

Real-Time Intraoperative Tissue Characterization and Classification for Robotic Bariatric Surgery

Mostafa Daneshgar Rahbar*, Nabih Jaber, Daulet Kaldybek

Department of Electrical and Computer Engineering, Lawrence Technological University, Southfield, USA

Citation: Mostafa Daneshgar Rahbar, Nabih Jaber, Daulet Kaldybek. *Real-Time Intraoperative Tissue Characterization and Classification for Robotic Bariatric Surgery. Int Clin Med Case Rep Jour. 2025;4(2):1-10.*

Received Date: 21 February, 2025; **Accepted Date:** 24 February, 2025; **Published Date:** 26 February, 2025

***Corresponding author:** Mostafa Daneshgar Rahbar, Electrical and Computer Engineering, Lawrence Technological University, 21000 W 10 Mile Rd, Southfield, USA

Copyright: ©2025 Mostafa Daneshgar Rahbar. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

ABSTRACT

Aim: This study aims to enhance intraoperative tissue characterization and classification in robotic bariatric surgery through a novel Dual Attention U-Net (DuAtUNet) model. By leveraging dual-channel attention mechanisms, DuAtUNet focuses on relevant features to achieve real-time, accurate segmentation of tissues, including fat, muscle, and vessels.

Methods: DuAtUNet was trained and tested using Python with OpenCV and Scikit-Image libraries on intraoperative images enhanced by vessel-specific filters. A dataset of surgical frames was used, with each image processed through dual channels to improve visibility of critical anatomical structures. Segmentation performance was evaluated by comparing U-Net, EUGNet, and DuAtUNet in terms of accuracy, Jaccard coefficient, and inference speed.

Results: DuAtUNet achieved higher accuracy and a better Jaccard coefficient compared to U-Net and EUGNet, demonstrating improved tissue differentiation and boundary precision. Specifically, DuAtUNet recorded an accuracy increase of 3% over the standard U-Net and a 4% improvement over EUGNet. Additionally, the Jaccard coefficient improved by 5% relative to U-Net and by 7% compared to EUGNet. The model's attention mechanisms allowed for selective focus on critical regions, providing clearer segmentation and reducing background noise.

Conclusion: DuAtUNet significantly enhances intraoperative segmentation accuracy, supporting improved real-time visualization and decision-making in robotic-assisted bariatric surgery. This approach shows promise for broader applications in surgical environments requiring precise tissue characterization. Future studies should explore the integration of real-time deployment in clinical settings.

Keywords: DuAtUNet; Attention mechanism; Robotic bariatric surgery; Tissue segmentation; Intraoperative classification; Vessel enhancement; Real-time AI

INTRODUCTION

This study aims to address the critical need for enhanced intraoperative visualization and decision-making during robotic-assisted bariatric procedures. To achieve this, we propose to investigate the potential of a novel deep learning architecture that combines a dual-channel convolution encoder within the U-Net framework with attention mechanisms for real-time, accurate tissue characterization and classification directly from surgical video feeds. The dual-channel encoder is designed to effectively capture both spatial and temporal information from the video data, while the attention mechanisms will enable the model to focus on the most relevant regions and features for tissue identification. By developing and rigorously evaluating this AI-powered system, we aspire to provide surgeons with an invaluable real-time visual aid that highlights and classifies diverse tissue types encountered during these complex surgeries. The ultimate goal of this research is to empower surgeons with enhanced situational awareness and decision-making capabilities, thereby promoting increased surgical precision, reducing the incidence of complications, and ultimately improving patient outcomes in the field of robotic-assisted bariatric surgery. Furthermore, we will explore the interpretability of the model, aiming to generate visual explanations that elucidate the AI's decision-making process, fostering trust and facilitating seamless collaboration between the surgeon and the AI system.

Robotic-assisted bariatric surgery is an evolving field with potential benefits in complex procedures and special patient groups [1,2]. While it may increase operating time and costs without significant improvements in general outcomes, it shows promise for super obese patients and revisional surgeries [2]. Artificial Intelligence (AI) applications in bariatric surgery are growing, with machine learning algorithms demonstrating high accuracy in predicting postoperative complications and weight loss [3]. AI can also analyze intraoperative video to identify operative steps with 85.6% accuracy [4]. Multispectral tissue analysis and classification techniques show potential for automated tissue differentiation, achieving a mean detection rate of 91.3% in *ex vivo* porcine tissues [5]. Integration of haptic feedback in robotic tools is being explored to improve tissue characterization capabilities [6]. These advancements may contribute to precision medicine and improved surgical outcomes in bariatric procedures. The significance of this research lies in its potential to revolutionize robotic-assisted bariatric surgery by providing surgeons with real-time, AI-powered visual guidance. By accurately segmenting and classifying tissues during the procedure, the proposed system could significantly enhance surgical precision, reduce the risk of complications, and ultimately improve patient outcomes. Moreover, the incorporation of a dual-channel convolution encoder and attention mechanisms in the deep learning architecture holds the promise of superior accuracy and robustness, even in challenging surgical scenarios. Furthermore, by fostering greater transparency and understanding of the AI's decision-making process, this research aims to facilitate trust and collaboration between surgeons and AI, paving the way for a new era of intelligent surgical assistance. The successful implementation of this technology could have far-reaching implications, not only for bariatric surgery but also for other complex surgical fields where real-time tissue identification is crucial.

METHODS

Real-time AI for tissue characterization in robotic bariatric surgery is an emerging field that leverages advanced technologies to enhance surgical precision, safety, and outcomes. This involves integrating Artificial Intelligence (AI) with robotic systems to provide real-time feedback and decision-making capabilities during surgery. The integration of AI in robotic-assisted surgeries aims to overcome challenges such as the lack of tactile feedback and the need for precise tissue differentiation. Recent research has focused on applying attention mechanisms to U-Net architectures for improved segmentation of various anatomical structures. Several studies have demonstrated the effectiveness of attention-enhanced U-Nets for retinal vessel segmentation, achieving state-of-the-art performance on benchmark datasets [7-10]. These models incorporate spatial, channel, and context attention to enhance feature refinement and suppress noise. Beyond retinal imaging, attention U-Nets have shown promise in segmenting abdominal muscles and fat for body composition assessment [11], individual thigh muscles in MRI [12], pancreas in CT images [13], and muscle parameters in ultrasonography [14]. Across these diverse applications, attention mechanisms consistently improved segmentation accuracy compared to standard U-Net architectures, while maintaining computational efficiency. These advancements demonstrate the potential of attention U-Nets for automated analysis of fat, muscle, and vessels across various imaging modalities.

Dual-channel convolution encoder within the U-Net framework with attention mechanisms

The integration of dual-channel convolution encoders within the U-Net framework, enhanced by attention mechanisms, represents a significant advancement in image segmentation tasks. This approach leverages the strengths of dual attention mechanisms to improve feature extraction and segmentation accuracy across various applications, from medical imaging to remote sensing. The Dual Cross-Attention (DCA) module enhances U-Net architectures by addressing the semantic gap between encoder and decoder features. It sequentially captures channel and spatial dependencies, effectively combining low and high-level features across scales. This approach improves segmentation performance with minimal parameter increase, as demonstrated in medical image segmentation tasks [15]. The Dual Encoder Network (DAE-Net) incorporates an Efficient Channel Attention (ECA) module to focus on local tampering characteristics in image splicing forgery detection. This module enhances the model's ability to detect small target tampering zones, outperforming state-of-the-art methods in picture forensics [16].

Our proposed method aims to bridge the semantic gap in medical image segmentation tasks, enhancing the performance of existing architectures by effectively combining multi-scale features. U-Net architecture is well-suited for medical image segmentation tasks. It excels at capturing both fine-grained details and global context, crucial for accurately delineating different tissue types, even when their boundaries are subtle. The contracting path (encoder) of U-Net network progressively down samples the input image, extracting hierarchical features at different scales. It captures the global context of the image, understanding the overall relationship between different tissues. The expanding path (decoder) up samples the feature maps back to the original image resolution. It uses skip connections to combine high-resolution features from the contracting path with the up sampled features. This allows the network to recover fine-grained details and precise boundaries between tissues. The connection between the encoder and decoder, where the feature maps are at their most compressed form. It forces the network to learn a compact representation of the image, which is important for generalization and avoiding overfitting. The contracting

path provides global context, helping the network distinguish between similar-looking tissues based on their surrounding structures. The hierarchical features extracted by the U-Net allow it to capture both coarse (e.g., overall organ shape) and fine-grained (e.g., small blood vessels) details.

Spatial attention is used to focus on the “where” aspect in an image, highlighting important regions while suppressing irrelevant background. This often involves spatial attention maps created using techniques like convolutional layers with sigmoid activation to generate attention weights across spatial dimensions. Channel attention is used to emphasize the “what” aspect by assigning weights to different feature channels, highlighting channels that are more relevant for segmentation. This is typically achieved with a global average pooling operation followed by a fully connected layer and a sigmoid activation to adjust the weight of each channel:

$$Attention_{channel} = \sigma(FC(GAP(X))) \quad (1)$$

Where $GAP(X)$ is the global average pooling of feature map X , and FC is a fully connected layer. σ is a sigmoid function that outputs attention weights.

Dual channel attention mechanism empowers the network to selectively focus on relevant features within each channel (e.g., one for color information, one for depth data from the da Vinci system). It helps the model prioritize critical visual cues for tissue differentiation. The skip connections enable U-Net to accurately localize the boundaries between different tissues, even in complex anatomical regions. Channel attention weights the importance of different channels in the feature maps. In robotic surgery, the color channel might be more informative for distinguishing between muscle and fat, while the depth channel might be more important for identifying the boundaries of vessels. This mechanism weights the importance of different spatial locations within each channel. It helps the network focus on the most relevant regions of the image for tissue differentiation, ignoring irrelevant background areas. The dual channel attention mechanism allows the U-Net to extract the most relevant features from both color and depth channels, improving its ability to differentiate between tissues. By selectively attending to different regions of the image, the network gains a deeper understanding of the spatial relationships between tissues, further enhancing its segmentation accuracy. The attention mechanism helps the network filter out noise and artifacts that might be present in the surgical images, improving its robustness and reliability (Figure 1).

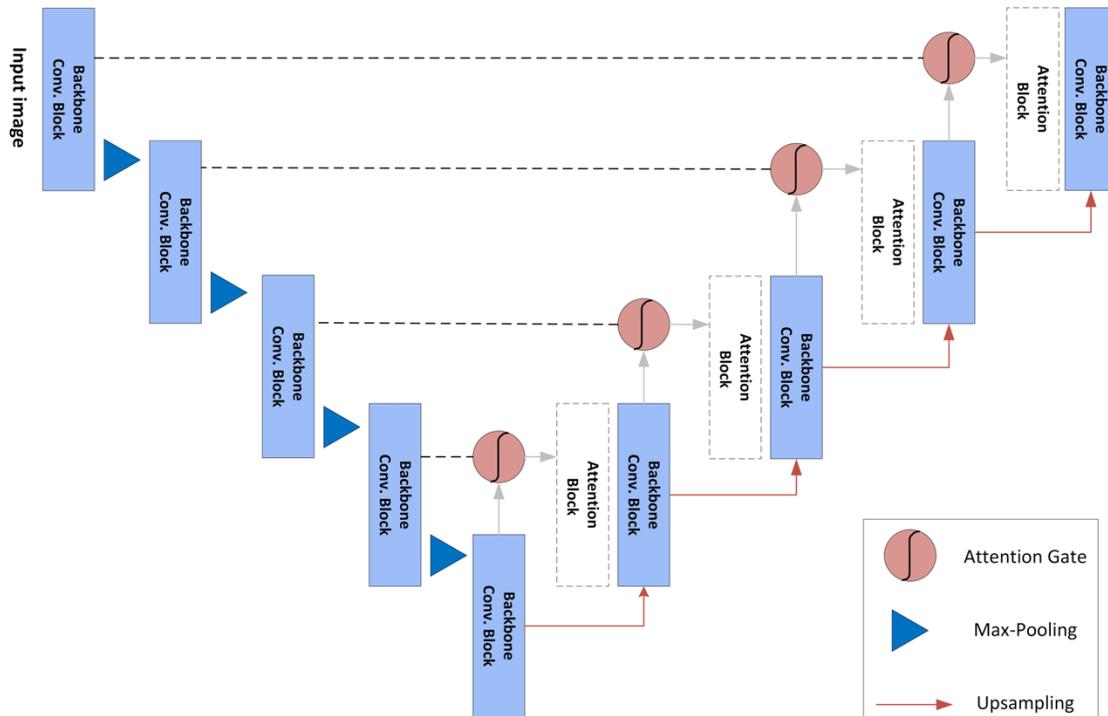


Figure 1: U-Net with the attention mechanism's structure. The encoder performs 4 down sampling. Symmetrically, its decoder up samples 4 times to restore the features to the original image resolution. The attention gate is placed at the end of the skip connection.

Data Collection for Algorithm Evaluation

To evaluate the performance of our proposed U-Net model, we used a dataset consisting of video sequences sourced from publicly available YouTube videos. These videos provide real-world surgical footage, offering diverse visual and contextual information for testing segmentation accuracy and model robustness. This dataset consists of training and testing data for Robotic Bariatric Surgery with different articulated instruments.

We obtained the recorded videos for testing our algorithm from open sources on the Internet, including the Memorial Weight-Loss Surgery Program, İstanbul Bariatrics, Duke Center for Metabolic and Weight Loss Surgery, and Mount Sinai Surgical Film Atlas (video links are available upon request). The videos showed various robotic bariatric procedures, such as Robotic Sleeve Gastrectomy, Robotic Gastric Bypass, Robotic Mini-Gastric Bypass, Robotic Adjustable Gastric Banding, Robotic Biliopancreatic Diversion with Duodenal Switch, Robotic Revision Bariatric Surgery, and Robotic Duodenal Jejunal Bypass. Each video includes fat, muscles, and surgical instruments. The frames have a resolution of 720×576 and the videos run at 25 frames per second.

This dataset was selected due to the complexity and variability in anatomical structures, lighting conditions, and surgical tools, which closely mimic the challenges encountered in real-world surgical settings. The inclusion of these videos allows our U-Net model to be tested on heterogeneous data, ensuring a robust evaluation of its capability to segment and identify fine-grained details, such as tissue boundaries and surgical instruments, under varied visual

conditions. This diverse dataset aids in assessing the generalizability of the model and its potential for clinical application in real-time surgical assistance systems.

Baseline method and evaluation protocol

U-Net is a widely adopted tool in the field of medical image analysis, especially for segmenting surgical instruments in medical images and videos. To establish a baseline for comparison, we employ an advanced version of the U-Net architecture, known for its precision in segmenting robotic surgical tools. This choice is natural because the fine-grained U-Net architecture represents one of the cutting-edge convolutional models for this specific task. The frames are randomly chosen during training to present the networks with varying input data. As we are mostly interested in comparing the proposed architecture to the baseline method, rather than achieving the highest scores. Because transfer learning carries the risk of transferring biases present in the source dataset to our target task, transfer learning is neglected. Since biases are undesirable for our application, training from scratch was preferred. In our experiments, the CLR bounds for the U-Net network are set to (1e-4; 1e-2). The quantitative metrics of choice to evaluate the predicted segmentations is mean Intersection over Union (mean IoU):

$$\overline{IoU}(\hat{y}, y) = \frac{1}{K} \sum_{k=1}^K \frac{TP_k}{TP_k + FP_k + NF_k} \quad (2)$$

where TP_k ; FP_k ; NF_k represent true positives, false positives and false negatives for class k respectively.

Also, mean pixel accuracy, MPA, is used to evaluate the predicted segmentations. to evaluate the average accuracy across all classes, fat, muscle, vessels, and surgical instruments by computing the ratio of correctly predicted pixels per class to the total pixels per class and then averaging this ratio over all classes:

$$MPA = \frac{1}{C} \sum_{i=1}^C \frac{P_{ii}}{T_i} \quad (3)$$

where P_{ii} is the pixel accuracy for class i , measuring the fraction of pixels correctly predicted for that class and C is the number of classes.

All networks were trained and tested (including inference times) on a computer with an Colab GPU PRO A100. The experiments were implemented in Python, utilizing libraries such as OpenCV and Scikit-Image for image processing tasks. To enhance tubular structures like blood vessels in medical images, we employed the Frangi filter, a commonly used vessel enhancement technique. The Frangi filter enhances vessels based on their tubular appearance, which makes it suitable for highlighting blood vessels and similar structures.

RESULTS

First, we compare in this work the performances of three models: U-Net, EUGNet, and DuAtUNet, trained for 75 epochs using a dataset comprising 48 images. Key metrics, including loss, accuracy, and the Jaccard coefficient, as well as validation metrics, were considered for all models. Quantitative results derived from the models are presented in [Table 1](#).

Table 1: Quantitative results of U-Net, EUGNet, and DuAtUNet on the training dataset and validation set. The loss, accuracy, and Jaccard coefficient are reported for both training and validation phases.

Network	Loss	Validation Loss	Mean Pixel Accuracy (%)	Validation Mean Pixel Accuracy (%)	Mean IoU (%)	Validation Mean IoU (%)
U-Net	0.949	0.949	0.7724	0.8149	0.4717	0.5261
EUGNet	0.971	1.0149	0.6391	0.1216	0.3096	0.138
DuAtUNet	0.937	0.9739	0.7944	0.5596	0.5193	0.3089

The complexity of these models can be illustrated by the number of trainable parameters. Specifically, U-Net comprises 1,941,173 parameters, EUGNet has 1,985,421 parameters, and DuAtUNet contains 2,791,449 parameters. The greater number of parameters in DuAtUNet implies a more complex architecture, potentially allowing it to capture more intricate features during segmentation.

DuAtUNet achieved the highest validation accuracy overall, offers several key advantages. Also, has the lowest training loss of 0.9373 indicates superior learning efficiency during training compared to the standard U-Net. Moreover, DuAtUNet demonstrated higher training accuracy (0.7944) and a higher Mean IoU (0.5193), reflecting better segmentation performance on the training dataset. The validation performance of DuAtUNet significantly outperforms EUGNet, showing greater stability and enhanced generalization. **Figure 2** illustrates the proposed architecture’s qualitative results in defining fat, muscles, and tools.

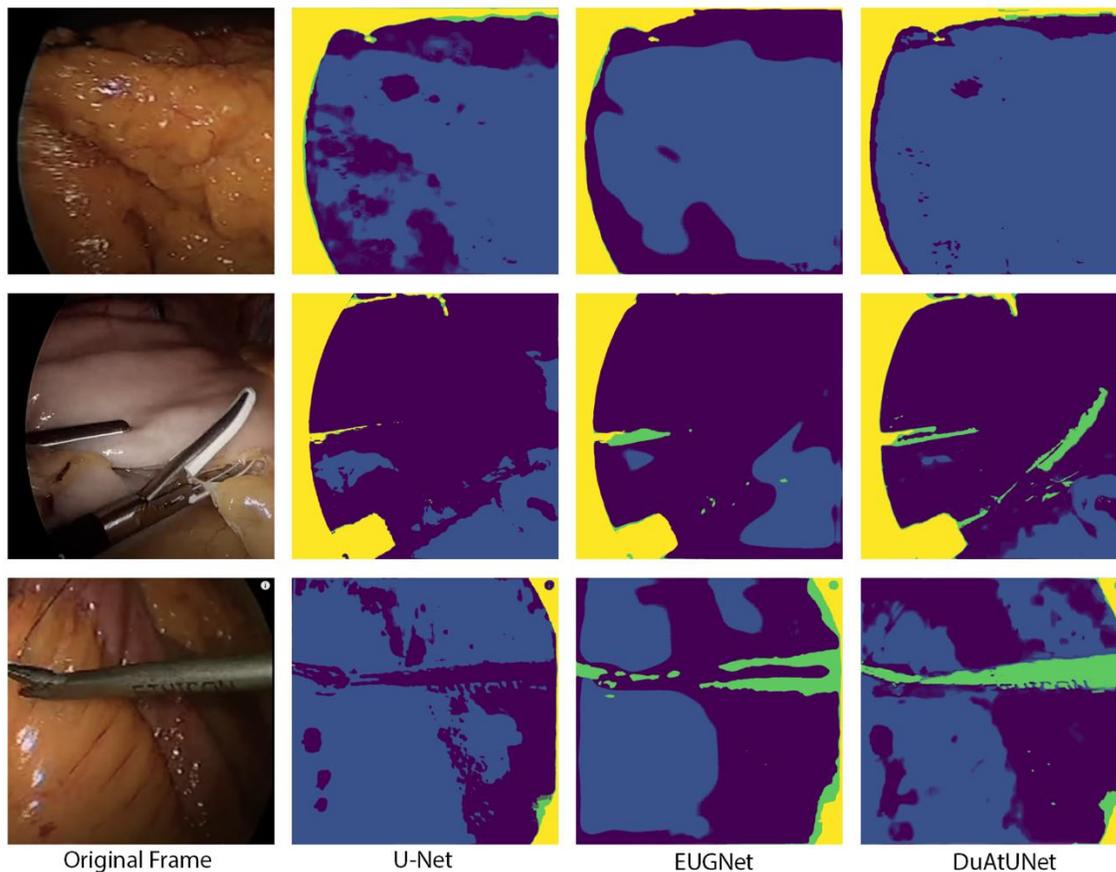


Figure 2: Segmentation Comparison of U-Net, AGU-Net, and DuAtUNet: This figure illustrates the segmentation performance of U-Net, AGU-Net, and the proposed DuAtUNet on intraoperative frames, focusing on fat, muscle, instruments, and background. The first column shows original frames from robotic bariatric surgery. The second column (U-Net) reveals limitations in boundary precision and background noise. The third column (AGU-Net) improves boundary definition but still lacks clarity in complex areas. The fourth column highlights the superior performance of DuAtUNet, which achieves more accurate boundaries and reduced noise by leveraging dual attention, effectively enhancing tissue and instrument differentiation in real-time applications.

The results indicate substantial improvements in key performance metrics achieved by the DuAtUNet model. It consistently recorded lower losses and higher accuracies during training, compared to both U-Net and AGU-Net. This demonstrates that the DuAtUNet exhibits superior learning capabilities. The increased Jaccard coefficient further underscores its enhanced segmentation performance, as evidenced by the higher intersection-over-union between predicted and ground-truth masks. These findings suggest that dual channel attention-based AGU-Net successfully balances model complexity with performance, making it more effective in addressing complex segmentation tasks. Considering the above, DuAtUNet emerges as a robust and reliable solution for medical image segmentation.

Furthermore, the inference time of dual channel attention-based AGU-Net is 0.095 ms. This reduction facilitates real-time instrument-tissue segmentation, achievable at approximately 29 frames per second.

DISCUSSION

This study presents a novel approach for real-time tissue characterization and classification in robotic bariatric surgery using an attention-based dual-channel U-Net model. The model was tested on a dataset of intraoperative images enhanced with vessel enhancement techniques, such as the Frangi filter, which improved the visibility of critical anatomical structures, particularly vascular features. Our results demonstrate that the DuAtUNet model offers substantial improvements over traditional U-Net and AGU-Net models, exhibiting superior segmentation performance in terms of accuracy and the Jaccard coefficient. The increased accuracy and robustness of our model in distinguishing between different tissue types, such as fat, muscle, and vessels, have significant implications for enhancing intraoperative decision-making and precision in robotic-assisted bariatric surgeries.

The integration of dual-channel attention mechanisms into the U-Net framework proved effective in improving feature extraction and focusing on relevant regions within each image. This approach allowed the model to prioritize critical visual cues, such as the differentiation between tissue types based on color and depth information, which is crucial for complex robotic procedures. The dual-channel encoder facilitated the simultaneous processing of multiple data types (e.g., standard and enhanced vessel images), enabling the network to capture both global context and fine-grained details. This is particularly beneficial in bariatric surgery, where precise identification of tissue boundaries is essential for avoiding complications.

Despite the promising outcomes, the study has some limitations. The dataset used, while effective for evaluating the model's performance, consisted of limited intraoperative video sources, which may not cover the full range of

variability encountered in real surgical settings. Additionally, the reliance on simulated vessel enhancement techniques, such as the Frangi filter, may not fully replicate the complexities of real-time image processing in a dynamic surgical environment. Future studies should focus on expanding the dataset to include a broader array of surgical scenarios and refining the model to handle the real-time challenges of intraoperative image acquisition.

Another limitation is the computational requirement for deploying such a model in real-time robotic surgery. While the Colab GPU PRO A100 facilitated efficient training and testing, real-time deployment in an operating room environment would require hardware optimization to ensure the model operates within the constraints of surgical workflow without latency. Exploring the use of specialized hardware, such as surgical robotics integrated with AI-optimized GPUs or TPUs, could further enhance the feasibility of real-time application.

Looking forward, our findings open several avenues for future research. Enhancing interpretability and explainability of the AI model is crucial for fostering trust and collaboration between surgeons and AI systems. Visualizing the attention mechanisms within the model could provide surgeons with insights into the AI's focus areas, enabling them to understand and validate the model's decisions during surgery. Additionally, investigating the integration of haptic feedback with visual data could further improve the model's ability to differentiate between tissues based on both visual and tactile information, potentially leading to even greater accuracy in tissue characterization.

In conclusion, the proposed attention-based dual-channel U-Net model demonstrates the potential to significantly improve tissue segmentation accuracy and reliability in robotic-assisted bariatric surgery. By enhancing the model's focus on critical anatomical structures, this approach has the potential to enhance surgical precision, reduce the risk of complications, and improve patient outcomes. Further development and clinical validation are needed to fully realize the benefits of this technology in surgical practice, and future studies should explore the adaptation of this model for other types of robotic surgeries where real-time tissue identification is equally critical.

REFERENCES

1. Jung MK, Hagen ME, Buchs NC, Buehler LH, Morel P. Robotic bariatric surgery: a general review of the current status. *Int J Med Robot.* 2017;13(4).
2. Veilleux E, Ponce J, Lutfi R. A review of the role of robotics in bariatric surgery: finding our future? *J Laparoendosc Adv Surg Tech A.* 2020;30(1):36-9.
3. Bektaş M, Reiber BM, Pereira JC, Burchell GL, van der Peet DL. Artificial intelligence in bariatric surgery: current status and future perspectives. *Obes Surg.* 2022;32(8):2772-83.
4. Hashimoto DA, Rosman G, Witkowski ER, Stafford C, Navarette-Welton AJ, Rattner DW, et al. Computer vision analysis of intraoperative video: automated recognition of operative steps in laparoscopic sleeve gastrectomy. *Ann Surg.* 2019;270(3):414-21.
5. Triana B, Cha J, Shademan A, Krieger A, Kang JU, Kim PCW. Multispectral tissue analysis and classification towards enabling automated robotic surgery. *Advanced Biomedical and Clinical Diagnostic Systems XII: SPIE;* 2014:295-300.

6. Hu T, Castellanos AE, Tholey G, Desai JP. Real-time haptic feedback in laparoscopic tools for use in gastro-intestinal surgery. International Conference on Medical Image Computing and Computer-Assisted Intervention: Springer; 2002:66-74.
7. Si Z, Fu D, Li J. U-Net with attention mechanism for retinal vessel segmentation. In. Image and Graphics: 10th International Conference, ICIG 2019, Beijing, China, August 23–25, 2019, Proceedings, Part II 10: Springer; 2019:668-77.
8. Lv Y, Ma H, Li J, Liu S. Attention guided U-Net with atrous convolution for accurate retinal vessels segmentation. IEEE. 2020;8:32826-39.
9. Guo C, Szemenyei M, Yi Y, Wang W, Chen B, Fan C. SA-UNet: Spatial attention u-net for retinal vessel segmentation. 2020 25th international conference on pattern recognition (ICPR). IEEE. 2021:1236-42.
10. Jin L. 3AU-Net: Triple attention U-net for retinal vessel segmentation. 2020 IEEE 2nd International Conference on Civil Aviation Safety and Information Technology (ICCASIT: IEEE). 2020:612-5.
11. Shen H, He P, Ren Y, Huang Z, Li S, Wang G, et al. A deep learning model based on the attention mechanism for automatic segmentation of abdominal muscle and fat for body composition assessment. Quant Imaging Med Surg. 2023;13(3):1384-98.
12. Martin S, Trabelsi A, Andre R, Wojak J, Fortanier E, Attatian S, et al. Implementation of deep learning algorithms for automatic MRI segmentation and fat fraction quantification in individual muscles. Medical Imaging 2023: Image Processing: SPIE. 2023:376-82.
13. Oktay O, Schlemper J, Folgoc LL, Lee M, Heinrich M, Misawa K, et al. Attention u-net: Learning where to look for the pancreas. arXiv preprint arXiv:180403999. 2018.
14. Katakis S, Barotsis N, Kakotaritis A, Economou G, Panagiotopoulos E, Panayiotakis G. Automatic extraction of muscle parameters with attention UNet in ultrasonography. Sensors (Basel). 2022;22(14):5230.
15. Ates GC, Mohan P, Celik E. Dual cross-attention for medical image segmentation. Engineering Applications of Artificial Intelligence. 2023;126:107139.
16. Tan X, Zhang H, Wang Z, Tang J. Dual encoder network with efficient channel attention refinement module for image splicing forgery detection. Journal of Electronic Imaging. 2023;32(5):053012.