Colorectal Polyp Segmentation Using U-Net Based Models

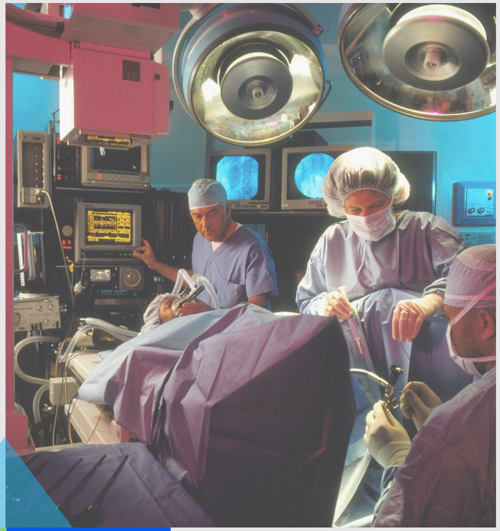
Eldar Eyvazov (2073221)
Antoni Valls Cifre (2072119)

Vision and Cognitive Systems

Index

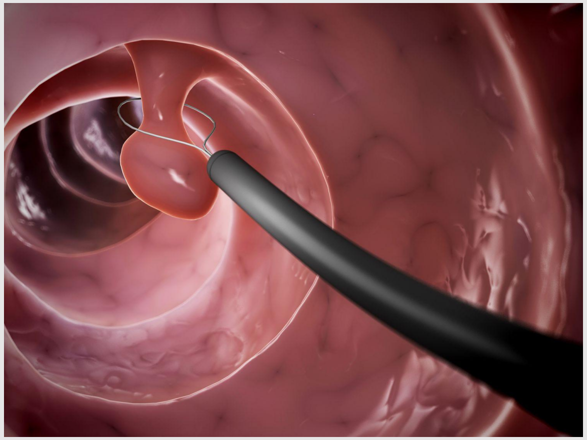
Introduction: Importance of Early Detection in Colorectal Cancer

- Colorectal cancer (CRC) ranks third in cancer-related mortality worldwide. The five-year survival rate of colon cancer stands at approximately 68%.
- Early detection and removal of precancerous polyps through colonoscopy greatly reduce the risk of invasive cancer.
- Polyps are abnormal tissue growths that can develop into CRC if not detected early.
- The accuracy of polyp detection during colonoscopy heavily depends on the quality of the procedure and the expertise of the colonoscopist.
- High-quality image segmentation is essential for accurately identifying polyps during colonoscopy, especially small ones.



Introduction: Overview of Our Research Project

- We evaluate state-of-the-art U-Net based models on the Kvasir-SEG dataset.
- We successfully model U-Net, as our benchmark method, ResUNet, Attention U-Net and UNet++
- We experiment with the improvements achieved through data augmentation.



Related Work

Traditional Polyp Detection Methods

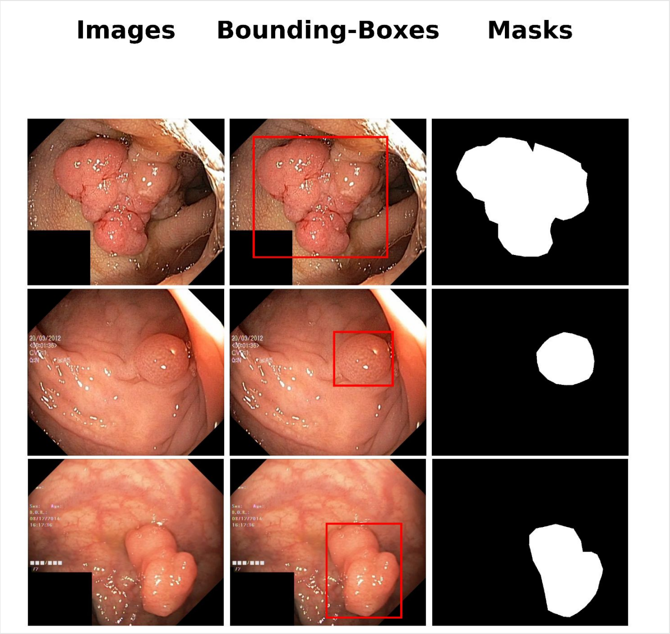
- Early approaches relied on manual inspection by colonoscopists, leading to variability in detection rates.
- Initial computer-aided detection (CAD) systems used hand-crafted features and classical machine learning algorithms.
- Techniques like edge detection, texture analysis, and shape modeling were used but struggled with polyp variability.

Evolution to Deep Learning Techniques

- Deep learning revolutionized medical image analysis, with CNNs showing remarkable performance.
- The U-Net architecture by Ronneberger et al. in 2015 became a cornerstone in medical image segmentation.
- Enhancements like multi-scale feature extraction, attention mechanisms, and ensemble learning improved accuracy.

Dataset Description

- We utilized the Kvasir-SEG dataset for the colorectal polyp segmentation task.
- The dataset includes 1,000 high-resolution polyp images captured using the ScopeGuide electromagnetic imaging system by Olympus Europe.
- Each image has corresponding segmentation masks (ROI) and bounding box annotations.
- The dataset features 700 large polyps (over 160 × 160 pixels), 323 medium-sized polyps (between 64 × 64 pixels and 160 × 160 pixels), and 48 small polyps (64 × 64 pixels or smaller).



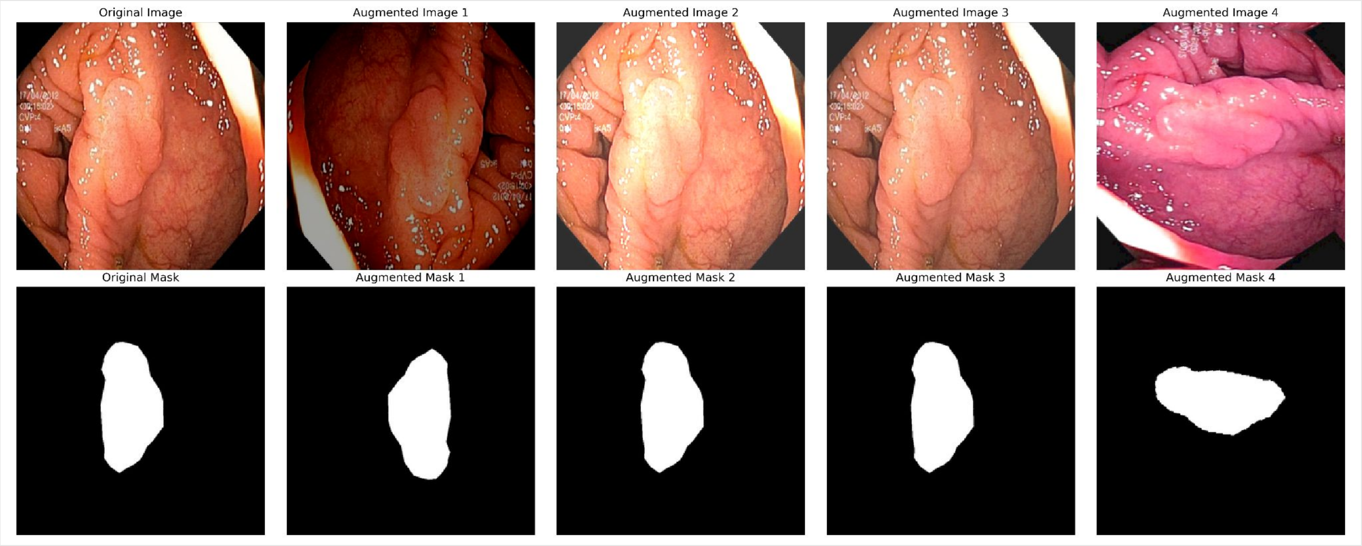
Preprocessing and Methodologies:

- All images and masks are resized to 256x256 pixels
- Splitting dataset (800+100+100 initial split)
- Data Normalization
- Dice Loss used for model training
- Early stopping and Learning rate scheduler (LR on plateau)
- Models are trained on both original and augmented datasets
- Adam Optimizer used for all models

Data augmentation:

- To enhance model performance, the original dataset of 800 train images was expanded to 3,200 training images through extensive data augmentation.

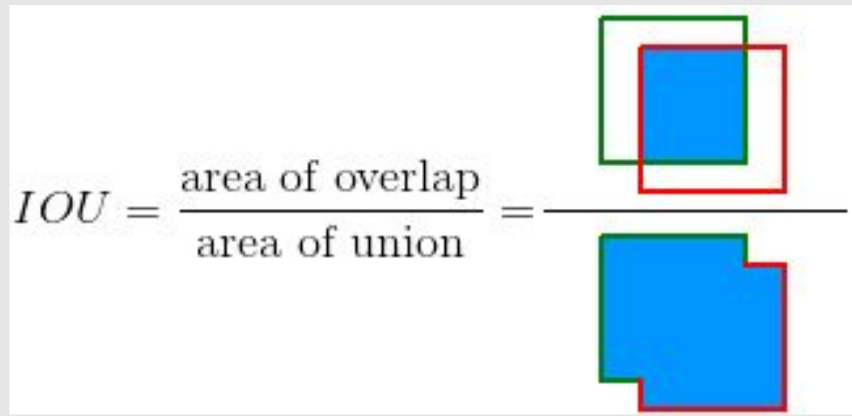
- The Albumentations library was employed for image augmentation, ensuring robust training by diversifying the dataset with synthetic variations.
- Augmentation techniques included rotations, flips, shifts, zooms, and changes in brightness and contrast.



Evaluation Metrics

Intersection over Union (IoU)

The IoU score is defined as the ratio of the intersection of the predicted segmentation mask and the ground truth mask to the union of these masks.

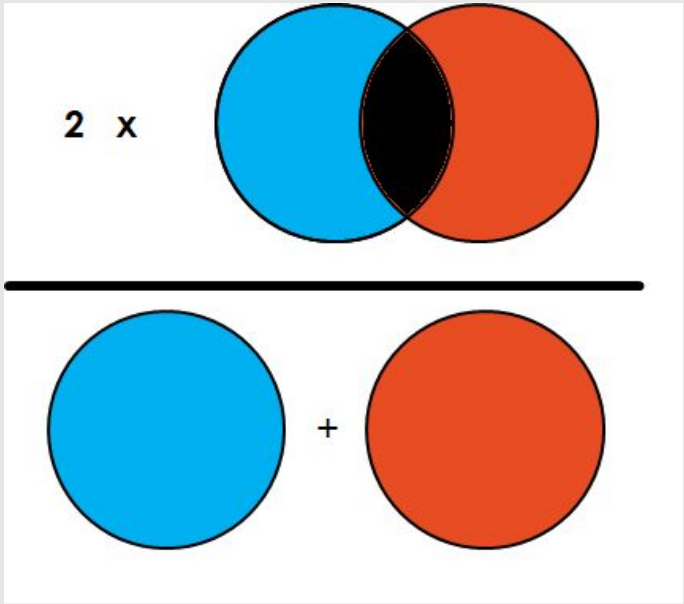


$$\mathrm{IoU} = \frac{|A \cap B|}{|A \cup B|}$$

Evaluation Metrics

Dice Score

The Dice score, also known as the Dice coefficient, is defined as twice the area of overlap between the predicted segmentation mask and the ground truth mask divided by the total number of pixels in both masks.



$$\text{Dice score} = \frac{2|A \cap B|}{|A|+|B|}$$

Evaluation Metrics

Pixel Accuracy

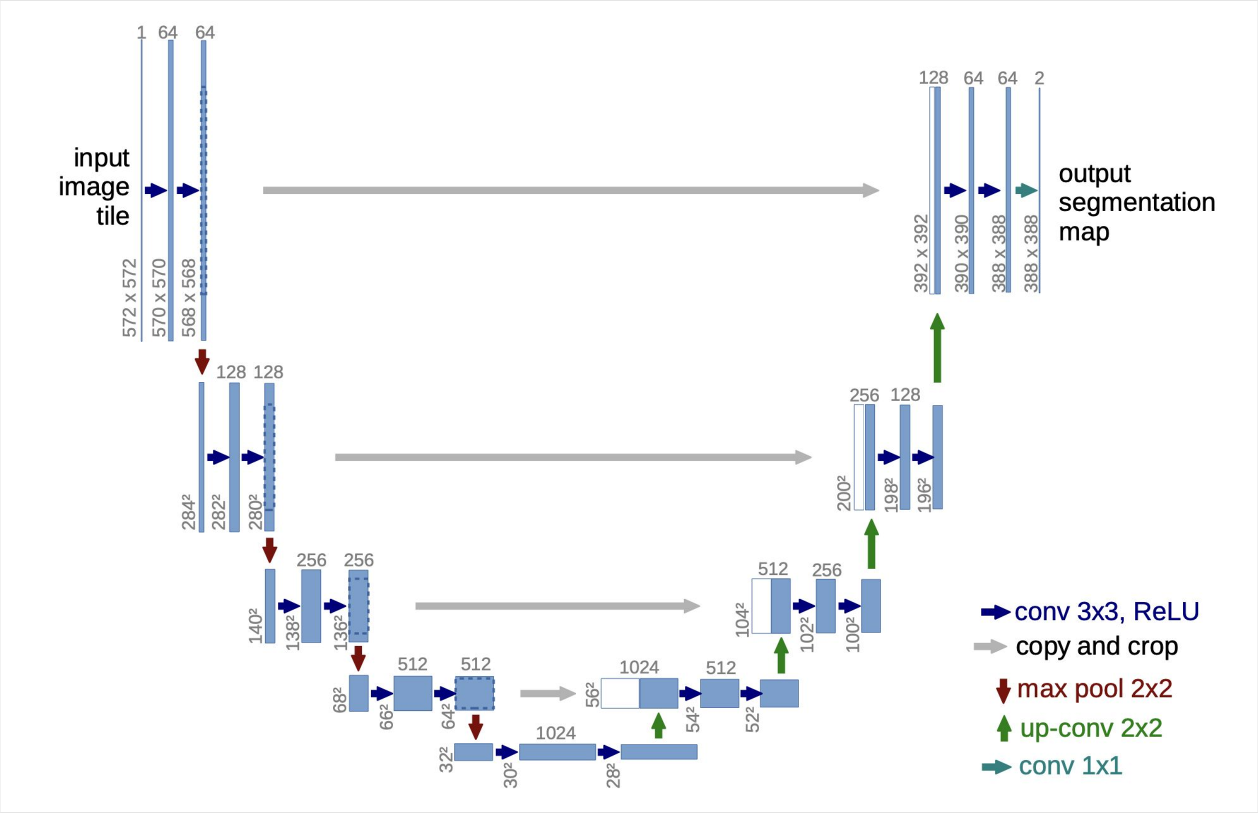
Pixel accuracy, or just accuracy, is defined as the ratio of the number of correctly predicted pixels to the total number of pixels.

$$\text{Accuracy} = \frac{\text{Number of correctly predicted pixels}}{\text{Total number of pixels}}$$

U-Net

- Introduced by Ronneberger et al. in 2015. Designed for biomedical image segmentation, particularly effective in segmenting images with limited training data.
- Features a symmetrical encoder-decoder architecture.
    - **Encoder Path:** Consists of convolutional layers followed by max-pooling for down-sampling.
    - **Decoder Path:** Comprises convolutional layers followed by up-sampling to recover spatial resolution.
- **Includes skip connections:** Between the encoder and decoder layers , which help preserve spatial information.
- Architecture with 1,747,489 trainable parameters

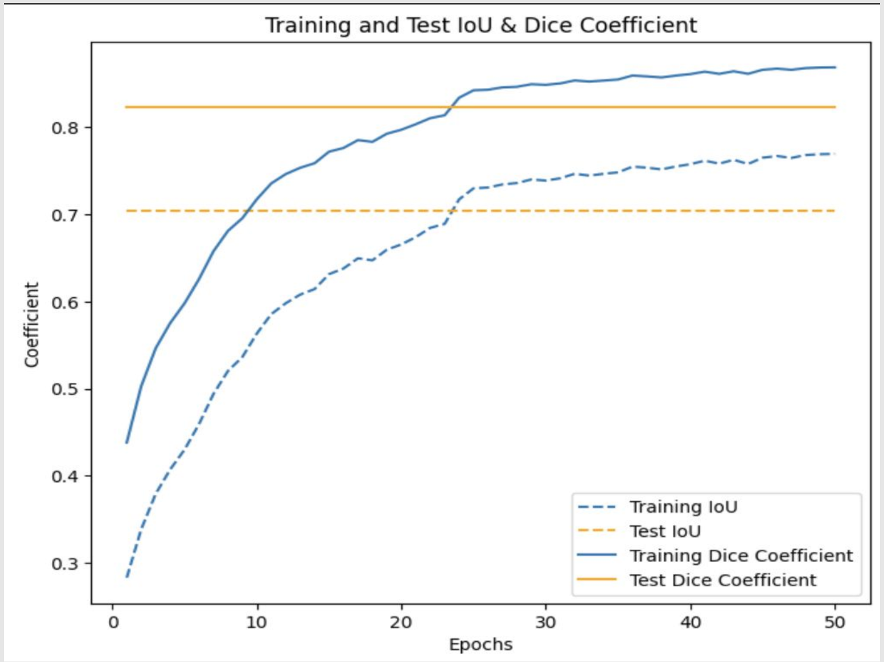
U-Net

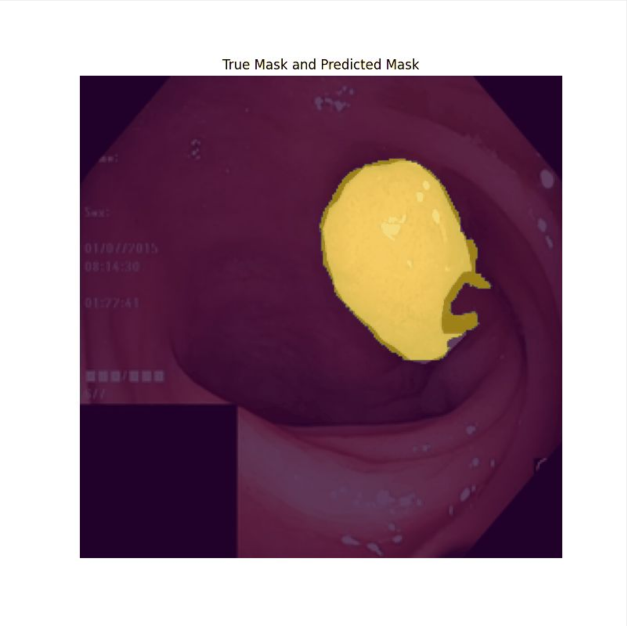


1
64
64
input
image
tile
572 x 572
570 x 570
568 x 568
128
128
284²
282²
280²
256
256
140²
138²
136²
512
512
68²
66²
64²
1024
32²
30²
28²
1024
512
56²
54²
52²
512
256
104²
102²
100²
256
128
200²
198²
196²
128
64
64
2
392 x 392
390 x 390
388 x 388
388 x 388
output
segmentation
map
conv 3x3, ReLU
copy and crop
max pool 2x2
up-conv 2x2
conv 1x1

U-Net: Results

Method	Loss	IoU	Dice	Accuracy
U-Net Non-Aug	0.3423	0.5123	0.6750	0.8874
U-Net Aug	**0.1767**	**0.7042**	**0.8233**	**0.9480**

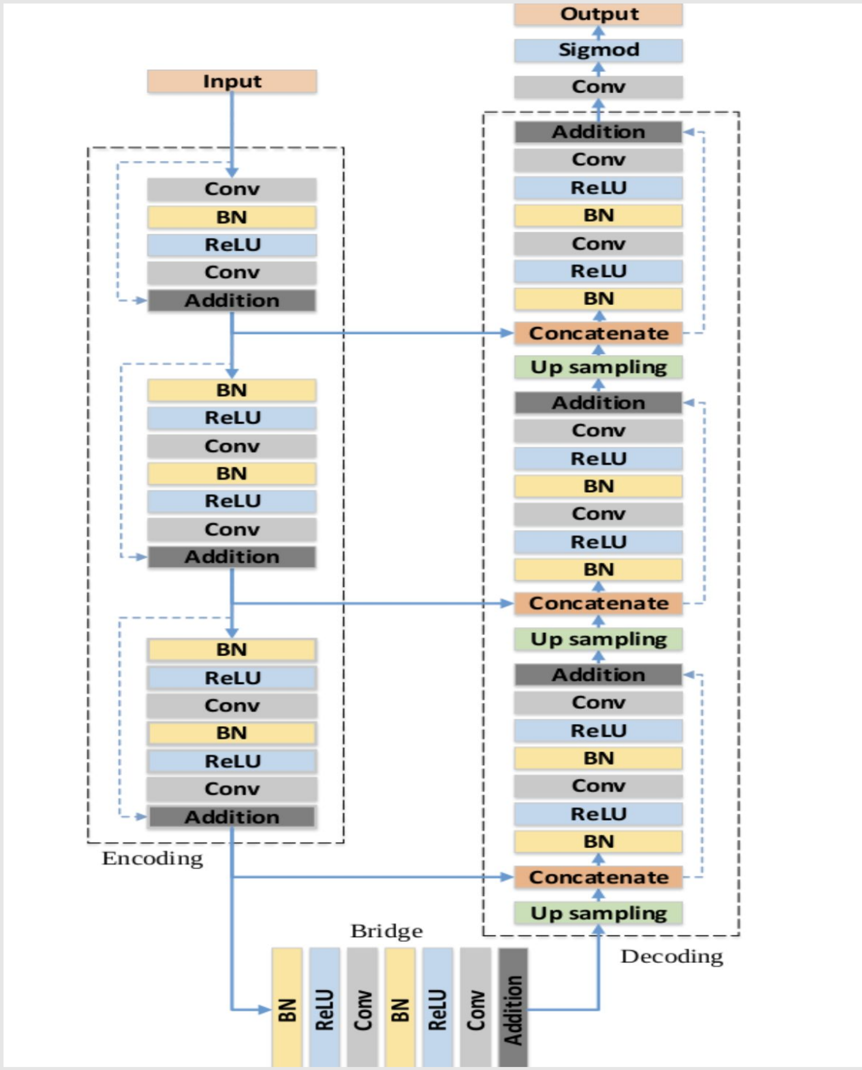


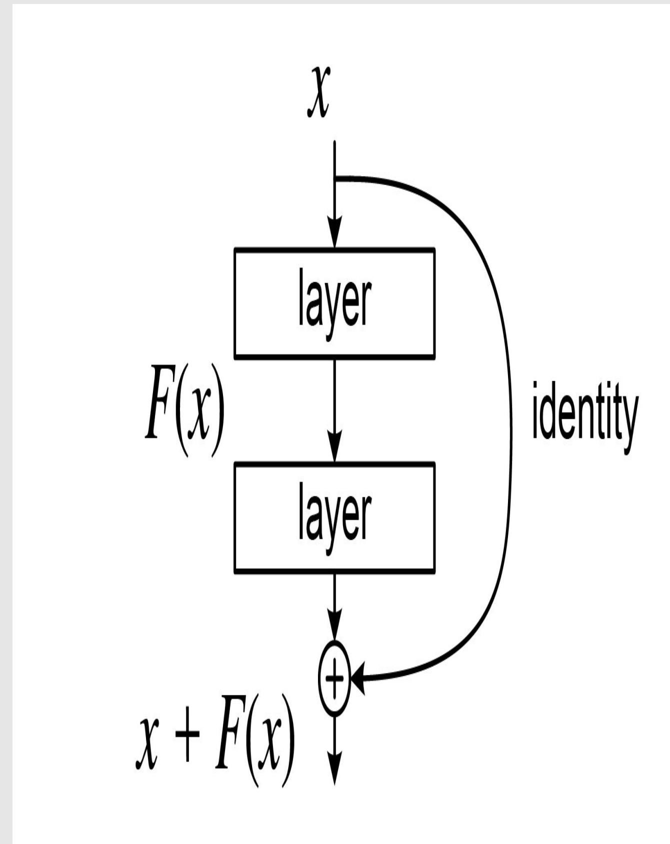


ResUNet

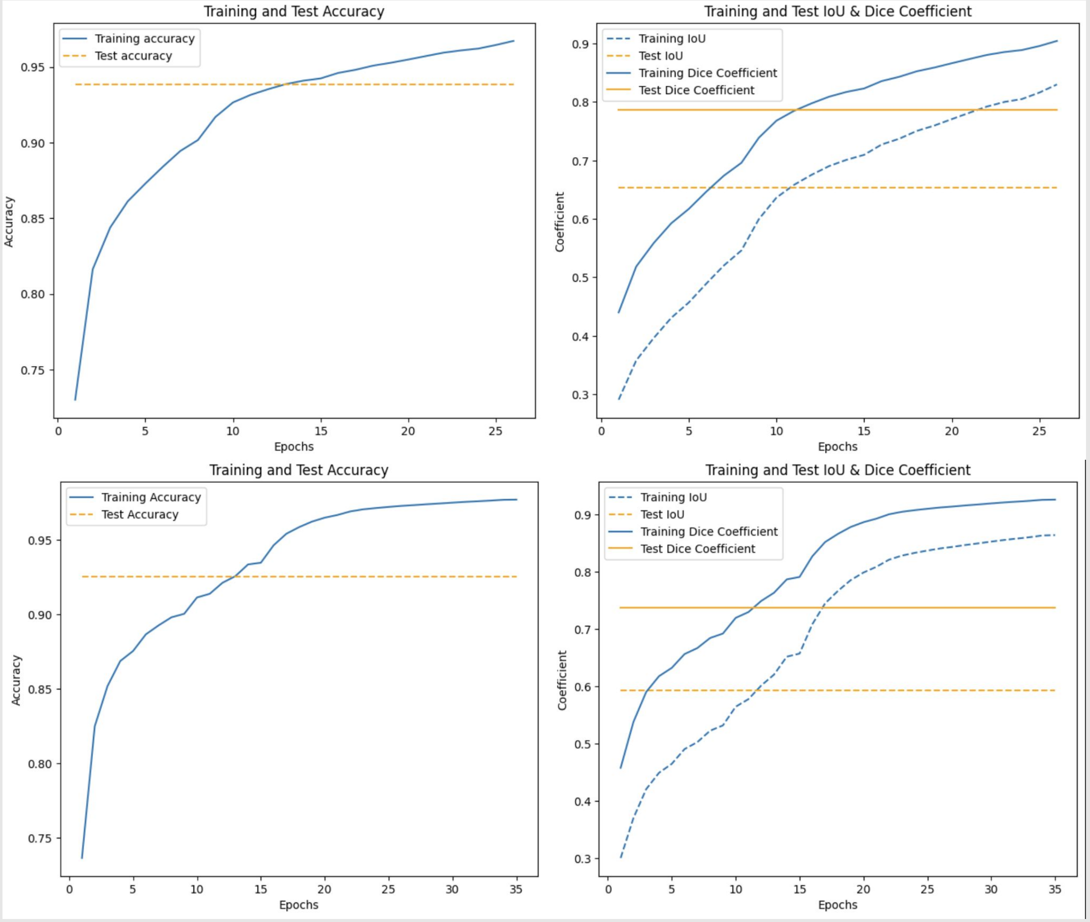
- **Developed from U-Net, incorporating residual blocks.**
- **Effective in deep image segmentation tasks, reducing vanishing gradient issues.**
- **Residual Block-based Encoder-Decoder:**
  - **Encoder Path:**
    - Residual blocks with convolutional layers, followed by max-pooling.
  - **Decoder Path:**
    - Residual blocks with convolutional layers, followed by up-sampling.
- **Skip Connections:**
  - Link encoder and decoder layers at the same resolution, preserving spatial information and enhancing gradient flow.
- **Residual Connections:**
  - Facilitate training deeper networks without degradation.
- **Trainable Parameters:**
  - Contains 7,597,377 trainable parameters after adjusting model architecture for our task at the same time considering computational efficiency

ResUNet


Input
Conv
BN
ReLU
Conv
Addition
BN
ReLU
Conv
BN
ReLU
Conv
Addition
BN
ReLU
Conv
BN
ReLU
Conv
Addition
Encoding
Bridge
BN
ReLU
Conv
BN
ReLU
Conv
Addition
Output
Sigmod
Conv
Addition
Conv
ReLU
BN
Conv
ReLU
BN
Concatenate
Up sampling
Addition
Conv
ReLU
BN
Conv
ReLU
BN
Concatenate
Up sampling
Addition
Conv
ReLU
BN
Conv
ReLU
BN
Concatenate
Up sampling
Decoding


x
layer
F(x)
layer
identity
x + F(x)

ResUNet: Results

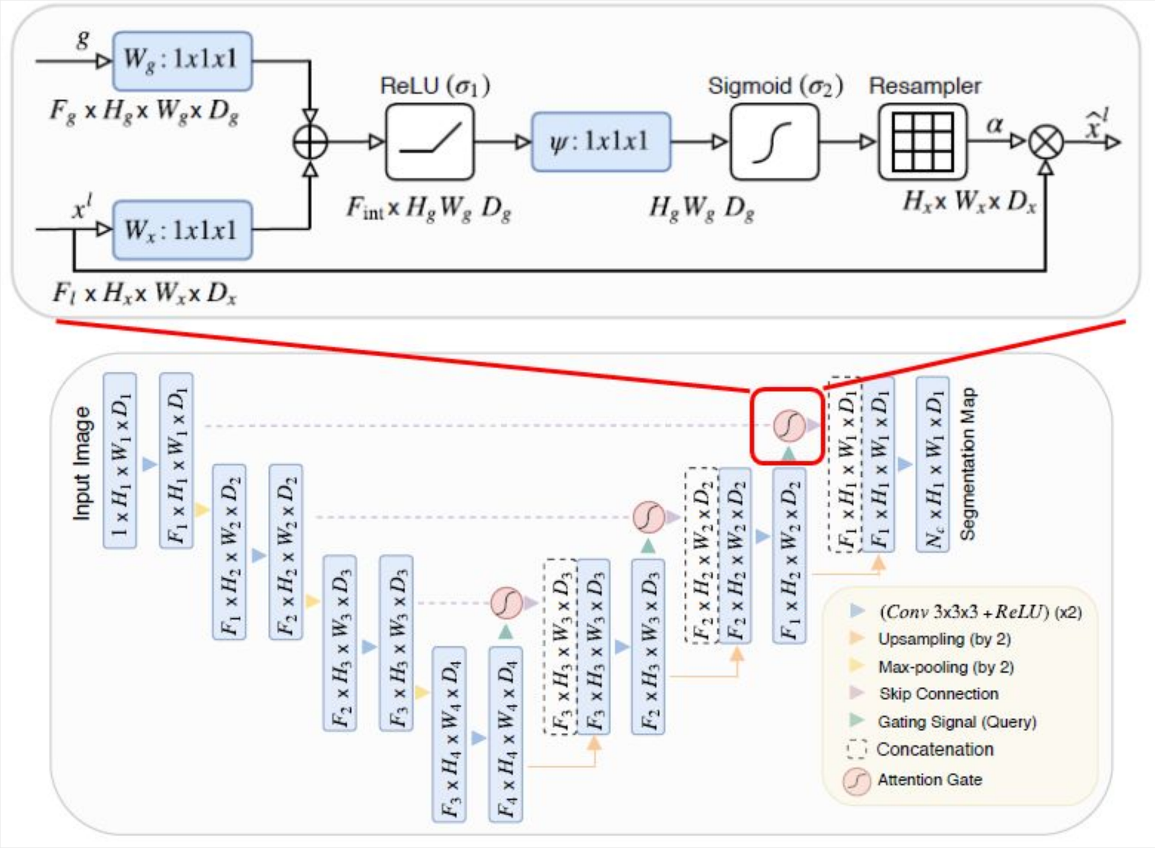


Method	Loss	IoU	Dice	Accuracy
ResUNet Non-Aug	0.4816	0.5935	0.7378	0.9251
ResUNet Aug	**0.3358**	**0.6536**	**0.7868**	**0.9382**

Attention U-Net

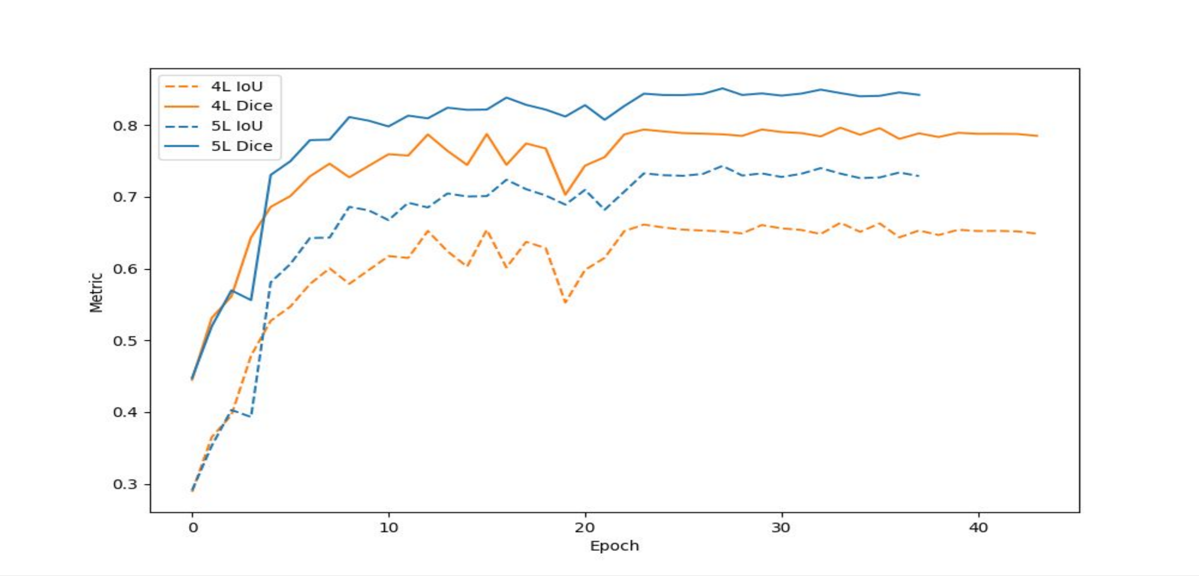
- Introduced by Oktay et al. in 2018.
- It enhances the standard U-Net model for medical image segmentation by integrating **attention gates** (AGs) into its architecture.
- These AGs are embedded into the skip connections to capture contextual information from different scales, focusing on relevant image regions while suppressing irrelevant ones.
- It has proved to be particularly suitable for tasks with large inter-patient variability in organ shape and size.
- We experimented with a 4-level model that has 2,264,899 trainable parameters, and a more complex 5-level model with 8,726,077 trainable parameters, highlighting the trade-off between model complexity and performance.

Attention U-Net



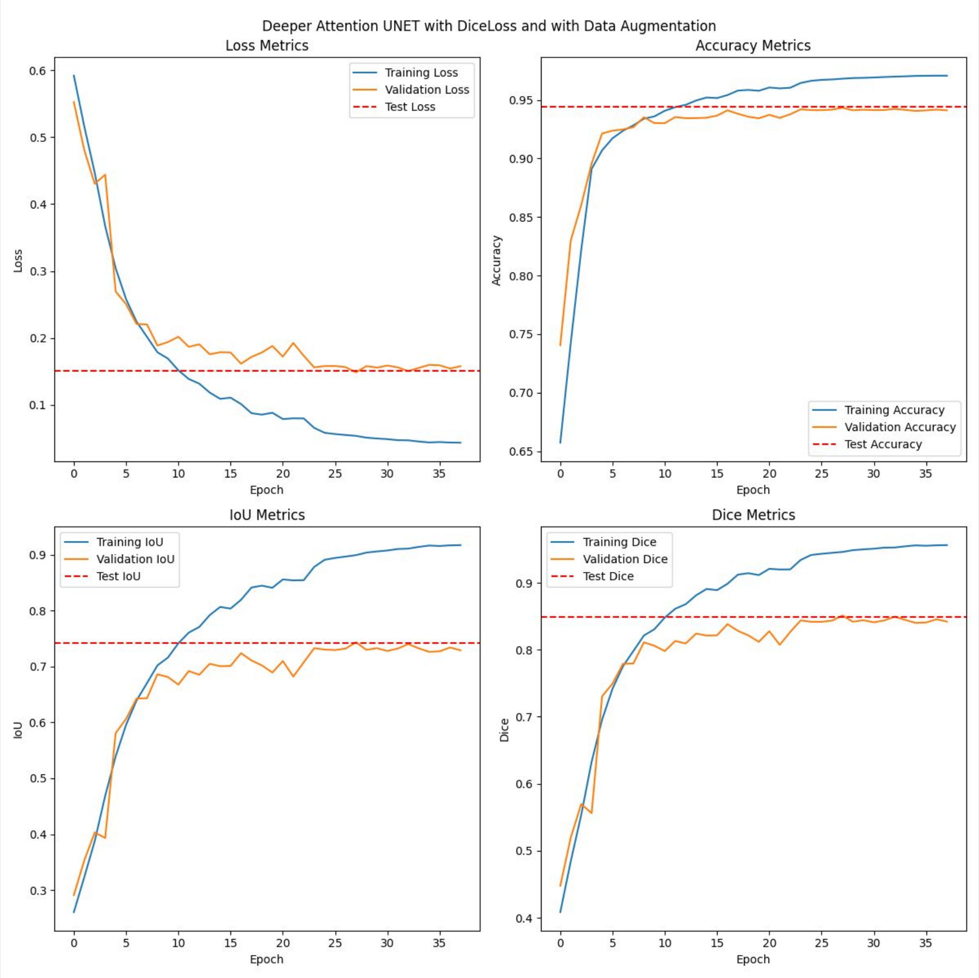
Attention U-Net: Results

Method	Loss	IoU	Dice	Accuracy
Non-Aug 4-L	0.2795	0.5671	0.7205	0.9154
Aug 4-L	0.2004	0.6696	0.7996	0.9299
Non-Aug 5-L	0.2147	0.6527	0.7853	0.9302
Aug 5-L	**0.1508**	**0.7422**	**0.8492**	**0.9440**



Validation set Dice and IoU score evolution when training.

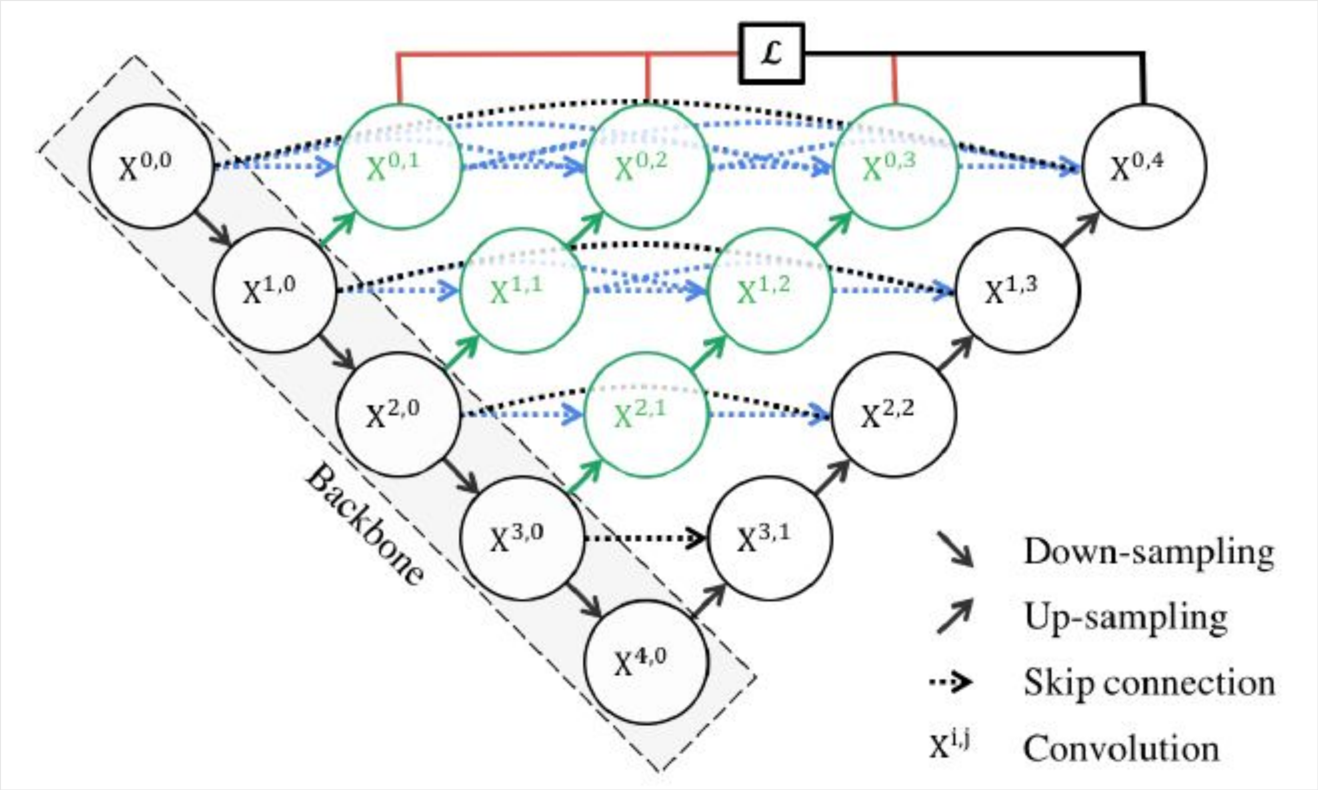
Attention U-Net: Results



UNet++

- Also known as Nested U-Net, was developed by Zhou et al. for medical image segmentation in 2018.
- It features a **deeply-supervised** encoder-decoder network with redesigned skip pathways that use dense convolutional blocks. These pathways bridge the semantic gap between the encoder and decoder, enhancing feature map fusion and simplifying the learning task for the optimizer.
- The deep supervision technique involves adding auxiliary classifiers to intermediate layers, improving training by mitigating the vanishing gradient problem and encouraging learning of discriminative features at multiple scales. We experimented UNet++ with and without DS.
- To optimize computational efficiency, we have pruned the original model to 4 levels, resulting in 2,264,833 trainable parameters.

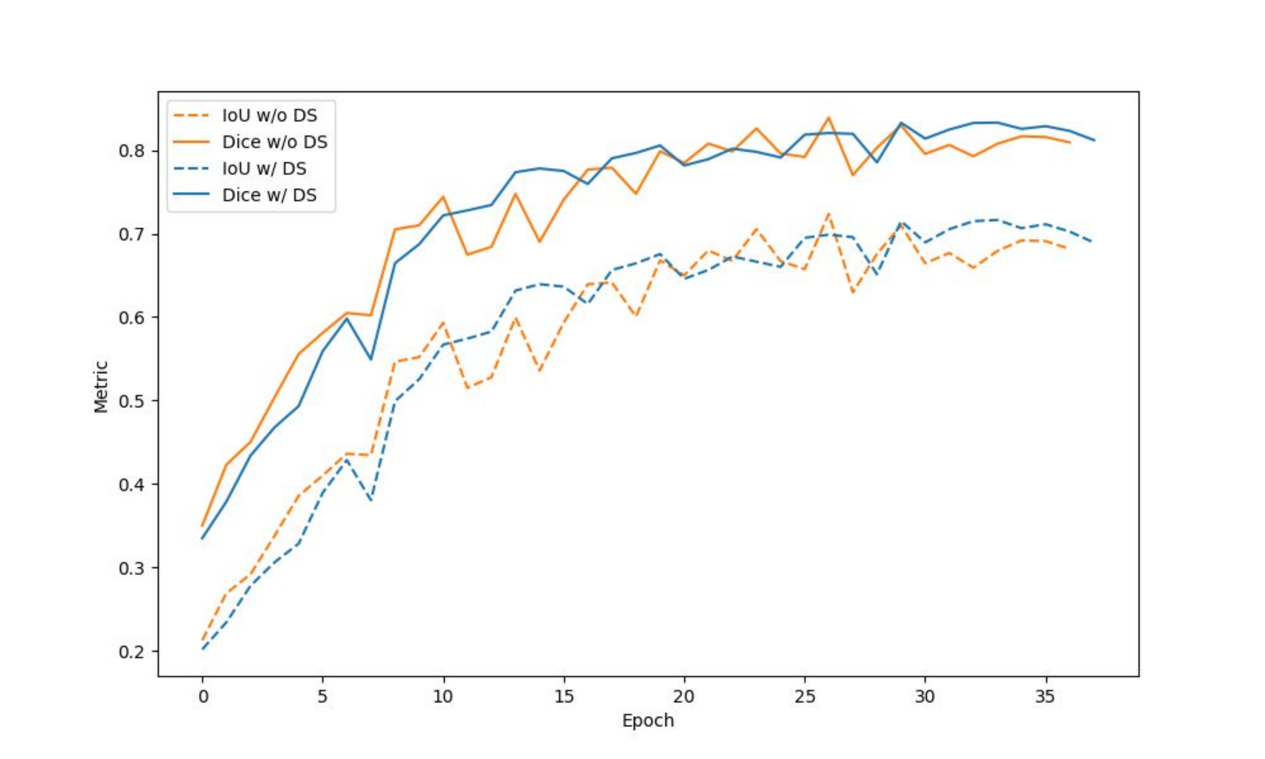
UNet++



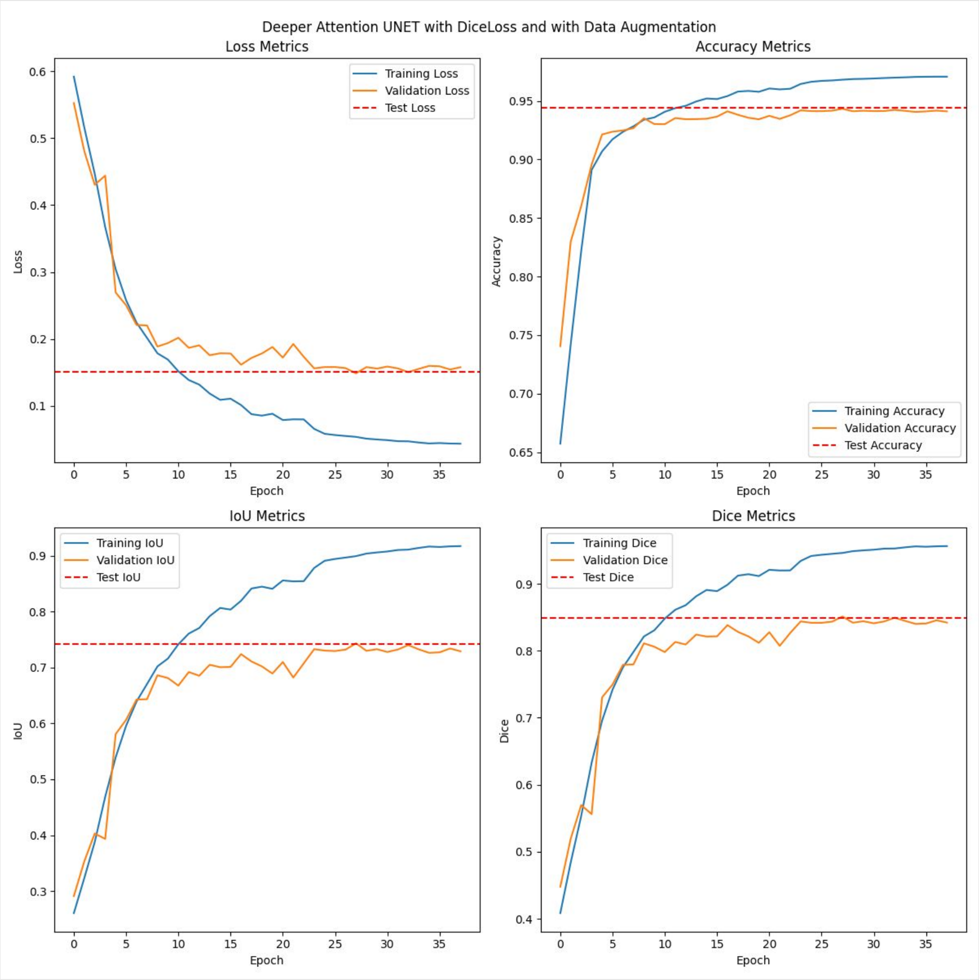
$\mathcal{L}$
$X^{0,0}$
$X^{0,1}$
$X^{0,2}$
$X^{0,3}$
$X^{0,4}$
$X^{1,0}$
$X^{1,1}$
$X^{1,2}$
$X^{1,3}$
$X^{2,0}$
$X^{2,1}$
$X^{2,2}$
$X^{3,0}$
$X^{3,1}$
$X^{4,0}$
Backbone
Down-sampling
Up-sampling
Skip connection
$X^{i,j}$ Convolution

UNet++: Results

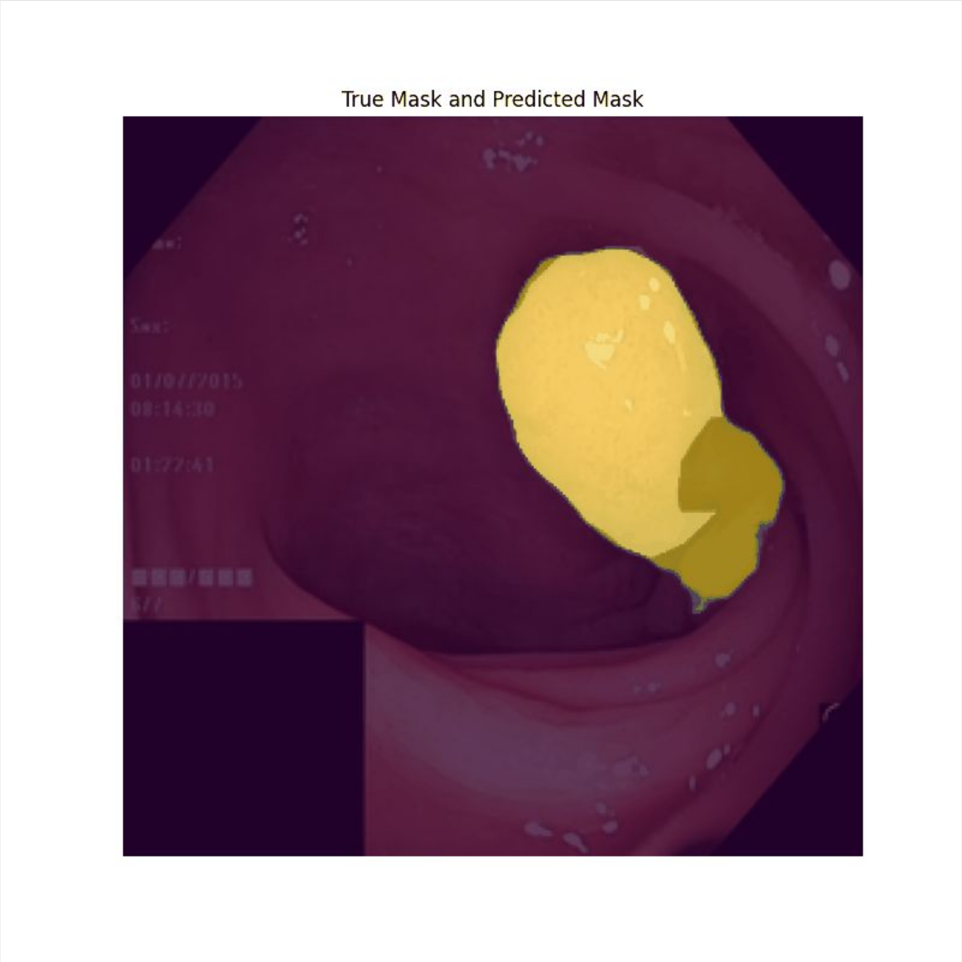
Method	Loss	IoU	Dice	Accuracy
Non-Aug w/o DS	0.3025	0.5378	0.6975	0.9166
Aug w/o DS	**0.1710**	0.7103	0.8290	0.9416
Non-Aug w/ DS	0.3467	0.5994	0.7471	0.9252
Aug w/ DS	0.2465	**0.7624**	**0.8639**	**0.9480**



UNet++: Results

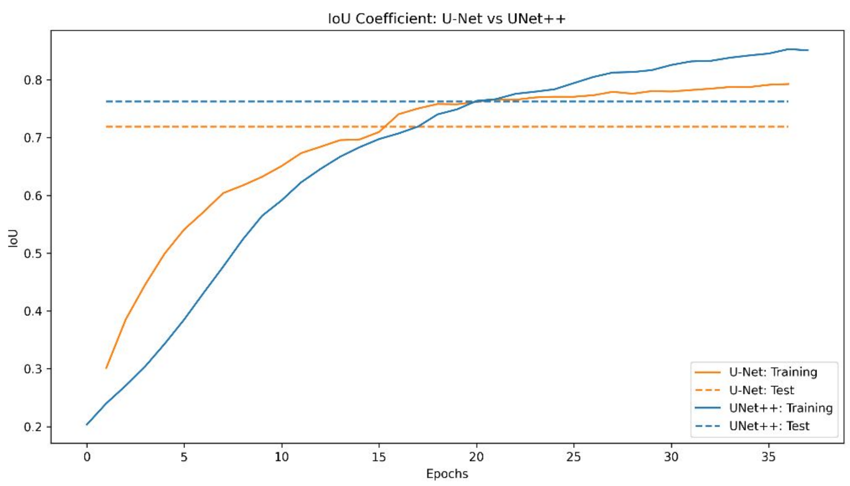


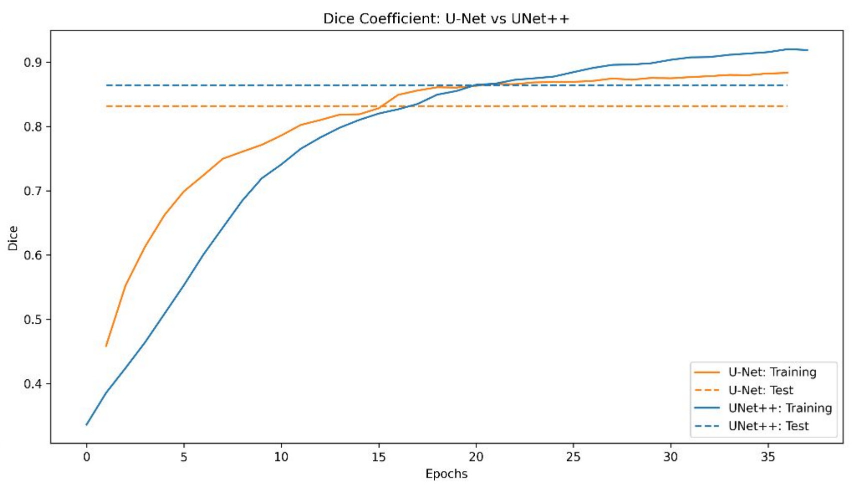
UNet++: Results

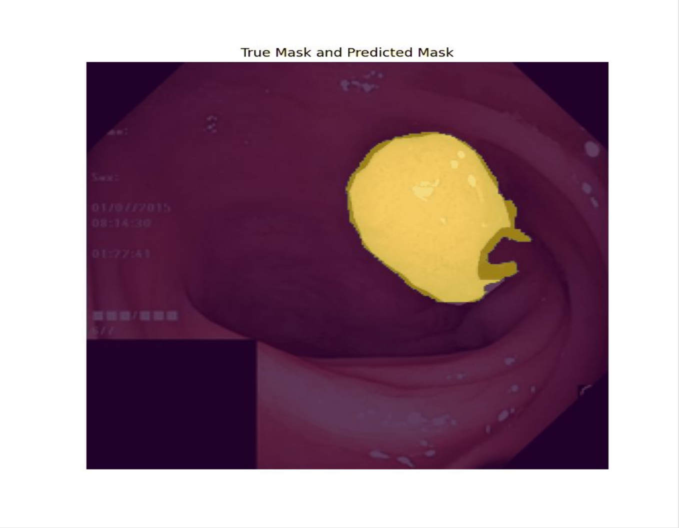


True mask and predicted mask from the UNet++ w/ DS

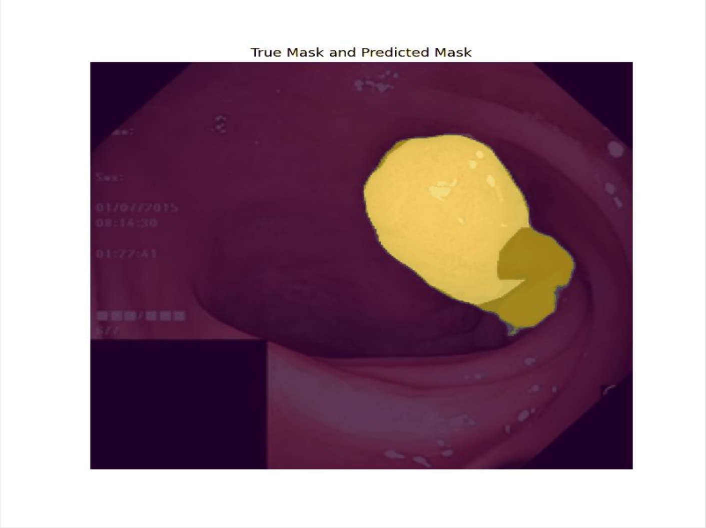
Comparing the Models







U-Net



UNet++

Comparing the Models

Model	Loss	IoU	Dice	Accuracy
U-NET	0.1767	0.7042	0.8233	0.9480
ResUNet	0.3358	0.6536	0.7868	0.9382
Attention UNet (5L)	**0.1508**	0.7422	0.8492	0.9440
UNet++ w/ DS	0.2465	**0.7624**	**0.8639**	**0.9489**

Test set results of models trained on augmented datasets.

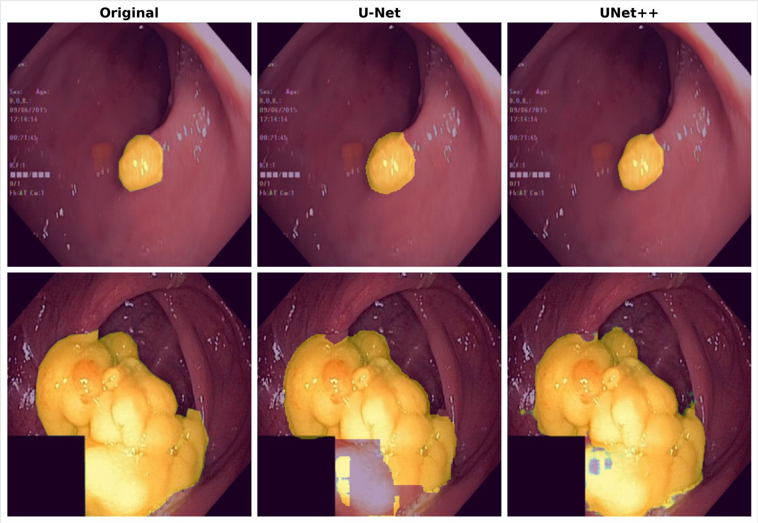
Conclusion

- Our study on polyp segmentation using the Kvasir-SEG dataset found that the UNet++ model, trained on augmented data with DS, significantly outperformed the baseline U-Net model. The superior IoU and Dice coefficients achieved by the UNet++ model demonstrate its enhanced segmentation accuracy. This underscores the advantages of employing a more advanced architecture and sophisticated training techniques.

- The ResUNet model did not perform better than the baseline U-Net when trained on augmented data. Similarly, the Attention U-Net with a 4-level architecture also underperformed relative to the U-Net. However, when configured with 5 levels, the Attention U-Net surpassed the performance of the U-Net.

- UNet++ model emerged as the best performer among all the models evaluated.

- Integrating such advanced models into clinical practice could improve CRC detection and diagnosis, reduce variability in polyp detection, and ultimately contribute to better patient outcomes.



Thank you.