

Real-Time Semantic Segmentation of the Ileocolic Pedicle from Laparoscopic Colectomy Images

Ziya Ata Yazıcı

Istanbul Technical University

Faculty of Computer and Informatics Engineering

Istanbul, Turkey

yaziciz21@itu.edu.tr

Abstract—Minimally Invasive Surgery (MIS) is the operation of using computer-aided systems to perform surgeries to increase the comfort of the patients and the recovery rate after the surgery. However, as a drawback, the visibility of the tissues becomes difficult for surgeons due to the limited space in the operation area. To this end, we investigate the Convolutional Neural Network-based segmentation models to perform real-time semantic segmentation of the Ileocolic Pedicle. We compose annotated dataset to perform the training of two state-of-the-art models, U-Net and Attention U-Net. Our results demonstrate that the trained models of U-Net achieved mIoU ranging between 0.948 - 0.957 and 0.748 - 0.957 mIoU for Attention U-Net. We further analyzed the inference time of the models and the implemented backbones. We observed that the inference time was ranging between 100 ms - 760 ms for U-Net, and 100 ms - 1600 ms for Attention U-Net. The achieved results indicate a promising real-time assistive system for surgeons to perform the operations.

Index Terms—Segmentation, medical image processing, convolutional neural networks, laparoscopy

I. INTRODUCTION

Colorectal cancer is the third most common cancer type worldwide with approximately 1.85 million new cases each year [1]. When it is required, the resection process of the colon (colectomy) can be performed both with open or minimally invasive surgery [2]. Colectomy is a surgical operation to treat colon diseases by removing a portion of the colon. The remaining parts are either attached or a new opening is created [3]. Minimally Invasive Surgery (MIS) is a surgical process in which medical instruments and an endoscopic camera is inserted into the patient's body through small cuts or orifices. Laparoscopy is a type of MIS method where surgeons observe the tissues and perform the surgery by controlling the robotic instruments through a computer interface [4]. Compared to conventional open surgery, this process increases the patient comfort and recovery rate [5].

Colectomy operation with laparoscopy is a difficult process that is performed by experienced surgeons only [6]. During the resection, the Ileocolic Pedicle - the group of Ileocolic Artery (ICA) and Ileocolic Vein (ICV) - should be properly identified to avoid hemorrhage – sudden blood loss; however, due to multiple factors of the scene (blood, fat tissue, lightning, etc.), this procedure become challenging [7].

Specialized deep-learning models are able to recognize and localize the patterns that are taught during the training phase.

The more specialized models are also capable of recognizing the tissues in a pixel-wise manner with semantic segmentation approaches. Semantic segmentation is a supervised learning method, in which the trained model predicts the class of each pixel in an image [7]. There have been multiple semantic segmentation models proposed for laparoscopy in colectomy. However, they mainly focus on the segmentation of the surgery instruments, and there is a lack of Ileocolic Pedicle, or more specifically, ICA and ICV segmentation studies.

In this study, collaborating with Istanbul University (Faculty of Medicine), we compose a dataset that includes 100 video frames of the colectomy operation of 2 patients. The pedicle regions of each selected video frame are annotated by the physicians, and the dataset will be used for semantic segmentation experiments. We propose benchmark results on our dataset by training and fine-tuning U-Net and Attention U-Net state-of-the-art transformer models to apply semantic segmentation of the Ileocolic Pedicle and provide a model that could be used in real-time operations. The models will be compared with multiple metrics such as Dice Coefficient, Intersection Over Union (Jaccard Index), Recall, Precision, F1-Score, and Inference Time. The details of the dataset and the models are given in the following subsections in detail. Dataset, Codes, and model weights can be accessed from the following GitHub page: <https://github.com/yaziciz/IleocolicPedicle>.

II. RELATED WORKS

In overall, most of the laparoscopic surgery segmentation studies have published their model results based on medical instrument segmentation, and popular challenges are generally hosted on this topic. However, due to the lack of tissue-based annotated datasets, semantic segmentation studies for tissue-specific model development – especially for colectomy surgery – are comparatively less in the literature [8].

Scheikl et al. [9] extract frames from the laparoscopy videos of the Endoscopic Vision Challenge 2019 [10] dataset and annotate them with respect to the organs and tissues in the scenes as a multi-label task. They choose multiple popular Convolutional Neural Network (CNN)-based segmentation models (U-Net [11], TernausNet [12], LinkNet [13], FCN [14], SegNet [15]) and their variations, such as with different backbones. Except for the U-Net model, they use ImageNet [16] pre-trained backbones. Their results show that LinkNet

variations with different ResNet [17] models achieve similar mean- IoU values. TernausNet with VGG-11 also results with a high mean intersection over union (IoU) result; however, other variations of TernausNet could not attain high mean-IoU, but high max-IoU. It is also indicated that the overall inference time is also low for a 25Hz laparoscope.

Kitaguchi et al. [18] use the images in LapSig300 [19] dataset, and annotates them concerning the inferior mesenchymal artery (IMA). The trains DeepLabV3+ with ResNet-256 backbone. They achieved 79.8% of the mean dice similarity coefficient (DSC) metric. According to their result, the trained model is capable of segmenting 12 FPS videos in real-time which is said to be considered an acceptable speed.

Kumazu et al. [20] annotate the connective tissue fibers gathered from a private video dataset of robot-assisted gastrectomy surgery. They train an original U-Net model and evaluate its performance with recall, F1/Dice coefficient, and questionaries to the surgeons. The model achieves a 54.9% mean F1/Dice value. According to the authors, the qualitative feedbacks of surgeons indicate that they are more convinced than the performance metrics.

Jha et al. [21] present benchmark results for real-time segmentation of the ROBUST-MIS challenge dataset by indicating their DSC and mIoU metric results. Additionally, an FPS metric is provided which decreases or increases depending on the inference time of the models. The experiments were applied for U-Net with ResNet-34 backbone, ResUNet [22], DeepLabV3+ [23] with MobileNetV2 [24] and ResNet-50 backbones, and ColonSegNet models. The results indicate that the ColonSegNet model could result in higher FPS values by sustaining its segmentation quality. Both DeepLabV3+ models could be used for the segmentation as they resulted in similar segmentation performance; however, the U-Net model achieves less DSC value with the same frame per second (FPS) as DeepLabV3+. The ResUNet model is comparably better than the U-Net model, but the FPS value is less than all the other models, which makes it not preferable for real-time applications.

III. DATASET

In the following sections, we present the details of the composed dataset and the used augmentations methods to increase the number of samples.

A. IleoPedSet

In our dataset - IleoPedSet (Table I, Figure 1) - we provide colectomy images that were recorded via laparoscopic tools for two patients. Each operation was recorded for 10 minutes with a resolution of 720 x 576 at 25 FPS. As the videos were observed, it was seen that the resection scene does not change strictly in a second, 25 frames. Therefore, it was agreed to select a single frame per second. From each patient, the surgeons hand-picked 50 frames that contains the ileocolic pedicle to be segmented, and the rest frames were skipped. The annotation procedure was applied via an open-source annotation tool, cvat.ai [25].



Fig. 1: Ileocolic pedicle annotated sample images from the IleoPedSet dataset

TABLE I: The IleoPedSet dataset

	# Samples	# Patients	Resolution
IleoPedSet	100	2	720 x 576

B. Data Pre-Processing

The dataset was split by 60%:40% as train and validation sets. We have created an additional test set by splitting the validation set in half. To further increase the size of the training set, we applied various positional augmentations.

Due to the limited number of samples in the training set, and to prevent overfitting [26], we applied augmentation methods to increase the 60 images to 300 frames. In Table II, we have indicated the performed augmentation methods by their parameters. The rotation is applied in a range of -5 to 5 degrees to represent the situations where the imaging tool records the scenes from different angles. We have applied the shear operation between -10 to 10 degrees to represent an alike condition as the rotation. The shift operation was applied both for the x and y axes by -10% to 10% of the image. The zoom operation was included to simulate the situations in which the position of the camera becomes too close or far from the tissue; therefore, it had a range of -30% to 30% of the image. Finally, a horizontal flip was randomly applied to the images. To comply with the models, all images were resized to 224 x 224 and the experiments were performed.

TABLE II: The applied augmentations

Augmentation	Parameter
Rotation	[-5, 5]
Shear	[-10, 10]
Width Shift	[-10%, 10%]
Height Shift	[-10%, 10%]
Zoom	[-30%, 30%]
Horizontal Flip	True

IV. METHODS

In the following sections, we detailed the models and the employed evaluation metrics which were used to assess the segmentation performances of the utilized models.

A. Architectures & Parameters

In this study, we trained two state-of-the-art U-Net [11] and Attention U-Net [27] models that perform semantic segmentation.

U-Net [11] is a convolutional neural network-based segmentation model with an encoder-decoder architecture. Each convolution operation in the encoder part consists of 3×3 kernels with unpadded inputs. The extracted feature matrices are feed-forwarded to a rectifier linear unit (ReLU), and down sampling is applied with 2×2 max-pooling layers with a stride of 2. In the decoder part, each layer doubles the resolution of the feature matrix by using up convolutions. At each layer, the feature matrix from the encoder part is concatenated with the up-sampled feature matrix. Therefore, the high-resolution features are sustained throughout the model, and the output mask becomes better localized.

Attention U-Net [27] is an updated version of the U-Net model, which implements attention gates on the decoder part of the model. These attention gates apply soft attention to the feature matrices by using the spatial features extracted from coarser levels. According to the paper, this improves the model's sensitivity to foreground pixels.

The learning rate was initially set to 0.001 and was updated based on the plateau in the validation loss. When necessary, the learning rate value was decreased by a factor of 0.8. The batch size was experimentally determined as 16. The models were trained for 200 epochs. The weights with the highest metric results were saved, and the experiments were performed on those weights. The Adam optimizer was preferred with $\beta_1 = 0.9$ and $\beta_2 = 0.999$. The L2 regularization term was used with a coefficient of 0.01 to discourage the model to overfit.

B. Metrics

In this study, we used mean Intersection over Union (mIoU) 1, Precision 2, recall 3, F1-Score 4, and mean Inference Time (mit) to evaluate the semantic segmentation performance of the model. In the scope of semantic segmentation; true positive (TP) is the correct classified positive pixels, false positive (FP) is the wrong classified negative pixels, true negative (TN) is the correct classified negative pixels and false negative (FN) is the wrong classified positive pixels. Referring to these terms, the following metrics can be represented as follows.

$$\text{mIoU} = \frac{1}{N} \sum_{i=1}^N \frac{TP_i}{TP_i + FP_i + FN_i} \quad (1)$$

$$\text{Precision} = \frac{TP}{TP + FP} \quad (2)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (3)$$

$$\text{F1-Score} = \frac{2TP}{2TP + FP + FN} \quad (4)$$

In order to optimize the model, we selected two loss functions that are popularly used in image segmentation literature, which are Dice Loss 5 and Jaccard Loss 6.

$$\text{Dice Loss} = 1 - \frac{2 \cdot \sum(p_{true} \cdot p_{pred})}{\sum p_{true}^2 + \sum p_{pred}^2} \quad (5)$$

$$\text{Jaccard Loss} = 1 - \frac{1}{N} \sum_{i=1}^N \frac{TP_i}{TP_i + FP_i + FN_i} \quad (6)$$

V. RESULTS

In this section, we will present the best achieved qualitative and quantitative semantic segmentation results of the U-Net and Attention U-Net models. As indicated in Table III, the best-performed model - the one with the highest mIoU and F1-Score results - has 0.957 mIoU, 0.909 precision, 0.982 recall, and 0.944 F1-Score results. The best model was achieved from the Attention U-Net architecture, it had a backbone of VGG-16, it was pre-trained on ImageNet, and Dice Loss was used. For the U-Net model, the mIoU values were ranging between 0.948 - 0.957 and 0.748 - 0.957 for the Attention U-Net model. The Attention U-Net performed poorly with non-pre-trained VGG-16 settings but reached the best score with the VGG-16, pre-trained Dice Loss settings. However, in general, the U-Net model performed consistently among all hyper-parameter settings compared to the Attention U-Net model.

As for the qualitative segmentation results, we investigate the segmentation performances of the best four models in Figure 2. Both U-Net and Attention U-Net models resulted in sufficient mask outputs to locate the tissue in the given image. The overall structure of the tissues was kept by all of the models, but the details were slightly covered more by the Attention U-Net model with VGG-16 backbone, pre-trained on ImageNet and Dice Loss (Figure 2, E). Since the mask results were satisfactory, we explored the quantitative results further to determine the best-performing model.

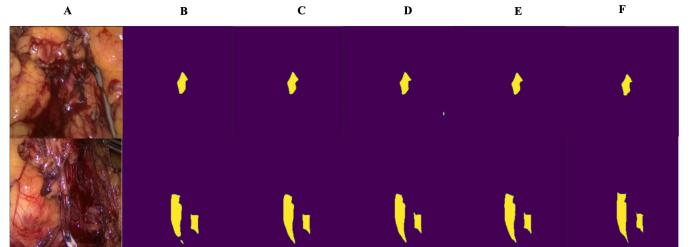


Fig. 2: The segmentation results of the selected best-performing models. A: Input image. B: U-Net, ResNet-50, Jaccard Loss. C: U-Net, ResNet-50, ImageNet, Jaccard Loss. D: Attention U-Net, VGG-16, ImageNet, Dice Loss. E: Attention U-Net, VGG-16, ImageNet, Jaccard Loss. F: Ground truth.

Referring to Table III, it can be observed that having a pre-trained model and fine-tuning the weights with a small dataset generally performed better compared to training from scratch. By comparing the Attention U-Net models that were trained with the Dice Loss and VGG-16 backbone, the fine-tuned model achieved comparably higher mIoU, Precision, Recall, and F1-Score than the model trained from scratch. Yet, we observed the opposite characteristics with the ResNet-50 versions. Training the both encoder and decoder with IleoPedSet reached better results.

We conducted the U-Net and Attention U-Net experiments for 200 epochs. By the end of the training, we observed an overall convergence to the local minima. However, in

TABLE III: The experiment results of the U-Net and Attention U-Net models

Model	Backbone	Pre Trained	Loss Function	mIoU	Precision	Recall	F1-Score
U-Net	VGG-16	-	DL / JL	0.955 / 0.950	0.858 / 0.884	0.991 / 0.985	0.920 / 0.932
	ResNet-50	ImageNet	DL / JL	0.956 / 0.948	0.900 / 0.885	0.985 / 0.988	0.941 / 0.934
Attention U-Net	VGG-16	-	DL / JL	0.955 / 0.957	0.908 / 0.890	0.983 / 0.984	0.944 / 0.940
	ResNet-50	ImageNet	DL / JL	0.951 / 0.957	0.896 / 0.896	0.985 / 0.986	0.938 / 0.939

models with ResNet-50 backbones, we noticed a stable plateau occurred for the gradients so that the validation loss becomes constant until the learning rate was reduced with the scheduler. The best four models' validation loss curves can be seen in Figure 3 and Figure 4 for U-Net and Attention U-Net, respectively.

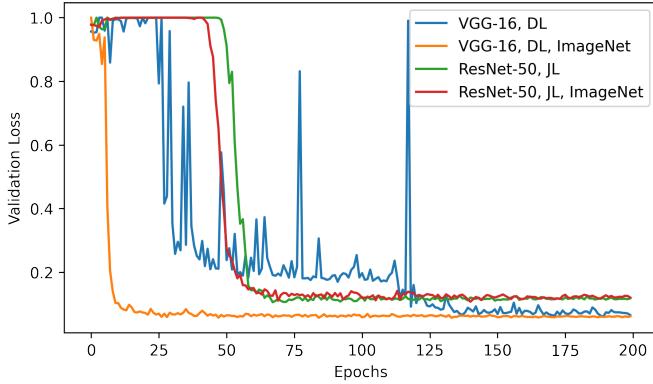


Fig. 3: The validation loss of the trained U-Net models over the epochs

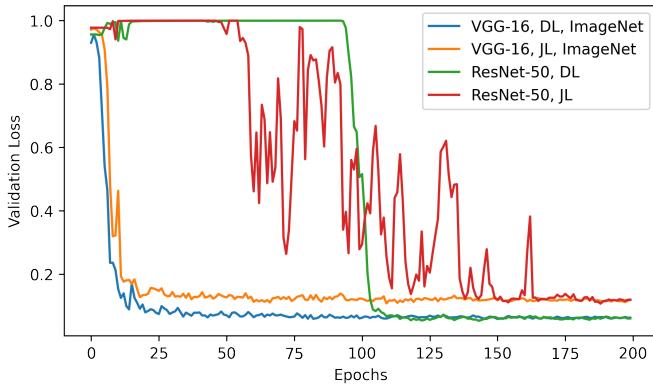


Fig. 4: The validation loss of the trained Attention U-Net models over the epochs

One of our main drives was to execute such segmentation with real-time performance. To further inspect the inference time of the models, we performed the test dataset's prediction

on two separate systems with Intel® Xeon® and Intel® Core™ i9 CPUs. The experiment was repeated for both VGG-16 and ResNet-50 backbones of the best-performed U-Net and Attention U-Net models. It was observed that (Figure 5) the U-Net model with VGG-16 and ResNet-50 backbones achieved 760 ms and 440 ms on Intel® Xeon®, 125 ms and 100 ms on Intel® Core™ i9 for the U-Net model, respectively. Attention U-Net was slower in inference time with 1600 ms and 1300 ms on Intel® Xeon®, 123 ms and 100 ms on Intel® Core™ i9, respectively. The best-achieved model, which was Attention U-Net, is capable of applying semantic segmentation at 10 FPS. However, depending on the CPU, the performance drastically decreases to 0.77 FPS. Yet, the U-Net model's performance changes between 10 FPS to 2.27 FPS.

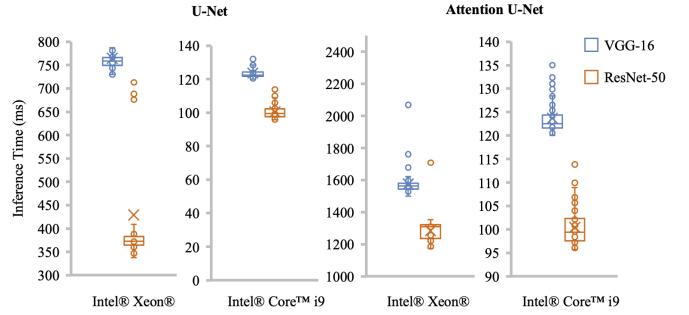


Fig. 5: The average inference time of the models on different CPUs

VI. DISCUSSION

During the colectomy operation, the proper and rapid segmentation of the ileocolic pedicle is important in the aspect of efficient surgery operations and supports the less experienced physicians in computer-assisted surgeries. Therefore, the model to be used in such systems should deliver high accuracy and performance. In our study, we observed that the Attention U-Net model with VGG-16 backbone, pre-trained on ImageNet and trained with Dice Loss had the highest segmentation accuracy among all method combinations, yet the difference between Attention U-Net and U-Net was not significant. Thus, the main metric to determine the most adequate model depends on the inference time. The architecture of the U-Net model makes it relatively lightweight than

the Attention U-Net model, thus the inference time becomes smaller. A similar characteristic can be observed in the scope of backbones. With the residual connections, the ResNet-50 model has fewer floating point operations (FLOP) than the VGG-16 model [17]. Therefore, the most efficient model could be considered as the U-Net model trained with ResNet-50 and trained from scratch with Dice Loss. Finally, it achieved 0.955 mIoU, 0.908 Precision, 0.983 Recall, and 0.944 F1-Score.

However, although 10 FPS could be a sufficient speed for a real-time surgery, it should be increased in case of a system that lacks a powerful CPU. To perform this, we should focus on more lightweight models by having a negative trade-off in segmentation accuracy. By having diverse trained models, the best-performing model can be determined by the surgeons in terms of inference time and segmentation success. Moreover, the IleoPedSet dataset should be increased in sample size and the distribution of the training and test sets should be done on a patient basis to observe the generalizability of the model utmost. In this study, although the train and test sets were randomly composed from the annotated images, the model might get biased from the patient-related details on the images to be segmented since both sets include the same patients.

VII. CONCLUSION

This study examines the real-time segmentation performances of the two state-of-the-art models, U-Net and Attention U-Net. It was observed that the U-Net model with ResNet-50 backbone and trained from scratch with Dice Loss provides the best performance with mIoU result of 0.955, 0.908 Precision, 0.983 Recall, and 0.944 F1-Score and 100 ms inference time, which is 10 frames per second. Although the Attention U-Net performed slightly better than the U-Net models, it suffered from real-time segmentation performance. During the surgery, the scene transition of the recorded tissue is rapid, and the model should be capable of annotating the tissue with a proper speed to not reduce the performance of the surgeon. Therefore, we will further investigate more lightweight model candidates, observe their semantic segmentation results, and determine the best-performing among them.

REFERENCES

- [1] C. Mattiuzzi, F. Sanchis-Gomar, and G. Lippi, “Concise update on colorectal cancer epidemiology,” *Annals of translational medicine*, vol. 7, no. 21, 2019.
- [2] N. Wasserberg *et al.*, “Laparoscopic colectomy for colorectal cancer,” *Isr Med Assoc J*, vol. 12, no. 9, pp. 572–576, 2010.
- [3] “Colectomy — johns hopkins medicine.”
- [4] A. Bartoli, T. Collins, N. Bourdel, and M. Canis, “Computer assisted minimally invasive surgery: is medical computer vision the answer to improving laparsurgery?,” *Medical hypotheses*, vol. 79, no. 6, pp. 858–863, 2012.
- [5] A. M. Lacy, J. C. García-Valdecasas, S. Delgado, A. Castells, P. Taurá, J. M. Piqué, and J. Visa, “Laparoscopy-assisted colectomy versus open colectomy for treatment of non-metastatic colon cancer: a randomised trial,” *The Lancet*, vol. 359, no. 9325, pp. 2224–2229, 2002.
- [6] A. Pigazzi, M. Hellan, D. R. Ewing, B. I. Paz, and G. H. Ballantyne, “Laparoscopic medial-to-lateral colon dissection: how and why,” *Journal of Gastrointestinal Surgery*, vol. 11, no. 6, pp. 778–782, 2007.
- [7] J. Kang, I.-k. Kim, S. I. Kang, S.-K. Sohn, and K. Y. Lee, “Laparoscopic right hemicolectomy with complete mesocolic excision,” *Surgical endoscopy*, vol. 28, no. 9, pp. 2747–2751, 2014.
- [8] “Endovis - grand challenge.”
- [9] P. M. Scheikl, S. Laschewski, A. Kisilenko, T. Davitashvili, B. Müller, M. Capek, B. P. Müller-Stich, M. Wagner, and F. Mathis-Ullrich, “Deep learning for semantic segmentation of organs and tissues in laparoscopic surgery,” *Current Directions in Biomedical Engineering*, vol. 6, no. 1, 2020.
- [10] T. Ross, A. Reinke, P. M. Full, M. Wagner, H. Kenngott, M. Apitz, H. Hempe, D. M. Filimon, P. Scholz, T. N. Tran, and et al., “Robust medical instrument segmentation challenge 2019,” May 2020.
- [11] 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*, pp. 234–241, Springer, 2015.
- [12] V. Iglovikov and A. Shvets, “Ternausnet: U-net with vgg11 encoder pre-trained on imagenet for image segmentation,” *arXiv preprint arXiv:1801.05746*, 2018.
- [13] A. Chaurasia and E. Culurciello, “Linknet: Exploiting encoder representations for efficient semantic segmentation,” in *2017 IEEE Visual Communications and Image Processing (VCIP)*, pp. 1–4, IEEE, 2017.
- [14] J. Long, E. Shelhamer, and T. Darrell, “Fully convolutional networks for semantic segmentation,” in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3431–3440, 2015.
- [15] V. Badrinarayanan, A. Kendall, and R. Cipolla, “Segnet: A deep convolutional encoder-decoder architecture for image segmentation,” *IEEE transactions on pattern analysis and machine intelligence*, vol. 39, no. 12, pp. 2481–2495, 2017.
- [16] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, “Imagenet: A large-scale hierarchical image database,” in *2009 IEEE conference on computer vision and pattern recognition*, pp. 248–255, Ieee, 2009.
- [17] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016.
- [18] D. Kitaguchi, N. Takeshita, H. Matsuzaki, T. Igaki, H. Hasegawa, S. Kojima, K. Mori, and M. Ito, “Real-time vascular anatomical image navigation for laparoscopic surgery: experimental study,” *Surgical Endoscopy*, vol. 36, no. 8, pp. 6105–6112, 2022.
- [19] D. Kitaguchi, N. Takeshita, H. Matsuzaki, T. Oda, M. Watanabe, K. Mori, E. Kobayashi, and M. Ito, “Automated laparoscopic colorectal surgery workflow recognition using artificial intelligence: experimental research,” *International Journal of Surgery*, vol. 79, pp. 88–94, 2020.
- [20] Y. Kumazu, N. Kobayashi, N. Kitamura, E. Rayan, P. Neculoiu, T. Misumi, Y. Hojo, T. Nakamura, T. Kumamoto, Y. Kurahashi, *et al.*, “Automated segmentation by deep learning of loose connective tissue fibers to define safe dissection planes in robot-assisted gastrectomy,” *Scientific Reports*, vol. 11, no. 1, pp. 1–10, 2021.
- [21] D. Jha, S. Ali, N. K. Tomar, H. D. Johansen, D. Johansen, J. Rittscher, M. A. Riegler, and P. Halvorsen, “Real-time polyp detection, localization and segmentation in colonoscopy using deep learning,” *Ieee Access*, vol. 9, pp. 40496–40510, 2021.
- [22] Z. Zhang, Q. Liu, and Y. Wang, “Road extraction by deep residual u-net,” *IEEE Geoscience and Remote Sensing Letters*, vol. 15, no. 5, pp. 749–753, 2018.
- [23] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, “Encoder-decoder with atrous separable convolution for semantic image segmentation,” in *Proceedings of the European conference on computer vision (ECCV)*, pp. 801–818, 2018.
- [24] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, “Mobilenetv2: Inverted residuals and linear bottlenecks,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 4510–4520, 2018.
- [25] CVAT.ai Corporation, “Computer Vision Annotation Tool (CVAT),” 9 2022.
- [26] P. Chlap, H. Min, N. Vandenberg, J. Dowling, L. Holloway, and A. Haworth, “A review of medical image data augmentation techniques for deep learning applications,” *Journal of Medical Imaging and Radiation Oncology*, vol. 65, no. 5, pp. 545–563, 2021.
- [27] O. Oktay, J. Schlemper, L. L. Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N. Y. Hammerla, B. Kainz, *et al.*, “Attention u-net: Learning where to look for the pancreas,” *arXiv preprint arXiv:1804.03999*, 2018.