

David Exler*, Nils Friederich, Martin Krüger, John Jbeily, Mario Vitacolonna, Rüdiger Rudolf, Ralf Mikut, and Markus Reischl

Data Synthesis Improves 3D Myotube Instance Segmentation

<https://doi.org/10.1515/cdbme-2026-0134>

Abstract: Myotubes are multinucleated muscle fibers serving as key model systems for studying muscle physiology, disease mechanisms, and drug responses. Mechanistic studies and drug screening thereby rely on quantitative morphological readouts such as diameter, length, and branching degree, which in turn require precise three-dimensional instance segmentation. Yet established pretrained biomedical segmentation models fail to generalize to this domain due to the absence of large annotated myotube datasets. We introduce a geometry-driven synthesis pipeline that models individual myotubes via polynomial centerlines, locally varying radii, branching structures, and ellipsoidal end caps derived from real microscopy observations. Synthetic volumes are rendered with realistic noise, optical artifacts, and CycleGAN-based Domain Adaptation (DA). A compact 3D UNet with self-supervised encoder pretraining, trained exclusively on synthetic data, achieves a mean Injective Panoptic Quality (IPQ) of 0.22 on real data, significantly outperforming three established zero-shot segmentation models, demonstrating that geometry-driven synthesis enables effective instance segmentation in annotation-scarce biomedical domains.

Keywords: Myotube, Data Synthesis, Self Supervised Learning, Segmentation, 3D Data

1 Introduction

In this paper, we show that existing state-of-the-art biomedical segmentation models, such as CellposeSAM [8],

PlantSeg [13], and StarDist [11], fail to generalize to 3D myotube instance segmentation. These models are predominantly trained on compact, roughly spherical cell morphologies and lack exposure to the highly elongated, branching, and mutually overlapping geometry of myotubes. Manual annotation of 3D myotube data is particularly challenging: individual instances can span hundreds of voxels in length, frequently touch or cross neighboring fibers, and exhibit locally varying thickness and ellipsoidal nuclear bulges, making unambiguous delineation extremely laborious even for domain experts. Recent research aims to alleviate annotation efforts by introducing data synthesis pipelines [1, 2, 3]. As no large annotated 3D myotube datasets, nor a myotube simulation pipeline exist, we tackle this challenge by presenting the first data synthesis pipeline for 3D myotube data and training exclusively on synthetic data. We validate our approach by benchmarking a compact UNet [10] with a self-supervised learning (SSL) pretrained encoder against the three established models, contributing a foundation for scalable analysis of myotubes as key model systems in muscle development, disease modeling, and drug screening [6, 7, 9, 12].

2 Methods

2.1 Synthesis

We derive geometrical characteristics from real myotube data (see Figs. 1a, b) and generate synthetic annotation volumes in three stages. First, polynomial centerlines are sampled in XY and Z using damped Chebyshev basis functions T_k of degree K with coefficients $\tilde{c}_k \sim \mathcal{U}(-1, 1)$ scaled by $k^{-\alpha}$ to suppress boundary oscillations. Local thickness is modulated by a smooth polynomial and a sinusoidal term controlled by γ and δ . Second, straight branching segments are inserted while preserving slope continuity. Third, ellipsoidal structures are placed along the centerline, with hollow ellipsoids along the shaft and solid caps at endpoints. Realistic fluorescence appearance is simulated by additive Poisson and Gaussian noise, salt-and-pepper textures, debris artifacts, anisotropic halos, and Gaussian PSF blurring (Figs. 1c, d). A CycleGAN is trained on unpaired real and synthetic images to generate domain-adapted (DA) images [3].

*Corresponding author: David Exler, Institute for Automation and Applied Informatics, Karlsruhe Institute of Technology, Hermann-von-Helmholtz-Platz 1, 76344, Eggenstein Leopoldshafen, Germany, e-mail: david.exler@kit.edu
 Nils Friederich, Martin Krüger, Ralf Mikut, Markus Reischl, Institute for Automation and Applied Informatics, Karlsruhe Institute of Technology, Hermann-von-Helmholtz-Platz 1, 76344, Eggenstein Leopoldshafen, Germany
 Nils Friederich, Institute of Biological and Chemical Systems, Karlsruhe Institute of Technology, Hermann-von-Helmholtz-Platz 1, 76344, Eggenstein Leopoldshafen, Germany
 John Jbeily, Mario Vitacolonna, Rüdiger Rudolf, CeMOS Research and Transfer Center, Technische Hochschule Mannheim, Paul-Wittsack-Straße 10, 68163 Mannheim, Germany

2.2 Segmentation Models

We train two compact residual 3D UNets with four encoder stages, each comprising two convolutional layers, a residual connection, max pooling, and normalization. The decoder progressively upsamples via transposed convolutions with skip connections and residual refinement blocks. Because domain-specific SSL pretraining can benefit segmentation models, the encoder of one model is pretrained on real myotube data using Fully Convolutional Masked Autoencoding (FCMAE) [14]: 3D input patches are randomly masked, and the network learns to reconstruct the missing regions. This SSL variant extends the encoder with additional trainable layers beyond the four shared stages. The other SSL-pretrained encoder layers are frozen during downstream optimization. We train four model variants: (1) UNet, (2) UNet on DA data, (3) SSL-pretrained UNet, and (4) SSL-pretrained UNet on DA data. Two output channels are learned to predict the foreground and centerline probabilities. Thresholding the foreground probability predicts a binary mask, and a seeded Watershed algorithm with the thresholded centerline probability separates the instances. We benchmark our model against three established zero-shot biomedical segmentation models (see Table 1).

Tab. 1: Segmentation models evaluated in this study. Three established zero-shot biomedical segmentation models are compared against our models trained exclusively on synthetic data. Training duration (Train) and inference duration (Infer.) per 3D image are reported. DA: domain-adapted training data, *: our models.

Model	Abbrev.	Param.	Infer.	Train
CellposeSAM	CSAM	304.6M	800s	
PlantSeg	psG	4M	55s	
StarDist	SD	1.4M	120s	
UNet*	UNet	5M	0.8s	20s
UNet DA*	UNet (DA)	5M	0.8s	20s
SSL UNet*	SSL	23M	1.0s	25s
SSL UNet DA*	SSL (DA)	23M	1.0s	25s

2.3 Dataset

We generate 200 individual synthetic instances across 30 images of $128 \times 1024 \times 1024$ pixels for our training dataset. To validate our synthesis pipeline, we manually annotate a subset of a new 3D myotube fluorescence microscopy dataset that we acquire according to the protocol reported in [4].

This dataset contains 17 images of shape $(z, 1024, 1024)$, with z ranging from 24 to 128. Due to the challenges described above, we perform sparse annotation, labeling $n = 40$ individual myotube instances across 17 volumetric stacks.

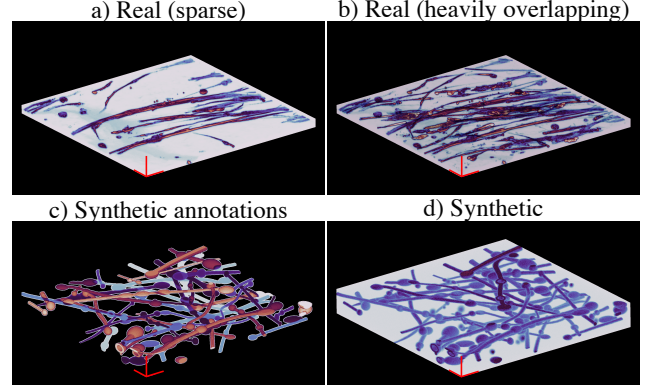


Fig. 1: 3D renderings illustrating the synthesis pipeline. Real myotube data ranges from sparsely distributed (a) to heavily overlapping configurations (b) from our new dataset. The pipeline generates instance-labeled annotation volumes (c) and corresponding synthetic fluorescence image stacks (d) that closely approximate real imaging conditions.

Instances are selected to cover a representative range of morphological configurations, including straight, curved, and branching myotubes, as well as varying degrees of neighbor overlap. These annotations serve exclusively as an independent test set; no annotated data is used during training.

3 Results & Discussion

3.1 Visual Inspection

Fig. 2 shows a representative 2D slice of the real dataset (a) alongside the sparse manual annotations (b) of this slice and the predictions from all models. CSAM (c) merges multiple instances into large, irregular blobs and responds strongly to background halo artifacts. psG (d) segments numerous small round regions, reflecting its training bias toward compact cell morphologies. While SD (e) covers the present myotubes well, it produces fragmented oval pieces instead of capturing the elongated myotube geometry. Marker 1) shows two round nucleus artifacts in the real image. All zero-shot models falsely predict these artifacts to be myotubes (marker 2). In contrast, our models (f-i) successfully detect the elongated structures, though it partially predicts shadow artifacts as foreground. The domain-adapted UNet (g) further reduces these false positives, producing cleaner foreground predictions that match the manual annotations (b) more closely. SSL DA (i) predicts minimal false positives. It is the only model avoiding the nucleus artifacts (marker 3). Fig. 3 provides a closer inspection of a single myotube instance featuring a thin, elongated body, a pronounced nuclear bulge, and a tapering distal end (a). CSAM (b) produces disconnected fragments at incorrect locations.

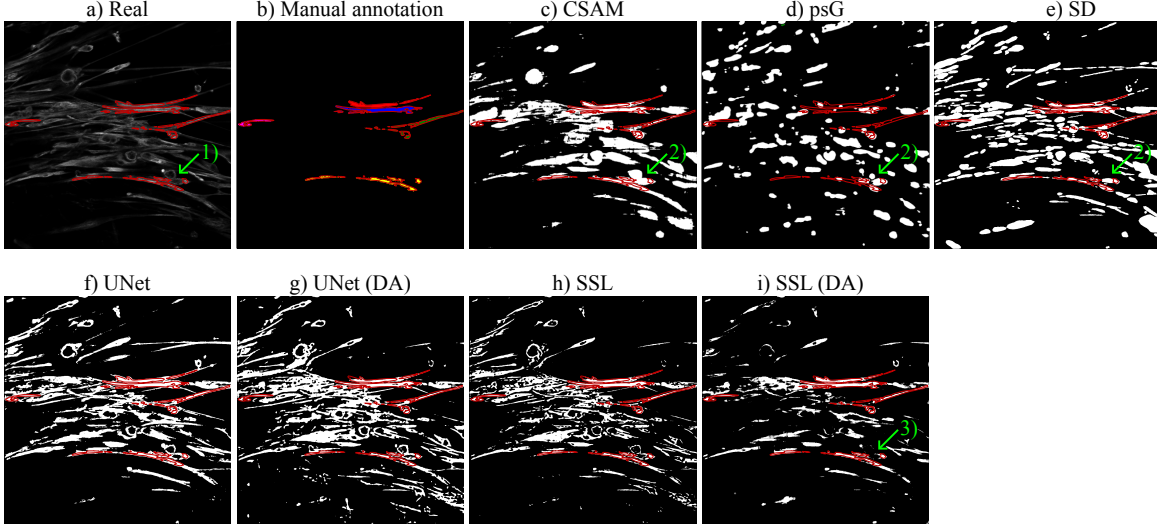


Fig. 2: Examples of 2D slices. a): Real, b): Corresponding sparse manual annotation, c)-e): Segmentation masks predicted by the pre-trained segmentation models, and f)-i): Segmentation masks by our models. Contours of the annotations are red in every figure.

psG (c) detects only the bright nuclear bulge while entirely missing the elongated body. SD (d) splits the instance into three separate fragments, recovering parts of the geometry but lacking continuity along the myotube axis. Both our UNet models (e-f) correctly capture both the elongated shaft and the nuclear bulge, with only the thin distal tip partially missed. The SSL prediction (g) also captures these relevant features, but is noisy on the edges. The domain-adapted SSL (h) produces the closest match to the real instance, preserving the overall shape, including the nuclear bulge with clean boundaries and correct identification of the z-offset in the middle of the body.

3.2 Quantitative Comparison

We evaluate all models using the Injective Panoptic Quality (IPQ) [5], which jointly measures segmentation accuracy and instance detection while penalizing incorrect splits, with values ranging from 0 to 1. Since only a subset of instances is annotated, the recall component is disabled. Standard metrics such as Dice or PQ assume complete annotations and would unfairly penalize models for detecting unannotated instances, making IPQ the appropriate choice here. An initial evaluation of our models on synthetic data shows $IPQ > 0.9$ for every model. Fig. 4 reports the mean IPQ with standard error over $n = 40$ annotated instances, with significance assessed via paired t-tests across instances. All of our models score significantly higher than all three zero-shot baselines ($p < 0.001$), despite having far fewer parameters than CSAM. Our domain-adapted SSL model achieves a mean IPQ of 0.22, demonstrating the substantial impact of our data synthesis pipeline.

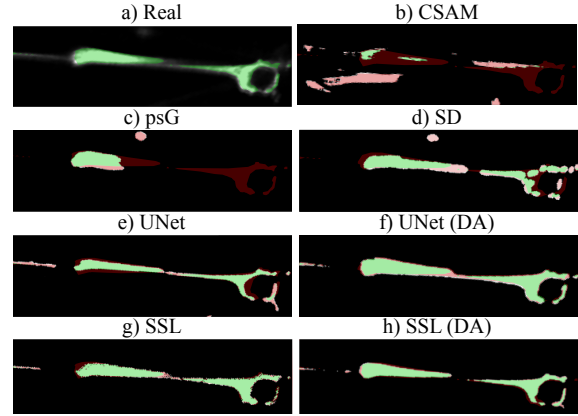


Fig. 3: Examples of another zoomed-in 2D slice. a): Real, b)-h): Segmentation masks predicted for this instance by the segmentation models included in this study. Green areas show correct segmentation, red areas are false positive or negative predictions.

The remaining gap to a perfect score reflects the domain shift between synthetic training data and real images, which we identify as the primary direction for future improvement.

3.3 Ablation Study

Because the results of Fig. 4 show large performance gaps between the SSL (DA) model (0.22) and our other models (0.034 - 0.039), we disentangle the contributions of DA and SSL pre-training by comparing our four model variants (Table 1). SSL pretraining without domain-adapted downstream data yields a marginal decline (0.039 to 0.038), suggesting that the pre-trained features align only with the real-domain.

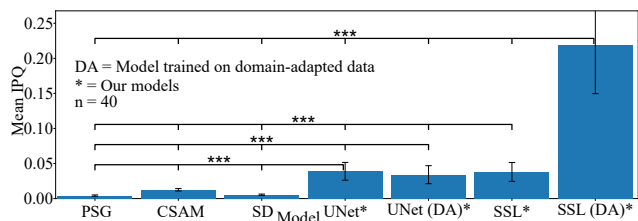


Fig. 4: Mean IPQ scores with standard error of the mean over $n = 40$ annotated instances for all evaluated models. Significance of pairwise differences between our best model (SSL (DA)) and each baseline was assessed via paired t-tests ($*** p < 0.001$). Our models (marked with *) consistently outperform all three zero-shot baselines despite having far fewer parameters. DA denotes training on domain-adapted synthetic data.

The features then transfer poorly under synthetic-domain downstream training. In contrast, combining SSL with DA leads to a substantial performance gain (0.22), outperforming all other variants by a large margin. Without a strong feature initialization, this additional appearance complexity can hinder learning of the underlying geometry and increase the risk of overfitting to GAN-specific artifacts, explaining the marginal performance drop in UNet (DA) relative to UNet. However, when combined with SSL pretraining, the encoder already provides robust real-domain feature representations, allowing the model to effectively exploit the reduced domain gap introduced by DA, which further explains the performance gap of SSL DA.

4 Conclusion

We present the first geometry-driven synthesis pipeline for 3D myotube data, enabling instance segmentation in a domain where no annotated training data exists. A compact SSL-pretrained UNet trained exclusively on synthetic data significantly outperforms three established zero-shot biomedical segmentation models, demonstrating that controlled geometry-driven synthesis can substitute for costly manual annotations. The approach generalizes to similar annotation-scarce domains such as 3D nerve fiber or mycelium networks. Future work will focus on closing the remaining domain gap, incorporating disease-specific morphological variations, and scaling toward a large, well-performing instance segmentation model for myotube data. The synthesis pipeline is publicly available at github.com/DavidExler/syn_myo.

Author Statement

The Helmholtz Association funds this project under the research school "Helmholtz Information and Data Science School for Health (HIDSS4Health)", and the program "Natu-

ral, Artificial and Cognitive Information Processing (NACIP)". RR was funded by the Baden-Württemberg-Stiftung, grant LSE-012, and the DFG, grant INST874/9-1. The authors state no conflict of interest. Informed consent has been obtained from all individuals included in this study. Ethical approval: Does not apply.

References

- [1] M. Böhland et al. "Unsupervised GAN epoch selection for biomedical data synthesis." In: *Current Directions in Biomedical Engineering*. Vol. 9. 1. De Gruyter. 2023, pp. 467–470.
- [2] R. Bruch et al. "Improving 3D deep learning segmentation with biophysically motivated cell synthesis." In: *Communications Biology* 8.1 (2025), p. 43.
- [3] R. Bruch et al. "Synthesis of large scale 3D microscopic images of 3D cell cultures for training and benchmarking." In: *PLOS ONE* 18.3 (2023), e0283828.
- [4] N. Couturier et al. "Aberrant evoked calcium signaling and nAChR cluster morphology in a SOD1 D90A hiPSC-derived neuromuscular model." In: *Frontiers in Cell and Developmental Biology* 12 (2024), p. 1429759.
- [5] D. Exler et al. "Bayesian Optimization for Design Parameters of 3D Image Data Analysis." In: *arXiv preprint arXiv:2602.15660* (2026).
- [6] S. F. Gilbert. *Developmental biology*. 10th ed. Sunderland, Massachusetts: Sinauer Associates, 2014.
- [7] W. H. Lewis and M. R. Lewis. "Behavior of cross striated muscle in tissue cultures." In: *American Journal of Anatomy* 22.2 (1917), pp. 169–194.
- [8] M. Pachitariu, M. Rariden, and C. Stringer. "Cellpose-SAM: superhuman generalization for cellular segmentation." In: *bioRxiv* (2025).
- [9] S. Park et al. "Quantitative analysis of the dexamethasone side effect on human-derived young and aged skeletal muscle by myotube and nuclei segmentation using deep learning." In: *Bioinformatics* 41.1 (2024), btae658.
- [10] O. Ronneberger, P. Fischer, and T. Brox. "U-Net: Convolutional Networks for Biomedical Image Segmentation." In: *Medical Image Computing and Computer-Assisted Intervention*. 2015, pp. 234–241.
- [11] M. Weigert et al. "Star-convex Polyhedra for 3D Object Detection and Segmentation in Microscopy." In: *The IEEE Winter Conference on Applications of Computer Vision (WACV)*. 2020.
- [12] A. Weisrock et al. "MyoFlnDer: an AI-based tool for myotube fusion index determination." In: *Tissue Engineering Part A* 30.19-20 (2024), pp. 652–661.
- [13] A. Wolny et al. "Accurate and versatile 3D segmentation of plant tissues at cellular resolution." In: *eLife* 9 (2020), e57613.
- [14] S. Woo et al. "ConvNeXt V2: Co-designing and scaling convnets with masked autoencoders." In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2023, pp. 16133–16142.