Time-symmetric tracking of yeast cells and other objects, mitigation of benchmark positional bias

Gergely Szabó

Theses of the PhD dissertation

supervisor:
Dr. András Horváth

Pázmány Péter Catholic University
Faculty of Information Technology and Bionics
Roska Tamás Doctoral School of Sciences and Technology

Budapest, 2024

Introduction

During my PhD studies, my primary objective was to develop a novel multi-object segmentation and tracking architecture for videomicroscopic recordings of budding yeast cells. This initial task, while highly specialized, led to the development of a generalizable solution with far-reaching implications. The architecture I designed, initially tailored for the detection, instance segmentation, and tracking of yeast cells, demonstrated significant advantages over existing tools, both in terms of performance and adaptability. This outcome extends beyond the original task, accommodating a wide variety of image modalities due to its data-driven approach. Through comprehensive evaluations, I have explored its strengths and limitations across diverse synthetic and semi-synthetic datasets. These results serve as the foundation of thesis group 1.

While designing the architecture, I identified a previously mostly overlooked bias in convolutional neural networks (CNNs), particularly related to object positioning in widely used image processing benchmarks. This bias, present even in standard evaluation subsets, has far-reaching implications for tasks that rely on CNNs for object recognition and segmentation. My investigation into this bias, supported by experiments with data manipulation techniques and architectural modifications, led to the discovery of simple solutions that nearly completely eliminate the bias, and hint at boundary conditions in CNNs as the root cause. These results serve as the foundation of thesis group 2.

Alongside these main contributions, which are reflected in my first-author publications, I have also participated in numerous studies as a second or lower author. While these are beyond the scope of my thesis points and my PhD study in general, I will provide a list of these works for the sake of transparency.

New Scientific Results

Thesis 1a

I developed a novel deep-learning-based multi-object instance segmentation and tracking architecture for videomicroscopic recordings of budding yeast cells. On a yeast tracking dataset collected by IFOM, the architecture achieved IoU-based segmentation and tracking F-scores of $[0.918 \pm 0.019,\ 0.917 \pm 0.016]$*, respectively. This performance surpassed that of competing state-of-the-art tools designed for this particular task, specifically Phylocell* $[0.881 \pm 0.020,\ 0.878 \pm 0.020]$ *and YeaZ* $[0.818 \pm 0.022,\ 0.807 \pm 0.023]$.
Corresponding publications: [Ar1, Ar2]

Segmentation and tracking of cells in videomicroscopic recordings are critical in various fields of biological research, including drug discovery, gene expression monitoring, and others. Although some solutions exist for the automatic or semi-automatic tracking of yeast cells, even machine-learning-based approaches often rely on specific architectural assumptions, such as classical pattern and position matching models and Markovian temporal forward tracking. While these solutions may be advantageous in fields like the automotive industry, where inference speed is crucial, in videomicroscopic tracking applications, prediction performance — particularly stable tracking — is far more important. Moreover, such limitations reduce the flexibility of these models, hindering their data-driven adaptability. Based on a request from the International Foundation of Medicine (IFOM), my objective was to develop a novel multi-object segmentation and tracking (MOTS) architecture designed to support data-driven generalization, improved prediction performance, and enhanced tracking consistency, even at the potential cost of inference speed. For further details on the proposed architecture, please refer to Thesis 1b.

The models used in the MOTS pipeline were trained on manually corrected and curated data provided by IFOM. Final performance evaluations were conducted at IFOM using samples that were not disclosed to me during model design and training. An example result for segmentation and tracking predictions is presented in Fig. 1, demonstrating near-perfect outcomes based on empirical assessment.

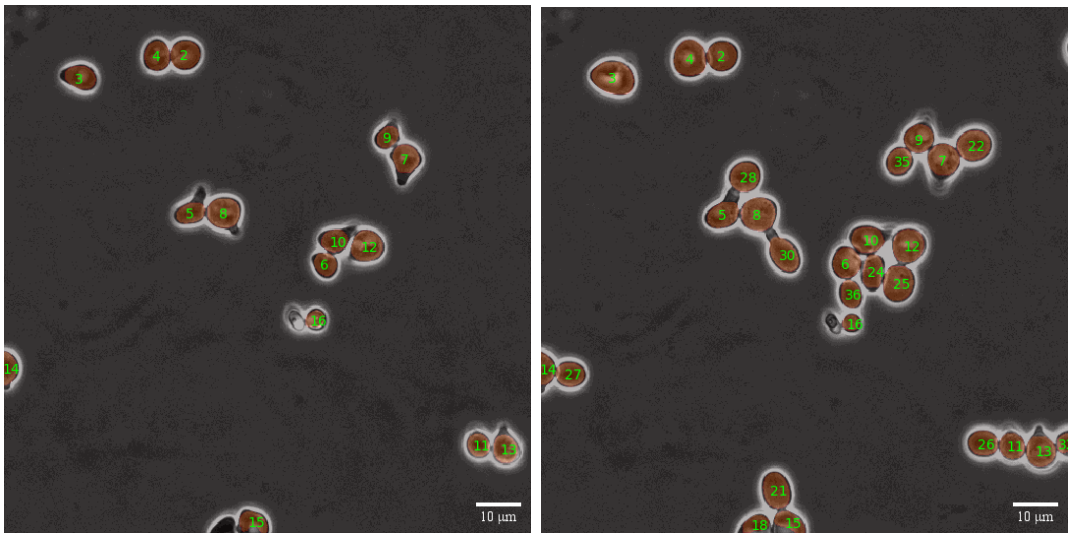


Figure 1: **Tracking Sample Results**
A side by side display of tracking results, demonstrating successful tracking of all cells in the same recording with a temporal difference of 8 frames. New cells were appropriately assigned new IDs while maintaining consistent tracking of existing cells.

For comparative evaluation, the tools Phylocell and YeaZ were chosen. Phylocell was selected as it was used to create the raw ground truth labels that were later manually corrected, while YeaZ exhibited the greatest architectural similarity to my design among the available tools, albeit still using a highly different architecture. Instance segmentation F-scores were measured using Intersection over Union (IoU), while tracking F-scores were assessed using link matching between consecutive frames. In case of link matching a prediction is considered true positive only when both segmentation instances were correctly matched and the track IDs of the matched instances were identical on both the ground truth and prediction sides, while false predictions include track loss on both the ground truth and prediction sides as well as ID switches.

Segmentation and tracking performance were evaluated with tracks leaving the field of view both enabled and disabled in the ground truth data, the former introducing a positive bias towards Phylocell designed to track these instances, whereas the latter provided an unbiased evaluation. The results, presented in Fig. 2, show that in the unbiased scenario the proposed pipeline far outperforms

both Phylocell $[0.881 \pm 0.020, 0.878 \pm 0.020]$ and YeaZ $[0.818 \pm 0.022, 0.807 \pm 0.023]$ in terms of segmentation and tracking F-scores $[0.918 \pm 0.019, 0.917 \pm 0.016]$, and it even exhibits superior performance with enabled border tracks.

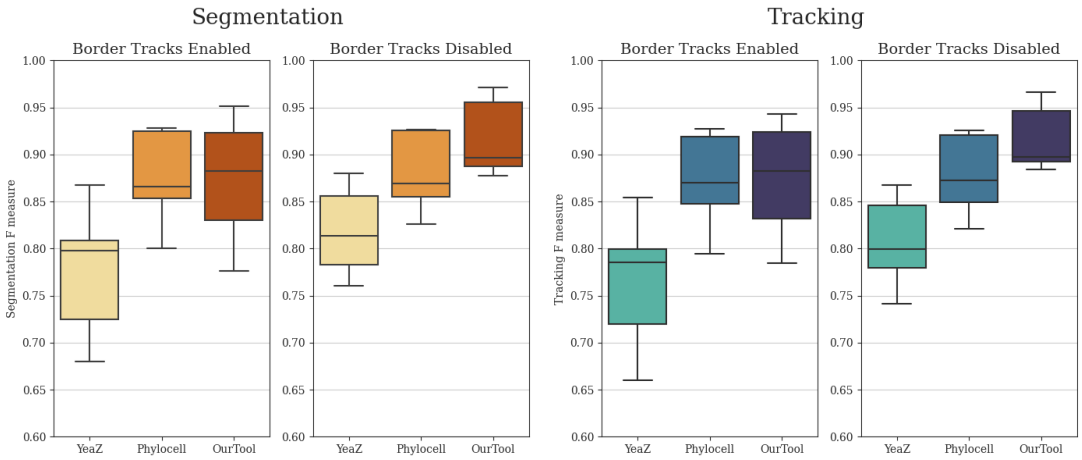


Figure 2: **Comparative Tool Evaluation**
Comparative evaluation of segmentation and tracking performances of Phylocell, YeaZ and the proposed pipeline based on F-measures of segmentation IOU and tracking link matches with enabled and disabled border track predictions. Disabled border tracks present an unbiased comparison between the tools, while enabled border tracks show results with the unaltered outputs.

Thesis 1b

The proposed architecture is inherently capable of reconstructing fragmented tracks due to its novel time-symmetric tracking approach, greatly improving tracking consistency, a critical requirement for accurate cell inheritance assignment. For instance, in scenarios with uniform random removal of every fifth object instance, the tracking-based reconstruction improved the tracking F-score from 0.404 ± 0.016 *to* 0.888 ± 0.013.
Corresponding publication: [Ar1]

The proposed architecture consists of the two major segments of instance segmentation and tracking. The tracking segment can be further divided into the steps: temporal local tracking, global assignment using depth-first-search-based track ID reduction, and a post-processing step for missed instance interpolation. The macro-architecture is illustrated from a data flow perspective in Fig. 3. The novelty of the architecture lies in its combined use of the unique time-symmetric local tracking and global assignment steps. This approach is based on the concept of tracking each object within its local temporal neighborhood — later referred to as the single-sided tracking range (TR) — calculating metric similarity between the offset local track predictions as illustrated in Fig. 4, and performing globally optimal assignment using the Hungarian method in a hierarchical order that decreases as the temporal offset increases. Following this, the architecture's capability to naturally reassign fragmented tracks within $2TR$ allows for the identification of missed object instances, which can be reinterpolated using the following equations:

$$\Delta c(x,y) = \frac{(t - t_{\text{last}})c_{\text{next}}(x,y) + (t_{\text{next}} - t)c_{\text{last}}(x,y)}{t_{\text{next}} - t_{\text{last}}} \tag{1}$$

$$S(x,y) = S_{\text{last}}(x,y) + c(x,y) \tag{2}$$

Where $S(x,y)$ and t represent the interpolated segmentation and its corresponding time, respectively. $S_{\text{last}}(x,y)$, $c_{\text{last}}(x,y)$, and t_{last} denote the segmentation, centroid, and time of the last occurrence of the object with the given ID, while $c_{\text{next}}(x,y)$ and t_{next} represent the centroid and time of the next occurrence of the object with the given ID.

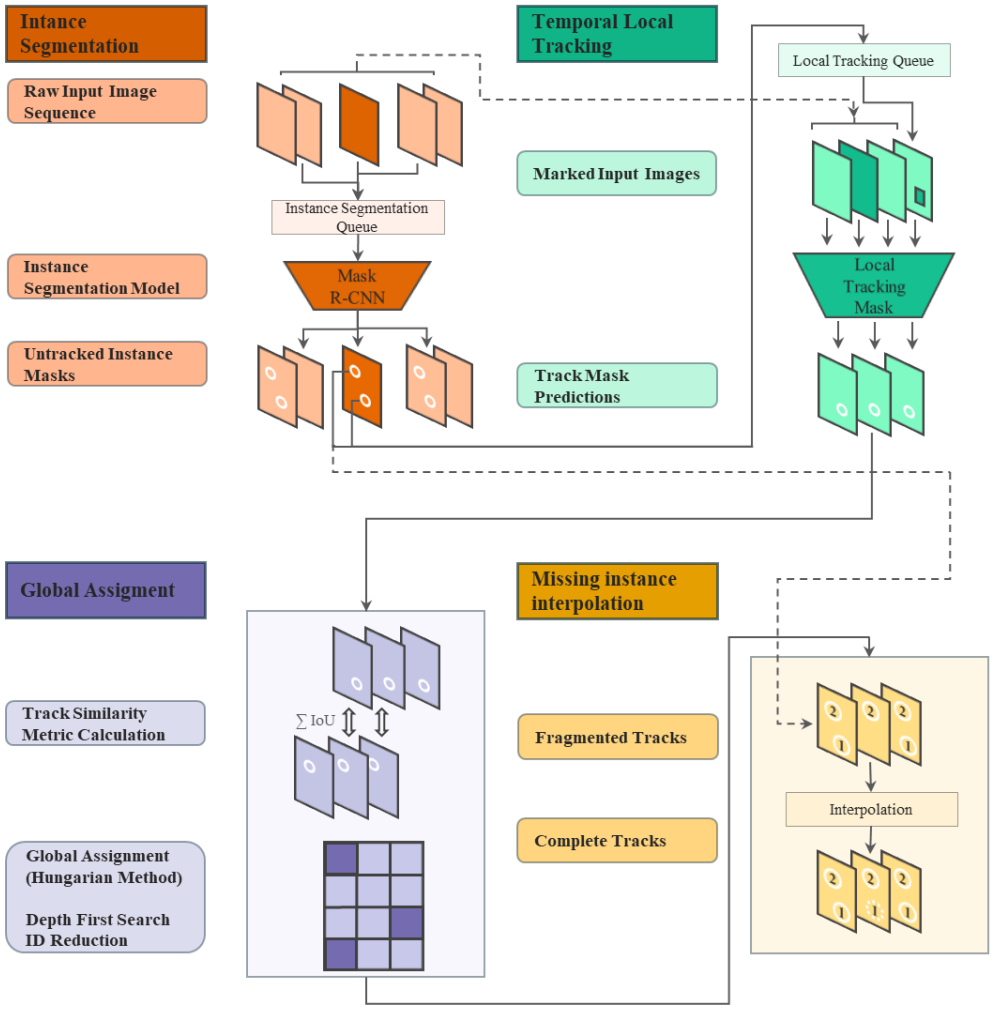


Figure 3: **Macro-architecture overview**
Data flow diagram of the proposed architecture, illustrating the process from raw input image sequence to finalized track predictions.

The reconstructive capabilities of the architecture are best represented through the experiments in which every 1:15 or 1:5 object was removed in a uniform random distribution or every 1:5 object was randomly removed but in continuous boxes. The results, shown in Fig. 5 and Fig. 6, illustrate how effectively the architecture preserves tracking continuity, especially with an appropriately chosen local tracking range (TR) model parameter.

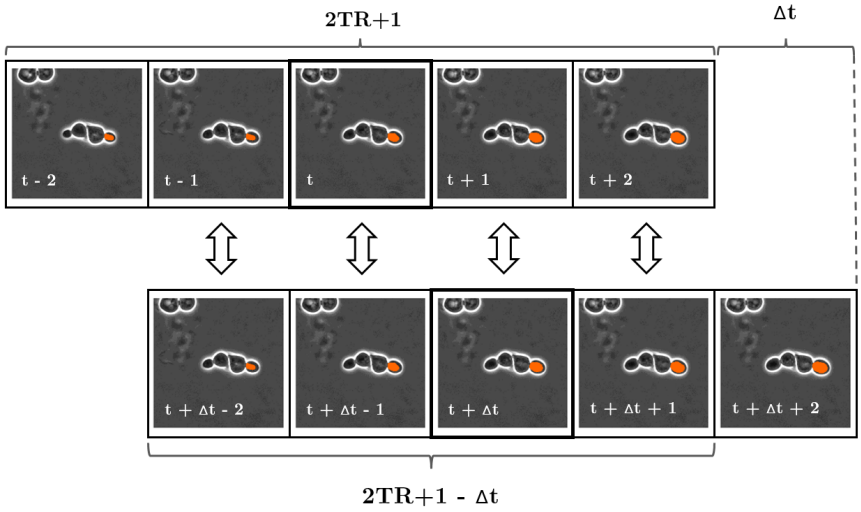


Figure 4: **Metric similarity structure**
Schematic structure of the metric similarity measurement step between $2TR + 1$ long local tracks of different cell instances on frames with a temporal distance of Δt. The solid lines indicate the central segmentation instances with a Δt temporal distance to be matched, while the arrows indicate the similarity metric between the segmentation estimates for each time frame. Subsequently, the individual metric results are averaged to obtain a single measure of similarity.

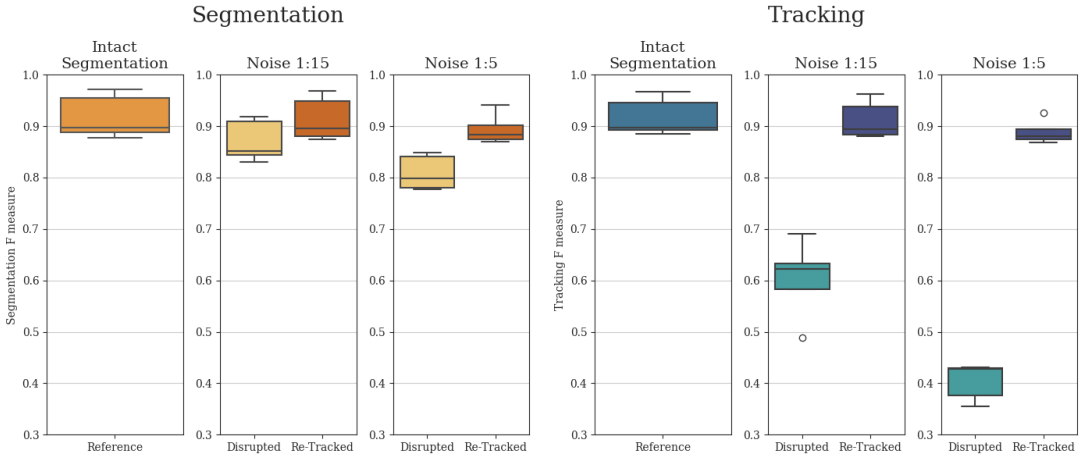


Figure 5: **Uniform random noise reconstruction**
Comparative evaluation of segmentation and tracking F-scores for Intact, Noise 1:15, and Noise 1:5 cases. Both segmentation and tracking results include baseline values disrupted by the given noise, as well as re-tracked values initiated with the disrupted segmentation.

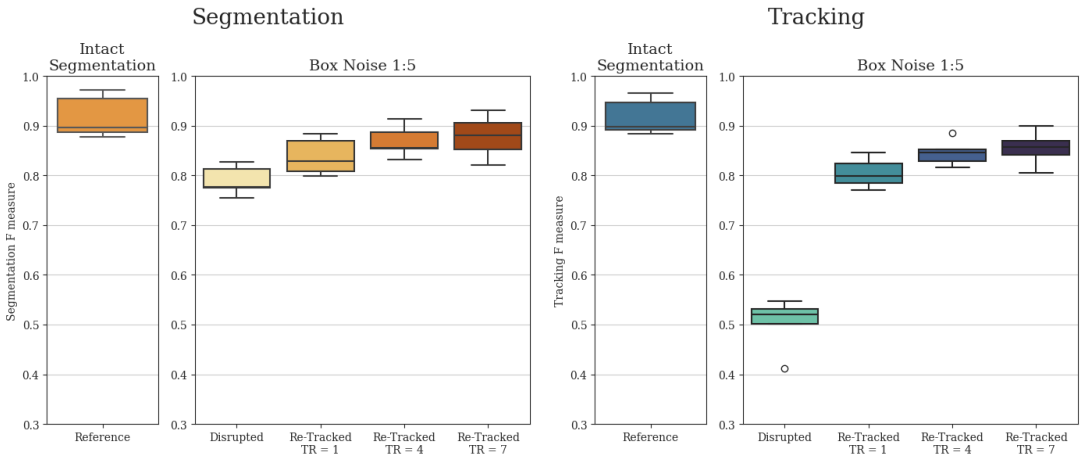


Figure 6: **Box noise reconstruction**
Comparative evaluation of segmentation and tracking F-scores for Intact and Box Noise 1:5 cases. Both segmentation and tracking results include baseline values disrupted by the noise, as well as re-tracked values initiated with the disrupted segmentation using local tracking ranges (TR) of 1, 4 and 7.

Thesis 1c

I evaluated the proposed architecture on various synthetic and semi-synthetic datasets, showcasing its robustness in addressing potential challenges in cell tracking environments and other natural tracking contexts, such as pedestrian tracking. The results demonstrate the versatility of the architecture, the reliability of its novel tracking approach, and can serve as a guide of expected performance on other datasets.
Corresponding publications: [Ar1, Ar2]

For the simulation and benchmark evaluation of varied cell tracking environments, I created five synthetic datasets due to the lack of high-quality publicly available cell segmentation and tracking datasets.

- "S. Arrows" simulates fast-moving arrows that represent movement direction and are rigid objects.
- "S. Amoeboids" simulates fast-moving objects with semi-random shapes undergoing continuous elastic changes, modeled using Perlin noise.
- "S. Amoeboids-PC" features objects similar to those in "S. Amoeboids" but incorporates simulated phase-contrast microscopy effects.
- "S. Amoeboids-PCC" builds upon "S. Amoeboids-PC" by introducing modified object motility through the promotion of clumping behavior.
- "S. Amoeboids-PCCA" is similar to "S. Amoeboids-PCC" but adds a static, highly disruptive noise randomized for each video.

In addition, all synthetic samples include image-wise and video-wise random background noise. Prediction samples are shown in Fig. 7, while segmentation and tracking F-scores for each dataset category are presented in Tab. 1. The results demonstrate that, even in substantially more complex scenarios compared to yeast tracking, the architecture achieves reliable object tracking. However, detection and instance segmentation performance are noticeably reduced in cases involving noisy or otherwise challenging recordings.

Partially building on "S. Arrows" and "S. Amoeboids", I created two additional scenarios, each consisting of multiple datasets that express the signature behaviors

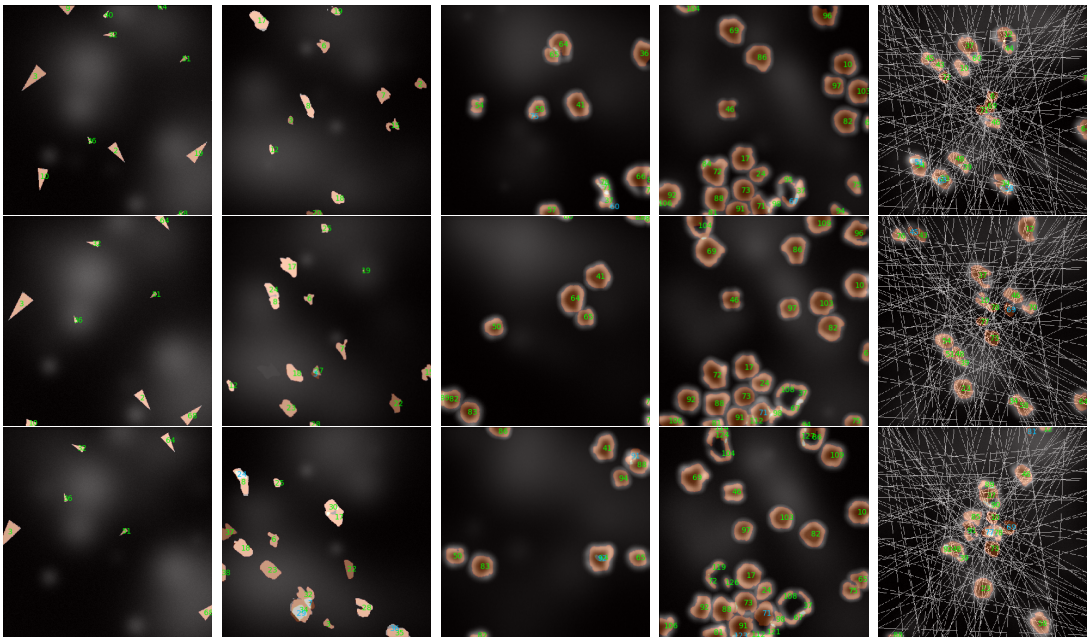


Figure 7: **Synthetic sets tracking sample results**
A display of segmentation and tracking results on various synthetic datasets with highly different object behaviors, ordered from left to right as "S. Arrows", "S. Amoeboids", "S. Amoeboids-PC", "S. Amoeboids-PCC", and "S. Amoeboids-PCCA." The displayed consecutive images are only 4 frames apart to show understandable results even in cases of extremely fast-moving objects.

	Segmentation F score	Tracking F score
S. Arrows	0.9185 ± 0.0057	0.8990 ± 0.0076
S. Amoeboids	0.7137 ± 0.0082	0.6605 ± 0.0093
S. Amoeboids-PC	0.6861 ± 0.0089	0.6662 ± 0.0098
S. Amoeboids-PCC	0.7726 ± 0.0101	0.7693 ± 0.0111
S. Amoeboids-PCCA	0.5078 ± 0.0129	0.5022 ± 0.0140
Yeast Reference	0.9234 ± 0.0136	0.9202 ± 0.0138

Table 1: **Synthetic sets numeric results**
Segmentation and Tracking F scores of the full pipeline for various synthetic datasets with vastly different object behaviors and challenges compared to natural yeast recordings.

at varying strengths. In the "Visual signaling" scenario, the synthetic arrows exhibit rapid turning behavior that can only be predicted through visual cues. This behavior is analogous to the movement of cars partially predictable by signal lights or the movement of pedestrians based on their eye direction. In the "Semi-

random positioning" scenario, the elastic changes of the synthetic amoeboids are disabled. However, the amoeboids display semi-random, completely unpredictable movement patterns. The aim of these scenarios is to provide baseline tests for potential real-world cases where visual cues and object motility are interrelated, or where stable tracking of objects must rely at least partially on morphological or other visual characteristics rather than solely on positional derivatives. For comparative prediction results on these datasets, please refer to Thesis 1d.

In addition to the fully synthetic datasets and the yeast tracking benchmark, I tested the proposed architecture on a variant of the MOTSynth-MOTS-CVPR22 benchmark dataset. In this benchmark, model training is performed on synthetic samples generated using the videogame GTA V, while evaluation is conducted on manually annotated real-world pedestrian tracking and segmentation data. Even without extensive hyperparameter tuning due to the dataset's extremely large size, the architecture achieved a HOTA score of **48.56**, nearly identical to the benchmark value of 48.8 submitted in 2022. This benchmark was based on the widely used state-of-the-art architecture by P. Bermann et al., "Tracking Without Bells and Whistles" (ICCV, 2019). Furthermore, the AssA score of **82.39** achieved by the proposed architecture far surpasses the benchmark value of 44.6, highlighting its exceptional tracking capability. This result suggests that substantially higher HOTA values could likely be achieved with improvements to the instance segmentation model alone.

Thesis 1d

I conducted an ablation study on the proposed architecture, separating the contributions of motility-based and morphology-based tracking, which the original architecture integrates seamlessly. The results revealed the individual impact of motility and morphology on tracking performance across the evaluated scenarios and highlighted the advantages of a tracking method that combines both. Furthermore, the proposed architecture consistently matched or outperformed the widely used Kalman filter in all scenarios. Notably, in the scenario with semi-randomized object motility, the architecture achieved a 3.49-fold improvement in association F-score, due to its ability to utilize all temporally local imaging information for tracking.

Corresponding publication: [Ar2]

Figure 8: **"Visual signaling" metric results**
Mean metric results of tracker models *Kalman*, *TS* and *TS-L2*, *TS-Shape* for scenario "Visual signaling" datasets *Synthetic Arrows*, *Synthetic Arrows TR-1* and *Synthetic Arrows TR-2*.

The modular design of the proposed architecture enabled the integrated implementation of different tracking models. I implemented two ablated variants of the proposed architecture, both using the same local trackers trained on the given dataset but modifying the global assignment step. In one variant, only the centroid Euclidean distance (L2) was considered during assignment, while

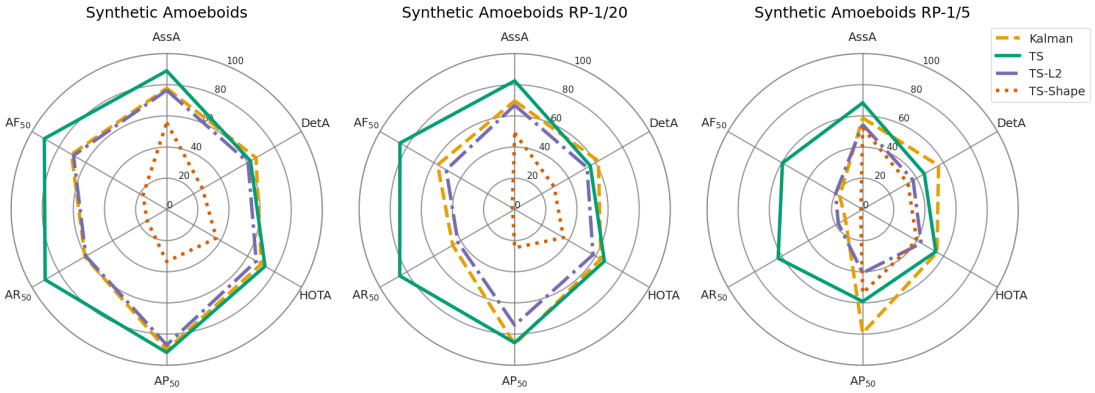


Figure 9: **"Semi-random positioning" metric results**
Mean metric results of tracker models *Kalman*, *TS*, *TS-L2* and *TS-Shape* for scenario "Semi-random positioning" datasets *Synthetic Amoeboids*, *Synthetic Amoeboids RP-1/20* and *Synthetic Amoeboids RP-1/5*.

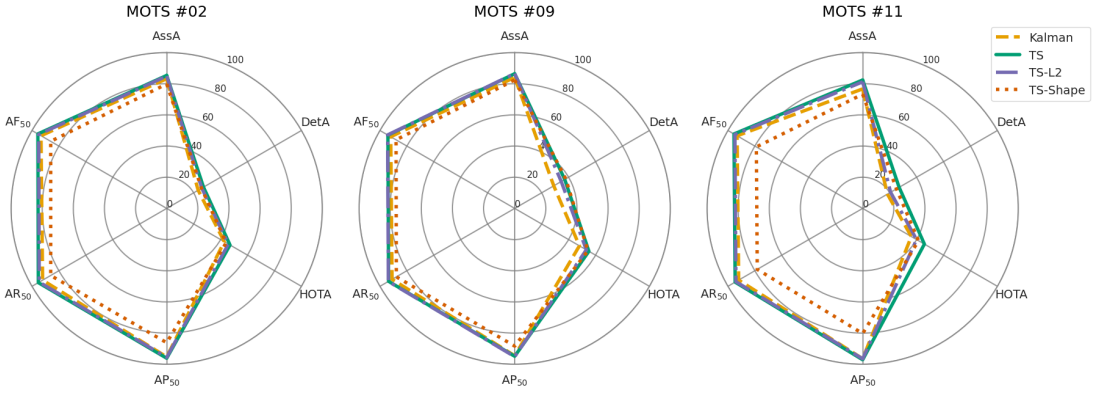


Figure 10: **MOTS challenge metric results**
Mean metric results of tracker models *Kalman*, *TS*, *TS-L2* and *TS-Shape* for the MOTS dataset samples.

in the other, only the centroid-aligned intersection over union (Shape) was used. Additionally, I implemented the widely used Kalman filter.

The "Visual signaling" and "Semi-random positioning" scenarios, as well as the MOTSynth-MOTS-CVPR22 dataset variant, were evaluated using the proposed time-symmetric tracking architecture (TS), its two ablated variants (TS-L2 and TS-Shape), and the Kalman filter. For these evaluations, I employed an improved metric system compared to the link-matching-based F-scores used for yeast cell tracking. While link-matching F-scores effectively signal tracking performance,

they are strongly correlated with segmentation performance, necessitating their combined interpretation. In contrast, for these scenarios, I measured association performance, where links were accounted for only when segmentation was successful on both sides, removing the correlation to better isolate the individual contribution of tracking. Metrics included association precision (AP_{50}), recall (AR_{50}), and F-scores (AF_{50}), all measured at a 50% IoU cutoff for segmentation correctness and scaled by 100 for consistency with the other statistics. These metrics provide a detailed insight into tracking performance. Additionally, I measured the bounding box-based industry standard DetA, AssA, and HOTA scores to provide a general performance assessment of detection, association, and overall performance, comparable to public benchmark results such as those on MOTSynth-MOTS-CVPR22.

The results for the "Visual signaling" scenario are presented in Fig. 8, detailing outcomes for no turns, infrequent turns (TR-1), and frequent turns (TR-2). The results for the "Semi-random positioning" scenario are shown in Fig. 9, covering no random movements, small random movements (RP-1/20), and large random movements (RP-1/5). The results for the MOTSynth-MOTS-CVPR22 dataset are presented in Fig. 10, showing performance for the three manually labeled evaluation samples individually. These results highlight the importance of the proposed tracker's ability to integrate positional and visual cues and demonstrate its superior performance compared to the Kalman filter, even in tasks for which it was neither specifically designed nor highly optimized. Notably, in the "Semi-random positioning" scenario, the proposed architecture achieved a 3.49-times higher association F-score (AF_{50}), reflecting a substantially better tracking performance for this particular task.

Thesis 2a

I demonstrated that widely used benchmark datasets, such as MS-COCO, exhibit an object positioning bias, strongly favoring objects located near the center of the image. This bias can result in prediction performance that is more than five orders of magnitude lower near the edges of the image, even when the objects are fully visible. The effects of this bias were analyzed for segmentation, detection and classification tasks, with a detailed localization of its impact based on prediction performance results and saliency maps.
Corresponding publication: [Ar3]

The foundation of most modern image-processing neural networks is the convolution operation, which is inherently translation-invariant. This property enables convolutional neural networks (CNNs) to exhibit translation-invariant behavior. While at the time of the publication this thesis point is based on, it was already known that the assumed translation-invariant behavior of CNNs does not always hold true, the underlying cause of this effect was not well understood, and the potential intensity and impact were not clearly demonstrated. Initial measurements I conducted revealed that training with object positioning biased towards the center of the images can have a profound impact on prediction performance near the edges. Specifically, these tests showed a more than 16,000-fold decrease in prediction performance, even when the objects remained fully within the field of view.

The significance of this effect is amplified by the strong object positioning bias present in popular and widely used benchmarks. This bias, being present in both training and testing datasets, remains undetectable during evaluation on the benchmark datasets. Its presence is only naturally revealed when the trained models partially fail in real-world applications lacking the bias.

To thoroughly investigate the extent and localization of this bias, I performed more detailed measurements across different image-processing tasks. For these experiments, I trained several U-net models on a controlled dataset using ImageNet samples as backgrounds to ensure the task remained non-trivial. Hand-drawn

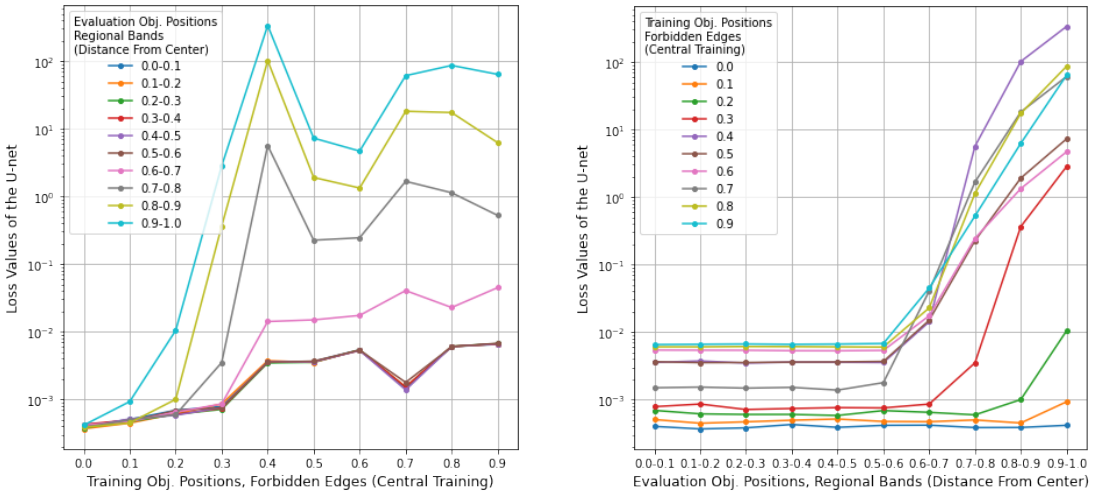
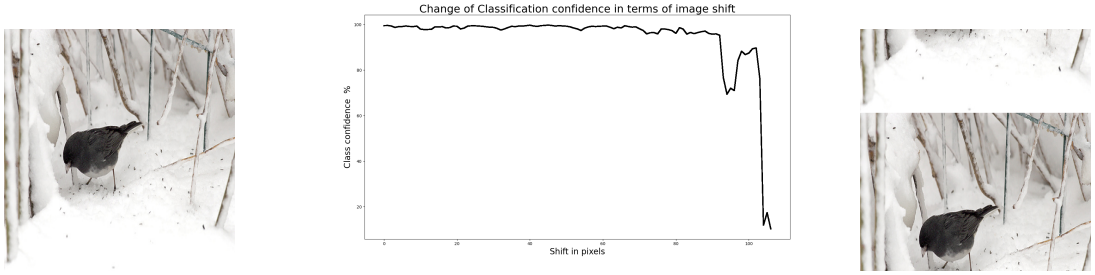


Figure 11: **Positional bias distributions**
Mean loss values of the U-Nets trained and evaluated on restricted regions. The loss values begin to increase drastically starting at the 0.5-0.6 evaluation band unless objects are specifically positioned near the edges during training. This indicates that more than 64 percent of an image will be segmented and classified with substantially worse performance in terms of area.



Junco Snowbird 74%

Junco Snowbird 0.7%
Ostrich 99%

Figure 12: **Classification bias sample**
Junco Snowbird sample showing the drastic effect of shifting objects to the boundary of the image on classification models, using the pretrained version of VGG-16 with ImageNet samples. The first column displays the original images which are classified correctly. The second column depicts how the classification confidence of the original class changes by shifting the image. The last column depicts the largest investigated shift. The classification confidence for the original class and the newly predicted class are displayed below the images along with their confidence scores.

MNIST numbers were overlaid onto these images for segmentation and classification, with their positions restricted to specific regions during training and

evaluation. During training, the objects were confined to centralized regions, while during evaluation, they were restricted to bands at specified distances from the image center. The results of these experiments are shown in Fig. 11.

Additionally, I assessed instance segmentation and classification tasks using pre-trained models, where objects were manually shifted closer to the image edges. A sample results for classification shown in Fig. 12. The results confirm that prediction performance drastically decreases near the edges of images for all evaluated tasks, unless objects were positioned near the image edges during training.

To further localize the bias and possibly reveal its cause, I calculated saliency map difference images based on shifted object positions using models trained with objects near the center or near the edge of the image on the previously described dataset with Image net samples as backround and MNIST samples as objects to be segmented and classified. Example results are presented in Fig. 13.

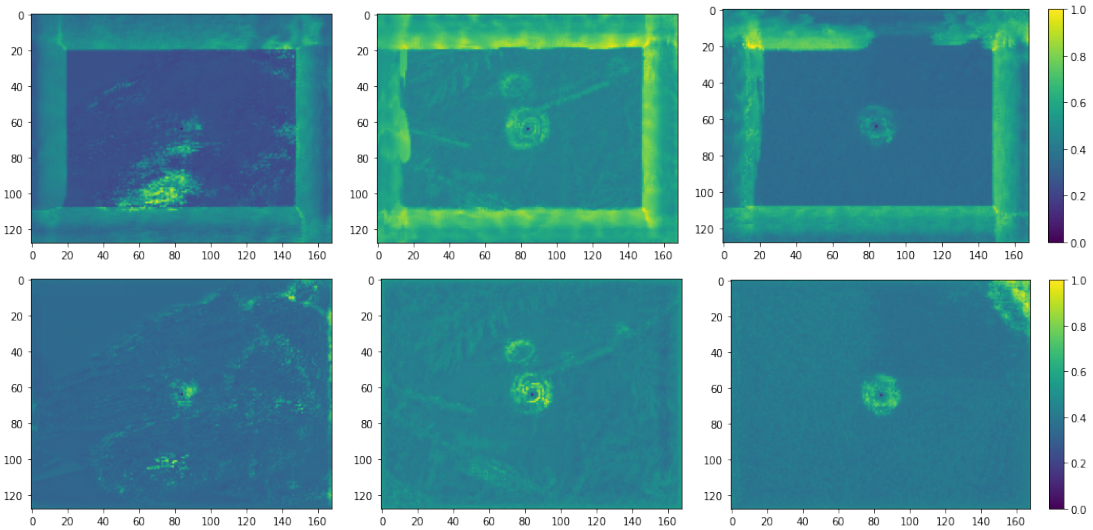


Figure 13: **Saliency map differences**
Normalized saliency map differences based on the shift vectors, with zero shift positioned at the center. The top row depicts the results using U-Nets trained with objects positioned in the middle, allowing for a central 30 percent, while the bottom row depicts the results using U-Nets trained with objects at the edges, prohibiting a central 70 percent. As observed, the CNNs trained on images containing objects solely at the center exhibit a distinct and large difference in their saliency maps when the objects are present at the edges. In contrast, the CNNs trained with objects positioned at the edges do not exhibit such differences.

Thesis 2b

I proposed architectural and data manipulation-based solutions to mitigate this bias. The most effective approach involved replacing zero-padding in all convolutional layers of the model with toroidal boundary conditions. This modification led to a performance improvement of more than 37,000 times in the most extreme cases. Corresponding publication: [Ar3]

Based on the measurement results described in Thesis 2a, I proposed several potential solutions to mitigate the object centering bias. Among these, two solutions demonstrated the best performance: randomly shifting training samples, even at the risk of partial object loss, and implementing toroidal boundary conditions in every convolutional layer. The latter introduces almost no drawbacks, aside from requiring an architectural modification.

	MS-COCO Orig	MS-COCO Shifted
Box mAP Orig.	33.6%	15.6%
Box mAP Shifted	31.7%	28.4%
Seg mAP Orig	. 31.4%	17.4%
Seg mAP Shifted	29.4%	25.3%

Table 2: **Detection bias mitigation**
Mean average precision (mAP) results for a Mask R-CNN network with a ResNet-50 backbone and a feature pyramid network with ROI align on the MS-COCO dataset. Two versions of the test set were evaluated: the original MS-COCO test set (MS-COCO Orig) and a modified version where a randomly selected object was always shifted to the boundary (MS-COCO Shifted). mAP results are reported for bounding box detection (Box) and instance segmentation tasks (Seg) under two different training conditions. Rows marked with (Orig.) show results on the unaltered MS-COCO dataset, while rows marked with (Shifted) show mAPs for networks trained on a dataset where a randomly selected object was always shifted to the boundary.

Results for classification, instance segmentation, and semantic segmentation tasks with classification, as described in Thesis 2a, were evaluated using these solutions either individually or in combination. These results are presented in Tab. 3, Tab. 2, and Fig. 14. The findings indicate that the use of toroidal boundary

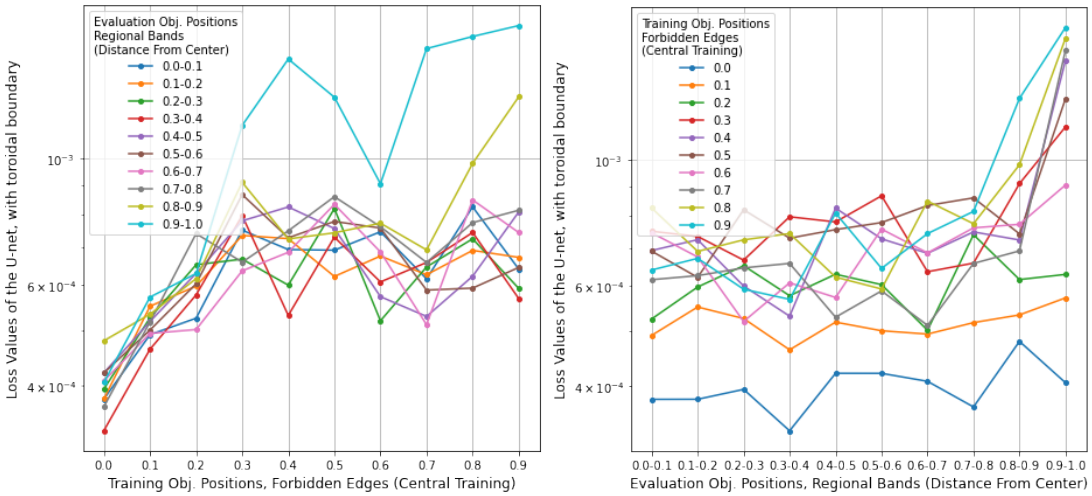


Figure 14: **Positional bias distributions - Toroidal boundary**
Mean loss values of the U-Nets trained on restricted regions, with toroidal boundary conditions in each convolutional layer. Compared to the values in Fig. 11, the relative increase of loss values based on training and evaluation positions is minimal.

conditions offers slightly superior performance compared to random shifts. This advantage likely arises because large random shifts can introduce new artifacts, such as partial removal of objects from the field of view. Nonetheless, both solutions far outperform the biased models and nearly eliminate the effects of the bias entirely.

Architecture	ImageNet	ImageNet Boundary
DenseNet121 Orig.	75.1%	3%
DenseNet121 Shifted	69.3%	66%
DenseNet121 Toroidal	74.2%	68%

Table 3: **Classification bias mitigation**
Top-1 test accuracies of DenseNet121 architecture on the original ImageNet test set (first column) and a manually created small subsample, where images have objects positioned at the boundary (second column). The model is evaluated in three variants: trained on the original ImageNet training set using zero-padding (Orig.), trained on a version of the dataset with objects shifted towards the boundaries (Shifted), and trained on the original dataset with toroidal boundary conditions (Toroidal).

Publications related to the thesis

[Ar1] G. Szabó, P. Bonaiuti, A. Ciliberto, and A. Horváth, "Enhancing cell tracking with a time-symmetric deep learning approach," *arXiv preprint arXiv:2308.03887*, 2023. (document)

[Ar2] G. Szabó, Z. Molnár, and A. Horváth (2024). "Post-Hoc MOTS: Exploring the Capabilities of Time-Symmetric Multi-Object Tracking", *arXiv preprint arXiv:2412.08313*, 2024 (document)

[Ar3] G. Szabó and A. Horváth, "Mitigating the bias of centered objects in common datasets," in *2022 26th International Conference on Pattern Recognition (ICPR)*, pp. 4786–4792, IEEE, 2022. (document)

Other publications of the author

[Au1] R. Bagdy-Bálint, G. Szabó, Ö. H. Zováthi, B. H. Zováthi, Á. Somorjai, C. Köpenczei, and N. K. Rózsa, "Accuracy of automated analysis in cephalometry," *Journal of Dental Sciences*, Elsevier, 2024.

[Au2] D. Babicz, S. Kontár, M. Peto, A. Fulop, G. Szabó, and A. Horváth, "Receptive field size optimization with continuous time pooling," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pp. 1449–1458, 2021.

[Au3] W. A. Bosbach, B. Németh, R. Zelei, J. F. Senge, B. Pasztor, L. Ebner, M. Szabo, S. Anderson, G. Szabo, P. Dlotko, A. Horvath, K. Daneshvar, and J. Heverhagen, "SCR 2023 poster presentation: An open access AI-based pattern recognition tool for application in MSK imaging – suitability and limitations of resources today [Poster]," *Swiss Congress of Radiology*, Zenodo, 2023. DOI: 10.5281/zenodo.7958527.

[Au4] G. Szabó, "Time-symmetric generalization of multi-object tracking," *PhD Proceedings Annual Issues of the Doctoral School Faculty of Information Technology and Bionics*, vol. 19, pp. 189–192, 2024. [Online]. Available: `https://m2.mtmt.hu/api/publication/35446463`.

[Au5] G. Szabó, "Time symmetric tracking of yeast cells using convolutional neural networks," *PhD Proceedings Annual Issues of the Doctoral School Faculty of Information Technology and Bionics*, vol. 18, pp. 145–148, 2023. [Online]. Available: `https://m2.mtmt.hu/api/publication/34176533`.

[Au6] G. Szabó, "The effects and mitigation of object centering bias in common datasets," *PhD Proceedings Annual Issues of the Doctoral School Faculty of Information Technology and Bionics*, vol. 17, pp. 191–194, 2022. [Online]. Available: `https://m2.mtmt.hu/api/publication/35446504`.

[Au7] G. Szabó, "Investigation of cell motility," *PhD Proceedings Annual Issues of the Doctoral School Faculty of Information Technology and Bionics*, vol. 16, pp. 169–172, 2021. [Online]. Available: `https://m2.mtmt.hu/api/publication/35446512`.