Exploring Deep Learning for Medical Image Segmentation:

A Study on the Kvasir-Instrument Dataset

Kandidatnr:2016

***Abstract*—This research goes into the use of deep learning for medical instrument segmentation, with an emphasis on the UNet architecture. The Kvasir-Instrument dataset is used in the study, and data augmentation techniques are used to improve the model robustness. The findings provide light on UNet's ability to handle difficult medical imaging jobs.**

I. Introduction

With the use of deep learning techniques, the landscape of medical image analysis is experiencing a transformational phase. This study investigates the use of the U-Net architecture for medical instrument segmentation, providing approaches for accurate medical picture analysis [1].

II. Dataset Overview

This analysis is based on a thorough examination of the Kvasir-Instrument dataset. The resilience of the U-Net model is a result of the broad range of the dataset as well as the careful selection of training and test sets [2].

1) Training Set: The training set, which represents a wide range of medical instrument settings, is critical in creating the U-Net model. Extensive curation enables exposure to a wide range of challenges, developing flexibility.

2) Test Set: The test set serves as a vital assessment metric, putting the model to the test using real-world events and previously unknown data.

A. Discussion

The diversity of the training set is critical to the model's generalisation ability. It is critical to recognise potential biases and limits in capturing the full variety of real-world situations. The function of the test set in ensuring model generalisation demonstrates the need of testing performance in actual scenarios.

III. Model Architecture: U-Net

Key to this task is the U-Net architecture, which is well known for its abilities in semantic segmentation. When combined with skip connections, the encoder-decoder structure makes it easier to extract complex characteristics that are essential for accurate and precise medical instrument segmentation [1].

A. Training Odyssey: Optimizing Model Parameters

The U-Net model goes through a demanding 10-epoch training procedure. The Adam optimizer orchestrates the iterative process of updating model parameters, led by the Binary Cross-Entropy with Logits Loss function.

B. Discussion

The U-Net architecture was well-chosen for medical picture segmentation. The model's capacity to understand sophisticated patterns is highlighted by the training process's effectiveness, as evidenced by the decreasing trend in average loss. However, issues for computational resources and training time must be taken into account. It has taken the Model a total of 50 minutes to run all 10 epochs, using 5 minute for each epoch.

IV. Data Augmentation

Three data augmentation strategies were carefully used to reinforce the U-Net model against overfitting and improve its flexibility [1], [3].

1) Random Rotation: Randomly rotating images mimics variations in instrument orientations, ensuring the model is robust to different viewing angles.

2) Horizontal and Vertical Flipping: Horizontal and vertical flipping diversify the dataset, exposing the model to variations in instrument positions and orientations.

3) Transpose: Transpose operations enhance the model's resilience by handling variations in spatial orientation, mirroring potential real-world scenarios.

A. Discussion

Augmenting the data is a crucial part of making the model more resilient. Rotation and flipping are two of the selected approaches that successfully boost the dataset's variety. But it's important to recognise the possible restrictions and artifacts that augmentation may produce. Although there are many more data augmentation techniques available, selecting these three does not guarantee that the model will perform well for all augmentations or provide findings as accurate as it did.

V. Results and Discussion:Training UNetModel

Ten epochs are used in the UNetModel training procedure. As the average loss gradually drops, the model is improving its ability to forecast the segmentation of medical instruments [1], [4].

```
Epoch [1/10], Average Loss: 0.2884
Epoch [2/10], Average Loss: 0.1682
Epoch [3/10], Average Loss: 0.1488
Epoch [4/10], Average Loss: 0.1507
Epoch [5/10], Average Loss: 0.1495
Epoch [6/10], Average Loss: 0.1450
Epoch [7/10], Average Loss: 0.1358
Epoch [8/10], Average Loss: 0.1334
```

```
Epoch [9/10], Average Loss: 0.1338
Epoch [10/10], Average Loss: 0.1332
```

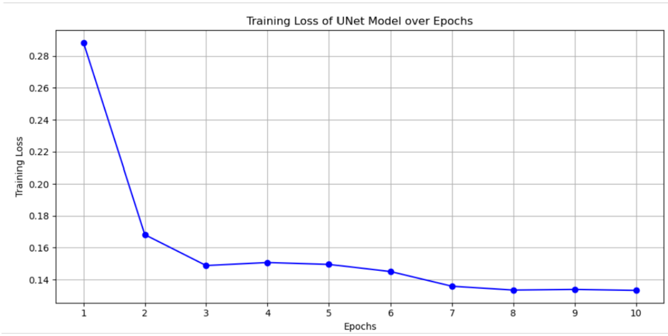


Fig. 1: Train Loss Visualisation

A. Testing UNetModel

UNetModel is evaluated on the evaluation dataset after the training phase. With an average test loss of 0.1123, the difference between the ground truth and anticipated masks is evident [1].

B. Discussion

The observed decrease in testing and training losses indicates that the U-Net architecture successfully learns to segment medical equipment when combined with the used data augmentation strategies. Promising results are shown in the model's ability to capture complex properties and generalise effectively to new data. The model is successful in this area because the final test loss of 0.1123 shows a fair degree of accuracy and demonstrates the model's capacity to provide correct predictions on new, never-before-seen data.

VI. Evaluation Metrics: Decoding Model Performance

A. Test Loss and Intersection over Union (IoU)

The difference between the ground truth and expected masks is measured by test loss. By quantifying the overlap between the real and anticipated segmentation masks, IoU sheds light on the accuracy of segmentation.

1) Data Augmentation Effects: One important method used to improve model generalisation is data augmentation, which exposes the model to different viewpoints on the training data. The goal of adding variations—like random flips, transpositions, and rotations—is to enhance the model's flexibility and replicate real-world situations, which is why we picked it for our case. Nonetheless, the following are some ways that the impacts of data augmentation may increase outcome variability:

a) Augmentation Magnitude: The degree to which data is augmented, such as the range of rotation angles or the extent of flipping, can impact the model's exposure to different variations. Small changes in augmentation parameters might lead to subtle alterations in the learned representations, influencing the final segmentation results.

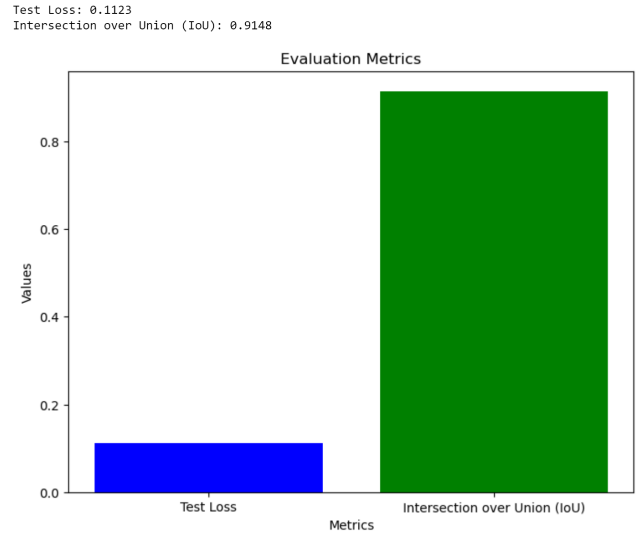


Fig. 2: Test Loss and Intersection over Union (IoU) with Visualisation

b) Augmentation Artifacts: While data augmentation is intended to enhance the model's ability to handle diverse scenarios, it may inadvertently introduce artifacts. For instance, aggressive rotations or flips could lead to unrealistic configurations that the model may struggle to generalise. The interplay between augmentation techniques and the dataset's intrinsic characteristics can introduce variations in model performance.

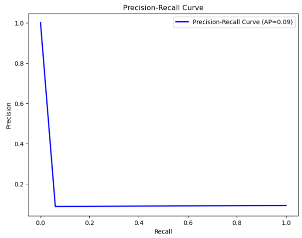
c) Overfitting and generalisation: By diversifying the training dataset, data augmentation is used to fight overfitting. The balance between augmentation and preventing overfitting, on the other hand, is tricky. If the augmentation is excessively forceful, the model may begin memorising enhanced samples from the training rather than learning the underlying patterns, thus affecting its generalisation to previously unknown data.

d) Randomness in Augmentation: The stochastic nature of certain augmentation operations, such as introducing random rotations or flips, can contribute to variability in model training. The randomness ensures that each epoch sees a slightly different version of the training data, introducing variability in the learned representations.

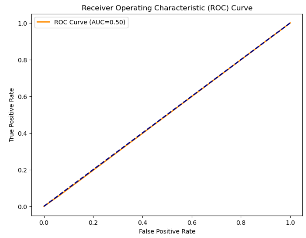
e) Discussion: Given these considerations, it's critical to strike and achieve a balance in augmentation techniques, ensuring they improve generalisation without generating artifacts or impairing the model's capacity to acquire and achieve meaningful patterns. The observed variability in results can be influenced by the nuanced interactions between data augmentation techniques, model architecture, and the inherent complexity of the medical instrument segmentation task.

B. Confusion Matrix, Average Precision, Area under ROC Curve, Dice Coefficient

These measures, which include segmentation accuracy overall, recall, and precision, provide complex insights into the functioning of the model.



(a) Precision-Recall Curve



(b) ROC Curve

VII. Visual Insights: Unraveling Model Predictions

A. Feature Map Exploration

Investigating feature maps at various layers sheds light on how well the model can identify and convey intricate patterns found in images obtained by medical instruments [5].

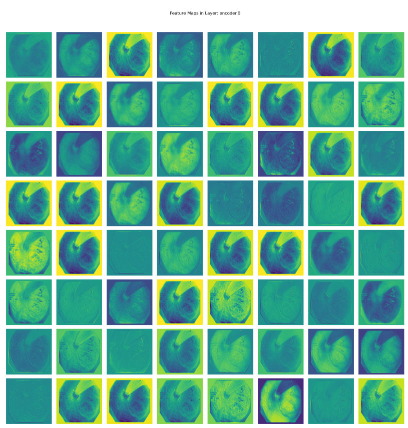


Fig. 4: Feature Map Exploration

B. Discussion

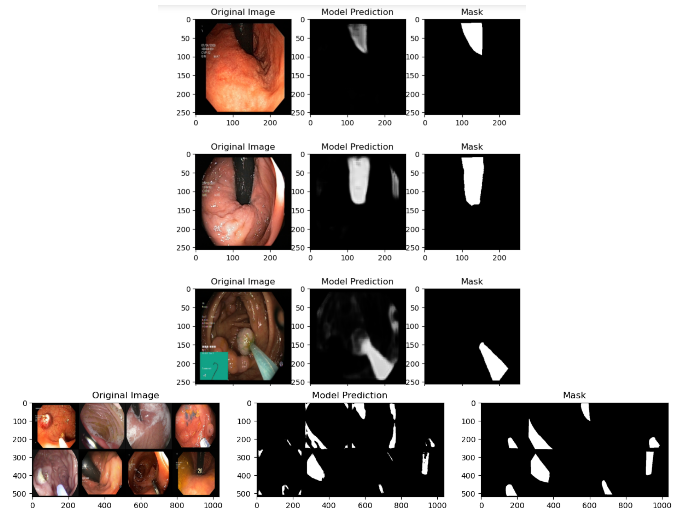
Exploration of feature maps enables a more in-depth knowledge of the hierarchical representation of medical instrument characteristics. Model predictions on the training dataset are visualised to demonstrate the usefulness of data augmentation in increasing model generalisation. The use of enhanced versions in the visualisation deepens the analysis.

C. Augmentation Impact Analysis

Analysing the effect of data augmentation on model predictions offers information on the strategies' efficacy in improving robustness and generalisation. It was able to provide a model forecast that was identical to the mask available for comparison.

VIII. Testing with Augmentation Visualisation

The testing loop evaluates the performance of the U-Net model on the evaluation dataset. It involves the following steps:



(a) Model Prediction Visualisation on Training Dataset

- The model is set to evaluation mode.
- For each batch of images and masks in the test dataset:
  - The original image undergoes processing by the model.
  - Augmented versions of the images are created using random rotations, flips, and transpositions.
  - The model predicts segmentation masks for both the original and augmented images.
  - The original image, its prediction, and the augmented images with their predictions are plotted for visualisation.
- The loop visualises predictions for three samples before exiting.

IX. Test Images and Augmented Version

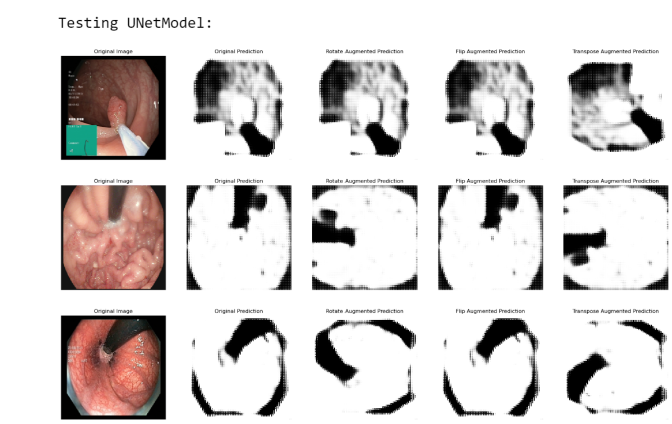


Fig. 6: Test Images and Augmented Version

a) Result: The results provide insights into how well the model generalises to different perspectives introduced by data augmentation, aiding in the assessment of its robustness and effectiveness on unseen data using the test image. this was just an extra to see how the test images would look in using the augmentation [1], [6].

X. Conclusion

This study demonstrates the efficacy of deep learning, specifically the U-Net architecture, in the our task of medical image segmentation using the Kvasir-Instrument dataset. The research highlights the importance of a diverse training set and the strategic application of data augmentation techniques to enhance the model's robustness.

The U-Net architecture proves to be well-suited for medical instrument segmentation, as evidenced by the model's effective learning of complex patterns during the 10-epoch training procedure. Despite computational challenges, the U-Net model showcases promising results in accurately segmenting medical instruments, as indicated by the decreasing trend in average loss.

Data augmentation strategies, including random rotation, horizontal and vertical flipping, and transpose operations, playing a crucial role in improving the model's resilience against overfitting. However, careful consideration of augmentation parameters and potential artifacts is essential to strike a balance and ensure meaningful pattern learning.

Test loss, Intersection over Union (IoU), confusion matrix, average accuracy, area under ROC curve, and Dice coefficient are among the evaluation metrics that give a thorough assessment of the model's performance. The observed results diversity highlights the intricate connections between data augmentation approaches, model design, and the intrinsic complexity of medical instrument segmentation.

Visual insights like feature map exploration and augmentation impact analysis help us understand how the algorithm detects complicated patterns in medical instrument pictures. The testing loop with visualisation of augmentation confirms the model's generalisation to multiple viewpoints introduced by data augmentation.

Finally, this study gives important insights into the use of deep learning for medical picture segmentation. The findings advance the field by highlighting the significance of careful dataset curation, model architecture selection, and augmentation technique in achieving accurate and being able to generalise outcomes in medical equipment segmentation tasks.

References

[1] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," *arXiv preprint arXiv:1505.04597*, 2015.

[2] Pogorelov, K. R. G. Konstantin Randel, and T. J. Carsten Eskeland, Sigrun Losada de Lange, "Kvasir: A multi-class image dataset for computer-aided gastrointestinal disease detection," *arXiv preprint arXiv:1904.02701*, 2019.

[3] C. Shorten and T. M. Khoshgoftaar, "A survey on image data augmentation for deep learning," *Journal of Big Data*, vol. 6, no. 1, p. 60, 2019.

[4] F. Milletari, N. Navab, and S.-A. Ahmadi, "V-net: Fully convolutional neural networks for volumetric medical image segmentation," in *2016 Fourth International Conference on 3D Vision (3DV)*, pp. 565–571, IEEE, 2016.

[5] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, and A. Torralba, "Learning deep features for discriminative localization," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2921–2929, 2016.

[6] L. Perez and J. Wang, "The effectiveness of data augmentation in image classification using deep learning," *arXiv preprint arXiv:1712.04621*, 2017.