DeepGI: An Automated Approach for Gastrointestinal Tract Segmentation in MRI Scans

Ye Zhang[1,*], Yulu Gong[2], Dongji Cui[3], Xinrui Li[4], and Xinyu Shen[5]
[1]University of Pittsburgh, Pittsburgh, USA
[2]Northern Arizona University, Flagstaff, USA
[3]Trine University, Phoenix, USA
[4]Cornell University, New York, USA
[5]Columbia University, Frisco, USA
[*]Corresponding author

ABSTRACT: Detecting gastrointestinal (GI) tract cancers accurately remains essential for improved radiotherapy outcomes. This study introduces an innovative deep learning model for automated segmentation of GI regions within MRI scans, featuring an architecture that combines Inception-V4 for classification, a UNet++ with VGG19 encoder for 2.5D segmentation, and an Edge UNet optimized for grayscale images. Detailed data preprocessing, including 2.5D data handling, is employed to enhance segmentation precision. Our model addresses the limitations of manual segmentation by providing a streamlined, high-accuracy solution that captures complex GI structures crucial for treatment planning.

Keywords: Deep Learning, MRI Segmentation, Inception-V4, UNet++, Edge UNet, Gastrointestinal Imaging

1. Introduction

Gastrointestinal (GI) cancers, encompassing malignancies within the stomach, colon, rectum, and small intestine, represent a significant challenge in global healthcare due to their high incidence and complex treatment require- ments. Radiotherapy is one of the primary treatment options for GI cancers, as it allows for precise delivery of radiation to targeted tumor areas while preserving surrounding healthy tissues. Recent advancements, such as Magnetic Resonance Linear Accelerator (MR-Linac) systems, enable real-time imaging during radiotherapy, providing clinicians with the flexibility to adapt to shifts in tumor or organ position. Despite these innovations, effective radiotherapy planning still depends heavily on accurate segmentation of GI organs within magnetic resonance imaging (MRI) scans, a process that remains predominantly manual.

Manual delineation of GI organs is labor-intensive and requires significant expertise. This segmentation pro- cess is also prone to inter-operator variability, where different clinicians may produce inconsistent segmentations for the same patient, impacting treatment precision. Therefore, a streamlined, automated segmentation ap- proach is essential to optimize radiotherapy planning and ensure uniformity across clinical practices. However, the unique anatomy of the GI tract, which includes organs that are highly variable in size, shape, and structure, presents a distinct challenge for automated segmentation models.

Machine learning, particularly deep learning, has gained prominence in medical imaging due to its ability to handle complex data patterns. Convolutional neural networks (CNNs) and other deep learning architectures have achieved substantial success in tasks like image classification and segmentation. Although general models like U-Net and its variants have shown effectiveness in medical segmentation, they often struggle to handle the anatomical diversity and fine structural details necessary for accurate GI tract segmentation. Moreover, the diversity of MRI scan types, including variations in resolution and slice thickness, adds to the complexity of designing a model that is both adaptable and robust.

In response to these challenges, this study introduces a comprehensive, automated approach for GI tract seg- mentation in MRI scans. Our model integrates a hybrid of advanced neural network architectures, each selected for its unique strengths in addressing specific aspects of the segmentation task. The primary contributions of this work are as follows Advanced Architecture Integration: We develop a hybrid model combining three state-of-the-art architectures: Inception-V4 for initial organ classification, UNet++ with VGG19 encoder for 2.5D data processing, and Edge UNet for precise grayscale segmentation. This ensemble approach enables the model to capture diverse structural details of the GI organs, providing a more comprehensive solution than existing single-architecture models.

Enhanced Data Preprocessing: Recognizing the variability in MRI data, we implement a robust pre- processing pipeline, including spatial and intensity augmentation techniques tailored for GI imaging. Addi- tionally, we incorporate a 2.5D processing method that utilizes depth information from consecutive slices, offering a richer representation of anatomical context compared to traditional 2D methods.

Reduction in Inter-Operator Variability: By automating the segmentation process, our model aims to minimize variability across different practitioners, thus ensuring consistent delineation of GI organs and enhancing treatment planning reliability.

Optimized Workflow for Radiotherapy Planning: This work provides a solution that has potential to significantly reduce the time and effort required by clinicians in radiotherapy planning, allowing for faster and more accurate treatment workflows. The integration of deep learning architectures tailored for specific tasks within the segmentation pipeline marks a significant step forward in radiotherapy applications.

This paper is structured as follows: Section ?? reviews recent developments in medical image segmentation with a focus on GI tract imaging. Section ?? outlines the proposed methodology, including the specific neural network architectures and data preprocessing techniques employed. Section ?? presents experimental results, with an evaluation of the model's performance across different MRI data types, followed by a discussion in Section. This work ultimately contributes to the field by providing a practical and effective tool that can support clinicians in achieving precise, automated GI tract segmentation, offering potential improvements in both radiotherapy planning and patient outcomes.

2. Related Work

The field of medical image segmentation, especially within gastrointestinal (GI) cancer diagnosis and treatment, has evolved with substantial contributions from deep learning approaches, particularly in automating segmenta- tion tasks. This review covers the foundational methodologies and breakthroughs, which form the basis for our model.

Kocak et al.1 explored the role of deep learning in medical imaging segmentation, with a focus on enhanc- ing interpretability to support clinical decision-making. This work provided foundational insights into the use of neural networks in medical imaging. Zhou et al.2 demonstrated the effectiveness of U-Net for capturing spatial context, highlighting its utility in medical imaging tasks that require high detail. Building on model efficiency, Lu et al.3 introduced a processor for analyzing power consumption in data cycles, showcasing the importance of preprocessing in reliable modeling outcomes. In segmentation tasks, Edge U-Net was enhanced with Holistically-Nested Edge Detection (HED),4 focusing on edge detection for better boundary preservation in complex structures.

Developments in intelligent systems also inform medical imaging. Tianbo et al.5 designed a swarm intelligence system integrating adaptive control and object recognition, illustrating how autonomous systems can enhance robustness. Zhang et al.6 applied a transformer module with evidential learning for pedestrian intent prediction, introducing interpretability enhancements that are advantageous for medical imaging. In addressing imbalanced datasets, Chen et al.7 proposed a two-stage classification strategy, improving feature alignment in medical segmentation, which inspired our multi-path approach.

Maccioni et al.8 discussed challenges in GI tract segmentation due to anatomical variation, emphasizing the need for adaptable segmentation models in radiotherapy planning. The U-Net architecture by Ronneberger et al.,9 designed for biomedical segmentation, introduced skip connections that retain spatial information critical in medical imaging. Expanding on this, the VGG architecture by Simonyan and Zisserman10 contributed a deep convolutional structure that serves as a reliable encoder for image recognition tasks.

In terms of reducing computational complexity, Szegedy et al.11 enhanced the Inception model series, making it suitable for complex segmentation tasks that demand efficient processing. SegNet12 introduced an encoder-decoder architecture that facilitated image segmentation through efficient down-sampling and up-sampling opera- tions, influencing segmentation tasks requiring high-resolution retention. Expanding beyond biomedical imaging, Zhang et al.13 presented a machine vision-based manipulator control system for robotics, highlighting precision handling techniques applicable to segmentation.

In breast cancer detection, Zhang et al.14 advanced binary classification by using novel pooling techniques, enhancing model accuracy. For face recognition, Liao et al.15 proposed the Attention Selective Network (ASN) to manage pose variations, a concept that can focus on clinically significant regions in medical images. Lin16 introduced a framework for Bruch's membrane segmentation in OCT images, supporting biomarker tracking in retinal diseases, which is relevant for improving segmentation in diagnostic imaging.

T Xiao et al.17 developed dGLCN, a dual-graph network, for Alzheimer's diagnosis, demonstrating how graph structures can enhance segmentation interpretability. In the field of point cloud classification, Hu et al.18 introduced M-GCN, leveraging multi-scale graph convolutional techniques for better feature fusion, which aligns with our focus on multi-dimensional data processing. Zeng et al.19 presented a two-phase Alzheimer's diagnostic framework, emphasizing feature alignment, a principle adapted in our preprocessing techniques.

In OCT technology, Chen et al.20 developed long-range SS-OCT for anterior eye imaging, facilitating detailed imaging useful in segmentation applications. Chen et al.21 further applied ultrahigh-resolution OCT for distin- guishing age-related changes in the retina, showcasing advancements that can be adapted to GI segmentation by emphasizing high-resolution boundaries. Quintana et al.22 introduced a method for scleral lens fitting based on OCT imaging, which accurately identifies transitional boundaries, an approach directly relevant for segmenting GI tract organs with complex structures.

Additionally, Wang and Xia23 proposed a volatility derivative framework, incorporating stochastic processes, to enhance model reliability in financial analysis, offering insights into model stability that influenced our en- semble design.

These works, while contributing valuable advancements, do not fully address the unique demands of GI tract segmentation. Our study builds on these foundations by introducing a multi-path approach that combines advanced architectures, targeted preprocessing, and robust ensemble techniques, specifically designed to enhance segmentation precision and efficiency in GI imaging.

3. METHODOLOGY

Our approach integrates multiple deep learning architectures in a synergistic model, specifically designed to address the challenges associated with gastrointestinal (GI) tract segmentation in MRI. By utilizing distinct architectures for classification, 2.5D processing, and grayscale segmentation, our model captures both high-level classification and fine-grained details necessary for accurate segmentation.

3.1 Model Architecture Overview

The proposed model comprises three interdependent pathways, each optimized for a unique aspect of the seg- mentation task. Illustrated in Figure 1, the architecture is designed to leverage the combined strengths of Inception-V4, UNet++ with VGG19 encoding, and Edge U-Net to achieve comprehensive segmentation cover- age across various MRI data types.

3.1.1Inception-V4 Pathway for Initial Classification

The first pathway employs the Inception-V4 architecture, a model known for efficiently processing complex image patterns. In our segmentation pipeline, Inception-V4 serves as the initial classifier, identifying key anatomical regions in MRI scans such as the colon, small intestine, and stomach. Only when these regions are detected does the model proceed to more detailed segmentation; otherwise, a blank mask is generated, streamlining the process and optimizing computational resources.

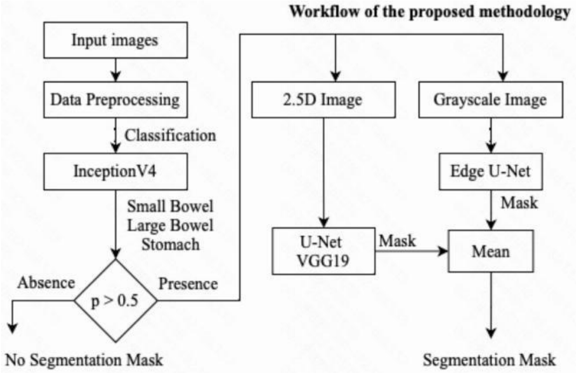


Figure 1. Overview of Model Architecture

3.1.22.5D UNet++ Pathway with VGG19 Encoder

For enhanced spatial awareness, the second pathway uses a 2.5D approach by stacking consecutive MRI slices to generate a multi-dimensional context. This 2.5D representation is processed using UNet++ with a VGG19 encoder, where the architecture captures intricate anatomical features while preserving essential contextual information. This pathway excels in segmenting regions with complex boundaries, enhancing overall model robustness.

3.1.3Edge U-Net for Grayscale Segmentation

The third pathway focuses on grayscale image processing, integrating an Edge U-Net framework equipped with Holistically-Nested Edge Detection (HED). This pathway detects edges and contours more effectively in grayscale data, which simplifies computational requirements and improves segmentation accuracy for organ boundaries. Grayscale data is particularly useful for highlighting edges without the interference of color information.

The outputs from each pathway are aggregated by averaging, which produces a unified segmentation map that combines both broad anatomical context and fine boundary detail. This ensemble approach ensures that each architecture contributes its strengths to the final segmentation result.

3.2 Data Preprocessing

To optimize model performance and ensure adaptability to diverse imaging conditions, a tailored data prepro- cessing pipeline was developed. As depicted in Figure 2, the preprocessing process includes both spatial and intensity augmentations, aimed at enhancing model generalization and improving segmentation accuracy across various MRI datasets.

3.2.1Spatial Augmentation Process

The spatial augmentation step standardizes the resolution of all input images to 320x384 pixels, facilitating consistency in feature extraction. Augmentation techniques, including random rotations, horizontal flipping, and elastic deformations, are applied to simulate various imaging conditions. These augmentations expand the training data and improve the model's resilience to distortions, occlusions, and positional variations in the target organs.

3.2.2Intensity Augmentation for Grayscale Images

A separate intensity augmentation process is applied to grayscale images to enhance the model's sensitivity to pixel-level variations, which are crucial for boundary detection. This step includes adjustments to brightness, contrast, and intensity, creating a richer dataset that enables the model to capture subtle features in grayscale MRI scans. These intensity adjustments support improved edge detection in grayscale images.

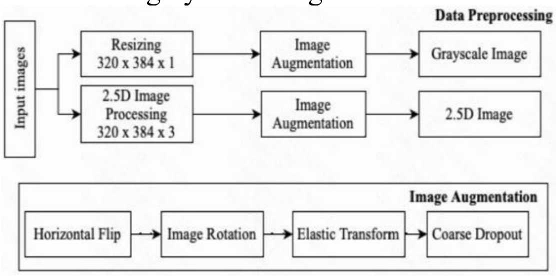
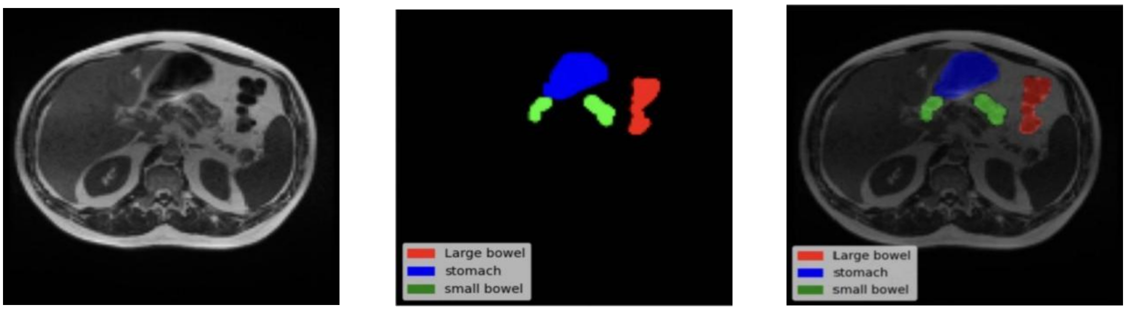


Figure 2. Data Preprocessing Workflow

3.2.32.5D Image Processing for Enhanced Context

In the 2.5D pathway, consecutive MRI slices are stacked to provide the model with added spatial context. This pseudo-3D representation combines depth information across slices without the computational demands of a full 3D model. This technique preserves key features between slices and enriches contextual detail, making it particularly effective in segmenting the GI tract's layered structures. Following augmentation, 2.5D images are input into the UNet++ pathway, enhancing the segmentation model's ability to handle complex anatomical boundaries.

To facilitate visual analysis and mask generation, regions are converted from 16-bit RLE encoding to pixel data, as illustrated in Figure 3. This approach enables enhanced visualization and validation of labeled regions within MRI scans.



(a)Original Image (b) Segmented Masks (c) MRI with Overlayed Masks

Figure 3. Visualization of Segmentation Masks

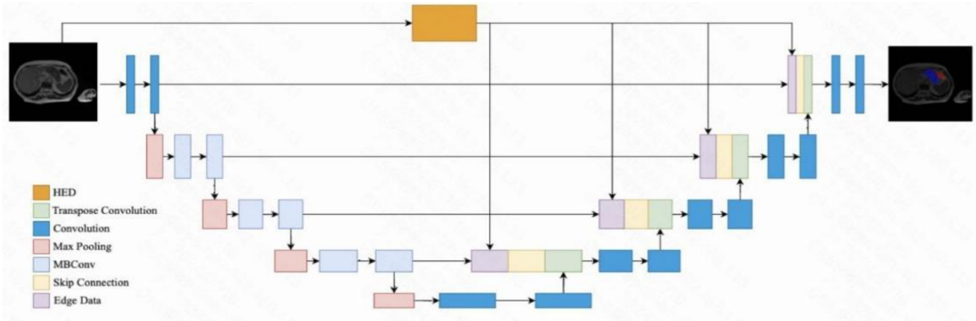


Figure 4. Edge U-Net Architecture with HED

3.3 Model Architectures

3.3.1Inception-V4 for Initial Classification

Inception-V4, developed for complex classification tasks, is used in our model as the initial classification layer. Designed to reduce computational load while retaining high accuracy, Inception-V4 incorporates various filter sizes, batch normalization, and residual connections, which together enable it to efficiently identify large-scale anatomical regions within MRI scans. These design elements facilitate the preliminary classification that guides subsequent segmentation processes.

3.3.2UNet++ with VGG19 Encoder for 2.5D Segmentation

Our segmentation architecture combines UNet++ with a VGG19 encoder to balance depth and detail. The VGG19 encoder extracts high-resolution features from the 2.5D image input, while UNet++'s nested skip connections allow for intricate feature integration across scales. These skip connections effectively maintain fine structural details, providing enhanced segmentation accuracy for GI regions with complex anatomy.

3.3.3Edge U-Net with HED for Grayscale Data

The Edge U-Net module is specifically designed to capture clear edges in grayscale MRI images. By incorporating HED, Edge U-Net can identify boundary features at multiple scales, as illustrated in Figure 4. The MBconv blocks in Edge U-Net allow for efficient processing while preserving edge integrity, which is critical for precise GI organ delineation.

3.4 Training and Parameter Settings

To optimize learning, the model is trained with a batch size of 16 and an initial learning rate of 0.001. An adaptive learning rate scheduler, coupled with the Adam optimizer, adjusts the rate throughout the training process. Training is conducted over 50 epochs, with early stopping applied if the validation loss does not improve over 10 consecutive epochs. These settings balance model performance and training efficiency.

3.5 Evaluation Metrics

3.5.1Dice Coefficient (DC)

The Dice Coefficient measures the similarity between predicted and actual segmentations, providing an overlap- based evaluation metric for segmentation quality. It is computed as:
where PM and OM represent the predicted and ground truth masks, respectively.

$$DC(PM, OM) = \frac{2 \times |PM \cap OM|}{|PM| + |OM|} \#(1)$$

3.5.23D Hausdorff Distance

The 3D Hausdorff Distance captures spatial dissimilarity between two segmentation masks by measuring the maximum separation between the closest points in the predicted and ground truth masks:

$$HD(PM, OM) = \max\left(\max_{pm,om} \min(|pm - om|)\right) \#(2)$$

This metric assesses the spatial accuracy of the segmentation, crucial for precise anatomical delineation.

3.5.3Composite Score

Combining Dice Coefficient and 3D Hausdorff Distance, a composite score evaluates overall segmentation per- formance:

Score = 0.4 × Dice Coefficient + 0.6 × 3D Hausdorff Distance (3)

This score provides a balanced view of model accuracy and robustness in delineating complex GI structures.

4. Experimental Results

Our proposed model was rigorously evaluated on both grayscale and 2.5D MRI datasets to assess its segmentation performance across different configurations. Results for each model architecture are presented below, highlighting the comparative advantages of each approach.

4.1 Dataset Description

The dataset used in this study includes MRI scans of the GI tract, annotated for key regions such as the colon, small intestine, and stomach. This dataset, divided into training and validation sets, was sourced from multiple institutions to ensure diversity and robustness in model evaluation.

4.2 Results on Grayscale Images

In the grayscale image experiments, different encoders were tested within the UNet and UNet++ architectures, alongside Edge U-Net as a baseline for edge-focused segmentation. The validation results are shown in Table 1.

Table 1. Grayscale Image Segmentation Results

Model	Encoder	Validation Score
UNet	ResNet50	0.71599
UNet	Inception-V4	0.71002
UNet	Xception	0.73761
UNet	EfficientNet-B0	0.68033
UNet	VGG19	0.78925
UNet++	ResNet50	0.7899
UNet++	Inception-V4	0.80095
UNet++	Xception	0.79711
UNet++	EfficientNet-B0	0.71372
UNet++	VGG19	0.80717
Edge UNet	-	0.84046

Edge U-Net achieved the highest score for grayscale images, reaching a validation accuracy of 0.84046, outper- forming both UNet and UNet++ configurations. The high performance of Edge U-Net suggests that edge-focused architectures are particularly suited for grayscale MRI data, where precise boundary detection is critical.

4.3 Results on 2.5D Images

In the 2.5D image experiments, UNet++ with various encoder configurations was evaluated. Validation scores are presented in Table 2.

Table 2: 2.5D Image Segmentation Results

Model	Encoder	Validation Score
UNet++	ResNet50	0.80138
UNet++	Xception	0.7961
UNet++	VGG19	0.84984

For 2.5D images, UNet++ with the VGG19 encoder showed the best performance, achieving a validation score of 0.84984. This demonstrates the value of combining VGG19’s feature extraction capabilities with 2.5D data to capture contextual depth, which enhances segmentation quality in regions with complex anatomical structures.

5. Conclusion

This study presents a comprehensive, automated approach for gastrointestinal (GI) tract segmentation in MRI scans, integrating advanced deep learning architectures to address the unique challenges associated with medical imaging for radiotherapy planning. By combining Inception-V4 for initial classification, UNet++ with VGG19 for 2.5D data processing, and Edge U-Net for edge-focused grayscale segmentation, the model achieves high segmentation accuracy across different MRI data types.

The experimental results demonstrate that Edge U-Net is particularly effective in grayscale segmentation tasks due to its edge detection capabilities, while UNet++ with VGG19 excels in 2.5D data processing, bene- fiting from additional depth context in the MRI slices. These complementary strengths support a multi-path approach, enabling precise segmentation across diverse imaging scenarios and reducing the inter-observer vari- ability commonly associated with manual segmentation.

This work contributes to the field by offering a flexible, high-performance segmentation solution tailored for GI imaging, ultimately aimed at improving the efficiency and consistency of radiotherapy planning. Future work may focus on further refining the ensemble model and exploring its application to other types of medical imaging to expand its utility in clinical practice.

References

[1]Kocak, B., Yardimci, A. H., Yuzkan, S., Keles, A., Altun, O., Bulut, E., Bayrak, O. N., and Okumus, A. A., "Transparency in artificial intelligence research: a systematic review of availability items related to open science in radiology and nuclear medicine," Academic Radiology 30(10), 2254–2266 (2023).
[2]Zhou, Z., Qian, X., Hu, J., Chen, G., Zhang, C., Zhu, J., and Dai, Y., "An artificial intelligence-assisted diagnosis modeling software (aims) platform based on medical images and machine learning: a development and validation study," Quantitative Imaging in Medicine and Surgery 13(11), 7504 (2023).
[3]Lu, J., Jiyu, W., Lu, N., and Wente, Z., "Decoupled modeling methods and systems," (Nov. 29 2022). US Patent 11,514,537.
[4]Chen, L., Li, J., Zou, Y., and Wang, T., "Etu-net: edge enhancement-guided u-net with transformer for skin lesion segmentation," Physics in Medicine & Biology 69(1), 015001 (2023).
[5]Tianbo, S., Weijun, H., Jiangfeng, C., Weijia, L., Quan, Y., and Kun, H., "Bio-inspired swarm intelligence: a flocking project with group object recognition," in [2023 3rd International Conference on Consumer Electronics and Computer Engineering (ICCECE)], 834–837, IEEE (2023).
[6]Zhang, Z., Tian, R., and Ding, Z., "Trep: Transformer-based evidential prediction for pedestrian intention with uncertainty," in [Proceedings of the AAAI Conference on Artificial Intelligence], 37 (2023).
[7]Chen, K., Zhuang, D., and Chang, J. M., "Supercon: Supervised contrastive learning for imbalanced skin lesion classification," arXiv preprint arXiv:2202.05685 (2022).
[8]Maccioni, F., Busato, L., Valenti, A., Cardaccio, S., Longhi, A., and Catalano, C., "Magnetic resonance imaging of the gastrointestinal tract: Current role, recent advancements and future prospectives," Diagnos- tics 13(14), 2410 (2023).
[9]Ronneberger, O., Fischer, P., and Brox, T., "U-net: Convolutional networks for biomedical image segmenta- tion," in [Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18], 234–241, Springer (2015).
[10]Simonyan, K. and Zisserman, A., "Very deep convolutional networks for large-scale image recognition,"
arXiv preprint arXiv:1409.1556 (2014).
[11]Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., and Wojna, Z., "Rethinking the inception architecture for computer vision," in [Proceedings of the IEEE conference on computer vision and pattern recognition], 2818–2826 (2016).
[12]Badrinarayanan, V., Kendall, A., and Cipolla, R., "Segnet: A deep convolutional encoder-decoder archi- tecture for image segmentation," IEEE transactions on pattern analysis and machine intelligence 39(12), 2481–2495 (2017).
[13]Zhang, Y., Wang, X., Gao, L., and Liu, Z., "Manipulator control system based on machine vision," in [International Conference on Applications and Techniques in Cyber Intelligence ATCI 2019: Applications and Techniques in Cyber Intelligence 7], 906–916, Springer (2020).
[14]Zhang, Q., Cai, G., Cai, M., Qian, J., and Song, T., "Deep learning model aids breast cancer detection,"
Frontiers in Computing and Intelligent Systems 6(1), 99–102 (2023).
[15]Liao, J., Kot, A., Guha, T., and Sanchez, V., "Attention selective network for face synthesis and pose- invariant face recognition," in [2020 IEEE International Conference on Image Processing (ICIP)], 748–752, IEEE (2020).
[16]Lin, J., Deep-learning Enabled Accurate Bruch's Membrane Segmentation in Ultrahigh-Resolution Spectral Domain and Ultrahigh-Speed Swept Source Optical Coherence Tomography, PhD thesis, Massachusetts In- stitute of Technology (2022).
[17]Xiao, T., Zeng, L., Shi, X., Zhu, X., and Wu, G., "Dual-graph learning convolutional networks for in- terpretable

alzheimer's disease diagnosis," in [International Conference on Medical Image Computing and Computer-Assisted Intervention], 406–415, Springer (2022).

[18]Hu, J., Wang, X., Liao, Z., and Xiao, T., "M-gcn: Multi-scale graph convolutional network for 3d point cloud classification," in [2023 IEEE International Conference on Multimedia and Expo (ICME)], 924–929, IEEE (2023).

[19]Zeng, L., Li, H., Xiao, T., Shen, F., and Zhong, Z., "Graph convolutional network with sample and feature weights for alzheimer's disease diagnosis," Information Processing & Management 59(4), 102952 (2022).

[20]Chen, S., Potsaid, B., Li, Y., Lin, J., Hwang, Y., Moult, E. M., Zhang, J., Huang, D., and Fujimoto, J. G., "High speed, long range, deep penetration swept source oct for structural and angiographic imaging of the anterior eye," Scientific reports 12(1), 992 (2022).

[21]Chen, S., Abu-Qamar, O., Kar, D., Messinger, J. D., Hwang, Y., Moult, E. M., Lin, J., Baumal, C. R., Witkin, A., Liang, M. C., et al., "Ultrahigh resolution oct markers of normal aging and early age-related macular degeneration," Ophthalmology Science 3(3), 100277 (2023).

[22]Quintana, C. L., Chen, S., Lin, J., Fujimoto, J. G., Li, Y., and Huang, D., "Anterior topographic limbal demarcation with ultrawide-field oct," Investigative Ophthalmology & Visual Science 63(7), 1195–A0195 (2022).

[23]Wang, L. and Xia, W., "Power-type derivatives for rough volatility with jumps," Journal of Futures Mar- kets 42(7), 1369–1406 (2022).