Semantic Segmentation

Jahnavi Challa IMT2020103, **Tutor**: Neelam Sinha.

Report date: 19th May 2023.

Abstract

In this study, we focus on the task of semantic segmentation on the CholecSeg8k dataset, which contains annotated laparoscopic images of liver and gallbladder tissue. Our goal is to develop and evaluate a segmentation model that can accurately predict pixel-level labels for these tissue regions. We train our model using a deep neural network architecture and evaluate its performance by comparing its predicted masks with the ground truth annotations provided in the dataset. To assess the accuracy of our model, we use three common evaluation metrics: F1 score, Intersection over Union (IoU), and Dice coefficient. Our experimental results demonstrate that our model achieves high scores for F1, IoU, and Dice coefficient, indicating that it can accurately segment liver and gallbladder tissue in laparoscopic images from the CholecSeg8k dataset. Our study contributes to the development of computer vision algorithms for medical image analysis and has implications for improving the accuracy and efficiency of laparoscopic surgeries.

1 Introduction

In this study, we focus on the task of semantic segmentation of liver and gallbladder tissue in laparoscopic images using the CholecSeg8k dataset. The accurate segmentation of these tissue regions is critical for laparoscopic surgeries, which are minimally invasive procedures that rely on real-time visualization of internal organs. Our goal is to develop and evaluate a deep neural network architecture that can accurately predict pixel-level labels for liver and gallbladder tissue in laparoscopic images. We use three common evaluation metrics, F1 score, Intersection over Union (IoU), and Dice coefficient, to assess the performance of our model.

Our experimental results show that our segmentation model achieves high scores for all three-evaluation metrics, indicating its accuracy in segmenting liver and gallbladder tissue in laparoscopic images. The results of our study are significant as they contribute to the development of computer vision algorithms for medical image analysis, which has the potential to improve the accuracy and efficiency of laparoscopic surgeries. The accurate segmentation of liver and gallbladder tissue in laparoscopic images can aid surgeons in planning and performing surgeries with greater precision and efficiency. Overall, our study demonstrates the potential of deep neural network architectures for accurate semantic segmentation of liver and gallbladder tissue in laparoscopic images and highlights the importance of continued research in this field for improving medical image analysis and surgical outcomes.

2 Dataset

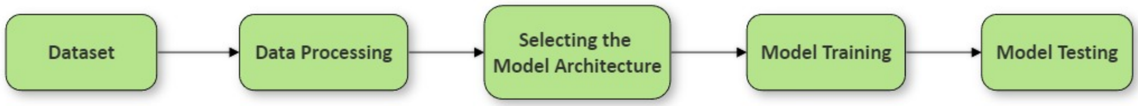
The CholecSeg8k dataset is a collection of annotated laparoscopic images of liver and gallbladder tissue, which has been made publicly available for research purposes. The dataset contains 8,000 images, each of size 576x720 pixels, captured during laparoscopic cholecystectomy surgeries.

Each image in the dataset is annotated with pixel-level labels for liver and gallbladder tissue regions. These annotations have been manually created by medical experts and are provided in the form of binary masks, where each pixel is assigned a label indicating whether it belongs to liver/gallbladder tissue or not.

The Cholec80k dataset includes surgeries performed by different surgeons, which introduces natural variability in techniques, skills, and approaches. This variability enhances the dataset’s representativeness and helps in evaluating the generalizability of algorithms across different surgical styles. The dataset presents various challenges commonly encountered in real surgical settings, such as variations in lighting conditions, tissue deformations, occlusions, and camera motion. These challenges make it a suitable benchmark for testing the robustness of surgical activity recognition algorithms.

3 Methodology

In this section, we describe the methodology employed for semantic segmentation of medical data, including data processing steps, selection of an appropriate model architecture, model training and finally evaluation of the model.



Flow Chat of the Methodology

1. **Data Processing:** We retrieve the file paths of the image and mask files using specific patterns (the image file ends with the name ’*_endo.png’ and the mask file ends with ’*_endo_watershed_mask.png’).

   Then we process each image file by opening it, converting it to grayscale, and resizing it to the target size. The resulting image is stored as a NumPy array in the images list. We preprocess the mask files in the same way as image files. Finally, the images and mask lists are converted to NumPy arrays.Then we add an extra dimension to the images array. This is typically done to match the expected input shape of the U-Net model. To ensure compatibility with the U-Net model, which expects input images to have a channel dimension, an additional dimension is added to the images array along the specified axis.

   Label encoding is performed because semantic segmentation models require numeric inputs rather than categorical labels. By converting the masks into numeric labels, it allows the model to understand and learn from the encoded information.

   Then, the mask array is reshaped to a 2D shape. This transformation converts the array to a single column, where each element represents a pixel value. We then apply Label encoding which assigns a unique integer label to each unique pixel value in the array. This step is typically done to convert categorical labels into a numeric representation that can be used for model training. We then reshape the mask array to its original shape to restore the shape or array.

We then normalize input images along the specified axis Which helps standardize the pixel values across the dataset. It brings the pixel values to a common scale and removes any potential biases that might arise due to variations in the input data.

2. **Model Architecture:** When considering semantic segmentation architectures for medical data, there are several options available. Some popular architectures include:

   (a) UNet

   (b) DeepLab

   (c) FCN (Fully Convolutional Network)

   (d) PSPNet (Pyramid Scene Parsing Network)

   While each of these architectures has its strengths, UNet stands out for medical semantic segmentation due to its effectiveness in capturing both local and global information.

   The UNet architecture is based on a fully convolutional neural network (FCN) and is particularly effective for pixel-wise segmentation tasks, where the goal is to classify each pixel in an image into different classes or segments.

   The key characteristic of the UNet architecture is its U-shape structure, which allows for both localization (capturing detailed information) and contextual information (capturing the global context) at different stages of the network. The architecture consists of two main parts: the **contracting path (encoder)** and the **expansive path (decoder)**.

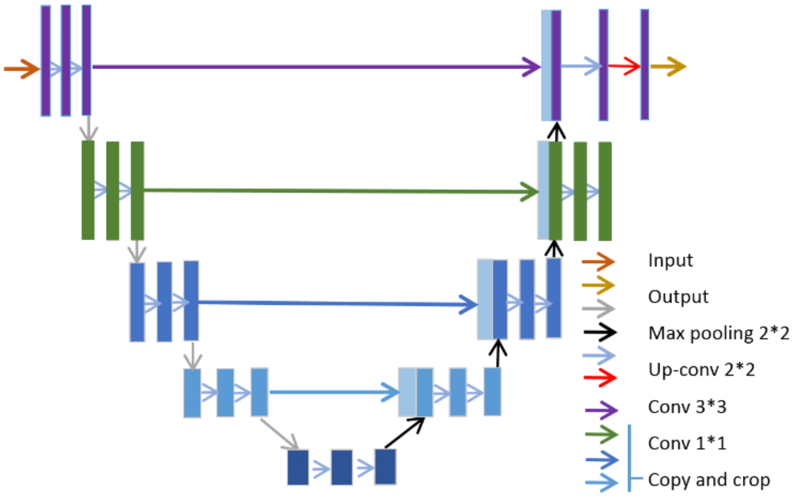


Figure 1: UNet Architecture

   Encoder path: In the contracting path or encoder of the UNet architecture, the input image is processed through multiple convolutional layers with pooling operations. The purpose of this path is to capture the context and extract high-level features from the input image.

   Convolutional layers increase the number of feature channels to learn complex representations. They apply filters to detect patterns. Pooling, like max pooling, reduces spatial resolution while preserving important features. It captures a larger receptive field and contextual information. The reduction in spatial resolution helps understand relationships in semantic segmentation. At the end of the contracting path, feature maps have reduced resolution but more channels. In the decoder, they are upsampled and combined for the final segmentation map.

Decoder path: The expansive path, also referred to as the decoder, is responsible for upsampling the feature maps from the contracting path and recovering the spatial resolution while reducing the number of feature channels. This path combines the low-level features from the contracting path with the upsampled features to provide detailed localization information.

The UNet decoder reconstructs lost spatial information using upsampling operations. It increases resolution while reducing feature channels for efficiency. **Skip connections** connect corresponding layers to access low-level features for localization. Feature maps from the contracting path and previous decoder layers are concatenated, combining local details and global context. This merger enables precise localization and accurate segmentation. The decoder outputs a segmentation map based on learned features from both paths.

Skip connections: The skip connections provide two main benefits:

First, they facilitate the propagation of detailed information from the contracting path to the expanding path, allowing the decoder to access the fine-grained spatial features. This enhances the network's ability to precisely localize object boundaries and capture small-scale structures.

Second, the skip connections aid in overcoming the information loss that may occur during the pooling and upsampling operations. Since the skip connections directly connect the layers, they provide a shortcut for the gradients to flow through the network during backpropagation. This helps in mitigating the vanishing gradient problem and ensures that the network can effectively learn and refine the segmentation results.

By combining both the local details from the skip connections and the global context captured by the expanding path, the UNet architecture can effectively leverage the complementary information and achieve accurate semantic segmentation. Skip connections play a crucial role in enhancing the localization capability of the network and improving its overall performance.

UNet is a popular choice for semantic segmentation for several reasons:

U-shape architecture: The UNet architecture, with its U-shaped structure, is specifically designed for semantic segmentation tasks. The contracting (encoder) and expanding (decoder) paths allow the network to capture both local details and global context effectively. The skip connections facilitate the combination of information from different scales, enabling precise localization of object boundaries.

High-resolution segmentation: UNet is capable of producing high-resolution segmentation maps. The skip connections help to recover spatial information, allowing the network to preserve fine-grained details and generate accurate pixel-wise predictions. This is particularly important for tasks where precise localization is crucial, such as medical image analysis.

Limited data requirements: UNet can achieve good segmentation performance even with limited training data. This is beneficial when labeled data is scarce or expensive to obtain, as is often the case in medical imaging. The architecture's ability to leverage both low-level and high-level features through skip connections helps in generalizing well from limited training samples.

Effective context capture: The contracting path in UNet captures contextual information by progressively reducing the spatial resolution and increasing the receptive field. This helps the network understand the relationships between different objects or regions in the image, leading to improved segmentation results.

Researchers have introduced modifications and improvements to the original architecture, such as adding residual connections, incorporating attention mechanisms, and using different types of convolutions, to further enhance its performance.

A few such variants would be:

UNet++, Inception UNet, Nested UNet, Dense UNet, Ensemble UNet, Attention UNet, Dilated UNet.

3. **Model Training:** Split the dataset into training and validation sets. Train the chosen model using the training set by feeding the input images and their corresponding masks. During training, optimize the model's parameters to minimize a defined loss function. The model is compiled with the Adam optimizer and categorical cross-entropy loss function and then Training data is fed to the model in mini-batches with a batch size of 16. The model is trained for 50 epochs. During the Training, progress and metrics such as loss and accuracy, are monitored.

4. **Model Evaluation:** Evaluate the trained model's performance using the validation set. Evaluation metrics include accuracy, F1 score, confusion matrix, and classification report.

   Dice Co-efficient: In the context of semantic segmentation of images, the Dice coefficient measures the overlap between the predicted segmentation and the ground truth segmentation.
   The formula for the Dice coefficient is:

   $$Dice = (2 * |X \cap Y|)/(|X| + |Y|)$$

   where X is the predicted segmentation, Y is the ground truth segmentation, $|X \cap Y|$ is the number of pixels where X and Y both have a positive value (i.e., both predict the same class) and $|X| + |Y|$ is the total number of pixels in both X and Y.

   IOU / Mean IOU: Similar to the Dice coefficient, IOU measures the overlap between the predicted segmentation and the ground truth segmentation.
   The formula for IOU is:

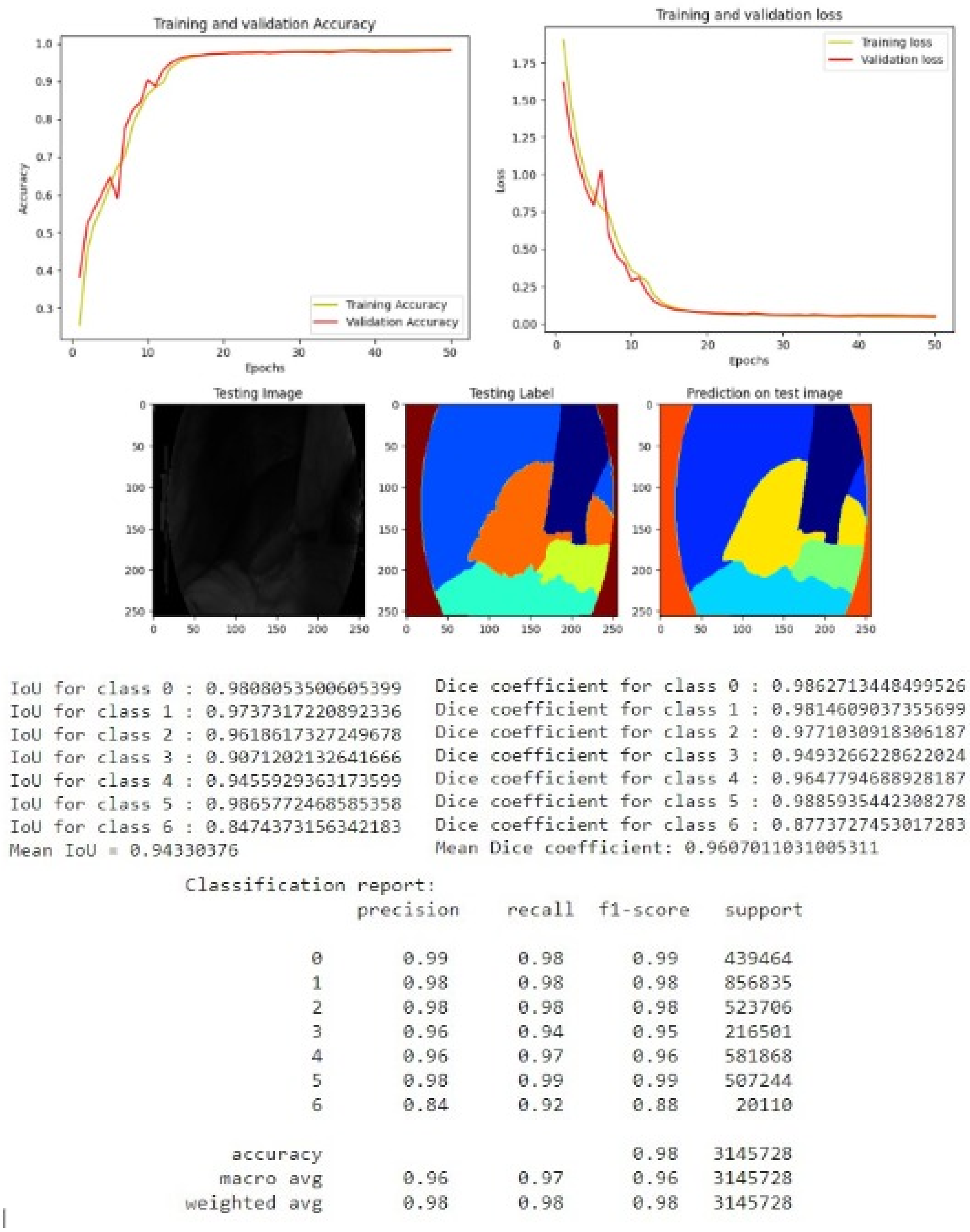
   $$IOU = |X \cap Y|/|X \cup Y|$$

   where X is the predicted segmentation, Y is the ground truth segmentation, $|X \cap Y|$ is the number of pixels where X and Y both have a positive value and $|X \cup Y|$ is the total number of pixels where X or Y have a positive value.

   One advantage of using IOU as an evaluation metric is that it is more interpretable than the Dice coefficient. It directly measures the ratio of the intersection of the predicted and ground truth segmentations to their union, which can be more easily understood and interpreted by researchers and practitioners.

4 Analysis and Results

1. UNet



```
IoU for class 0 : 0.9808053500605399      Dice coefficient for class 0 : 0.9862713448499526
IoU for class 1 : 0.9737317220892336      Dice coefficient for class 1 : 0.9814609037355699
IoU for class 2 : 0.9618617327249678      Dice coefficient for class 2 : 0.9771030918306187
IoU for class 3 : 0.9071202132641666      Dice coefficient for class 3 : 0.9493266228622024
IoU for class 4 : 0.9455929363173599      Dice coefficient for class 4 : 0.9647794688928187
IoU for class 5 : 0.9865772468585358      Dice coefficient for class 5 : 0.9885935442308278
IoU for class 6 : 0.8474373156342183      Dice coefficient for class 6 : 0.8773727453017283
Mean IoU = 0.94330376                     Mean Dice coefficient: 0.9607011031005311
```

```
Classification report:
              precision    recall  f1-score   support

           0       0.99      0.98      0.99    439464
           1       0.98      0.98      0.98    856835
           2       0.98      0.98      0.98    523706
           3       0.96      0.94      0.95    216501
           4       0.96      0.97      0.96    581868
           5       0.98      0.99      0.99    507244
           6       0.84      0.92      0.88     20110

    accuracy                           0.98   3145728
   macro avg       0.96      0.97      0.96   3145728
weighted avg       0.98      0.98      0.98   3145728
```

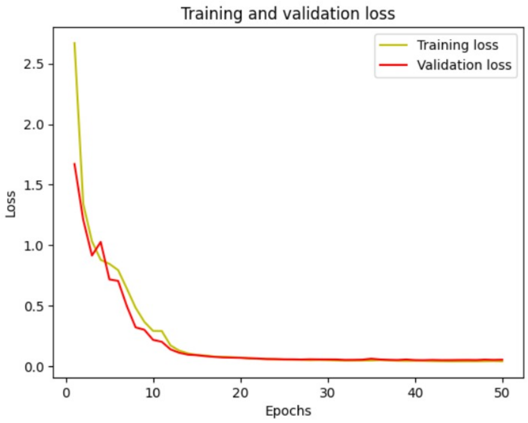
Figure 2: UNet Predictions

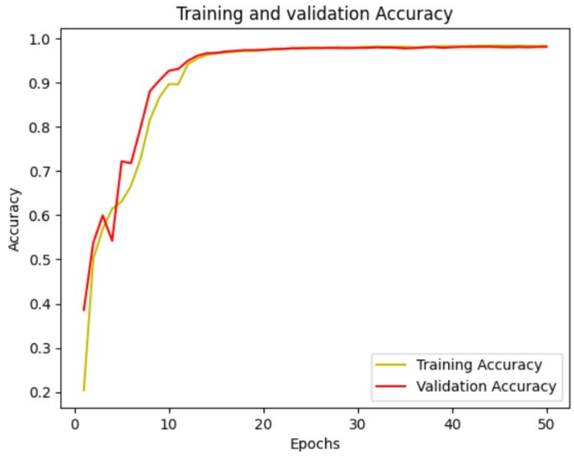
2. UNet++

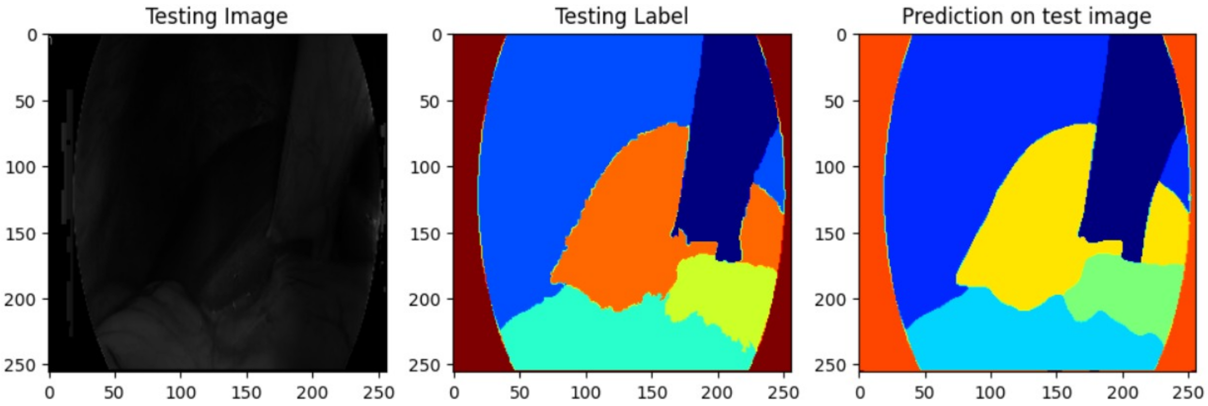
Here's how the additions in UNet++ help in achieving higher accuracy compared to UNet:

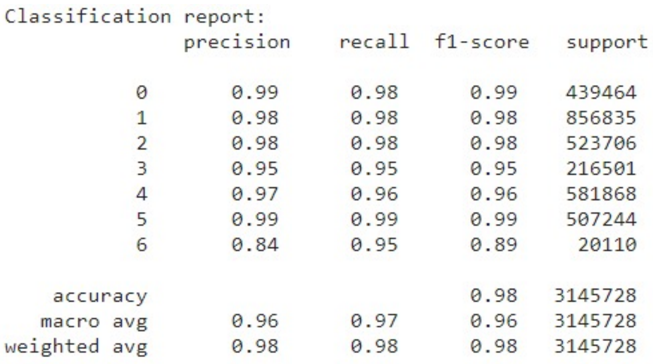
- **Nested skip connections:** Allow the network to capture features at different scales, incorporating both local details and global context for improved accuracy.

- **Deep supervision:** Produces predictions at multiple resolutions, guiding the learning process and enabling the model to refine segmentation at different levels of detail.
- **Dense skip connections:** Enable efficient information flow, reducing information loss and facilitating the capture of fine details for increased accuracy.
- **Deep aggregation module:** Aggregates features from different scales, enhancing the integration of multi-scale information and improving the representation of complex structures.
- **Improved feature fusion:** Combines local details and global context effectively, enabling a comprehensive understanding of the image and capturing both fine-grained and high-level information for precise segmentation.





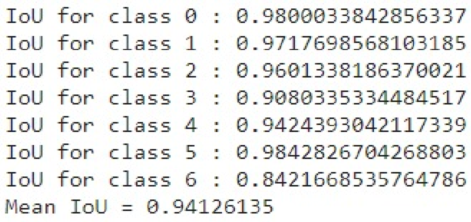




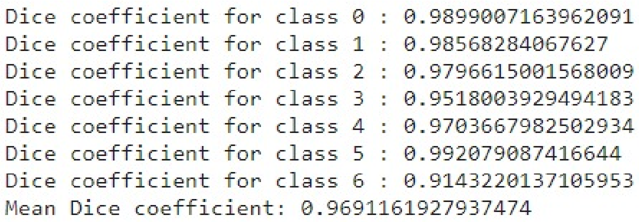
```
Classification report:
              precision    recall  f1-score   support

           0       0.99      0.98      0.99    439464
           1       0.98      0.98      0.98    856835
           2       0.98      0.98      0.98    523706
           3       0.95      0.95      0.95    216501
           4       0.97      0.96      0.96    581868
           5       0.99      0.99      0.99    507244
           6       0.84      0.95      0.89     20110

    accuracy                           0.98   3145728
   macro avg       0.96      0.97      0.96   3145728
weighted avg       0.98      0.98      0.98   3145728
```



```
IoU for class 0 : 0.9800033842856337
IoU for class 1 : 0.9717698568103185
IoU for class 2 : 0.9601338186370021
IoU for class 3 : 0.9080335334484517
IoU for class 4 : 0.9424393042117339
IoU for class 5 : 0.9842826704268803
IoU for class 6 : 0.8421668535764786
Mean IoU = 0.94126135
```



```
Dice coefficient for class 0 : 0.9899007163962091
Dice coefficient for class 1 : 0.98568284067627
Dice coefficient for class 2 : 0.9796615001568009
Dice coefficient for class 3 : 0.9518003929494183
Dice coefficient for class 4 : 0.9703667982502934
Dice coefficient for class 5 : 0.992079087416644
Dice coefficient for class 6 : 0.9143220137105953
Mean Dice coefficient: 0.9691161927937474
```

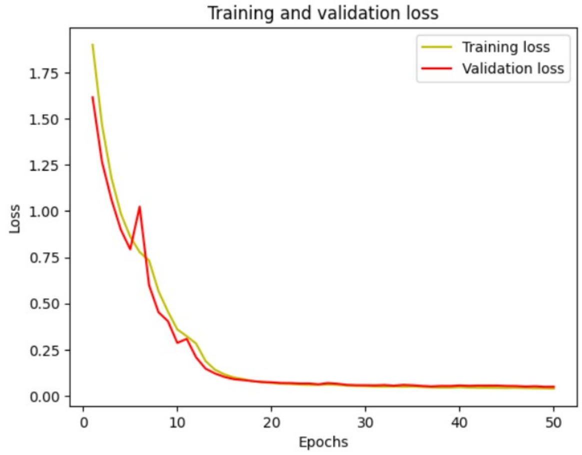
3. Nested UNet

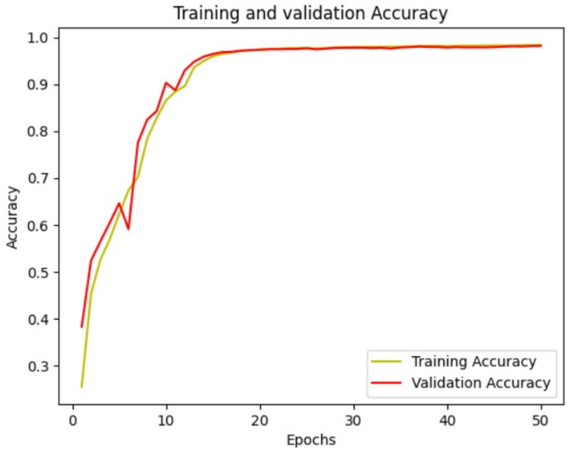
Nested UNet outperforms UNet in semantic segmentation due to its unique architecture. The addition of densely connected skip pathways in Nested UNet enhances the feature extraction process, allowing for the capture of more detailed and complex information from medical images. These skip connections facilitate the efficient flow of information between different levels of the network, enabling the model to leverage multi-scale features effectively.

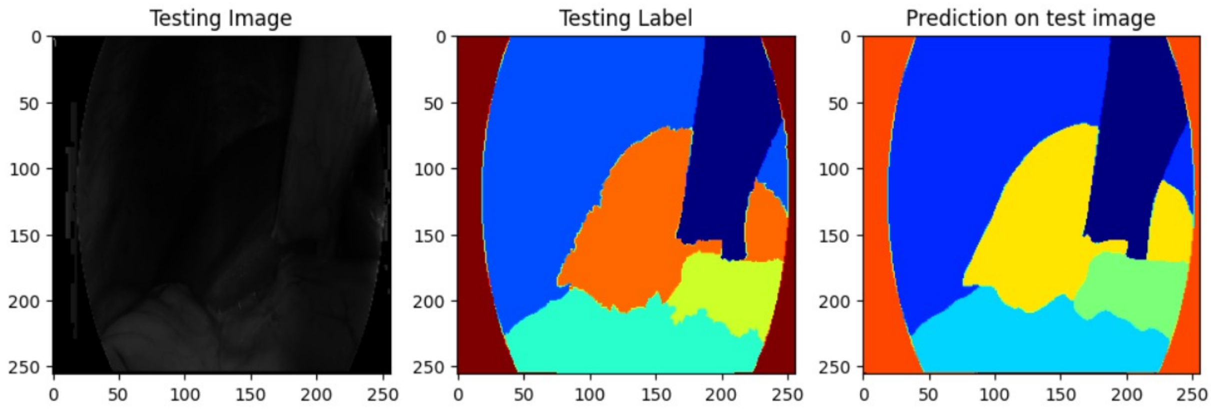
By incorporating dense skip connections, Nested UNet can access features from multiple levels of the contracting path. This enables the model to capture a broader context and utilize more contextual information, leading to improved segmentation accuracy. The dense connectivity helps in preserving fine details and capturing global spatial relationships, which are crucial for accurately delineating object boundaries in medical images.

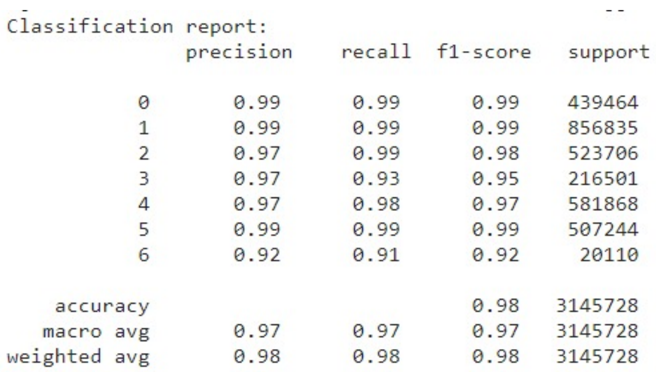
Nested UNet also benefits from its hierarchical representation learning. The architecture enables the model to learn representations at multiple scales, starting from shallow levels and gradually progressing to deeper levels. This hierarchical learning process allows the model to capture both local details and global context, improving its understanding of the image semantics and enhancing segmentation accuracy.

In summary, the densely connected skip pathways, improved feature extraction, enhanced context modeling, handling of complex structures, and hierarchical representation learning contribute to the superior performance of Nested UNet compared to UNet in semantic segmentation tasks for medical images.





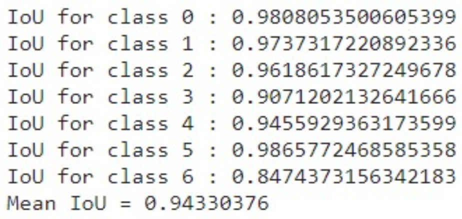




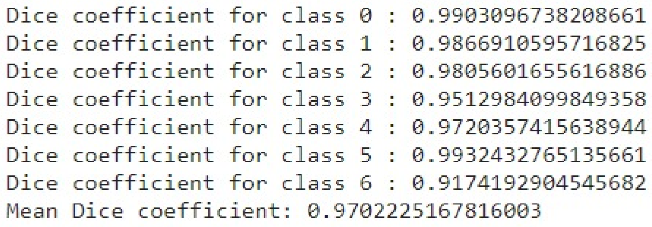
```
Classification report:
              precision    recall  f1-score   support

           0       0.99      0.99      0.99    439464
           1       0.99      0.99      0.99    856835
           2       0.97      0.99      0.98    523706
           3       0.97      0.93      0.95    216501
           4       0.97      0.98      0.97    581868
           5       0.99      0.99      0.99    507244
           6       0.92      0.91      0.92     20110

    accuracy                           0.98   3145728
   macro avg       0.97      0.97      0.97   3145728
weighted avg       0.98      0.98      0.98   3145728
```



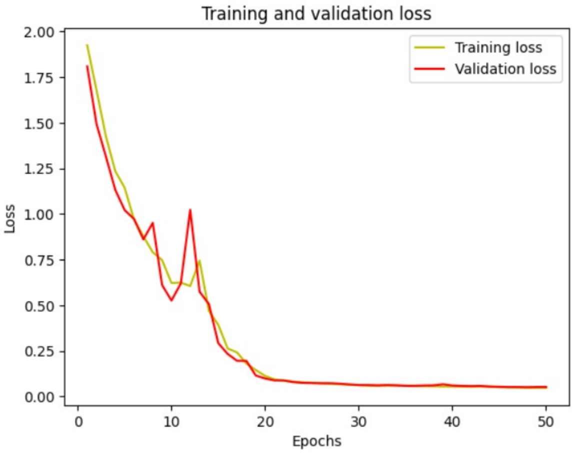
```
IoU for class 0 : 0.9808053500605399
IoU for class 1 : 0.9737317220892336
IoU for class 2 : 0.9618617327249678
IoU for class 3 : 0.9071202132641666
IoU for class 4 : 0.9455929363173599
IoU for class 5 : 0.9865772468585358
IoU for class 6 : 0.8474373156342183
Mean IoU = 0.94330376
```

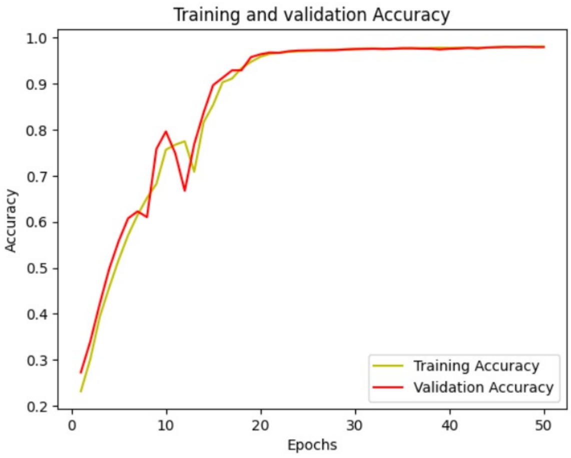


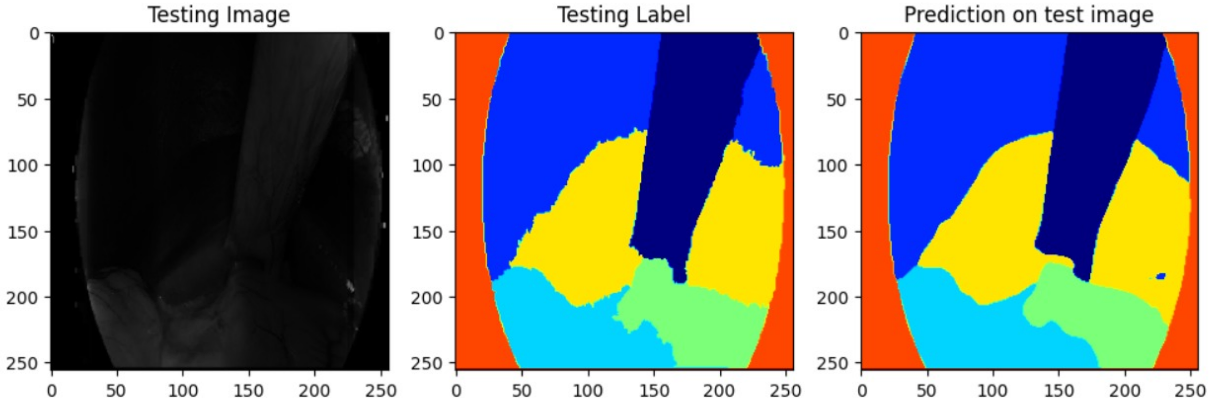
```
Dice coefficient for class 0 : 0.9903096738208661
Dice coefficient for class 1 : 0.9866910595716825
Dice coefficient for class 2 : 0.9805601655616886
Dice coefficient for class 3 : 0.9512984099849358
Dice coefficient for class 4 : 0.9720357415638944
Dice coefficient for class 5 : 0.9932432765135661
Dice coefficient for class 6 : 0.9174192904545682
Mean Dice coefficient: 0.9702225167816003
```

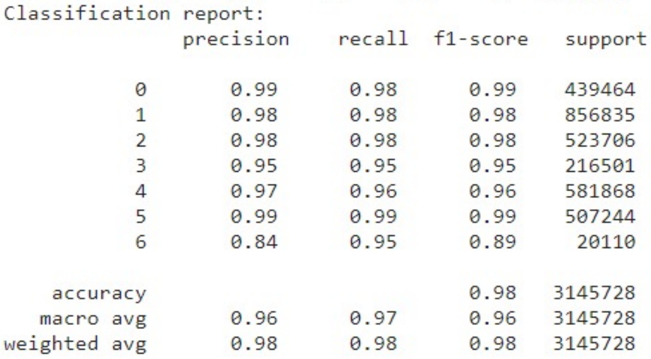
4. Inception UNet

- **Multi-scale information:** Inception modules capture features at different scales, enabling the model to understand the context at various levels and improve segmentation accuracy.
- **Enhanced feature extraction:** By combining skip connections with Inception modules, the Inception U-Net can access multi-scale information and fine-grained details, leading to improved feature extraction.
- **Reduced computational complexity:** The Inception U-Net incorporates bottleneck layers to reduce parameters and computational complexity while preserving feature representations, striking a balance between accuracy and efficiency.





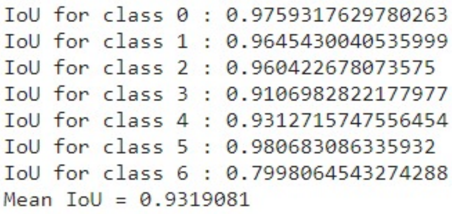




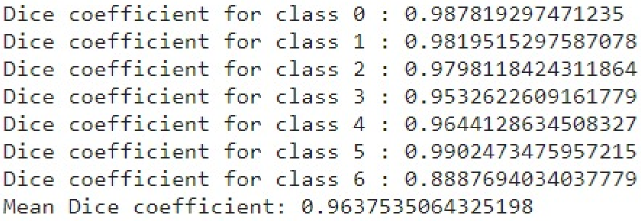
```
Classification report:
              precision    recall  f1-score   support

           0       0.99      0.98      0.99    439464
           1       0.98      0.98      0.98    856835
           2       0.98      0.98      0.98    523706
           3       0.95      0.95      0.95    216501
           4       0.97      0.96      0.96    581868
           5       0.99      0.99      0.99    507244
           6       0.84      0.95      0.89     20110

    accuracy                           0.98   3145728
   macro avg       0.96      0.97      0.96   3145728
weighted avg       0.98      0.98      0.98   3145728
```

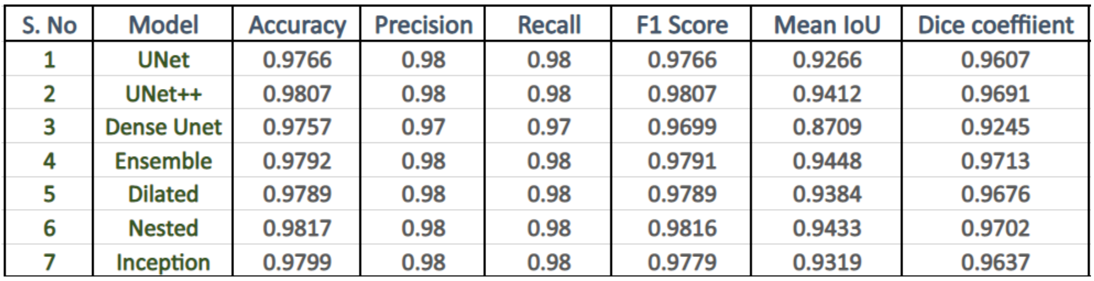


```
IoU for class 0 : 0.9759317629780263
IoU for class 1 : 0.9645430040535999
IoU for class 2 : 0.960422678073575
IoU for class 3 : 0.9106982822177977
IoU for class 4 : 0.9312715747556454
IoU for class 5 : 0.980683086335932
IoU for class 6 : 0.7998064543274288
Mean IoU = 0.9319081
```



```
Dice coefficient for class 0 : 0.987819297471235
Dice coefficient for class 1 : 0.9819515297587078
Dice coefficient for class 2 : 0.9798118424311864
Dice coefficient for class 3 : 0.953262260916179
Dice coefficient for class 4 : 0.9644128634508327
Dice coefficient for class 5 : 0.9902473475957215
Dice coefficient for class 6 : 0.8887694034037779
Mean Dice coefficient: 0.9637535064325198
```

Final observations for UNet and its variants:



S. No	Model	Accuracy	Precision	Recall	F1 Score	Mean IoU	Dice coeffiient
1	UNet	0.9766	0.98	0.98	0.9766	0.9266	0.9607
2	UNet++	0.9807	0.98	0.98	0.9807	0.9412	0.9691
3	Dense Unet	0.9757	0.97	0.97	0.9699	0.8709	0.9245
4	Ensemble	0.9792	0.98	0.98	0.9791	0.9448	0.9713
5	Dilated	0.9789	0.98	0.98	0.9789	0.9384	0.9676
6	Nested	0.9817	0.98	0.98	0.9816	0.9433	0.9702
7	Inception	0.9799	0.98	0.98	0.9779	0.9319	0.9637

5 Conclusion

In conclusion, the application of semantic segmentation techniques in medical imaging has shown promising results for accurate and precise delineation of anatomical structures and abnormalities. Through the evaluation of different semantic segmentation models, namely UNet, UNet++, Nested UNet, and Inception UNet, it is evident that architectural modifications play a crucial role in improving the segmentation performance.

The UNet architecture serves as a strong baseline, demonstrating competitive segmentation accuracy by capturing both local and global context through its encoding and decoding pathways, along with skip connections. Based on the experiment comparing UNet, UNet++, Nested UNet, and Inception UNet for semantic segmentation of medical images UNet variants (UNet++, Nested UNet, Inception UNet) achieved greater accuracies compared to the original UNet model. The UNet variants (UNet++, Nested UNet, and Inception UNet) outperformed the original UNet model in terms of segmentation accuracy for medical images. These variants, which incorporate enhancements such as skip connections, and nested or inception-like architectures, have demonstrated their capability to capture both local and global context, resulting in more precise and accurate segmentation results.

Overall, the evaluation highlights the importance of architectural enhancements in the semantic segmentation of medical images. These advancements contribute to better localization and delineation of structures, thereby aiding clinical decision-making, treatment planning, and computer-aided diagnosis. It is essential to consider the specific requirements of the medical imaging task at hand when selecting the most appropriate model architecture. Additionally, further research and investigation into customized architectural modifications and tailored training strategies can lead to even more advanced and accurate semantic segmentation models for medical image analysis. By leveraging the power of semantic segmentation, medical professionals can benefit from improved visualization, quantitative analysis, and automated detection of abnormalities, ultimately enhancing patient care and clinical outcomes.

6 References

1. CholecSeg8k

2. U-Net_1

3. U-Net variants

4. U-Net_2