

<https://doi.org/10.1038/s42003-025-07469-2>

# Improving 3D deep learning segmentation with biophysically motivated cell synthesis



Roman Bruch<sup>1</sup>✉, Mario Vitacolonna<sup>2,3</sup>, Elina Nürnberg<sup>2,3</sup>, Simeon Sauer<sup>2,4</sup>, Rüdiger Rudolf<sup>2,3</sup> & Markus Reischl<sup>1</sup>

Biomedical research increasingly relies on three-dimensional (3D) cell culture models and artificial-intelligence-based analysis can potentially facilitate a detailed and accurate feature extraction on a single-cell level. However, this requires for a precise segmentation of 3D cell datasets, which in turn demands high-quality ground truth for training. Manual annotation, the gold standard for ground truth data, is too time-consuming and thus not feasible for the generation of large 3D training datasets. To address this, we present a framework for generating 3D training data, which integrates biophysical modeling for realistic cell shape and alignment. Our approach allows the *in silico* generation of coherent membrane and nuclei signals, that enable the training of segmentation models utilizing both channels for improved performance. Furthermore, we present a generative adversarial network (GAN) training scheme that generates not only image data but also matching labels. Quantitative evaluation shows superior performance of biophysical motivated synthetic training data, even outperforming manual annotation and pretrained models. This underscores the potential of incorporating biophysical modeling for enhancing synthetic training data quality.

Biomedical and pharmaceutical research rely increasingly on three-dimensional cell culture models, such as spheroids and organoids. Leveraging optical tissue clearing alongside 3D-microscopy and AI-based image analysis enables the extraction of unprecedented detail at the single-cell level in 3D. As such, regional analysis of events like cell division and cell death can be performed. Most commonly, AI-based image analysis requires the manual generation of annotated datasets for model training, evaluation and subsequent model selection. To reach this surplus of information from 3D-cell culture whole mounts, the segmentation of the original 3D datasets needs to be of maximal precision, which requests the availability of a sufficient amount of high-quality ground truth datasets. Additionally, for accurate segmentation of rare cell states, it is important that such objects are represented within the dataset. While manual expert annotation is often considered the gold standard for ground truth data, its labor-intensive nature, especially for voluminous 3D data, makes it impractical, usually taking long time periods to create a sufficient dataset. Therefore, new and realistic modes of generating 3D-ground-truth data are needed.

Several methods have been proposed for the synthesis of 2D and 3D training data for nuclei<sup>1–7</sup> and cell<sup>8,9</sup> segmentation. Most of these approaches employ generative adversarial networks (GANs). An overview of different

GAN variants is provided in ref. 8. We have recently presented SimOptiGAN, a method to create training data based on a combination of simulation and deep-learning-based optimization, allowing the introduction of rare elements. It generates data with a few real-world nuclei and their subsequent assembly in virtual spheroids, including the implementation of synthetic optical features, such as noise, signal loss in depth, and point-spread function before optimization<sup>3</sup>. This gave surprisingly good results in terms of training efficacy. However, the method placed nuclei in a rather arbitrary manner, leaving doubts concerning the accuracy compared to the real-world distribution. For example, in cancer cell spheroids, dividing cells are often more concentrated toward the border of the 3D culture, while necrotic or apoptotic cells are more abundant toward the spheroid center. Moreover, cells oftentimes exhibit a preference for tangential orientation relative to the outer hull<sup>9</sup>. Therefore, a consideration of biophysical features and real-world distribution of cells with specific features might further increase the fidelity of synthetic datasets. This enhancement could furthermore foster greater consistency between the synthetic and real-image domains, thereby mitigating the risk of unintentionally generating unwanted structures through deep-learning-based transformations introducing inconsistencies between image and label data<sup>6,10</sup>. However, correct

<sup>1</sup>Institute for Automation and Applied Informatics, Karlsruhe Institute of Technology, Eggenstein-Leopoldshafen, Germany. <sup>2</sup>Institute of Molecular and Cell Biology, Mannheim University of Applied Sciences, Mannheim, Germany. <sup>3</sup>CeMOS, Mannheim University of Applied Sciences, Mannheim, Germany. <sup>4</sup>CHARISMA, Mannheim University of Applied Sciences, Mannheim, Germany. ✉e-mail: [roman.bruch@kit.edu](mailto:roman.bruch@kit.edu)

physical modeling of cell structures is a difficult and complex task. To this date, no existing method incorporates physical modeling to enhance the organization of cell and nuclei structures.

Several approaches have been developed to simulate 3D cell cultures in a biophysically realistic manner, each tailored to various scientific approaches for studying a broad spectrum of biological phenomena<sup>11–15</sup>. In particular, a widely used method to simulate the temporal evolution of cell morphology in 2D- or 3D-cell cultures is the Cellular Potts Model (CPM)<sup>16,17</sup>. Utilizing a grid-based system of pixels (2D) or voxels (3D), the CPM assigns each unit to specific cells or the extracellular medium. By modeling the total free energy of the system and applying the Metropolis algorithm in a Monte Carlo simulation, the initial cell shapes evolve toward their thermodynamic equilibrium<sup>16–18</sup>. This technique is used in exploring a variety of biological processes, including viral infection<sup>19</sup>, vascularization<sup>20,21</sup>, cell sorting and migration<sup>22–28</sup>, morphological changes<sup>29–31</sup> and cancer research<sup>32–41</sup>. The application of the CPM or alternative biophysically driven models to replicate the morphology of cell cultures for the creation of synthetic training data, as detailed in this study, has not yet been reported in scientific literature.

To improve the cell arrangement in synthetic image data, we introduce a framework to generate 3D training data of nuclei and membrane signals, integrating biophysical modeling to achieve realistic cell alignment and orientation (see Fig. 1). Unlike existing approaches, our framework facilitates the creation of coherent membrane and nuclei signals, enabling the training of segmentation models utilizing both signals simultaneously for improved segmentation performance. Three approaches for generating nuclei signals (SimOptiGAN+, Mem2NucGAN-P, Mem2NucGAN-U) are presented and compared against a previously developed synthesis method, SimOptiGAN. The random nuclei placement process of SimOptiGAN is improved in SimOptiGAN+ by utilizing simulated cell borders to control

the nuclei placement process. Mem2NucGAN-P and Mem2NucGAN-U utilize generative adversarial networks (GANs) to generate nuclei signals based on the simulated cell borders. For this, a training and post-processing scheme is introduced, allowing the additional generation of nuclei labels corresponding to the generated signal images. This adaptation can also be utilized for other image modalities. For Mem2NucGAN-P, the training is performed with paired images, while Mem2NucGAN-U utilizes unpaired images. We furthermore prove superior performance of synthetic nuclei training data against manually annotated data and a pretrained Cellpose model, using three manually corrected ground truth patches for comparison. To the best of our knowledge, this is also the first time that synthetic training data of SimOptiGAN is evaluated on fully annotated image data.

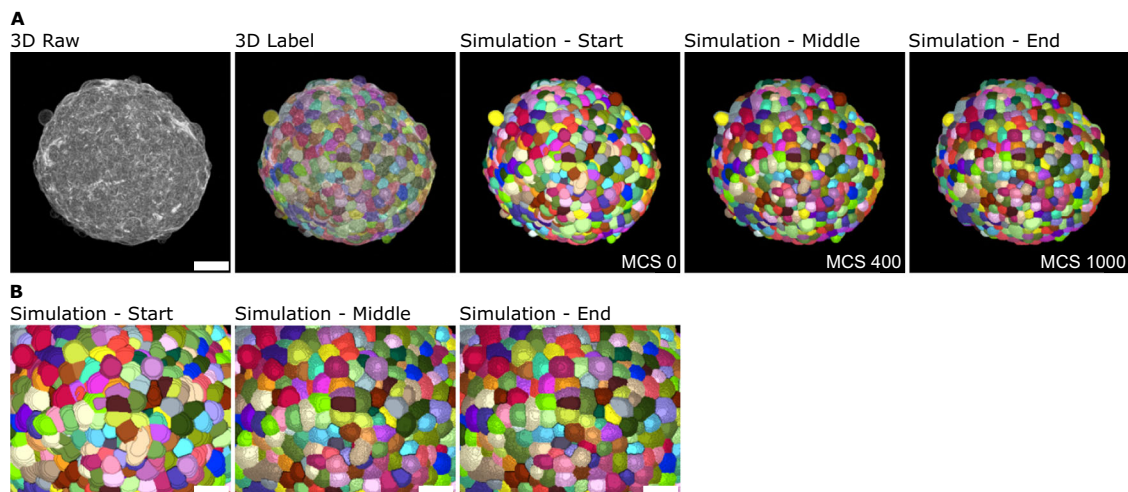
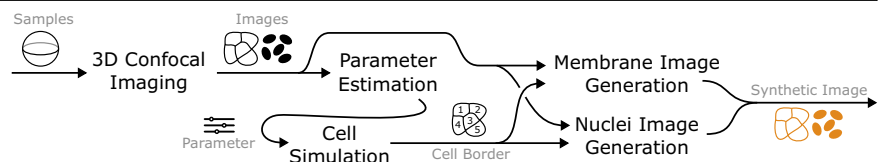
## Results

### Biophysical modeling enables marker transformation

To integrate meaningful, biophysical information into the generation of synthetic images, CPM simulations were developed using 3D label masks obtained from automated whole mount segmentation of real image data (Fig. 2). The detection of membranes relied on minimal training on single optical planes, acknowledging a balance between invested time and accepting some inaccuracies in boundary detection. Subsequently, morphological features such as volume, surface area, and shape descriptors were extracted from individual labels to analyze the distribution of cellular characteristics across the spheroid. Label masks obtained from whole mount spheroids were then transformed into a starting configuration for a CPM simulation. Furthermore, each cell's target volume and surface area were preset to their individual starting values.

Before simulating the entire spheroid, a parameter estimation was conducted on a small subsection to identify optimal combinations of contact energies  $J^{(c-c)}$  and  $J^{(c-m)}$ , as well as the volume and surface constraints  $\lambda_V$  and

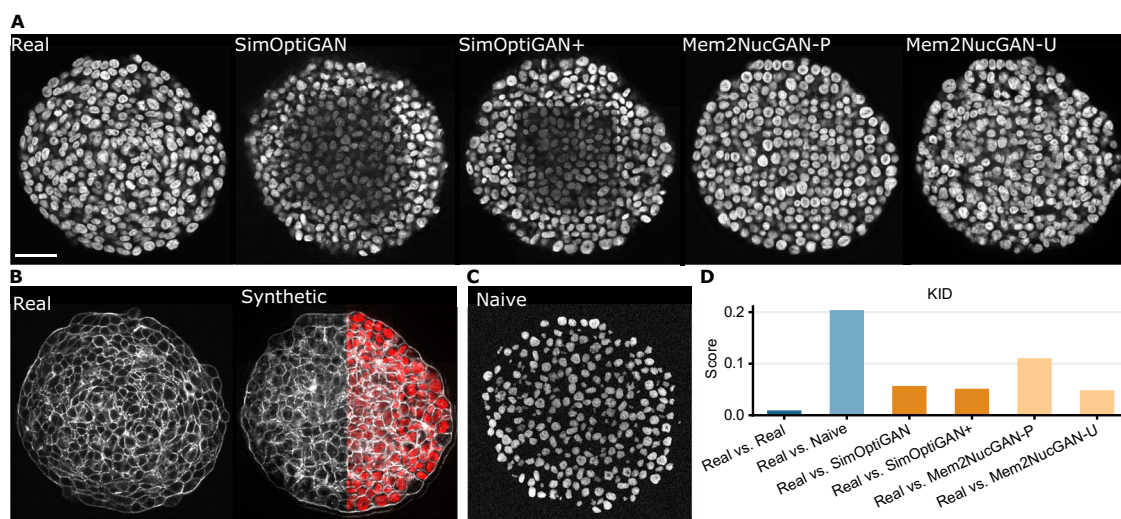
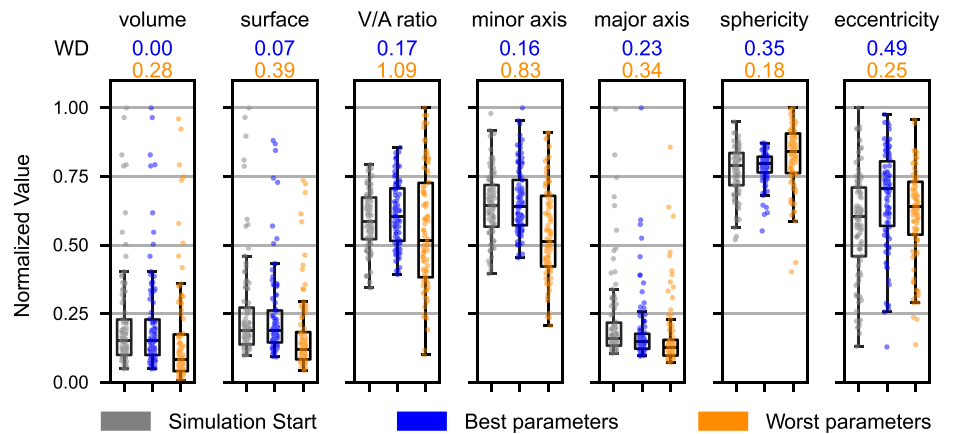
**Fig. 1 | Presented workflow for generation of 3D synthetic images featuring nuclei and membrane signals.** 3D cell cultures are imaged using confocal microscopy. Afterward, parameters are extracted based on the recorded images, which are then utilized during cell simulation to generate synthetic cell border images. Finally, synthetic nuclei and membrane images are generated on the basis of the simulated cell border image.



**Fig. 2 | Biophysically driven 3D spheroid simulations from real-image data reveal cell morphology variability across Monte Carlo steps (MCS), while preserving the distribution of morphological features throughout the spheroid.** **A** 3D maximum projection of membrane marker SiR-actin (gray, far left), overlay with label masks

(second from the left), and snapshots of the corresponding CPM simulation of morphological changes at the start (middle, 0 MCS), middle (second from the right, 400 MCS) and end (far right, 1000 MCS) of the simulation. **B** Zoom images of the simulation steps shown in (A). Scale bars: **A** 25  $\mu\text{m}$ , **B** 15  $\mu\text{m}$ .

**Fig. 3 | Comparison of morphological feature distributions at the start and end of simulations reflecting changes in cell morphology, for the best and worst parameter sets according to the metric  $m$ , Eq. (3).** Data is derived from a manually segmented image patch. Individual Wasserstein distances, quantifying the variations between the simulation's start and end for each feature, are marked above the plots for the best (blue,  $\lambda_V = 10.0$ ,  $\lambda_A = 0.001$ ,  $f^{(c-c)} = 2.0$ ,  $f^{(c-m)} = 55.0$ ) and worst (orange,  $\lambda_V = 0.001$ ,  $\lambda_A = 10.0$ ,  $f^{(c-c)} = 10.0$ ,  $f^{(c-m)} = 10.0$ ) parameter sets. Boxes indicate median and quartiles of data, while whiskers encompass all values within 1.5 times the IQR. All individual data points are shown.



**Fig. 4 | Comparison of real and synthetic images.** **A** Image slices of real and synthetic 3D nuclei images. SimOptiGAN uses a random process for nuclei arrangement, while SimOptiGAN+, Mem2NucGAN-P, and Mem2NucGAN-U incorporate biophysical modeling for a realistic arrangement. **B** Image slices of real and synthetic membrane signals. The synthetic membrane signal is generated, as described in the “Methods” section, based on the same simulated cell borders used in the nuclei synthesis methods SimOptiGAN+, Mem2NucGAN-P, and Mem2NucGAN-U. Consequently, the membrane signal exhibits a consistent cell

arrangement, as demonstrated by the overlay of synthetic membrane with nuclei generated using SimOptiGAN+. **C** A preview of naively generated data used as the worst example for KID evaluation in **(D)**. These naive images are preliminary outputs of SimOptiGAN, generated by the simulation pipeline before undergoing deep learning optimization. **D** Comparison of synthetic nuclei images with real counterparts based on the Kernel Inception Distance (KID). Lower scores represent a greater similarity between real and synthetic signals. Scale bar: 50  $\mu\text{m}$ .

$\lambda_A$ . These parameters influenced cellular morphology, enabling changes in shape and position while preserving the overall distribution of features within the simulated spheroid. Evaluation of individual parameter combinations was performed by minimizing a metric  $m$  that compares the starting and ending points of the simulation by using the Wasserstein distance of the morphological feature distribution  $W_k$  and the mean intersection over union (IoU) of all cells, as described in “Methods”.

Figure 3 displays the distribution of selected morphological features across individual cells at the start of the simulation based on real-world data. It also presents the results for these features after 1000 Monte Carlo steps (MCS), comparing two sets of parameters that received the highest and lowest rankings in the parameter scans.

After parameter estimation, sets of parameters resulting in the smallest values of metric  $m$  were selected for simulations of whole spheroids. These simulations were then conducted for 1000 MCS. As a result of the simulations, 3D cell border images were received, where each cell is represented as a set of voxels with the same intensity. Representative 3D images of an HT-29 spheroid, including fluorescent membrane markers and a segmented

label mask, along with the corresponding 3D simulation images from the start (0 MCS), middle (400 MCS), and end (1000 MCS) of the simulation, are displayed in Fig. 2A. Zoomed images from the respective time points of the simulation highlight the morphological changes of individual cells throughout the simulation.

### Synthetic data matches structure of real counterparts

A human approach for the quality evaluation of synthetic image data is the visual comparison with their real counterpart. However, the visual differences perceived by a human may differ from those that are relevant to segmentation models. The Kernel Inception Distance (KID)<sup>42</sup> allows the comparison of images in the feature space of a network, emphasizing differences more relevant for the segmentation task. Lower distance values indicate a closer similarity between real and synthetic images, representing a better result. Figure 4A shows 2D samples of synthetic image data generated with newly introduced methods SimOptiGAN+, Mem2NucGAN-P, and Mem2NucGAN-U as well as a previous method, SimOptiGAN, which lacks biophysical modeling, compared with a real counterpart. Additionally, the



KID between generated 3D images and real counterparts is provided (Fig. 4D). For comparison, the KID between real samples, indicating the best possible result, and between real and naively generated 3D images, indicating a suboptimal result, is also shown (Fig. 4C). The procedure for generating the naive data is described in the “Methods: KID” section.

By visual comparison, the synthesis methods showed differences in nuclei morphology and arrangement, as well as in brightness and texture. The nuclei in the data generated with SimOptiGAN, i.e., without biophysical modeling, appeared less sharp and featured a more uniform texture. By incorporating biophysical modeling (SimOptiGAN+), the nuclei arrangement became more structured, with features like small holes in the spheroid. Both images, i.e., those of SimOptiGAN and SimOptiGAN+, also exhibited reduced brightness in the center region. However, the generated image with SimOptiGAN+, shows a hard patch transition in the center region, giving the impression of a rectangle-shaped brightness reduction. This hard transition may result from using the already trained optimization model of SimOptiGAN. Combined with slight variations in image statistics between training and inference data, hard transitions are more likely to appear. Data generated with Mem2NucGAN-P displayed rather large and more roundish nuclei. Furthermore, a light checkerboard texture pattern was visible. Data generated with Mem2NucGAN-U resulted in the smallest KID when compared to real data. However, some nuclei morphed into each other, showing no clear outline between them.

Figure 4B shows a synthetic membrane generated on the same biophysical simulated cell borders as the nuclei images generated with SimOptiGAN+, Mem2NucGAN-P and Mem2NucGAN-U (Fig. 4A). Both signals feature the same cell arrangement and can be combined. This is illustrated by overlaying synthetic membrane signals with nuclei signals generated using SimOptiGAN+ (Fig. 4B).

The comparison of the generated images using the KID measure (Fig. 4D) shows only small differences between SimOptiGAN, SimOptiGAN+, and Mem2NucGAN-U compared to the difference with Mem2NucGAN-P. All methods, however, show drastically lower KID scores compared to a naive approach. Apart from the central dark regions seen in the nuclei images by SimOptiGAN and SimOptiGAN+, the KID measure appeared to correlate well with the visual impression.

### Synthetic data can outperform manual annotation and universal models in terms of segmentation performance

The segmentation performance of the introduced methods was evaluated using the SEG (segmentation) and DET (detection) metrics as utilized in the Cell Tracking Challenge<sup>43</sup>. While the SEG metric focuses on the overlap between segmentation and ground truth, the DET metric evaluates the results on a cell level, i.e., whether an object is correctly detected or not. To compare the segmentation performance of different training datasets, the 3D StarDist<sup>44</sup> segmentation model was used. The performance was calculated based on three ground truth images of size  $25 \times 128 \times 128 \text{ px}^3$  to  $51 \times 128 \times 128 \text{ px}^3$  containing a total of 1001 nuclei. Approximately 37.5 h were required for ground truth generation. Each segmentation model training, except for the pretrained Cellpose nuclei model, was repeated six times, and the mean values of the resulting metrics are reported. Figure 5A shows the obtained segmentation performances for the introduced training data generation methods.

The model labeled as Maximum was trained on the ground truth dataset, later used for testing. This model was considered as a maximum of the segmentation performance for the StarDist architecture on this dataset, as train and test data were identical. However, generating training data through manual annotation of 3D image data is a challenging and time-consuming process. Cross-validation indicates results obtained by using a leave-one-out approach on the ground truth dataset, i.e., using each patch as the test set once while training the model on the remaining two patches. For inference of the Cellpose nuclei model<sup>45</sup>, only the nuclei size parameter was adapted to match the size of the ground truth data. The remaining models were all trained with synthetic datasets. SimOptiGAN was generated using our preliminary synthesis method presented in ref. 3, which does not

incorporate biophysical modeling. Conversely, the datasets labeled SimOptiGAN+, Mem2NucGAN-P, and Mem2NucGAN-U incorporate biophysical modeling. To demonstrate the superiority of SimOptiGAN(+) approaches utilizing the hybrid approach of imaging simulation and GAN-based optimization, a comparison was performed with a direct transformation of label data. To ensure a fair comparison, noise and brightness reduction were applied to the label data of SimOptiGAN prior to GAN training and inference. This dataset is referred to as binary transformation.

The results revealed significant variations in the segmentation performance among the different models. The maximum SEG score for this data was 0.593 (Maximum). Using Cross-validation, the score dropped to 0.505. Cellpose Nuclei attained a SEG score of 0.332. The SimOptiGAN model yielded a SEG score of 0.495 which is improved to 0.534 - the best score apart from Maximum - by incorporating biophysical modeling (SimOptiGAN+). Both Mem2NucGAN-P and Mem2NucGAN-U demonstrated notably lower SEG scores, with values of 0.044 and 0.103, respectively. Using SimOptiGAN's labels as the basis for the transformation (binary transformation), a score of 0.291 was achieved.

The DET scores across all models exhibited a slight increase compared to the SEG scores, given that only detection errors are penalized. Nonetheless, the disparity between the models remained consistent. The Maximum and Cross-Validation scores for the DET metric were 0.891 and 0.785, respectively. The Cellpose Nuclei model attained a DET score of 0.512. SimOptiGAN+ again improved the result from 0.769 (SimOptiGAN) to 0.8. Both Mem2NucGAN-P and Mem2NucGAN-U led to scores of 0.061 and 0.155, respectively. Data generated with binary transformation achieved a DET score of 0.437.

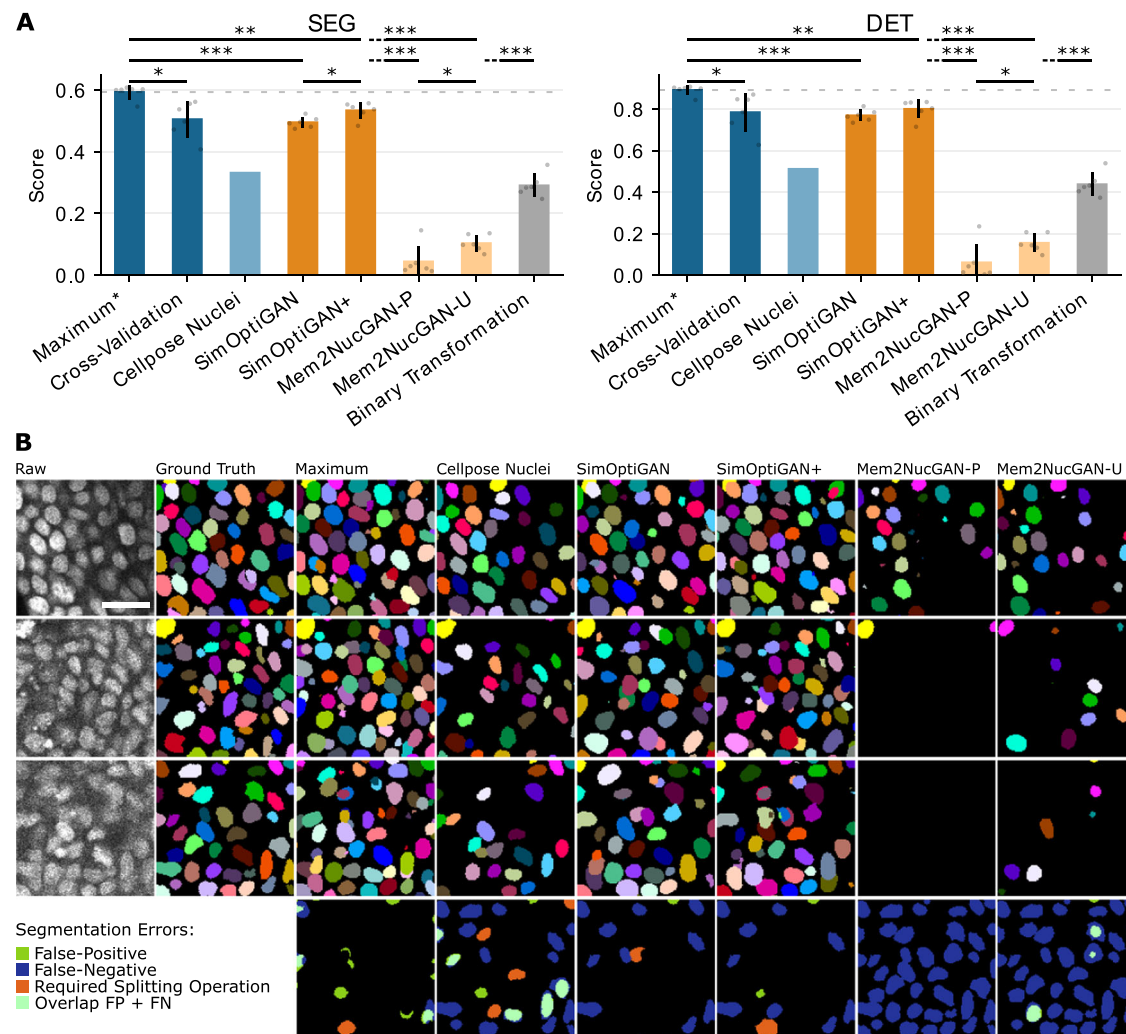
Figure 5B shows the segmentation results in comparison with the raw image and the ground truth obtained by manual annotation. A 2D slice is shown for each of the three ground truth patches. The patch locations were selected to feature three image regions: the upper, middle, and lower regions of the 3D image stack. As a result, the brightness and signal-to-noise ratio decrease over the three patches. At first glance, it can be seen that the models Cellpose Nuclei, Mem2NucGAN-P, and Mem2NucGAN-U had difficulties detecting nuclei in darker image regions. With the segmentation model Mem2NucGAN-P, only two nuclei were detected in the slice of the second patch and none in the slice of the third patch. This matches the results of DET measure. The model Maximum detected most of the nuclei, directly followed by the SimOptiGAN+ and SimOptiGAN model.

Further analysis was conducted regarding the segmentation scores of the Mem2NucGAN-U approach. To see whether a reduction in brightness in the training data can result in improved segmentation scores, a new dataset was created by applying the brightness reduction function of SimOptiGAN to the existing images generated with Mem2NucGAN-U. This resulted in a twofold improvement in the segmentation scores (SEG: 0.203 and DET: 0.332) in comparison to the default Mem2NucGAN-U.

### Using synthetic data drastically reduces manual effort for training data generation

By employing the presented framework for synthetic data generation, users can drastically reduce the manual effort required to generate a sufficient amount of training data for 3D segmentation models. While the time for manual annotation is reduced, computation time and GPU hardware are required. A comparison of the times needed for synthetic training data generation with the proposed methods is given in Table 1.

The biophysical simulation used to create realistic cell borders involved a one-time segmentation of membrane signals to automatically extract essential parameters. For this, the pretrained Cellpose Cyto2 model was optimized using the human-in-the-loop annotation method, requiring 4 h of annotation time. Subsequent parameter estimation for biophysical simulation needed 13 h of calculation time. Both the membrane segmentation and simulation parameter estimation were one-time processes. After these initial steps, the user can theoretically generate an infinite number of simulated cell borders. The computation for the biophysical simulation took 148 h to produce four 3D cell border images, which served as the basis for



**Fig. 5 | Comparison of segmentation results of nuclei segmentation models trained with different types of training data.** **A** SEG and DET segmentation scores of nuclei segmentation models trained with different types of training data. Scores can range from zero (worst possible) to one (best possible). Three manually corrected image patches from different image regions serve as test data. Maximum indicates a model trained on the test data and is considered an upper boundary. Based on this score, a dotted gray horizontal line is drawn to indicate this maximum. Dark blue color indicates training data generated by manual annotation. The nuclei model provided by Cellpose is depicted in light blue color. Dark and light orange colors indicate the pure use of synthetic training data generated with physical simulation-based and GAN-based approaches, respectively. Gray color indicates the use of synthetic data generated by the GAN-based transformation of synthetic nuclei

labels. The error bars represent the standard deviation across six training runs of the segmentation models. Since the Cellpose nuclei model is pretrained, no standard deviation is provided. **B** Qualitative comparison of nuclei segmentation results. Representative single optical sections of ground truth patches are shown for enhanced clarity. The first and second columns display the raw image signal and its corresponding ground truth, while subsequent columns show the segmentation masks obtained with the segmentation models. Additionally, the last row visualizes the DET-related errors of the third row, including false-negative, false-positive, and required splitting operations. A complete visualization of DET errors across all ground truth patches is given in the Supplementary material (Fig. A2). Scale bar: 25  $\mu$ m.

generating realistic nuclei and membrane signals. As SimOptiGAN does not utilize biophysical simulated cell borders, the associated annotation and calculation times were not applicable.

For generating nuclei signals with SimOptiGAN and SimOptiGAN+, manual annotation of a few nuclei was necessary to populate the nuclei prototype database. In this study, only 10 nuclei were annotated as prototypes, requiring 22.5 min for annotation. The most resources for SimOptiGAN and SimOptiGAN+ were needed for training the optimization network (99.5 h and 42 GB of GPU memory). For the other parts of the SimOptiGAN and SimOptiGAN+ pipeline (prototype generation and imaging simulation), only 0.5 GB of RAM and a computation time of less than 2 min were required. In contrast to previous experiments<sup>3</sup>, calculations of SimOptiGAN (and SimOptiGAN+) were directly performed at the desired output resolution, which drastically improved calculation time and memory consumption.

Mem2NucGAN-P and Mem2NucGAN-U did not necessitate additional manual annotation but also needed training of GAN-based models. Training of the transformation network Mem2NucGAN-P required 20.5 h and 16.5 GB GPU memory, while Mem2NucGAN-U consumed 59.5 h and 31 GB GPU memory.

Initial membrane segmentation and subsequent parameter estimation and biophysical simulation were performed on a workstation equipped with an AMD Ryzen 9 5950X CPU, 128 GB RAM, and an NVIDIA GeForce RTX 3060 Ti 8 GB GPU. All remaining calculations were performed on a server computer equipped with an AMD EPYC 7252 CPU, 64 GB RAM, and an NVIDIA A6000 48 GB GPU.

## Discussion

To the best of our knowledge, this is the first time that biophysical simulation has been incorporated into the process of generating synthetic

**Table 1 | Comparison of the required times to generate training data**

Method	Manual annotation	Computation preparation	Computation inference
Maximum	37.5 h	-	-
SimOptiGAN	0.4 h	99.5 h	<0.1 h
SimOptiGAN+	4.4 h	112.5 h	148 h
Mem2NucGAN-P	4 h	33.5 h	148 h
Mem2NucGAN-U	4 h	72.5 h	148 h

The Cellpose nuclei segmentation model is not included, as no training data is required for the usage of this model. Manual annotation defines the time required to manually generate or correct label masks. Preparation and inference time include the computation time required for the generation of synthetic images. While preparation includes one-time calculations such as parameter estimation for biophysical simulation and GAN-based model training, data generation comprises the remaining calculation time to generate four synthetic images.

membrane and nuclear signals. Both synthetic signals were based on the same simulated cell borders and therefore allowed the combination of both signals. This was utilized to train segmentation models with both membrane and nuclei signals as input, which, as preliminary results showed, improved segmentation performance for membrane signals.

We further introduced a GAN training scheme that enables the generator to not only transform image signals but, more importantly, also generate the corresponding labels. This allowed GANs to generate nuclei signals and their labels based on membrane signals, which are afterward available as training data for segmentation models. This approach could potentially be used in the inverse direction, i.e., for the generation of membrane signals and labels based on real nuclei signals or even for other image modalities like the transformation between nuclei and the proliferation marker Ki-67. This concept is also of potential interest in other application fields, such as virtual staining. In this field it enables the generation of synthetic marker signals, such as Ki-67, along with the corresponding segmentation masks.

The incorporation of biophysical modeling visually improved the arrangement of nuclei compared to a random placement approach. Visual inspection of the synthetic images generated by various methods (see Fig. 4) revealed notable differences in the brightness distribution. Specifically, images produced using the SimOptiGAN and SimOptiGAN+ showed a decreasing brightness deeper inside the spheroid, whereas those generated by the Mem2NucGAN methods maintained a nearly constant brightness throughout the image. One possible explanation for this discrepancy could stem from the utilization of different training datasets for the optimization model used in SimOptiGAN(+) and the transformation models used in Mem2NucGAN. The training data for the latter models consisted of real images exhibiting only a minimal reduction in brightness in deeper regions. Additionally, the absence of brightness information in simulated binary membranes, coupled with a patch-based transformation approach, rendered the Mem2NucGAN models mostly incapable of incorporating a decreasing brightness in deeper regions.

Such discrepancies undoubtedly affected the segmentation quality of nuclei in darker image regions. As observed in the visual inspection of the segmentation results (see Fig. 5), both Mem2NucGAN models encountered challenges in detecting nuclei in darker image regions. Both of the SimOptiGAN methods performed far better under such conditions, indicating that the brightness reduction improved the segmentation performance. The KID measure appears unaffected by differences in image brightness, which may account for some of the discrepancy compared to the segmentation scores. One reason for the brightness invariance might be the underlying InceptionV3 Net, which was trained on a diverse classification dataset where brightness information is irrelevant for the class outcome.

Another factor contributing to the lower segmentation scores of the Mem2NucGAN models could be attributed to the label generation process. As the binary labels were only an additional output of the generator models, a strict alignment between image and label data may not be given for all

instances. Although the post-processing step, used to derive instance labels, visually improved the quality of the label data, it may have introduced additional discrepancies between labels and generated image data. The fact that, despite the low segmentation score of the Mem2NucGAN-U method, its KID is the lowest overall may indicate potential label inconsistency. Since KID does not account for label consistency, a model might achieve favorable KID results while producing poor segmentation scores if the generated labels are inconsistent with the corresponding synthetic images.

The additional membrane segmentation step required for Mem2NucGAN-P might have affected the task of transformation learning due to errors in the membrane segmentation. We also tested additional approaches that did not require segmented membrane signals for training the generator and instead directly used the raw membrane signals (see Supplementary material Fig. A1, *Real Membrane* → *Nuclei*). However, these involved pre-processing of the simulated binary membrane to match real membrane signals. We tested several pre-processing variants, including the imaging simulation and a GAN transformation model to generate realistic membrane signals, however, no nuclei with realistic textures could be obtained.

In the additional experiment, in which a separate brightness reduction was applied to the resulting images of Mem2NucGAN-U, the segmentation scores improved by a factor of two. That indicates that, indeed, the missing brightness reduction in the training data has a significant influence on segmentation performance. Nevertheless, SimOptiGAN(+) continues to exhibit superior segmentation performance. In the visualized overlay of generated nuclei images and their corresponding labels (provided in the Supplementary material), it appears that the label consistency of both Mem2NucGAN methods does not match that of SimOptiGAN.

The comparison of the hybrid approach of SimOptiGAN, e.g., combining imaging simulation and GAN-based optimization, with a direct transformation of label data (binary transformation) demonstrated that the hybrid approach of SimOptiGAN led to significantly higher segmentation scores than the direct transformation of label data. The approach of binary transformation can also be considered an alternative to Mem2NucGAN-U, where the transformation task is simplified by predefining the shapes and positions of nuclei. However, the fundamental concept of the proposed Mem2NucGAN approaches, namely that a GAN model can generate realistic nuclei morphologies based on the provided membrane signals, is lost.

The segmentation scores showed that synthetic training data can not only outperform the pretrained Cellpose model but also training data generated by manual annotation. It can be argued that by increasing the data size of the manual training data, the results of the Cross-Validation model would be improved. However, this would also strongly increase the amount of time required for annotation. Ideally, smaller patch sizes are preferred to label a greater number of patches, thereby enhancing heterogeneity. However, the requirement for a minimal patch size in most segmentation models restricts the number of patches that can be labeled within a limited time-frame. In the case of rare objects, a reduced number of patches can lead to an underrepresentation or even a complete absence of such objects in the training dataset. Synthetic training data, on the other hand, can be generated in large quantities with minimal user interaction and thus with minimal time investment. SimOptiGAN and SimOptiGAN+ additionally allow for an overrepresentation of rare objects and thus improve segmentation performance for these objects. Both SimOptiGAN and SimOptiGAN+ also require some annotation for the extraction of nuclei prototypes. However, the time investment here is rather small. For the generated synthetic data, only 20 nuclei were labeled instead of the 1001 nuclei present in the three ground truth patches.

The improvements achieved through biophysical modeling of cell arrangements come with certain limitations. Currently, some degree of manual annotation is still required for parameter estimation and to define the initial configuration for the biophysical simulation. Nevertheless, the amount of manual labor required is significantly reduced compared to the extensive effort needed for the manual annotation of image patches. Another potential limitation lies in the diversity of nuclei within the images



generated by SimOptiGAN(+). As these methods rely on a database of nuclei prototypes, the diversity is inherently limited by the available prototypes. To mitigate this, augmentation techniques are employed, and users are advised to select a diverse set of nuclei for prototype extraction to enhance the variability in the generated data.

No clear statement can be given as to why the incorporated biophysics improved the results of the segmentation. Apart from the placement process of the nuclei, the pipelines were identical, including the used parameters. For example, the same nuclei prototypes, imaging-simulation parameters and even the same optimization model were used. Reasonable explanations for the improvement may be due to a more fitting nuclei size or a more realistic orientation and arrangement of nuclei. Both could potentially lead to a better transformation during the optimizing step.

We would like to emphasize that the proposed methods are not limited to the 3D spheroid dataset presented in our study. They are, in fact, transferable to other visually distinct datasets. The methodology has the potential to be extended to other complex structures, such as organoids, which are becoming increasingly important in biomedical research for applications such as disease modeling and drug discovery. However, adapting the simulation to account for additional biophysical effects may be necessary to effectively model these more intricate structures.

It is important to note that the performance of the model may vary depending on the degree of difference in cell morphology, orientation, and density between the datasets. The larger the difference between these factors, the more likely it is that the segmentation performance will decrease. If the differences are substantial, we recommend generating new training data specific to the new dataset to ensure optimal results. For SimOptiGAN, the annotation of new nuclei prototypes can be skipped, if the new data shows similar nuclear morphology and texture. In such cases, a reparameterization of the simulation pipeline and retraining of the optimization network can be sufficient. As SimOptiGAN+ is based on the results of biophysical simulation, adapting cell arrangements requires the generation of new cell borders. For the biophysical simulation, it might be sufficient to adapt the initial starting configuration and parameters like expected cell volume. However, if differences are too large, one has to consider extracting parameters from the new data.

## Conclusion

In contrast to existing approaches, our three approaches incorporated biophysical simulation to improve the fuzzy similarity between real-world and synthetic data, thus also enhancing the training quality of segmentation algorithms. Furthermore, the segmentation model trained with our synthetic data outperformed both the pretrained Cellpose Nuclei model and a model trained with manually annotated data. This use of synthetic training data enabled precise single-cell level analysis of 3D cell cultures. Further experiments examining the influence of parameters such as cell size and density on segmentation performance could contribute to further improvements in the generation of synthetic data. Looking ahead, we plan to test the proposed method on other cell models, like organoids and validate our GAN-based methods, Mem2NucGAN-P and Mem2NucGAN-U, on different image modalities such as nuclei and Ki-67. In future work, we will also assess the segmentation performance of models trained with combined synthetic nuclei and membrane signals. Furthermore, emerging foundation models, such as those described in ref. 46, could potentially eliminate the necessity for manual parameter estimation in biophysical simulations. Instead, these models could be leveraged to automatically derive parameters for novel cell types.

## Methods

### Biophysically motivated simulation of cell boundaries

Cell boundaries were generated in a biophysically motivated simulation. To this end, a 3D Cellular Potts Model (CPM) was set up using the *CompuCell3D* implementation<sup>18</sup> (Version 4.3.2, Revision 0). The simulation approach outlined below is described in more detail in ref. 47. As the initial configuration of the simulation, we started from a real-world confocal image

stack of a mono-culture spheroid with fluorescence-labeled cell membranes, as described below in section “Methods—Dataset”.

Cell boundaries in this image stack were segmented with *Cellpose 2.0*<sup>45</sup>, using a *cyto2* model<sup>48</sup> retrained with a human-in-the-loop approach. For applying the model to 3D images, the integrated stitching method of Cellpose was utilized. This approach involves performing segmentation on individual slices, followed by IoU-based stitching to generate a complete 3D segmentation. A detailed description of the procedure is given in ref. 47. The resulting 3D label image was then further processed: Label masks with fewer than 5 voxels were removed, as they result almost certainly from segmentation errors. Labels spanning less than four image planes in the z-direction were removed for the same reason. A closing and a dilation operation were applied to fill the resulting holes in the image. Image up-sampling through nearest neighbor interpolation was applied in the z-direction to ensure an isotropic voxel size of 0.5682  $\mu\text{m}^3$ . A list of morphological features  $F_k$  was extracted for each cell, where the index  $k$  runs over 7 distinct features: cell volume, surface area, volume-to-surface (V/A) ratio, minor and major axis lengths, sphericity and eccentricity.

The label image was then converted into a file format that specifies the initial configuration of a CPM simulation in *CompuCell3D*<sup>47</sup>. The system Hamiltonian  $H$ , which models the total free energy of the cell conglomerate, includes terms for the adhesion energy between neighboring cells, surface free energy contributions of cells in contact with the medium, and penalty terms for the deviation of volume  $V_i$  and surface area  $A_i$  of cell  $i$  from their target values:

$$H = \sum_i \lambda_V \left( V_i - V_i^{(\text{target})} \right)^2 + \sum_i \lambda_A \left( A_i - A_i^{(\text{target})} \right)^2 \quad (1)$$

$$+ \sum_{i,j} J^{(c-c)} A_{i,j}^{(c-c)} + \sum_i J^{(c-m)} A_i^{(c-m)} \quad (2)$$

Here,  $\lambda_V$  and  $\lambda_A$  are penalty parameters that determine the strength of the volume and area constraints.  $A_{i,j}^{(c-c)}$  is the contact area between cell  $i$  and  $j$ ,  $A_i^{(c-m)}$  the contact area of cell  $i$  with the surrounding medium, and  $J^{(c-c)}$  and  $J^{(c-m)}$  are the respective surface tension parameters. The target values  $V_i^{(\text{target})}$  and  $A_i^{(\text{target})}$  for volume and area were assigned individually to each cell. Here, the target values of cell  $i$  were taken from the *actual* values of cell  $j$  in the initial configuration, using a random permutation  $P_{ij}$  of all cells. This way, each cell has target values *different* from its initial values, ensuring some dynamics in the simulation; on the other hand, the distribution of target values and original values over all cells is identical, ensuring a realistic volume/area distribution of the ensemble throughout the simulation.

Moreover, a temperature parameter  $T$  needs to be set, which determines the strength of the fluctuations around the energy minimum in the equilibrium state.

In sum, the simulation model contained four parameters:  $\lambda_V$ ,  $\lambda_A$ ,  $J^{(c-c)}$ ,  $J^{(c-m)}$ . Their values were fixed by a parameter scan on a hand-segmented subset of a spheroid of  $101 \times 250 \times 250 \text{ px}^3$  ( $z, y, x$ ), with the objective to minimize the following metric  $m$ <sup>47</sup>:

$$m = \left( \frac{1}{N_{\text{Features}}} \sum_k W_k \right) \cdot \left( \frac{1}{N_{\text{Cells}}} \sum_i \text{IoU}_i \right) \quad (3)$$

Here,  $W_k$  is the Wasserstein distance of the empirical distribution of the morphological feature  $F_k$  at the beginning of the simulation versus its end. Small values of  $W_k$  thus ensure that throughout the simulation, the cell ensemble retains its morphological characteristics.  $\text{IoU}_i$ , on the other hand, is the intersection over union of cell  $i$  in the beginning vs. the end of the simulation; i.e., it measures to what extent a cell moves in the simulation. Small values of  $\text{IoU}_i$  hence ensure that individual cells change their position and/or shape in the simulation, to avoid simulation results with little movement or even a complete “freeze.” The resulting parameter values are listed in Table 2.

**Table 2 | Simulation parameters for CompuCell3D**

Parameter	Parameter scan values	Best
$\lambda_V$	0.001, 2.0, 4.0, 6.0, 8.0, 10.0	10.0
$\lambda_A$	0.001, 2.0, 4.0, 6.0, 8.0, 10.0	0.001
$\int_{(c-c)}$	0.0001, 2.0, 4.0, 6.0, 8.0, 10.0	2.0
$\int_{(c-m)}$	10.0, 55.0, 100.0	55.0
$T$	–	30.0
Potts Neighbor Order	–	3
Contact Neighbor Order	–	4

With those parameters, the system was simulated over 1000 MCS with different random seeds in CompuCell3D. The number of MCS was selected because, beyond this point, the system approaches thermodynamic equilibrium, resulting in only minimal changes in cellular morphology. The resulting label masks were converted into cell border images, serving as a template for the actual generation of synthetic images, as described below.

### Membrane synthesis

To produce authentic membrane signals derived from the simulated binary membranes, we employed the CycleGAN architecture<sup>49</sup>. Apart from the 3D adaptation of the generator and discriminator models, the CycleGAN architecture is similar to the original one presented in ref. 49. As training data, unpaired image data consisting of four images showing real membrane signals and four images showing simulated binary membranes were utilized.

### Nuclei synthesis—from membrane to nuclei

The main objective was the generation of synthetic nuclei data for the training of segmentation models. We analyzed several ways of creating nuclei images based on the simulated cell borders (see Fig. 6; a comprehensive overview is given in the Supplementary material Fig. A1). One approach is based on placing nuclei into the simulated cell borders, while the other two approaches utilize a GAN-based transformation between membrane-like signals and nuclei signals. These two approaches are subdivided based on the model that enables the transformation.

**Simulation-based approach with random nuclei placement: SimOptiGAN.** Our pipeline, first introduced in ref. 3, enables the generation of realistic-looking 3D nuclei images. It relies on a database containing nuclei prototypes, which are extracted by manual annotation of a few nuclei in ideally high-resolution images. The prototypes represent cutout images of nuclei, where only the annotated region is visible, as already described in ref. 3. During the *Prototype Generation*, a 3D image of a cell culture is generated by placing nuclei prototypes into the image.

Afterward, an *Imaging Simulation* is performed to simulate the effects of the recording process with a microscope. Lastly, during the *Optimization*, a CycleGAN is used to post-process the generated image.

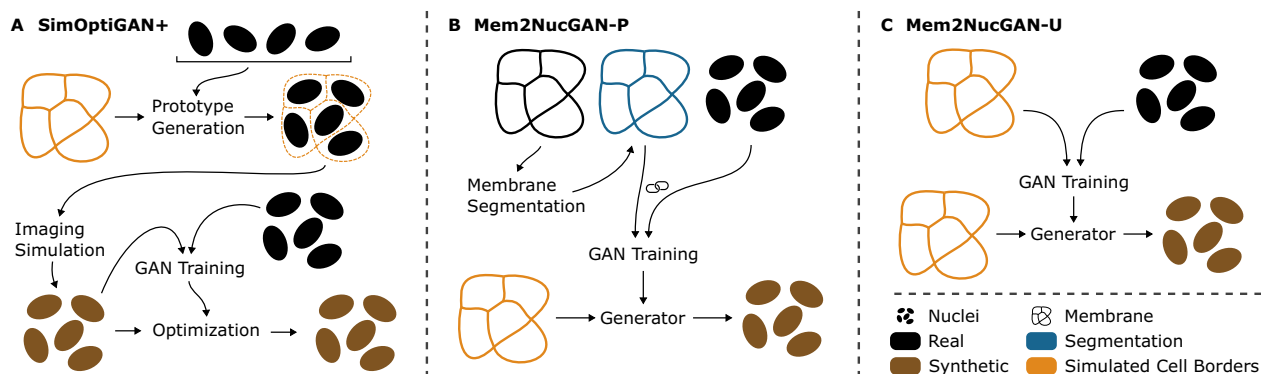
**Enhanced nuclei placement for simulation-based approach: SimOptiGAN+.** With SimOptiGAN, the arrangement of nuclei is not physically motivated but rather follows a random pattern. For SimOptiGAN+, the pipeline is adapted to utilize the cell borders generated by biophysical simulation to improve the placement process of the nuclei prototypes during the *Prototype Generation* step (see Fig. 6A). The procedure is described in the following.

For each cell of the simulated cell structure, a random nucleus prototype is chosen from the database. This prototype is then rotated to align with the cell's orientation and scaled to match a predefined volume relative to the cell's volume. As nuclei in the real world are not strictly located in the center of a cell, the placement position is selected based on a random distribution with its highest likelihood in the cell's center. For a selected position, the overlap between the nucleus prototype and the cell is assessed. If it is larger than a specified percentage threshold of the nucleus prototype's volume, the nucleus is placed. If the overlap is smaller than the threshold, a new position is selected and reassessed. This process continues until either a suitable position is found or the maximum number of attempts is reached. In the latter case, a new nucleus prototype is chosen from the database, and the positioning procedure is repeated. A maximum number of nuclei prototypes are tested before the cell is skipped and no nucleus is placed in this cell.

Once all cells are processed, the *Prototype Generation* is completed. Afterward, the *Imaging Simulation* is used to simulate the optical and sensor effects of a microscope. This includes a brightness reduction in deeper regions, the convolution with a point-spread function, a downsampling step, and the calculation and application of noise. Subsequently, a deep-learning-based post-processing is performed during the *Optimization* step to further increase the realism of the image. For this step, the same network can be used as for the regular pipeline. Finally, a synthetic nuclei image is obtained with the corresponding labels.

**Binary membrane to nuclei with paired data: Mem2NucGAN-P.** This approach utilizes a GAN-based transformation for the generation of realistic nuclei signals based on the biophysically simulated cell boundary images (see Fig. 6B). Our previous experiments indicated that GANs can learn the generation of nuclei signals with correct orientation and morphology by only utilizing membrane signals. Using such a transformation model, which is trained on real data, would thus allow the generation of nuclei signals with realistic properties corresponding to the membrane signal.

However, as the biophysically simulated cell borders are binary signals, the transformation will fail due to the domain gap with the initial training data (real membrane signals). To overcome this, we train the transformation



**Fig. 6 | Overview of presented methods for the generation of nuclei images based on simulated cell labels.** SimOptiGAN+ (A) uses the simulated cell borders to improve the nuclei placement of a previously presented simulation-based pipeline,

while Mem2NucGAN-P (B) and Mem2NucGAN-U (C) rely on a GAN-based transformation of simulated cell borders.



model with segmented instead of raw membrane signals. This requires the training of a *Membrane Segmentation* model. As manual annotation of membrane images is not feasible, synthetic training data is generated by an additional transformation model, *Synthetic Cell Borders*  $\rightarrow$  *Membrane* (see Supplementary material Fig. A1). This CycleGAN-based transformation model allows the generation of realistic membrane signals based on the biophysical simulated cell borders by utilizing unpaired images as training data. Data pairs of biophysical simulated cell borders and synthetic membrane signals generated with *Synthetic Cell Borders*  $\rightarrow$  *Membrane* are then used to train the *Membrane Segmentation* model. In this study, we relied on the pretrained cyto2 model of Cellpose for the segmentation of membrane signals.

By utilizing the *Membrane Segmentation* model, real-image pairs of segmented membrane and corresponding raw nuclei signals can be generated (see Fig. 6B). This data is then used to train a conditional GAN (cGAN) representing the main transformation model *Binary Membrane*  $\rightarrow$  *Nuclei*. We utilized the pix2pix training approach and discriminator models, but with ResNet generator models, as used by CycleGAN. Finally, this main transformation model is used to generate synthetic nuclei signals based on biophysically simulated cell borders.

#### Binary membrane to nuclei with unpaired data: Mem2NucGAN-U.

While the transformation model of the previous approach is trained on real data and therefore can learn morphological correlations between membrane and nuclei, it is a complex procedure. The required segmentation model and its training introduce additional error sources. Furthermore, differences between segmented membranes used during training and simulated membrane labels used during the inference could decrease the quality of the generated nuclei signals. We therefore tested an additional method that is based on a CycleGAN, which does not require paired data for training (see Fig. 6C). The additional segmentation of membrane signals is skipped, and simulated membrane labels in combination with real nuclei signals are directly used for training. During inference, the model transforms the same input data as used during training, i.e., simulated membrane labels, into synthetic nuclei images.

**Adapted GAN training for the additional generation of matching labels.** Mem2NucGAN-P and Mem2NucGAN-U utilize GANs to obtain nuclei images based on membrane-like signals. For the training of segmentation models, corresponding label data for the generated image data is required. However, conventional GAN training methods for image generation lack the ability to output corresponding labels alongside the generated images. To circumvent this constraint, existing approaches resort to generating synthetic labels separately, such as through random ellipse placement, before using GANs to transform

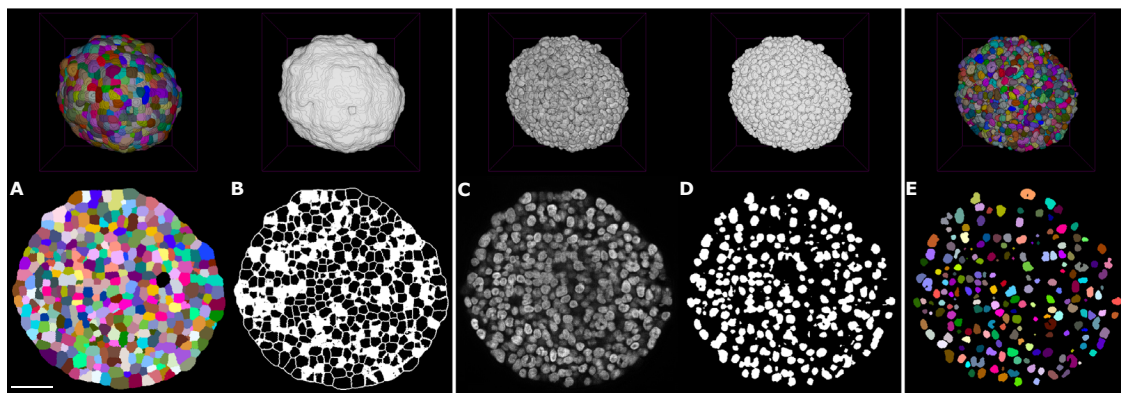
them into image data. However, this approach is not feasible for our task, as we aim to transform between two distinct image modalities—membrane to nuclei (Fig. 7B, C)—while only having access to membrane labels via biophysically simulated cell borders (Fig. 7A). Hence, we require the GAN in both Mem2NucGAN methods to produce both synthetic nuclei signals and their corresponding labels (Fig. 7C, D). To our knowledge, no existing training scheme facilitates the additional generation of corresponding labels. Therefore, we introduce a GAN structure to generate nuclei images along with matching label masks. The new structure can be applied to both cGAN and CycleGAN (see Fig. 8). First, the adaptation of the cGAN is described, while later the methodology is transferred to the CycleGAN.

The last layer of the generator  $G$  is altered to generate a two-channel image. The first channel contains the nuclei signal prediction, while the second channel contains a binary label mask corresponding to the nuclei signal in the first channel (see Fig. 7C, D). The generator  $G$  thus maps the image condition  $x$  and a noise vector  $z$  to a synthetic image  $\hat{y}$  and its label  $\hat{v}$ :  $G: \{x, z\} \rightarrow \{\hat{y}, \hat{v}\}$ . As the two output channels are based on the same decoder path, i.e., sharing the same feature maps, the alignment between the image and label data is promoted. During training, the discriminator  $D$  receives fake samples consisting of the condition  $x$  (i.e., the membrane signals) and the first channel of the generator output  $\hat{y}$  (i.e., nuclei signals) and real samples consisting of the condition  $x$  (i.e., the membrane signals) and real nuclei signals  $y$ .

In addition to the traditional training scheme, an extra discriminator  $D_{\text{Seg}}$  is introduced. While the default discriminator only penalizes the quality of the nuclei channel, the purpose of the second discriminator is to penalize the generated binary label mask. The discriminator  $D_{\text{Seg}}$  itself is trained to distinguish between real ( $v$ ) and fake ( $\hat{v}$ ) binary labels. We want to emphasize that, in contrast to the regular discriminator  $D$ ,  $D_{\text{Seg}}$  receives no additional condition. This enables the use of arbitrary, non-paired binary label masks for the training of  $D_{\text{Seg}}$ , diminishing the demand for high-quality label masks in two key aspects. First, with the absence of a corresponding nuclei signal, any overlooked nuclei within the label masks become less conspicuous, thereby minimizing confusion during model training. Second, the use of a binary label mask eliminates issues regarding under- and over-segmentation errors.

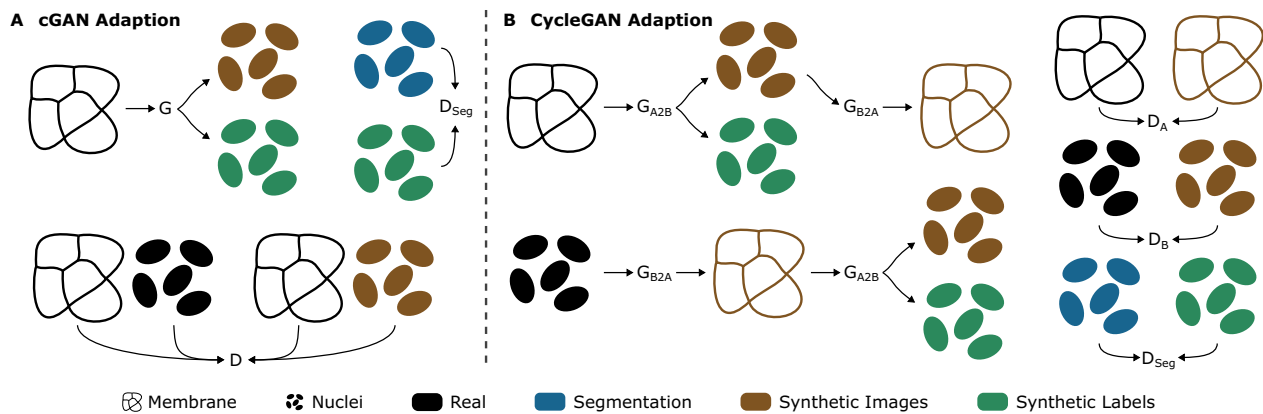
The original optimization function of the cGAN, consisting of an adversarial GAN loss  $\mathcal{L}_{\text{cGAN}}(G, D)$  and a reconstruction loss  $\mathcal{L}_{L1}(G)$  (weighted by a parameter  $\lambda_{L1}$ ) is appended with an additional GAN loss  $\mathcal{L}_{\text{sGAN}}(G, D_{\text{Seg}})$  resulting from the discriminator  $D_{\text{Seg}}$ :

$$\min_G \max_{D, D_{\text{Seg}}} \mathcal{L}_{\text{cGAN}}(G, D) + \mathcal{L}_{\text{sGAN}}(G, D_{\text{Seg}}) + \lambda_{L1} * \mathcal{L}_{L1}(G), \text{ with} \quad (4)$$



**Fig. 7 | Generation of synthetic images and labels using GAN-based methods.** The simulated cell borders (A) are converted into binary membrane signals (B), which are then utilized by the generator to produce synthetic nuclei signals (C) along with corresponding binary labels (D). Subsequently, a post-processing step is employed

to extract instance labels (E) from the generated binary labels. The binary membrane signals (B) partially show large white regions. These are a result of the lower resolution in the z-direction, leading to cell borders/membranes that completely lie in one z-plane. Scale bar: 50  $\mu\text{m}$ .



**Fig. 8 | Overview of two proposed GAN training schemes for the generation of not only synthetic images but also matching labels.** Both methods utilize an additional discriminator  $D_{Seg}$  to generate additional labels. **A** and **B** show the adapted training schemes for a cGAN and CycleGAN, respectively.

$$\mathcal{L}_{cGAN}(G, D) = \mathbb{E}_{x,y} [\log D(y, x)] + \mathbb{E}_{x,z} [\log (1 - D(\hat{y}, x))], \quad (5)$$

$$\mathcal{L}_{L1}(G) = \mathbb{E}_{x,y,z} [\|y - \hat{y}\|_1] \text{ and} \quad (6)$$

$$\mathcal{L}_{sGAN}(G, D_{Seg}) = \mathbb{E}_v [\log D_{Seg}(v)] + \mathbb{E}_{x,z} [\log (1 - D_{Seg}(\hat{v}))]. \quad (7)$$

$$\mathcal{L}_{GAN_B}(G_{A2B}, D_B) = \mathbb{E}_y [\log D_B(y)] + \mathbb{E}_x [\log (1 - D_B(\hat{y}))], \quad (10)$$

$$\mathcal{L}_{cyc}(G_{A2B}, G_{B2A}) = \mathbb{E}_x [\|x - \tilde{x}\|_1] + \mathbb{E}_y [\|y - \tilde{y}\|_1] \text{ and} \quad (11)$$

$$\mathcal{L}_{GAN_S}(G_{A2B}, D_{Seg}) = \mathbb{E}_v [\log D_{Seg}(v)] + \mathbb{E}_x [\log (1 - D_{Seg}(\hat{v}))]. \quad (12)$$

As the generator  $G$  produces binary labels masks (Fig. 7D), a post-processing step is required to derive instance labels (Fig. 7E) required for segmentation training. In the first step, the predicted label mask undergoes a thresholding procedure to obtain a true binary label mask. Afterward, the binary membrane signal which is used as input for the generator is subtracted from the binary label mask to split the labels. Then, a binary opening operation is performed to remove potential artifacts. Utilizing the simulated cell label mask, each foreground voxel in the binary label mask is assigned the corresponding cell label. Thus, foreground voxels in the binary label mask that are zero in the cell label mask, i.e., artifacts, are automatically removed.

The CycleGAN structure is altered similarly to the cGAN. The generator  $G_{A2B}$  is tasked with the generation of a two-channel output based on a single-channel input featuring a binary membrane:  $G_{A2B} : x \rightarrow \{\hat{y}, \hat{v}\}$ . The first channel  $\hat{y}$  represents the nuclei signal, while the second channel  $\hat{v}$  represents the binary labels. Since there are no segmentations available for real nuclei images, the generator  $G_{B2A}$  only receives a nuclei image  $\hat{y}$  and no label mask  $G_{B2A} : y \rightarrow \hat{x}$ . Similarly, the discriminator  $D_B$ , only assesses nuclei images to distinguish between real and fake samples. Discriminator  $D_A$ , distinguishes between real and synthetic membrane signals. Image consistency between domains is achieved by encouraging cycle consistency, i.e., by penalizing differences between  $x$  and  $\tilde{x} = G_{B2A}(\hat{y})$ , and  $y$  and  $\tilde{y} = G_{A2B}(\hat{x})$ . To penalize the generated label masks, an additional discriminator  $D_{Seg}$  is introduced, evaluating real and fake binary label masks. As only unpaired binary label masks are required, the same benefits as for the adapted cGAN apply here. The original optimization function of the CycleGAN, consisting of the GAN loss  $\mathcal{L}_{GAN_A}$ ,  $\mathcal{L}_{GAN_B}$  and the cyclic loss  $\mathcal{L}_{cyc}$ , is appended with the GAN segmentation loss  $\mathcal{L}_{GAN_S}$  resulting from the discriminator  $D_{Seg}$ :

$$\min_{G_{A2B}, G_{B2A}} \max_{D_A, D_B, D_{Seg}} \mathcal{L}_{GAN_A}(G_{B2A}, D_A) + \mathcal{L}_{GAN_B}(G_{A2B}, D_B) + \mathcal{L}_{GAN_S}(G_{A2B}, D_{Seg}) + \lambda_{cyc} \mathcal{L}_{cyc}(G_{A2B}, G_{B2A}), \text{ with} \quad (8)$$

$$\mathcal{L}_{GAN_A}(G_{B2A}, D_A) = \mathbb{E}_x [\log D_A(x)] + \mathbb{E}_y [\log (1 - D_A(\hat{x}))], \quad (9)$$

## Quality evaluation measures

**Kernel inception distance (KID).** To assess the similarity of 3D microscopic images generated by the presented methods to their real counterparts, we employ the Kernel Inception Distance (KID)<sup>42</sup>. As the KID is based on a pretrained model, designed to process 2D images, the measure cannot be directly applied to 3D images. We therefore calculate the KID measure slice by slice. The mean results for all three major plane directions ( $xy$ ,  $xz$ , and  $yz$ ) are used to calculate a final mean value of the KID measure.

Given the unlimited range of the KID measure, interpreting the results can be challenging. To provide context, we include the KID value between real and naively generated reference images. These reference images are interim outputs from SimOptiGAN, produced by the simulation pipeline prior to deep learning optimization.

**Segmentation.** Raw confocal datasets were converted to multichannel TIFF files using Fiji<sup>26</sup>. 3D segmentation of nuclei and membrane staining (SiR-actin) was performed using Cellpose (V2.2), a deep learning-based instance segmentation tool<sup>39</sup>. To prepare the hand-annotated training data, spheroids of each cell type were first pre-segmented by the pre-trained *nuclei* and *cyto2* models in Cellpose, using the two fluorescent markers for DAPI and SiR-actin, as an initial step. From these, three patches for each marker with a size of  $32 \times 128 \times 128 \text{ px}^3$  ( $z$ ,  $y$ ,  $x$ ) were extracted and manually corrected using the Segmentor software<sup>20</sup>. The locations of the patches were selected to feature different image qualities. The patches represent the top, middle, and bottom regions of the spheroid, where each patch shows a reduced signal-to-noise ratio. Supervised training from scratch was performed as described in ref. 38 using the command line interface for Cellpose. Please note that the sample of the spheroid used to generate the ground truth was from a previous experiment. Thus, clearing effectiveness slightly differs between the image samples used to generate synthetic data and the image sample from which the ground truth was generated.

The suitability of different types of synthetic data for use as training data for segmentation models is tested by training multiple segmentation models. For each presented approach of generating synthetic data, a 3D StarDist<sup>44</sup> segmentation model is trained with four synthetic images. To

assess the segmentation performance of the models, we utilize two metrics: SEG (Segmentation) and DET (Detection), as used in the Cell Tracking Challenge (CTC)<sup>43</sup>. The SEG measure evaluates the accuracy of segmentation by quantifying the agreement between the predicted and ground truth segmentation masks. In contrast, the DET metric focuses on the accuracy of detection, measuring how accurately the model identifies objects of interest within the image while considering factors like false-positives, false-negatives, and under-segmented objects. The implementation provided by the CTC allows the direct calculation of the segmentation performance in 3D. Both measures can range from zero to one, representing worst to best result. The performance was evaluated by segmenting the whole image and subsequently cropping the segmented image to match the ground truth patches. A direct segmentation of image crops would lead to many incomplete nuclei, potentially decreasing the segmentation quality. By segmenting the entire image, all nuclei remain fully visible, ensuring a more comprehensive assessment of segmentation performance.

## Statistics and reproducibility

Biophysical parameter estimation and the initialization of the starting configuration for the CPM were based on a single recording of a spheroid cell culture. Four biophysical simulations were subsequently performed, all utilizing the same parameters but differing in their random seed values to introduce variability. The resulting cell border images from these simulations served as the foundation for performing the biophysically motivated synthesis with SimOptiGAN+, Mem2NucGAN-P, and Mem2NucGAN-U. Each corresponding deep-learning model was trained only once due to the large number of models and the extensive training time required.

Statistical analysis was conducted to compare differences in mean performance among the tested segmentation models. Each StarDist segmentation model was trained six times ( $n = 6$ ), and performance was evaluated based on the mean results from three ground truth patches extracted from a single spheroid recording. A one-sided Welch's  $t$ -test, which does not assume equal variances or a normal distribution, was employed due to the small sample size ( $n = 6$ ), making the test more conservative. An increase in sample size was not feasible due to the extensive training time required for each model. Predefined significance levels were set as follows:  $*p < 0.05$ ,  $**p < 0.01$ , and  $***p < 0.001$ .

## Dataset

**Sample preparation.** HT-29 colon cancer cells (ATCC) were cultured in McCoy's 5A medium (Capricorn) supplemented with 10 % FBS and 1 % Pen/Strep and maintained in a humidified incubator at 37 °C with 5 % CO<sub>2</sub> fumigation. For spheroid generation, cells were detached using Trypsin/EDTA and seeded onto the ULA plates at a concentration of  $5 \times 10^2$  cells per well.

Spheroids ( $n = 10$ ) were transferred after 4 days of culturing to Eppendorf tubes, washed once with phosphate-buffered saline (PBS, Sigma Aldrich), and fixed with 4% paraformaldehyde (PFA, Carl Roth) for 1 h at 37 °C, followed by two washes with PBS containing 1% FBS for 5 min each. To remove traces of fixative, spheroids were quenched with 0.5 M glycine (Carl Roth) in PBS for 1 h at 37 °C with gentle shaking. Spheroids were then incubated for 30 min in a penetration buffer containing 0.2% Triton X-100, 0.3 M glycine, and 20% DMSO (all Carl Roth) in PBS to enhance the penetration of antibodies and nuclear stains. Spheroids were then incubated in a blocking buffer (0.2% Triton X-100, 1% BSA, 10% DMSO in PBS) for 2 h at 37 °C with gentle shaking. Samples were then stained with SiR-actin (Spirochrome/Tebubio SC001, 1:1000) and DAPI (Sigma Aldrich D9542-1MG, 1:1000) ON at 37 °C in antibody buffer (0.2% Tween 20, 10 µg/mL heparin (both Sigma-Aldrich), 1% BSA, 5% DMSO in PBS) with gentle shaking. Samples were then washed 5 times for 10 min each in wash buffer (0.2% Tween-20, 10 µg/mL heparin, 1% BSA) with gentle shaking and cleared with FUNGI clearing solution (50% glycerol (vol/vol), 2.5 M fructose, 2.5 M urea, 10.6 mM Tris Base, 1 mM EDTA) ON as previously described<sup>8</sup>. Cleared samples were transferred to 18-well µ-slides (Ibidi) in

the same solution and kept in the microscope room for several hours to allow for temperature adjustment.

The HT-29 colon cancer cell line used in our study is not listed in the ICLAC database of commonly misidentified cell lines. The cells were obtained from the American Type Culture Collection (ATCC), a reputable and widely recognized source for authenticated cell lines. In our laboratory, we routinely test all cell lines, including HT-29, for mycoplasma contamination to ensure the integrity of our experimental results. Additionally, the HT-29 cell line has been authenticated using Short Tandem Repeat (STR) profiling, following the global standard ANSI/ATCC ASN-0002.1-2021, to confirm its identity and eliminate concerns about misidentification.

**Imaging and data description.** Spheroids were imaged using an inverted Leica TCS SP8 confocal microscope (Leica Microsystems CMS, Mannheim, Germany) equipped with an HC PL APO 20× /0.75 IMM CORR objective, 488 nm, 561 nm, and 633 nm lasers, and Leica Application Suite X software. All image stacks were acquired with comparable settings, using Immersion Type F (Leica Microsystems, RI 1.52) as immersion fluid, with a resolution of  $1024 \times 1024$  px<sup>2</sup>, a z-step size of 1 µm, a laser intensity of  $1 \rightarrow 1.5\%$  and a gain setting of 600 to avoid overexposure of pixels. All image stacks were acquired with z-compensation to compensate for depth-dependent signal loss.

## Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

## Data availability

Real and resulting synthetic image data, trained models, segmentation results, along with the ground truth data employed in this paper, are available at Zenodo: <https://doi.org/10.5281/zenodo.11240362><sup>50</sup>. Data not included in this repository are available from the corresponding author upon reasonable request. The source data behind the graphs in the paper can be found in Supplementary Data 1.

## Code availability

The code for the presented methods, as well as for the evaluation metrics, can be found on GitHub: biophysical simulation: <https://github.com/s-sauer/tissue-simulation-pipeline>, cell data synthesis and evaluation: [https://github.com/bruchr/cell\\_synthesis](https://github.com/bruchr/cell_synthesis)<sup>51</sup>.

Received: 30 April 2024; Accepted: 6 January 2025;

Published online: 11 January 2025

## References

1. Fu, C. et al. Three dimensional fluorescence microscopy image synthesis and segmentation. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2221–2229 (2018).
2. Dunn, K. W. et al. DeepSynth: three-dimensional nuclear segmentation of biological images using neural networks trained with synthetic data. *Sci. Rep.* **9**, 18295 (2019).
3. Bruch, R. et al. Synthesis of large scale 3D microscopic images of 3D cell cultures for training and benchmarking. *PLoS ONE* **18**, e0283828 (2023).
4. Yao, K. et al. Analyzing cell-scaffold interaction through unsupervised 3D nuclei segmentation. *Int. J. Bioprinting* **8**, 495 (2022).
5. Wu, L. et al. NISNet3D: three-dimensional nuclear synthesis and instance segmentation for fluorescence microscopy images. *Sci. Rep.* **13**, 9533 (2023).
6. Eschweiler, D., Rethwisch, M., Jarchow, M., Koppers, S. & Stegmaier, J. 3D fluorescence microscopy data synthesis for segmentation and benchmarking. *PLoS ONE* **16**, e0260509 (2021).



7. Eschweiler, D. et al. Denoising diffusion probabilistic models for generation of realistic fully-annotated microscopy image datasets. *PLoS Comput. Biol.* **20**, e1011890 (2024).
8. Saxena, S. & Teli, M. N. Comparison and analysis of image-to-image generative adversarial networks: a survey. Preprint at <https://arxiv.org/abs/2112.12625> (2021).
9. Desmaison, A. et al. Impact of physical confinement on nuclei geometry and cell division dynamics in 3D spheroids. *Sci. Rep.* **8**, 8785 (2018).
10. Böhland, M., Scherr, T., Bartschat, A., Mikut, R. & Reischl, M. Influence of synthetic label image object properties on GAN supported segmentation pipelines. In *Proc. 29th Workshop Computational Intelligence*, 289–305 (2019).
11. Amerreh, M., Edwards, R., Akbari, M. & Nadler, B. In-silico modeling of tumor spheroid formation and growth. *Micromachines* **12**, 749 (2021).
12. Bowers, H. J., Fannin, E. E., Thomas, A. & Weis, J. A. Characterization of multicellular breast tumor spheroids using image data-driven biophysical mathematical modeling. *Sci. Rep.* **10**, 11583 (2020).
13. Ozik, J. et al. High-throughput cancer hypothesis testing with an integrated PhysiCell-EMES workflow. *BMC Bioinformatics* **19**, 483 (2018).
14. Barham, K., Spencer, R., Baker, N. C. & Knudsen, T. B. Engineering a computable epiblast for in silico modeling of developmental toxicity. *Reprod. Toxicol.* **128**, 108625 (2024).
15. Zhang, Y. et al. Computational modeling to determine the effect of phenotypic heterogeneity in tumors on the collective tumor-immune interactions. *Bull. Math. Biol.* **85**, 51 (2023).
16. Graner, F. & Glazier, J. A. Simulation of biological cell sorting using a two-dimensional extended Potts model. *Phys. Rev. Lett.* **69**, 2013–2016 (1992).
17. Glazier, J. A. & Graner, F. Simulation of the differential adhesion driven rearrangement of biological cells. *Phys. Rev. E Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.* **47**, 2128–2154 (1993).
18. Swat, M. H. et al. Multi-scale modeling of tissues using CompuCell3D. *Methods Cell Biol.* **110**, 325–366 (2012).
19. Ferrari Gianlupi, J. et al. Multiscale model of antiviral timing, potency, and heterogeneity effects on an epithelial tissue patch infected by SARS-CoV-2. *Viruses* **14**, 605 (2022).
20. Shirinifard, A. et al. Adhesion failures determine the pattern of choroidal neovascularization in the eye: a computer simulation study. *PLoS Comput. Biol.* **8**, e1002440 (2012).
21. Svoboda, D. et al. Vascular network formation in silico using the extended cellular potts model. In *Proc. 2016 IEEE International Conference on Image Processing (ICIP)*, 3180–3183 (IEEE, 2016).
22. Libby, A. R. G. et al. Automated design of pluripotent stem cell self-organization. *Cell Syst.* **9**, 483–495 (2019).
23. Mulberry, N. & Edelstein-Keshet, L. Self-organized multicellular structures from simple cell signaling: a computational model. *Phys. Biol.* **17**, 066003 (2020).
24. Guisoni, N., Mazzitello, K. I. & Diambra, L. Modeling active cell movement with the Potts model. *Front. Phys.* **6**, 61 (2018).
25. Niculescu, I., Textor, J. & de Boer, R. J. Crawling and gliding: a computational model for shape-driven cell migration. *PLoS Comput. Biol.* **11**, e1004280 (2015).
26. Scianna, M. & Preziosi, L. A cellular Potts model for analyzing cell migration across constraining pillar arrays. *Axioms* **10**, 32 (2021).
27. Thomas, G. L., Fortuna, I., Perrone, G. C., Graner, F. & de Almeida, R. M. C. Shape-velocity correlation defines polarization in migrating cell simulations. *Phys. A Stat. Mech. Appl.* **587**, 126511 (2022).
28. Tikka, P. et al. Computational modelling of nephron progenitor cell movement and aggregation during kidney organogenesis. *Math. Biosci.* **344**, 108759 (2022).
29. Link, R. & Schwarz, U. S. Simulating 3D cell shape with the cellular Potts model. *Methods Mol. Biol.* **2600**, 323–339 (2023).
30. Hirway, S. U., Lemmon, C. A. & Weinberg, S. H. Multicellular mechanochemical hybrid cellular potts model of tissue formation during epithelial-mesenchymal transition. *Comput. Syst. Oncol.* **1**, e1031 (2021).
31. Albert, P. J. & Schwarz, U. S. Dynamics of cell shape and forces on micropatterned substrates predicted by a cellular Potts model. *Biophys. J.* **106**, 2340–2352 (2014).
32. Li, J. F. & Lowengrub, J. The effects of cell compressibility, motility and contact inhibition on the growth of tumor cell clusters using the cellular Potts model. *J. Theor. Biol.* **343**, 79–91 (2014).
33. Osborne, J. M. Multiscale model of colorectal cancer using the cellular Potts framework. *Cancer Inform.* **14**, S19332 (2015).
34. Szabó, A. & Merks, R. M. H. Cellular potts modeling of tumor growth, tumor invasion, and tumor evolution. *Front. Oncol.* **3**, 87 (2013).
35. Drasdo, D. & Höhme, S. A single-cell-based model of tumor growth in vitro: monolayers and spheroids. *Phys. Biol.* **2**, 133–147 (2005).
36. Kim, Y. & Othmer, H. G. A hybrid model of tumor-stromal interactions in breast cancer. *Bull. Math. Biol.* **75**, 1304–1350 (2013).
37. Norton, K.-A., Jin, K. & Popel, A. S. Modeling triple-negative breast cancer heterogeneity: effects of stromal macrophages, fibroblasts and tumor vasculature. *J. Theor. Biol.* **452**, 56–68 (2018).
38. Smallbone, K., Gatenby, R. A., Gillies, R. J., Maini, P. K. & Gavaghan, D. J. Metabolic changes during carcinogenesis: potential impact on invasiveness. *J. Theor. Biol.* **244**, 703–713 (2007).
39. Shirinifard, A. et al. 3D multi-cell simulation of tumor growth and angiogenesis. *PLoS ONE* **4**, e7190 (2009).
40. Rubenstein, B. M. & Kaufman, L. J. The role of extracellular matrix in glioma invasion: a cellular Potts model approach. *Biophys. J.* **95**, 5661–5680 (2008).
41. Jeanquartier, F., Jean-Quartier, C., Cemernek, D. & Holzinger, A. In silico modeling for tumor growth visualization. *BMC Syst. Biol.* **10**, 59 (2016).
42. Bińkowski, M., Sutherland, D. J., Arbel, M. & Gretton, A. Demystifying MMD GANs. Preprint at <https://arxiv.org/abs/1801.01401> (2018).
43. Ulman, V. et al. An objective comparison of cell-tracking algorithms. *Nat. Methods* **14**, 1141 (2017).
44. Weigert, M., Schmidt, U., Haase, R., Sugawara, K. & Myers, G. Star-convex polyhedra for 3D object detection and segmentation in microscopy. In *Proc. IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, 3666–3673 (2020).
45. Stringer, C., Wang, T., Michaelos, M. & Pachitariu, M. Cellpose: a generalist algorithm for cellular segmentation. *Nat. Methods* **18**, 100–106 (2021).
46. Cui, H. et al. scGPT: toward building a foundation model for single-cell multi-omics using generative AI. *Nat. Methods* **21**, 1470–1480 (2024).
47. Nürnberg, E. et al. From in vitro to in silico: a pipeline for generating virtual tissue simulations from real image data. *Front. Mol. Biosci.* **11**, 1467366 (2024).
48. Pachitariu, M. & Stringer, C. Cellpose 2.0: how to train your own model. *Nat. Methods* **19**, 1634–1641 (2022).
49. Zhu, J.-Y., Park, T., Isola, P. & Efros, A. A. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proc. IEEE International Conference on Computer Vision (ICCV)*, 2242–2251 (2017).
50. Bruch, R. et al. Data corresponding to paper: improving 3D deep learning segmentation with biophysically motivated cell synthesis <https://zenodo.org/records/11240362> (2024).
51. Bruch, R. et al. Github code: bruchr/cell\_synthesis: v1.0.0 <https://zenodo.org/records/14536298> (2024).

## Acknowledgements

This work was funded by the German Federal Ministry of Education and Research (BMBF) grant 01IS21062B. R.B. and R.R. were funded by the Carl-Zeiss Foundation, project DigiFIT. This work was funded by the German

Federal Ministry of Education and Research (BMBF) as part of the Innovation Partnership M2Aind, projects M2OGA (03FH8I02IA) and Drugs4Future (13FH8I05IA) within the framework Starke Fachhochschulen-Impuls für die Region (FH-Impuls). This work is supported by the Helmholtz Association Initiative and Networking Fund on the HAICORE@KIT partition. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

### Author contributions

R.B. and M.R. developed and implemented algorithms. R.B. and E.N. invented the approach to generate synthetic data. E.N. implemented and performed biophysical simulations. M.V. developed the biological sample preparation and performed subsequent imaging. R.B. designed and conducted the experiments. R.B. and M.R. were responsible for the manuscript. M.R., R.R. and S.S. conceived and supervised the project. All authors discussed the results and implications, participated in writing, and commented on the manuscript at all stages.

### Funding

Open Access funding enabled and organized by Projekt DEAL.

### Competing interests

The authors declare no competing interests.

### Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s42003-025-07469-2>.

**Correspondence** and requests for materials should be addressed to Roman Bruch.

**Peer review information** *Communications Biology* thanks Ervin Tasnadi, Gani Rahmon and the other anonymous reviewer(s) for their contribution to the peer review of this work. Primary handling editors: Aylin Bircan, Laura Rodriguez Perez.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2025