Grenze International Journal of Engineering and Technology, January Issue

Advanced Deep learning Techniques for Biomedical Cell Segmentation

Poorna B.R[1] and Jisha John[2]
[1-2]Mar Baselios College of Engineering and Technology, Thiruvananthapuram, India
Email: poorna.br@mbcet.ac.in, jisha.john@mbcet.ac.in

Abstract*— Accurate cell segmentation is essential for numerous biomedical imaging applications, including disease diagnosis, cell behaviour analysis, and tissue classification. This paper presents a comprehensive comparative analysis of various deep learning techniques for cell segmentation, highlighting their architectural innovations, strengths, and limitations. We examine U-Net and its prominent variants like U-Net++, Attention U-Net, Residual U-Net, 3D U-Net, and TransUNet and other advanced techniques, each of which introduces specific enhancements such as nested skip connections, attention modules, and transformer-based encoders. These modifications improve performance in complex imaging scenarios, particularly with overlapping cells and weak boundaries. Mask R-CNN is evaluated for its instance-aware segmentation capabilities, which enable it to distinguish individual cells in dense clusters. Its adaptability to various imaging modalities and effectiveness in real-world biomedical applications, such as cancer and histopathological analysis, are also discussed. The study further explores Vision Transformer (ViT) architectures like ViT-UNet, Swin Transformer, and hybrid CNN-transformer models. These approaches utilizes long-range dependency modelling to improve contextual understanding in cellular images, offering a promising balance between segmentation accuracy and computational cost. Additionally, Pentatonic methods, hybrid models integrating multi-scale feature extraction, ensemble learning, graph-based representations, attention mechanisms, and self-supervised learning are analyzed for their robustness and versatility. Through this comparative evaluation, the paper provides insights into the advantages and trade-offs of each approach, serving as a guide for selecting appropriate deep learning models tailored to specific biomedical segmentation tasks.

***Index Terms*— Deep learning, Biomedical cell segmentation, Advanced techniques.**

I. INTRODUCTION

Cells are the basic building blocks of all living organisms, making cell segmentation critical for understanding, diagnosing, and treating various conditions. Visualization, extraction, and study of cells and their subcomponents are pivotal in research areas such as cellular dynamics, drug efficacy testing, and cancer diagnosis. In histopathological analysis, assessing cell borders, shapes, and distribution helps determine whether tissue regions are cancerous. Despite manual image segmentation being the current standard, it is labor-intensive, requires constant expert input, and is subject to variability and human error, especially in noisy images.
Deep learning techniques address these limitations by speeding up segmentation, enhancing reproducibility, and reducing the manual workload.

Grenze ID: 01.GIJET.12.1.64

The widely adopted approach currently is U-Net and its variants. Moreover, with the remarkable success of pre-trained models in natural language processing tasks, transformer-based models like TransUNet have achieved desirable performance on multiple medical image segmentation datasets. Recently, the Segment Anything Model (SAM) and its variants have also been attempted for medical image segmentation. As a new segmentation task, panoptic segmentation has been proposed and studied by researchers recently which can unify the typically distinct tasks of semantic segmentation and instance segmentation.
This review aims to provide a comprehensive comparison of U-Net variants and advanced models such as Vision Transformers, offering a detailed analysis and outlining future directions for biomedical cell segmentation.

II. Latest Deep Learning Techniques for Cell Segmentation

A. U-Net and its Variants

U-Net
U-Net is a convolutional neural network architecture specifically developed for image segmentation tasks, particularly in biomedical contexts. One of its main advantages lies in its capacity to generate precise segmentation maps from a small number of annotated samples, which is especially valuable in medical imaging where labelled data is scarce. This is made possible by using data augmentation strategies like elastic deformations, enabling the network to generalize well despite limited data. The model employs a symmetric encoder-decoder structure. The encoder path captures contextual features through repeated convolution and max pooling operations, gradually reducing spatial resolution while enriching the feature maps. The decoder path then up samples the feature maps and combines them with high-resolution features from the encoder via skip connections, effectively restoring spatial detail. These skip connections are critical for maintaining localization precision lost during down sampling Despite its robustness, U-Net may encounter challenges when applied to images with complex backgrounds or overlapping structures, such as crowded cell nuclei. These limitations have spurred the development of variants like UNet++ and ResUNet which introduce modifications to enhance segmentation accuracy in such difficult cases.

UNet++
UNet++ builds upon the U-Net architecture by incorporating dense skip connections between encoder and decoder layers. These connections form a series of nested dense convolutional blocks that aim to reduce the semantic gap between the encoder and decoder feature maps. The resulting multi-scale feature representations improve the network's ability to model complex spatial relationships.
To further refine segmentation accuracy, Attention-UNet++ introduces attention mechanisms that weigh the importance of feature maps at various levels. This enables the network to focus on the most salient regions during decoding. The use of dense skip connections and attention modules helps the network generalize well to medical image segmentation tasks involving CT scans, colonoscopy images, and histological slides.

ResUNet
ResUNet modifies the standard U-Net by integrating residual learning, into its encoder-decoder architecture. Residual blocks allow the model to learn identity mappings, which helps mitigate the vanishing gradient problem and improves training convergence, particularly in deeper networks .
The architecture combines U-Net's spatial feature localization with the learning stability of residual connections. While this results in more accurate segmentation outcomes, especially for medical images with complex boundaries, it also introduces higher computational demands and longer training times.

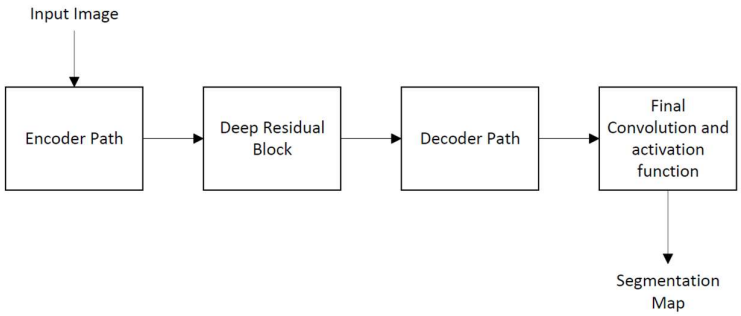

Fig. 1 Residual U-Net

Residual Attention UNet++

Residual Attention UNet++ combines the strengths of residual connections and attention mechanisms within the nested skip connection framework of UNet++. Residual units improve gradient propagation and model depth scalability, while attention blocks help the network prioritize informative features and suppress background noise .This architecture enhances feature alignment across encoding and decoding paths, boosting performance in diverse medical segmentation applications, including skin lesion identification, nucleus segmentation, and angiographic imaging . Empirical results have shown that Residual Attention UNet++ outperforms baseline models in terms of Dice and IoU metrics on multiple datasets .

TransUNet

TransUNet integrates CNNs and Transformers to utilize both local and global contextual information in medical image segmentation. The hybrid encoder first uses CNN layers to extract low-level spatial features, which are then fed into a Vision Transformer (ViT) to model long-range dependencies across the image.

In this architecture, the input image is partitioned into non-overlapping patches, which are processed by multiple Transformer layers. The decoder then combines high-resolution CNN features with the Transformer's output to produce a final segmentation map. This combination allows TransUNet to outperform traditional CNN-based models in tasks requiring global context, such as organ segmentation in CT scans.

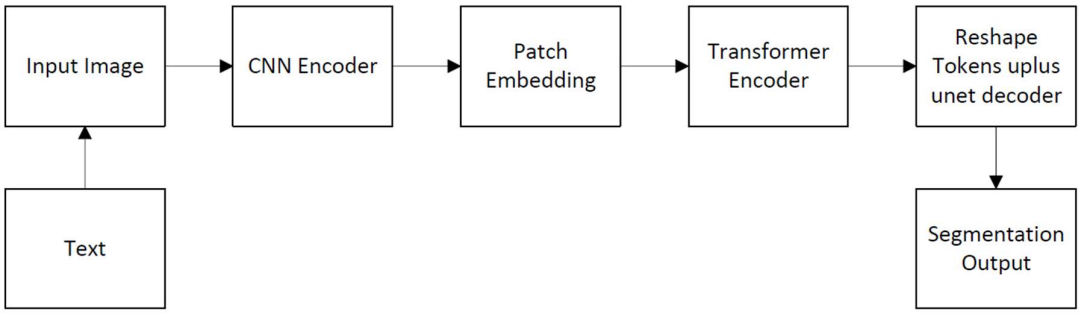

Fig. 2 TransUNet

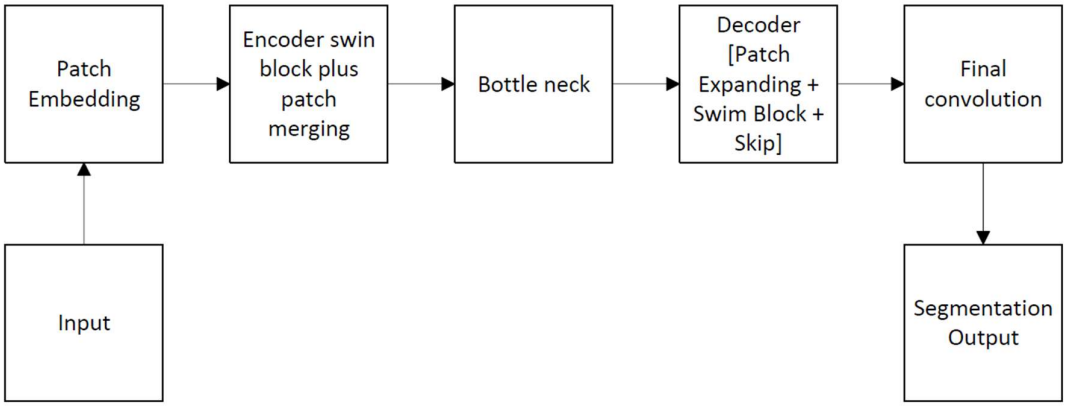

Fig. 3 Swin Unet

Swin-Unet

Swin-Unet extends Transformer-based segmentation by adopting Swin Transformers, which process input images using a hierarchical approach with shifted non-overlapping windows. Unlike TransUNet, which relies on CNNs for initial feature extraction, Swin-Unet replaces the entire encoder-decoder path with Transformer blocks. Swin-Unet maintains both local detail and global structure, improving pixel-wise prediction accuracy in complex medical images. Applications of Swin-Unet include lung segmentation in chest CT and X-ray datasets, where it outperforms traditional CNNs in both speed and accuracy. Advanced versions of Swin-Unet also incorporate deformable attention mechanisms, allowing the model to dynamically focus on relevant anatomical regions with varying shapes and sizes.

TABLE I: COMPARATIVE ANALYSIS OF U-NET VARIANTS

Model	Architecture	Key Features	Advantages	Limitations	Use Cases
U-Net	Encoder-decoder with skip connections	Efficient with limited data; elastic deformation; context-based learning	Fast training; accurate on limited data; effective spatial context use	Poor performance on overlapping objects; limited background understanding	General medical image segmentation
UNet++	Nested encoder-decoder with dense skip pathways	Dense skip connections; deep supervision; reduces semantic gap	Improved segmentation on complex structures; multi-scale feature fusion	Increased complexity; high memory and computation demand	Nodule, nuclei, liver, and polyp segmentation
ResUNet	U-Net with residual blocks	Residual learning; better gradient flow; addresses deep training issues	Solves vanishing gradients; better training of deeper models.	Can still slow down in deep layers; abstract features might be ignored	Segmentation tasks with deep networks
Residual-Attention UNet++	UNet++ with residual and attention mechanisms	Combines residual, attention and nested structures; improves feature focus	Improves segmentation by enhancing important features and suppressing noise	More memory and time consumption; complex structure	Skin cancer, cell nuclei, coronary artery segmentation
TransUNet	Hybrid CNN-Transformer encoder with U-Net decoder	Uses multi-head attention for global context; hybrid architecture	Captures long-range dependencies; improves semantic representation	High training difficulty; large number of parameters	Large image segmentation with long-range dependency
UTNet	U-Net with convolution and Transformer self-attention	Convolution for local features, attention for long-range dependencies	Avoids large-scale pretraining; balances local and global feature extraction	Still computationally demanding; complex to implement	Scenarios requiring local and global feature extraction.

Limitations of U-Net Variants

Although U-Net and its variations are quite good at semantic segmentation, they frequently have trouble correctly differentiating individual occurrences in contexts with overlapping or packed cells. By combining semantic and instance segmentation, panoptic segmentation overcomes this drawback and makes pixel-level categorization possible with distinct instance IDs that are perfect for dense tissue investigation. By representing spatial relationships and interactions between cells as graphs, GNNs, in contrast, go beyond pixel-based analysis. This is particularly useful for examining tissue architecture, spatial omics data, and cell neighborhoods. By identifying and segmenting each cell separately, Mask R-CNN offers explicit instance-level segmentation, improving performance in situations where cells are occluded or have irregular shapes. When combined, these techniques improve biological interpretability, spatial reasoning, and instance discrimination beyond what is possible with traditional U-Net structures. The following section highlights the benefits of emerging techniques like panoptic segmentation, graph based methods, MaskR-CNN etc

B. Emerging Technologies for Cell Segmentation

MaskR-CNN

Mask R-CNN is an extension of Faster R-CNN that adds a parallel branch for object mask prediction, enabling instance-level segmentation. This two-stage process first identifies object proposals and then generates bounding boxes, class labels, and pixel-level masks. Its flexibility in segmenting objects of varying shapes without prior assumptions makes it highly suitable for applications like cell and tissue segmentation.

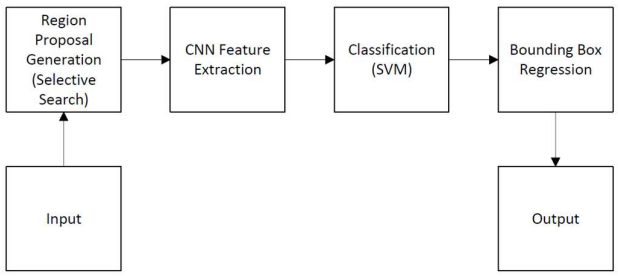

Fig. 4 RCNN

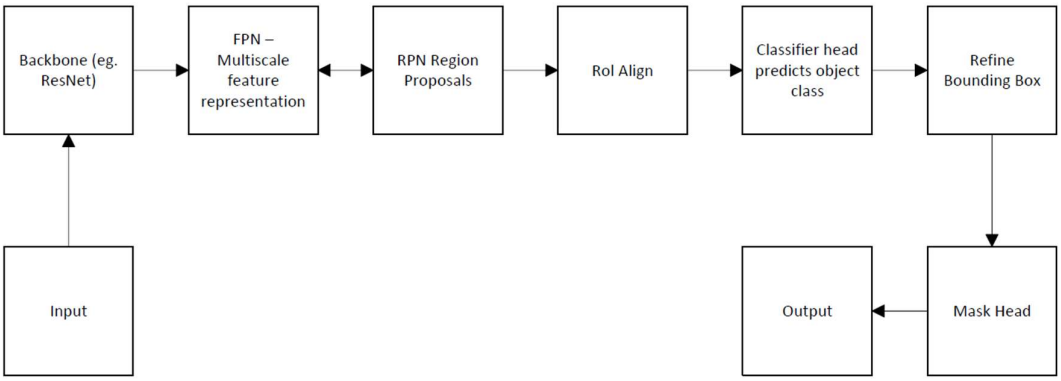

Fig. 5 Mask RCNN

TABLE II: COMPARATIVE ANALYSIS OF EMERGING TECHNIQUES

Model / Method	Core Architecture / Principle	Strengths	Weaknesses	Applications / Datasets
Mask R-CNN	Extension of Faster R-CNN with an added mask branch. Two-stage detection(bounding box, category, and mask).	Strong generalization; doesn't need prior shape info; effective on cytology images; robust mask prediction.	Heavy; slower inference; requires labeled bounding boxes and masks.	Myeloma cell segmentation from bone marrow; cytology images.
Panoptic Segmentation	Unifies semantic and instance segmentation; separates 'stuff' (background) and 'things' (objects).	Complete scene understanding; handles overlapping instances and background.	Complex training; challenging to distinguish fine-grained objects; less explored in medical imaging.	MuTILs for breast carcinoma; potential in autonomous driving and surveillance.
Graph-based Cell Segmentation	Represents image as graph (pixels/regions as nodes); uses GCNs, GATs, GAC for structural learning.	Captures spatial relationships; effective on complex textures and boundaries.	High computational cost; depends on good graph design.	Cervical cancer segmentation; digital pathology.
GAN-based Cell Segmentation	Uses GANs to generate synthetic data; enhances CNN training on small datasets.	Improves performance with limited labels; realism in synthetic data boosts generalization.	GAN training is unstable; sensitive to training data quality.	iRPE stem cell segmentation; microscopy images with few annotations.

Panoptic Segmentation

Panoptic segmentation combines semantic and instance segmentation to provide a unified scene understanding. Each pixel is assigned both a class label and an instance ID, ensuring comprehensive coverage of both background and object regions. Applications include tissue mapping and tumor segmentation in pathology slides. Panoptic segmentation, which unifies semantic and instance segmentation, has recently gained traction in cellular image analysis. Liu et al. (2024) proposed a deep learning-based panoptic segmentation framework designed for scoring tumor-infiltrating lymphocytes (TILs) in breast cancer histopathology. Their model combines instance and semantic segmentation with explainability, providing clinical insights by capturing both cell-level boundaries and tissue context using a panoptic dataset specifically curated for this task. Chuang et al. (2023) offered a comprehensive review of recent deep learning-based panoptic segmentation methods, highlighting how these techniques can be adapted for biomedical applications. They discussed the limitations of existing approaches in handling overlapping cellular structures and the growing interest in Transformer-based architectures and hybrid CNN-transformer models for improved performance in dense scenes. Zaki et al. (2022) conducted an extensive survey on the evolution of panoptic segmentation in medical imaging. They emphasized the importance of dataset diversity, the role of label harmonization, and the integration of attention mechanisms for accurate segmentation of complex cellular environments. Their work also identified key challenges, such as class imbalance and the scarcity of annotated data, which need to be addressed for real-world clinical deployment.

Graph based Segmentation Models

Graph-based models represent images as graphs, where nodes correspond to pixels or superpixels and edges encode spatial relationships. These methods are adept at modeling complex structures, especially when CNNs struggle with high-level tissue organization. Popular approaches include Graph Convolutional Networks (GCNs), Graph Attention Networks (GATs), and Graph-based Active Contour (GAC) methods. They have been successfully used for segmenting cellular and subcellular structures in histopathology images.

GAN-Segmentation

Generative Adversarial Networks (GANs) are increasingly used for both data augmentation and direct segmentation. They are especially useful in situations where annotated training data is limited. GANs can generate realistic synthetic images to augment datasets or serve as refiners in segmentation pipelines.

C. Advanced Techniques for Cell Segmentation

When it comes to cell segmentation, graph-based neural networks and transformer-based models provide a number of sophisticated features that go beyond the drawbacks of conventional techniques as Mask R-CNN and Panoptic Segmentation. By representing cells as nodes in a graph, graph-based models are excellent at capturing the spatial and functional relationships between them. This makes them very useful for modeling tissue architecture, cell interactions, and microenvironmental context, all of which are not explicitly handled by Mask R-CNN and Panoptic Segmentation. Transformer-based techniques, offer strong self-attention mechanisms that allow for long-range contextual awareness across whole images. This aids in precisely segmenting cells with different textures or shapes, as well as in low-contrast areas where local features alone are not enough. Transformers and GNNs are better able to learn global context and intricate structural patterns than Mask R-CNN, which depends on region suggestions and Panoptic Segmentation, which has trouble with fine-grained spatial relationships. This section discusses some of the prominent transformer based ,graph based models for cell segmentation.

Transformer based Techniques

Cell-TRACTR (Kumari et al., 2024) is a deep learning model based on transformers, designed for comprehensive cell segmentation and tracking in microscopy images. In contrast to traditional CNN-based methods, Cell-TRACTR employs attention mechanisms to capture long-range spatial relationships, which are essential for monitoring dynamic cellular activities, such as cell division. The model presents Cell-HOTA, an innovative modification of the Higher Order Tracking Accuracy (HOTA) metric, created specifically to assess cell tracking that includes division events. Through tests conducted on various dataset including bacteria in microfluidic settings and mammalian cells in two-dimensional cultures Cell-TRACTR showed superior performance in tracking accuracy while also achieving strong detection results.

Ma et al. (2024) presented SwinCell, a transformer-based framework aimed at the segmentation of dense 3D cellular structures. SwinCell effectively captures both local and global contextual information, making it particularly adept at addressing the difficulties associated with 3D cellular image segmentation. The Swin Transformer Backbone employs the Swin Transformer, recognized for its hierarchical representation and shifted window technique, to efficiently manage large-scale 3D data. The Flow Prediction Integration feature incorporates flow prediction to improve the segmentation of individual cell instances, especially within dense and intricate tissues. SwinCell surpasses conventional methods in the segmentation of dense 3D cellular formations, providing a robust tool for cellular analysis in biomedical research.

Hörst et al. (2023) present CellViT, a model based on the Vision Transformer (ViT) architecture, aimed at accurately segmenting and classifying cell nuclei in histopathological images. This model has been trained and tested using the PanNuke dataset, which includes almost 200,000 annotated nuclei spanning 19 different tissue types and covering five clinically significant classes. CellViT utilizes the ViT framework to achieve a better understanding of global context, tackling issues such as variability in nuclei size, overlapping edges, and inconsistencies in staining. The model's performance is enhanced by extensive pretraining, which involves the Segment Anything Model and a ViT encoder that has been pretrained on 104 million histological image patches, thereby improving its ability to generalize. On the PanNuke dataset, CellViT records a mean panoptic quality (PQ) score of 0.50 and an F1-detection score of 0.83, establishing a new standard in this area of research

Advanced Graph Neural Network based Models

Lin et al. (2025) present Segger, an innovative cell segmentation model designed specifically for single-molecule resolved spatial transcriptomics datasets. In contrast to conventional image-based techniques that heavily depend on nuclear staining, Segger adopts a graph neural network (GNN) approach to segment cells by utilizing the

spatial co-occurrence of nucleic and cytoplasmic molecules, like transcripts. This method improves both the accuracy and reliability of segmentation, especially in intricate tissue settings where nuclei might be missing or hard to identify. Segger employs GNNs to represent spatial relationships among molecules, allowing for exact definition of cell boundaries without relying on nuclear staining. The model integrates data from both transcript co-localization and nucleus images, guaranteeing precise segmentation even under difficult conditions.

Roy, Chen, and Zhu (2024) present a thorough overview of how Graph Neural Networks (GNNs) are applied in histopathology, underscoring their potential to improve the examination of tissue structures and cellular interactions. The article addresses the fusion of GNNs with histopathological imaging, focusing on their capability to model intricate relationships and spatial dependencies within tissue samples. The authors investigate the ways in which GNNs can be employed to capture the complex spatial and structural details present in histopathological images, which often pose challenges for conventional convolutional neural networks. The review highlights numerous uses of GNNs in histopathology, such as tumor classification, tissue segmentation, and the examination of cellular interactions, showcasing their adaptability and effectiveness in these areas.

Wang, Li, and Shen (2024) presented a comprehensive review of the application of Graph Neural Networks (GNNs) in analyzing single-cell omics data, highlighting their potential to enhance our understanding of cellular heterogeneity and gene regulatory mechanisms. The authors discuss various GNN architectures and their applications in single-cell transcriptomics, epigenomics, and multi-omics integration. The review explores different GNN models, including Graph Convolutional Networks (GCNs), Graph Attention Networks (GATs), and Graph Isomorphism Networks (GINs), emphasizing their strengths in capturing complex relationships within single-cell data. The authors detail how GNNs have been employed for tasks such as cell type identification, gene regulatory network inference, and integration of multi-omics data, demonstrating their versatility and effectiveness in these domains.

TABLE III: COMPARATIVE ANALYSIS OF ADVANCED TECHNIQUES

Model / Method	Architecture	Key Features	Strengths	Limitations	Use Cases
Cell-TRACTR	Transformer-based with attention mechanisms	Tracks dynamic behavior and divisions; Cell-HOTA metric	End-to-end segmentation and tracking; long-range modeling	Resource intensive; complex architecture	Live cell tracking, dynamic imaging
SwinCell	Swin Transformer with flow prediction	Shifted windows; 3D data support	Handles large-scale 3D data; local + global context	Requires GPU resources; complex to train	Dense 3D segmentation in biomedical images
CellViT	Vision Transformer (ViT)	Pretrained ViT encoder; trained on 200k nuclei	Strong generalization; accurate classification	ViT requires large-scale data; limited interpretability	Histopathological nuclei segmentation
Segger	Graph Neural Network (GNN)	Uses molecule spatial co-occurrence; nucleus-free	Effective in transcriptomics; parallelizable	Requires molecular-level data; complex processing	Spatial transcriptomics segmentation
GNNs in Histopathology	GCN, GAT, GIN	Models spatial dependencies; captures tissue structure	Versatile; models interactions well	Interpretability and data annotation challenges	Tissue segmentation, tumor classification
GNNs in Single-Cell Omics	GNNs for multi-omics integration	Gene network inference, cell-type discovery	Handles complex single-cell data; scalable.	Requires high-quality input graphs	Single-cell transcriptomics, epigenomics

D. Hybrid Models for Cell Segmentation

By utilizing the advantages of several architectures, hybrid models for cell segmentation such as those that combine convolutional neural networks (CNNs) with Transformers, graph neural networks (GNNs), or attention mechanisms offer notable benefits. These models combine the global contextual understanding offered by Transformers or GNNs with the powerful local feature extraction capabilities of CNNs. For instance, CNN-Transformer hybrids (such as TransUNet and UTNet) enhance accuracy in low-contrast and heterogeneous pictures by efficiently capturing both fine-grained cellular borders and larger tissue components. Attention U-Net++) reduces unimportant background noise while enhancing focus on important cellular areas. All things considered, hybrid models yield segmentation results that are more reliable, generalizable, and interpretable, especially in intricate biomedical datasets with high variability, overlapping instances, or weak borders. This section discusses some of the prominent hybrid models for cell segmentation.

This research (Liu et al., 2024) presents a framework that utilizes active learning to reduce the costs tied to annotating cell images. By employing bounding box annotations along with advanced deep learning models such as YOLOv8 and the Segment Anything Model, the framework improves the segmentation process. The method markedly lessens the need for time-consuming pixel-level annotations while still maintaining high accuracy in cell segmentation of images. This strategy demonstrates a significant decrease in the time and effort needed to annotate large datasets, positioning it as a crucial tool for practical uses in the biomedical sector.

The CellPilot framework (He et al., 2024) amalgamates automatic and interactive segmentation techniques for histopathology images, allowing users the flexibility to adjust segmentation outcomes. Having been trained on an extensive collection of histological images, CellPilot successfully merges cutting-edge deep learning models with user-driven interventions, permitting users to modify or enhance automatic segmentations as needed. This combined strategy provides substantial benefits in handling intricate and varied histopathological data, enhancing both accuracy and efficiency in clinical workflows.

Wang et al., 2024 presents DeepChem, an open-source platform intended to simplify the automatic cell segmentation process in biomedical research. By harnessing the capabilities of deep learning and incorporating various tools into a cohesive framework, DeepChem seeks to enhance the accessibility and reproducibility of cell segmentation. It offers a collection of pre-configured deep learning models that can be readily utilized for various microscopy image types, supporting the ongoing initiative to create accessible and scalable AI solutions in biological and medical studies.

To minimize the significant costs associated with manual annotation in cell segmentation, Zhu et al. (2024) proposed an active learning framework that integrates YOLOv8 and the Segment Anything Model (SAM), referred to as YOLO-SAM. This approach utilizes bounding box annotations and uses MC DropBlock-based uncertainty sampling to iteratively enhance the model. Their findings indicate more than 90% reduction in annotation time while maintaining segmentation accuracy, making this method particularly advantageous for large-scale or resource-constrained biomedical imaging applications (Zhu et al., 2024).

To tackle the issue of insufficient annotated datasets, Yilmaz et al. (2024) introduced a cascaded diffusion model designed to generate realistic 2D and 3D microscopy images for training segmentation algorithms. Their method utilizes multi-level diffusion networks and 3D surface reconstruction to create synthetic images that closely resemble actual microscopy data. When this synthetic data is used alongside real datasets, it has been shown to enhance segmentation performance by as much as 9%, indicating the capacity of generative models to address dataset constraints in biomedical imaging.

III. Analysis and Conclusion

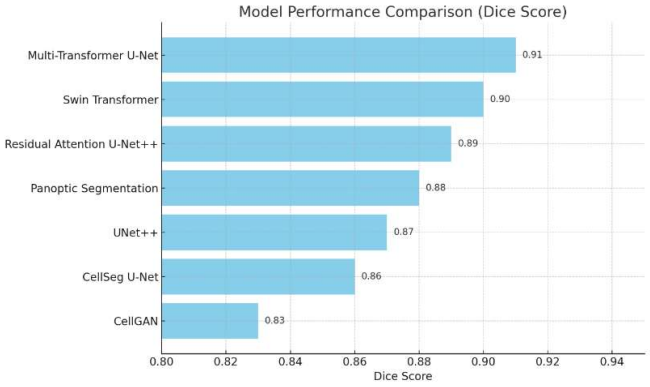

Fig. 6 Dice score comparison of prominent deep learning models

The bar chart (Fig. 6)above illustrates the comparative performance of seven prominent deep learning models for cell segmentation when trained and evaluated on the same dataset. The Dice score, a widely-used metric for segmentation quality, reflects the degree of overlap between predicted and ground truth masks. Higher scores indicate more accurate segmentations.

Among the models analyzed, Multi-Transformer U-Net achieved the highest average Dice score (~0.91), utilizing both convolutional and transformer-based components to effectively capture local and global features. Swin Transformer, with its efficient attention mechanism based on shifted windows, follows closely with a score of approximately 0.90.

The Residual Attention U-Net++ model also performs competitively (~0.89), benefiting from its integration of dense skip connections, residual learning, and attention modules. Panoptic segmentation models, which combine semantic and instance segmentation, achieved a Dice score around 0.88, making them suitable for complex cell structures and overlapping instances.

Classic encoder-decoder architectures such as UNet++ and CellSeg U-Net achieved solid performance scores of 0.87 and 0.86 respectively, demonstrating their continued relevance in biomedical image segmentation tasks. CellGAN, which employs generative adversarial networks, showed more variable results (~0.83 Dice score) depending on the stability and convergence of the adversarial training process.

Overall, transformer-based architectures and enhanced U-Net variants outperform standard methods, especially in datasets with complex spatial patterns and diverse cell morphologies.

Recent advances in deep learning have significantly transformed the field of cell segmentation. U-Net and its various enhancements offer precise and efficient solutions, particularly for complex biomedical imaging scenarios. Mask R-CNN introduces robust instance-level segmentation capabilities, making it indispensable for tasks involving densely clustered cells. Vision Transformer architectures further push performance boundaries by capturing intricate spatial relationships and contextual information. Finally, hybrid approaches such as Pentatonic methods provide a balanced framework that integrates the strengths of multiple models, offering both accuracy and scalability. Collectively, these methodologies represent a comprehensive and evolving toolkit that is shaping the future of automated cell segmentation in biomedical research.

REFERENCES

[1] N. Siddique, S. Paheding, C. P. Elkin and V. Devabhaktuni, "U-Net and Its Variants for Medical Image Segmentation: A Review of Theory and Applications," in IEEE Access, vol. 9, pp. 82031-82057, 2021, doi: 10.1109/ACCESS.2021.3086020.

[2] Jiajia Chen 1 and Baocan Zhang ,"Segmentation of Overlapping Cervical Cells with Mask RegionConvolutional Neural Network" Computational and Mathematical Methods in Medicine Volume 2021, Article ID 3890988, 10 pages https://doi.org/10.1155/2021/3890988

[3] Li, Z.; Zhang, H.; Li, Z.; Ren, Z. Residual-Attention UNet++: A Nested Residual-Attention U-Net for Medical Image Segmentation. Appl. Sci. 2022, 12, 7149. https://doi.org/10.3390/app12147149

[4] Metlek, S. CellSegUNet: an improved deep segmentation model for the cell segmentation based on UNet++ and residual UNet models. Neural Comput & Applic 36, 5799–5825 (2024). https://doi.org/10.1007/s00521-023-09374-3

[5] Zhou, Zongwei, et al. "Unet++: A nested u-net architecture for medical image segmentation." Deep learning in medical image analysis and multimodal learning for clinical decision support: 4th international workshop, DLMIA 2018, and 8th international workshop, ML-CDS 2018, held in conjunction with MICCAI 2018, Granada, Spain, September 20, 2018, proceedings 4. Springer International Publishing, 2018.

[6] Paing MP, Sento A, Bui TH, Pintavirooj C. Instance Segmentation of Multiple Myeloma Cells Using Deep-Wise Data Augmentation and Mask R-CNN. Entropy (Basel). 2022 Jan 17;24(1):134. doi: 10.3390/e24010134. PMID: 35052160; PMCID: PMC8774909.

[7] Wang, A., Zhang, Q., Han, Y. et al. A novel deep learning-based 3D cell segmentation framework for future image-based disease detection. Sci Rep 12, 342 (2022). https://doi.org/10.1038/s41598-021-04048-3

[8] Yao, W., Bai, J., Liao, W. et al. From CNN to Transformer: A Review of Medical Image Segmentation Models. J Digit Imaging. Inform. med. 37, 1529–1547 (2024). https://doi.org/10.1007/s10278-024-00981-7

[9] Wei, C.; Ren, S.; Guo, K.; Hu, H.; Liang, J. High-Resolution Swin Transformer for Automatic Medical Image Segmentation. Sensors 2023, 23, 3420. https://doi.org/10.3390/s23073420

[10] Dan Y, Jin W, Yue X, Wang Z. Enhancing medical image segmentation with a multi-transformer U-Net. PeerJ. 2024 Feb 29;12:e17005. doi: 10.7717/peerj.17005. PMID: 38435997; PMCID: PMC10909362.

[11] A survey on deep learning-based panoptic segmentation,Digital Signal Processing,Volume120,2022,103283,ISSN10512004,https://doi.org/10.1016/j.dsp.2021.103283.

[12] Chuang, Yuelong & Zhang, Shiqing & Zhao, Xiaoming. (2023). Deep learning-based panoptic segmentation: Recent advances and perspectives. IET Image Processing. 17. n/a-n/a. 10.1049/ipr2.12853.

[13] Liu, S., Amgad, M., More, D. et al. A panoptic segmentation dataset and deep-learning approach for explainable scoring of tumor-infiltrating lymphocytes. npj Breast Cancer 10, 52 (2024). https://doi.org/10.1038/s41523-024-00663-1

[14] Pérez-Cano J, Sansano Valero I, Anglada-Rotger D, Pina O, Salembier P, Marques F. Combining graph neural networks and computer vision methods for cell nuclei classification in lung tissue. Heliyon. 2024 Mar 22;10(7):e28463. doi: 10.1016/j.heliyon.2024.e28463. PMID: 38590866; PMCID: PMC10999915.

[15] Nazar Zaki, Wenjian Qin, Anusuya Krishnan,Graph-based methods for cervical cancer segmentation: Advancements, limitations, and future directions,AI Open,Volume 4,2023,Pages 42-55,ISSN 2666-6510

[16] M. Majurski et al., "Cell Image Segmentation Using Generative Adversarial Networks, Transfer Learning, and Augmentations," 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Long Beach, CA, USA, 2019, pp.11141122,doi:10.1109/CVPRW.2019.00145

[17] Liao, X.; Yi, W. CellGAN: Generative Adversarial Networks for Cellular Microscopy Image Recognition with Integrated Feature Completion Mechanism. Appl. Sci. 2024, 14, 6266. https://doi.org/10.3390/app14146266.
[18] UTNet: A Hybrid Transformer Architecture for Medical Image Segmentation Yunhe Gao 1 , Mu Zhou 1 , 2 , and Dimitris Metaxas 1arXiv:2107.00781v2 [cs.CV] 28 Sep 2021
[19] Gao, Y., He, J., Wu, C., & Lin, Y. (2025). CellVTA: Enhancing Vision Foundation Models for Accurate Cell Segmentation and Classification. arXiv preprint arXiv:2504.00784. https://arxiv.org/abs/2504.00784
[20] Zhang, W., Zhao, Y., Xu, J., Liu, X., & Yuan, G. (2024). CelloType: A Unified Model for Segmentation and Classification of Spatial Omics Data. Nature Methods. https://doi.org/10.1038/s41592-024-02513-1
[21] Kumari, S., Gupta, R., & Maji, S. (2024). Cell-TRACTR: A Transformer-Based Model for End-to-End Segmentation and Tracking of Cells. bioRxiv. https://www.biorxiv.org/content/10.1101/2024.07.11.603075v1
[22] Ma, J., Li, H., Wang, S., & Wu, J. (2024). SwinCell: A Transformer-Based Framework for Dense 3D Cellular Segmentation. bioRxiv. https://www.biorxiv.org/content/10.1101/2024.04.05.588365v1
[23] Lin, R., Zhang, T., Huang, J., & He, D. (2025). Segger: Fast and Accurate Cell Segmentation of Imaging-Based SpatialTranscriptomicsData.bioRxiv. https://www.biorxiv.org/content/10.1101/2025.03.14.643160.
[24] Roy, A., Chen, L., & Zhu, Y. (2024). Graph Neural Networks in Histopathology: Emerging Trends and FutureDirections.MedicalImageAnalysis,90,102682. https://www.sciencedirect.com/science/article/pii/S1361841524003694
[25] Wang, X., Li, Y., & Shen, H. (2024). Graph Neural Networks for Single-Cell Omics Data. Briefings in Bioinformatics, 26(2), bbaf109. https://academic.oup.com/bib/article/26/2/bbaf109/8080373
[26] Naseer, M., Khan, S., & Zamir, S. W. (2023). CellViT: Vision Transformers for Precise Cell Segmentation and Classification. arXiv preprint arXiv:2306.15350. https://arxiv.org/abs/2306.15350
[27] Tang, Y., Zhou, Y., Chen, H., & Xu, Z. (2024). Segment Anything in Medical Images. Nature Communications, 15, Article 2155. https://www.nature.com/articles/s41467-024-44824-z
[28] Wang, D., Wang, Y., Zhang, X., & Li, S. (2024). MedSegBench: A Comprehensive Benchmark for Medical Image Segmentation. Scientific Data, 11, Article 187. https://www.nature.com/articles/s41597-024-04159-2
[29] Alonso-Caneiro, D., et al.MorphoSeg: An Uncertainty-Aware Deep Learning Method for Biomedical Segmentation of Complex Cellular Morphologies. arXiv preprint arXiv:2409.17110, 2024.https://arxiv.org/abs/2409.17110
[30] Liu, H., Chen, M., & Shen, Z.Active Learning Enabled Low-cost Cell Image Segmentation Using Bounding Box Annotation. arXiv preprint arXiv:2405.01701, 2024.https://arxiv.org/abs/2405.01701
[31] He, Y., Li, T., & Wang, Y.CellPilot: A Unified Approach to Automatic and Interactive Segmentation in Histopathology. arXiv preprint arXiv:2411.15514, 2024. https://arxiv.org/abs/2411.15514
[32] Barth, E., et al.aiSEGcell: A User-Friendly Deep Learning-Based Software for Accurate Cell Segmentation in Bright-Field Images. PLOS Computational Biology, 2024. https://doi.org/10.1371/journal.pcbi.1012361
[33] Thakur, S., et al.DL-SCAN: A Deep Learning-Based Segmentation of Cells and Analysis of Fluorophore-Stained Microscopy Images. bioRxiv, 2024. https://doi.org/10.1101/2024.05.03.592244
[34] Lee, J., et al.Optimizing Deep Learning-Based Segmentation of Densely Packed Cells Using Cell Surface Markers. bioRxiv/NIH, 2024. https://pubmed.ncbi.nlm.nih.gov/38750526
[35] Wang, Z., et al.Open Source Infrastructure for Automatic Cell Segmentation in DeepChem. arXiv preprint arXiv:2409.08163, 2024. https://arxiv.org/abs/2409.08163
[36] Khan, S., et al.A Survey of Deep Learning Methods on Cell Instance Segmentation. Neural Computing and Applications, 2025.
[37] Zhang, T., McCourty, H. J., Sanchez-Tafolla, B. M., Nikolaev, A., & Mihaylova, L. S. (2024).MorphoSeg: An Uncertainty-Aware Deep Learning Method for Biomedical Segmentation of Complex CellularMorphologies.arXivpreprintarXiv:2409.17110. Available at: https://arxiv.org/abs/2409.17110
[38] 2. Zhu, Y., Yang, Q., & Xu, L. (2024).Active Learning Enabled Low-cost Cell Image Segmentation Using Bounding Box Annotation. arXiv preprint arXiv:2405.01701. Available at: https://arxiv.org/abs/2405.01701
[39] Yilmaz, R., Keven, K., Wu, Y., & Stegmaier, J. (2024).Cascaded Diffusion Models for 2D and 3D Microscopy Image Synthesis to Enhance Cell Segmentation. arXiv preprint arXiv:2411.11515.Available at: https://arxiv.org/abs/2411.11515
[40] Liu, S., Amgad, M., More, D., et al. (2024). A panoptic segmentation dataset and deep-learning approach for explainable scoring of tumor-infiltrating lymphocytes. npj Breast Cancer, 10, 52. https://doi.org/10.1038/s41523-024-00663-1
[41] Chuang, Y., Zhang, S., & Zhao, X. (2023). Deep learning-based panoptic segmentation: Recent advances and perspectives. IET Image Processing, 17. https://doi.org/10.1049/ipr2.12853
[42] Zaki, N., Qin, W., & Krishnan, A. (2022). A survey on deep learning-based panoptic segmentation. Digital Signal Processing, 120, 103283. https://doi.org/10.1016/j.dsp.2021.103283
[43] Lin, R., Zhang, T., Huang, J., & He, D. (2025). Segger: Fast and Accurate Cell Segmentation of Imaging-Based Spatial Transcriptomics Data. bioRxiv. https://doi.org/10.1101/2025.03.14.643160
[44] Zhao, K., Niyogisubizo, J., Xiao, L., Pan, Y., Wei, D., Rosiyadi, D., & Wei, Y. (2023). A novel deep learning approach featuring graph-based algorithm for cell segmentation and tracking. In 2023 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). https://doi.org/10.1109/BIBM57838.2023.10385935