

Colorectal Polyps Segmentation Using U-Net Based Models

Eldar Eyyazov

eldar.eyvazov@studenti.unipd.it

Antoni Valls Cifre

antonи.vallscifre@studenti.unipd.it

Abstract

Colorectal cancer (CRC) is a major cause of cancer-related deaths globally, emphasizing the need for accurate polyp detection and segmentation. In this project, we aimed to evaluate and improve polyp segmentation using U-Net based models. We began with the U-Net, one of the most popular segmentation models used in the medical image segmentation domain, and then explored advanced architectures including ResUNet, Attention UNet and UNet++.

First, we assessed the performance of these models on the original Kvasir-SEG dataset. To achieve better performance, we subsequently trained the models on augmented data. Among the models, the UNet++ with deep supervision (DS) and augmented data outperformed the others, demonstrating superior segmentation accuracy with higher IoU and Dice coefficients.

Our findings indicate that the UNet++ model with DS and augmented data is the most effective approach for polyp segmentation, offering enhanced accuracy and potential for improved CRC detection and diagnosis in clinical settings. This highlights the value of advanced architectures and training techniques in medical image analysis.

1. Introduction

Colorectal Cancer (CRC) is one of the leading causes of cancer-related deaths worldwide. It has the third highest mortality rate among all cancers. The overall five-year survival rate of colon cancer is around 68% [1]. Early detection and removal of precancerous polyps during colonoscopy significantly reduce the risk of developing invasive colon cancer. Polyps are abnormal tissue growths in the colon lining that can develop into the CRC at late stage, so their early detection is crucial for cancer prevention and treatment [2].

Colorectal polyps can be classified by size into three categories: diminutive (5 mm), small (6 to 9 mm), and advanced (large) (10 mm) [3]. Larger polyps are typically easier to detect and remove. However, there is a significant risk associated with small and diminutive colorectal polyps.

Colonoscopy is the most effective procedure for detecting and removing polyps. It involves a visual examination

of the colon using a colonoscope, a flexible tube with a camera and light at the end. However, the accuracy of polyp detection during colonoscopy heavily depends on the quality of the procedure and the expertise of the endoscopist. Despite advancements, a significant number of polyps, particularly small and flat ones, can be missed during routine examinations .

In this context, high-quality image segmentation plays a critical role. Image segmentation involves partitioning an image into meaningful regions, allowing for precise identification and localization of polyps. Automated image segmentation techniques can assist endoscopists by highlighting suspicious areas, ensuring that even subtle polyps are not overlooked. This not only enhances the detection rate but also reduces the variability in polyp detection among different endoscopists .

The necessity of high-quality image segmentation in colonoscopy cannot be overstated. It enhances the diagnostic accuracy, reduces the risk of missing lesions, and ultimately contributes to better patient outcomes. As technology continues to evolve, integrating advanced image segmentation algorithms into routine clinical practice will become increasingly feasible, providing valuable support to medical professionals in the fight against colon cancer.

In this paper, we evaluate state-of-the-art U-Net based methods on the Kvasir-SEG dataset [4] to provide a comprehensive benchmark for colonoscopy images. We successfully model U-Net, ResUNet, AttentionUNet and UNet++ for polyp detection and experiment with the improvements achieved through data augmentation.

The rest of the paper is organized as follows: Section 2 presents related work in the field. In Section 3, we describe the dataset. Section 4 details the architecture of the models, the methodology for polyp detection, and the metrics used. Section 5 presents the results, and Section 6 provides the conclusion.

2. Related work

The application of deep learning techniques in medical image analysis, particularly in colorectal polyp detection, has seen significant advancements in recent years. Traditional methods for polyp detection relied heavily on manual

inspection by endoscopists, leading to variability in detection rates and missed polyps, especially diminutive ones. To address these challenges, various computer-aided detection (CAD) systems have been developed [5].

Early approaches to polyp detection utilized hand-crafted features and classical machine learning algorithms. Techniques such as edge detection, texture analysis, and shape modeling were employed to identify polyp regions in colonoscopy images [6] and videos [7]. However, these methods struggled with the variability and complexity of polyp appearances, leading to suboptimal performance.

The advent of deep learning revolutionized the field of medical image analysis. Convolutional Neural Networks (CNNs) have demonstrated remarkable performance in medical segmentation tasks [8, 9], including polyp detection [3], and have been the preferred approach for participants in public challenges [10, 11].

The U-Net architecture, introduced by Ronneberger et al. in 2015 [12], has become a cornerstone in medical image segmentation due to its ability to capture both local and global features through a symmetric encoder-decoder structure with skip connections. Recent studies have explored various enhancements to the basic U-Net architecture. For instance, the use of multi-scale feature extraction, attention mechanisms, and ensemble learning has led to significant improvements in segmentation accuracy. Extensive literature exists on polyp segmentation using U-Net variants [13, 14, 15]. In this study, we focus on ResUNet [16], Attention UNet [17], and UNet++ [18], using the original U-Net as the baseline model.

3. Dataset

We utilized the Kvasir-SEG dataset [4] for the colorectal polyp segmentation task. This dataset is part of an initiative promoting open and reproducible research results. It comprises 1,000 high-resolution polyp images captured using the ScopeGuide electromagnetic imaging system by Olympus Europe, along with their corresponding segmentation masks (ROI) and bounding box annotations. The open-access dataset can be easily downloaded for research purposes at: <https://datasets.simula.no/kvasir-seg/>.

The Kvasir-SEG dataset supports both segmentation and detection tasks. The ground truth masks are used for segmentation, while the bounding box information facilitates detection. In detail, the dataset includes 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) [3]. Some images can have more than one polyp.

Figure 1 illustrates sample images from the Kvasir-SEG dataset, showing bounding boxes (second column) and annotated masks (third column) for selected samples.

Images Bounding-Boxes Masks

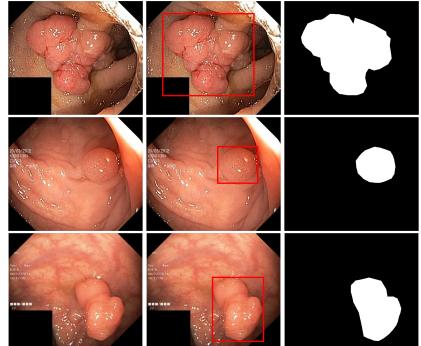


Figure 1. Visualization of three samples from the Kvasir-SEG dataset, along with their corresponding bounding-boxes and masks.

3.1. Data augmentation

To enhance our model performance, we expanded our dataset to 3,800 images through extensive data augmentation. We employed the well-regarded Albumentations library for image augmentation [19], ensuring robust training by diversifying our dataset with synthetic variations.

4. Methodology

In this section, we will introduce the methodologies employed for each model, detailing the architectures and the hyperparameters selected for our experiments.

Our study involves several state-of-the-art U-Net based models, including U-Net, ResUNet, Attention U-Net, and UNet++. Each model has been tailored to the task of colorectal polyp segmentation, with careful consideration given to their unique architectural features and the specific hyperparameters that optimize their performance.

We have evaluated the performance of all models on both the original and augmented datasets to ensure a comprehensive analysis.

4.1. U-Net

U-Net, introduced by Olaf Ronneberger, Philipp Fischer, and Thomas Brox in 2015 [12], revolutionized biomedical image segmentation. Its innovative U-shaped architecture, featuring an encoder-decoder design, excels at capturing contextual information through downsampling and achieving precise localization with upsampling. The use of skip connections between corresponding encoder and decoder layers preserves spatial details, making U-Net particularly effective even with small datasets. Its early success in accurately segmenting medical images has made it a cornerstone in the field.

In our research, U-Net serves as our baseline model, specifically a classic 4-level U-Net with 1,747,489 trainable parameters. Our initial focus is on evaluating the pure performance of various models. As outlined in the methodology section, the first stage involves assessing the baseline performance without any enhancements. In the second stage, we aim to boost performance through data augmentation techniques. This structured approach allows us to systematically evaluate and improve polyp segmentation models.

Figure 2 depicts the U-Net architecture, which highlights its encoder-decoder structure and skip connections. We will analyze the trade-off between model performance and complexity to find a suitable solution for real-world applications. This involves balancing the performance and efficiency of the models to ensure they are practical and effective for deployment in clinical settings.

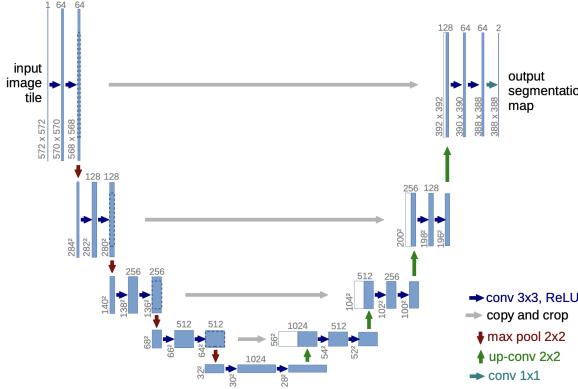


Figure 2. Architecture of U-Net model [12].

4.2. ResUNet

ResUNet is an advanced version of the U-Net architecture that incorporates residual connections, inspired by ResNet [16]. This model addresses some limitations of the original U-Net by improving gradient flow during training and enhancing the ability to learn deeper, more complex features. In our research, ResUNet serves as our second model. We will compare its performance to our baseline U-Net model. Following our workflow, we will train ResUNet on both the original and augmented datasets. This approach allows us to evaluate the pure performance of ResUNet and assess the impact of data augmentation techniques, as outlined in our methodology section.

The ResUNet architecture maintains the U-shaped structure of U-Net but integrates residual blocks. Figure 3 illustrates the architecture of the ResUNet model.

To ensure a reasonable training time while maintaining performance, we have reduced the number of filters in each layer. After these adjustments, our version of ResUNet

comprises 7,597,377 trainable parameters. We aim to determine whether these advancements in ResUNet can deliver improved performance compared to the baseline U-Net model. By analyzing the trade-off between model performance and complexity, we seek to find a balance that offers practical and effective solutions for real-world applications.

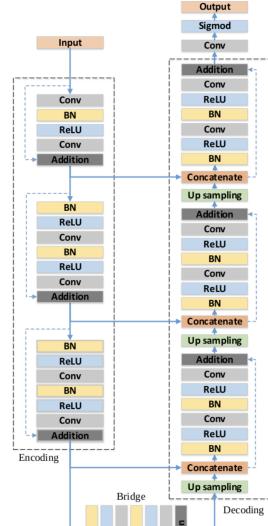


Figure 3. Architecture of ResUNet model [16].

4.3. Attention UNet

Attention U-Net, as introduced by Oktay Et. Al. [17], incorporates attention gates (AGs) to enhance the standard U-Net model for medical image segmentation tasks. By embedding AGs into the skip connections of the U-Net, the model efficiently captures contextual information from different scales, focusing on relevant regions of the image by suppressing irrelevant areas and highlighting salient features without needing external tissue or organ localization modules. The integration of AGs into the U-Net architecture is computationally efficient and improves prediction accuracy by leveraging contextual information from coarser scales. This approach proves to be particularly suitable for tasks with large inter-patient variability in organ shape and size, such as polyp segmentation.

In this model, we experimented with varying the depth of the network to assess its impact on performance. Specifically, we compared the performance of a 4-level Attention UNet with that of a 5-level Attention UNet. The 4-level model contains 2,264,899 trainable parameters, while the 5-level model is significantly more complex, with 8,726,077 trainable parameters. Figure 4 depicts a 4-level Attention UNet and its attention gate.

We have compared both models to provide insights into their computational demands and suitability for real-time

applications. Our findings contribute to the broader discussion on the trade-offs between model complexity and performance in deep learning architectures.

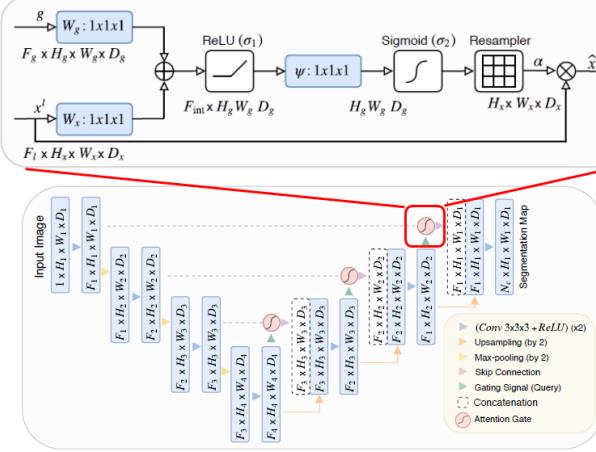


Figure 4. Diagram of the Attention U-Net architecture with a 4-level structure and a detailed illustration of the attention gate. Image from the Oktay Et. Al. paper [17].

4.4. UNet++

UNet++, also known as Nested U-Net, was developed by Zhou et al. [18] for medical image segmentation. This model features a deeply-supervised encoder-decoder network architecture. One of its key innovations is the redesigned skip pathways that utilize dense convolutional blocks to bridge the semantic gap between the encoder and decoder. This design enhances feature map fusion, ensuring that the feature maps are semantically similar and thus making the learning task easier for the optimizer.

Deep supervision in U-Net++ involves adding auxiliary classifiers to intermediate layers of the network, allowing for gradient flow directly from these classifiers. This technique improves training by mitigating the vanishing gradient problem and encourages the learning of more discriminative features at multiple scales. It results in more accurate and robust feature representations. Figure 5 shows a 5-level UNet++. To optimize computational efficiency, we have pruned the original model to 4 levels, resulting in 2,264,833 trainable parameters.

We have trained the UNet++ model in both deep supervision and non-deep supervision settings. We compared the performance of these two configurations to evaluate their effectiveness in the polyp segmentation task.

4.5. Metrics

In this subsection, we present the metrics used for evaluating the performance of our image segmentation model for colorectal polyps. The primary metrics include the Intersection over Union (IoU) score, Dice score, and accuracy.

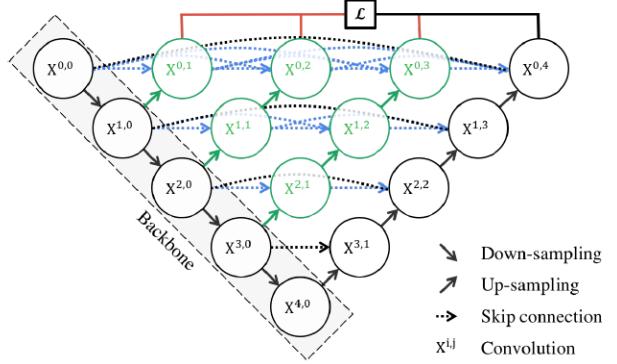


Figure 5. 5-level UNet++ architecture as an encoder-decoder connected through a series of nested dense convolutional blocks. In the graphical abstract, black indicates the original U-Net, green and blue show dense convolution blocks on the skip pathways, and red indicates deep supervision. Image from the Zhou Et. Al. paper [18].

4.5.1 Intersection over Union (IoU) score

The IoU score is a common evaluation metric for image segmentation tasks. It is defined as the ratio of the intersection of the predicted segmentation mask and the ground truth mask to the union of these masks. Mathematically, it is expressed as:

$$\text{IoU} = \frac{|A \cap B|}{|A \cup B|} \quad (1)$$

where A is the predicted segmentation mask and B is the ground truth mask. The IoU score ranges from 0 to 1, with 1 indicating perfect overlap.

4.5.2 Dice score

The Dice score, also known as the Dice coefficient, is another widely used metric for image segmentation. It 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. The formula for the Dice score is:

$$\text{Dice} = \frac{2|A \cap B|}{|A| + |B|} \quad (2)$$

Similar to the IoU score, the Dice score ranges from 0 to 1, with 1 indicating perfect agreement between the predicted and ground truth masks.

We employed the Dice score as our primary loss function during model training. We experimented with alternative metrics such as IoU score, binary cross-entropy, and a weighted combination of BCE with Dice loss. Ultimately, we found that utilizing the Dice loss yielded superior results.

4.5.3 Accuracy

Accuracy is a basic yet important metric for evaluating the performance of an image segmentation model. It is defined as the ratio of the number of correctly predicted pixels to the total number of pixels. The accuracy can be calculated using the following formula:

$$\text{Accuracy} = \frac{\text{Number of correctly predicted pixels}}{\text{Total number of pixels}} \quad (3)$$

Higher accuracy values indicate better performance of the segmentation model.

These metrics provide a comprehensive evaluation of the model's performance in segmenting colorectal polyps, helping to ensure accurate and reliable results.

5. Experiments

In this section, we detail the selected hyperparameters for each model and present their corresponding results. Additionally, we provide a comparative analysis to highlight the importance of data augmentation and the impact of different architectures on performance. The dataset images were resized to 256x256 pixels for all experiments.

5.1. U-Net

Starting with our baseline model, the U-Net, we first trained the model on the original dataset to establish a foundational performance evaluation. Having already mentioned the image sizes earlier, we also tried higher resolutions like 312x312 pixels, but they did not offer substantial improvements and would increase computational cost. The Adam optimizer was used with a batch size of 12, identified as optimal after testing various sizes, and a learning rate of 1e-4. To prevent overfitting, a learning rate scheduler and early stopping were employed.

Subsequently, we trained the model on the augmented dataset using the same hyperparameters. This step was intended to enhance the model's performance by increasing the diversity of the training data. Comparing the results, we observed a significant improvement in performance on the augmented dataset, underscoring the effectiveness of data augmentation. The performance comparison on the set is given in Table 1.

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

Table 1. Performance metrics for the U-Net model on the original and augmented datasets on the test set.

5.2. ResUNet

As defined in our methodology, we utilized the ResUNet model with the same experimental setup as the U-Net. The ResUNet was trained on both the original and augmented datasets using same image size as U-Net. We applied the same hyperparameters and training strategies, including the Adam optimizer, a batch size of 12, a learning rate of 1e-4, a learning rate scheduler, and early stopping. This approach aimed to enhance the model's performance by increasing the diversity of the training data. Comparing the results, we observed a significant improvement in performance on the augmented dataset. The performance comparison is given in Table 2.

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

Table 2. Performance metrics for the ResUNet model on the original and augmented datasets on the test set.

Comparing the non-augmented performance of the U-Net and ResUNet models, we observe that the U-Net achieved a lower test loss. However, the ResUNet outperformed the U-Net in terms of IoU, Dice coefficient, and test accuracy, and required fewer epochs to converge until the early stopping condition was triggered (35 epochs for ResUNet vs. 48 epochs for U-Net).

However, the situation is different for the augmented models. Despite the improvements observed with augmentation, the U-Net augmented model outperformed the ResUNet augmented model in every metric. This suggests that while ResUNet benefits from augmentation, the U-Net model is better able to leverage the augmented data to achieve superior performance. The ResUNet model has a more complex structure compared to the U-Net, which may require better fine-tuning to achieve optimal performance. However, due to the GPU resources available in Colab, we were unable to experiment with different parameters such as larger batch sizes or deeper configurations. This limitation might have impacted the ResUNet's ability to fully leverage the augmented data.

5.3. Attention U-Net

We have trained both 4-level and 5-level versions of the Attention U-Net on original and augmented datasets. After experimenting with a range of hyperparameters, we selected a batch size of 8 for all training runs. Additionally, we utilized the Dice loss function to handle class imbalances and ensure precise segmentation performance. The Adam optimizer was chosen for its efficiency and adaptive learning rate capabilities. To prevent overfitting and ensure optimal model performance, we employed an early-stopping mech-

anism, which monitored the validation loss and halted training when no improvement was observed.

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

Table 3. Performance metrics for different Attention U-Net models on the test set.

The results shown in Table 3 indicate that data augmentation and increased model depth significantly enhance the performance of Attention U-Net models. The Augmented Deeper Attention U-Net demonstrated the best metrics across all evaluation parameters. Overall, combining deeper architecture with data augmentation results in more accurate and robust models. Figure 6 illustrates the evolution of metrics for both 4-level and 5-level Attention UNet models trained on the augmented dataset. The deeper 5-level model shows superior performance with faster convergence.

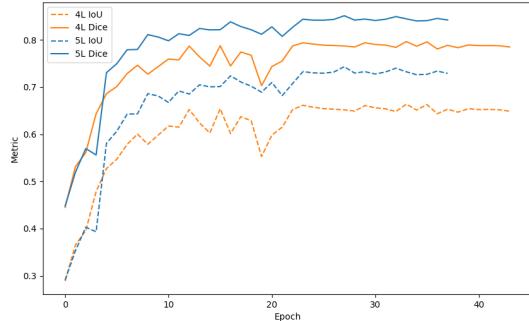


Figure 6. Evolution of Validation set IoU (dashed lines) and Dice (solid lines) scores for 4-level and 5-level Attention UNet models.

5.4. UNet++

We trained the UNet++ model on both the original and augmented datasets. After experimenting with various hyperparameters, we standardized the batch size to 16 for all training sessions. Similar to the Attention U-Net, our approach involved using the Dice loss function, the Adam optimizer, and implementing early stopping.

Additionally, we evaluated the model’s performance under different conditions, comparing the effectiveness of deep supervision versus standard training.

Table 4 reveals significant benefits from data augmentation and deep supervision techniques. Augmented models consistently outperform their non-augmented counterparts across all evaluated metrics. Moreover, deep supervi-

sion significantly enhances model effectiveness compared to models trained without it. These results underscore the critical roles of augmentation strategies and deep supervision in improving the performance of semantic segmentation models like UNet++. It is noteworthy that models with deep supervision tend to exhibit higher loss due to the aggregation of losses at multiple scales, reflecting their comprehensive learning approach.

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

Table 4. Comparison of test performance metrics for UNet++ models with and without deep supervision (DS), trained on augmented and non-augmented datasets.

6. Conclusion

In our study on polyp segmentation using the Kvasir-SEG dataset, we found that the UNet++ model, trained on augmented data with deep supervision (DS), significantly outperformed the baseline U-Net model. The superior IoU and Dice coefficients achieved by the UNet++ model demonstrate its enhanced segmentation accuracy, underscoring the advantages of employing a more advanced architecture and sophisticated training techniques compared to the baseline U-Net model. Figure 7 illustrates the predicted masks by U-Net and UNet++.

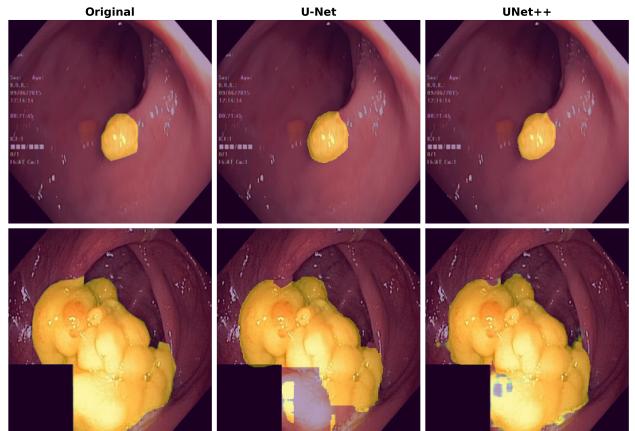


Figure 7. Comparison of original and predicted masks for two polyps. The first column shows the original images, the second column presents the masks predicted by U-Net, and the third column displays the masks predicted by UNet++. The top sample illustrates that UNet++ provides a better fit for small polyps, while the bottom sample shows that U-Net fails to capture part of a large polyp which UNet++ accurately identifies.

The ResUNet model performed worse 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. Ultimately, the UNet++ model emerged as the best performer among all the models evaluated, consistently delivering superior segmentation results.

References

- [1] J. Asplund, J. H. Kauppila, F. Mattsson, and J. Lagergren. Survival trends in gastric adenocarcinoma: A population-based study in sweden. *Annals of Surgical Oncology*, 25(9):2693–2702, Sep 2018. Epub 2018 Jul 9.
- [2] Gilberto Lopes, Mariana C. Stern, Sarah Temin, Ala I. Sharara, Andres Cervantes, Ainhoa Costas-Chavarri, Rena Engineer, Chisato Hamashima, Gwo Fuang Ho, Fidel David Huitzil, Mona Malekzadeh Moghani, Govind Nandakumar, Manish A. Shah, Catherine Teh, Sara E. Vázquez Manjarez, Azmina Verjee, Rhonda Yantiss, and Marcia Cruz Correa. Early detection for colorectal cancer: Asco resource-stratified guideline. *Journal of Global Oncology*, (5):1–22, 2019. PMID: 30802159.
- [3] D. Ha, 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*, 9:40496–40510, Mar 2021.
- [4] Debesh Jha, Pia H. Smedsrød, Michael A. Riegler, Pål Halvorsen, Thomas de Lange, Dag Johansen, and Håvard D. Johansen. Kvasir-seg: A segmented polyp dataset, 2019.
- [5] Ronald A Castellino. Computer aided detection (cad): an overview. *Cancer Imaging*, 5(1):17–19, 2005.
- [6] Stefan Ameling, Stephan Wirth, Dietrich Paulus, Gerard Lacey, and Fernando Vilariño. Texture-based polyp detection in colonoscopy. pages 346–350, 01 2009.
- [7] Sokratis A Karkanis, Dimitris K Iakovidis, Dimitris E Maroulis, Dimitris A Karras, and Michalis Tzivras. Computer-aided tumor detection in endoscopic video using color wavelet features. *IEEE Transactions on Information Technology in Biomedicine*, 7(3):141–152, 2003.
- [8] Nima Tajbakhsh, Jun Y Shin, Suryakanth R Gurudu, Robert T Hurst, Christopher B Kendall, Michael B Gotway, and Jianming Liang. Convolutional neural networks for medical image analysis: Full training or fine tuning? *IEEE Transactions on Medical Imaging*, 35(5):1299–1312, 2016.
- [9] Hoo-Chang Shin, Holger R Roth, Mingchen Gao, Le Lu, Ziyue Xu, Isaac Nogues, Jianhua Yao, Daniel Mollura, and Ronald M Summers. Deep convolutional neural networks for computer-aided detection: Cnn architectures, dataset characteristics and transfer learning. *IEEE Transactions on Medical Imaging*, 35(5):1285–1298, 2016.
- [10] Joan Bernal et al. Comparative validation of polyp detection methods in video colonoscopy: Results from the miccai 2015 endoscopic vision challenge. *IEEE Transactions on Medical Imaging*, 36(6):1231–1249, 2017.
- [11] Sharib Ali et al. An objective comparison of detection and segmentation algorithms for artefacts in clinical endoscopy. *Scientific Reports*, 10(1):1–15, 2020.
- [12] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation, 2015.
- [13] Michael Yeung, Evis Sala, Carola-Bibiane Schönlieb, and Leonardo Rundo. Focus u-net: A novel dual attention-gated cnn for polyp segmentation during colonoscopy. *Computers in Biology and Medicine*, 137:104815, 2021.
- [14] Krushi Patel, Andrés M. Bur, and Guanghui Wang. Enhanced u-net: A feature enhancement network for polyp segmentation. In *2021 18th Conference on Robots and Vision (CRV)*, pages 181–188, 2021.
- [15] Huisi Wu, Zebin Zhao, and Zhaoze Wang. Meta-unet: Multi-scale efficient transformer attention unet for fast and high-accuracy polyp segmentation. *IEEE Transactions on Automation Science and Engineering*, pages 1–12, 2023.
- [16] Zhengxin Zhang, Qingjie Liu, and Yunhong Wang. Road extraction by deep residual u-net. *IEEE Geoscience and Remote Sensing Letters*, 15(5):749–753, May 2018.
- [17] Ozan Oktay, Jo Schlemper, Loic Le Folgoc, Matthew Lee, Matthias Heinrich, Kazunari Misawa, Kensaku Mori, Steven McDonagh, Nils Y Hammerla, Bernhard Kainz, Ben Glocker, and Daniel Rueckert. Attention u-net: Learning where to look for the pancreas, 2018.
- [18] Zongwei Zhou, Md Mahfuzur Rahman Siddiquee, Nima Tajbakhsh, and Jianming Liang. Unet++: A nested u-net architecture for medical image segmentation, 2018.
- [19] Alexander Buslaev, Vladimir I. Iglovikov, Eugene Khvedchenya, Alex Parinov, Mikhail Druzhinin, and Alexandr A. Kalinin. Albumentations: Fast and flexible image augmentations. *Information*, 11(2), 2020.