cells, leading to errors. The importance of cell segmentation lies in its ability to quantify cell morphology, analyze cellular interactions, and support high-throughput screening in drug development (Durkee et al., 2021). Additionally, it aids in understanding developmental processes and immune responses by characterizing specific cell populations (Padovani et al., 2022).

In the past decade, many deep-learning-based approaches (Khalid et al., 2021a; Stringer et al., 2020; Edlund et al., 2021; Khalid et al., 2022a; Schwendy et al., 2020; Khalid et al., 2021b) have been developed for cell segmentation. However, these approaches require fully labeled datasets for training, where each cell boundary is delineated by experts, which is both

[a] https://orcid.org/0000-0001-9274-3757
[b] https://orcid.org/0000-0001-8780-253X
[c] https://orcid.org/0000-0003-4785-2588
[d] https://orcid.org/0009-0009-2170-1459
[e] https://orcid.org/0009-0004-5180-9704
[f] https://orcid.org/0000-0002-1021-2467
[g] https://orcid.org/0000-0003-3799-6094
[h] https://orcid.org/0000-0002-6100-8255
[i] https://orcid.org/0000-0002-4239-6520

Table 1: Comparison of supervision time between Full Mask and Tracking, and the SAT method. The SAT method involves point annotation only in the first frame. SAT (5 points per cell) saves significant time compared to full mask tracking, making it approximately 206 times faster.

Method	Time per Cell per Frame (s)		Total Time (A+T) for 100 Frames (min)	Times Faster than Full Mask (x)
	Segmentation/Point Annotation (A)	Tracking (T)		
Full Mask and Tracking	46	0.438	77.40	-
SAT (N=3)	$3 \times 0.9 = 2.7$	0	0.225	≈ 344
SAT (N=5)	$5 \times 0.9 = 4.5$	0	0.375	≈ 206
SAT (N=10)	$10 \times 0.9 = 9$	0	0.75	≈ 103

time-consuming and expensive. On the LIVECell dataset (Edlund et al., 2021), the largest publicly available dataset in the cell segmentation domain, it takes approximately 46 seconds to draw a segmentation mask for each cell. Given that the dataset comprises over 1.6 million cells, this process is extremely time-consuming and costly. Additionally, when cell tracking is added, annotators face further challenges. In simple scenarios, tracking each cell across 100 frames takes around 43.8 seconds. However, in more complex scenarios, this can take up to 98 seconds per cell (Phasefocus, nd). To address these significant time and resource demands, this study introduces an automated pipeline called SAT (Segment and Track Anything) for microscopy, which is specifically designed for cell segmentation and tracking in sequences of microscopic images. The proposed method uses point annotations in the first frame, then automatically segments and tracks cells across the sequence. This approach significantly reduces annotation time and minimizes the need for expert knowledge. By removing the need for detailed masks and manual tracking, this method streamlines the workflow. The reliance on simple point annotations makes it more accessible for researchers with varying levels of expertise, thereby democratizing the use of advanced cell tracking and segmentation techniques in the field of microscopy. This efficiency enables faster experimental turnaround and large-scale studies, reducing labor costs. Table 1 compares the supervision time required for cell segmentation and tracking using the traditional Full Mask and Tracking method versus the SAT method with different numbers of points per cell. The Full Mask and Tracking method takes 46 seconds per cell per frame for segmentation and 0.438 seconds for tracking, totaling 77.40 minutes for 100 frames. In contrast, the SAT method reduces initial annotation time by using point annotations in the first frame. For example, SAT (N=5) takes 4.5 seconds per cell per frame, totaling 0.375 minutes over 100 frames, making it about 206 times faster than the Full Mask and Tracking method. SAT (N=3) and SAT (N=10) are 344 and 103 times faster, respectively, demonstrating the efficiency of SAT compared to traditional full segmentation methods. The SAT pipeline mitigates this issue by requiring only point annotations per cell in the first frame to perform segmentation and tracking automatically for the entire sequence. This makes the SAT method highly practical for large-scale datasets, where traditional manual annotation would be extremely time-consuming and expensive. The main contributions of this study are as follows:

- Introduction of the SAT (Segment and Track Anything for Microscopy) pipeline, which leverages point annotations in the first frame to automate cell segmentation and tracking, significantly reducing the time and effort required compared to traditional methods.
- Comprehensive evaluation of the SAT pipeline on subsets of two extensive and diverse cell tracking datasets: the Cell Tracking Challenge (CTC) (Maška et al., 2023) and the Cell Tracking with Mitosis Detection Challenge (CTMC) (Anjum and Gurari, 2020) datasets, demonstrating the method's robustness and generalization capability.
- Achieving high tracking accuracy, with Multiple Object Tracking Accuracy (MOTA) exceeding 80%, and demonstrating time savings of over 100 times compared to full mask annotation methods.

2 LITERATURE REVIEW

2.1 Existing Cell Segmentation and Tracking Approaches

There are numerous studies on cell segmentation and tracking (Edlund et al., 2021; Stringer et al., 2020; Jelli et al., 2023; Khalid et al., 2023; Maška et al., 2023) which require full masks for training or some

form of weak supervision. Khalid et al. (Khalid et al., 2022b) introduced a cell segmentation method using only bounding boxes and point annotations, reducing annotation time and resources, though manual input is still required. Segmentation-first approaches focus on segmentation before linking detections across frames (Malin-Mayor et al., 2023), and unsupervised tracking methods aim to reduce reliance on labeled data (Maška et al., 2023).

2.2 Challenges in Microscopy Applications

Microscopy images often have noisy, low-contrast environments, making it difficult for models trained on natural image datasets to generalize effectively (Wang et al., 2023). Traditional segmentation methods struggle to handle different imaging modalities, such as phase contrast or fluorescence microscopy, which vary in contrast and clarity (Stringer et al., 2020). Additionally, variations in cell shapes and appearances require extensive retraining, limiting the scalability of these models across microscopy domains (Yazdi and Khotanlou, 2024).

2.3 Limitations of the Segment Anything Model (SAM)

The Segment Anything Model (SAM) (Kirillov et al., 2023) by MetaAI performs exceptionally well on natural scenes but struggles with microscopic images due to their complexity, low contrast, and noise (Archit et al., 2023). Domain-specific training and advanced pre-processing are needed to enhance SAM's applicability in microscopy.

Based on these limitations, the proposed work seeks to overcome the challenges faced by traditional methods. By utilizing point annotations in the first frame of microscopy sequences, the approach substantially reduces the need for fully labeled datasets and automates the segmentation and tracking of cells throughout all frames.

3 SAT: SEGMENT AND TRACK ANYTHING PIPELINE

The key technical component of the proposed pipeline is the Segment Anything Model (SAM). SAM was pre-trained on diverse images and fine-tuned using the LIVECell dataset for microscopy. It includes an image encoder, prompt encoder, and mask decoder, which work together to produce accurate segmentation masks from point prompts. The SAT (Segment and Track Anything for Microscopy) pipeline is divided into four main components: Query Points Selection, Point Tracking, Segmentation, and Point Tracking Reinitialization (Rajič et al., 2023). Below is a detailed explanation of each module, referring to Figure 3.

3.1 Query Points Selection

In the first step of SAT, query points are selected in the first video frame to denote the target object (positive points) and non-target regions (negative points). The user can provide these points interactively or derive them from a ground truth mask using various sampling techniques, including Random Sampling, K-Medoids Sampling, Shi-Tomasi Sampling (Shi et al., 1994), and Mixed Sampling. Each method ensures good coverage and robustness, significantly affecting the model's performance.

$$P = \{p_1, p_2, \ldots, p_n\} \tag{1}$$

For K-Medoids, let C represent the set of clusters, and P_m be the medoid points:

$$P_m = \{\text{medoid}(c_i) \mid c_i \in C\} \tag{2}$$

The objective function to minimize K-Medoids clustering is:

$$\text{minimize} \sum_{i=1}^{k} \sum_{p \in c_i} \|p - \text{medoid}(c_i)\| \tag{3}$$

3.2 Point Tracking

This module propagates the selected query points across all video frames using point trackers. This propagation generates point trajectories and occlusion scores, ensuring that the points follow the objects throughout the video. Point tracker, PIPS (Harley et al., 2022) is employed due to its robustness in handling long-term tracking challenges such as occlusion and reappearance of objects.

$$P_t = \{p_{t,1}, p_{t,2}, \ldots, p_{t,n}\} \tag{4}$$

The tracking function T predicts the position of points in the next frame:

$$P_{t+1} = T(P_t) \tag{5}$$

3.3 Segmentation

Using the point trajectories obtained from the tracking module, the Segment Anything Model (SAM)

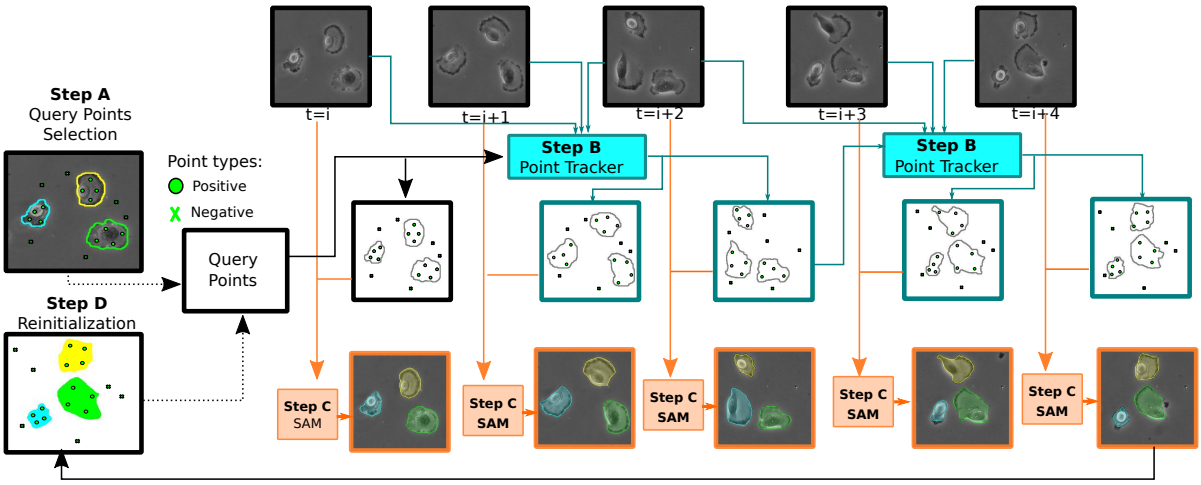

Figure 1: SAT (Segment and Track Anything for Microscopy) Pipeline. The SAT pipeline extends image segmentation models to microscopy videos through four steps: **A. Query Points Selection**, where positive and negative points are defined by the user or a ground truth mask; **B. Point Tracking**, which propagates points across video frames using point trackers, predicting trajectories and occlusion scores; **C. Segmentation**, where the Segment Anything Model (SAM) uses these trajectories to generate per-frame mask predictions; and **D. Point Tracking Reinitialization**, an optional step to reinitialize query points, improving tracking reliability and addressing newly visible cell segments.

(Kirillov et al., 2023), which is finetuned on the LIVECell dataset, generates per-frame segmentation masks. The SAM model, which comprises an image encoder, a prompt encoder, and a mask decoder, utilizes the non-occluded points as prompts to segment the object of interest in each frame accurately.

$$M_t = \text{SAM}(I_t, P_t) \tag{6}$$

where I_t is the input image at frame t, and P_t is the set of propagated points.

3.4 Point Tracking Reinitialization

This step involves reinitializing the query points periodically using the predicted masks. Reinitialization helps to remove unreliable points and add new points to object segments that become visible in later frames, thereby improving the accuracy and robustness of the segmentation over time.

$$P_t = \text{Reinitialize}(M_t) \tag{7}$$

Reinitialization occurs at intervals defined by the prediction horizon h:

$$P_{t+h} = \text{Reinitialize}(M_{t+h}) \tag{8}$$

While SAT integrates existing modules like SAM for segmentation and PIPS for point tracking, its novelty lies in optimizing these components specifically for microscopy images. By minimizing the manual effort with point annotations in the first frame and automating the rest of the segmentation and tracking process, SAT improves both accuracy and efficiency. Additionally, the reinitialization step enhances robustness in tracking, and addressing occlusions and the appearance of new cells over time.

Table 2: Statistics of the LIVECell dataset used for fine-tuning the Segment Anything Model.

Dataset	Train		Val		Test	
	Img	Cells	Img	Cells	Img	Cells
LIVECell	3253	1,018,576	570	181,609	1564	462,261

4 DATASET

For fine-tuning the Segment Anything Model (SAM) (Kirillov et al., 2023), the LIVECell dataset (Edlund et al., 2021) (Table 2) was exclusively used. LIVECell is a comprehensive dataset with label-free live-cell images and detailed annotations, making it ideal for refining SAM's segmentation capabilities. Leveraging LIVECell for fine-tuning enhances the model's performance and applicability to real-world microscopy images by providing high-quality annotations and diverse cell types. This approach equips the model to handle unique challenges posed by microscopic images, such as low contrast, high noise, various modalities, and complex cell structures.

Two datasets are used to evaluate the generalization of the proposed methodology for cell segmentation and tracking. The first is the Cell Tracking Challenge (CTC) dataset (Maška et al., 2023), which includes 2D and 3D time-lapse sequences of various mi-

croscopy videos, including Bright Field, Phase Contrast, and Differential Interference Contrast (DIC). It contains 20 sequences, 10 of which are 2D, with a total of 8,017 frames and an average cell density of 33.12 cells per image. The second is the Cell Tracking with Mitosis Detection Challenge (CTMC) dataset (Anjum and Gurari, 2020), comprising over 1.5 million images across 86 videos of 14 different cell lines, annotated with bounding boxes. Unlike CTC, CTMC does not provide segmentation masks, posing additional challenges for segmentation. To evaluate the method across diverse conditions, 4 sequences from the CTC dataset and 6 from the CTMC dataset were randomly selected. These subsets represent various imaging modalities, cell types, and capture intervals, providing a robust basis for testing the method's generalization.

5 EVALUATION METRICS

To assess the performance of the proposed pipeline for cell tracking, five distinct evaluation metrics are used, each providing a unique perspective on the results.

5.1 Multiple Object Tracking Accuracy

Multiple Object Tracking Accuracy (MOTA) (Bernardin and Stiefelhagen, 2008) measures the overall accuracy of the tracker and the detection.

$$\text{MOTA} = 1 - \frac{\sum_t (m_t + fp_t + mme_t)}{\sum_t g_t}$$

Here, m_t represents the total number of misses, fp_t the total number of false positives, and mme_t the total number of mismatches. Misses occur when a cell in the ground truth is not detected. False positives occur when a cell is detected but not present in the ground truth. Mismatches occur when a cell is incorrectly related to another cell.

5.2 Identification F1 Score

Identification F1 (IDF1) (Ristani et al., 2016) calculates a one-to-one mapping between ground truth trajectories and prediction trajectories.

$$\text{IDF1} = \frac{2 \cdot \text{IDTP}}{2 \cdot \text{IDTP} + \text{IDFP} + \text{IDFN}}$$

IDTP (Identity True Positives) is the number of true positive ID matches. IDFP (Identity False Positives) denotes the false positive IDs. IDFN (Identity False Negatives) is the number of false negative IDs.

5.3 Identity Switches

Identity Switches (IDs) (Bernardin and Stiefelhagen, 2008), also known as Mismatches, refer to the number of times a trajectory incorrectly changes from one ground truth object to another. A lower number of identity switches indicates a more reliable tracking system.

5.4 Mostly Tracked

If an object is successfully tracked for at least 80% of its lifespan, it is considered Mostly Tracked (MT) (Leal-Taixé et al., 2015). This metric evaluates the robustness of a tracking algorithm in maintaining the continuity of an object's identity.

5.5 Mostly Lost

If an object is tracked for 20% or less of its lifespan, it is considered Mostly Lost (ML) (Leal-Taixé et al., 2015). High ML scores suggest the tracker struggles with challenges, leading to frequent identity losses. This metric helps identify and address the limitations of tracking algorithms.

6 EXPERIMENTAL SETUP

Two different experimental settings are designed to evaluate the performance of the proposed pipeline for cell tracking from various aspects. The first setting, namely SAT Evaluation on Diverse Modalities and Intervals Using the CTC Dataset, assesses the performance of the SAT pipeline across various imaging modalities and time intervals, using annotated 2D and 3D sequences from the Cell Tracking Challenge dataset to determine its effectiveness in diverse cell tracking scenarios. In the second experimental setting, namely SAT Generalization Analysis Using CTMC's Wide-Ranging Cell Types, the SAT pipeline's ability to generalize is evaluated using the diverse cell lines and extensive imaging conditions provided by the CTMC dataset.

To fine-tune the Segment Anything Model (SAM) (Kirillov et al., 2023) on LIVECell data, an iterative training scheme was used (Archit et al., 2023). Mini-batches of input images and ground-truth segmentations were sampled with annotations using random positive points or bounding boxes. Key hyperparameters included a batch size of two, dice loss for masks, L2 loss for IOU, and the ADAM optimizer (Kingma and Ba, 2014) with a learning rate of 10^{-5}, adjusted using ReduceLROnPlateau. Models were trained for

100,000 iterations, with partial updates for 25,000 and fine-tuning for 10k iterations. Training was conducted on an A100 GPU with 80 GB of VRAM, using the Vision Transformer (ViT-h) (Dosovitskiy et al., 2020) for robust image segmentation. The implementation utilized PyTorch (Paszke et al., 2019) and the torch-em library (Pape, 2023).

6.1 Experimental Setting 1: SAT Evaluation on Diverse Modalities and Intervals Using the CTC Dataset

In this experimental setting, the performance of the SAT pipeline is assessed across various imaging modalities and time intervals using the Cell Tracking Challenge (CTC) dataset. The CTC dataset includes annotated 2D and 3D time-lapse video sequences of fluorescent counterstained nuclei, as well as 2D Bright Field, Phase Contrast, and Differential Interference Contrast (DIC) microscopy videos. This diverse dataset provides a comprehensive evaluation of the SAT pipeline's effectiveness in different cell-tracking scenarios. The table 3 shows the results for this setting. The sequences evaluated include:

- **PhC-C2DH-U373 (01)**: This sequence contains 61 frames and is a Phase Contrast microscopy video of U373 cells, captured at 10-minute intervals. MOTA is 79.82%, indicating high tracking accuracy. IDF1 is 89.00%, reflecting excellent identity preservation. There are no identity switches (IDS 0.0). MT is 71.43% and ML is 0.0%.
- **PhC-C2DH-U373 (02)**: This sequence contains 12 frames and is a Phase Contrast microscopy video of U373 cells, captured at 10-minute intervals. MOTA is 82.75%, indicating improved accuracy. IDF1 is 91.94%, reflecting excellent identity preservation. There are no identity switches (IDS 0.0). MT is 100.0%, indicating perfect tracking, and ML is 0.0%.
- **Fluo-N2DH-GOWT1**: This sequence contains 38 frames and is a fluorescence microscopy video of GOWT1 cells, captured at 30-minute intervals. MOTA is 88.24%, demonstrating very high accuracy. IDF1 is 91.20%, indicating excellent identity preservation. There are no identity switches (IDS 0.0). MT is 79.17% and ML is 8.34%.
- **Fluo-N2DH-SIM+**: This sequence contains 10 frames and is a fluorescence microscopy video of SIM+ cells, captured at 30-minute intervals. MOTA is 83.03%, showing high accuracy. IDF1 is 88.73%, reflecting excellent identity preservation. There are no identity switches (IDS 0.0). MT is 83.34% and ML is 0.0%.

Overall, the average metrics across sequences are MOTA of 83.46, IDF1 of 90.22, IDS of 0.0, MT of 83.45%, and ML of 2.08%, demonstrating high tracking accuracy, excellent identity preservation, no identity switches, most trajectories being well tracked, and very few trajectories being mostly lost.

6.2 Experimental Setting 2: SAT Generalization Analysis Using CTMC's Wide-Ranging Cell Types

In this experimental setting, the performance of the SAT pipeline is evaluated using the diverse cell lines and extensive imaging conditions provided by the Cell Tracking with Mitosis Detection Challenge (CTMC) dataset. The table 4 shows the results for this setting. The sequences evaluated include:

- **PL1Ut-run05**: This sequence contains 371 frames. It consists of phase-contrast images of PL1Ut cells, a rat hepatoma cell line. MOTA is 93.12%, IDF1 is 96.56%, with IDS 0.0%. MT is 100.0% and ML is 0.0%.
- **A-10-run01**: This sequence contains 305 frames. It consists of phase-contrast images of A-10 cells, a rat smooth muscle cell line. MOTA is 80.79%, IDF1 is 90.03%, with IDS 0.0. MT is 80.0% and ML is 0.0%.
- **LLC-MK2-run03**: This sequence contains 89 frames. It consists of phase-contrast images of LLC-MK2 cells, a monkey kidney epithelial cell line. MOTA is 96.18%, IDF1 is 98.08%, with IDS 0.0. MT is 100.0% and ML is 0.0%.
- **APM-run05**: This sequence contains 130 frames. It consists of phase-contrast images of APM cells, a human peripheral blood mononuclear cell line. MOTA is 61.85%, IDF1 is 82.38%, with IDS 0.0. MT is 75.0% and ML is 0.0%.
- **U2O-S-run03**: This sequence contains 100 frames. It consists of phase-contrast images of U2O-S cells, a human osteosarcoma cell line. MOTA is 68.93%, IDF1 is 84.17%, with IDS 0.0. MT is 75.0% and ML is 0.0%.
- **OK-run01**: This sequence contains 57 frames. It consists of phase-contrast images of OK cells, an opossum kidney epithelial cell line. MOTA is 57.77%, IDF1 is 78.11%, with IDS 0.0. MT is 60.0% and ML is 6.67%.

Overall, the average metrics across sequences are MOTA of 76.44%, IDF1 of 88.22%, IDS of 0.0%, MT

Table 3: Results for the SAT Evaluation on Diverse Modalities and Intervals Using the CTC Dataset. Higher values are better for MOTA, IDF1, and MT, indicated by upward arrows (↑). Lower values are better for IDS and ML, indicated by downward arrows (↓).

Sequence	Modality	Images	Cells	Points (N)	MOTA ↑	IDF1 ↑	IDS ↓	MT ↑	ML ↓
PhC-C2DH-U373 (01)	Phase Contrast	61	427	6P - 3N	79.82	89.00	0.0	71.43	0.0
PhC-C2DH-U373 (02)	Phase Contrast	12	58	6P - 3N	82.75	91.94	0.0	100.0	0.0
Fluo-N2DH-GOWT1	Fluorescence	38	799	3P - 3N	88.24	91.20	0.0	79.17	8.34
Fluo-N2DH-SIM+	Fluorescence	10	271	3P - 3N	83.03	88.73	0.0	83.34	0.0
Average/Total	-	121	1,555	-	**83.46**	**90.22**	**0.0**	**83.45**	**2.08**

Table 4: Results for the SAT Generalization Analysis Using CTMC's Wide-Ranging Cell Types. Higher values are better for MOTA, IDF1, and MT, indicated by upward arrows (↑). Lower values are better for IDS and ML, indicated by downward arrows (↓).

Sequence	Modality	Images	Cells	Points (N)	MOTA ↑	IDF1 ↑	IDS ↓	MT ↑	ML ↓
PL1Ut-run05	Phase Contrast	371	742	30P - 3N	93.12	96.56	0.0	100.0	0.0
A-10-run01	Phase Contrast	305	1,525	20P - 3N	80.79	90.03	0.0	80.00	0.0
LLC-MK2-run03	Phase Contrast	89	445	25P - 3N	96.18	98.08	0.0	100.0	0.0
APM-run05	Phase Contrast	130	443	25P - 3N	61.85	82.38	0.0	75.00	0.0
U2O-S-run03	Phase Contrast	100	396	15P - 3N	68.93	84.17	0.0	75.00	0.0
OK-run01	Phase Contrast	57	841	30P - 3N	57.77	78.11	0.0	60.00	6.67
Average/Total	-	1,173	5,447	-	**76.44**	**88.22**	**0.0**	**81.67**	**1.12**

of 81.67%, and ML of 1.12%, demonstrating good tracking accuracy, identity preservation, no identity switches, most trajectories being well tracked, and very few trajectories being mostly lost.

7 ANALYSIS AND DISCUSSION

This section explains the results of the two experimental settings, highlighting the SAT pipeline's ability to generalize across different modalities and cell lines. In the first setting, SAT Evaluation on Diverse Modalities and Intervals Using the CTC Dataset, the pipeline was tested on the Cell Tracking Challenge (CTC) dataset, which includes diverse imaging modalities. Table 3 shows the SAT pipeline achieves 83.46% MOTA, 90.22% IDF1, and zero Identity Switches (IDS), demonstrating consistent tracking across modalities. In the second setting, SAT Generalization Analysis Using CTMC's Wide-Ranging Cell Types, the pipeline was evaluated using the Cell Tracking with Mitosis Detection Challenge (CTMC) dataset, featuring diverse cell lines and extensive imaging conditions. The results, shown in Table 4, reveal that the SAT pipeline maintains good tracking accuracy with an average MOTA of 76.44% and an IDF1 of 88.22%, indicating effective generalization to different cell types and high identity preservation. The variation in MOTA across different cell types can be attributed to the challenges presented by different imaging modalities and cell cultures. For instance, the PhC-C2DH-U373 cell culture, imaged with phase-contrast microscopy, poses difficulties due to low contrast, causing cells to merge with the background as they grow. In contrast, the Fluo-N2DH-GOWT1 culture, captured with fluorescence microscopy, offers higher contrast, making segmentation and tracking significantly easier. These factors contribute to the variation in MOTA and are reflected in the observed performance across different datasets.

Figure 2 illustrates tracking results for the first setting with sequences PhC-C2DH-U373 and Fluo-N2DH-SIM+. The top row shows the ground truth, while the bottom row shows the SAT pipeline predictions. For PhC-C2DH-U373, all 7 cells are correctly segmented initially, with one cell missed at t + 300 minutes but recovered at t + 600 minutes and tracked till the last frame at t + 885 minutes. The SAT pipeline achieves a MOTA of 78.2%, IDF1 of 89.0%, 0 IDS, 71.4% MT, and 0.0% ML. For Fluo-N2DH-SIM+, all cells are correctly segmented initially, with one cell missed at t + 58 minutes and another at t + 116 minutes, both recovered at t + 174 minutes. The pipeline achieves a MOTA of 83.0%, IDF1 of 88.7%, 0 IDS, 83.3% MT, and 0.0% ML.

Figure 3 shows tracking results for the second setting with sequences LLC-MK2-run03 and A-10-run01. The top row shows ground truth bounding boxes, while the bottom row shows SAT pipeline predictions with both bounding boxes and segmentation masks. For LLC-MK2-run03, all cells are correctly segmented across all frames, achieving a MOTA of 96.2%, IDF1 of 98.1%, 0 IDS, 100.0% MT, and 0.0% ML. For A-10-run01, all cells are correctly segmented initially, with cell number 2 missed at t + 119 minutes but recovered at t + 149 minutes. The pipeline

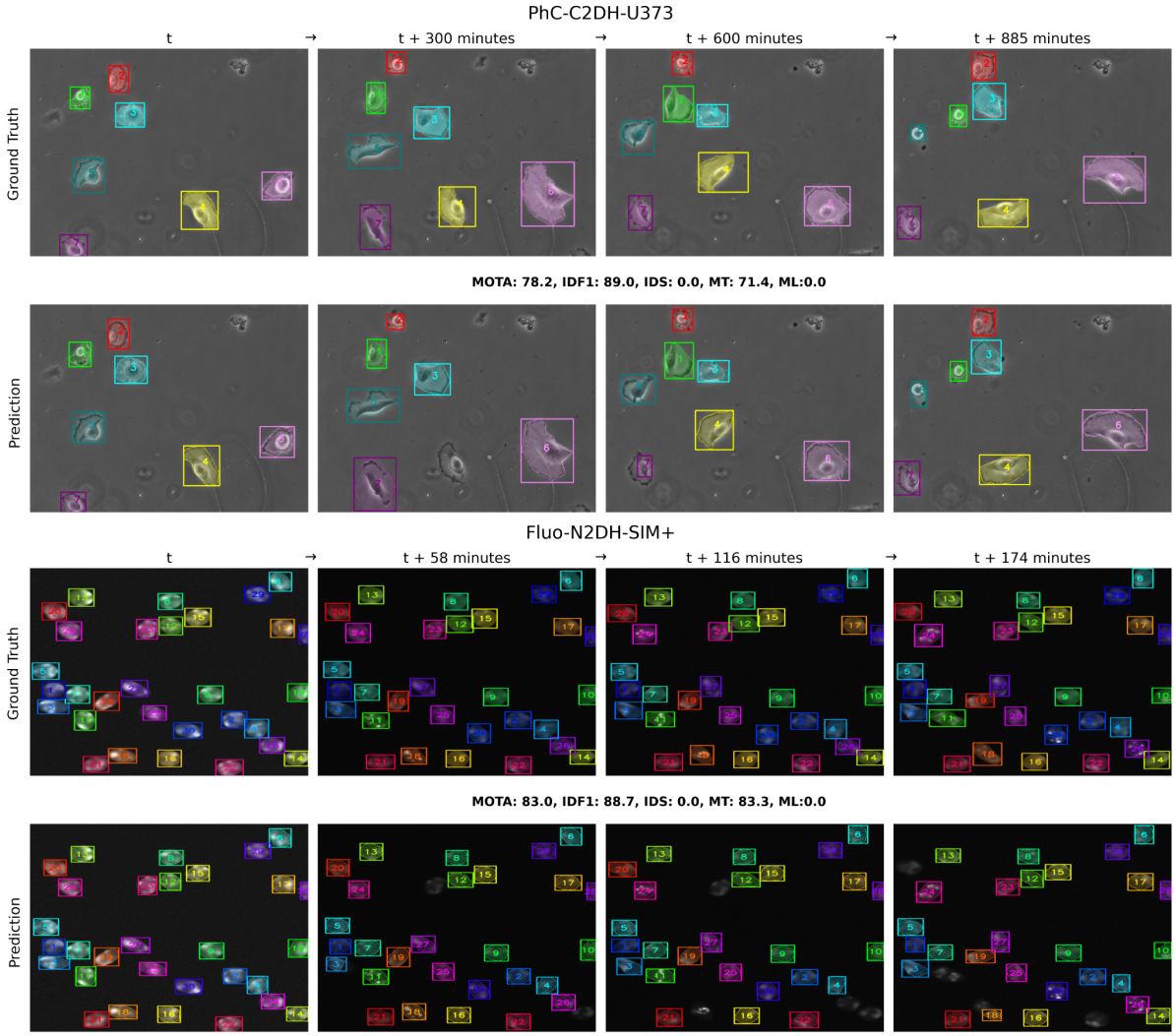

Figure 2: Tracking results for experimental setting 1 with two sequences, PhC-C2DH-U373 and Fluo-N2DH-SIM+. The top row shows ground truth, and the bottom row shows SAT pipeline predictions with evaluation scores above the prediction row.

achieves a MOTA of 80.8%, IDF1 of 90.0%, 0 IDS, 80.0% MT, and 0.0% ML.

While direct comparison with other tracking methods is not entirely feasible due to the unique nature of the proposed approach, an additional experiment with ByteTrack was conducted to offer some insights. ByteTrack, trained on the same LIVECell dataset as SAT, failed to detect any cells when applied to the CTMC dataset. To investigate further, ByteTrack was trained on a subset of the CTMC dataset (with no overlap with the test set) and tested on the remaining sequences. ByteTrack's performance was lower than SAT, with an average MOTA of 29.3 compared to 76.4, and higher IDS and ML scores. These findings emphasize the superior adaptability and robustness of SAT, which can generalize effectively to unseen datasets without retraining, a limitation observed in traditional methods like ByteTrack (Zhang et al., 2022).

Overall, the proposed SAT pipeline demonstrates strong generalization across different modalities and cell lines, achieving high tracking accuracy and identity preservation. This pipeline significantly impacts the biological and biomedical research community by automating cell segmentation and tracking, reducing the need for expert knowledge and manual intervention. It enhances accuracy, consistency, and speeds up data annotation, benefiting cancer research, drug development, and stem cell studies. SAT's broad applicability with minimal retraining makes it a versatile tool, driving new insights and improving research efficiency.

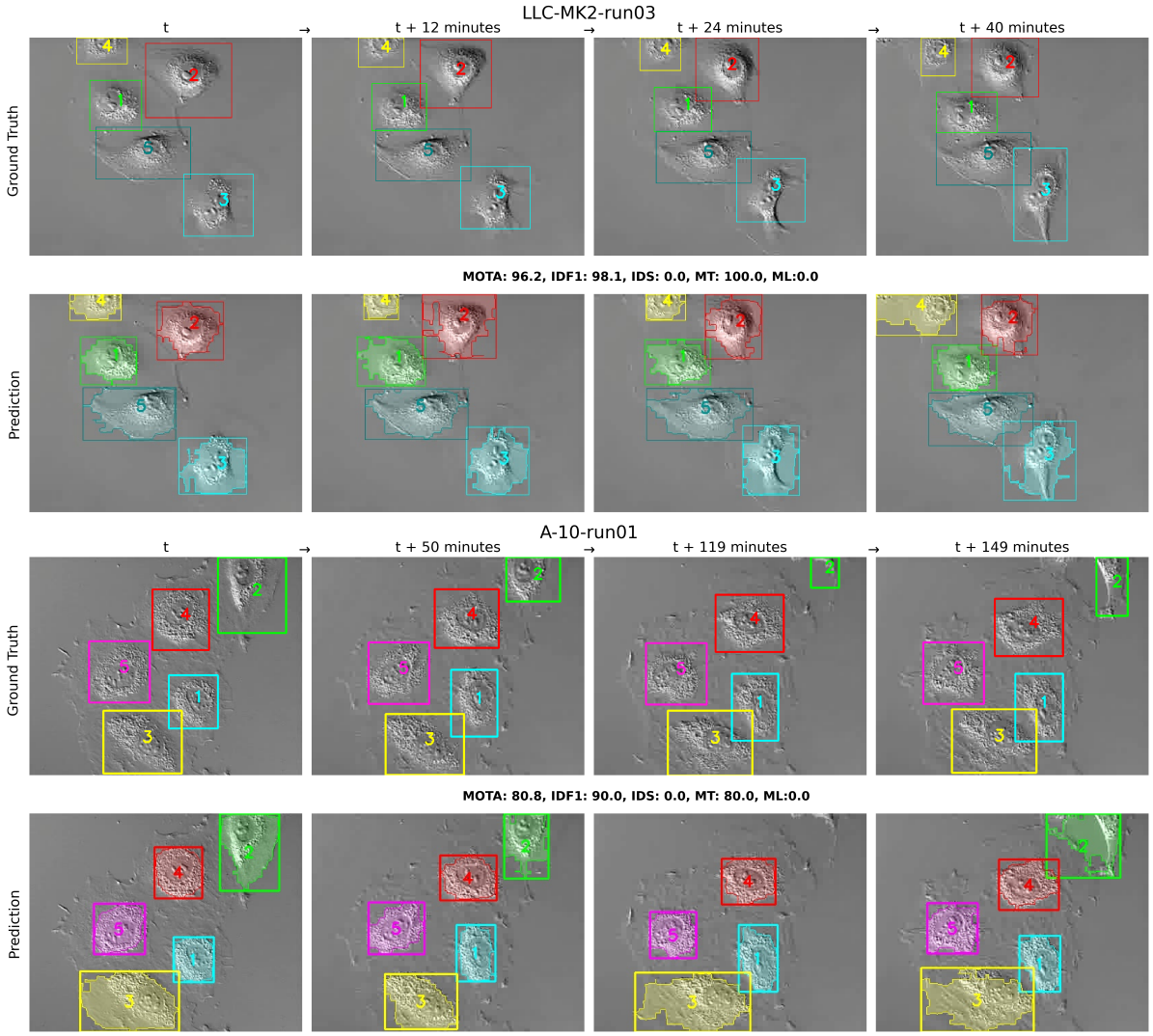

Figure 3: Tracking results for experimental setting 2 with two sequences, LLC-MK2-run03 and A-10-run01. The top row shows ground truth, and the bottom row shows SAT pipeline predictions with evaluation scores above the prediction row.

8 CONCLUSION

This study introduces a pipeline for Cell Segmentation and Tracking using only point annotations in the first frame of the sequence. The SAT pipeline demonstrates strong generalization and robustness across diverse imaging modalities and cell types, achieving over 80% Multiple Object Tracking Accuracy (MOTA) in evaluations on two diverse datasets. This highlights its effectiveness in various cell-tracking scenarios. The pipeline achieves high tracking accuracy and identity preservation, effectively handling different imaging conditions and extensive cell line variations. By automating cell segmentation and tracking, SAT reduces expert intervention and enhances research efficiency. This automation streamlines annotation, benefiting large-scale studies in cancer research, drug development, and stem cell studies. With improved efficiency and accuracy in cell tracking, the SAT pipeline opens the door to new insights and breakthroughs, facilitating scientific discoveries and their application in clinical settings.

REFERENCES

Anjum, S. and Gurari, D. (2020). Ctmc: Cell tracking with mitosis detection dataset challenge. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*.

Aramini, B., Masciale, V., Grisendi, G., Bertolini, F., Maur, M., Guaitoli, G., Chrystel, I., Morandi, U., Stella, F., Dominici, M., et al. (2022). Dissecting tumor growth: the role of cancer stem cells in drug resistance and recurrence. *Cancers*.

Archit, A., Nair, S., Khalid, N., Hilt, P., Rajashekar, V., Freitag, M., Gupta, S., Dengel, A., Ahmed, S., and Pape, C. (2023). Segment anything for microscopy. *bioRxiv*.

Bernardin, K. and Stiefelhagen, R. (2008). Evaluating multiple object tracking performance: the clear mot metrics. *EURASIP Journal on Image and Video Processing*.

Chou, T.-C., You, L., Beerens, C., Feller, K. J., Storteboom, J., and Chien, M.-P. (2023). Instant processing of large-scale image data with fact, a real-time cell segmentation and tracking algorithm. *Cell Reports Methods*, 3(11).

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., et al. (2020). An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*.

Durkee, M. S., Abraham, R., Clark, M. R., and Giger, M. L. (2021). Artificial intelligence and cellular segmentation in tissue microscopy images. *The American journal of pathology*, 191(10):1693–1701.

Edlund, C., Jackson, T. R., Khalid, N., Bevan, N., Dale, T., Dengel, A., Ahmed, S., Trygg, J., and Sjögren, R. (2021). Livecell—a large-scale dataset for label-free live cell segmentation. *Nature methods*.

Harley, A. W., Fang, Z., and Fragkiadaki, K. (2022). Particle video revisited: Tracking through occlusions using point trajectories. In *European Conference on Computer Vision*. Springer.

Jelli, E., Ohmura, T., Netter, N., Abt, M., Jiménez-Siebert, E., Neuhaus, K., Rode, D. K., Nadell, C. D., and 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(6):659–676.

Khalid, N., Froes, T. C., Caroprese, M., Lovell, G., Trygg, J., Dengel, A., and Ahmed, S. (2023). Pace: Point annotation-based cell segmentation for efficient microscopic image analysis. In *International Conference on Artificial Neural Networks*. Springer.

Khalid, N., Koochali, M., Rajashekar, V., Munir, M., Edlund, C., Jackson, T. R., Trygg, J., Sjögren, R., Dengel, A., and Ahmed, S. (2022a). Deepmucs: A framework for co-culture microscopic image analysis: From generation to segmentation. In *2022 IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI)*. IEEE.

Khalid, N., Munir, M., Edlund, C., Jackson, T. R., Trygg, J., Sjögren, R., Dengel, A., and Ahmed, S. (2021a). Deepcens: An end-to-end pipeline for cell and nucleus segmentation in microscopic images. In *2021 International Joint Conference on Neural Networks (IJCNN)*. IEEE.

Khalid, N., Munir, M., Edlund, C., Jackson, T. R., Trygg, J., Sjögren, R., Dengel, A., and Ahmed, S. (2021b). Deepcis: An end-to-end pipeline for cell-type aware instance segmentation in microscopic images. In *2021 IEEE EMBS International Conference on Biomedical and Health Informatics (BHI)*. IEEE.

Khalid, N., Schmeisser, F., Koochali, M., Munir, M., Edlund, C., Jackson, T. R., Trygg, J., Sjögren, R., Dengel, A., and Ahmed, S. (2022b). Point2mask: A weakly supervised approach for cell segmentation using point annotation. In *Medical Image Understanding and Analysis: 26th Annual Conference, MIUA 2022, Cambridge, UK, July 27–29, 2022, Proceedings*. Springer.

Kingma, D. P. and Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A. C., Lo, W.-Y., et al. (2023). Segment anything. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*.

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

Malin-Mayor, C., Hirsch, P., Guignard, L., McDole, K., Wan, Y., Lemon, W. C., Kainmueller, D., Keller, P. J., Preibisch, S., and Funke, J. (2023). Automated reconstruction of whole-embryo cell lineages by learning from sparse annotations. *Nature biotechnology*.

Maška, M., Ulman, V., Delgado-Rodriguez, P., Gómez-de Mariscal, E., Nečasová, T., Guerrero Peña, F. A., Ren, T. I., Meyerowitz, E. M., Scherr, T., Löffler, K., et al. (2023). The cell tracking challenge: 10 years of objective benchmarking. *Nature Methods*.

Newman, R. H., Fosbrink, M. D., and Zhang, J. (2011). Genetically encodable fluorescent biosensors for tracking signaling dynamics in living cells. *Chemical reviews*, 111(5):3614–3666.

Padovani, F., Mairhörmann, B., Falter-Braun, P., Lengefeld, J., and Schmoller, K. M. (2022). Segmentation, tracking and cell cycle analysis of live-cell imaging data with cell-acdc. *BMC biology*.

Pape, C. (2023). torch-em: Deep learning based semantic and instance segmentation for 3d electron microscopy and other bioimage analysis problems based on pytorch.

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., et al. (2019). Pytorch: An imperative style, high-performance deep learning library. In *Advances in neural information processing systems*.

Phasefocus (n.d.). Manual tracking - an inconvenient truth. Accessed: 2024-07-01.

Rajič, F., Ke, L., Tai, Y.-W., Tang, C.-K., Danelljan, M., and Yu, F. (2023). Segment anything meets point tracking. *arXiv preprint arXiv:2307.01197*.

Ristani, E., Solera, F., Zou, R., Cucchiara, R., and Tomasi, C. (2016). Performance measures and a data set for multi-target, multi-camera tracking. In *European conference on computer vision*. Springer.

Schwendy, M., Unger, R. E., and Parekh, S. H. (2020). Evican–a balanced dataset for algorithm development in cell and nucleus segmentation. *Bioinformatics*.

Shi, J. et al. (1994). Tomasi. good features to track. In *1994 Proceedings of IEEE conference on computer vision and pattern recognition*. sn.

Stringer, C., Wang, T., Michaelos, M., and Pachitariu, M. (2020). Cellpose: a generalist algorithm for cellular segmentation. *Nature Methods*.

Wang, R., Butt, D., Cross, S., Verkade, P., and Achim, A. (2023). Bright-field to fluorescence microscopy image translation for cell nuclei health quantification. *Biological Imaging*, 3:e12.

Yazdi, R. and Khotanlou, H. (2024). A survey on automated cell tracking: challenges and solutions. *Multimedia Tools and Applications*.

Zhang, Y., Sun, P., Jiang, Y., Yu, D., Weng, F., Yuan, Z., Luo, P., Liu, W., and Wang, X. (2022). Bytetrack: Multi-object tracking by associating every detection box. In *European conference on computer vision*. Springer.