Segmentation Task On Kvasir-SEG Dataset

Thang Duc Nong[1,2], Tam Trong Nguyen[1,2], Tuan Dat Nguyen[1,2],
Tai Huu Le[1,2], Dang Duy Huynh Tran[1,2], Thien Bao Tat Nguyen[1,2]

[1] University of Information Technology, Ho Chi Minh City, Vietnam.
[2] Vietnam National University, Ho Chi Minh City, Vietnam.

Contributing authors: 21522593@gm.uit.edu.vn; 21521405@gm.uit.edu.vn;
21522754@gm.uit.edu.vn; 21522562@gm.uit.edu.vn; 21521922@gm.uit.edu.vn;
thienntb@uit.edu.vn;

Abstract

In this study, we address the critical task of polyp segmentation in medical imaging, which is essential for the early detection and treatment of colorectal cancer. We explore and compare the performance of several convolutional neural network (CNN) based models, including SegNet, DuckNet, and U-Net, along with various U-Net variants. To enhance the quality of the input images, we employ preprocessing techniques such as specular highlight removal. Our study utilizes Kvasir-SEG dataset - a comprehensive dataset, annotated by medical experts, to train and validate our models. Performance metrics such as Dice coefficient, Intersection over Union (IoU), Sensitivity and Specificity are utilized to quantify the segmentation accuracy. Our experimental results demonstrate that DuckNet and various U-Net variants consistently achieve the highest performance, providing superior segmentation accuracy and robustness. This study contributes to the ongoing efforts in improving automated medical image analysis and highlights the potential of advanced CNN architectures and image enhancement techniques in enhancing diagnostic workflows.

Keywords: Segmentation, CNN, Medical image, Polyp

1 Introduction

Colorectal cancer (CRC) is the world's third-most common cancer, causing thousands of deaths every year. In 2022, more than 1.9 million cases were diagnosed. CRC is the second most common cause of cancer death, accounting for over 900,000 deaths per year (9.3% of total cancer deaths)[1]. The majority of polyps found in human body are located in nose, colon, uterus, and stomach. Researchers believe that polyps are mostly harmless. However, many studies have found that certain types of polyps can be dangerous and lead to cancer. Colonoscopy is the primary method for screening and preventing polyps from developing into colorectal cancer. However, detecting polyps through colonoscopy requires highly skilled endoscopists and high level of eye-hand coordination. Recent studies have shown that 25% of polyps were missed during colonoscopy[2]. Segmenting polyps is a crucial task in the early diagnosis of colon polyps for preventing colorectal cancer. The size of the segmented polyps directly impacts the miss rate in colonoscopy. Large polyps mostly can be detected easily, while small polyps, which are tiny and difficult to see, account for most of the miss rate. Different methods have been proposed with the goal of accurate polyp segmentation. There are three different groups of approaches for polyp segmentation that already exist. The first approach uses image processing-based segmentation without using any learning method. The second group uses methods that first extract features, then use classifiers for segmentation. The last group belongs to methods using convolutional neuronal networks (CNN) to perform the segmentation work.

In this study, we explore and evaluate the performance of CNN-based models such as SegNet[3], U-Net[4], and DuckNet[5], along with various U-net variants for segmentation and detection tasks. To improve the performance of the models, we apply preprocessing to enhance the quality of the input images using specular highlight removal method. We validate the performance of the presented models using the Kvasir-SEG dataset[6].

The rest of our study can be summarized as follows. Section 2 provides the related work in the literature on colorectal cancer and colonoscopy image preprocessing and polyp segmentation approaches. Section 3 provided a detailed description of Kvasir-Seg dataset. Section 4 presents data preprocessing methods and architectures of models. Section 5 describes evaluation metrics, results, and experimental analysis. Finally, Section 6 concludes the paper and discusses our study.

2 Related works

2.1 Preprocessing Methods for Colonoscopy Images

Image quality is one factor that impacts the performance of polyp detection. Endoscopic images can be affected by several unwanted factors such as the performing physician's skill, equipment limitations, and certain environmental conditions. Preprocessing is important in enhancing image quality and increasing results' accuracy in endoscopic image studies. Processing endoscopic images introduces common challenges, such as black masks, ghost colors, interlacing, specular highlights, and uneven lighting. Specular highlights, the bright spots reflecting off tumors or polyps in captured images, create significant challenges for algorithms. Mitigating their impact can be effectively achieved through detection or inpainting methods designed to eliminate these highlights. In their study, the authors Sánchez-González, A., & Soto, B. G. Z. (2017).[7] summarized the most commonly used preprocessing methods for endoscopic images such as the removal of ghost colors, removal of interlacing effects, lighting normalization, removal of specular highlights, removal of black borders. The effectiveness of these preprocessing steps is demonstrated by the improved performance of the model, with six out of the eight vision models achieving better F1-Score. Thai, Triet M., et al.,[8] utilized several preprocessing techniques in their study such as removing specular highlights and black borders and achieving promising results for their research on MEDVQA for the gastrointestinal tract. Overall, the enhancement process helps to improve the F1-Score by at least 0.4% and up to 1.11% on VQA performance.

2.2 Deep learning for polyp segmentation

The integration of deep learning models for medical image processing has achieved impressive results compared to the use of traditional image processing techniques. These advancements are particularly evident in the tasks of polyp detection and segmentation, where deep learning models are applied to endoscopic images. By leveraging these advanced models, the accuracy in identifying and segmenting polyps has increased significantly. Moreover, deep learning techniques have enabled more robust and reliable automated analysis, reducing the dependence on manual interpretation and thereby improving diagnostic efficiency. Guo, Y., Bernal, J., & J. Matuszewski, B.[9] proposed a polyp segmentation algorithm, inspired by FCN, DeepLab, and Global Convolutional networks, developed based on fully convolutional network models, that was originally developed for the Endoscopic Vision Gastrointestinal Image Analysis (GIANA) polyp segmentation challenges. This algorithm, based on fully convolutional network models, involved examining various network configurations, design parameters, data augmentation approaches, and polyp characteristics. Notably, the study highlighted the significance of data augmentation and careful parameter selection. Consequently, the method achieved state-of-the-art results with near real-time performance. As a result, it secured the top spot in the polyp segmentation sub-challenge at the 2017 GIANA challenge and second place in the standard image resolution segmentation task at the 2018 GIANA challenge. Hosseinzadeh Kassani, S., et al.[10] conducted a study comparing the performance of various deep learning architectures, including ResNet, DenseNet, InceptionV3, InceptionResNetV2, and SE-ResNeXt, as feature extractors in the encoder part of a U-Net architecture. Their findings indicated that the DenseNet169 feature extractor, when combined with the U-Net architecture, outperformed the other architectures. Safarov, S., and Whangbo, T. K.[11] developed a fully automated polyp segmentation model named A-DenseUNet, which aims to assist endoscopists in identifying colorectal disease more efficiently. The model is designed to adapt to the unknown depth of the network by sharing multiscale encoding information across different levels of the decoder side. It employs multiple dilated convolutions with various atrous rates to maintain a large field of view without increasing computational cost and to prevent loss of spatial information that would lead to dimensionality reduction. The A-DenseUNet model achieved a 90% Dice coefficient score on the Kvasir-SEG dataset and a 91% Dice coefficient score on the CVC-612 dataset. It outperformed other models such as UNet++, ResUNet, U-Net, PraNet, and ResUNet++ in polyp segmentation tasks.

3 Data

The Kvasir-SEG dataset builds on the original Kvasir dataset, the pioneering multi-class dataset for detecting and classifying gastrointestinal (GI) tract diseases

3.1 The original Kvasir dataset

3.1.1 Data collection

The data, collected using equipment shown in Figure 1(c) at Vestre Viken Health Trust (VV) in Norway, comes from four hospitals serving 470,000 people. One of these, Bærum Hospital, has a large gastroenterology department that provides training data, which will expand the dataset over time. Images are meticulously annotated by medical experts from VV and the Cancer Registry of Norway (CRN). CRN, part of the South-Eastern Norway Regional Health Authority and affiliated with Oslo University Hospital Trust, conducts cancer research and manages national cancer screening programs to detect cancers or pre-cancerous lesions early, aiming to prevent cancer deaths

(a) Colonoscopy (b) Gastroscopy (c) A colonoscope

Fig. 1: Various types of endoscopy examinations

3.1.2 Dataset details

The original Kvasir dataset contains 8,000 images annotated and verified by experienced endoscopists, divided into 8 classes of 1,000 images each, depicting anatomical landmarks, pathological findings, and endoscopic procedures in the GI tract. To enhance quality, 13 polyp images were replaced. This dataset supports various tasks such as image retrieval, machine learning, deep learning, and transfer learning. Anatomical landmarks include the Z-line, pylorus, and cecum; pathological findings cover esophagitis, polyps, and ulcerative colitis. It also contains images of polyp removal procedures (”dyed and lifted polyp” and ”dyed resection margins”).

First released in 2017, the Kvasir dataset was used in the Multimedia for Medicine Challenge at the MediaEval Benchmarking Initiative to develop methods for multiclass classification of endoscopic findings in the large bowel. Due to frame-wise annotations, the dataset was limited to frame classification.

The dataset includes images with resolutions ranging from 720x576 to 1920x1072 pixels, organized in folders by content. Some classes feature a green picture-in-picture showing the endoscope's position using an electromagnetic imaging system (ScopeGuide, Olympus Europe), which aids image interpretation but requires careful handling for endoscopic findings detection.

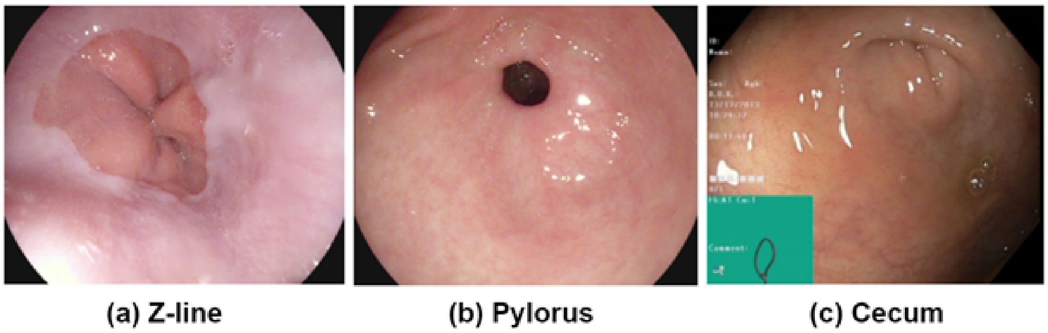
Anatomical Landmarks

An anatomical landmark in the GI tract is a recognizable feature visible through an endoscope, crucial for navigation and as a reference point for findings. These landmarks can also be common sites for pathology like ulcers or inflammation. A thorough endoscopic report should include descriptions and images of key anatomical landmarks.

Z-line. The Z-line marks the transition between the esophagus and stomach, visible as a border where the white esophageal mucosa meets the red gastric mucosa in Figure 2(a). Recognizing the Z-line is important for diagnosing diseases such as gastro-esophageal reflux and serves as a reference point for esophageal pathology.

Pylorus. The pylorus is the area around the opening from the stomach to the duodenum, regulated by circumferential muscles. Identifying the pylorus is necessary for endoscopic navigation to the duodenum. A complete gastroscopy inspects both sides of the pyloric opening for ulcers, erosions, or stenosis. Figure 2(b) shows a normal pylorus, appearing as a smooth, round opening surrounded by pink mucosa.

Cecum. The cecum is the most proximal part of the large bowel, and reaching it confirms a complete colonoscopy, a quality indicator. The cecum's key feature is the appendiceal orifice, which, along with the electromagnetic scope tracking system's configuration, proves cecal intubation when documented. Figure 2(c) shows the appendiceal orifice as a crescent-shaped slit, with a green picture-in-picture indicating the scope's position.



(a) Z-line (b) Pylorus (c) Cecum

Fig. 2: Sample of Anatomical Landmarks

Phatological findings

A pathological finding in the gastrointestinal tract is an abnormal feature visible through an endoscope, indicating damage or changes in the normal mucosa. These findings may signal ongoing disease or potential precursors to conditions like cancer. Detecting and classifying these pathologies is crucial for initiating appropriate treatment and follow-up.

Esophagitis. Esophagitis is an inflammation of the esophagus, seen as breaks in the mucosa near the Z-line. Figure 3(a) shows red mucosal tongues projecting into the white esophageal lining. The inflammation grade is determined by the length of mucosal breaks and the proportion of the circumference involved, commonly caused by gastroesophageal reflux, vomiting, or hernia. Early detection is essential for treatment to relieve symptoms and prevent complications. Computer detection could help assess severity and automate reporting.

Polyps. Polyps are mucosal outgrowths in the bowel, appearing as flat, elevated, or pedunculated lesions distinguishable by color and surface pattern. Figure 3(b) illustrates a typical polyp. While most are harmless, some can develop into cancer, making detection and removal vital for preventing colorectal cancer. Since polyps can be missed during examinations, automatic detection could enhance examination quality. The green boxes in the image illustrate the endoscope's configuration, aiding in locating the polyp within the bowel. Computer-aided detection would improve diagnosis, assessment, and reporting.

Ulcerative Colitis. Ulcerative colitis is a chronic inflammatory disease of the large bowel, significantly affecting quality of life. Diagnosis relies on colonoscopic findings, with inflammation ranging from mild to severe. Mild disease presents with swollen, red mucosa, while moderate cases show prominent ulcerations. Figure 3(c) shows ulcerative colitis with bleeding, swelling, and ulceration of the mucosa, with a white fibrin coating over the wounds. An automatic assessment system would enhance the accuracy of grading disease severity.

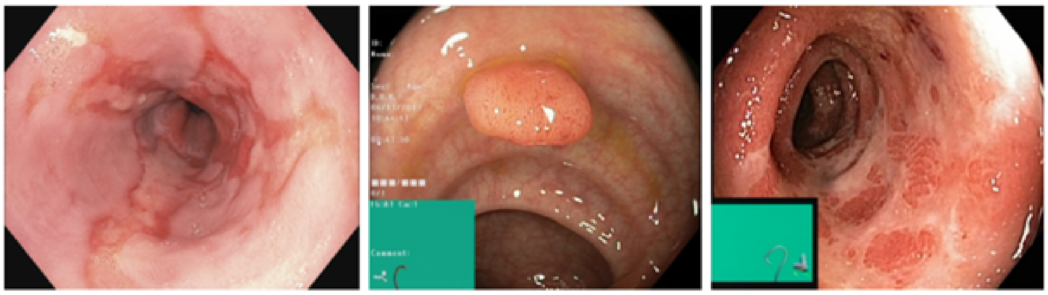

(a) Esophagitis (b) Polyp (c) Ulcerative colitis

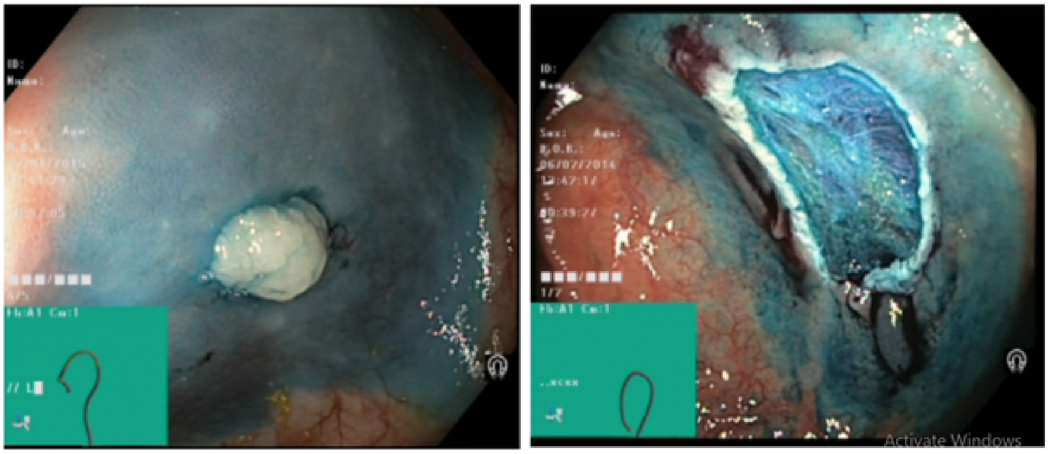
Fig. 3: Sample of Phatological findings

Polyp removal

Polyps in the large bowel can be precursors to cancer and are typically removed during endoscopy. One common technique is endoscopic mucosal resection (EMR), which involves injecting liquid beneath the polyp to lift it from the underlying tissue before removing it with a snare. This lifting reduces the risk of damage to deeper layers of the GI wall. Staining dye, such as diluted indigo carmine, is used to highlight the polyp margins for accurate removal. Computer detection of dyed polyps and resection sites is crucial for developing computer-aided reporting systems.

Dyed and Lifted Polyps. Figure 4(a) shows a polyp lifted with saline and indigo carmine, with its light blue margins clearly visible against the darker mucosa. Automatic reporting could include the success of the lift and the presence of non-lifted areas, which might indicate malignancy.

Dyed Resection Margins. Evaluating resection margins is vital to ensure complete polyp removal, as residual tissue can lead to regrowth and potentially malignancy. Figure 4(b) illustrates the resection site post-polyp removal. Automatic recognition of these sites would be valuable for reporting systems and assessing the completeness of polyp removal.



(a) Dyed and Lifted Polyp (b) Dyed Resection Margin

Fig. 4: Sample of Polyp removal

3.2 The Kvasir-SEG Dataset Details

To address the high incidence of colorectal cancer, they focused on the polyp class from the Kvasir dataset for initial investigation. The segmented version, Kvasir-SEG, was published in 2020 and contains annotated polyp images and corresponding masks, totaling 46.2 MB.
The Kvasir-SEG dataset includes two folders: one for images and one for masks, each containing 1,000 images with resolutions ranging from 332x487 to 1920x1072 pixels. Bounding boxes for the images are stored in a JSON file. Thus, the dataset consists of an image folder, a masks folder, and a JSON file. Each image and its corresponding mask. share the same filename and are encoded using JPEG compression, allowing for online browsing.

3.3 Mask Extraction

The entire Kvasir polyp class was uploaded to Labelbox for segmentation. Labelbox, a tool for labeling regions of interest (ROI) in images, was used to outline the polyp regions. An engineer and a medical doctor manually annotated the margins of all polyps in 1,000 images, with the annotations reviewed by an experienced gastroenterologist. Figure 5 shows example frames with polyps marked in green.

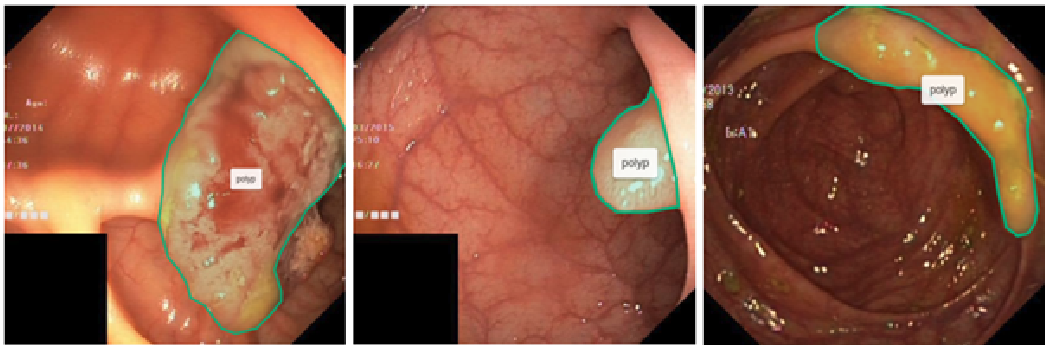


Fig. 5: Example frames from the Kvasir dataset with polyp tissue marked by green outlines

After annotation, the files were exported to create masks for each image. The exported JSON file contained details about the image and the coordinates for generating the masks. Using these coordinates, contours were drawn on a black background and filled with white to create 1-bit depth masks. Figure 6 shows example images, their segmentation masks, and bounding boxes from the Kvasir-SEG dataset. The white areas represent polyp regions, while the black background represents non-polyp tissue. Some original images contained endoscope position markers from ScopeGuide (Olympus), shown as small green boxes in one of the bottom corners. These markers, irrelevant to segmentation, were replaced with black boxes in the Kvasir-SEG dataset.

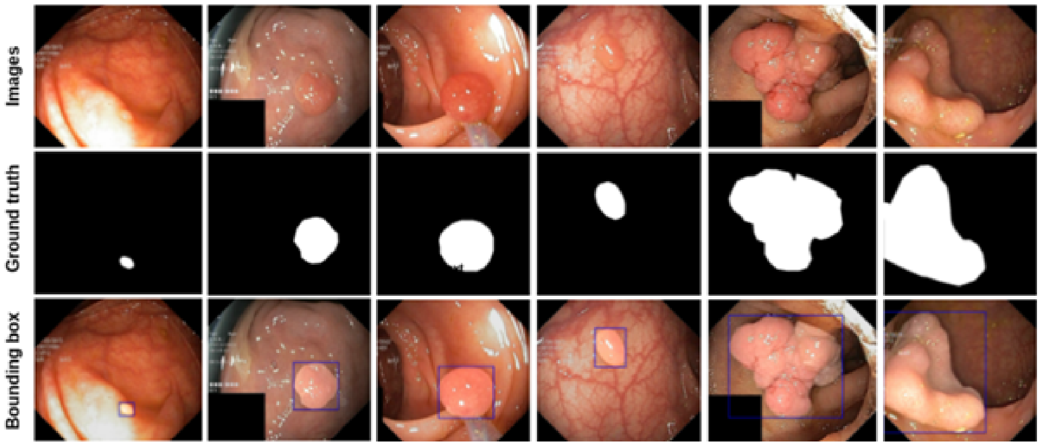


Fig. 6: Examples of polyp images, their corresponding masks and bounding box from Kvasir-SEG

3.4 Applications of the Dataset

The Kvasir-SEG dataset is designed for researching and developing advanced methods for polyp segmentation, detection, localization, and classification. It serves as both a training and validation dataset, enabling comparisons of computer vision algorithms. This dataset aids in creating state-of-the-art solutions for colonoscope images from various manufacturers, potentially reducing polyp miss rates and improving examination quality. Additionally, Kvasir-SEG is suitable for general segmentation and bounding box detection research, complementing datasets from various fields, both medical and beyond.

4 Experimental methods

4.1 Data preprocessing

In medical endoscopic images, the quality of images plays a crucial role in identifying the pathology of machine learning models. However, despite remarkable advancements in endoscopic technology, images are still often affected by specularities, causing undesired bright spots that obscure important details and degrade image quality. Removing these specular reflections helps improve image sharpness, contrast, balances brightness, and enables machine learning models to easily identify disease points.

The images of polyps in the Kvasir dataset are no exception. In Figure 7., we can observe images containing numerous specular and imbalanced lighting.

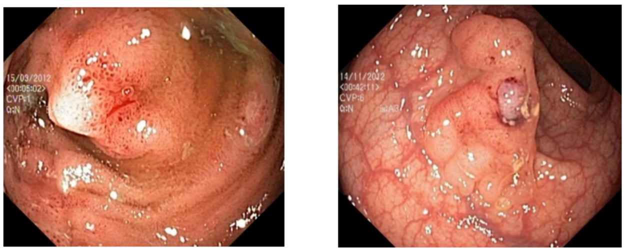


Fig. 7: The images contain many glare points in the Kvasir dataset

To address this issue, we combine various preprocessing methods such as GaussianBlur, expanding glare spots, etc., to optimize the removal of specular, avoid noise in the images, and make the images smoother.

- **Specular highlight detection:** Reading images from the dataset in two forms: color images and grayscale images, and resizing them to 256x256, we compute the intensity values $\boldsymbol{m}$ and the saturation $\boldsymbol{s}$ of each pixel. Then, we assign these values to the grayscale image according to the formula:
  - Brightness intensity:

$$m[\mathrm{ri}, \mathrm{ci}] = \mathrm{round}\left(\frac{R+G+B}{3}\right) \tag{1}$$

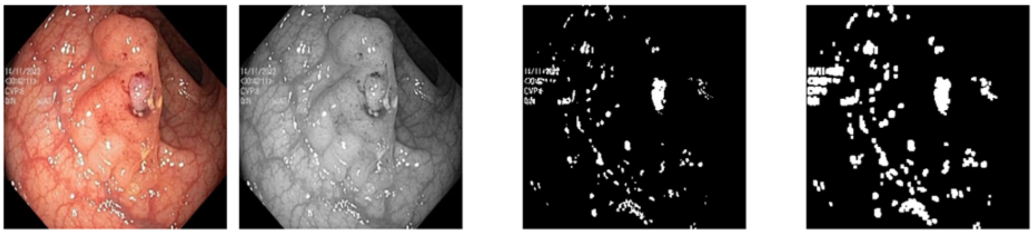
  - And saturation:

$$\begin{aligned} s_1 &= \mathrm{clip}(B+R, 0, 255) && (2)\\ s_2 &= 2 \times G && (3)\\ \text{If } s_1 \geq s_2: \quad \mathrm{s}[\mathrm{ri}, \mathrm{ci}] &= 1.5 \times \mathrm{clip}(R - m[\mathrm{ri}, \mathrm{ci}], 0, 255) && (4)\\ \text{If } s_1 < s_2: \quad \mathrm{s}[\mathrm{ri}, \mathrm{ci}] &= 1.5 \times \mathrm{clip}(m[\mathrm{ri}, \mathrm{ci}] - B, 0, 255) && (5) \end{aligned}$$

  - Where:
    - R, G, B respectively represent the Red, Green, and Blue color components of a pixel in the original image.
    - $\mathrm{s}[\mathrm{ri}, \mathrm{ci}]$ denotes the computed saturation value for the pixel at row index i and column index i.
    - $m[\mathrm{ri}, \mathrm{ci}]$ denotes the computed brightness intensity value for the pixel at row index i and column index i.
    - The round function is used to round the resulting value to the nearest integer.
    - The $\mathrm{clip}(x, a, b)$ function is used to constrain the value of x within the range $[a, b]$.

  After computing the intensity and saturation for each pixel, we create a copy of the grayscale image to store images containing highlight pixels. Next, we check whether each pixel is a highlight pixel by setting separate thresholds and comparing them with the intensity and saturation values of each pixel: the intensity threshold $\boldsymbol{m}$ is half of the maximum intensity value, and the saturation threshold $\boldsymbol{s}$ is one-third of the maximum saturation value. If the brightness intensity of a pixel is greater than or equal to the intensity threshold and the saturation of the pixel is less than or equal to the saturation threshold, then that pixel is considered a specular highlight pixel.
- **Specular enlarged:** To ensure a smooth image processing, improve brightness balance, and enhance the handling of highlights while avoiding halo effects, we perform specular enlargement by moving a 3x3 pixel window over the image. If this window contains a specular highlight pixel at the center position, all pixels within that window are marked as specular highlight pixels.
- **Brightness balance and smoothing:** Starting from the original image, we create a smoothed image using a Gaussian filter. Then, we replace the expanded highlight regions in the original image with smoothed pixels, reducing the glare effect and balancing the brightness across the image.
- **Telea inpainting:** Finally, we employ the Telea algorithm[12], an effective inpainting technique widely used in various image processing applications, to handle the smoothed specular highlight regions in the image.

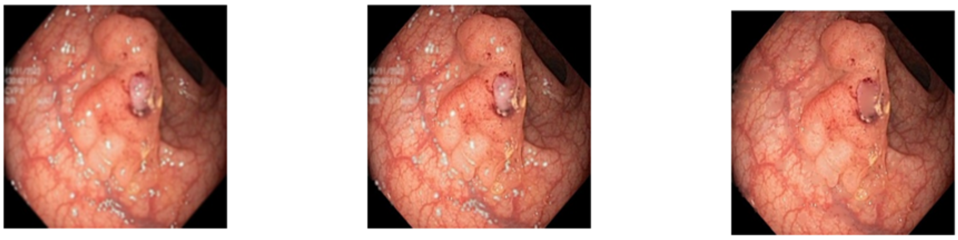
The final image has undergone processing to handle highlight pixels, smoothing, and brightness balancing. For details on the preprocessing steps, please refer to Figure 8.



1. Original and greyscale image

2. Specular highlight detection

3. Specular enlarging



4. Gaussian Blur image

5. Restore original image

6. Utilize the Telea algorithm for image inpainting

Fig. 8: The process of specular highlight removal

4.2 Models

4.2.1 SegNet

SegNet is an image segmentation architecture that uses an **encoder-decoder** type of architecture. It is designed to take an image as input and produce a pixel-wise label map as output. The encoder captures high-level features by applying convolutional and pooling layers. The key innovation in SegNet is its decoder network, which uses the spatial pooling indices generated during the max-pooling in the encoder phase to upsample and produce a segmentation map. This form of up-sampling is both memory-efficient and helps preserve the fine-grained details in the output.

- **Encoder Network:** Contains 13 convolutional layers, derived from the VGG16 network, without fully connected layers to retain higher resolution feature maps. Each encoder layer includes convolution, batch normalization, ReLU activation, and 2x2 max-pooling.
- **Decoder Network:** Uses stored max-pooling indices to up-sample feature maps, followed by convolution with a trainable decoder filter bank and batch normalization. The final decoder layer outputs a high-dimensional feature map.

4.2.2 UNet

UNet is a neural network architecture developed in 2015 for biomedical image segmentation, known for its U-shaped encoder-decoder structure. It enhances semantic segmentation tasks by efficiently learning from fewer training samples.

- **Encoder Network:** Extracts features using a sequence of blocks with two 3x3 convolutions followed by ReLU activation and a 2x2 max-pooling to halve spatial dimensions and reduce computational costs.
- **Skip Connections:** Link corresponding encoder and decoder blocks, providing additional information and aiding gradient flow during backpropagation.
- **Bridge:** Connects the encoder and decoder networks with two 3x3 convolutions followed by ReLU activations.
- **Decoder Network:** Reconstructs the segmentation mask by up-sampling through transpose convolutions, concatenating with skip connections, and applying two 3x3 convolutions followed by ReLU activations. The final layer uses a 1x1 convolution with sigmoid activation to produce the segmentation mask.

This architecture allows for detailed and precise segmentation, especially useful in biomedical applications.

4.2.3 DuckNet

DuckNet is a convolutional neural network designed for polyp image segmentation, based on the U-Net architecture but with several significant modifications. Here are the key aspects of the architecture:

- **Encoder-Decoder Structure:** The overall architecture follows the encoder-decoder structure of U-Net, commonly used for image segmentation tasks.
- **Duck Blocks:** Instead of the traditional 3x3 convolutional blocks used in U-Net, DuckNet employs a novel Duck block at each step, except for the last one. The Duck block is designed to capture more detailed features while compromising on finer low-level details. Duck Blocks includes components:
  - **Residual Block:** Simulates kernel sizes of 5x5, 9x9, and 13x13 using combinations of one, two, and three Residual blocks, respectively.
  - **Midscope and Widescope Blocks:** Utilize dilated convolutions to simulate larger kernels (7x7 and 15x15) while reducing the number of parameters. These blocks aim to capture higher-level features.
  - **Separated Block:** Combines 1xN and Nx1 kernels to simulate NxN behavior, addressing the concept of ”diagonality” to maintain spatial details related to diagonal patterns in the image.

The DuckNet architecture aims to improve the accuracy and efficiency of polyp segmentation by leveraging advanced convolutional techniques and maintaining essential image details throughout the processing steps.

4.2.4 UNet Xception

UNet Xception is a UNet Xception-style model that incorporates depthwise separable convolutions from the Xception architecture into the U-Net framework. This integration can enhance the performance of U-Net by improving its efficiency and potentially its accuracy. The main changes are:

- **Encoder:** Uses Xception-style depthwise separable convolutions instead of traditional convolutions to capture more complex features with fewer parameters.
- **Decoder:** Similarly, it may use depthwise separable convolutions to reconstruct the high-resolution segmentation map.

4.2.5 UNet 3+

UNet 3+ is an advanced version of the traditional U-Net architecture designed to address some of its limitations and enhance its performance for image segmentation tasks. It introduces several innovations that improve feature extraction and segmentation accuracy.

- **Full-Scale Skip Connections:** Integrates low-level details with high-level semantics from feature maps across different scales to capture fine and coarse details.
- **Deep Supervision:** Uses hierarchical representations and side outputs at each decoder stage, connected to a hybrid loss function for improved accuracy.

5 Experimental results

5.1 Evaluation Metrics

We briefly review several popular metrics for polyp segmentation tasks, i.e., Dice coefficient (Dice)[13], Intersection over Union (IoU)[7], Sensitivity, and Specificity.

First, let’s introduce some parameters. TP represents True Positives, which indicates the number of positive samples that the model correctly predicts as positive. FP represents False Positives, which represents the number of negative samples that the model incorrectly predicts as positive. FN represents False Negatives, which represents the number of positive samples that the model incorrectly predicts as negative.

- **Dice:** Dice coefficient, also known as the F1 score, is a measure of the overlap between two sets, with a range of 0 to 1. A value of 1 indicates a perfect overlap, while 0 indicates no overlap.

$$\text{Dice} = \frac{2 \cdot \text{TP}}{2 \cdot \text{TP} + \text{FP} + \text{FN}}$$

- **IoU:** Intersection over Union, also known as the Jaccard coefficient, is used to measure the overlap between the predicted result and the ground truth target. It is commonly used in handling imbalanced datasets.

$$\text{IoU} = \frac{\text{TP}}{\text{TP} + \text{FP} + \text{FN}}$$

- **Sensitivity**: Sensitivity, also known as Recall or true positive rate, measures the proportion of true positive predictions among all actual positive instances.

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

- **Specificity**: Specificity measures the model's ability to recognize negative samples. Specifically, Specificity can be calculated using the following formula:

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}}$$

5.2 Result

We evaluated 5 models SegNet, UNet, DuckNet, UNet Xception, and UNet 3+ using 4 prior metrics on both processed and unprocessed datasets. The results are shown in Table 1

Unpreprocessed:

Model	**DICE**	**IOU**	**SENS**	**SPEC**
SegNet	0.5510	0.4171	0.6661	0.9216
UNet	0.4367	0.3137	0.6464	0.8493
DuckNet	0.4961	0.3706	0.7895	0.8322
UNet Xception	0.7499	0.6519	0.7706	**0.9762**
UNet 3+	**0.8211**	**0.7444**	**0.8630**	0.9734

Preprocessed:

Model	**DICE**	**IOU**	**SENS**	**SPEC**
SegNet	0.6285	0.5032	0.6947	0.9508
UNet	0.4290	0.3075	0.6880	0.8150
DuckNet	0.7498	0.6599	0.7577	0.9789
UNet Xception	0.5986	0.4871	0.6540	0.9609
UNet 3+	**0.8648**	**0.8023**	**0.8776**	**0.9855**

Table 1: Performance of segmentation models on Kvasir datasets

In general, the UNet and UNet Xception models perform better on the unprocessed dataset, whereas the SegNet, DuckNet, and UNet 3+ models show improved results on the preprocessed dataset.

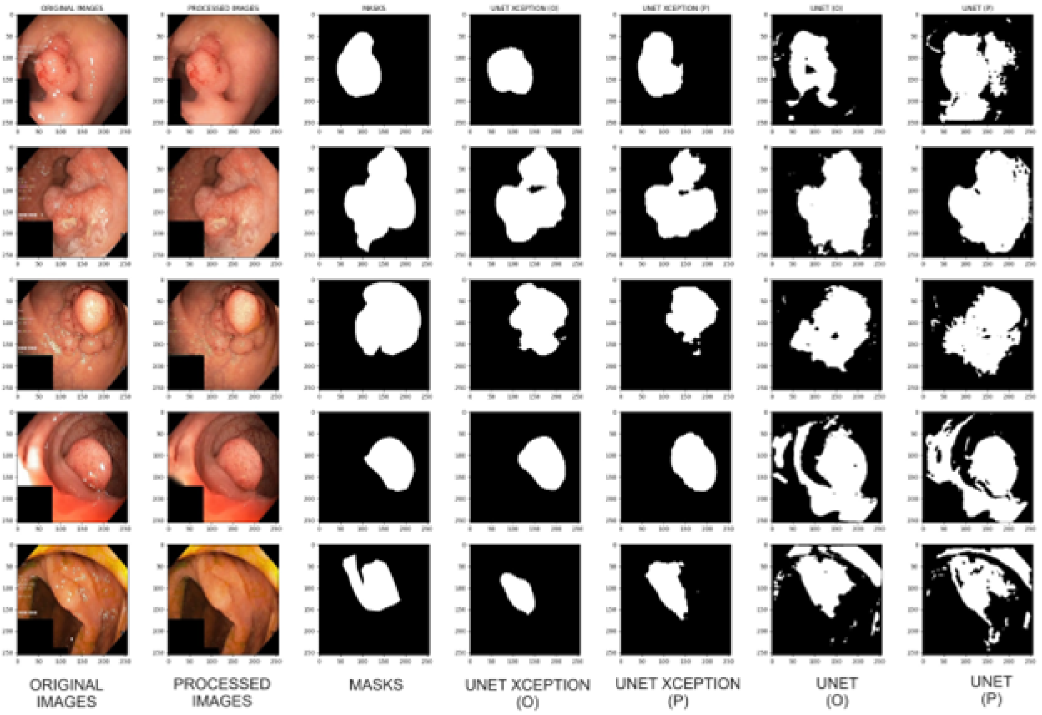


Fig. 9: Some prediction results of UNET and UNET XCEPTION

According to Figure 9, UNet shows the lowest performance in this setting, particularly in terms of IoU and Dice, indicating poor segmentation accuracy. Additionally, UNet mispredicts many negative pixels as positive, as evidenced by its low specificity (SPEC).

Looking at Figure 10, like UNet, DuckNet trained on the original dataset (DuckNet(O)) also incorrectly predicts many negative pixels as positive. However, DuckNet trained on the processed dataset (DuckNet(P)) performs very well, showing significant improvement in its predictions. SegNet is less likely to mispredict negative pixels as positive, as indicated by its high specificity (SPEC), but it also misses many positive pixels, which is reflected in its low sensitivity (SENS).

UNet 3+ consistently outperforms other models in both unprocessed and preprocessed datasets, making it the most robust model in this comparison. UNet Xception and DuckNet also perform well, particularly after preprocessing for DuckNet. SegNet benefits significantly from preprocessing but still lags behind the top models. UNet shows the least improvement or even a slight decline in performance after preprocessing, indicating potential issues with either the model's architecture or its compatibility with the preprocessing techniques used.

This analysis highlights the importance of choosing appropriate preprocessing techniques for different models, as preprocessing can have a positive impact on some models while negatively affecting others.

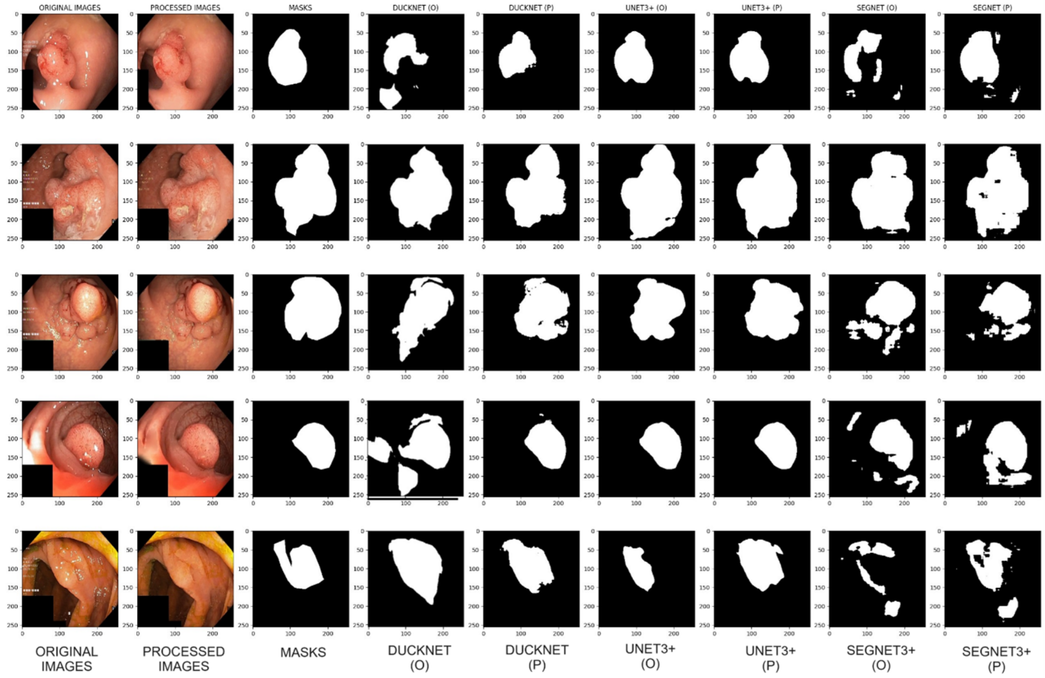


Fig. 10: Some prediction results of SEGNET, DUCKNET and UNET3+

6 Conclusion

Our investigation into the performance of various CNN-based models, including SegNet, DuckNet, and U-Net variants, for polyp segmentation in medical imaging has yielded significant insights. By incorporating preprocessing techniques such as specular highlight removal, we enhanced the quality of input images, leading to improved segmentation accuracy. The comprehensive evaluation, based on metrics like Dice coefficient, Intersection over Union (IoU), Sensitivity and Specificity, demonstrated that DuckNet and the U-Net variants consistently outperform other models, offering superior accuracy and robustness. However, some models, such as traditional U-Net and SegNet, did not perform as well, highlighting the need for more advanced architectures and preprocessing techniques.

These results highlight the effectiveness of DuckNet and U-Net variants in enhancing polyp segmentation, which is critical for the early detection and treatment of colorectal cancer. This underscores the importance of leveraging advanced network architectures alongside effective image enhancement techniques in medical image analysis. Future research will aim to further refine these models and explore their application to other types of medical imaging and segmentation tasks. Additionally, efforts will be made to investigate the real-time deployment and integration of these models into clinical workflows to improve diagnostic efficiency and patient outcomes.

References

[1] Bray, F., Laversanne, M., Sung, H., Ferlay, J., Siegel, R.L., Soerjomataram, I., Jemal, A.: Global cancer statistics 2022: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: a cancer journal for clinicians **74**(3), 229–263 (2024)

[2] Leufkens, A., Van Oijen, M., Vleggaar, F., Siersema, P.: Factors influencing the miss rate of polyps in a back-to-back colonoscopy study. Endoscopy, 470–475 (2012)

[3] Badrinarayanan, V., Kendall, A., Cipolla, R.: Segnet: A deep convolutional encoder-decoder architecture for image segmentation. IEEE transactions on pattern analysis and machine intelligence **39**(12), 2481–2495 (2017)

[4] Krithika Alias AnbuDevi, M., Suganthi, K.: Review of semantic segmentation of medical images using modified architectures of unet. Diagnostics **12**(12), 3064 (2022)

[5] Dumitru, R.-G., Peteleaza, D., Craciun, C.: Using duck-net for polyp image segmentation. Scientific reports **13**(1), 9803 (2023)

[6] Jha, D., Smedsrud, P.H., Riegler, M.A., Halvorsen, P., De Lange, T., Johansen, D., Johansen, H.D.: Kvasir-seg: A segmented polyp dataset, 451–462 (2020). Springer

[7] Rahman, M.A., Wang, Y.: Optimizing intersection-over-union in deep neural networks for image segmentation, 234–244 (2016). Springer

[8] Thai, T.M., Vo, A.T., Tieu, H.K., Bui, L.N., Nguyen, T.T.: Uit-saviors at medvqa-gi 2023: Improving multimodal learning with image enhancement for gastrointestinal visual question answering. arXiv preprint arXiv:2307.02783 (2023)

[9] Guo, Y., Bernal, J., J. Matuszewski, B.: Polyp segmentation with fully convolutional deep neural networks—extended evaluation study. Journal of Imaging **6**(7), 69 (2020)

[10] Hosseinzadeh Kassani, S., Hosseinzadeh Kassani, P., Wesolowski, M.J., Schneider, K.A., Deters, R.: Automatic polyp segmentation using convolutional neural networks, 290–301 (2020). Springer

[11] Safarov, S., Whangbo, T.K.: A-denseunet: Adaptive densely connected unet for polyp segmentation in colonoscopy images with atrous convolution. Sensors **21**(4), 1441 (2021)

[12] Telea, A.: An image inpainting technique based on the fast marching method. Journal of graphics tools **9**(1), 23–34 (2004)

[13] Shamir, R.R., Duchin, Y., Kim, J., Sapiro, G., Harel, N.: Continuous dice coefficient: a method for evaluating probabilistic segmentations. arXiv preprint arXiv:1906.11031 (2019)