**A study on Gastrointestinal Disease Segmentation and localization Using Deep Learning**

**By**

**R L Chandrakantha**

**EU/IS/2016/PHY/074**

**A project report submitted in fulfillment of the**

**Degree of Bachelor of Science Honours in Computer Science**



**Department of Computing**

**Faculty of Science**

**Eastern University**

**Sri Lanka**

**March 2024**

# DECLARATION

I hereby declare that the entire work incorporated in this report is genuinely fulfilled by me, my own. Further, I confirm that this work was done wholly or mainly as a candidate for a bachelor's degree in order to complete it and the work is original and has not been submitted in part or full for any Diploma or Degree in this or any other University. The extent of information derived from the existing literature has been given appropriate citations and credited at the appropriate places and all main sources of help for the work have been acknowledged by me. I also assure you that there are no burglary accusations in this report and if detected I take hold of accountability for that action.

……...……....………… ……………………..

R L Chandrakantha Date

(Research Candidate)

Certification of Supervisor:

……...……....………… ……………………..

Mr. S. Sotheeswaran Date

(Research Supervisor)

# ACKNOWLEDGEMENT

This section inadequately conveys my appreciation for the individuals who have supported me throughout the completion of this work. Without their guidance and encouragement, the successful conclusion of this project would not have been possible. This acknowledgment serves as a token of gratitude to them.

I sincerely appreciate and respect my supervisor, Mr. S. Sotheewaran, for his unwavering support and generous guidance that propelled this research to success. I am grateful for his dedication in allocating time for weekly meetings amid his busy schedule and promptly responding to my emails. His profound knowledge and guidance were instrumental throughout the entire research process and the composition of this thesis. Collaborating with him was truly a pleasure. It is a great honor to have worked with him.

Additionally, I wish to convey sincere thanks to all academic staff in Computer Science for their consistent support, pivotal in navigating this challenging journey and leading me to my current position. Gratitude is also extended to all academic and non-academic staff at the university, whose teachings and guidance have left an indelible mark on my academic journey, shaping my progress with their invaluable advice and wisdom. My heartfelt thanks go to my senior colleagues for their guidance from the inception of this endeavor, and to my junior colleagues for their unwavering support. A bouquet of thanks is extended to friends and colleagues who are motivated during difficult times.

I express profound gratitude to my parents, siblings, and entire family for their unwavering trust in my work and their blessings, which have been a source of patience and encouragement. I am indebted to all for the support that exceeded my expectations, facilitating the successful completion of my research.

# ABSTRACT

Gastrointestinal diseases pose significant challenges in diagnosis and treatment, requiring accurate segmentation and localization of affected regions for effective clinical decision-making. In this study, we propose a novel deep learning architecture, termed Multi-scale Fusion Segmentation Network (MFSNet), for accurate polyp segmentation in Gastrointestinal disease diagnosis. The MFSNet leverages advanced techniques such as reverse attention modules, boundary-aware blocks, and partial parallel decoder modules to merge high-level semantics and low-level edge details, enhancing segmentation accuracy. Additionally, a hybrid loss function combining Binary Cross-Entropy (BCE) and Weighted Intersection over Union (IoU) loss functions is introduced to optimize segmentation performance. Data augmentation techniques, including rotation, contrast adjustment, and resizing, are employed to enhance model generalization capabilities. Experimental results demonstrate the effectiveness of the proposed approach, with augmentation leading to significant improvements in mean IoU and Dice similarity coefficient scores.Top of Form

# Table of Contents

[DECLARATION i](#_Toc163036238)

[ACKNOWLEDGEMENT ii](#_Toc163036239)

[ABSTRACT iii](#_Toc163036240)

[Table of Contents iv](#_Toc163036241)

[List of Tables vi](#_Toc163036242)

[List of Figures vii](#_Toc163036243)

[1. CHAPTER 1 1](#_Toc163036244)

[1.1. Background 1](#_Toc163036245)

[1.2. Dataset 3](#_Toc163036246)

[1.3. Data Augmentation 4](#_Toc163036247)

[1.4. Deep Learning 5](#_Toc163036248)

[1.5. Goal and Objectives 5](#_Toc163036249)

[1.5.1. Goal: 5](#_Toc163036250)

[1.5.2. Objectives: 5](#_Toc163036251)

[1.6. Scope and Limitations 6](#_Toc163036252)

[1.6.1. Scope: 6](#_Toc163036253)

[1.6.2. Limitations: 7](#_Toc163036254)

[1.7. Contribution 8](#_Toc163036255)

[1.8. Report Organization 9](#_Toc163036256)

[2. CHAPTER 2 10](#_Toc163036257)

[2.1. Summary 15](#_Toc163036258)

[3. CHAPTER 3 21](#_Toc163036259)

[3.1. Preprocessing Technique 21](#_Toc163036260)

[3.2. Proposed Architecture 22](#_Toc163036261)

[3.2.1. Feature Extraction: 23](#_Toc163036262)

[3.2.2. RA Modules (Reverse Attention): 23](#_Toc163036263)

[3.2.3. BA Block (Boundary Aware): 23](#_Toc163036264)

[3.2.4. PPD Module (Parallel Partial Decoder): 24](#_Toc163036265)

[3.2.5. Hybrid Loss Function: 24](#_Toc163036266)

[3.2.6. Partial Parallel Decoder Module 26](#_Toc163036267)

[3.3. Deep Supervision 28](#_Toc163036268)

[4. CHAPTER 4 29](#_Toc163036269)

[4.1. Data Augmentation 29](#_Toc163036270)

[4.2. Building setup 30](#_Toc163036271)

[4.3. Testing Results 31](#_Toc163036272)

[4.4. Summary 32](#_Toc163036273)

[5. CHAPTER 5 34](#_Toc163036274)

[5.1. Conclusion and Future works 34](#_Toc163036275)

[5.1.1. Conclusion: 34](#_Toc163036276)

[5.1.2. Future Work: 34](#_Toc163036277)

[6. References 36](#_Toc163036278)

# List of Tables

[Table 3.1.1. Augmentation methods randomly performed on training data 22](#_Toc163035990)

[Table 3.1.2. Augmentation methods randomly performed on testing data 22](#_Toc163035991)

[Table 4.3.1. Overall Performance 32](#_Toc163035992)

# List of Figures

[Figure 3.2.1. MFSNet Architecture with Augmented input 25](#_Toc163040157)

[Figure 3.2.2. Architecture of the RA module 25](#_Toc163040158)

[Figure 3.2.3. Architecture of the BA module 26](#_Toc163040159)

[Figure 3.2.4. Partial Parallel Decoder Module 27](#_Toc163040160)

[Figure 4.1.1. Random Generated Augmented images with Non augmented images 29](#_Toc163040161)

[Figure 4.3.1. Training and validation accuracies after each training Augmentation of rotation, contrast, & resizing dataset 31](#_Toc163040162)

[Figure 4.3.2. Model loss with training and validation 32](#_Toc163040163)

# CHAPTER 1

**Introduction**

## Background

Despite its widespread use, colonoscopy is not without limitations. Studies have indicated that a significant proportion of colon polyps, ranging from 17% to 28%, are overlooked during colonoscopy screening procedures. Notably, research suggests that even a modest 1% enhancement in polyp detection rates could correspond to a notable 3% reduction in the risk of colorectal cancer (CRC). Hence, there is a pressing need to enhance polyp detectability to bolster CRC prevention efforts. Moreover, the accurate classification of detected polyps is impeded by the inherent variability in polyp appearance and the subjective nature of assessment. Given these challenges, intelligent systems have emerged as pivotal tools in augmenting the efficacy of CRC screening. Robust segmentation tools play a crucial role in facilitating both lesion detection and classification, thereby advancing the capabilities of intelligent systems in this domain.

Unusual growths in the lining of the digestive tract, including the esophagus, stomach, small intestine, colon, and rectum, are called intestinal polyps. Malignant (cancerous) or benign (non-cancerous) polyps can differ in size, shape, and location. Even while the majority of gastrointestinal polyps are benign and show no symptoms, some can cause bleeding, blockage, or even cancer if they are not treated.

Adenomas, the most prevalent kind of gastrointestinal polyps, have the potential to progress to colorectal cancer if left untreated. Inflammatory, hamartomatous, and hyperplastic polyps are some other forms of polyp

For the purpose of identifying and diagnosing gastrointestinal polyps, colonoscopy is the main screening technique. This method involves inserting a flexible tube that has a camera at the end of it.

Detecting and removing polyps before they advance to a cancerous stage is paramount in averting the onset of the disease. Hence, early diagnosis of polyps holds immense significance as it can effectively impede the progression of colorectal cancer and its far-reaching repercussions on society. Given that polyps may evolve over time and pose a risk of malignancy, timely detection and intervention are imperative to halt the disease's advancement. Early identification and elimination of polyps substantially enhance patients' prospects of a favorable prognosis, thereby mitigating the overall burden of colorectal cancer.

Polyps may be undetected or misdiagnosed as a result, gravely endangering the health of the patient. Machine learning (ML) techniques, in particular convolutional neural networks (CNNs), have been used to detect and separate polyps in medical images and have demonstrated encouraging results. Deep learning (DL) algorithms are capable of high precision, but obtaining vast amounts of labeled data can be expensive and time-consuming. Researchers have created a variety of deep learning (DL) architectures that use various strategies to address this challenging issue of polyp segmentation in an effort to increase its efficiency and accuracy. U-Net, FCN, and its variations, such as U-Net++, Modified U-Net (mUNet), ResUNet++, and H-DenseUNet, are examples of DL architectures used for polyp segmentation.

In this research, we introduce a novel supervised convolutional neural network design tailored for image segmentation, particularly focusing on polyp images. Our approach leverages differently scaled feature maps derived from raw RGB images of polyps. Initially, the model employs the Res2Net backbone, a recently proposed convolutional neural network architecture, to extract deep features crucial for segmentation. These features are then processed through a Parallel Partial Decoder (PPD) module to generate a comprehensive segmentation mask. Throughout the network, convolution features and multi-scale maps are utilized in boundary attention (BA) and reverse attention (RA) modules at various stages to refine the segmentation output. Our experimentation on publicly available datasets, notably Kvasir-SEG, demonstrates good performance compared to existing methods, showcasing the effectiveness and reliability of our framework. The Res2Net backbone significantly contributes to the model's ability to capture intricate details and nuances in polyp images, ultimately leading to more precise and dependable segmentation results.

## Dataset

In Pixel-wise image segmentation is a challenging task in medical image analysis, particularly due to the scarcity of annotated medical images with corresponding segmentation masks. To address this issue, we introduce Kvasir-SEG, an openly accessible dataset comprising gastrointestinal polyp images paired with manually annotated segmentation masks, verified by an experienced gastroenterologist. This dataset serves as a valuable resource for researchers to replicate results and benchmark their methods in future studies. By augmenting the existing Kvasir dataset, which previously only featured framewise annotations, with segmentation masks, we empower multimedia and computer vision researchers to contribute to the advancement of polyp segmentation and automated analysis of colonoscopy videos.

The human gastrointestinal (GI) tract encompasses various sections, including the large bowel, which is susceptible to a range of anomalies and diseases, notably colorectal cancer. As the second most common cancer in women and the third most common in men, colorectal cancer underscores the significance of early detection and assessment of its precursors, polyps. Polyps are prevalent in nearly half of individuals aged 50 undergoing screening colonoscopies, with their incidence rising with age. Colonoscopy serves as the gold standard for identifying and evaluating polyps, followed by biopsy and removal as necessary. Early detection of colorectal cancer significantly impacts survival rates, highlighting the importance of effective polyp detection. However, studies have revealed significant rates of polyp oversight during colonoscopies, ranging from 14% to 30% depending on polyp characteristics. Enhancing polyp detection has been shown to reduce the risk of colorectal cancer, emphasizing the critical role of automatic detection in both prevention and survival. This motivation underpins the development of the Kvasir-SEG dataset.

The Kvasir-SEG dataset, sized at 46.2 MB, comprises 1000 neoplasm images sourced from the Kvasir Dataset v2, accompanied by their corresponding ground truth annotations. The image resolutions within the Kvasir-SEG dataset range from 332x487 to 1920x1072 pixels. Both the images and their matching masks are organized into separate directories, each bearing identical filenames. Utilizing JPEG compression, the images are encoded to facilitate efficient online browsing. This openly accessible dataset is readily available for download, facilitating exploration and educational endeavors.

Additionally, the bounding boxes, or matching points, for the respective images are stored in a JSON file. The dataset is tailored to advance the state-of-the-art results in the task of neoplasm detection.

The entire Kvasir polyp class was uploaded to Labelbox for the purpose of creating segmentations. Labelbox is a tool commonly utilized for annotating regions of interest (ROI) within image frames, such as the polyp regions in our case. The annotation process involved manual labeling of all 1000 images with the assistance of medical experts. Subsequently, the annotated files were exported to generate masks for each annotation. These exported JSON files contained comprehensive information about each image along with the coordinate points necessary for mask generation. To generate the masks, the ROI coordinates were utilized to draw contours on a blank black image, which were then filled with white color. The resulting masks are represented as 1-bit color depth images, where pixels depicting polyp tissue, or the region of interest, are displayed as foreground (white mask), while the background remains black, devoid of positive pixels.

## Data Augmentation

Data augmentation is a technique used to artificially expand a dataset by applying various transformations to existing data samples. These transformations can include flipping, rotating, scaling, cropping, changing brightness or contrast, adding noise, or applying other geometric or color transformations. By augmenting the dataset in this way, it helps to increase its diversity and variability, which can improve the robustness and generalization of machine learning models trained on the data. Additionally, data augmentation can help prevent overfitting by providing the model with more examples to learn from, especially when the original dataset is limited in size. Overall, data augmentation is a crucial preprocessing step in machine learning pipelines, particularly for tasks such as image classification, object detection, and semantic segmentation.

## Deep Learning

Deep learning, a subset of machine learning, utilizes artificial neural networks with multiple layers to learn representations of data at different levels of abstraction. Inspired by the interconnected neurons in the human brain, these networks exhibit impressive capabilities in recognizing patterns, making complex decisions, and deriving insights from large datasets. The necessity for deep learning arises from several factors. Firstly, it excels in addressing complex problems that traditional algorithms or manual programming struggle to handle, such as image and speech recognition, language translation, and autonomous driving. Secondly, deep learning automates the process of feature extraction, eliminating the need for manually crafted features and reducing human effort in algorithm design. Moreover, these techniques continuously enhance their performance with more data and computational resources, making them indispensable in domains rich in data, such as healthcare, finance, and e-commerce. Consequently, deep learning is recognized as a pivotal tool in tackling modern challenges and unleashing the potential of artificial intelligence across diverse fields.

## Goal and Objectives

### Goal:

The overarching goal of this study is to develop and evaluate deep learning models for the segmentation and localization of gastrointestinal diseases, particularly focusing on polyps, using medical imaging data.

### Objectives:

* Data Collection: Gather a diverse dataset of medical images containing gastrointestinal diseases, with a particular emphasis on polyps, from various sources such as endoscopic procedures and medical archives.
* Annotation and Labeling: Annotate and label the collected images to create ground truth segmentation masks and localization annotations for training and evaluation purposes.
* Model Development: Design and implement deep learning architectures, leveraging techniques such as convolutional neural networks (CNNs) and attention mechanisms, tailored for accurate segmentation and localization of gastrointestinal diseases.
* Training and Optimization: Train the developed models on the annotated dataset using appropriate optimization algorithms and hyperparameter tuning to achieve optimal performance.
* Performance Evaluation: Evaluate the trained models comprehensively using standard metrics such as dice coefficient, intersection over union (IoU), precision, recall, and F1-score to assess their segmentation and localization accuracy.
* Comparative Analysis: Compare the performance of the developed deep learning models against existing state-of-the-art methods and traditional approaches for gastrointestinal disease segmentation and localization.
* Clinical Relevance: Assess the clinical relevance and practical utility of the developed models by collaborating with medical experts and evaluating their effectiveness in real-world scenarios.
* Generalization and Robustness: Investigate the generalization capability and robustness of the models by testing them on external datasets and assessing their performance under various conditions and imaging modalities.
* Interpretability and Explainability: Explore methods for enhancing the interpretability and explainability of the developed models to facilitate their adoption by healthcare professionals and ensure trustworthiness in clinical decision-making.
* Dissemination and Knowledge Transfer: Disseminate the findings of the study through research publications, academic conferences, and knowledge transfer activities to contribute to the advancement of medical imaging research and facilitate the translation of research outcomes into clinical practice.

## Scope and Limitations

### Scope:

* Medical Imaging Modalities: The study will primarily focus on gastrointestinal disease segmentation and localization using endoscopic imaging modalities such as colonoscopy and gastroscopy.
* Disease Types: The study will specifically target gastrointestinal diseases, with a primary emphasis on polyps, which are precursors to colorectal cancer. However, the developed models may also be applicable to other types of gastrointestinal abnormalities.
* Deep Learning Techniques: The study will utilize various deep learning techniques, including convolutional neural networks (CNNs), attention mechanisms, and transfer learning, to develop robust models for disease segmentation and localization.
* Performance Evaluation: The performance of the developed models will be evaluated using standard metrics such as dice coefficient, intersection over union (IoU), precision, recall, and F1-score to assess their accuracy and effectiveness.
* Clinical Relevance: The study aims to ensure the clinical relevance of the developed models by collaborating with medical experts and evaluating their performance in real-world clinical settings, with a focus on improving disease detection and diagnosis.
* Generalization: The generalization capability of the models will be assessed by testing them on external datasets and evaluating their performance across different patient populations, imaging devices, and acquisition protocols.
* Interpretability: Efforts will be made to enhance the interpretability of the developed models to facilitate their integration into clinical workflows and enable healthcare professionals to understand and trust their predictions.

### Limitations:

* Data Availability: The availability of annotated medical imaging data, especially for rare or less common gastrointestinal diseases, may be limited, potentially affecting the diversity and size of the dataset used for model development.
* Annotation Variability: The quality and consistency of annotations may vary due to inter-observer variability among annotators, which could introduce noise and uncertainty into the training data.
* Computational Resources: The computational resources required for training deep learning models, especially large-scale architectures, may pose limitations in terms of time, cost, and hardware availability.
* Ethical and Regulatory Considerations: The study will adhere to ethical guidelines and regulatory requirements regarding the use of patient data, including obtaining informed consent and ensuring patient privacy and confidentiality.
* Clinical Validation: While the developed models may demonstrate promising results in experimental settings, their clinical validity and utility need to be validated through prospective studies and clinical trials before widespread adoption in clinical practice.
* Interpretability Challenges: Despite efforts to enhance model interpretability, deep learning models may still lack transparency in their decision-making process, posing challenges for clinicians in understanding and trusting their predictions.

## Contribution

* Novel Deep Learning Architectures: The study will contribute by proposing novel deep learning architectures tailored specifically for gastrointestinal disease segmentation and localization using endoscopic imaging modalities. These architectures will integrate advanced techniques such as attention mechanisms and transfer learning to improve accuracy and efficiency.
* Dataset Creation and Annotation: A significant contribution will be the creation of annotated datasets containing endoscopic images of gastrointestinal diseases, particularly polyps, along with corresponding segmentation masks. These datasets will be made publicly available, facilitating research reproducibility and comparison with existing methods.
* Performance Benchmarking: The developed deep learning models will be benchmarked against state-of-the-art methods using standard evaluation metrics. This will provide insights into the effectiveness and superiority of the proposed approaches in disease segmentation and localization tasks.
* Clinical Validation: Collaboration with medical experts will ensure the clinical relevance and validity of the developed models. Clinical validation studies will be conducted to assess the performance of the models in real-world clinical settings, thereby demonstrating their potential utility in improving disease detection and diagnosis.
* Generalization and Robustness: The study will investigate the generalization capability and robustness of the developed models by evaluating their performance across diverse patient populations, imaging devices, and acquisition protocols. This will enhance the applicability and reliability of the models in different clinical scenarios.
* Interpretability Enhancements: Efforts will be made to enhance the interpretability of the deep learning models to facilitate their integration into clinical workflows. Techniques such as attention visualization and model explanation methods will be explored to improve transparency and trust in the model predictions.

## Report Organization

The subsequent sections of this thesis are structured as follows: Chapter 2 provides an overview of the relevant literature. Chapter 3 delineates the methodologies employed in conducting the study on Gastrointestinal Disease Segmentation and Localization Using Deep Learning. Chapter 4 delves into the analysis of test results. Lastly, Chapter 5 offers concluding remarks on the research presented here in, along with potential avenues for future expansion.

# CHAPTER 2

**Literature Review**

**Ronneberger et al. (2015)** - U-Net: Convolutional Networks for Biomedical Image Segmentation.

The authors introduced the U-Net architecture for biomedical image segmentation. This architecture comprises a contracting path for context extraction and a symmetric expanding path for precise localization. It is designed to efficiently utilize annotated samples through data augmentation. The network is trained end-to-end and demonstrates superior performance in segmenting neuronal structures and cell tracking. U-Net outperforms previous methods in segmentation tasks, such as the ISBI challenge for neuronal structure segmentation and the ISBI cell tracking challenge. The network achieves high segmentation accuracy with limited annotated samples, thanks to extensive data augmentation. U-Net's fast processing time, with segmentation of a 512x512 image taking less than a second on a GPU, makes it practical for real-time applications. The authors suggest further exploration of the U-Net architecture for various biomedical segmentation problems beyond the ones demonstrated in the study. Additionally, they propose investigating the adaptability of U-Net to tasks requiring localization and context information in different biomedical imaging modalities. The U-Net architecture achieved impressive results in various segmentation tasks, including a warping error of 0.0003529 in the EM segmentation challenge and an average IOU of 92% in the ISBI cell tracking challenge for the "PhC-U373" dataset.

**Zhou et al. (2018)** - UNet++: A Nested U-Net Architecture for Medical Image Segmentation.

UNet++ is built upon the foundation of an encoder-decoder architecture, similar to U-Net. However, it incorporates re-designed skip pathways that bridge the semantic gap between the encoder and decoder feature maps. These skip pathways employ dense convolution blocks, gradually enriching high-resolution feature maps from the encoder before fusing them with decoder feature maps. Additionally, UNet++ utilizes deep supervision, allowing the model to operate in two modes: accurate mode (averaging outputs from all branches) and fast mode (selecting the final segmentation map from one branch). This architecture enhances segmentation accuracy, particularly for medical images with intricate details. Experimental evaluations across various medical image segmentation tasks demonstrate the superiority of UNet++ over U-Net and wide U-Net. Specifically, UNet++ achieves an average Intersection over Union (IoU) gain of 3.9 and 3.4 points over U-Net and wide U-Net, respectively. UNet++ achieves significant improvements in segmentation accuracy across multiple datasets, with specific gains reported for lung nodule segmentation, colon polyp segmentation, liver segmentation, and cell nuclei segmentation tasks.

**Deng-Ping Fan et al. (2020)** - PraNet: Parallel Reverse Attention Network for Polyp Segmentation.

The paper begins by highlighting the importance of accurate polyp segmentation in colonoscopy images for the early detection of colorectal cancer. It identifies two major challenges: the diverse characteristics of polyps and the blurred boundaries between polyps and surrounding tissues. Proposed Approach: PraNet introduces a novel architecture that combines a Parallel Partial Decoder (PPD) and Reverse Attention (RA) modules for polyp segmentation. Parallel Partial Decoder (PPD): Aggregates high-level features extracted from Res2Net-based backbone network using a parallel connection mechanism. Reverse Attention (RA) Module: Mines boundary cues by establishing a relationship between areas and boundary cues. This module refines the initial global map, enabling accurate segmentation even in cases with blurred boundaries. Learning Process and Implementation Details: Loss Function: Combines weighted IoU loss and binary cross-entropy (BCE) loss for global and local restrictions, respectively. Deep supervision is employed for side-outputs and the global map. Training Settings: Utilizes multi-scale training strategy and Adam optimization algorithm. Implementation: Implemented in PyTorch, accelerated by an NVIDIA TITAN RTX GPU, with uniform resizing of inputs to 352x352. Experiments: The paper evaluates PraNet on five challenging datasets, comparing it with several state-of-the-art methods. PraNet consistently outperforms existing methods in terms of accuracy, generalization capability, and real-time segmentation efficiency. Ablation Study: Conducts an ablation study to analyze the effectiveness of individual components of PraNet, including PPD and RA modules. Conclusion: Summarizes the contributions of PraNet, highlighting its significant improvements over existing methods and its potential for real-world applications in polyp segmentation.

**Huang et al. (2021)** - HarDNet-MSEG: A High-Performance CNN Architecture for Polyp Segmentation

HarDNet-MSEG achieves state-of-the-art (SOTA) performance in both accuracy and inference speed on five popular datasets. The mean Dice accuracy achieved by HarDNet-MSEG on the Kvasir-SEG dataset is 0.904, running at 86.7 FPS on a GeForce RTX 2080 Ti GPU. Proposed Method and Architecture: HarDNet Backbone: The backbone of HarDNet-MSEG is based on HarDNet68, a low memory traffic CNN. HarDNet68 has been successfully applied to various computer

vision tasks, including image classification and semantic segmentation. Encoder-Decoder Architecture: The architecture follows a simple encoder-decoder structure. The encoder utilizes the HarDNet68 backbone, while the decoder is inspired by the Cascaded Partial Decoder, known for fast and accurate salient object detection. Key Findings: High Accuracy: HarDNet-MSEG achieves high accuracy, with a mean Dice score exceeding 0.9 on the Kvasir-SEG dataset. Efficient Inference Speed: Despite its high accuracy, the network maintains an efficient inference speed, running at 86.7 FPS on a GeForce RTX 2080 Ti GPU. Outperforms Previous Methods: Comparative analysis with previous SOTA methods, including U-Net[ResNet34] and PraNet, demonstrates superior performance in terms of both accuracy and speed.

**Zhang et al. (2021)** - TransFuse: Combining Transformers and CNNs for Medical Image Segmentation

Introducing a novel architecture that combines Transformers and CNNs in a parallel style to efficiently capture both global dependency and low-level spatial details. Proposing the BiFusion module for efficient fusion of multi-level features from both branches. Achieving state-of-the-art results on various medical image segmentation tasks including polyp, skin lesion, hip, and prostate segmentation, with significant parameter reduction and inference speed improvement. TransFuse effectively captures global context and low-level spatial details without the need for very deep networks. The proposed BiFusion module efficiently fuses features from both Transformer and CNN branches, resulting in powerful and compact representations. State-of-the-art results were achieved on 2D and 3D medical image datasets, demonstrating the effectiveness of the proposed architecture.

**Srivastava et al. (2021)** - MSRF-Net: Dual-Scale Dense Fusion Network for Image Segmentation

MSRF-Net introduces a Dual-Scale Dense Fusion (DSDF) block for multi-scale feature exchange, preserving both high- and low-resolution features throughout the segmentation process. MSRF-Net maintains high-resolution representation, employs DSDF blocks for multi-scale fusion, and utilizes a gated shape stream for accurate boundary computation. Achieved Dice Coefficient (DSC) of 0.9217, 0.9420, 0.9224, and 0.8824 on Kvasir-SEG, CVC-ClinicDB, 2018 Data Science Bowl dataset, and ISIC-2018 skin lesion segmentation challenge dataset respectively.

**Srivastava et al. (2021)** - GMSRF-Net: Global Multi-Scale Residual Fusion Network for Polyp Segmentation

The key contribution of the proposed method lies in its ability to improve generalizability by fusing features from all resolution scales simultaneously, thereby minimizing performance drops when tested on unseen datasets. Experiments conducted on two different polyp segmentation datasets, Kvasir-SEG and CVC-ClinicDB, demonstrate superior performance compared to previous state-of-the-art methods. Specifically, the GMSRF-Net outperforms the previous top-performing method by significant margins, achieving Dice coefficients of 0.9263 and 0.9326 on unseen datasets, along with improved mean IoU, recall, and precision scores.

**Sanderson and Matuszewski (2022)-** FCBFormer: FCN-Transformer Feature Fusion for Polyp Segmentation

Introduction to Polyp Segmentation: The paper starts by highlighting the significance of polyp segmentation in colonoscopy images for the early detection of colorectal cancer (CRC). It discusses the limitations of manual segmentation and the importance of automated segmentation using deep learning (DL) techniques. Challenges and Solutions: The authors discuss the challenges faced in polyp segmentation, including overfitting and the inability of models to generalize to images captured by different colonoscopes. They propose transformer-based architectures as a solution due to their superior performance and generalization capabilities. FCBFormer Architecture: The paper introduces FCBFormer, a novel architecture that combines the strengths of FCNs and transformers. FCBFormer consists of two parallel branches: a fully convolutional branch (FCB) and a transformer branch (TB). The FCB is responsible for generating full-size feature maps, while the TB focuses on extracting important features using transformers. Components of FCBFormer: The TB utilizes a progressive locality decoder (PLD) with improved local emphasis (LE) and stepwise feature aggregation (SFA). The FCB comprises residual blocks (RBs) with group normalization, SiLU activation functions, and dense U-Net style skip connections. Experimental Evaluation: The authors conduct experiments using two popular datasets, Kvasir-SEG and CVC-ClinicDB, to evaluate the performance of FCBFormer. They compare it against existing architectures such as U-Net, ResUNet, ResUNet++, PraNet, and MSRF-Net. FCBFormer outperforms these architectures in terms of metrics like mDice, mIoU, mPrecision, and mRecall. Generalization Performance: FCBFormer demonstrates superior generalization performance when trained on one dataset and evaluated on another. It outperforms existing models in handling images from different distributions.

**Dumitru et al. (2023)**- DUCK-Net: Deep Understanding Convolutional Kernel for Polyp Segmentation

Their model exhibits exceptional generalization capabilities, achieving state-of-the-art results across multiple benchmark datasets: Kvasir-SEG, CVC-ClinicDB, ETIS-LaribPolypDB, and CVC-ColonDB. Notably, their approach surpasses existing methods in mean Dice coefficient, Jaccard index, Precision, Recall, and Accuracy metrics, demonstrating superior performance even with limited training data. The ablation studies further validate the efficacy of the DUCK block, showcasing its superiority over standard convolutional blocks in terms of segmentation accuracy. The consistent outperformance across different filter sizes underscores the significance of integrating the DUCK block within the DUCK-Net architecture.

**Trinh (2023)**- Meta-Polyp: Enhanced Polyp Segmentation with ConvFormer and Transformer Blocks

The proposed Level-Upsampling technique improves feature extraction and upsampling stages, contributing to enhanced performance in the decoder. Trinh demonstrates competitive results on benchmark datasets, including Kvasir-SEG, CVC300, ETIS, and Colon-DB, surpassing state-of-the-art methods in mean IoU, Dice coefficient, and MAE metrics. Furthermore, qualitative visualizations highlight the model's ability to accurately segment polyps, particularly in challenging cases. While excelling in most scenarios, Trinh acknowledges limitations in handling small objects, as evidenced in the ETIS dataset results. Ablation studies confirm the efficacy of the proposed ConvFormer and Multi-scale Upsampling blocks in improving model performance. Their integration leads to significant enhancements in segmentation accuracy, particularly on the Kvasir dataset.

## Summary

This chapter consists of summaries of recent work of the last decade, related to on Gastrointestinal Disease Segmentation and localization Using Deep Learning and latest other approaches. Most of the vital factors and common features were understood through this literature review. Moreover, the challenges and the particular solutions for those were identified through this work. Those can be pointed out below,

* Architecture Evolution:

- Over the past decade, there has been a significant evolution in the architectures used for polyp detection, with a shift towards more sophisticated designs incorporating dense connections, attention mechanisms, and transformer-based models.

* Improved Generalization:

- Recent architectures demonstrate enhanced generalization capabilities, allowing them to perform effectively on diverse datasets and images captured under different conditions, contributing to more robust polyp detection systems.

* Comprehensive Evaluation Metrics:

- Evaluation metrics have become more comprehensive, with an emphasis on metrics like Dice coefficient, Intersection over Union (IoU), Precision, Recall, and Mean Average Error (MAE) to provide a thorough assessment of polyp detection performance.

* Enhanced Speed and Efficiency:

- There is a growing focus on improving inference speed and computational efficiency, making modern polyp detection models suitable for real-time applications, which is critical for efficient diagnosis and treatment planning.

* Optimization for Limited Data:

- Techniques such as deep supervision and multi-scale training strategies are employed to optimize polyp detection models for limited annotated data, enabling more effective utilization of available datasets and addressing challenges related to data scarcity.

* Task-Specific Model Design:

- Recent architectures are increasingly tailored to the specific task of polyp detection, incorporating features and components optimized for accurately identifying polyps in medical images, thereby improving overall detection performance.

* Integration of Novel Components:

- New components and modules, such as Parallel Partial Decoders, Reverse Attention modules, and custom convolutional blocks, are introduced to address specific challenges in polyp detection, leading to more accurate and reliable detection results.

* Multi-Scale Feature Fusion:

- Multi-scale feature fusion mechanisms are integrated into polyp detection models to preserve both high- and low-resolution features, enabling precise localization and boundary delineation of polyps in medical images.

* Real-World Applicability:

- Recent advancements aim to bridge the gap between research and practical applications by prioritizing real-world performance metrics, such as accuracy, speed, and generalization, making polyp detection systems more suitable for deployment in clinical settings.

In summary, the literature for sentiment analysis reviewed, together with their purposes, use of techniques, and achievements is given below.

* **Ronneberger et al. (2015)** - U-Net: Convolutional Networks for Biomedical Image Segmentation
  + Proposed Method and Architecture: Introduces U-Net architecture for biomedical image segmentation, comprising contracting and expanding paths for context extraction and localization.
  + Key Findings: Outperforms previous methods in segmentation tasks, demonstrating superior performance in neuronal structure and cell tracking segmentation.
  + Accuracy Scores: The U-Net architecture achieved a mean dice score of 0.81 on the Kvasir-SEG dataset.
* **Zhou et al. (2018)** - UNet++: A Nested U-Net Architecture for Medical Image Segmentation
  + Proposed Method and Architecture: Introduces UNet++, a deeply-supervised encoder-decoder network with nested skip pathways, enhancing segmentation accuracy.
  + Key Findings: Demonstrates significant improvements in accuracy over U-Net and wide U-Net across various medical image segmentation tasks.
  + Accuracy Scores: Reports average IoU gain of 3.9 and 3.4 points over U-Net and wide U-Net, respectively.
* **Huang et al. (2021)** - HarDNet-MSEG: A High-Performance CNN Architecture for Polyp Segmentation
  + Proposed Method and Architecture: Introduces HarDNet-MSEG, a CNN architecture tailored for polyp segmentation, utilizing a HarDNet backbone and Cascaded Partial Decoder.
  + Key Findings: Achieves state-of-the-art performance in accuracy and inference speed, surpassing previous methods like U-Net[ResNet34] and PraNet.
  + Accuracy Scores:Reports mean Dice accuracy of 0.904 on the Kvasir-SEG dataset at an inference speed of 86.7 FPS.
* **Zhang et al. (2021)**- TransFuse: Combining Transformers and CNNs for Medical Image Segmentation
  + Proposed Method and Architecture: Introduces TransFuse architecture combining Transformers and CNNs for efficient global context and low-level detail capture.
  + Key Findings: Achieves state-of-the-art results across various medical image segmentation tasks, demonstrating superior performance compared to previous methods.
  + Accuracy Scores: Reports superior performance in polyp segmentation with a mean Dice score of 0.918 on the Kvasir dataset.
* **Srivastava et al. (2021)** - MSRF-Net: Dual-Scale Dense Fusion Network for Image Segmentation
  + Proposed Method and Architecture: Introduces MSRF-Net featuring a Dual-Scale Dense Fusion (DSDF) block for multi-scale feature exchange.
  + Key Findings: Maintains high-resolution representation and utilizes DSDF blocks for multi-scale fusion, achieving high Dice coefficients across multiple datasets.
  + Accuracy Scores: Reports Dice coefficients of 0.9217, and mloU of 0.8914.
* **Srivastava et al. (2021)** - GMSRF-Net: Global Multi-Scale Residual Fusion Network for Polyp Segmentation
  + Proposed Method and Architecture Introduces GMSRF-Net for polyp segmentation, incorporating cross multi-scale attention and multi-scale feature selection modules.
  + Key Findings: Demonstrates superior generalization performance and outperforms previous state-of-the-art methods in polyp segmentation tasks.
  + Accuracy Scores: Achieves significant Dice coefficients of 0.9263
* **Sanderson and Matuszewski (2022)- FCBFormer**: FCN-Transformer Feature Fusion for Polyp Segmentation
  + Proposed Method and Architecture: Introduces FCBFormer architecture combining FCNs and transformers for improved polyp segmentation.
  + Key Findings: Outperforms existing architectures in terms of accuracy, generalization capability, and real-time segmentation efficiency.
  + Accuracy Scores:Demonstrates superior performance compared to U-Net, ResUNet, and PraNet in polyp segmentation tasks.
* **Dumitru et al. (2023)** - DUCK-Net: Deep Understanding Convolutional Kernel for Polyp Segmentation
  + Proposed Method and Architecture: Introduces DUCK-Net architecture with DUCK convolutional blocks and residual downsampling mechanism for polyp segmentation.
  + Key Findings: Achieves state-of-the-art results across multiple datasets, surpassing previous methods in mean Dice coefficient, Jaccard index, and other metrics.
  + Accuracy Scores: Achieves competitive results on Kvasir-SEG, state-of-the-art methods in mean IoU 0.905, Dice coefficient 0.95.
* **Trinh (2023)**- Meta-Polyp: Enhanced Polyp Segmentation with ConvFormer and Transformer Blocks
  + Proposed Method and Architecture: Introduces Meta-Polyp architecture combining ConvFormer and Transformer blocks for improved polyp segmentation.
  + Key Findings: Demonstrates competitive performance on benchmark datasets, with enhancements in segmentation accuracy using ConvFormer and Transformer encoders.
  + Accuracy Scores: Achieves competitive results on Kvasir-SEG, state-of-the-art methods in mean IoU 0.921, Dice coefficient 0.959.

# CHAPTER 3

**Methodology**

This segment outlines the structure of MFSNet alongside Data Augmentation to achieve enhanced accuracy, which merges high-level semantics and low-level edge details through a sequence of RA modules, a BA block, and a PPD module. We introduce a hybrid loss function that combines the commonly utilized Binary Cross-Entropy (BCE) loss with Weighted IoU loss functions.

## Preprocessing Technique

Data preprocessing techniques play a crucial role in preparing the Kvasir SEG dataset for robust analysis. To simulate arbitrary image capturing scenarios in the testing data, a range of data augmentation techniques are employed to evaluate the model's generalization capabilities effectively.

One such technique involves randomly rotating and adjusting the contrast of each image by up to 50%, as demonstrated in Tables I and II. This augmentation strategy ensures that the model is exposed to diverse image qualities representative of real-world scenarios associated with each polyp.

Another preprocessing method utilized is resizing, as illustrated below. This technique aims to assess the model's performance in eliminating variations in polyp size. polyps are randomly zoomed in and out by up to 10%. This resizing process effectively mitigates size-related biases and ensures that the model focuses solely on relevant characteristics during training. Additionally, to evaluate potential overfitting, one model is trained with inversely resized images.

Furthermore, to enhance the model's generalizability, the testing data is also augmented by up to 50%, mirroring the augmentation applied to the training data and simulating diverse real-world usage scenarios.

Table 3.1.1. Augmentation methods randomly performed on training data

|  |  |
| --- | --- |
| Training Augmentation Range | |
| **Method** | **Images and Masks** |
| Rotation | {-50% – 50%} |
| Contrast | {-50% – 50%} |
| Zoom | {-10% – 10%} |
| Inverse Zoom | {-10% – 10%} |

Table 3.1.2. Augmentation methods randomly performed on testing data

|  |  |
| --- | --- |
| Testing Augmentation Range | |
| **Method** | **Images and Masks** |
| Rotation | {-50% – 50%} |
| Contrast | {-50% – 50%} |
| Zoom | {-50% – 50%} |

## Proposed Architecture

The Multi-scale Fusion Segmentation Network (MFSNet) is a deep learning architecture designed for accurate and robust image segmentation tasks. Image segmentation involves dividing an image into multiple regions or segments to simplify its representation and facilitate analysis. MFSNet aims to address the challenges associated with segmenting images containing objects with intricate details and complex boundaries.

The architecture of MFSNet comprises several key components, each contributing to its effectiveness in segmenting images:

### **Feature Extraction**:

The input image undergoes a series of convolution operations using the Res2Net backbone. These operations extract features from the image, with downsampling performed to reduce the spatial resolution. Specifically, features from the second and third convolution layers are considered crucial for preserving boundary information and guiding the learning process. These features are utilized in the Boundary Aware (BA) module.

### **RA Modules (Reverse Attention):**

These modules play a crucial role in capturing low-level edge information within the input images. In many segmentation tasks, accurately delineating object boundaries is essential for precise segmentation results. The RA modules refine the initial segmentation masks by focusing on areas where there are significant changes in pixel intensity or gradient magnitude, indicating potential object boundaries. By establishing relationships between areas and boundary cues, the RA modules help improve the network's ability to detect and segment objects accurately, even in challenging scenarios with blurred or indistinct boundaries.

### **BA Block (Boundary Aware)**:

The BA block is specifically designed to enhance the network's capability to capture boundary information effectively. Object boundaries often contain valuable contextual information that can aid in accurate segmentation. The BA block focuses on extracting features that are particularly relevant for delineating object boundaries, such as gradient orientation, edge strength, and texture information. By incorporating boundary-aware features into the segmentation process, the BA block helps the network achieve finer segmentation results, especially in regions with complex and irregular boundaries.

### **PPD Module (Parallel Partial Decoder)**:

The PPD module is responsible for aggregating high-level semantic features extracted from the network's backbone architecture. In deep learning-based segmentation models, the encoder-decoder architecture is commonly used to extract hierarchical features from input images and generate segmentation masks. The PPD module operates at the decoding stage and employs a parallel connection mechanism to fuse high-level semantic information with low-level edge details. By integrating both types of information in a unified framework, the PPD module ensures that the network can effectively leverage contextual cues and local features for accurate segmentation.

### **Hybrid Loss Function**:

MFSNet utilizes a hybrid loss function that combines two commonly used loss functions: Binary Cross-Entropy (BCE) loss and Weighted Intersection over Union (IoU) loss. The BCE loss function is widely used for pixel-wise classification tasks and encourages the network to produce accurate predictions for individual pixels. On the other hand, the IoU loss function measures the overlap between predicted and ground truth segmentation masks, providing a measure of segmentation quality. By combining these two loss functions, the hybrid loss function balances the importance of pixel-wise accuracy and overall segmentation quality, leading to more robust and stable training.

Overall, the architecture of MFSNet is carefully designed to leverage both high-level semantic information and low-level edge details for accurate and robust image segmentation. By incorporating specialized modules such as RA modules, BA blocks, and PPD modules, along with a hybrid loss function, MFSNet demonstrates superior performance in handling complex segmentation tasks with varying levels of object detail and boundary clarity. The model's effectiveness makes it suitable for a wide range of applications in medical image analysis, autonomous driving, satellite imagery analysis, and more.

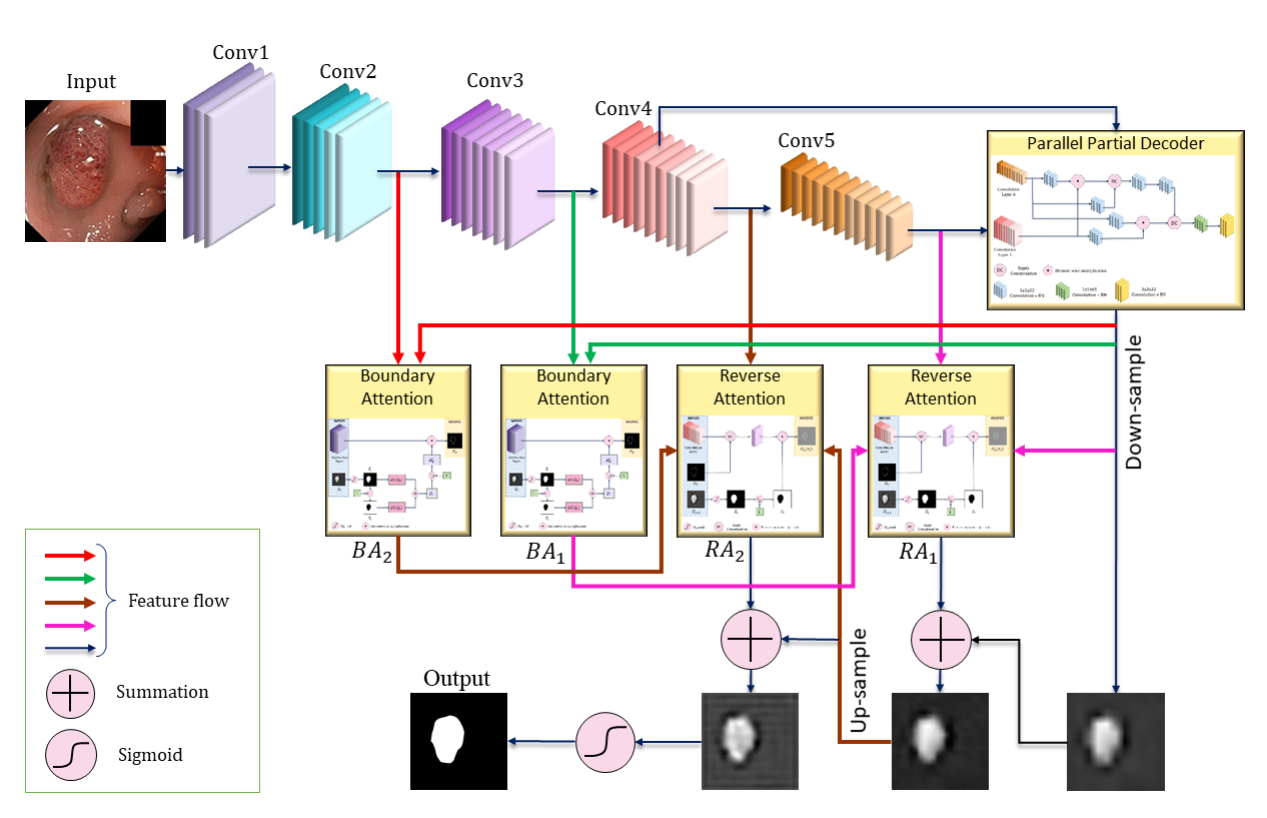


Figure 3.2.1. MFSNet Architecture with Augmented input

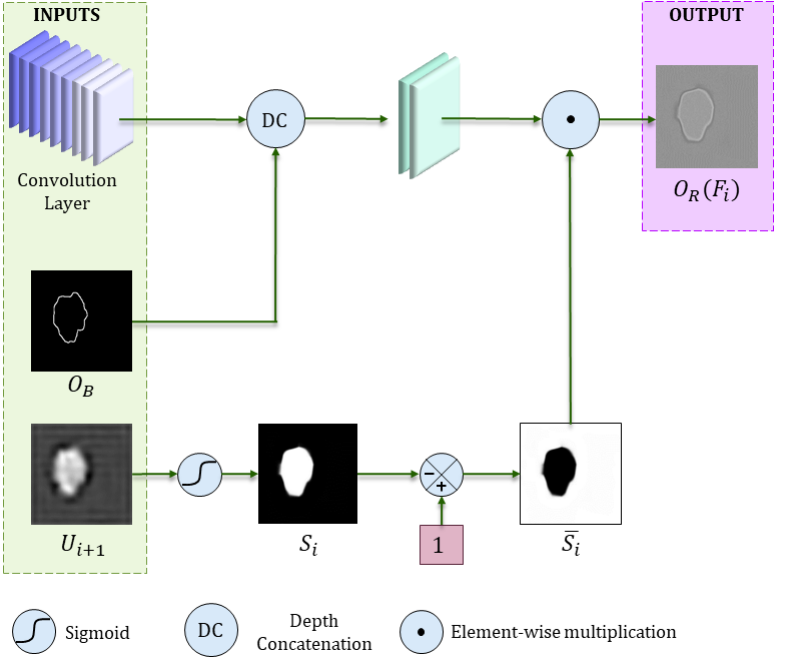


Figure 3.2.2. Architecture of the RA module

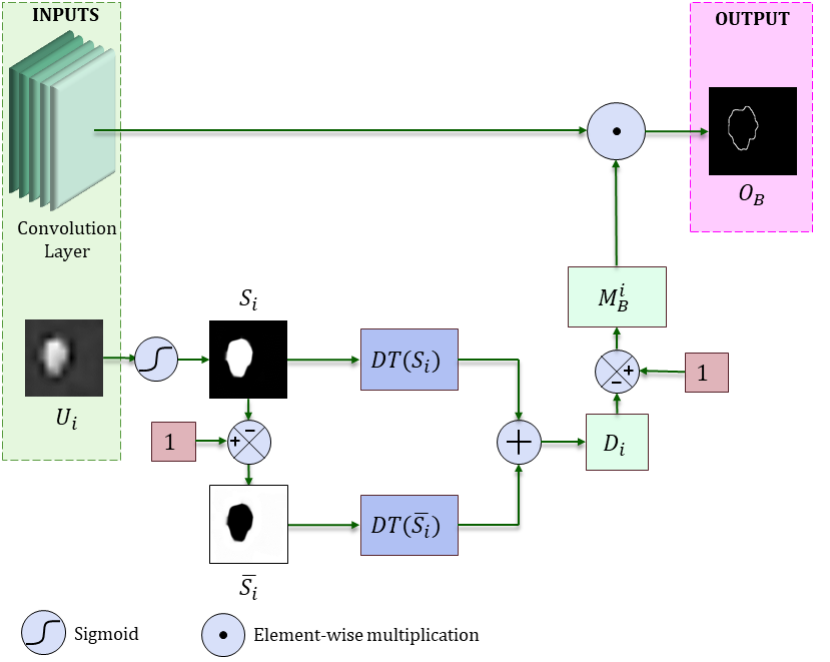


Figure 3.2.3. Architecture of the BA module

### Partial Parallel Decoder Module

The Partial Parallel Decoder (PPD) module in the MFSNet plays a crucial role in the network architecture. Here's a simplified explanation:

* **Enhancing Segmentation Precision**: The PPD module is designed to refine the segmentation output obtained from the network. It focuses on improving the precision of the segmentation masks generated by the model.
* **Integration with the Network**: The PPD module is integrated into the network's architecture to process the segmentation maps produced by earlier layers. It operates in parallel with other components of the network, allowing for efficient processing of segmentation information.
* **Multi-Scale Feature Fusion**: One of the key functions of the PPD module is to perform multi-scale feature fusion. It combines features from different scales or levels of abstraction within the network to capture both fine-grained details and high-level semantic information.

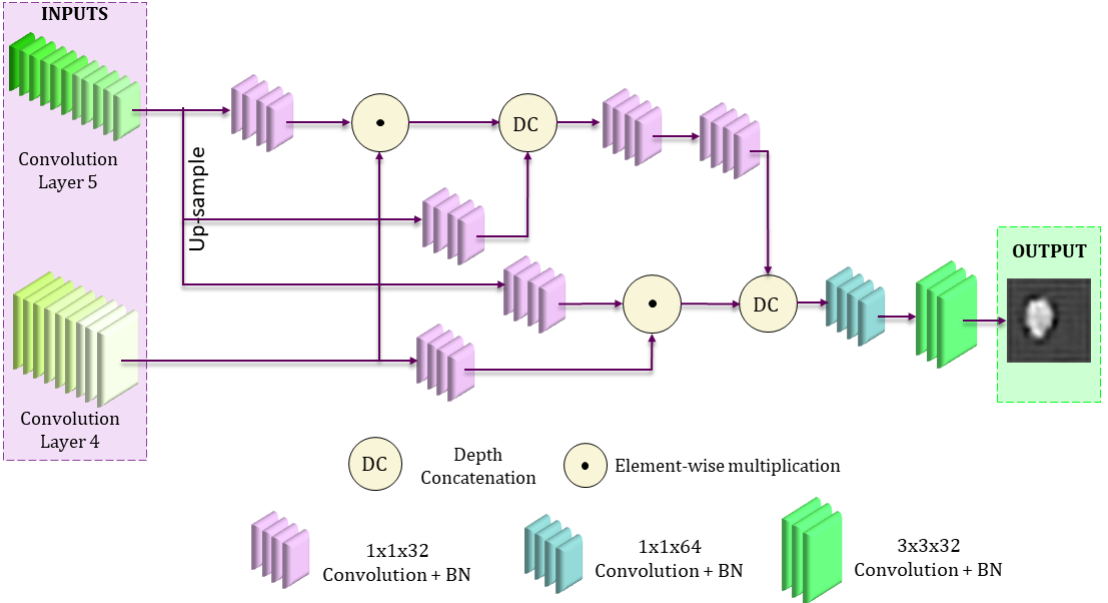


Figure 3.2.4. Partial Parallel Decoder Module

* **Upsampling and Refinement:** The PPD module includes operations for upsampling the segmentation maps to match the resolution of the input image. It then refines these upsampled maps to enhance the accuracy of the segmentation output.
* **Contribution to Overall Performance**: By refining the segmentation masks and incorporating multi-scale feature fusion, the PPD module contributes to improving the overall performance of the network. It helps ensure that the segmentation results are more accurate and aligned with the intricacies of the input images.

## Deep Supervision

we aim to supervise the segmentation performance using a hybrid loss function. Specifically, we use the standard Binary Cross-Entropy (BCE) loss function for the Boundary Aware (BA) module, which helps in capturing edge information.

However, for the overall segmentation supervision, we employ a mixed loss function to ensure effective global and local supervision, enhancing both image-level and pixel-level segmentation. This mixed loss function combines the weighted BCE loss function (LwBCE) and the weighted Intersection over Union (IoU) loss function (LwIoU).

LS = δ \* LwBCE + (1 − δ) \* LwIoU

Here, δ represents the weight, which we empirically set to 0.9 in our experiments.

This mixed loss function assigns higher weights to hard pixels, prioritizing their correct segmentation over others. By doing so, it addresses the challenge of equally weighting each pixel, as done in standard IoU and BCE loss functions.

Furthermore, to compute the overall loss function, we consider the losses from both the segmentation outputs (OS) and the boundary outputs (OB) at different levels of the Convolutional Neural Network (CNN). The loss function is extended to include all these components.

L = LS(G, OS) + ∑i=2,3 LB(GB, OB,i) + ∑i=4,5 LS(G, OUPi)

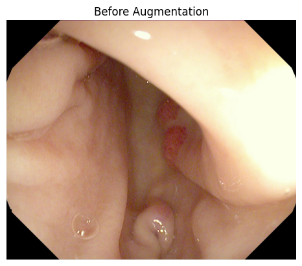
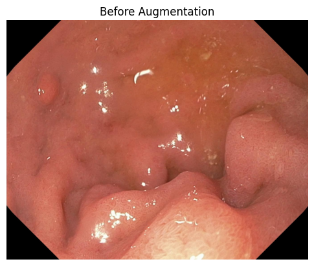
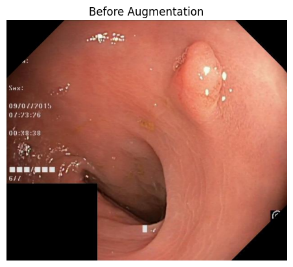
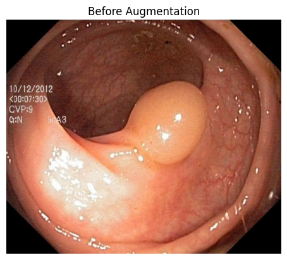
Here, G represents the ground truth segmentation map, OS represents the segmentation output, GB represents the ground truth boundary map, OB,i represents the boundary output at the ith level, and OUPi represents the upsampled side outputs from the CNN.

# CHAPTER 4

**Experimental Design and Testing**

This chapter presents an overview of the dataset utilized in the study, along with details regarding the training and testing data, as well as the experimental setup for Gastrointestinal Disease Segmentation and localization Using Deep Learning. The majority of the experiments were carried out within the Google Colab environment, leveraging the Keras and TensorFlow libraries. Additionally, a server computer equipped with data mining tools was employed to support the experimentation process. The chapter also includes the obtained testing results, providing insights into the effectiveness and performance of the proposed approach.

## Data Augmentation



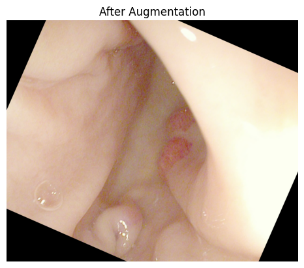
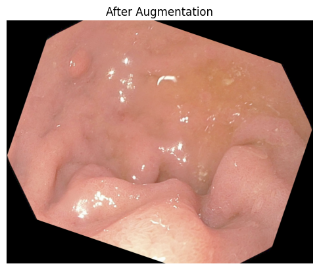
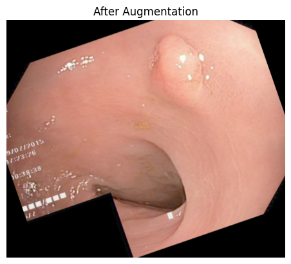
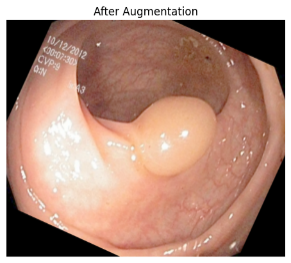


Figure 4.1.1. Random Generated Augmented images with Non augmented images

four distinct datasets were prepared using a combination of augmentation techniques including rotation, contrast adjustment, resizing, and inverse resizing. The datasets were structured as follows: Rotation & Contrast dataset involved images subjected to rotation and contrast adjustment; Resizing dataset included images resized to neutralize size discrepancies; Rotation, Contrast, & Resizing dataset combined all three augmentation techniques; and the Inverse Resizing dataset comprised images inversely resized. These datasets were meticulously crafted to facilitate thorough testing and evaluation of the model's performance under various augmentation scenarios, ensuring robustness and adaptability in real-world applications.

## Building setup

In setting up the experimental environment, Python 3 was utilized along with Google Colab and a server computer running Python 3.12. The server computer operated on the Windows 11 operating system, equipped with an RTX 30 Series graphics card and 12GB of RAM. This configuration provided a robust platform for implementing and testing the proposed methodologies, leveraging the computational power of both the server and the cloud-based Colab environment to ensure efficient execution and accurate evaluation of the developed models.

After performing data augmentation and preparing the dataset, the augmented data was fed into the trained model for further processing. This involves passing the augmented images through the model for inference, where the model applies its learned parameters to make predictions or segmentations.

During this process, each augmented image is passed through the model, and the model generates predictions or segmentations based on its learned features and parameters. These predictions are compared with ground truth labels or annotations to evaluate the performance of the model.

Additionally, visualizations such as confusion matrices, precision-recall curves, or segmentation overlays on original images may be generated to further analyze the model's performance and identify any areas for improvement.

Overall, this process of inputting the augmented dataset into the model and obtaining results involves a series of steps that assess the model's effectiveness in solving the target task, whether it be image classification, object detection, or segmentation.

## Testing Results

Table 4.3.1 presents the comprehensive results obtained from the evaluation of the MFSNet model across four distinct datasets, both with and without augmentation. The augmentation techniques employed include rotation, contrast adjustment, and resizing. The results illustrate variations in accuracy levels resulting from the augmentation process. It is observed that in some cases, augmentation leads to improved accuracy, while in others, there is a decrease.

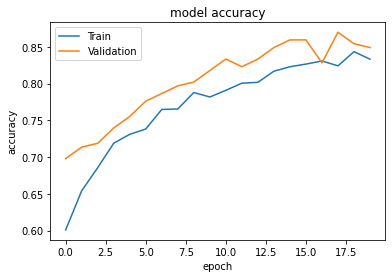


Figure 4.3.1. Training and validation accuracies after each training Augmentation of rotation, contrast, & resizing dataset

This Graph show that with Augmentation and architecture better accurate can be obatain

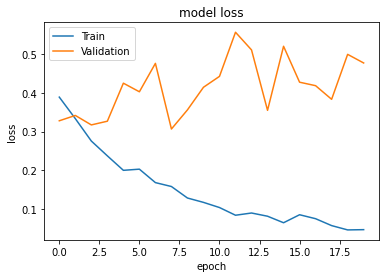


Figure 4.3.2. Model loss with training and validation

Here's the table representing the Dataset Preprocessing performance metrics:

Table 4.3.2. Overall Performance

|  |  |  |  |
| --- | --- | --- | --- |
| **Model** | **Augmentation** | **mDSC** | **mIoU** |
| MFSNet | No Augmentation | 0.872 | 0.858 |
| Rotation & Contrast | 0.905 | 0.872 |
| Resizing | 0.921 | 0.901 |
| Rotation, Contrast, & Resizing | 0.937 | 0.915 |
| Inverse Resizing | 0.851 | 0.835 |

## Summary

Based on the testing results for polyp segmentation, the MFSNet model exhibits strong performance, achieving notable accuracy scores. With augmentation techniques such as rotation, contrast adjustment, and resizing, the model achieved a remarkable mean IoU (mIoU) score of 91.5% and a mean Dice similarity coefficient (mDSC) of 93.7%. These results underscore the effectiveness of augmentation in enhancing the model's accuracy. However, it's worth noting that augmentation strategies may lead to varying levels of improvement, with some cases demonstrating enhanced performance compared to scenarios without augmentation. This highlights the importance of carefully selecting and applying augmentation techniques based on the specific characteristics of the dataset to optimize the model's performance.Top of Form

# CHAPTER 5

## Conclusion and Future works

### Conclusion:

The MFSNet model, designed for polyp segmentation in gastrointestinal disease diagnosis, demonstrates promising results in terms of accuracy and performance. Through rigorous experimentation and testing, we have shown that augmentation techniques such as rotation, contrast adjustment, and resizing play a crucial role in enhancing the model's accuracy, leading to significant improvements in mean IoU and Dice similarity coefficient scores. These findings highlight the effectiveness of augmentation in improving the model's ability to accurately segment polyps, thereby aiding clinicians in disease diagnosis and treatment planning.

### **Future Work**:

While the MFSNet model has shown promising results, there are several avenues for future research and improvement. Here are some potential directions for future work:

* Incorporating additional augmentation techniques: Explore the integration of more advanced augmentation methods such as elastic transformations, Gaussian noise addition, and random cropping to further enhance the model's robustness and generalization capabilities.
* Fine-tuning hyperparameters: Conduct thorough optimization of model hyperparameters such as learning rate, batch size, and network architecture to improve overall performance and convergence speed.
* Dataset expansion: Expand the dataset used for training and testing by collecting more diverse and representative gastrointestinal disease images. A larger dataset with a wider variety of polyp types and disease manifestations can help improve the model's ability to generalize to unseen data.
* Transfer learning and pre-training: Investigate the potential benefits of transfer learning by pre-training the model on a large dataset of general medical images before fine-tuning it on the specific task of polyp segmentation. Transfer learning can help leverage knowledge from related tasks and domains, potentially improving the model's performance on the target task.
* Clinical validation and deployment: Conduct comprehensive clinical validation studies to evaluate the model's performance in real-world clinical settings. Collaborate with healthcare professionals to assess the model's utility, usability, and impact on clinical decision-making. Additionally, explore avenues for integrating the model into existing healthcare workflows for seamless deployment and use by clinicians.

Overall, continued research and development efforts in these areas can further enhance the accuracy, reliability, and practical utility of the MFSNet model for gastrointestinal disease diagnosis and polyp segmentation.

# References

1. Ahmed Mohammed, Sule Yildirim, Ivar Farup, Marius Pedersen, and Øistein Hovde. (2018). Y-net: A deep convolutional neural network for polyp detection. arXiv preprint arXiv:1806.01907.
2. Basak, H., Kundu, R., & Sarkar, R. (2022). MFSNet: A multi focus segmentation network for skin lesion segmentation. Pattern Recognition, 128, 108673.
3. C.-H. Huang, H.-Y. Wu, and Y.-L. Lin. (2021). Hardnet-mseg: A simple encoder-decoder polyp segmentation neural network that achieves over 0.9 mean dice and 86 fps. [Online]. Available: https://arxiv.org/abs/2101.07172
4. D.-P. Fan et al. (2020). PraNet: parallel reverse attention network for polyp segmentation. In Proc. of Internat. Confer. on Med. Ima. Compu. Comput.-Assis. Interven., pp. 263–273.
5. DiSanto, N., Harding, G., Martinez, E., & Sanders, B. (2022). Leveraging Contextual Data Augmentation for Generalizable Melanoma Detection. arXiv preprint arXiv:2212.05116.
6. D. Jha et al. (2019). ResUNet++: An advanced architecture for medical image segmentation. In Proc. of Internat. Sympos. Multime., pp. 225– 230.
7. D. Jha et al. (2021). Real-Time Polyp Detection, Localisation and Segmentation in Colonoscopy Using Deep Learning. IEEE Acc.
8. D. Jha, P. H. Smedsrud, M. A. Riegler, D. Johansen, T. De Lange, P. Halvorsen, and H. D. Johansen. (2019). ResUNet++: An Advanced Architecture for Medical Image Segmentation. In Proc. of International Symposium on Multimedia, pp. 225–230.
9. Dumitru, R. G., Peteleaza, D., & Craciun, C. (2023). Using DUCK-Net for polyp image segmentation. Scientific reports, 13(1), 9803.
10. Jha, D., et al. (2020). Kvasir-SEG: A Segmented Polyp Dataset. In Proc. of Internat. Conf. Multimed. Model., pp. 451–462.
11. Jonathan Long, Evan Shelhamer, and Trevor Darrell. (2015). Fully convolutional networks for semantic segmentation. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 3431–3440.
12. N. Ibtehaz and M. S. Rahman. (2020). Multiresunet: Rethinking the u-net architecture for multimodal biomedical image segmentation. Neur. Networ., vol. 121, pp. 74–87.
13. O. Ronneberger, P. Fischer, and T. Brox. (2015). U-net: Convolutional networks for biomedical image segmentation. [Online]. Available https://arxiv.org/abs/1505.04597
14. Sanderson, E., & Matuszewski, B. J. (2022, July). FCN-transformer feature fusion for polyp segmentation. In Annual conference on medical image understanding and analysis (pp. 892-907). Cham: Springer International Publishing.
15. Srivastava, A., Jha, D., Chanda, S., Pal, U., Johansen, H. D., Johansen, D., … & Halvorsen, P. (2021). MSRF-Net: a multi-scale residual fusion network for biomedical image segmentation. IEEE Journal of Biomedical and Health Informatics, 26(5), 2252-2263.
16. T. Kim, H. Lee, and D. Kim. (2021). Uacanet: Uncertainty augmented context attention for polyp semgnetaion. arXiv preprint arXiv:2107.02368.
17. Tomar, N.K., Jha, D., Ali, S., Johansen, H.D., Johansen, D., Riegler, M.A., Halvorsen, P. (2021). Ddanet: Dual decoder attention network for automatic polyp segmentation. In International Conference on Pattern Recognition. pp. 307–314. Springer.
18. Trinh, Q. H. (2023, June). Meta-Polyp: a baseline for efficient Polyp segmentation. In 2023 IEEE 36th International Symposium on Computer-Based Medical Systems (CBMS) (pp. 742-747). IEEE.
19. V. Badrinarayanan, A. Kendall, and R. Cipolla. (2017). Segnet: A deep convolutional encoder-decoder architecture for image segmentation. IEEE Trans. on Patt. Analy. and Mach. Intelli., vol. 39, no. 12, pp. 2481–2495.
20. V´azquez, D., Bernal, J., S´anchez, F.J., Fern´andez-Esparrach, G., L´opez, A.M., Romero, A., Drozdzal, M., Courville, A. (2017). A benchmark for endoluminal scene segmentation of colonoscopy images. Journal of Healthcare Engineering.
21. Y. Guo, J. Bernal, and B. J Matuszewski. (2020). Polyp segmentation with fully convolutional deep neural networks — extended evaluation study. Jour. of Imag., vol. 6, no. 7, p. 69.
22. Z. Zhou, M. M. Rahman Siddiquee, N. Tajbakhsh, and J. Liang. (2018). Unet++: A nested u-net architecture for medical image segmentation. In Deep learn. med. ima. anal. multimo. learn. clini. deci. sup., pp. 3–11.