Surgical Instrumentation Segmentation using U-Net with ResNet Backbone

Halil İbrahim Akgün

April 2025

1 Introduction

In this task, we aim to solve the problem of surgical instrumentation segmentation, which is a critical component in various computer-assisted interventions. The challenge involves processing RGB video frames and performing pixel-level semantic segmentation to generate masks for surgical instruments, including smaller and thin objects such as surgical clips, suturing threads, and needles.

2 Methodology

We employed a **U-Net** architecture [1] for the semantic segmentation task, as it is widely used for biomedical image segmentation tasks. The encoder is based on the ResNet-34 model [2], pretrained on ImageNet, which helps to improve feature extraction and model performance. The U-Net architecture was implemented using the *Segmentation Models PyTorch (SMP)* library [3], which provides an easy-to-use interface for building and training segmentation models. This library is particularly suitable for biomedical segmentation tasks, and it supports various backbones such as ResNet, EfficientNet, and VGG.

2.1 Model Architecture

The model consists of the U-Net architecture with the ResNet-34 encoder. The encoder, pretrained on ImageNet, extracts hierarchical features from the input RGB images, and the U-Net decoder reconstructs the segmented output. The final output consists of pixel-wise classification, generating masks for various surgical instruments.

2.2 Training and Evaluation

The model was trained on the provided training dataset using a batch size of 32, with an Adam optimizer and a learning rate of 1×10^{-4}. The loss function used was the Cross-Entropy loss, which is suitable for multi-class segmentation tasks. The training process was carried out for a maximum of 100 epochs, with early stopping triggered if the validation loss did not improve for 10 consecutive epochs.

The model's performance was evaluated using the Intersection over Union (IoU) metric, which measures the overlap between the predicted segmentation mask and the ground truth mask. The final model was selected based on the best validation loss. The trained model achieved a mean IoU of **0.72** on the test set, demonstrating effective segmentation performance across the surgical instrument classes. All the code is provided in [4].

3 Segmentation Masks Format

Segmentation masks are provided as `.png` images at the same resolution as the original video frames. The grayscale value of each pixel corresponds to a semantic class. There are 10 classes. The mapping between grayscale pixel values and semantic labels is as follows:

RGB Value	**Segmentation Class**
0	Background
1	Tool Clasper
2	Tool Wrist
3	Tool Shaft
4	Suturing Needle
5	Thread
6	Suction Tool
7	Needle Holder
8	Clamps
9	Catheter

Table 1: Mapping between grayscale pixel values and segmentation classes.

4 Results

The trained model achieved a mean IoU of 0.72 on the test set, demonstrating good performance in segmenting the surgical instruments. Visual results show that the model effectively identifies both large and small instruments, including suturing threads and surgical clips.

4.1 Visualization of Predictions

Figure 1-6 shows example predictions, where the input image is displayed along with the ground truth mask and the predicted mask. The model successfully segments the surgical instruments, including thin objects such as needles. The input image, ground truth mask, and predicted mask are shown side by side.

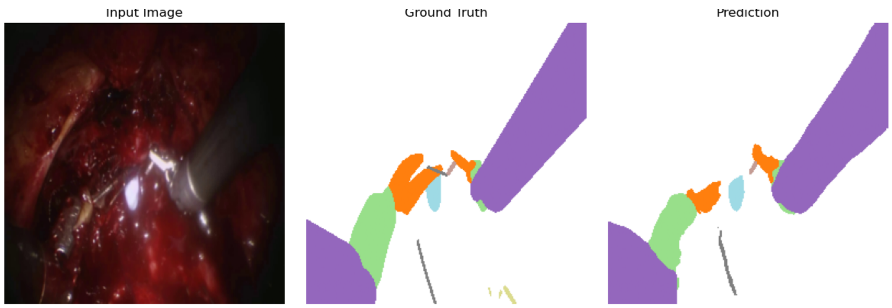

Figure 1: Example of model predictions.

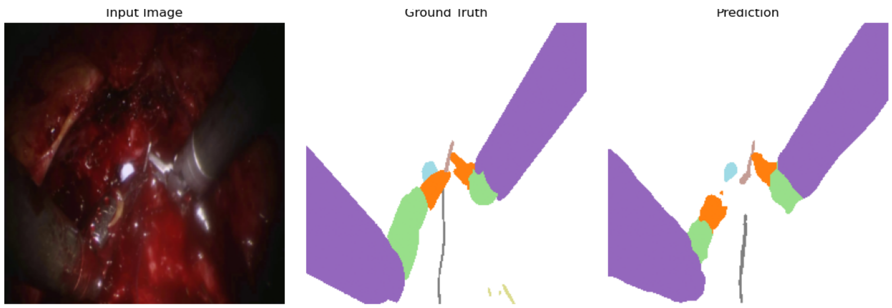

Figure 2: Example of model predictions.

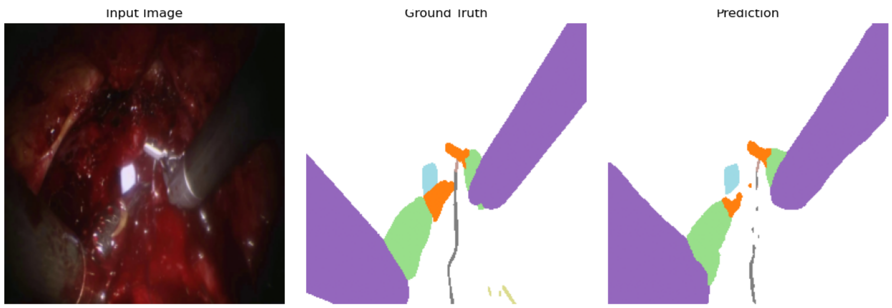

Figure 3: Example of model predictions.

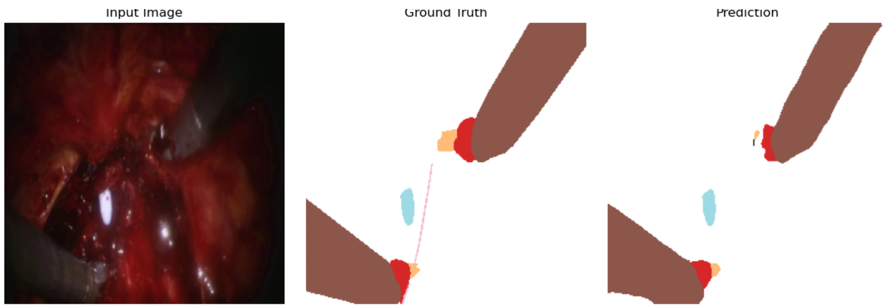

Figure 4: Example of model predictions.

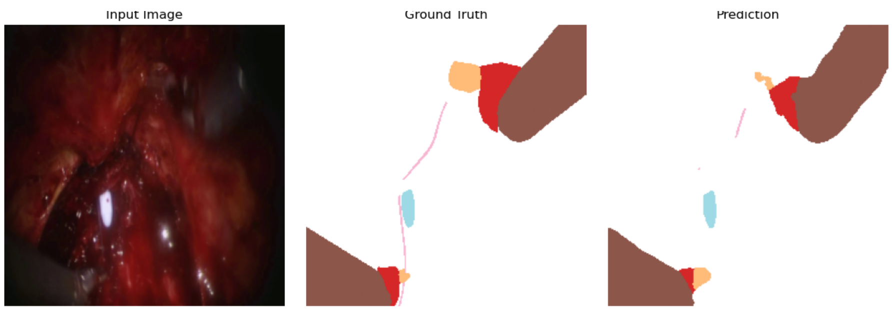

Figure 5: Example of model predictions.

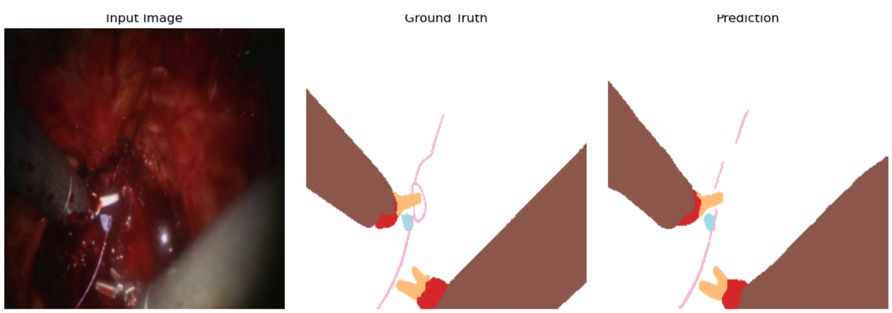

Figure 6: Example of model predictions.

5 Conclusion

The U-Net model with ResNet-34 encoder, implemented using the Segmentation Models PyTorch (SMP) library, provides effective segmentation of surgical instruments. Future work could involve exploring more complex architectures and backbones for improved accuracy. Also hyperparameter optimization will be done.

References

[1] O. Ronneberger, P. Fischer, and T. Brox, *U-net: Convolutional networks for biomedical image segmentation*, in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III*, 2015, pp. 234-241. Springer.

[2] 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*, 2016, pp. 770-778.

[3] P. Iakubovskii, *Segmentation Models Pytorch*, 2019. GitHub repository. [Online]. Available: `https://github.com/qubvel/segmentation_models.pytorch`.

[4] H. İbrahim, *Surgical Instrument Segmentation*, 2025. GitHub repository. [Online]. Available: `https://github.com/halilvibrahim/surgical-instrument-segmentation`.