A Multi-Scale Inception-UNet with Structure-Aware Evaluation for Branch-Preserving Segmentation of Organoids

Author Sandra H. Andrusca[1] SANDRA.ANDRUSCA@TUM.DE
Author Christopher D. Kießling [1] CHRISTOPHER.KIESSLING@TUM.DE
Author Andreas R. Bausch [1] ABAUSCH@MYTUM.DE
[1] *Heinz Nixdorf Chair in Biophysical Engineering of Living Matter*
Center for Protein Assemblies (CPA)
Technical University of Munich (TUM)
85748 Garching, Germany

Editors: Under Review for MIDL 2026

Abstract

Branched organoids exhibit increasingly complex morphologies as they progress from simple spheroid states to highly ramified structures, making topology-preserving segmentation essential for quantitative biological analysis. Capturing thin protrusions and maintaining branch continuity remains challenging for classical UNet-based architectures, particularly in brightfield imaging where fine structures are easily blurred or disconnected.

In this work, we present a multi-scale Inception-UNet designed to capture the heterogeneous spatial scales of branched organoids through parallel convolutional paths with complementary receptive fields. As a model system, we analyze brightfield pancreatic ductal adenocarcinoma (PDAC) organoids, a system known for strong morphological heterogeneity and invasive branching behavior (Randriamanantsoa, 2022), cultured using high-throughput Patternoid assays (Kurzbach, 2025) that enable standardized imaging and robust quantitative analysis.

To assess segmentation quality beyond region overlap, we combine Dice with a structure-aware skeleton-based Dice score that directly probes branch integrity and topological continuity. Across deterministic seeds and strictly separated organoid positions, the Inception-UNet achieves the highest region-based Dice (0.8035 ± 0.0076) and skeleton-based Dice (0.2513 ± 0.0156), and most importantly, the strongest preservation of branch continuity compared to UNet and UNet++. These improvements become increasingly pronounced with growing morphological complexity.

Overall, our results demonstrate that multi-scale feature extraction combined with topology-aware evaluation substantially improves segmentation of branched organoids and provides a robust foundation for downstream morphological and invasion-related analyses.

Keywords: semantic segmentation, organoid imaging, PDAC, multi-scale architectures, UNet, topology preservation, skeleton-aware loss, brightfield microscopy, multi-seed evaluation, branch morphology

1. Introduction

Branched organoids have emerged as an important model system for studying epithelial morphogenesis, invasion, and structural organization. As these systems transition from compact spheroids into highly ramified architectures, branch integrity and topological structure become central descriptors of their biological behavior (Clevers, 2016; Boj, 2015). Pancreatic

ductal adenocarcinoma (PDAC) organoids in particular exhibit pronounced morphological heterogeneity and complex, invasive branching dynamics (Randriamanantsoa, 2022), making them a demanding but informative benchmark for topology-preserving segmentation.

In this work, we analyze PDAC organoids imaged using high-throughput Patternoid assays (Kurzbach, 2025), which provide standardized 3D culture conditions and longitudinal brightfield acquisition across large organoid populations. The resulting morphologies contain both thin protrusions and broader structural compartments, requiring segmentation methods that can jointly capture fine-scale and coarse-scale features.

Classical UNet architectures (Ronneberger et al., 2015) remain the dominant approach for biomedical image segmentation, yet their single-scale convolutional blocks often struggle to preserve thin branches in brightfield organoids. UNet++ (Zhou et al., 2018) improves feature propagation through nested skip connections, while Inception-style multi-scale representations (Szegedy et al., 2015) and residual designs (He et al., 2016) have proven effective for capturing heterogeneous spatial patterns in natural images. However, these ideas have not been systematically adapted to branched organoid morphologies, where preserving topological continuity is essential for downstream quantitative analysis.

To address these limitations, we introduce a multi-scale Inception-UNet that integrates parallel convolutional paths with complementary receptive fields within each encoder stage. This design allows the network to model both thin, elongated branches and larger morphological compartments simultaneously. Because region-based Dice (Long et al., 2015) does not fully reflect structural fidelity, we complement it with a skeleton-based Dice score that directly probes centerline continuity and branch preservation. All models are trained within a deterministic multi-seed PyTorch pipeline (Paszke et al., 2019) to ensure reproducibility.

Under identical training and evaluation settings with strictly separated organoid positions, the Inception-UNet achieves the highest Dice and skeleton-based Dice among all compared architectures and shows the clearest improvements in branch continuity. These results highlight the importance of multi-scale feature extraction and topology-aware evaluation for analyzing the complex morphologies of branched organoids.

Our contributions are threefold:

- We introduce a multi-scale Inception-UNet architecture tailored to the branched morphology of PDAC organoids cultured in high-throughput Patternoid assays, and evaluate it within a deterministic multi-seed analysis pipeline with strictly separated organoid positions.
- We establish a topology-aware evaluation protocol that combines region-based Dice with a skeleton-based Dice metric to specifically probe branch continuity and structural integrity beyond conventional overlap measures.
- We provide an initial analysis of a skeleton-aware auxiliary loss that augments standard region-based supervision with topology-focused signals, illustrating how explicit branch-aware objectives can further enhance structural fidelity.

1.1. Dataset

We analyze brightfield microscopy data of pancreatic ductal adenocarcinoma (PDAC) organoids, a system known for strong morphological heterogeneity and invasive, branched growth

patterns (Randriamanantsoa, 2022). The dataset is generated using the Patternoid high-throughput culture and imaging platform (Kurzbach, 2025), which enables standardized organoid formation, longitudinal acquisition, and consistent brightfield imaging across large numbers of samples.

Each organoid is recorded as an hourly time-lapse sequence, with every time point captured as a brightfield z-stack comprising approximately 100 slices to cover the full axial extent of the culture. For segmentation analysis, each stack is converted into a 2D standard-deviation (STD) projection computed over the in-focus portion of the z-stack. This projection enhances local intensity variations and improves the visibility of thin, elongated branches that are often indistinct in raw brightfield slices.

The dataset includes organoids spanning three established PDAC morphological subtypes (Boj, 2015; Clevers, 2016), ranging from compact cystic structures to highly branched invasive phenotypes. Since our task is binary semantic segmentation (foreground vs. background), all subtypes are pooled to maximize morphological diversity and support generalizable topology-aware learning. Representative examples are shown in Figure 1.

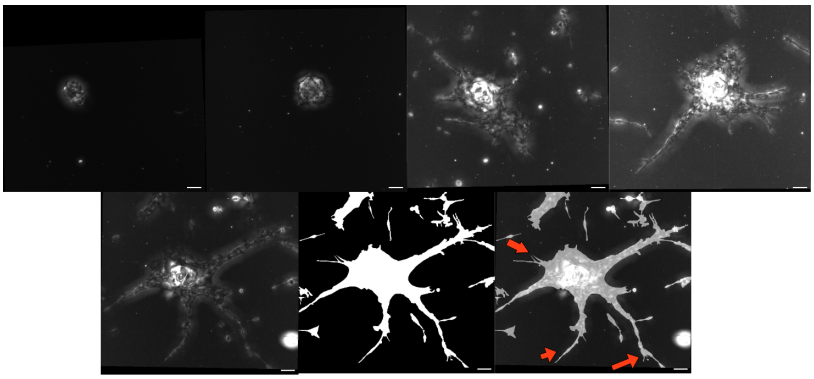

Figure 1: Representative standard-deviation (STD) projections of brightfield z-stacks of PDAC organoids imaged in Patternoid assays. **Top row:** Example organoids at four representative time windows (seeding, early phase around day 1, around day 2, and late/endpoint phase), illustrating the progression from compact spheroid-like structures to highly branched invasive morphologies. **Bottom row:** Zoomed late-stage organoid shown as raw STD projection (left), corresponding ground-truth mask (center), and overlay (right), with red arrows highlighting thin branches and complex regions that are particularly challenging to segment. Scale bar: 50 µm.

To avoid temporal and positional leakage, we enforce strict separation of organoid positions: all time points from a given organoid are assigned to exactly one of the training, validation, or test sets. Evaluation is therefore performed on entirely unseen organoids rather than additional time points of already observed ones.

Manual pixel-wise annotations were generated by trained experts on the projected images. The final dataset comprises distinct annotated training, validation, and test sets

that jointly cover early, intermediate, and late stages of organoid development. The exact number of images per split is reported in Section 2.

1.2. Preprocessing

Each brightfield z-stack is first rigidly registered to correct for sample drift and stage motion, ensuring spatial alignment across slices. We then restrict the stack to in-focus slices with sufficient structural contrast and compute a 2D standard-deviation (STD) projection per time point. This enhances high-frequency intensity variations and makes thin, elongated branches more visible than in raw slices or mean-intensity projections.

Projected frames are cropped to regions containing the organoid of interest, normalized, converted to 8-bit grayscale, and resized or cropped to a fixed resolution of 128×128 pixels. Data augmentation consists of random flips, small rotations, and mild brightness or contrast changes, restricted to transformations that preserve organoid topology. The same preprocessing and augmentation pipeline is applied to all model architectures.

1.3. Inception-UNet Architecture

Our model builds on the classical UNet encoder–decoder topology with skip connections, but replaces the standard encoder blocks with multi-scale Inception modules that capture features at different spatial scales (Figure 2). The network follows a contracting path that reduces spatial resolution while increasing feature dimensionality, and an expanding path that progressively restores spatial detail.

Each Inception encoder block comprises four parallel branches:

1. a 1×1 convolution for channel projection,
2. a standard 3×3 convolution ($d = 1$) for local context,
3. a dilated 3×3 convolution with depth-dependent dilation $d \in \{2, 4\}$ for mid- and long-range context,
4. an asymmetric 1×7 followed by 7×1 convolution to further expand the receptive field at moderate parameter cost.

The branch outputs are concatenated and passed through a squeeze-and-excitation (SE) block with global pooling and channel-wise gating, allowing the model to emphasize informative scales adaptively. A final 1×1 convolution fuses the reweighted features, which are then forwarded to the next encoder stage and to the decoder via skip connections.

The decoder follows a standard UNet design with upsampling (transpose convolutions or bilinear upsampling followed by a 3×3 convolution), concatenation with the corresponding encoder features, and two 3×3 refinement convolutions with non-linear activations. All convolutional layers are followed by batch normalization and a rectified linear unit (ReLU), except for the final 1×1 layer that maps to a single-channel logits map; a pointwise sigmoid is applied at inference to obtain foreground probabilities. Encoder–decoder depth and base channel width are matched for UNet, UNet++, and Inception-UNet to ensure a fair architectural comparison.

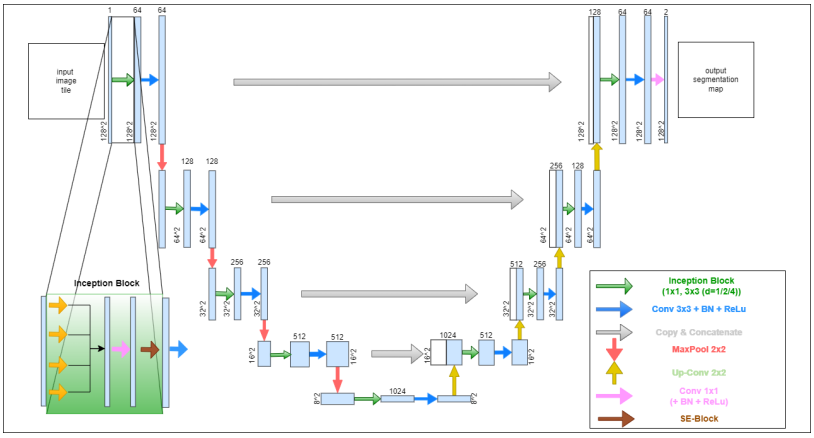

Figure 2: Overview of the proposed Inception-UNet. Each encoder stage uses a multi-scale Inception block with four parallel branches (1×1, 3×3, dilated 3×3, and asymmetric 1×7–7×1). Outputs are concatenated, reweighted by an SE block, and fused with a 1×1 convolution. The decoder follows the standard UNet design with upsampling, skip connections, and 3×3 refinement convolutions.

1.4. Structure-Aware Evaluation

Conventional region-based metrics such as Dice or Intersection-over-Union (IoU) primarily quantify area overlap between prediction and ground truth. While widely used in biomedical image segmentation, they are largely insensitive to topological errors and failures in thin or elongated structures. For branched PDAC organoids, such errors can directly distort downstream analyses of branch length, continuity, and growth trajectories.

To better capture structural fidelity, we complement region-based Dice with a skeleton-based Dice score. This metric evaluates the agreement between centerline representations of the predicted and ground-truth masks and thus focuses on the continuity of narrow branches rather than on area overlap alone. Skeletons are computed by applying a morphological thinning operation to the binary ground-truth and predicted masks (e.g. via standard skeletonization algorithms).

Let S_{gt} and S_{pred} denote the binary skeletons of the ground-truth and predicted segmentations, respectively. The skeleton-based Dice score is defined as

$$\text{Dice}_{\text{skel}} = \frac{2|S_{\text{gt}} \cap S_{\text{pred}}|}{|S_{\text{gt}}| + |S_{\text{pred}}| + \varepsilon}, \tag{1}$$

where ε is a small constant to ensure numerical stability. This formulation penalizes breaks, discontinuities, and missing branches, which have little effect on the standard Dice but are critical for faithfully capturing organoid morphology.

In practice, skeleton-based Dice reveals structural differences between models that region - based metrics may obscure, highlighting cases where segmentations have similar area overlap but differ substantially in the connectivity of thin protrusions. All models in this study are therefore assessed using both region-based Dice and skeleton-based Dice to provide a more topology-aware view of segmentation performance.

1.5. Exploratory Skeleton-Aware Loss

While our primary analysis relies on region-based Dice and skeleton-based Dice at evaluation time, we also ask whether explicit structural supervision during training can further improve branch continuity. To this end, we explore a lightweight skeleton-aware loss that augments conventional region-based objectives with an auxiliary term computed on organoid skeletons.

Let M_{gt} and M_{pred} denote the ground - truth and predicted masks, and let $S_{\text{gt}} = \text{Skel}(M_{\text{gt}})$ and $S_{\text{pred}} = \text{Skel}(M_{\text{pred}})$ denote their corresponding skeletons obtained via morphological thinning. The skeleton-aware term is defined as

$$\mathcal{L}_{\text{skel}} = \text{BCE}\left(S_{\text{pred}}, S_{\text{gt}}\right), \tag{2}$$

where BCE denotes pixel-wise binary cross-entropy. This term penalizes discrepancies between predicted and reference centerlines and explicitly encourages continuity in thin or elongated branches that may be underrepresented in purely region-based supervision.

The overall training loss is given by

$$\mathcal{L} = \alpha\,\mathcal{L}_{\text{Dice}} + \beta\,\mathcal{L}_{\text{BCE}} + \gamma\,\mathcal{L}_{\text{skel}}, \tag{3}$$

with weighting coefficients α, β, and γ chosen empirically. In this work, we treat this formulation as an exploratory extension rather than a fully tuned objective: the goal is to assess whether adding topology-focused supervision can provide measurable gains in continuity-based metrics.

Initial results indicate that the skeleton-aware loss can yield consistent improvements in skeleton-based Dice, with smaller or model-dependent effects on region-based Dice. A more systematic study of topology-aware loss design and hyperparameter choices is left for future work.

1.6. Training Procedure

All models are trained and evaluated under a multi-seed protocol using seeds $\{0, 1, 2\}$. For each seed, we initialize all relevant random number generators in Python, NumPy, and PyTorch to control stochastic components such as weight initialization, data shuffling, and augmentation. To keep training efficient on our hardware, we allow multi-threaded backends and tune the number of CPU workers to saturate the available cores, which yields numerically stable but not strictly bitwise-identical runs.

We use the Adam optimizer with an initial learning rate selected via hyperparameter search. Hyperparameters are tuned using Optuna (Akiba et al., 2019) on a validation-based objective, with UNet serving as the reference architecture. The search space includes the initial learning rate, the batch size, and the base number of feature channels in the first encoder layer. Each Optuna trial is trained with seeds $\{0, 1, 2\}$ for a reduced number of epochs, and the mean validation Dice across these seeds is used as the trial score. The best-performing configuration from this search—a small learning rate, a batch size of two, and a base width of 64 channels—is then fixed and used for all subsequent experiments with UNet, UNet++, and the proposed Inception-UNet to ensure a fair comparison.

Full training is performed for a fixed number of 75 epochs for each model and seed, without early stopping. All models are trained with a composite loss consisting of Dice

and binary cross-entropy (BCE); in the exploratory skeleton-aware setting, the skeleton-based BCE term described in Section 1.5 is added with a small weight. Training and validation splits are defined at the level of organoid positions to prevent temporal leakage. After training, we select the checkpoint with the best validation Dice for each seed and report mean and standard deviation over the three seeds on the held-out test set. Test-time predictions are binarized using a fixed probability threshold of 0.5 for all models and seeds.

1.7. Baseline Models

We compare the proposed Inception-UNet against two widely used architectures for biomedical image segmentation: the classical UNet and the nested UNet++. All baselines are matched to the proposed model in depth, encoder–decoder layout, and channel width to ensure that differences in performance primarily reflect architectural design rather than model capacity or training hyperparameters.

UNet. The standard UNet follows a five-level encoder–decoder structure with symmetric skip connections (Ronneberger et al., 2015). Each encoder stage applies two 3×3 convolutions with ReLU activations and 2×2 max-pooling for downsampling; the decoder mirrors this using transposed convolutions for upsampling, followed by two 3×3 convolutions and a final 1×1 layer to produce a single-channel logits map. Batch normalization is applied after each convolution, and the base number of feature channels is chosen via the hyperparameter search described in Section 1.6 and reused across all models.

UNet++. UNet++ extends UNet by introducing nested dense skip connections that refine encoder features before fusion in the decoder (Zhou et al., 2018). This design aims to reduce the semantic gap between encoder and decoder representations and to improve the propagation of fine structural details. Apart from the modified skip topology, encoder and decoder blocks are kept identical to those of the standard UNet to maintain architectural comparability.

Training Conditions. UNet, UNet++, and Inception-UNet are all trained under identical conditions, including optimizer and learning-rate settings, batch size, data augmentation, composite Dice+BCE loss (with optional skeleton-aware term), train/validation/test splits, and the deterministic multi-seed protocol. This setup enables a fair comparison focused on architectural effects.

2. Results

2.1. Quantitative Evaluation

We evaluate UNet, UNet++, and the proposed Inception-UNet on the held-out test set using region-based Dice and the structure-aware skeleton-based Dice (Section 1.4). All values are reported as mean ± standard deviation across three deterministic seeds.

Table **??** summarizes the results. Across seeds, the Inception - UNet attains the highest overall Dice, followed by UNet and UNet++. Absolute differences remain moderate but are consistent, indicating that the multi-scale encoder yields more reliable overlap performance across organoid morphologies.

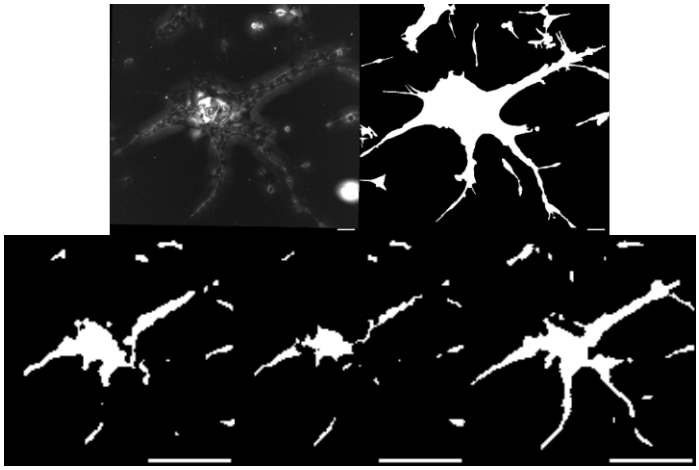

Figure 3: Qualitative comparison of segmentation performance. **Top:** Raw brightfield image with the corresponding ground-truth mask overlaid. **Bottom:** Predicted segmentation masks for the same sample generated by UNet (left), UNet++ (center), and the proposed Inception-UNet (right). Compared to the baselines, the Inception-UNet more faithfully preserves thin branches and overall topological continuity. Scale bar: 50 µm.

Table 1: Region-based Dice across time windows (mean ± std over three deterministic seeds).

Model	1–12h	24–30h	48–55h	60–70h
UNet	0.828 ± 0.014	0.779 ± 0.003	0.775 ± 0.004	0.760 ± 0.005
UNet++	0.846 ± 0.008	0.782 ± 0.000	0.717 ± 0.014	0.705 ± 0.008
Inception U-Net	0.817 ± 0.010	0.790 ± 0.005	0.814 ± 0.004	0.777 ± 0.009

Skeleton-based Dice reveals clearer separations between architectures. It penalizes discontinuities and missing branches that standard Dice largely overlooks, and consistently ranks the Inception-UNet highest. This indicates that the proposed model not only improves region overlap but also better preserves thin, elongated structures and branch continuity compared to UNet and UNet++.

2.2. Analysis of Structural Fidelity

Standard Dice captures region overlap but overlooks topological defects such as broken or discontinuous branches. As illustrated in Figure **??**, UNet and UNet++ frequently exhibit small gaps in thin protrusions, whereas the Inception-UNet produces skeletons that follow ground - truth centerlines more closely and maintain branch connectivity more reliably. Skeleton-based Dice quantifies these differences and reveals structural discrepancies that are largely invisible to region-based metrics.

Table 2: Skeleton-based Dice across time windows (mean ± std over three deterministic seeds).

Model	1–12h	24–30h	48–55h	60–70h
UNet	0.184 ± 0.022	0.247 ± 0.003	0.270 ± 0.011	0.245 ± 0.011
UNet++	0.220 ± 0.012	0.255 ± 0.004	0.251 ± 0.011	0.240 ± 0.001
Inception U-Net	0.185 ± 0.008	0.263 ± 0.013	0.303 ± 0.032	0.280 ± 0.009

2.3. Stability Across Seeds

All models show low seed-to-seed variation ($< 1\%$ Dice and similarly small for Skeleton-Dice), indicating numerically stable training. The performance ranking remains consistent across seeds, with Inception-UNet leading both Dice and Skeleton-Dice and UNet++ showing slightly higher variability in highly branched morphologies.

2.4. Ablation: Architecture vs. Structure-Aware Metrics

Region-based Dice yields small differences between architectures, reflecting its sensitivity to thick organoid regions. Skeleton-based Dice, in contrast, amplifies architectural distinctions and consistently favors the Inception-UNet, demonstrating that its multi-scale encoder more effectively preserves thin branches and topological continuity.

2.5. Exploratory Skeleton-Aware Loss

We also tested an auxiliary skeleton-aware loss that adds BCE supervision on centerlines. Preliminary results indicate modest improvements in Skeleton-Dice, particularly in thin branches, while region-based Dice remains unchanged. Although exploratory, these findings highlight the potential of topology-focused objectives for future segmentation pipelines.

3. Discussion

The results of this study demonstrate that accurate and topology-preserving segmentation of highly branched PDAC organoids from brightfield microscopy is feasible, despite low contrast and limited annotated data. While region-based Dice remains the standard metric in biomedical image segmentation, our structure-aware evaluation clearly shows that models with similar Dice scores can differ substantially in their ability to preserve thin branches, maintain connectivity, and reconstruct the underlying topology.

Within this setting, the proposed Inception-UNet achieves the highest overall region-based Dice and skeleton-based Dice across deterministic seeds, and shows the strongest gains in later, highly branched stages. This indicates that the multi-scale encoder with parallel receptive fields is well suited to the heterogeneous spatial scales present in organoid morphologies. In contrast, UNet++ does not offer clear advantages under the constraints of this dataset, suggesting that nested skip refinements are less effective when fine structural cues are weak or scarce in brightfield projections.

An important outcome of this work is the combination of a deterministic multi-seed training protocol with topology-aware evaluation. The low variance across seeds confirms that the pipeline is numerically stable and that observed differences between architectures are robust rather than driven by stochastic effects. At the same time, skeleton-based Dice exposes structural discrepancies that are largely invisible to area-based overlap metrics, highlighting the need for topology-aware assessment in applications where branch integrity is critical.

Overall, this study establishes a reproducible, topology-aware benchmark for bright-field PDAC organoid segmentation and presents a multi-scale Inception-UNet that learns to segment complex branching morphologies while better preserving topological structure. As future work explores richer topology-preserving objectives and additional datasets, we expect such architectures and evaluation protocols to become a foundation for downstream analyses of organoid growth, invasion, and morphodynamic behavior.

4. Outlook

Future work will focus on developing topology-preserving evaluation measures beyond Dice and Skeleton-Dice to more precisely quantify branch integrity and subtle structural defects in complex organoid morphologies.

Acknowledgments

We thank Sophie Kurzbach for providing the experimental data. We acknowledge the support by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) through the project BA 2029/15-1 and by the European Research Council under the European Union's Horizon 2020 research and innovation programme (grant agreement no. 810104-PoInt).

References

Takuya Akiba, Shotaro Sano, Toshihiko Yanase, Takeru Ohta, and Masanori Koyama. Optuna: A next-generation hyperparameter optimization framework. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2019.

Sylvia F. et al. Boj. Organoid models of human and mouse ductal pancreatic cancer. *Cell*, 160:324–338, 2015.

Hans Clevers. Modeling development and disease with organoids. *Cell*, 165:1586–1597, 2016.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.

S. et al. Kurzbach. Patternoid: A high-throughput platform for standardized 3D organoid culture and longitudinal imaging. *Lab on a Chip*, 2025.

Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015.

Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Kopf, Edward Yang, Zachary DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. PyTorch: An imperative style, high-performance deep learning library. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2019.

Samuel et al. Randriamanantsoa. A novel pancreatic ductal adenocarcinoma organoid model reveals interpatient morphological and transcriptional heterogeneity. *Nature Communications*, 2022.

Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, 2015.

Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper

with convolutions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015.

Zongwei Zhou, Md Mahfuzur Rahman Siddiquee, Nima Tajbakhsh, and Jianming Liang. UNet++: A nested U-Net architecture for medical image segmentation. In *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, 2018.

Appendix A. Training Plots

Appendix B. Box Plots Results without Skeleton Loss

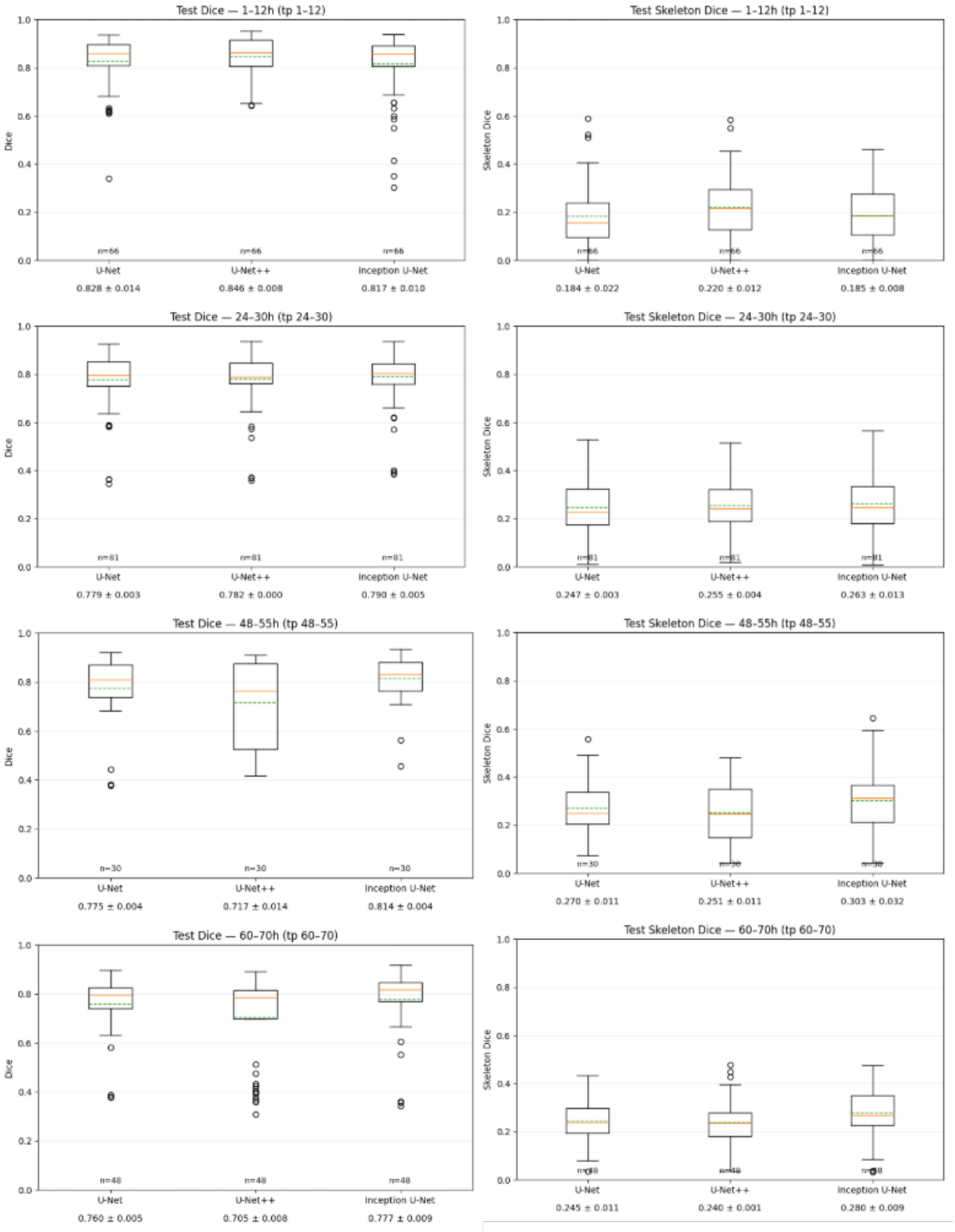

Figure 4: Boxplot comparison of region-based Dice (left) and skeleton-based Dice (right) for UNet, UNet++, and Inception-UNet across four time windows (1–12h, 24–30h, 48–55h, 60–70h). Each box summarizes per-sample test performance over three deterministic seeds. Skeleton-based Dice reveals clearer differences between models, with Inception-UNet showing the strongest branch preservation in the later windows.