Journal of Information Systems Engineering and Management
2025, 10(18s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Improved Semantic Segmentation in Medical Imaging Using U-Net and Attention Mechanisms

Mamdouh Gomaa[a, b,1], Ahmed Rabie[a,2], Al Hussien Seddik Saad[a,3], Osman Ali Sadek[a,4], Moheb R. Girgis[a,5]
[a] Department of Computer Science, Faculty of Science, Minia University, Egypt
[b] Department of Computer Science, Faculty of Information Technology, Amman Arab University, 11953, Amman, Jordan
[1] mamdouh.gomaa@mu.edu.eg, m.gomaa@aau.edu.jo, [2] arabie@mu.edu.eg, [3] al.hussien_seddik@mu.edu.eg, [4] osman.ibrahim@mu.edu.eg, [5] moheb.girgis@mu.edu.eg

ARTICLE INFO

Received: 15 Dec 2024
Revised: 26 Jan 2025
Accepted: 15 Feb 2025

ABSTRACT

Image segmentation poses a significant challenge in the field of medical image analysis aiming to extract valuable information and enhance clinical diagnosis accuracy. The main aim of this paper is to explore the use of the baseline U-Net architecture and another model that combines the baseline U-Net with attention mechanisms called OrganFocusU-Net model in organ semantic segmentation and differentiation in laparoscopic hysterectomy. These models involve leveraging the UD Ureter-Uterine Artery-Nerve dataset, which is a comprehensive collection from laparoscopic surgeries, accompanied by corresponding multiclass masks and capable of pixel-wise detection and differentiation of three key organs: ureter, uterine artery, and nerves, with a specific emphasis on accurately distinguishing the ureter from the other organs. The experiments showed that the baseline U-Net model on the augmented dataset has a mean IoU score of 79.04%, while the proposed OrganFocusU-Net model achieved a mean IoU score of 79.52% on the augmented dataset, indicating its effectiveness in accurately distinguishing critical organs.

Keywords: Image Segmentation; Deep Learning; U-Net Architecture; Attention Mechanisms; Intersection over Union (IoU)

1. INTRODUCTION

Image segmentation means splitting an image into several segments or regions. It is an important subject in computer vision, which can be applied in object detection and medical imaging. In the dynamic and densely populated world, the prevalence of lifestyle-related diseases has increased significantly, disrupting normal human routines. This surge necessitates advanced approaches in medical image segmentation, particularly dealing with diverse tissues [1].

End-to-end learning with high-resolution medical images enhances segmentation accuracy, yet challenges arise from network depth, excessive parameters, and limited receptive fields in deep architectures. The absence of multi-scale contextual information further compromises segmentation performance due to variations in the sizes and shapes of regions of interest, so the incorporation and consolidation of multi-scale features become crucial for enhancing medical image segmentation performance [2].

The uterine arteries serve as the primary vessels responsible for providing blood to the uterus. These arteries emit branches that distribute blood to different segments of the uterus, playing a pivotal role in maintaining blood supply during physiological processes, such as the modifications in the endometrium throughout the menstrual cycle and the expansion of the uterus during pregnancy [3].

Convolutional neural networks (CNNs), notably the U-Net [4] and its variants, have emerged as leaders in this domain due to their exceptional performance. Characterized by 'U-shaped' architecture, these models consist of an encoder for global representation learning and a decoder for the gradual decoding of learned representations into pixel-wise segmentation. However, the limited encoding performance of CNN-based models stems from their localized receptive fields [5]. The encoder learns contextual features, reducing the resolution of medical images through convolution and pooling operations. Conversely, the decoder restores image resolution using an upsampling operation while enhancing abstract representation through convolution operations. Skip-connections in U-Net's architecture employ aggregation functions, either concatenation [6] or addition [7].

Despite widespread use in CNN methods for medical image segmentation [6,8], U-Net architectures suffer from an excessive number of parameters. Cascaded strategies on U-Net have been explored, particularly in brain tumor segmentation [9,10] addressing overlapping label challenges. Cascaded U-Net involves multiple encoder-decoder networks to address segmentation complexities, such as the work by Baid et al. [10], where the first network segments the whole tumor, and the second focuses on the tumor core and enhancing prediction of tumor. However, applying cascading U-Net to solve multi-class segmentation problems introduces complications.

Residual connections [11] in cascaded U-Net aim to prevent vanishing gradient issues in deeper networks. However, to address increased training parameters, researchers have replaced residual networks with dense connections, enabling short connections between layers and reducing parameters [12]. Dense connected U-Net models have been developed for multiple organ segmentation challenges [7,13].

Upon conducting a thorough examination of the available literature, several problems have been identified: A noteworthy concern observed in existing research work is the disregard for clinically attained databases. To increase the applicability of the proposed model, it is imperative to broaden the investigation by incorporating additional benchmark and clinically acquired datasets [14]. This inclusion ensures enabling the proposed model to adeptly handle a diverse array of real-world data. Many existing studies have predominantly concentrated on the classification process [15-17], with minimal attention directed toward segmentation. Only a few of studies have delved into the segmentation process [18-20]. However, the achieved accuracy rates in these studies are relatively modest, indicating the necessity for further enhancements in this domain [21].

Identifying these shortcomings in the existing literature underscores the need to address these limitations and enhance current approaches. The proposed model endeavors to overcome these challenges by incorporating clinically attained databases and prioritizing accurate and effective segmentation of laparoscopic surgery.

In this research, we introduce an approach to enhance semantic segmentation in laparoscopic surgery. Our contributions can be summarized as follows:

• **Custom Dataset Utilization:** Throughout our study, we leverage the UD Ureter-Uterine Artery-Nerve Dataset [22], a comprehensive collection of 586 high-resolution RGB images with corresponding masks. This dataset, meticulously annotated by gynecological experts, it forms the cornerstone of our experimentation. By training and evaluating models on this specific dataset, we ensure that our solutions are tailored to the intricacies of laparoscopic hysterectomy procedures.

• **Proposed Model (OrganFocusU-Net):** A proposed model, called OrganFocusU-Net, is developed, which combines the robust U-Net architecture with exploring attention mechanisms, to enhance the precision and accuracy of organ segmentation. This model specifically addresses the challenge of distinguishing critical organs during laparoscopic hysterectomy, showcasing the capability to intelligently focus on the most relevant regions within an image.

• **Evaluation of the performance of the baseline U-Net and the proposed OrganFocusU-Net model:** We evaluate and validate the segmentation models against ground truth masks, providing quantitative insights into their effectiveness.

The rest of the paper is organized as follows, Section 2 provides a review of related work; Section 3 presents the proposed model; Section 4 presents the experimental results and discussion; and finally, Section 5 presents the paper's conclusions and future work.

2. RELATED WORK

The limitations associated with manual and semi-automated techniques in biomedical segmentation have led to the introduction of fully automated approaches. Utilizing automated techniques for image segmentation has emerged as an alternative to manual processing, facilitating faster and more efficient patient examinations by healthcare professionals.

Fully convolutional networks (FCN) introduced by J. Long et al. in [23], and U-Net - introduced by Ronneberger et al. in [24] are based on a common key concept: skip connections. FCN sums upsampled feature maps with feature maps skipped from encoder while U-Net concatenates inserting convolutions and nonlinearities between each upsampling step. Following the intuition from DenseNet architecture Huang et al in [12] and Li et al. In [25], H-

denseunet for liver and liver tumor segmentation was introduced. Drozdzal et al. [26] systematically explored the importance of skip connections and introduced short skip connections within the encoder. Though there are minor architectural differences between the above approaches, they tend to fuse semantically dissimilar feature maps of encoder and decoder sub-networks, which compromise the segmentation performance based on experiments.

Other related models used for segmentation include GridNet, which is an encoder-decoder model where feature maps are set in a grid fashion. It does not have upsampling layers between the skip connections and also does not represent UNet++. Mask-RCNN is an important meta-framework that offers object detection, classification, and segmentation. The UNet++ could be fitted as its backbone architecture with just simple skip connections combined with nested dense skip pathways.

Another model called Pseudo-Mask Guided Feature Aggregation Network (PG-FANet) used in [29], it's a novel semi-supervised learning framework designed for histopathology image segmentation. It employs a two-stage sub-network architecture that aggregates multi-scale and multi-stage features, incorporating inter- and intra-uncertainty regularization to enhance prediction consistency in a teacher-student model.

To support this interest, the authors in [30] present a novel architecture for the diagnosis and segmentation of skin lesions using dermoscopy images named Cascade Knowledge Diffusion Network (CKDNet). It consists of three sub-networks: an initial coarse-level segmentation network, a classification network, and a fine-level segmentation network. It incorporates innovative feature entanglement modules (Entangle-Cls and Entangle-Seg) to facilitate knowledge diffusion between tasks, enhancing performance in both classification and segmentation.

A solution proposed in [31] presents a contour-aware network for 3D multi-organ segmentation in CT scans, specifically targeting abdominal organs. It employs two loss functions: Region Dice Loss for overall segmentation accuracy and Contour Cross Entropy Loss for precise boundary detection. The proposed method evaluated on the BVTAMOS dataset consisting of 110 annotated CT scans of 14 abdominal organs, achieves a Dice Similarity Coefficient (DSC) of 83.32% and a 95 %Hausdorff Distance (95HD) of 3.63 mm, outperforming several state-of-the-art techniques.

The work in [32] discusses the development and performance of a novel 3D medical image segmentation model called UNesT, which employs a hierarchical transformer-based approach to effectively capture local and global information in high-resolution medical images. UNesT has demonstrated state-of-the-art performance across multiple challenging datasets, including whole brain segmentation with 133 tissue classes and a newly created renal sub-structures CT dataset. The model outperforms previous methods, including an ensemble of models, by efficiently aggregating spatially adjacent patches and addressing the challenges of data inefficiency in medical imaging. Additionally, the authors emphasize the clinical utility of their work through accurate volumetric analysis and robust reproducibility, while also providing public access to their codes and trained models.

In the context of Diabetic Retinopathy (DR), an improved U-Net for segmentation has been proposed, replacing max-pooling functions with convolutional functions to retain multiple feature-related details [33]. The system demonstrated satisfactory outcomes in terms of Correspondence Ratio (CR) and Dice Similarity Coefficient (DSC) coefficients compared to FCN and Max-pooling U-Net models [19].

3. PROPOSED ORGANS SEGMENTATION FRAMEWORK

The detection of the three labeled organs: ureter, uterine artery, and nerves, or distinguishing the ureter from the other organs, is considered the main concern and challenge when the surgeons perform laparoscopic hysterectomy.

This paper proposes a framework for automatic organ segmentation and differentiation during laparoscopic hysterectomy using different semantic segmentation models.

Figure 1 shows the basic framework for organs semantic segmentation. This framework consists of three components: the dataset, data preprocessing, semantic segmentation model. These components are described below:

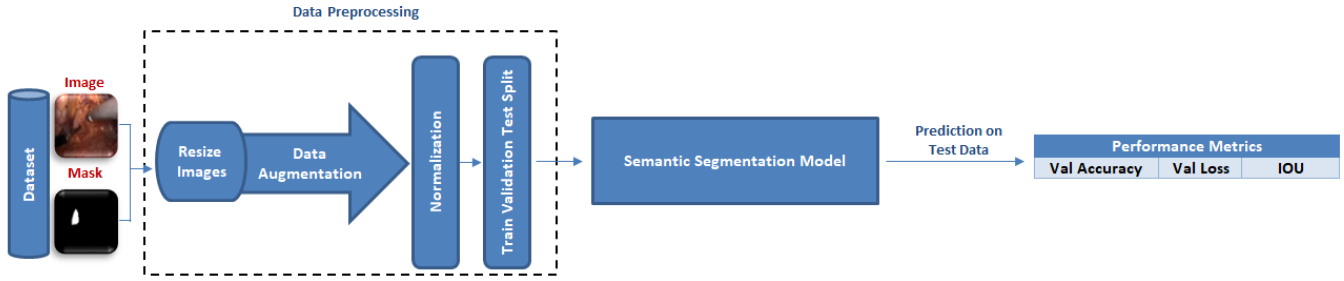


Figure 1. The basic framework for organs segmentation

3.1. Dataset

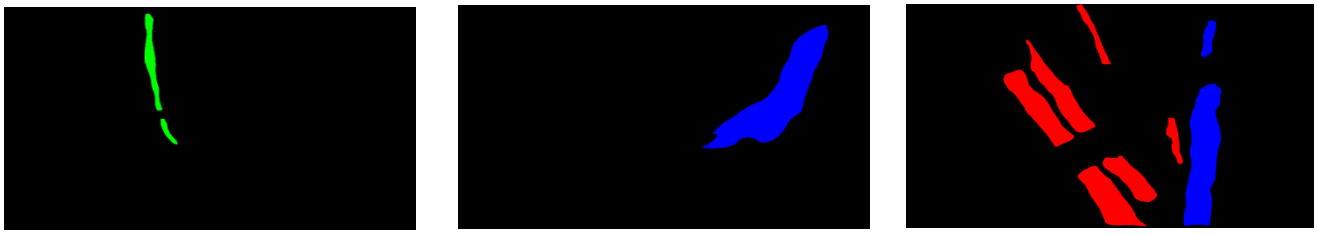
The dataset used in this work is the UD Ureter-Uterine Artery-Nerve dataset [22], which consists of 586 high-resolution RGB images obtained from 38 laparoscopic surgeries, The labels are as follows: 0 - background, 1 - uterine artery, 2 - ureter, 3 - nerve. These images are accompanied by corresponding masks designed for both binary and multiclass semantic segmentation. Gynecological experts from the University of Debrecen meticulously annotated this dataset.

The primary objective of the UD Ureter-Uterine Artery-Nerve dataset is to serve as a valuable resource for automatic organ segmentation. This dataset can be utilized to train semantic segmentation models, enabling pixel-wise detection of the three labeled organs: ureter, uterine artery, and nerves. Furthermore, the dataset is versatile enough to train models to distinguish the ureter from other organs, addressing a crucial concern and challenge faced by surgeons during laparoscopic hysterectomy.

3.2. Data preprocessing

The following section focuses on preparing the dataset for subsequent model training and evaluation. It involves organizing the dataset into **'images** and **masks'**, standardizing image sizes, encoding labels, and splitting data.

a. Organize the dataset into **'images** and **masks'** for structured access to ensure a consistent pairing of images and masks, a challenging step that we successfully addressed. This difficulty stemmed from the observation that all mask images are rendered as black images. So, to make sure that each image is accompanied by the proper mask, the pixel values in each mask were adjusted so that the mask becomes visible, as shown in Figure 2. But it should be noted that the presented work uses the original images and masks without pixel adjustment.



(a) A mask image that includes a nerve (shown in green)

(b) A mask image that includes a ureter (shown in blue)

(c) A mask image that includes a uterine artery (shown in red) and ureter (shown in blue)

Figure 2. Example of mask images after adjusting their pixels

b. Resize the images to 128 x 128 and standardize resolution using nearest-neighbor interpolation. This step ensures uniform input dimensions for the models, facilitating comparisons [34].

c. Data augmentation [35] is performed to expand the dataset and improve model robustness. the data augmentation technique was employed in this framework to generate additional images from the original dataset. After applying data augmentation, the number of images increased to 1,218, and the number of masks of each class before and after applying data augmentation is shown in Table 1.

Table 1. Number of masks of each class before and after applying data augmentation

Class Name	Before Data Augmentation	After Data Augmentation
Uterine artery	210 masks	429 masks
Ureter	254 masks	530 masks

Nerves	183 masks	388 masks

d. Normalize the images to ensure consistent intensity values, reducing the risk of model biases towards specific intensity ranges.

e. Encoding organ pixels to map organ label classes to numerical values, as shown in Table 2.

Table 2. Organ classes encoding

Class Name	Label (Pixel Value)
Background	0
Uterine artery	1
Ureter	2
Nerves	3

f. Data Split: To effectively train and evaluate the models, the dataset is split 80% for training, 10% for validation, and 10% for testing, and test sets, in a manner that maintains class balance.

3.3. Semantic segmentation model

In the semantic segmentation stage, we applied and tested two models: the baseline U-Net and the proposed model which is called OrganFocusU-Net.

3.3.1. Baseline U-Net model

The segmentation model used to perform pixel-wise detection and differentiation of organs is U-Net architecture, which is a widely used baseline model for semantic segmentation.

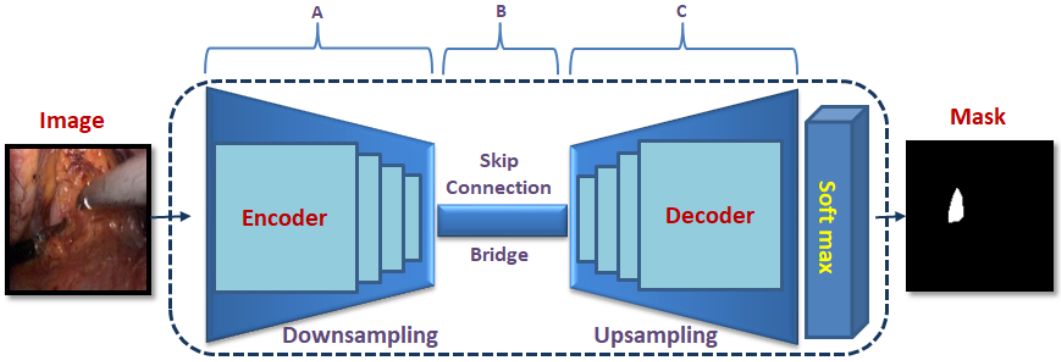


Figure 3. Architecture of baseline U-Net model

The U-Net architecture [4] is a CNN architecture that is commonly used for image segmentation tasks. As shown in Figure 3, it includes four parts:

1. Encoder Path: Part (A) of U-Net consists of a contracting path or an encoder. This path is responsible for capturing the context and extracting features from the input image. The encoder is composed of a series of convolutional layers followed by rectified linear unit (ReLU) activations and max-pooling operations. These operations progressively reduce the spatial dimensions of the input while increasing the number of feature channels.
2. Bottleneck: At part (B) of U-Net, there is a bridge layer that connects the encoder to the decoder. This bridge layer retains high-level abstract features learned by the encoder. Skip connections are employed, connecting corresponding layers from the encoder to the decoder. These skip connections help in preserving fine-grained details during the upsampling process.
3. Decoder Path: Part (C) of U-Net is an expansive path or a decoder. This path is responsible for upsampling the features to generate the final segmentation map. The decoder uses transposed convolutions (also known as deconvolutions or upsampling) to increase the spatial resolution of the feature maps. Each block in the decoder consists of transposed convolutions, followed by concatenation with the corresponding feature maps from the encoder, and then applying convolutional and activation layers.
4. Output Layer: The final part of the U-Net is a 1x1 convolutional layer with softmax activation function. This layer produces the segmentation map, where each pixel in the map represents the likelihood of belonging to the target class.

3.3.2. The proposed semantic segmentation model

The aim of the proposed semantic segmentation model, OrganFocusU-Net, is to enhance the precision and accuracy of organ segmentation in medical imaging by combining the proven power of the U-Net architecture with exploring attention mechanisms. This model specifically addresses the challenge of distinguishing critical organs during laparoscopic hysterectomy, showcasing the capability to intelligently focus on the most relevant regions within an image.

OrganFocusU-Net effectively captures intricate features and structures within images while addressing the challenge of distinguishing critical organs like the ureter, uterine artery, and nerves during laparoscopic hysterectomy. The OrganFocusU-Net architecture employs a series of convolutional and attention blocks that enable it to intelligently focus on the most relevant regions within an image. This architecture of the proposed model enables it to capture fine details and differentiate between organs, which makes it well-suited for applications where pixel-wise detection and differentiation of organs are paramount, revolutionizing the way complex surgical procedures are approached.

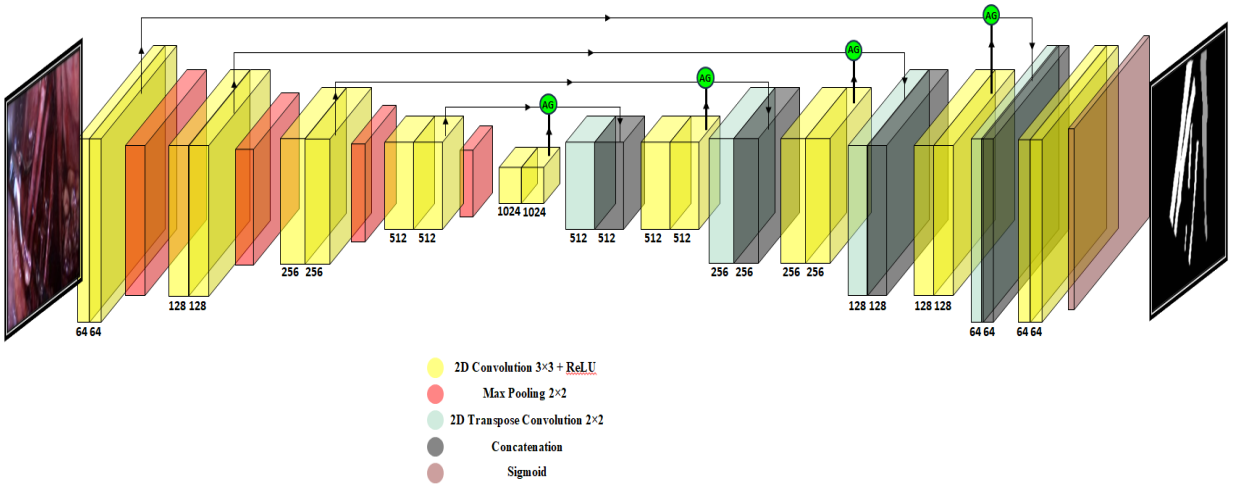


Figure 4. OrganFocusU-Net model architecture

The main components of OrganFocusU-Net architecture, shown in Figure 4, are described below:

1. Convolutional Block: The Conv Block function defines a pair of convolutional layers with batch normalization and ReLU activation. This block is responsible for capturing and processing features within the network.
2. Encoder Block: The Encoder Block function combines a convolutional block with max-pooling. It captures features and reduces spatial dimensions to create feature maps at different scales.
3. Attention Gate: The Attention Gate function calculates attention weights based on the feature maps from the encoder and decoder. It uses two convolutional layers and applies a sigmoid activation to produce attention masks.
4. Decoder Block: The Decoder Block function consists of an upsampling step, an attention gate, and concatenation. This block takes high-level features from the encoder and the attention-guided features from the decoder to refine the segmentation.

The OrganFocusU-Net model works as follows:

- It starts with an input layer followed by an encoder section. The encoder gradually reduces spatial dimensions while capturing important features.
- After the bottleneck, the decoder section brings the features back to the original resolution using the decoder blocks, where the attention gates refine the information at each stage.
- The output layer uses a 1x1 convolution with softmax activation function to produce a probability map for pixel-wise classification.

4. EXPERIMENTAL RESULTS AND DISCUSSION

This section presents the results of the experiments that we have conducted to evaluate the performance of the baseline U-Net model and the proposed OrganFocusU-Net model on the UD Ureter-Uterine Artery-Nerve Dataset. The presented models were trained and tested using python version (3.7.11). All experiments were conducted using

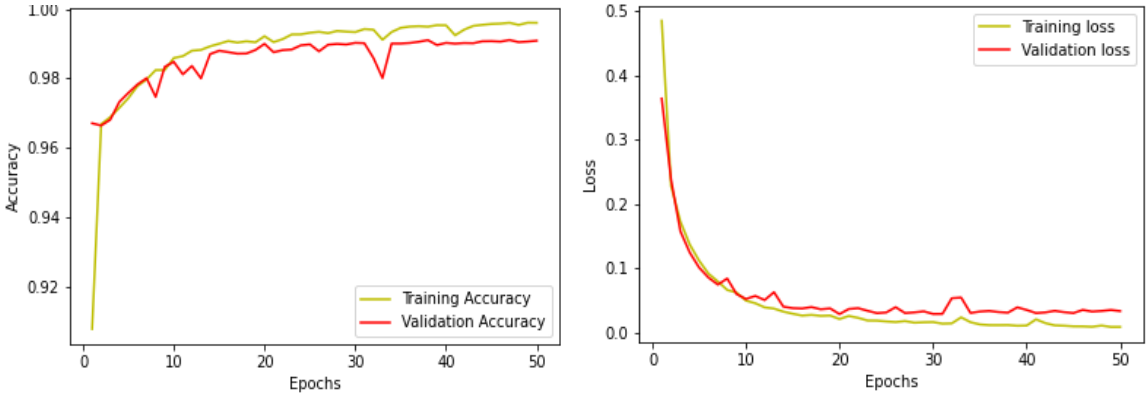
a single NVIDIA GeForce RTX 2060 GPU, Intel Core i7 10750H CPU and 32GB RAM. The models were trained for 50 epochs, which were enough for convergence with a learning rate (LR) of 0.0001 and a batch size of 4.

To evaluate the performance of the presented models, we calculate the Intersection over Union (IoU) score [36], which is a common metric that provides a quantitative measure of the model's segmentation accuracy. It is used to assess the overlap between predicted and ground truth segmentation masks, and is calculated as follows:

$$\text{IoU} = \frac{\text{Area of Overlap}}{\text{Area of Union}}$$

In addition, for each of the presented models, we will show the result of the model's segmentation and prediction on a sample image.

Figure 5 (a) presents the training and validation accuracies and Figure 5 (b) illustrates the training loss and validation loss for baseline U-Net. Figure 6 (a) and Figure 6 (b) show the training and validation accuracies and losses for OrganFocusU-Net model, respectively.



(a) (b)

Figure 5. (a) Training accuracy and validation accuracy for baseline U-Net model, (b)Training loss and validation loss for baseline U-Net Model

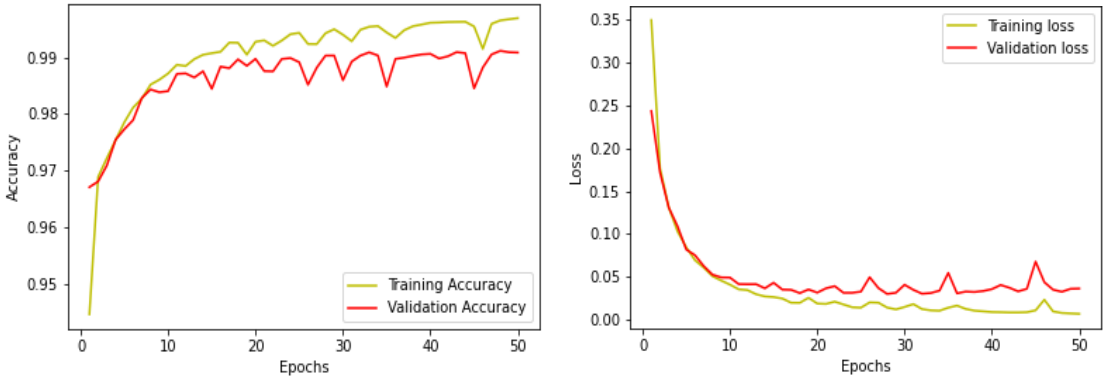


Figure 6. (a) Training accuracy and validation accuracy, (b) Training loss and validation loss for OrganFocusU-Net.

Table 3 presents the mean IoU score of the baseline U-Net model and proposed model which achieved 79.04% and 79.52% respectively, also in this table we displayed the class-wise results for two models, class-wise means the mean IoU for each class from four classes. Table 4 shows the classification report for baseline U-Net model and proposed model.

Table 3. Mean IoU class-wise scores

Class Name	Baseline U-Net mean IoU (%)	Proposed mean IoU (%)
Background	99.05	99
Uterine artery	72.92	72.5
Ureter	77.1	75.43
Nerves	67.08	71.03
Mean IoU	**79.04**	**79.52**

Both the baseline U-Net model and the proposed model achieved very high accuracy in segmenting the background class, with performance exceeding 99% across all metrics. However, as shown in Table 4, the main differences emerge in the segmentation of smaller anatomical structures such as the uterine artery, ureter, and nerves. For the uterine artery, the proposed model demonstrated higher precision (91.17% vs. 87.57%) but slightly lower recall (78.06% vs. 81.34%), leading to comparable F1-scores between the two models. In the case of the ureter, both models achieved similar balanced results with F1-scores of approximately 86%. For the nerve class, the proposed model yielded a clear improvement in recall (84.55% vs. 82.31%), which contributed to a higher overall F1-score (83.06% vs. 80.3%). Overall, these results highlight the ability of the proposed model to improve the trade-off between precision and recall, particularly for more challenging structures, while preserving high performance across all classes.

Table 4. Classification report for baseline U-Net model and proposed model

Model	Class	Precision (%)	Recall (%)	F1-score (%)
Baseline U-Net model	Background	99.34	99.71	99.52
	Uterine Artery	87.57	81.34	84.34
	Ureter	93.73	81.29	87.07
	Nerve	87.38	82.31	80.3
Proposed Model	Background	99.31	99.71	99.51
	Uterine Artery	91.17	78.06	84.11
	Ureter	91.17	81.39	86
	Nerve	81.63	84.55	83.06

The following Figure 7 depicts the model's segmentation and prediction on the sample test image using baseline U-Net model.

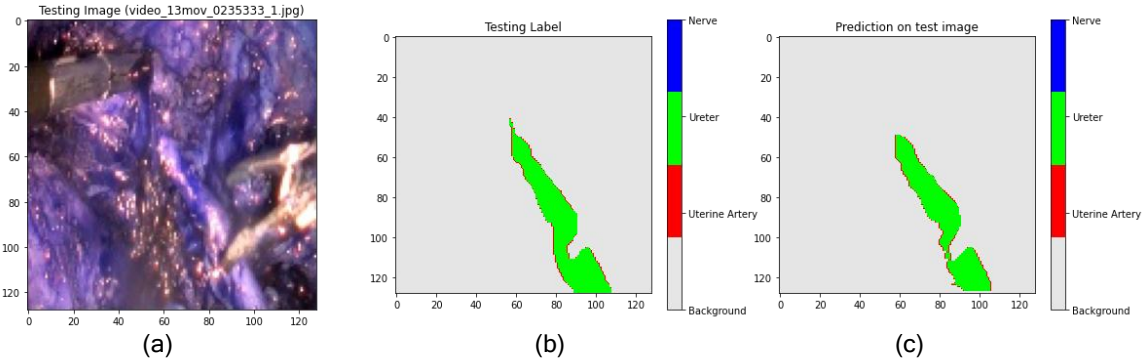


Figure 7. The baseline U-Net model prediction on the sample test image, (a) the original test image, (b) the ground truth of test image, (c) the prediction result

Figure 8 displays the OrganFocusU-Net model's segmentation and prediction on the same test image which is used in baseline U-Net.

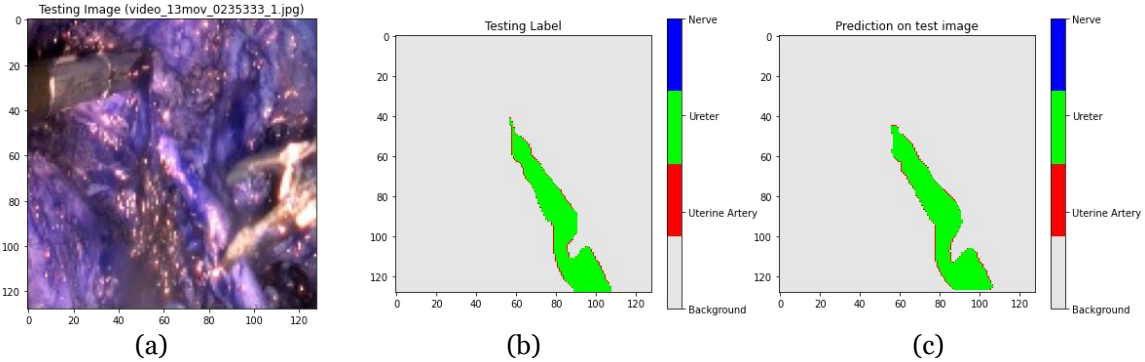


Figure 8. The proposed model prediction on the same test image, (a) the original test image, (b) the ground truth of test image, (c) the prediction result

Comparing the results in Figures 7 and 8, we can observe the enhanced segmentation capability achieved by combining the U-Net architecture with attention mechanisms, instead of using only the baseline U-Net model.

5. CONCLUSION AND FUTURE WORK

In this paper, we have proposed models for automatic organ segmentation and differentiation during laparoscopic hysterectomy using semantic segmentation models. Leveraging the UD Ureter-Uterine Artery-Nerve Dataset, which is capable of pixel-wise detection and differentiation of critical organs, namely the ureter, uterine artery, and nerves with a specific emphasis on accurately distinguishing the ureter from other organs.

Our investigations encompassed the implementation of the baseline U-Net model, and OrganFocusU-Net model, which combines the baseline U-Net with attention mechanisms, for enhanced precision in organ segmentation.

Experiments have been conducted to evaluate the performance of the baseline U-Net model and the proposed OrganFocusU-Net model on the UD Ureter-Uterine Artery-Nerve Dataset, it indicated that OrganFocusU-Net achieved a high mean IoU score on the augmented dataset, followed by the baseline U-Net, which signifies the capability of OrganFocusU-Net in accurately segmenting ureter in images.

While the proposed model represents a significant advancement in laparoscopic organ segmentation, ongoing research and development are essential for addressing the challenges and advancing the capabilities of these models for practical clinical applications.

REFRENCES

[1] W. Jakob, H.-C. Breit, M. T. Meyer, M. Pradella, D. Hinck, A. W. Sauter, T. Heye, D. Boll, J. Cyriac, S. Yang, M. Bach, and M. Segeroth, "TotalSegmentator: Robust Segmentation of 104 Anatomical Structures in CT Images," *Radiology: Artificial Intelligence*, 2022. doi:10.1148/ryai.230024.

[2] P. Ahmad, H. Jin, R. Alroobaea, S. Qamar, R. Zheng, F. Alnajjar, and F. Aboudi, "MH UNet: A Multi-Scale Hierarchical Based Architecture for Medical Image Segmentation," *IEEE Access*, vol. 9, pp. 148384–148408, 2021. doi:10.1109/ACCESS.2021.3122543.

[3] S. Hafez, "Comparative Placental Anatomy: Divergent Structures Serving a Common Purpose," *Progress in Molecular Biology and Translational Science*, vol. 145, pp. 1–28, 2017. doi:10.1016/bs.pmbts.2016.12.001.

[4] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, pp. 234–241, 2015. doi:10.1007/978-3-319-24574-4_28.

[5] H. Hu, Z. Zhang, Z. Xie, and S. Lin, "Local Relation Networks for Image Recognition," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 15-20 June 2019, pp. 3464–3473. doi:10.48550/arXiv.1904.11491.

[6] F. Isensee, P. Kickingereder, W. Wick, M. Bendszus, and K. H. Maier-Hein, "Brain Tumor Segmentation and Radiomics Survival Prediction: Contribution to the BRATS 2017 Challenge," *arXiv*, 2018. doi:10.48550/arXiv.1802.10508.

[7] M. Khened, V. A. Kollerathu, and G. Krishnamurthi, "Fully Convolutional Multi-Scale Residual Densenets for Cardiac Segmentation and Automated Cardiac Diagnosis Using Ensemble of Classifiers," *Med. Image Anal.*, vol. 51, pp. 21–45, Jan. 2019. doi:10.1016/j.media.2018.10.004.

[8] G. Wang, T. Song, Q. Dong, M. Cui, N. Huang, and S. Zhang, "Automatic Ischemic Stroke Lesion Segmentation from Computed Tomography Perfusion Images by Image Synthesis and Attention-Based Deep Neural Networks," 2020. doi:10.1016/j.media.2020.101787.

[9] M. Ghaffari, A. Sowmya, and R. Oliver, "Brain Tumour Segmentation Using Cascaded 3D Densely-Connected U-Net," arXiv, 2020. doi:10.48550/arXiv.2009.07563.

[10] U. Baid, N. A. Shah, and S. Talbar, "Brain Tumor Segmentation with Cascaded Deep Convolutional Neural Network," in *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*, A. Crimi and S. Bakas, Eds. Cham, Switzerland: Springer, 2020, pp. 90–98. doi:10.1007/978-3-030-46643-5_9.

[11] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 7-12 June 2015. doi:10.48550/arXiv.1512.03385.

[12] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, "Densely Connected Convolutional Networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Honolulu, HI, USA, 21-26 July 2017, pp. 2261–2269. doi:10.48550/arXiv.1608.06993.

[13] J. Dolz, I. B. Ayed, and C. Desrosiers, "Dense Multi-Path U-Net for Ischemic Stroke Lesion Segmentation in Multiple Image Modalities," arXiv, 2018. doi:10.48550/arXiv.1810.07003.

[14] N. Sambyal, P. Saini, R. Syal, and V. Gupta, "Modified U-Net Architecture for Semantic Segmentation of Diabetic Retinopathy Images," *Biocybernetics and Biomedical Engineering*, vol. 40, no. 3, pp. 1094–1109, 2020. doi:10.1016/j.bbe.2020.05.006.

[15] X. Li, X. Hu, L. Yu, L. Zhu, C.-W. Fu, and P.-A. Heng, "CANet: Cross-Disease Attention Network for Joint Diabetic Retinopathy and Diabetic Macular Edema Grading," *IEEE Transactions on Medical Imaging*, vol. 39, no. 5, pp. 1483–1493, May 2020. doi:10.1109/TMI.2019.2951844.

[16] R. K. Singh and R. Gorantla, "DMENet: Diabetic Macular Edema Diagnosis Using Hierarchical Ensemble of CNNs," *PLoS One*, vol. 15, no. 2, e0220677, 2020. doi:10.1371/journal.pone.0220677.

[17] L. Luo, D. Xue, and X. Feng, "Automatic Diabetic Retinopathy Grading via Self-Knowledge Distillation," *Electronics*, vol. 9, no. 9, p. 1337, 2020. doi:10.3390/electronics9091337.

[18] Y. Fu, J. Chen, J. Li, D. Pan, X. Yue, and Y. Zhu, "Optic Disc Segmentation by U-Net and Probability Bubble in Abnormal Fundus Images," *Pattern Recognition*, vol. 117, p. 107971, 2021. doi:10.1016/j.patcog.2021.107971.

[19] S. Kundu, V. Karale, G. Ghorai, G. Sarkar, S. Ghosh, and A. K. Dhara, "Nested U-Net for Segmentation of Red Lesions in Retinal Fundus Images and Sub-Image Classification for Removal of False Positives," *J. Digit. Imaging*, vol. 35, no. 5, pp. 1111–1119, 2022. doi:10.1007/s10278-022-00629-4.

[20] C. Kou, W. Li, W. Liang, Z. Yu, and J. Hao, "Microaneurysms Segmentation with a U-Net Based on Recurrent Residual Convolutional Neural Network," *J. Med. Imaging*, vol. 6, no. 2, 2019, Art. no. 025008. doi:10.1117/1.JMI.6.2.025008.

[21] A. Meshal and D. Gupta, "Segmentation of Diabetic Retinopathy Images Using Deep Feature Fused Residual with U-Net," *Alexandria Engineering Journal*, vol. 83, pp. 307-325, 2023. doi:10.1016/j.aej.2023.10.040.

[22] N. Norbert, "UD Ureter-Uterine Artery-Nerve Dataset," *IEEE Dataport*, July 10, 2023. doi:10.21227/q2dd-yt09.

[23] J. Long, E. Shelhamer, and T. Darrell, "Fully Convolutional Networks for Semantic Segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 7-12 June 2015, pp. 3431–3440. doi:10.1109/TPAMI.2016.2572683.

[24] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, pp. 234–241, 2015. doi:10.1007/978-3-319-24574-4_28.

[25] X. Li, H. Chen, X. Qi, Q. Dou, C.-W. Fu, and P.-A. Heng, "H-DenseUNet: Hybrid Densely Connected UNet for Liver and Liver Tumor Segmentation from CT Volumes," 2017. doi:10.1109/TMI.2018.2845918.

[26] M. Drozdzal, E. Vorontsov, G. Chartrand, S. Kadoury, and C. Pal, "The Importance of Skip Connections in Biomedical Image Segmentation," in *LABELS/DLMIA 2016: Lecture Notes in Computer Science*, G. Carneiro et al., Eds. Cham, Switzerland: Springer, 2016, vol. 10008, pp. 179–187. doi:10.1007/978-3-319-46976-8_19.

[27] D. Fourure, R. Emonet, E. Fromont, D. Muselet, A. Tremeau, and C. Wolf, "Residual Conv-Deconv Grid Network for Semantic Segmentation," arXiv, 2017. doi:10.48550/arXiv.1707.07958.

[28] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," in *2017 IEEE International Conference on Computer Vision (ICCV)*, Venice, Italy, 2017, pp. 2961–2969. doi:10.48550/arXiv.1703.06870.

[29] Jin Q, Cui H, Sun C, et al. "Inter- and Intra-Uncertainty Based Feature Aggregation Model for Semi-Supervised Histopathology Image Segmentation," Expert Systems with Applications, vol. 238, p. 122093, 2024. doi:10.1016/j.eswa.2023.122093.

[30] Jin Q, Cui H, Sun C, et al. "Cascade Knowledge Diffusion Network for Skin Lesion Diagnosis and Segmentation," Applied Soft Computing, vol. 99, p. 106881, 2021. doi:10.1016/j.asoc.2020.106881.

[31] Gao H, Lyu M, Zhao X, et al. "Contour-Aware Network with Class-Wise Convolutions for 3D Abdominal Multi-Organ Segmentation," Medical Image Analysis, vol. 87, p. 102838, 2023. doi:10.1016/j.media.2023.102838.

[32] Yu X, Yang Q, Zhou Y, et al. "UNesT: Local Spatial Representation Learning with Hierarchical Transformer for Efficient Medical Segmentation," Medical Image Analysis, vol. 90, p. 102939, 2023. doi:10.1016/j.media.2023.102939.

[33] D. Yang, G. Liu, M. Ren, B. Xu, and J. Wang, "A Multi-Scale Feature Fusion Method Based on U-Net for Retinal Vessel Segmentation," *Entropy*, vol. 22, no. 8, p. 811, 2020. doi:10.3390/e22080811.

[34] N. Jiang and L. Wang, "Quantum Image Scaling Using Nearest Neighbor," *Quantum Inf. Process.*, vol. 14, pp. 1559–1571, 2015. doi:10.1007/s11128-014-0841-8.

[35] E. Goceri, "Medical Image Data Augmentation: Techniques, Comparisons," *Artificial Intelligence Review*, vol. 56, pp. 12561–12605, 2023. doi:10.1007/s10462-023-10453-z.

[36] A. Simonelli, S. R. Bulo, L. Porzi, M. Lopez-Antequera, and P. Kontschieder, "Disentangling Monocular 3D Object Detection," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, 15-20 June 2019, pp. 1991-1999. doi:10.48550/arXiv.1905.12365.