Critique Open Research & Review (CORR)

eISSN: 3023-3267
Volume. 04, Issue. 02, pp. 16-26, April 2026"

Hybrid Attention-Convolution Framework with Shape-Sensitive Optimization for Improved Three-Dimensional Partitioning in Medical and Cellular Imaging

Dr. Haruto Nakamura
Faculty of Interdisciplinary Studies Kyoto International University of Science Kyoto, Japan

Dr. Yuki Matsumoto
Center for Applied Multidisciplinary Innovation Osaka Institute of Integrated Technology Osaka, Japan

Article received: 19/02/2026, Article Accepted: 27/03/2026, Article Published: 07/04/2026

ABSTRACT

Accurate three-dimensional partitioning of medical and cellular imaging data remains a fundamental challenge in biomedical image analysis due to the complex morphology, multi-scale structures, and high variability present in volumetric datasets. Conventional convolutional neural networks have demonstrated strong performance in segmentation tasks; however, they often struggle to capture long-range dependencies and global contextual relationships required for precise boundary delineation in three-dimensional environments. Transformer-based architectures address global context modeling but frequently introduce high computational complexity and insufficient spatial detail preservation when applied to volumetric data. To overcome these limitations, this study proposes a hybrid attention-convolution framework combined with a shape-sensitive optimization strategy designed to enhance structural consistency and boundary accuracy in three-dimensional segmentation of medical and microscopic images.

The proposed framework integrates convolutional feature extraction with multi-head attention mechanisms to jointly capture local spatial patterns and global contextual dependencies. A multi-branch hybrid encoder is developed to fuse convolutional and transformer-based representations, enabling robust feature learning across multiple scales. In addition, a shape-sensitive loss formulation is introduced to improve segmentation accuracy by enforcing geometric consistency using curvature-aware and distance-based constraints. This optimization strategy allows the model to preserve fine anatomical details and maintain topological correctness, which are critical for applications such as organoid analysis, tumor boundary detection, and volumetric clinical imaging.

The effectiveness of the proposed approach is evaluated through extensive experiments on three-dimensional medical and cellular datasets. Comparative analysis with state-of-the-art architectures, including U-Net variants, transformer-based segmentation models, and multi-aperture fusion networks, demonstrates consistent improvements in segmentation accuracy, boundary preservation, and structural stability. The results indicate that combining attention-driven global modeling with convolutional spatial learning and shape-sensitive optimization provides a balanced and computationally efficient solution for complex volumetric segmentation tasks.

This work contributes a unified segmentation framework that advances current research in medical image analysis by improving three-dimensional partitioning performance while maintaining scalability and robustness across heterogeneous imaging modalities.

Keywords: 3D medical image segmentation, hybrid neural networks, attention mechanism, convolutional networks, shape-aware loss, volumetric imaging, transformer models, biomedical image analysis, deep learning segmentation.

INTRODUCTION

Three-dimensional image partitioning plays a central role in modern medical and cellular imaging analysis, enabling quantitative evaluation of anatomical structures, pathological regions, and microscopic cellular formations. Advances in imaging technologies such as magnetic resonance imaging, computed tomography, and high-resolution microscopy have produced increasingly

complex volumetric datasets that require precise and reliable segmentation algorithms for downstream analysis. Accurate segmentation is essential for clinical diagnosis, treatment planning, organoid modeling, and high-content biological profiling, yet the complexity of volumetric data continues to pose significant challenges for computational methods (Ronneberger et al., 2015; Xing and Yang, 2016).

Early deep learning approaches for image segmentation relied primarily on convolutional neural networks, with the U-Net architecture becoming one of the most influential frameworks due to its encoder–decoder design and skip connections that preserve spatial information (Ronneberger et al., 2015). Extensions such as 3D U-Net enabled volumetric processing by replacing two-dimensional operations with three-dimensional convolutions, allowing improved performance in medical imaging tasks involving volumetric structures (Çiçek et al., 2016). Subsequent developments introduced self-configuring architectures such as nnU-Net, which automatically adapts network configuration to different datasets and has achieved strong results across a wide range of biomedical segmentation benchmarks (Isensee et al., 2021).

Despite their success, convolution-based architectures exhibit inherent limitations when modeling long-range dependencies in large volumetric images. Convolutional kernels operate within local receptive fields, which restricts the ability to capture global context, particularly in cases where anatomical structures extend across large spatial regions. To address this limitation, transformer-based models were introduced into medical image segmentation, inspired by their success in natural language processing and computer vision (Dosovitskiy et al., 2020). Architectures such as TransUNet and MISSFormer integrate self-attention mechanisms to enhance global feature representation, enabling improved segmentation performance in complex scenarios (Chen et al., 2021; Huang et al., 2023).

While transformer-based methods improve contextual modeling, they often require high computational resources and may lose fine spatial details due to tokenization and patch-based processing. This issue becomes more pronounced in three-dimensional segmentation, where volumetric data significantly increases memory requirements. Recent studies have explored hybrid architectures that combine convolutional operations with attention mechanisms to balance efficiency and accuracy (Xie et al., 2021; Lee et al., 2022). These approaches aim to retain the local feature extraction strength of convolutional networks while leveraging the global reasoning capability of transformers.

Another critical challenge in volumetric segmentation is preserving structural integrity and geometric consistency. Standard loss functions such as cross-entropy and Dice loss focus primarily on pixel-wise accuracy and often fail to maintain correct topology, leading to fragmented or irregular boundaries. To address this issue, shape-aware and topology-preserving loss functions have been proposed, including Hausdorff distance-based optimization, curvature-aware penalties, and persistent homology constraints (Karimi and Salcudean, 2020; Clough et al., 2022; Xing et al., 2022). These methods improve boundary accuracy but are rarely integrated into hybrid attention–convolution architectures designed for three-dimensional imaging.

Recent research in multi-aperture and attention-fusion networks has shown that combining multiple feature extraction pathways can significantly enhance segmentation accuracy in volumetric medical images (Shabani et al., 2024; Sohaib et al., 2025). Multi-branch designs allow the network to analyze structures at different spatial scales while maintaining computational efficiency. However, existing approaches still struggle to simultaneously achieve global context modeling, local detail preservation, and shape-consistent optimization within a unified framework.

Motivated by these limitations, this study proposes a hybrid attention–convolution framework with shape-sensitive optimization for improved three-dimensional partitioning in medical and cellular imaging. The proposed method integrates convolutional and transformer-based feature extraction within a multi-branch architecture, combined with a geometry-aware loss formulation that enhances structural accuracy during training. This design aims to provide a balanced solution capable of handling complex volumetric data without sacrificing computational efficiency or boundary precision.

The main objectives of this research are threefold. First, to design a hybrid segmentation architecture that effectively combines local convolutional learning with global attention modeling. Second, to develop a shape-sensitive optimization strategy that improves boundary accuracy and topological consistency in three-dimensional segmentation. Third, to evaluate the proposed framework against existing state-of-the-art models using volumetric medical and microscopic imaging datasets.

The scope of this work focuses on three-dimensional segmentation in clinical and cellular imaging, including organoid datasets, volumetric medical scans, and microscopy images. The proposed approach is intended to be generalizable across different imaging modalities while maintaining high accuracy and structural reliability.

The significance of this research lies in its potential to improve automated analysis in biomedical applications where precise segmentation is essential. Enhanced three-dimensional partitioning can support better disease

diagnosis, improved biological modeling, and more accurate quantitative analysis in both clinical and research environments. By integrating hybrid feature learning with shape-aware optimization, the proposed framework contributes to the advancement of deep learning–based segmentation methods for complex volumetric imaging.

2. Literature Review

Accurate segmentation of medical and cellular images has been a long-standing research problem in computer vision and biomedical engineering. The development of deep learning has significantly advanced the state of the art, particularly through convolutional neural networks, transformer-based architectures, and hybrid models that combine multiple feature learning strategies. This section reviews prior work related to convolutional segmentation networks, transformer-based models, hybrid attention–convolution frameworks, and shape-sensitive optimization methods, with emphasis on studies relevant to three-dimensional medical and microscopic image partitioning.

The introduction of the U-Net architecture marked a major milestone in biomedical image segmentation due to its encoder–decoder structure and skip connections that allow precise localization while preserving contextual information (Ronneberger et al., 2015). U-Net demonstrated strong performance on limited datasets and became the foundation for many subsequent segmentation models. To support volumetric data, the architecture was extended to 3D U-Net, which replaced two-dimensional operations with three-dimensional convolutions, enabling improved performance on volumetric medical images such as MRI and CT scans (Çiçek et al., 2016). Later, nnU-Net introduced an adaptive framework capable of automatically configuring network parameters for different datasets, showing that proper architectural adaptation plays a critical role in segmentation performance (Isensee et al., 2021). These convolution-based models remain highly effective for local feature extraction but are limited in capturing long-range dependencies due to the restricted receptive field of convolutional kernels.

To address the limitations of convolutional networks, transformer-based architectures were introduced for visual tasks. The Vision Transformer demonstrated that self-attention mechanisms can model global relationships across an image by processing tokens representing image patches (Dosovitskiy et al., 2020). Inspired by this concept, several studies applied transformers to medical image segmentation. TransUNet integrated transformer layers within the U-Net encoder to improve global context modeling while retaining spatial detail through convolutional decoding (Chen et al., 2021). An improved version further refined the architecture by analyzing the interaction between convolutional and transformer modules to enhance segmentation accuracy (Chen et al., 2024). Similarly, MISSFormer proposed an efficient transformer-based encoder designed specifically for segmentation tasks, achieving improved performance by combining multi-scale attention with hierarchical feature extraction (Huang et al., 2023).

Although transformer-based models provide better contextual understanding, they often require high computational cost, especially when processing volumetric data. To overcome this limitation, hybrid architectures have been proposed that combine convolutional operations with attention mechanisms. SegFormer introduced a lightweight transformer-based design that maintains efficiency while preserving multi-scale features (Xie et al., 2021). UX-Net further improved volumetric segmentation by incorporating large-kernel convolutional operations with hierarchical attention, enabling effective modeling of three-dimensional structures without excessive computational overhead (Lee et al., 2022). These hybrid approaches demonstrate that combining convolution and attention allows better balance between local detail extraction and global reasoning.

Recent research has also explored multi-branch and multi-aperture feature extraction strategies for volumetric segmentation. Multi-aperture transformer networks have been shown to improve segmentation of clinical and microscopic images by processing multiple receptive fields simultaneously (Sohaib et al., 2025). Similarly, fusion-based architectures that combine transformer and convolutional pathways provide enhanced feature representation across different spatial scales (Shabani et al., 2024). These methods are particularly useful in medical and cellular imaging, where objects vary significantly in size and shape. The use of attention modules such as convolutional block attention has also been shown to improve feature refinement by selectively focusing on informative regions (Woo et al., 2018).

In addition to architectural improvements, loss function design plays a crucial role in segmentation accuracy. Traditional loss functions such as cross-entropy and Dice loss optimize pixel-wise classification but do not guarantee geometric consistency. This limitation often leads to irregular boundaries or disconnected structures in three-dimensional segmentation. To address this problem, several shape-sensitive optimization strategies have been proposed. Hausdorff distance–based loss functions reduce boundary error by penalizing large deviations between predicted and ground-truth contours (Karimi and Salcudean, 2020). Modified Hausdorff distance metrics have also been used to improve object matching accuracy in segmentation tasks (Dubuisson and Jain, 1994). Curvature-aware loss formulations further enhance boundary smoothness by enforcing geometric constraints during training (Xing et al., 2022).

Topology-preserving loss functions have been introduced to maintain structural correctness in segmentation results.

Persistent homology–based loss functions ensure that predicted structures maintain the same topological properties as the ground truth, preventing fragmentation and unrealistic shapes (Clough et al., 2022). Similarly, topology-aware loss functions such as clDice preserve tubular and elongated structures, which are common in biomedical images (Shit et al., 2021). These methods demonstrate that incorporating geometric and topological information into optimization can significantly improve segmentation reliability.

Another important direction in segmentation research involves the use of large-scale pretraining and foundation models. The Segment Anything framework demonstrated that general-purpose segmentation models can be trained on large datasets and adapted to various tasks (Kirillov et al., 2023). Extensions of this approach to medical imaging have shown promising results, but adapting general models to volumetric biomedical data remains challenging due to domain differences and the need for high-resolution 3D processing (Ma et al., 2024; Bui et al., 2024). Self-supervised transformer pretraining has also been explored to improve performance in 3D medical image analysis by learning representations without extensive annotation (Tang et al., 2022).

In cellular and organoid imaging, segmentation accuracy is particularly important because small structural differences may correspond to significant biological changes. Studies on organoid models and three-dimensional cell cultures highlight the need for precise volumetric analysis to understand morphological behavior and disease progression (Srivastava et al., 2020; Han et al., 2010). Deep learning models designed for organoid profiling have shown that specialized architectures and loss functions are required to handle complex cellular structures (Sohaib et al., 2025; Winkelmaier and Parvin, 2021). These findings indicate that segmentation frameworks must preserve fine structural details while maintaining global consistency.

Despite the progress achieved by existing methods, several limitations remain. Pure convolutional networks lack global context modeling, transformer-based models are computationally expensive for volumetric data, and many hybrid approaches do not incorporate shape-aware optimization. In addition, most existing segmentation frameworks focus on either architectural improvement or loss function design, but rarely combine both within a unified system. This gap suggests the need for a framework that simultaneously integrates hybrid attention–convolution feature extraction with geometry-aware optimization for accurate three-dimensional segmentation.

Based on these observations, the present study proposes a hybrid attention–convolution framework combined with shape-sensitive optimization to address the limitations of existing segmentation models. By integrating multi-branch feature learning, attention-based global modeling, and geometry-aware loss functions, the proposed approach aims to achieve improved accuracy, structural consistency, and robustness in three-dimensional partitioning of medical and cellular imaging data.

3. Proposed Framework Overview

The objective of the proposed study is to design a unified segmentation framework capable of accurately partitioning complex three-dimensional medical and cellular images while maintaining structural consistency and computational efficiency. To achieve this goal, a hybrid attention–convolution architecture is combined with a shape-sensitive optimization strategy that enforces geometric correctness during training. The framework is designed to overcome three major limitations observed in existing methods: insufficient global context modeling in convolutional networks, high computational cost in transformer-based models, and lack of structural constraints in standard loss functions. The proposed model integrates these components into a multi-branch encoder–decoder pipeline that can effectively process volumetric data across different imaging modalities.

The overall framework consists of four main components: a hybrid encoder that combines convolutional and attention-based feature extraction, a multi-scale feature fusion mechanism, a geometry-aware loss formulation, and a three-dimensional decoder that reconstructs the segmentation map. Each component is designed to complement the others, enabling the network to learn both local spatial details and global contextual relationships while maintaining shape consistency.

3.1 Hybrid Attention–Convolution Architecture

Convolutional neural networks are effective for capturing local spatial features because convolutional kernels operate on neighboring pixels, allowing the model to detect edges, textures, and fine structural details. However, their limited receptive field makes it difficult to capture long-range dependencies in large volumetric images. Transformer-based attention mechanisms, on the other hand, allow global interactions between distant regions but may lose precise spatial information due to tokenization and patch-based processing (Dosovitskiy et al., 2020; Chen et al., 2021).

To balance these characteristics, the proposed architecture uses a hybrid encoder that combines convolutional blocks with attention modules. The convolutional pathway extracts low-level and mid-level spatial features, while the attention pathway captures global relationships between different regions of the volume. These two feature streams are fused at multiple stages to produce a richer representation of the input data. Similar hybrid designs have shown improved performance in medical segmentation tasks by leveraging the complementary strengths of convolution and

attention mechanisms (Huang et al., 2023; Lee et al., 2022).

The encoder is organized into hierarchical stages, where each stage reduces spatial resolution while increasing feature depth. At each level, convolutional layers perform local feature extraction, followed by an attention block that models global dependencies. The attention block computes relationships between feature tokens across the entire volume, allowing the network to understand structural context beyond the local neighborhood. This design improves segmentation accuracy in cases where anatomical structures extend across large regions.

3.2 Multi-Branch Feature Extraction and Fusion

Medical and cellular images often contain structures with highly variable size and shape. Small cellular components and large anatomical regions may appear in the same volume, making single-scale feature extraction insufficient. To address this challenge, the proposed framework introduces a multi-branch feature extraction strategy inspired by multi-aperture and fusion-based networks (Shabani et al., 2024; Sohaib et al., 2025).

In the multi-branch encoder, multiple convolutional kernels with different receptive fields operate in parallel. One branch uses small kernels to capture fine details, another uses larger kernels to capture coarse structures, and a third branch incorporates attention-based processing for global context. The outputs of these branches are fused using a weighted aggregation module that learns the relative importance of each feature stream.

Feature fusion is performed using channel-wise attention to emphasize informative features while suppressing noise. This mechanism allows the network to adaptively focus on relevant regions of the volume. Similar attention-based fusion strategies have been shown to improve segmentation performance in complex biomedical datasets (Woo et al., 2018). By combining multi-scale convolutional features with attention-based context modeling, the proposed framework provides a robust representation suitable for volumetric segmentation.

3.3 Shape-Sensitive Optimization Model

While architectural improvements enhance feature representation, segmentation accuracy also depends strongly on the loss function used during training. Standard loss functions such as cross-entropy and Dice loss measure pixel-wise differences between predicted and ground-truth labels, but they do not enforce geometric correctness. As a result, segmentation outputs may contain irregular boundaries, disconnected regions, or unrealistic shapes, especially in three-dimensional images.

To address this limitation, a shape-sensitive optimization model is introduced. This model combines multiple loss components designed to enforce boundary accuracy, geometric smoothness, and topological consistency. The first component is a distance-based loss that penalizes deviations between predicted and true boundaries using the Hausdorff distance, which has been shown to improve boundary precision in medical segmentation (Karimi and Salcudean, 2020; Dubuisson and Jain, 1994).

The second component is a curvature-aware penalty that encourages smooth and anatomically plausible surfaces. Curvature-based loss functions reduce jagged edges and improve the continuity of segmented structures (Xing et al., 2022). This is particularly important in volumetric imaging, where small boundary errors can propagate across slices and produce unrealistic three-dimensional shapes.

The third component is a topology-preserving constraint that ensures structural correctness. Persistent homology–based loss functions maintain the connectivity and topology of segmented objects, preventing fragmentation or merging of structures that should remain separate (Clough et al., 2022). For tubular and elongated structures, topology-aware losses such as clDice are incorporated to preserve thin connections that are often lost in standard optimization (Shit et al., 2021).

The final loss function is a weighted combination of these components together with Dice loss, allowing the network to optimize both pixel-level accuracy and global structure simultaneously.

3.4 Three-Dimensional Decoder and Reconstruction

After feature extraction and fusion, the decoder reconstructs the segmentation map at full resolution. The decoder follows an encoder–decoder design similar to U-Net but incorporates attention-guided skip connections. These connections transfer high-resolution features from the encoder to the decoder, ensuring that spatial detail is preserved during reconstruction (Ronneberger et al., 2015).

Each decoding stage consists of up-sampling, convolutional refinement, and attention-based feature selection. The up-sampling operation restores spatial resolution, while convolutional layers refine the feature map. Attention modules are used to filter irrelevant information and enhance important structures. This design improves the accuracy of boundary reconstruction, especially in complex volumetric datasets.

To support three-dimensional data, all operations in the decoder are implemented using volumetric convolutions. This allows the network to maintain spatial consistency across slices and generate smooth three-dimensional segmentation results. Similar volumetric architectures have been shown to outperform two-dimensional

approaches in medical imaging tasks (Çiçek et al., 2016; Zhou, 2023).

3.5 Three-Dimensional Segmentation Pipeline

The complete segmentation pipeline begins with preprocessing, where volumetric images are normalized and resized to a fixed resolution. The processed data are then passed through the hybrid encoder, where multi-scale features are extracted using convolutional and attention-based branches. The fused features are forwarded to the decoder, which reconstructs the segmentation map using attention-guided skip connections.

During training, the shape-sensitive loss function evaluates both pixel-level accuracy and structural correctness. This ensures that the network learns not only to classify each voxel correctly but also to maintain realistic shapes and boundaries. During inference, the trained model produces a three-dimensional partition of the input volume, which can be used for quantitative analysis, visualization, or clinical decision support.

The proposed framework provides a unified solution that integrates hybrid feature learning, multi-scale fusion, and geometry-aware optimization. By combining these components, the model aims to achieve higher accuracy, improved boundary preservation, and better structural consistency than existing segmentation approaches

4. Experimental Setup

To evaluate the effectiveness of the proposed hybrid attention–convolution framework with shape-sensitive optimization, extensive experiments were conducted on three-dimensional medical and cellular imaging datasets. The experimental design focuses on comparing the proposed model with representative convolutional, transformer-based, and hybrid segmentation architectures. The evaluation considers segmentation accuracy, boundary precision, structural consistency, and computational efficiency to demonstrate the advantages of the proposed approach in volumetric partitioning tasks.

4.1 Dataset Description

The experiments were performed using three-dimensional medical and microscopic imaging datasets representing different segmentation challenges. These datasets include volumetric clinical scans, microscopy images, and organoid imaging data. Such datasets contain structures with large variability in size, shape, and intensity distribution, making them suitable for evaluating segmentation robustness. Previous studies have shown that accurate segmentation of organoids and cellular structures requires models capable of preserving fine morphological details while maintaining global consistency (Srivastava et al., 2020; Sohaib et al., 2025).

The volumetric datasets were preprocessed using normalization and resizing to ensure consistent spatial resolution across samples. Intensity normalization was applied to reduce variations caused by different imaging devices. Data augmentation techniques, including rotation, scaling, and flipping, were used to improve model generalization. Similar preprocessing strategies have been widely used in medical image segmentation research to reduce overfitting and improve performance (Isensee et al., 2021).

4.2 Implementation Details

The proposed framework was implemented using a three-dimensional encoder–decoder architecture with hybrid attention–convolution blocks. The encoder consists of multiple hierarchical stages with convolutional layers followed by attention modules. Each stage reduces spatial resolution while increasing feature depth, enabling the network to capture both local and global information. Multi-branch convolutional blocks with different kernel sizes were used to extract features at multiple scales.

The attention mechanism was implemented using multi-head self-attention, which allows the model to learn relationships between distant regions of the volume. Feature fusion was performed using channel attention to emphasize informative features. Similar attention-based feature selection methods have been shown to improve segmentation accuracy in previous studies (Woo et al., 2018; Huang et al., 2023).

The decoder follows a volumetric U-Net–style structure with skip connections that transfer high-resolution features from the encoder. Each decoding stage includes up-sampling, convolutional refinement, and attention-guided feature selection. This design ensures that spatial details are preserved during reconstruction. Volumetric convolutions were used in all layers to maintain three-dimensional consistency (Çiçek et al., 2016; Zhou, 2023).

Training was performed using the proposed shape-sensitive loss function, which combines Dice loss, distance-based loss, curvature-aware penalty, and topology-preserving constraints. The combination of these loss components allows the model to optimize both voxel-level accuracy and structural correctness. Previous research has shown that distance-based and topology-aware losses significantly improve segmentation boundaries (Karimi and Salcudean, 2020; Clough et al., 2022; Xing et al., 2022).

The model was trained using stochastic gradient descent with adaptive learning rate scheduling. Training continued until convergence, and the best model was selected based on validation accuracy.

4.3 Evaluation Metrics

Segmentation performance was evaluated using standard metrics commonly used in medical image analysis. The Dice similarity coefficient was used to measure overlap between predicted and ground-truth segmentation. Intersection-over-union was used to evaluate region accuracy, while Hausdorff distance measured boundary error. These metrics provide complementary information about segmentation quality and are widely used in previous studies (Karimi and Salcudean, 2020; Isensee et al., 2021).

In addition to accuracy metrics, structural consistency was evaluated by analyzing boundary smoothness and topology preservation. This is important in volumetric segmentation, where small boundary errors can produce large structural artifacts. Computational efficiency was also measured to ensure that the hybrid architecture does not introduce excessive overhead compared with existing models.

4.4 Comparative Analysis

The proposed framework was compared with several representative segmentation models, including convolutional networks, transformer-based architectures, and hybrid models. U-Net and 3D U-Net were used as baseline convolutional methods due to their strong performance in medical imaging (Ronneberger et al., 2015; Çiçek et al., 2016). nnU-Net was included because of its adaptive configuration capability (Isensee et al., 2021).

Transformer-based models such as TransUNet, MISSFormer, and SegFormer were used to evaluate the effect of attention-based global modeling (Chen et al., 2021; Huang et al., 2023; Xie et al., 2021). UX-Net and nnFormer were included as advanced volumetric architectures designed for three-dimensional segmentation (Lee et al., 2022; Zhou, 2023).

Hybrid and multi-aperture networks were also included for comparison because they combine convolutional and attention-based features. These models have shown strong performance in volumetric segmentation tasks and provide a relevant benchmark for the proposed method (Shabani et al., 2024; Sohaib et al., 2025).

The comparison focuses on three main aspects: segmentation accuracy, boundary preservation, and structural consistency. These criteria are essential for evaluating segmentation performance in medical and cellular imaging.

5. Results

The experimental results demonstrate that the proposed hybrid attention–convolution framework with shape-sensitive optimization achieves consistent improvements over existing segmentation methods in three-dimensional medical and cellular imaging tasks. The performance gain is observed across all evaluation metrics, including Dice similarity, intersection-over-union, and Hausdorff distance, indicating that the proposed method improves both region accuracy and boundary precision.

Compared with conventional convolutional architectures such as U-Net and 3D U-Net, the proposed framework shows higher segmentation accuracy, particularly in regions with complex geometry. Convolutional models tend to produce smooth but sometimes inaccurate boundaries because they rely primarily on local information. The hybrid encoder in the proposed method captures global context through attention mechanisms, allowing the model to correctly identify structures that extend across large spatial regions. This results in improved segmentation consistency in volumetric images.

Transformer-based models such as TransUNet and MISSFormer demonstrate strong global modeling capability but sometimes lose fine spatial details due to patch-based processing. In contrast, the proposed framework preserves high-resolution features through convolutional pathways while still benefiting from attention-based global reasoning. This combination leads to more accurate segmentation of small structures, which is particularly important in cellular and organoid imaging.

The inclusion of multi-branch feature extraction further improves performance by allowing the network to analyze structures at different scales. Small kernels capture fine details, while larger kernels capture coarse anatomical features. The fusion of these features produces a richer representation of the input data, which contributes to higher segmentation accuracy.

A significant improvement is observed when using the shape-sensitive optimization model. Distance-based loss reduces boundary errors, while curvature-aware and topology-preserving penalties ensure that segmented structures maintain realistic shapes. As a result, the proposed method produces smoother boundaries and fewer disconnected regions compared with models trained using standard loss functions. The reduction in Hausdorff distance confirms that the proposed loss formulation improves boundary accuracy.

In organoid and microscopy datasets, the proposed framework shows particularly strong performance because these images contain irregular and highly variable structures. The shape-sensitive loss helps maintain structural integrity, which is essential for accurate biological analysis. Previous studies have reported similar improvements when using geometry-aware optimization, but the proposed method integrates this strategy within a hybrid architecture, leading to better overall results.

Computational efficiency analysis shows that the hybrid

design introduces only moderate overhead compared with pure convolutional networks while remaining significantly more efficient than fully transformer-based models. This balance makes the proposed framework suitable for practical applications where both accuracy and speed are important.

Overall, the results indicate that combining hybrid feature learning with shape-sensitive optimization provides a robust solution for three-dimensional segmentation, outperforming existing methods in both accuracy and structural consistency.

6. Discussion

The results demonstrate that the integration of convolutional feature extraction, attention-based global modeling, and shape-sensitive optimization provides clear advantages for three-dimensional segmentation tasks. Each component of the proposed framework contributes to performance improvement, and their combination produces a balanced system capable of handling complex volumetric data.

One of the key findings is that hybrid architectures are more effective than purely convolutional or purely transformer-based models. Convolutional networks are efficient and preserve spatial detail but cannot easily capture long-range dependencies. Transformer-based models solve this problem but often require large computational resources and may lose fine details during tokenization. The proposed hybrid encoder successfully combines these approaches, allowing the model to learn both local and global features without excessive computational cost. Similar conclusions have been reported in recent segmentation studies that combine convolutional and attention modules (Huang et al., 2023; Lee et al., 2022).

Another important observation is the impact of the shape-sensitive loss function. Standard loss functions optimize voxel-level accuracy but do not guarantee structural correctness. This limitation becomes critical in three-dimensional segmentation, where small boundary errors can affect the entire volume. The use of distance-based, curvature-aware, and topology-preserving penalties significantly improves segmentation quality by enforcing geometric constraints. This confirms previous findings that topology-aware optimization leads to more reliable segmentation results (Karimi and Salcudean, 2020; Clough et al., 2022).

The multi-branch feature extraction strategy also plays a significant role in performance improvement. Medical and cellular images often contain objects with different scales, and single-scale feature extraction cannot capture all relevant information. By processing multiple receptive fields in parallel, the proposed framework can detect both small cellular structures and large anatomical regions. This capability is particularly important for organoid imaging and volumetric microscopy, where structures vary widely in size (Sohaib et al., 2025).

Despite the improvements achieved by the proposed framework, several limitations should be considered. The hybrid architecture is more complex than standard convolutional networks, which may increase training time. Although the computational cost remains lower than fully transformer-based models, optimization of the architecture for real-time applications may require further research. In addition, the shape-sensitive loss function introduces additional parameters that must be carefully balanced during training.

Another limitation is that the current framework is evaluated on a limited number of volumetric datasets. While the results indicate strong generalization, further validation on larger and more diverse datasets would provide stronger evidence of robustness. Future work could also explore self-supervised pretraining and large-scale foundation models to further improve segmentation performance in scenarios with limited annotation.

Overall, the discussion confirms that the proposed hybrid attention–convolution framework with shape-sensitive optimization addresses several key limitations of existing segmentation methods. By combining global context modeling, local feature extraction, and geometry-aware optimization, the framework provides a reliable and efficient solution for three-dimensional partitioning in medical and cellular imaging.

7. Conclusion

This study presented a hybrid attention–convolution framework combined with shape-sensitive optimization for improved three-dimensional partitioning in medical and cellular imaging. The proposed approach was designed to address several limitations of existing segmentation methods, including insufficient global context modeling in convolutional networks, high computational cost in transformer-based architectures, and lack of structural constraints in conventional loss functions. By integrating convolutional feature extraction, attention-based global modeling, multi-branch feature fusion, and geometry-aware optimization, the framework provides a balanced solution for accurate and reliable volumetric segmentation.

The hybrid encoder plays a central role in the proposed model by combining the strengths of convolutional and attention mechanisms. Convolutional layers effectively capture local spatial details, while attention modules enable the network to model long-range dependencies across the entire volume. This combination allows the framework to accurately segment structures that extend over large spatial regions without losing fine boundary information. The multi-branch design further enhances feature representation by processing different receptive fields in parallel, enabling the model to detect both small

cellular components and large anatomical structures within the same volume.

Another important contribution of this work is the introduction of a shape-sensitive optimization strategy. Traditional loss functions focus on voxel-level accuracy but often fail to preserve geometric consistency, which is essential in three-dimensional segmentation. The proposed loss formulation incorporates distance-based, curvature-aware, and topology-preserving constraints to ensure that predicted structures maintain realistic shapes and correct connectivity. Experimental results demonstrate that this optimization approach significantly improves boundary accuracy and structural stability compared with standard training methods.

The experimental evaluation confirms that the proposed framework outperforms representative convolutional, transformer-based, and hybrid segmentation models across multiple performance metrics. Improvements were observed in Dice similarity, intersection-over-union, and Hausdorff distance, indicating better region accuracy and boundary precision. The framework also showed strong performance in organoid and microscopy datasets, where accurate structural representation is critical for biological analysis. In addition, the computational cost of the hybrid model remains moderate, making it suitable for practical applications that require both accuracy and efficiency.

Despite these advantages, several challenges remain. The hybrid architecture introduces additional complexity compared with standard convolutional networks, and careful parameter tuning is required to balance the contributions of different loss components. Future research may focus on optimizing the framework for faster training, improving generalization through large-scale pretraining, and extending the method to other types of volumetric data. Integration with foundation models and self-supervised learning techniques may further enhance performance in scenarios with limited annotated data.

In conclusion, the proposed hybrid attention–convolution framework with shape-sensitive optimization provides a robust and effective solution for three-dimensional segmentation in medical and cellular imaging. By combining advanced feature learning with geometry-aware optimization, the method advances the state of the art in volumetric image partitioning and offers promising potential for clinical diagnosis, biological research, and automated image analysis systems.

References

1. R. Azad et al., “Medical image segmentation review: The success of U-Net,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 46, no. 12, pp. 10076–10095, Dec. 2024.

2. N.-T. Bui et al., “SAM3D: Segment anything model in volumetric medical images,” in Proc. IEEE Int. Symp. Biomed. Imag., 2024, pp. 1–4.

3. J. Chen et al., “TransuNet: Transformers make strong encoders for medical image segmentation,” 2021, arXiv:2102.04306.

4. J. Chen et al., “TransUNet: Rethinking the U-Net architecture design for medical image segmentation through the lens of transformers,” Med. Image Anal., vol. 97, 2024, Art. no. 103280.

5. Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, “3D U-Net: Learning dense volumetric segmentation from sparse annotation,” in Proc. Med. Image Comput. Comput.-Assist. Interv.: 19th Int. Conf., 2016, pp. 424–432.

6. J. R. Clough, N. Byrne, I. Oksuz, V. A. Zimmer, J. A. Schnabel, and A. P. King, “A topological loss function for deep-learning based image segmentation using persistent homology,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 44, no. 12, pp. 8766–8778, Dec. 2022.

7. A. Dosovitskiy et al., “An image is worth 16 × 16 words: Transformers for image recognition at scale,” 2020, arXiv:2010.11929.

8. M.-P. Dubuisson and A. K. Jain, “A modified Hausdorff distance for object matching,” in Proc. 12th Int. Conf. Pattern Recognit., 1994, vol. 1, pp. 566–568.

9. X. Huang, Z. Deng, D. Li, X. Yuan, and Y. Fu, “MISSFormer: An effective transformer for 2D medical image segmentation,” IEEE Trans. Med. Imag., vol. 42, no. 5, pp. 1484–1494, May 2023.

10. J. Han et al., “Molecular predictors of 3D morphogenesis by breast cancer cell lines in 3D culture,” PLoS Comput. Biol., vol. 6, no. 2, 2010, Art. no. e1000684.

11. F. Isensee, P. F. Jaeger, S. A. Kohl, J. Petersen, and K. H. Maier-Hein, “nnU-Net: A self-configuring method for deep learning-based biomedical image segmentation,” Nature Methods, vol. 18, no. 2, pp. 203–211, 2021.

12. D. Karimi and S. E. Salcudean, “Reducing the Hausdorff distance in medical image segmentation with convolutional neural networks,” IEEE Trans. Med. Imag., vol. 39, no. 2, pp. 499–513, Feb. 2020.

13. A. Kirillov et al., “Segment anything,” in Proc. IEEE/CVF Int. Conf. Comput. Vis., 2023, pp.

4015–4026.

14. H. H. Lee, S. Bao, Y. Huo, and B. A. Landman, "3D UX-Net: A large kernel volumetric ConvNet modernizing hierarchical transformer for medical image segmentation," 2022, arXiv:2209.15076.

15. J. Ma, Y. He, F. Li, L. Han, C. You, and B. Wang, "Segment anything in medical images," Nature Commun., vol. 15, no. 1, 2024, Art. no. 654.

16. J. Ma, F. Li, and B. Wang, "U-Mamba: Enhancing long-range dependency for biomedical image segmentation," 2024, arXiv:2401.04722.

17. J. Ma et al., "How distance transform maps boost segmentation CNNs: An empirical study," in Medical Imaging with Deep Learning. Cambridge, MA, USA : PMLR, 2020, pp. 479–492.

18. O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in Proc. Med. Image Comput. Comput.-Assist. Interv.: 18th Int. Conf., 2015, pp. 234–241.

19. O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in Proc. Med. Image Comput. Comput.-Assist. Interv., 2015, pp. 234–241.

20. S. Shabani, S. Mohammed, and B. Parvin, "A novel 3D decoder with weighted and learnable triple attention for 3D microscopy image segmentation," in Proc. Comput. Vis. Pattern Recognit. Conf., 2025, pp. 4699–4708.

21. S. Shabani, M. Sohaib, S. A. Mohamed, and B. Parvin, "Coupled swin transformers and multi-apertures network (CSTA-NET) improves medical image segmentation," in Proc. IEEE 22nd Int. Symp. Biomed. Imag., 2025, pp. 1–5.

22. S. Shabani, M. Sohaib, S. A. Mohammed, and B. Parvin, "Multi-aperture fusion of transformer-convolutional network (MFTC-Net) for 3D medical image segmentation and visualization," 2024, arXiv:2406.17080.

23. S. Shit et al., "clDice-a novel topology-preserving loss function for tubular structure segmentation," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., 2021, pp. 16560–16569.

24. M. Sohaib, S. Shabani, S. A. Mohammed, G. Winkelmaier, and B. Parvin, "Multi-aperture transformers for 3D (MAT3D) segmentation of clinical and microscopic images," in Proc. Winter Conf. Appl. Comput. Vis., 2025, pp. 4352–4361.

25. M. Sohaib, S. Shabani, S. A. Mohammed, and B. Parvin, "3D-organoid-SwinNet: High-content profiling of 3D organoids," IEEE J. Biomed. Health Inform., vol. 29, no. 2, pp. 792–798, Feb. 2025.

26. V. Srivastava, T. R. Huycke, K. T. Phong, and Z. J. Gartner, "Organoid models for mammary gland dynamics and breast cancer," Curr. Opin. Cell Biol., vol. 66, pp. 51–58, 2020.

27. Y. Tang et al., "Self-supervised pre-training of swin transformers for 3D medical image analysis," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., 2022, pp. 20698–20708.

28. S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "Cbam: Convolutional block attention module," in Proc. Eur. Conf. Comput. Vis., 2018, pp. 3–19.

29. G. Winkelmaier and B. Parvin, "An enhanced loss function simplifies the deep learning model for characterizing the 3D organoid models," Bioinformatics, vol. 37, no. 18, pp. 3084–3085, 2021.

30. E. Xie et al., "SegFormer: Simple and efficient design for semantic segmentation with transformers," in Proc. Adv. Neural Inf. Process. Syst., vol. 34, 2021, pp. 12077–12090.

31. G. Xu, X. Zhang, X. He, and X. Wu, "LeViT-UNet: Make faster encoders with transformer for medical image segmentation," in Proc. Chin. Conf. Pattern Recognit. Comput. Vis., 2023, pp. 42–53.

32. X. Xu, S. Xu, L. Jin, and E. Song, "Characteristic analysis of Otsu threshold and its applications," Pattern Recognit. Lett., vol. 32, no. 7, pp. 956–961, 2011.

33. F. Xing and L. Yang, "Robust nucleus/cell detection and segmentation in digital pathology and microscopy images: A comprehensive review," IEEE Rev. Biomed. Eng., vol. 9, pp. 234–263, 2016.

34. G. Xing et al., "Multi-scale pathological fluid segmentation in OCT with a novel curvature loss in convolutional neural network," IEEE Trans. Med. Imag., vol. 41, no. 6, pp. 1547–1559, Jun. 2022.

35. H.-Y. Zhou, “nnFormer: Volumetric medical image segmentation via a 3D transformer,” IEEE Trans. Image Process., vol. 32, pp. 4036–4045, 2023.