Identification, Segmentation and Localization of GI Polyps in Colonoscopy Using Deep Learning Techniques

By

Abdusamad Shakeel

0000330559 MS 20 (CE)

Supervisor

Dr. Ali Hassan

Department of Computer & Software Engineering

A thesis submitted in partial fulfillment of the requirements for the degree of Masters of Science in Computer Engineering (MS CE)

In

College of Electrical and Mechanical Engineering (CEME) ,

National University of Sciences and Technology (NUST),

Islamabad, Pakistan.

(August 7, 2024)

Annex A

THESIS ACCEPTANCE CERTIFICATE

Certified that final copy of MS/MPhil thesis written by NS **Abdusamad Shakeel** Registration No. **00000330559**, of College of E&ME has been vetted by undersigned, found complete in all respects as per NUST Statutes/Regulations, is free of plagiarism, errors and mistakes and is accepted as partial fulfillment for award of MS/MPhil degree. It is further certified that necessary amendments as pointed out by GEC members of the scholar have also been incorporated in the thesis.

Signature: _____

Name of Supervisor: Dr Ali Hassan

Date: 07-08-2024

Signature of HOD: _____
(Dr Usman Qamar)
Date: 07-08-2024

Signature of Dean: _____
(Brig Dr Nasir Rashid)
Date: 07 AUG 2024

Dedicated To

My Wife

Acknowledgments

I am sincerely grateful to the Almighty Allah, whose divine guidance and blessings have enabled me to undertake and successfully accomplish this task. His unwavering support, bravery, and wisdom have been the cornerstone of my journey. I humbly acknowledge his role in making things easier for me, and I attribute all praise and glory to Him.

I am truly humbled and thankful to my supervisor, Dr. Ali Hassan, for his invaluable support, inspiration, and guidance throughout my academic endeavors. His unwavering commitment to my growth, both academically and personally, has been instrumental in shaping my success. I am lucky to have had such a caring mentor who invested time and effort into lifting my spirits and providing vital advice.

I am indebted to my entire thesis committee for their cooperation, wise comments, and expertise that has contributed significantly to the refinement of my work. I am thankful for their commitment to academic excellence and their intention to share their knowledge.

I would also like to show heartfelt gratitude to my beloved parents, whose unwavering support and love have been my guiding light. Their support, sacrifices, and confidence in my abilities have been the bedrock of my accomplishments. Finally, I want to extend my appreciation to all my friends and well-wishers who have stood by me and provided assistance and encouragement during this journey. Their presence and support have been a source of strength and inspiration.

Abdusamad Shakeel

Contents

List of Figures

List of Tables

LIST OF ABBREVIATIONS

Abstract

Computers have improved the old colonoscopy procedures as they are helpful in the early detection of the problem and help in pin-pointing the issue by highlighting the main areas of the interest. With the improvement in computer technology, various computer vision methods have been developed which can be applied on polyp datasets to automatically detect the polyps by outlining its boundaries. But, it is still an on-going challenge to find the best method for the early detection of polyps. Finding the best method for polyps' detection not only ensures that the results generated by this method are reproducible but also helpful in completing the tasks of Automated identification and segmentation of polyps with greater accuracy. Moreover, it is important to consider that the ideal method should be consistent with industrial standards and should has results that are comparable to other used methods. The purpose of this research is to identify a benchmark method, by analysing several different methods, by publicly available dataset i.e. Kvasir-SEG which contains hundreds of images obtained from colonoscopy that can be used for detection, segmentation and pin-pointing of polyps. By doing this the speed and accuracy of the benchmark method is also evaluated. Most methods which are already proposed in literature perform well but they do not provide accurate results thus these methods cannot be used as a benchmark method. For identification and localization of polyps, ColonSegNet, method has been used in this study which provided satisfactory results with a mean AP and mean IoU of 0.8000 and 0.8100 respectively. This method has the highest speed of 180 frame per second (fps) which is more than enough to prove that the method is useful and can be used as a benchmark method. Not only this, the proposed method has also performed well in segmentation task by achieving 182.38 fps average speed with a comparable 0.8206 dice coefficient. The detail comparison carried out in this study by using different and advanced methods highlights that it is critically important to benchmark deep learning techniques to automate polyps

identification and segmentation process. The benchmarking can help us to transform clinical practices, which are already in use, to be more accurate and efficient. Thus, the chances of missing polyps' detection during clinical examination may reduce and a significant improvement in the outcome of clinical examination can be achieved.

Keywords: Kvasir-SEG, polyp detection, localization, ColonSegNet, deep learning, colonoscopy

Chapter 1

Introduction

1.1 Motivation

Among all cancers the death rate of Colorectal Cancer (CRC) (Colorectal Cancer) is the third highest rate. For cancer of colon the survival rate is about 68% and for stomach cancer there is only 44% survival rate in five years [25]. To decrease the death rate based on CRC, searching and removing of dysplasias (cells that have a higher chance of becoming cancer) is one of the best methods. In the presence of these anomalies, it is very important to detect the polyps in the colon this is because these polyps at later stage can be develop into the CRC. So that for survival the CRC detection in the early stage is very crucial.

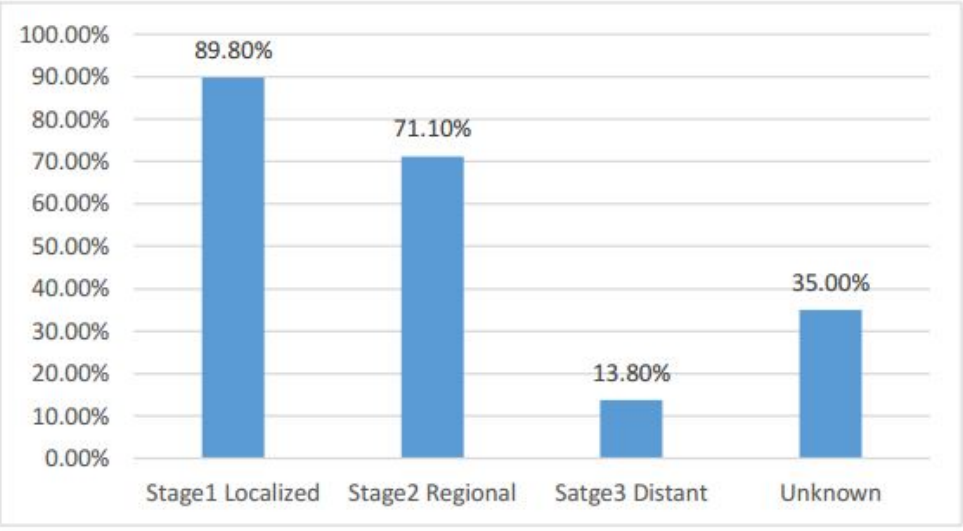

Figure 1.1: *Colorectal cancer survival rate for five years at different stages.*

With the passage of time lifestyle changes, regular colon screening takes place for CRC preventions. To reduce the number of CRC cases and also diagnosis the CRC various studies indicates that the population-wide screening is valuable [5]. The medical process in which medical health professionals (endoscopists) examines polyp and then operate it by using flexible endoscope is known as colonoscopy. In earliest methods, it is the best procedure for the examination and removal of the polyps. That's why the clonoscopic screening is one of the best and popular technique for diagnosis of polyps among different gastroenterologists.

The different developed polyps are mostly cause of the colon cancer which is unwanted. In early time, in most cases CRC avoided when the polyps were detected [39]. And if the detected polyps were not removed before further development in polyp, the risk of becoming the cancer was increased and after 5 years risk is 2.5%, after 10 years risk is 8%. After 20 years the risk is alarming and it increased to 24% [23]. From cancerous polyps almost 20% colon cancer are derived. This contains precancerous cells that might be further developed and become a cancer. After treatment CRC can be avoided for 5 years, although this depends on the ability of endoscopist to detect and remove the polyps. Overall, the polyp removal process is uncomfortable and disturbing process for the patients so that some people not take interest to examined at all.

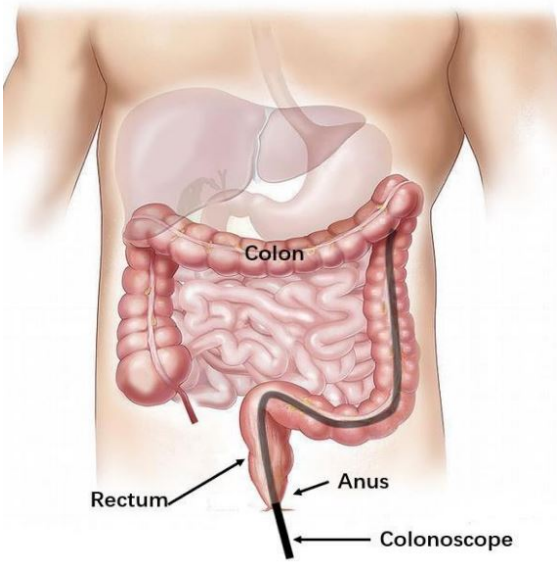

Figure 1.2: *Colonoscopy Procedure Explanation.*

The datasets that are stored and recorded during the examination in the hospitals are put to use the various artificial intelligence (AI) artificial intelligence techniques for diag-

nosing the gastrointestinal tract diseases and also detection and removal of the polyps. These artificial intelligence techniques facilitate the doctors, reduce the manual work and build the complete solution that improve the detection rate of polyps. Otherwise, the polyps can be easily missed by the less qualified medical practitioner.

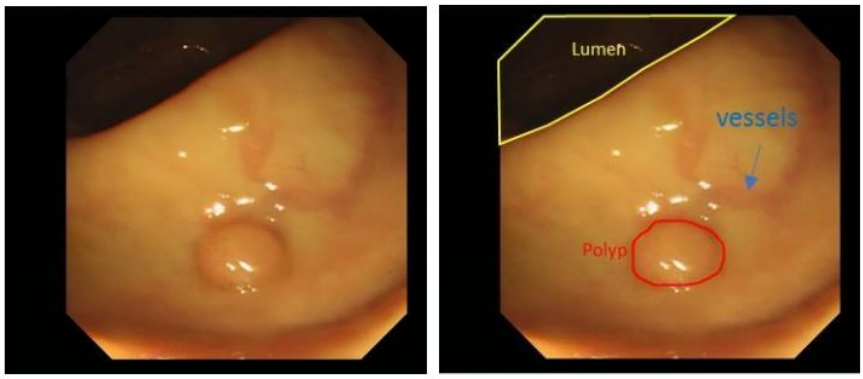

Figure 1.3: *An image of a polyp with a brief explanation of its structure that is visible.*

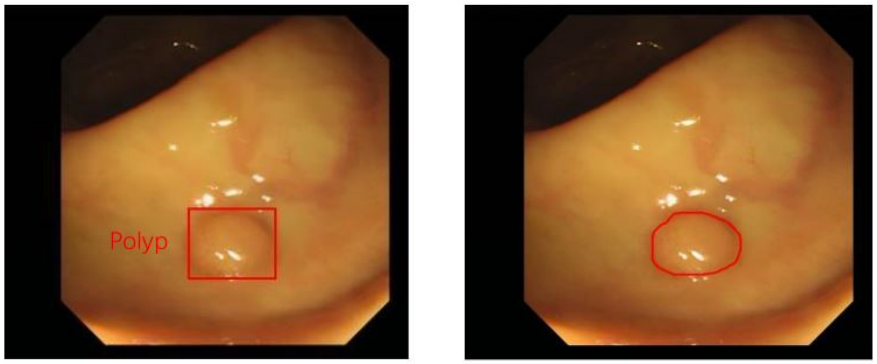

Figure 1.4: *On left polyp localization and detection on right polyp segmentation.*

1.2 Problem Statement

As discussed in motivation section of thesis, the main purpose of this research is to explore methods of improvements in already available automated detection systems of polyps as these systems have a great potential for improvements to give better accuracy in outcomes. The identification and then localization of polyps in HD frames obtained from colonoscopic determination these models will served for utilization and configuration of manufacturing of diagnostic systems which is based on deep learning methods. EIR (name of healing goddess in Norwegian method) [59] is the basis of this entire

system and multimedia research system which serve as a tool for Gestrointestinal (GI) tract polyps' detection. In this research Kvasir-SEG; publicly available dataset was used which contains images having annotations obtained from real time colonoscopy examination [20].

Following are the research objectives was addressed and discussed in this thesis.

Can deep learning method (Artificial intelligence technique) used for detection and localization of polyps gives us a well-defined, complete, correct and accurate solution of any medical problem and having comparable results with results obtained from manually diagnosis of same polyps. For trainings different models many samples have been required by used algorithms relies on deep learning techniques. Here as discussed earlier, Kvasir-SEG has been used as algorithm for trainings and testing of models which contains annotated images for polyps' objects present in bounding boxes [1].

To address above mentioned questions of research, we set following objectives.

1. Training model for identification and localization of object in obtained medical HD frames of GI (Gastrointestinal tract) has been implemented.
2. The results obtained from training models on unseen images from gastrointestinal tract has been validated and analyzed and then compare them with values of ground truth.
3. Compare the resulting evaluation and detection values of training models for object with the already present highly developed standard models.
4. Finally, the model that has been detected can be used to solve other medically related other problems like classification and segmentation of polyps.

1.3 Limitation and Scope

The polyp detection system has many factors regarding limitation when we design and implement the system. It's a big challenge to classify the polyp, this means that still there is much need to do work that determines which feature gives the best result in detection of polyp and to find how these features can maximize the polyp detection rate. In our research, we choose the features that are available, and these are not much resource demanding and suited for the detection system and the implementation of fast

algorithm for the classification of the polyps that does not affect on the FPS (Frames Per Second). In the research, Kvasir-SEG dataset that is publicly available has to be used for our research and this dataset has to be categorized into two main parts i.e. train and test. For the cross-validation stage these two main splits has to be used. The different features are extracted from the Kvasir-SEG dataset that have available, it means that the desirable performance and accuracy is also limited and this depends on the dataset and the two main splits. We found some images that are not suitable for the research and these can also affect the results and output results.

For this research, our main concern was on the polyps and how can the system detect the polyps with focusing on the performance and accuracy, however there are many types of diseases of the GI tract.

It's a challenging task for us this is because there are several diseases can occur in the colon which can hide the different polyps. We are facing different challenges focusing on the large scale publicly available datasets for the medical research of GI diseases.

For the collection of data different hospitals used different equipments. These equipments can be of different light conditions, different resolutions and other characteristics. Form this scenario that uses of different equipment the images quality in terms of resolutions, light condition and other attributes are different. It's a big challenge for us to develop a system that can neglect and remove these differences and constraints. While our system aims to detect the polyps, our goal is to create such a modular system that can be helpful the other diseases of GI tract in future.

We still need to focused so that our system should be general and can be extended for the identification of other diseases when the required dataset is available in future. The system should also be checked by using different cases to ensure that the system is generic and efficient to detect the different diseases of GI tract.

1.4 Research and Method

Machine Learning, image processing, deep learning are the different fields of scientific research. When we design our system there is still gap for the further improvement for the identification and categorization of the polyps. In order to obtain the desired results, first we have to study and target on the processing of images, machine learning

and deep learning. This has to do for the sake of getting knowledge and understanding of:

- How a computer experiences the dataset.
- How image processing highlights the region of interest (ROI) in the frames and also which features can be helpful for this.
- How the deep learning algorithm is applied for the system and how we can used this for desirable results.

These tools have to be evaluated with the physician by discussing the system and the development of system for the physicians as users. Additionally, by using different deep learning techniques we developed a system for the training of the classifier. The system and the classifier both are coded together. The Kvasir-SEG dataset has been ready by the implementation of the software for the feature extraction. This is very necessary in order to get the results and to found how the system can be further improved in the future.

We have experimented with different parameters and techniques for the object identification and made the strongest matches of the output. Finally, evaluation of the methods is done with the machine learning techniques, FPS and the performance of the identification system.

1.5 Main Contribution

The summary of our main contribution is given as:

1. ColonSegNet which serve as an architecture of encoder-decoder for segmentation of medical images from colonoscopy as it is very good for processing images with remarkable speed and also highly competitive regarding performance.
2. There is a comparison of the publicly available Kvasir-SEG dataset with the already available SOTA methods that provides detection, localization, and segmentation of GI polyps with an exceptional accuracy and speed.

3. On selected dataset i.e. Kvasir-SEG, very strong and powerful benchmark for polyps' detection and localization has been established [39][48][59]. These benchmarks can then be used to generate medically acceptable and reliable techniques.

4. The values of performance obtained from detection, localization and then segmentation of polyps are analyzed by visual AI metrics.

5. The best and worst performing situations has been deeply studied and presented in this thesis that will enable us to analyze both successful and failed aspects of method to expedite development of algorithm.

Chapter 2

Background and Relevant Work

The gastrointestinal tract contains all those organs that are required for digestion process. It has upper and lower tracts. The upper tracts include mouth, esophagus and then stomach. After upper tract the lower tracts started which has small and large intestine or colon. The last part of gastrointestinal tract (GI) Is known as large intestine or colon. The main function of colon is to form stool by removing waste substances or extra water, salts and different nutrients from body. Colon or large intestine has four main parts which are as follows.

- **The ascending colon:** It is the starting point of colon which travels up to the right part of abdomen.
- **The transverse colon:** It is present after the ascending colon and hepatic flexure and move throughout the abdomen.
- **The descending colon:** It begins after transverse colon and splenic flexure.
- **The sigmoid colon:** The last part of large intestine which connects to rectum is the sigmoid colon.

2.1 Diseases of gastrointestinal tract

Following are the diseases of GI tract that can influence its proper functioning. The treatment, cost, and time of recovery is different for each disease which depends on severity of that disease.

- Ulcerative Colitis (UC)
- Colorectal cancer
- Irritable bowel syndrome (IBS)
- Colonic diverculitis

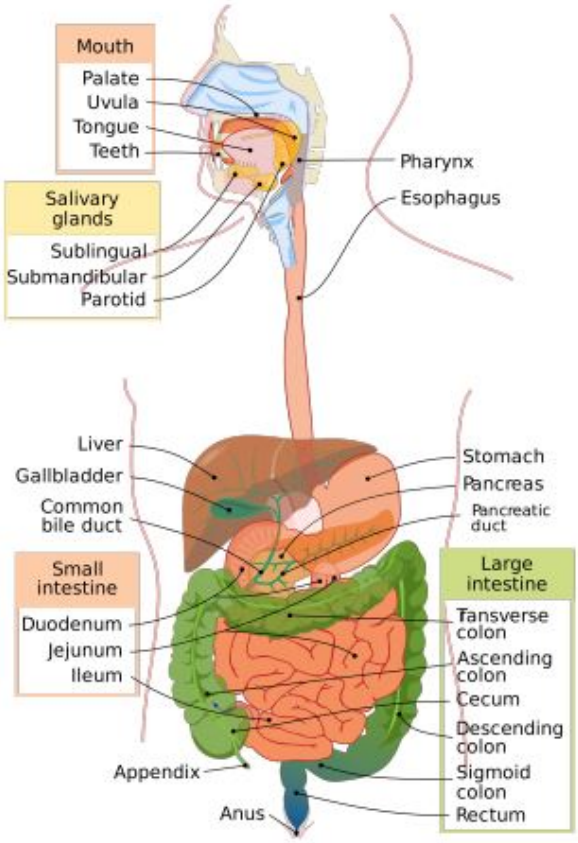

Figure 2.1: *GI Tract of Human*

2.2 Polyps in GI tract

Formation of polyps or lesions occur in the innermost lining of colon which is known as mucosa. Polyps are basically unwanted changes in tissues of large intestine. The development of polyps mainly depends on age, unhealthy lifestyle (smoking), genetics and weight of people. Some of these polyps are not very dangerous but some polyps will turn into cancer with the passage of time. So, early detection and then removal of GI polyps should be done to avoid any kind of loss.

The are generally 2 main groups of colon/large intestine polyps. Neoplastic Polyps which consists of the hyperplastic polyps, pseudo or inflammatory polyps and hamartomatous polyps (HPs) while the adenomas (most common type) and serrated polyps that do not usually transform into cancer are neoplastic polyps. Generally, possibility of transforming neoplastic polyps into cancer is directly related to size of polyps i.e. longer the polyps present in body more is the chance of them to convert into cancer. The purpose of this research is the early identification of colon polyps which would be helpful in medical field.

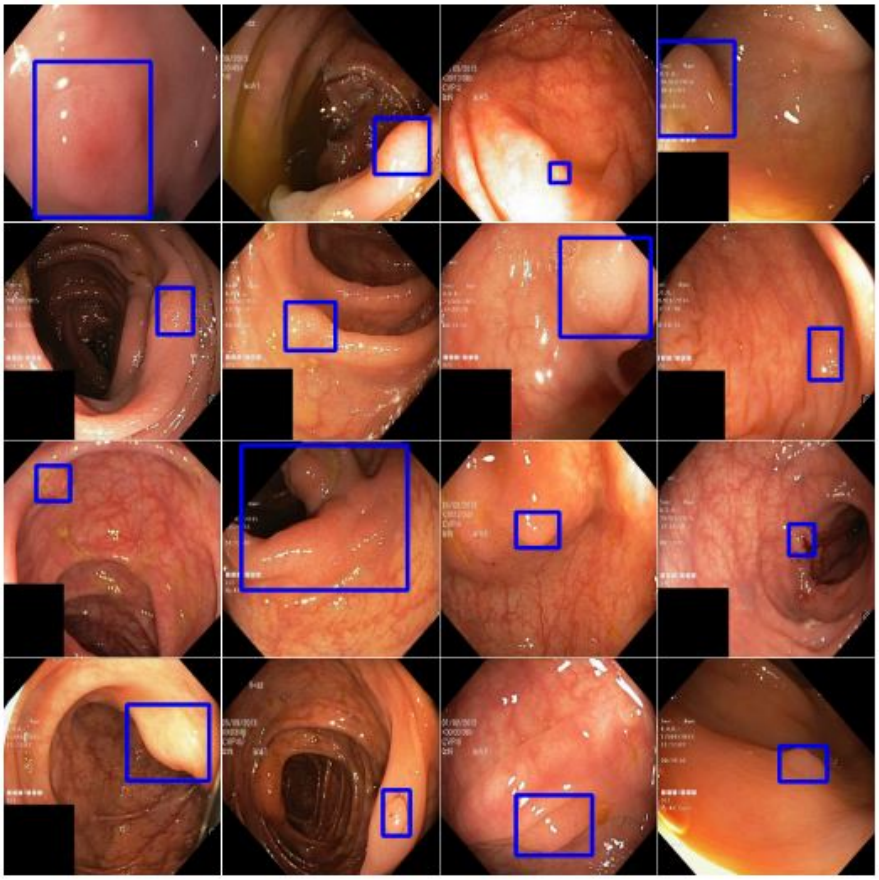

Figure 2.2: *Sessile and Flat Polyps*

CRC which is also known as colon or rectum cancer depending on its development area. It started with smaller non-cancerous growths (polyps) on tissues of inside colon. With the passage of time if these polyps remain undetected then some of these polyps will grow in size due to abnormal cell divisions and become colon cancer. However, even during this stage if polyps can be detected by different screening methods, then they can be prevented from further spreading.

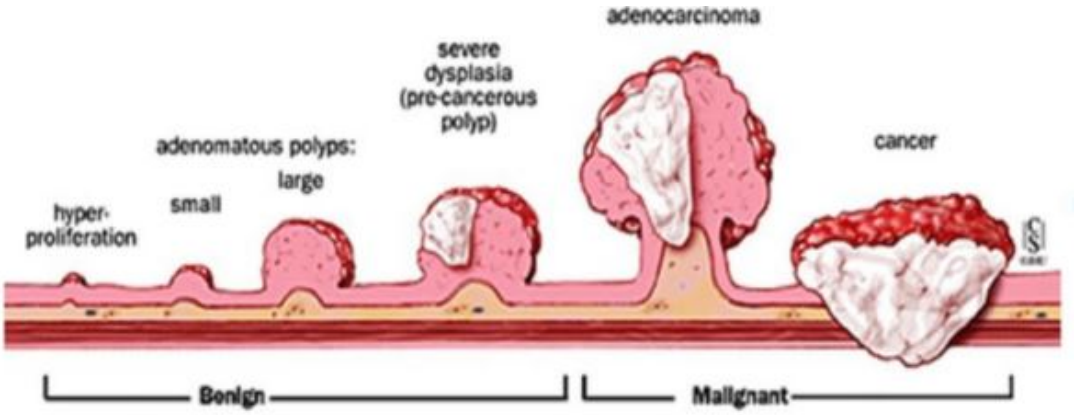

Figure 2.3: *Development stages from polp to cancer*

There are five stages of CRC [3]. The Stage A starts when polyp does not appear from the most internal layer, the mucosa, of the colon. During the stage B of the CRC, the polyp starts to grow from the mucosa but during this stage the polyp has not yet reached the lymph nodes or any other areas which are associated with the colon. In another stage i.e. Stage C the polyp starts to grow out of the colon wall and reach to the neighboring lymph nodes and other organs of the body. The polyp then starts to grow into a thicker layer of the muscle during this process the muscle contracts and then force the contents of the intestine with itself. During Stage D of the CRC, the polyp reaches the tissues and then grows into or through all these tissues and ultimately starts to surround the colon or the rectum area. In this stage of the CRC the growth caused by the polyp can be dangerous and even fatal. The surface of visceral peritoneum is attacked by polyp which grows into its surface and then comes out from all the layers of the colon. The polyp then starts to attach with or spreading to the organs which are nearby such as lungs and liver and various other inner structures of the body.

The Table 2.1 shows the rate of survival at different stages from colorectal cancer. The result present in figure shows the importance of detection of polyps at earlier stage. According to the results, patients of stage 0 & 1 (A & B) have survival rate of 93.2% and 77% respectively While, the survival rate for patients at last stage is about only 6.6% which is so alarming within 5 years of specific time period which reflects the importance of polyp's detection at earlier stages. The reported number of cases also increases from 26,727 to 106,040 (170%) from stage 0 to last stage that means only small number of patients have detected the presence of polyps and it may be due to lack of any symptom of polyps at earlier stage and when the symptoms start to appear people notice them

and go to medical health professionals for check-up. So, Routine check-up for polyps is very necessary for people after certain age. The early detection and treatment will drastically reduce the number of polyps that will finally grow to CRC stage.

Diagnosis Stages	**Spoted Cases**	**% cases (%)**	**% cases excl. non-identified (%)**	**5-year survival Rate (%)**	**Interval of Confidence (95%)**
A	26,727	8.7	13.2	93.2	92.5 - 93.9
B	74,784	24.2	36.9	77.0	76.4 - 77.5
C	72,806	23.6	35.9	47.7	47.1 - 48.3
D	28,377	9.2	14.0	6.6	6.1 - 7.0
non-identified	106,040	34.3	-	35.4	35.0 - 35.8
Total	**308,734**	**100.0**	**100.0**	**50.7**	**50.4 - 51.0**

Table 2.1: *Rate of survival (1996-2006) at different stages from CRC*

2.3 Methods of Screening

Gastrointestinal tract is often at risk of various infections, cancers, diseases and inflammations. Early detection and then suitable treatments help to recover patient timely, improve prognosis of disease and reduce death rate due to GI tract diseases. There are several methods of screening are present and utilize them at clinical setup but each of these methods have some issues and difficulties that are also discussed. The cancer Facts and Figures 2020, a publication of American Cancer Society's (ACS) has provided statistics of mortality due to colorectal cancer which showed that the mortality or death rate due to CRC in 1970 was 54% more than the mortality rate in 2017. This improvement in statistics is due to improvement in methods of treatment and advancement in screening methods that will result in early detection of disease and abnormalities which further help to cure them at earlier stages and prevent them from turning into cancer. According to guidelines of American Cancer Society 2018, it is recommended that the persons above age of 45years should do their regular checkups and gastrointestinal screenings because colorectal cancer have no early symptoms until it reaches to any se-

rious stage so, in order to prevent them from gastrointestinal diseases or timely recovery from early treatment if detected. Following are the current SOTA methods of screening for gastrointestinal tract. For prevention and timely treatment of GI tract diseases, the American Cancer Society (ACS) 2018, recommended that people above 45 years of age should regularly visits medical health professionals for their proper checkup and screening process of colon and rectum. As diseases or CRC of GI tract does not show any symptoms unless it reaches to its advance stage where treatment could not possible, so it is considerd as one of the main issues related to CRC. Hence, timely screening of GI tract is necessary for prevention and advancement of polyps to any serious stage. Following are the details of some standard GI tract Screening methods with their associated specific characteristics i.e. efficiency, cost, comfort level for patients and also their possible challenges that are discussed in detail.

2.3.1 Colonoscopy

In order to examine the lower GI tract, we use the colonoscopy procedure. In the rectum and large intestine which is also called colon, colonoscopy is used for the detection of any changes and medical abnormalities. Strong medicines or anti-anxiety medicines are given for the colonoscopy. A flexible long tube which is called colonoscope has to be inserted in the rectum and a small camera shows the different pictures of the large intestine on the screen. In case of any medical problem, colonoscopy is used for the detection of different symptoms and intestinal signs. Colonoscopy is also used for the identification and screening of polyps as well as CRC. Through the colonoscopy the removal of polyps and for taking the different tissues samples has also been performed[68].

2.3.2 Flexible Sigmoidoscopy

To examine the lower part of the large intestine we used flexible sigmoidoscopy. For sigmoidoscopy a small flexible tube which is called sigmoidoscope is inserted into the rectum during the process of sigmoidoscopy. In this process generally the strong medicines are not required before the process. A small camera is placed on the top of the sigmoidoscope which shows the images to the doctors inside the rectum. The colon of the sigmoid and colon of descending which is the 50cm of the colon (large intestine). The whole large intestine cannot be shown on the camera through the process of sigmoi-

doscopy. So that, this process not suitable to detect a single polyp in the large intestine and can be develop to colon cancer [2].

2.3.3 Virtual Colonoscopy

In order to examine the large intestine for cancer and polyps which later becomes colon cancer, a minimally invasive technique is used for the colon examination which is called virtual colonoscopy. In later techniques, a tiny camera is inserted through the tube and doctor examine the large intestine onto the camera. However, if we talk about the virtual endoscopy a CT scan of the patient has to be done and for the generation of the multiple cross-sectional images to examine the abnormal changes in the large intestine. After this process and CT scan all the images have to combined and digitally manipulated to take the complete and digitally report of the rectum and large intestine. So that virtual colonoscopy canbe utilized for the colon cancer CRC[36].

As we discuss that in the virtual colonoscopy there is no need of tube or colonoscope, so that virtual colonoscopy is more preferable for some patients because of no need to insert camera into the rectum. An also there is no need to get the proper medicines and patients can normally go to their home and get back to normal daily life routines. As there is no need of the medicines so there is lower risk to the peoples to get harmful reactions through the medicines in the traditional process of colonoscopy. The virtual colonoscopy is less time taken process as we compared to the SOTA colonoscopy. In order to get the information about the polyps and other diseases outside the large intestine, virtual colonoscopy has more benefits for this purpose.

2.3.4 Wireless Capsule endoscopy

In the process of the wireless capsule endoscopy, a small tiny camera which is the wireless is used and the small camera is easily fit into the small capsule of vitamin sized which can be swallowed easily. The camera containing small vitamin sized capsule is inserted in to the large intestine and can be moved easily into the intestine and make hundreds of images and these images transferred to the device that is linked to the wearable belt. The advantages of wireless capsule endoscopy also includes the examination of small intestine which is otherwise very difficult in the other standard processes of endoscopy. Through the process of exertion, the swallowed capsule is out of the large intestine after

24 to 48 hours. So that we can say this process of endoscopy is much safe process as compared to the other methods of endoscopy [57].

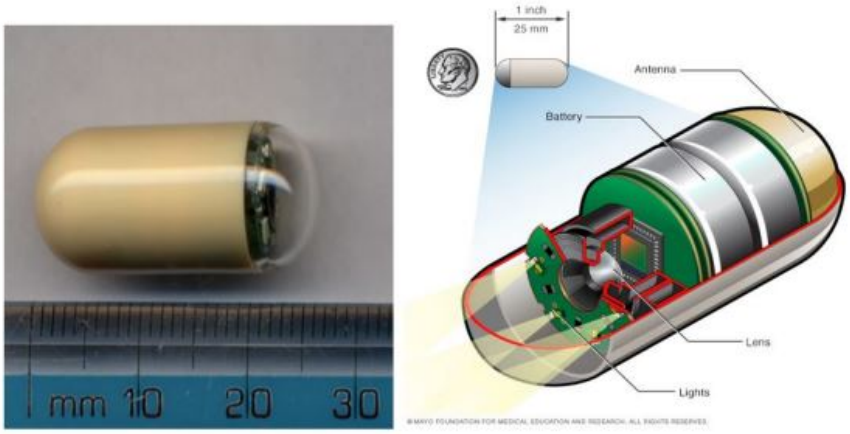

Figure 2.4: *Wireless Capsule Endoscopy*

2.4 Artificial Intelligence in the endoscopy of GI tract

The disease of the GI tract can be identified and diagnosis through the GI endoscopy which is the golden way of treatment. Across the world around about 2.9 million of diseases of GI tract every year has been detected and diagnosis. These diseases include stomach and colorectal cancers etc. Some of these diseases has been diagnosed through the better and highly performed endoscopy and other screening techniques. These diseases affected the quality of life of an individual because the death rate is about to 66% making CRC is the second largest cause of the death.[63].

In the previous 10 to 15 years, the quality and performance of the different endoscopes has been improved. But the result and performance of the examination of endoscope is highly depends on the level of expertise, attitude of work, knowledge of polyps and set of skills of the endoscopists. There are different reasons of get missed polyp rates as we use old instruments, old monitors, less preparation of the colon, stress and human errors done by the endoscopists. There is 20% chance of miss the polyps during the process of colonoscopy.

In the research field of medical, AI has very great potential to identify and then to remove the colon polyps because CRC associated the health risk. And AI also perform good in polyp detection because if the polyps cannot be detected by the model, then it

developed to the CRC. So that, if we use the different techniques of AI in the medical field and the identification and removal of polyps with better and improved results.

2.5 Techniques of Artificial Intelligence

Technical revolution has been brought in different fields by the artificial intelligence. There are number of uses and applications of AI techniques in medical field. Artificial intelligence has been studied and researched in the gastroenterology field which is the medical field with the AI inverse applications like as risk identification, pathological detection and diagnosis. This topic is very interesting in the endoscopy and put the more potential in gastroenterological field. Artificial intelligence has completely changed all the areas of modern endoscopy form screening of cancer to generation of report automatically.[46].

If we talk in a simple manner, AI referred to the intelligence that is produced by the different machines and this is inverse of the naturally produced intelligence of animals and human beings.[67]. The Fig 2.5 reflects hierarchical structure of AI domains. Machine Learning is the branch of AI where the patterns of the dataset has been utilized to learn varoius algorithms as an programming alternative. [28][51]. A sub field of the ML is (RL) Representation Learning in which the training algorithm pick the best feature which is utilized for the data classification. Now the () is the type of representation learning in which the combination of different features has been taken that are represent the different hierarchical structures for the data that classification result of output image. If we talk in a simple manner, AI referred to the intelligence that is produced by the different machines and this is inverse of the naturally produced intelligence of animals and human beings.[67]. The Fig 2.5 reflects hierarchical structure of AI domains. Machine Learning is the branch of AI where the patterns of the dataset has been utilized to learn varoius algorithms as an programming alternative. [28][51]. A sub field of the ML is (RL) Representation Learning in which the training algorithm pick the best feature which is utilized for the data classification. Now, the Deep Learning (DL) is a representation learning in which the combination of different features has been taken that represent various hierarchical structure of data.

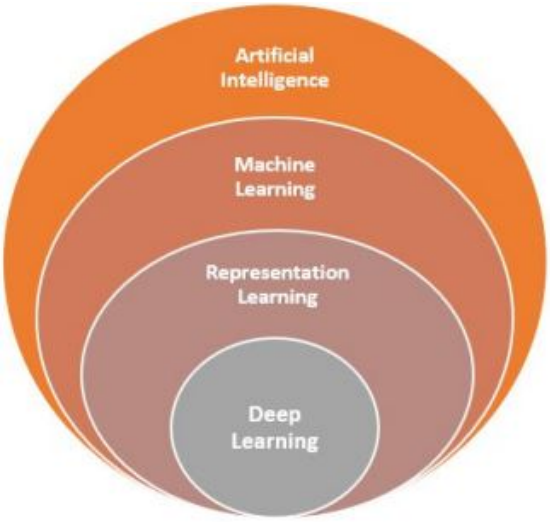

Figure 2.5: *Hierarchy of AI Domains*

2.5.1 Deep learning techniques

Deep learning algorithms are the type of machine learning algorithms. In the article [6] describe the deep learning as deep learning consists of specific characteristics of the input training data from the training input data and for the extraction of higher level it contains the multiple layers. If we talk in the field of image processing, the edges could be identified by the lower layer and different human concepts like digits, faces and numbers etc. are identified by the higher layer [70]. The Fig 2.6 depicts the hierarchy of deep learning algorithm that contains the different layers. These layers are used for the classification and extraction of features to and also feature extraction from the input data.

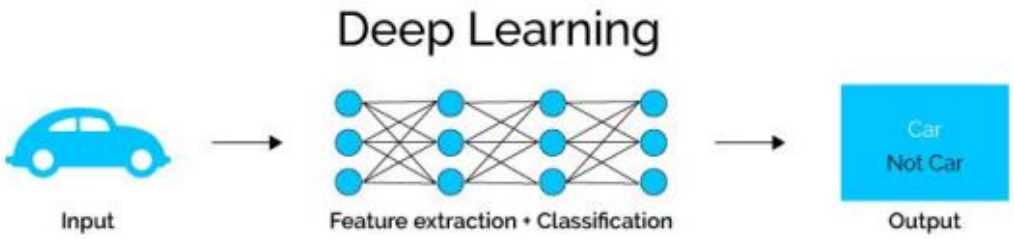

Figure 2.6: *Arcitecture of Deep Learning*

2.5.2 Convolutional Neural Network

The class of neural network which is specialized in the analysis of visual image like as classification and recognition of the images is called Convolutional neural network (CNN). A human brain starts the processing a huge amount of data when we look to an image. If we talk about the visual field each neuron is further connected to the other neurons and working in the respective field. As the human brain neuron work only its respective field CNN neuron also process the data that is present in its relevant field. The organization of CNN layers happen in such a way that they first of all they identify the patterns that are simple like as curves, lines etc. and after this they identify the complex patterns like as objects, faces etc [56].

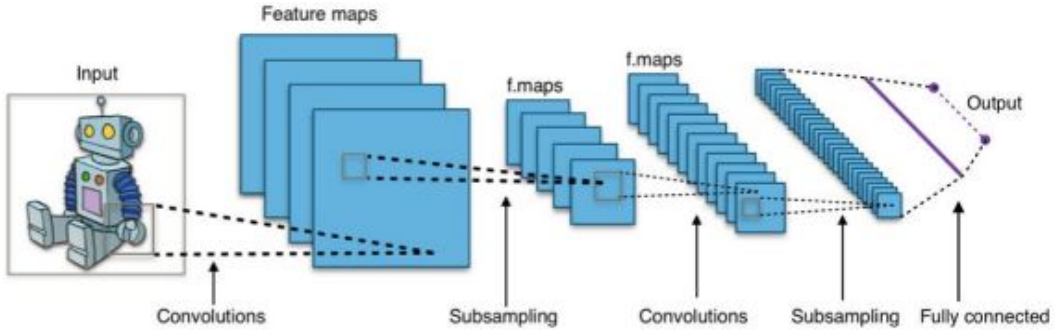

Figure 2.7: *Architecture of CNN*

A CNN contains pooling, convolutional and completed connected layers. To extract the different features pooling and convolutional layers are used as a main component of CNN, the completed connected layers are served for the classification purpose. For the extraction of key features the multiple featured maps has been created during the filter applying phase of the input image. This step of pre-processing is called convolution. In order to make the CNN more successful and to get the bigger section of the input image the featured map has been compressed to the smaller size by the pixels of pooling. The pooling and the convolutional layers iterated many times. By combining all the features, the fully connected layers produce the final result [69].

2.5.3 Computer Vision

In order to analyze, visualize and understand the different aspects of videos and images in the computer CV (Computer Vision) is used to obtain better results. This problem has

been simply solved by the humans trivially with the impression that it is straightforward. Due to change in the physically world and the complexity of visually environment it still open and an unsolved problem [71].The problem has been solved for computers by the use of CV. It is basically the branch of AI and ML that uses the different specialized methods and used the general learning models. The main application of CV is to identify and recognize the different things in an image.

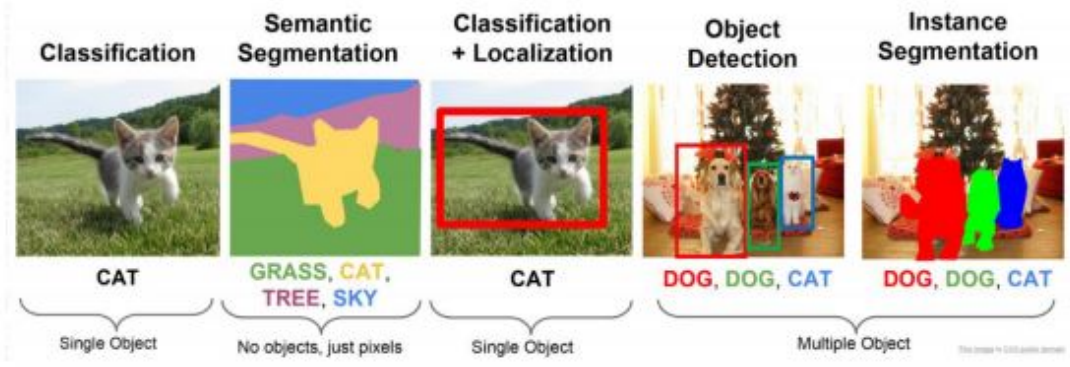

Figure 2.8: *Tasks related to Computer Vision*

Object recognition includes different tasks such as detection of objects in an image. Image classification targets the prediction of object's class founding in an picture. The localization of object targets to identify the location of the objects in an image and also draw a bounding box across the object. The localization and classification together known as object detection.

Following is the input and output explanation for computer vision.

Image Classification: It predicts the type or class of the object in the given image. Image containing one object is taken as input. While the image with proper labelling of class and specifying the categorization of input image is termed as output.

Object Localization: It identifies the object in an image and shows its location with bounding box. Image taken as an input with the single object. Output is an image with different bounding boxes that specify the location of the object.

Object Detection: It locates the existence of the object with bounding box and also find the class of the object present in an image. Image taken as an input with the single object. Output is an image of bounding boxes with label of class assigned to each bounding box and also class of the objects that are identified.

Object Segmentation: It shows the specified pixels that are belongs to the different objects in the image. Image taken as an input with the multiple object. Output is an image of the highlighted different bounding boxes for the objects in the picture.

The major concern of this research is to identify the polyps and then localize them in the input image during the process of the endoscopy. Following this, next section explain the various object identification and localization SOTA methods.

2.6 Object Detection Models

As the previous section described that the detection of object is the process of the identification and then labelling of the object in the video or in an image. The different object identification methods have been used to train the vast amount of data and then this data turns to process the new data. The component of the object detection methods is the bounding boxes which are their key component used to make a square or rectangle box around the object in an image or in a video. The bounding boxes identifies the different edges of the detected object in the image or video. The bounding box also labeled so that it identified the object is a cycle a dog or a person. If there are multiple objects in image or in a video the bounding boxes are overlapped and this shows that the image has multiple objects.

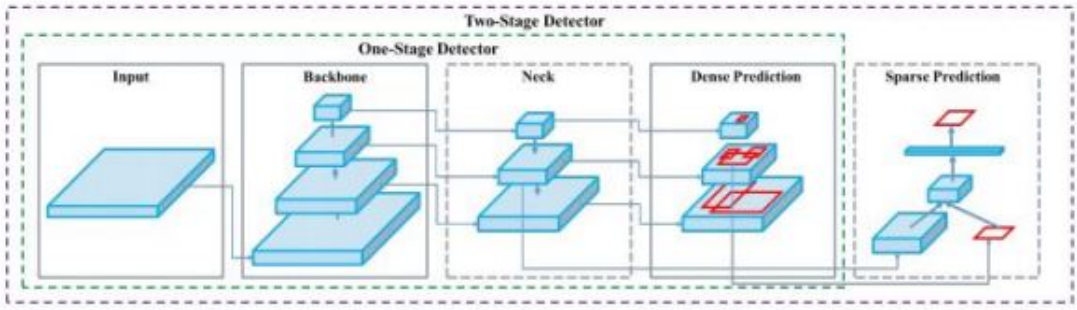

Figure 2.9: *Structure of two stage object detector*

Machine learning and deep learning are two main techniques that are used for the detection of the objects in image or video. the deep learning technique is used in the previous SOTA methods this is because it requires the minimum human interception. If we use the deep learning for the object detection than we have to use the idea of convolutional neural networks CNNs. The CNNs used the idea as neurons of the an

individual brain. If we learn the neural network it depends on the training of data which can be supervised, semi supervised or unsupervised. Bu the use of less manually engineering the deep neural network has ability to detect the single or number of objects in the image or in a video and give the best and accurate results.

Different identification methods are used to identify the type of the object and also produces the bounding boxes around the object which are done by the localization methods. If we talk about the categories of the object identification models so there are mainly two categories of object detection models which are briefly described.

Classification-based algorithms: CBA has two stages first stage is to select the ROIs (Region of Interest) in an image or in a video having higher chances of detecting the object. In second stage the CNN is applied on the ROIs for the identification of the objects. There is a disadvantage of this method is that CNN has to be applied on every ROI so that it may be slow and become more costly. R-CNN, Fast R-CNN and Faster R-CNN are the models of classification-based algorithms.

Regression-based algorithms: In RBAs there is no need for the selection of the ROIs. In these methods the bounding boxes and they the type from whole image is predicted at once. This process makes the RBAs faster than the CBAs. YOLO is used for the RBAs methods which is the faster then all the CBAs models.

2.6.1 R-CNN

Girshich, et al. has described R-CNN in paper [7] in 2014. This model has based on the CNN and used for the identification and localization of the objects in HD frames and videos. These models have been used in different stat-of-the-art methods. The Figure 2.10 reflects the structure of R-CNN. The R-CNN includes the different modules that are described in the following:

1. **Region Proposal:** For the generation and extraction of the region proposals that are the independent on the class of the object.

2. **Feature Extractor:** This module is used to extract feature from every candidate by using CNN.

3. **Classifier:** This module is used for the classification of the object from the specific known classes.

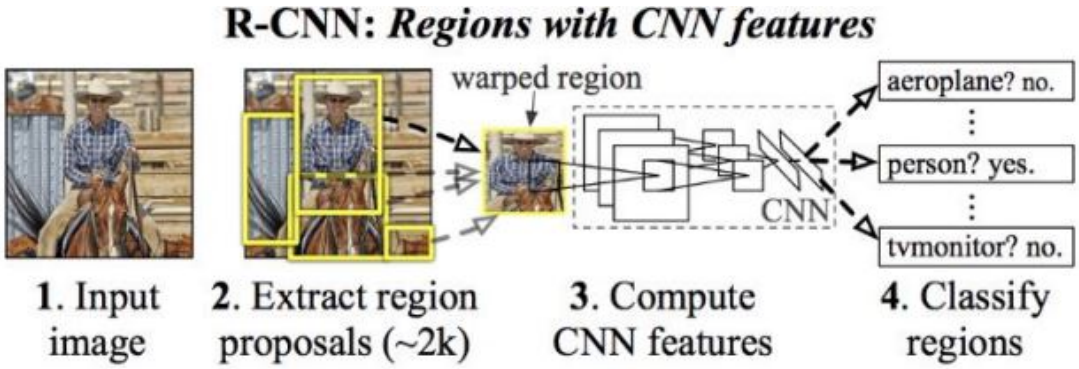

Figure 2.10: *Overview of R-CNN object detector*

For the identification of the interested objects in an image a selective search algorithm which is the Computer Vision algorithm is used. This algorithm proposes the regions or the bounding boxes for the detection of the objects. R-CNN also used the algorithm for the region proposals. R-CNN is one of the simplest techniques of CNN so that is has some disadvantages in detection of the objects. The data has been trained in the R-CNN in a pipeline which is the slow process and costly. And also, it takes time in the CNN-based feature extraction from each algorithm of proposed regions. And the identification of the object by using R-CNN has less speed.

2.6.2 Fast R-CNN

The fast R-CNN model has been presented in 2015 in the paper [12]. This paper has been presented to address the problems in terms of speed of R-CNN. The team of researchers has been proposed a method which is much more simpler method as compared to series of extraction and learning of the regions and then classify them directly. The structure of the Fast R-CNN is shown the Figure 2.11. The architecture shows that image is taken as an input, a set of proposed regions and then this set passes through the deep CNN. In order to extract the features or-trained CNN is used fir this purpose. Custom layer which is also known as ROI pooling layer is the last layer of CNN and is used for the extraction of specific information from the desired region. After that, the results obtained from the CNN is examined by the fully inter-related layers. This fully connected layer divides into two parts one part is for prediction of class and other class is for the results of the bounding box. This process repeats many times for each ROI in the given input image. For this method training of data and make the final prediction is much faster

than R-CNN. But each and every input image required a set of regions which has to be proposed by the model.

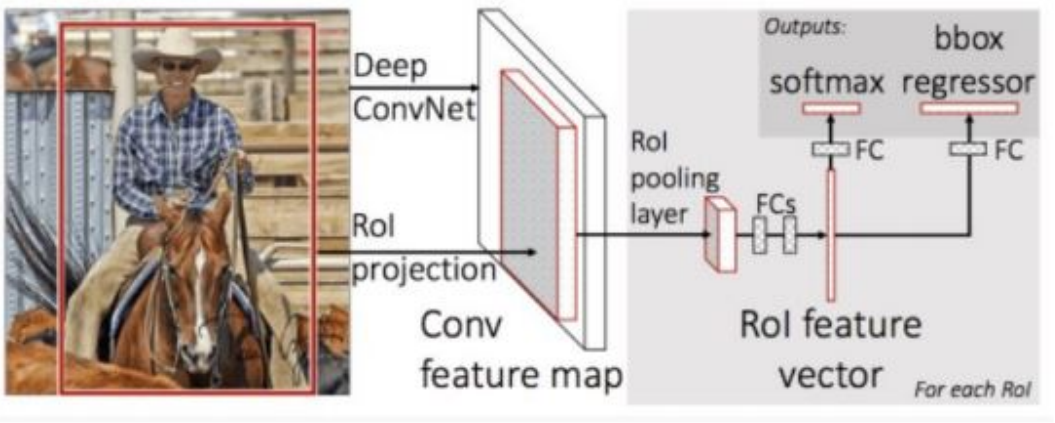

Figure 2.11: *Structure of Fast R-CNN*

2.6.3 Faster R-CNN

In 2016 an improved method is proposed in the paper [14]. In this model the design of the model Fast R-CNN model has been improved for data identification and training of the object with reference to speed. The goal for this model is to change the proposals of regions for the data training is called RPN (Region Proposal Network). Single design model has been converted into fasr R-CNN model by RPM. It will reduce the number of proposed regions and improves the speed of the data training and performance in the various SOTA methods.

The structure of the Faster R-CNN model has been shown in the Figure 2.12. RPN is used in fast R-CNN as the attention algorithm which helps to provide information about area of the interest. The RPN gets the results from the pre-trained CNN and then passes through the small model over the map of features with the different results of the proposed regions with the class prediction. Then the proposal of regions from the output of the RPN is in the form of bounding boxes which are the pre-defined shapes and sizes of the boxes for the determination of the object and improved the regions proposals. For the determination of the object the prediction of the class is in binary form which tells yes or no for the presence of the object which is called objectness of the proposed region.

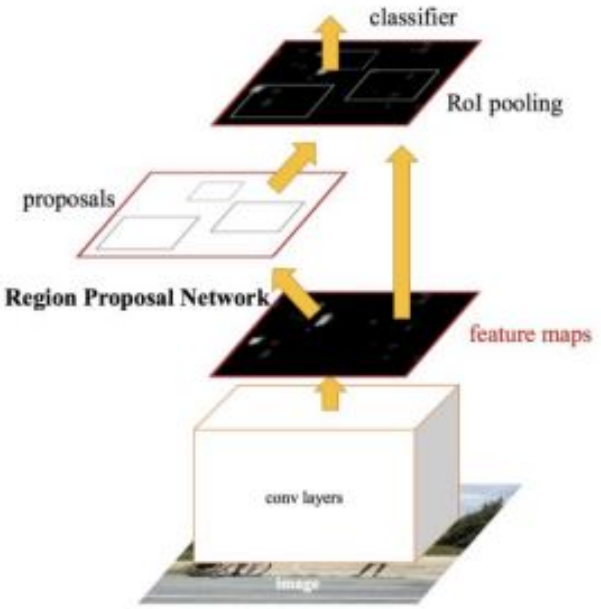

Figure 2.12: *Structure of Faster R-CNN*

2.6.4 YOLO

YOLO (You Only Looks Once) is a object identification system which is mainly depdends on different techniques of deep learning[18]. YOLO belongs to the RBAs (Regression based algorithms). This method divides the class probabilities and the bounding boxes. That's why this model is perfume well on the object detection. YOLO used the unified technique of detection which segregates different parts of the identified object into a single feed forward neural network.

The methodology of the YOLO algorithm is shown in the Figure 2.13. The YOLO algorithm takes the image as an input and then divides this input t several grids and after this calculated the probability for every grid for the object presence in the corresponding grid. After this the YOLO combines all the grids and make groups of the surrounding grids into a single object. As similar to the previous, the training of the model has been performed where the ground truth image is used to compare the identified object center and the model's weights has been adjusted according to this. In this paper [10] the results achieved are mAp 63.4 and 45 FPS which is significantly higher result in contrast to the other SOTA methods. The YOLO was written originally in the custom framework which is called darknet and it is written in machine languages which is understandable by the computer directly. With the passage of time YOLO was developed since the starting

version release in 2016 and after this YOLOv2, YOLOv3 and other are developed.

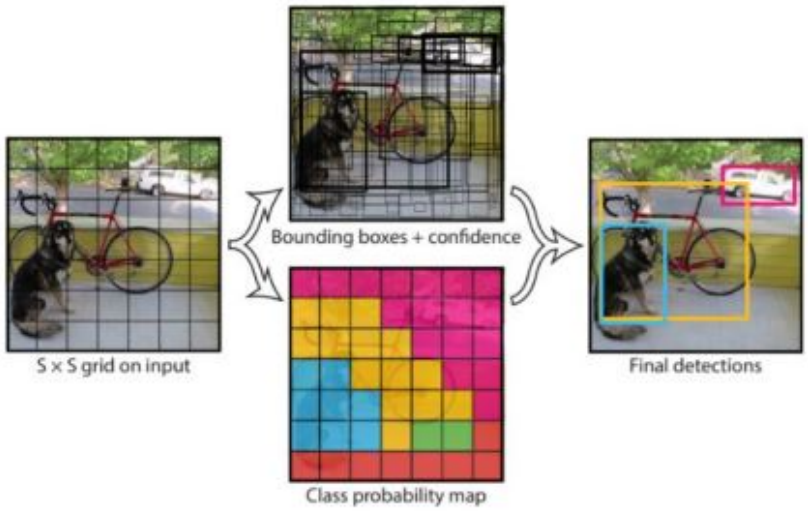

Figure 2.13: *Methodology of YOLO Technique*

In the different research papers, the authors make comparisons the results of the different methods with YOLO. Faster R-CNN and SSD which are faster and have improved accuracy rate when compared to the YOLO it shows that the YOLO is best one for the object detection. And also, the comparison tells us that the other methods take more time than YOLO. After this, the second version of YOLO which is the YOLOv2 is developed and can be identified more than 9000 different categories of the objects. Then YOLOv3 was introduced which is the three times faster than SSD and have better features then previously methods. YOLOv4 is the improved version of YOLOv3 which has 10% and 12% better AP and average speed in FPS. Likewise, YOLOv5 is the improved variant of YOLOv3 and YOLOv4 but its speed is similar to both of them. The table 2.1 depicts the comparison of various methods of object identification which are basically depend on different techniques of deep learning in terms of mAP and FPS.

2.7 Related Work

Qadir H.A at all (2013) et al. instead of using binary mask used 2D Gaussian mask to detect different types of tiny, flat polyps more efficiently by reducing false positives [65]. In this paper they used CNN based different encoder decoder variants that are basically used for segmentation of objects by using 2D Gaussian mask as real images. The 2D

Gaussian mask create a well-defined difference between polyps and their similar false positives. The proposed method was applied on two different datasets i.e. ETIS-LARIB and on CVC-ColonDB with precision 86.12% ,88.35%, recall 86.54%, and FI-Score 91%, 86.33% and 89.65% respectively.

Podlasek Jeremi at all (2020) et al. created a system on CNN architecture to identify polyps in real time using single GPU (Graphic Processing Unit) [64]. These are tested on both public datasets and clinical examination recordings. The study included 165 recordings of colonoscopy procedures and 2678 gathered images. Total of 81962 polyp frames were used for training the system. After training, the system is tested using videos from 42 colonoscopies and four public datasets i.e. CVC-ClinicDB, CVC-ColonDB, Hyper-Kvasir, and ETIS-Larib. Videos were used for identification of polyps and false positive rate (with 3% false positive rate identified 94% polyps and additional polyps also that missed during initial video examination) and datasets are used for F1 score (performance were varying from 0.727 to 0.942 F1 score.)

Sornapudi Sudhir at all (2019) et al. proposed a system that was a modified region-based CNN by creating mask around the polyps that are identified from the still frames [42]. From the images of the polyps the features were extracted by using Resnet-51 and Resnet-101 model trough the techniques of fine tuning and feature extraction. From Resnet-51 and Resnet-101 models it's clear that if the model is deeper than it extracts more fruitful features from the images of polyps. One more good thing for this system is that for every polyp it produces the exact segmentation. The proposed system applied on three public datasets and these produced the F1 and F2 scores. These datasets include g CVC-ColonDB (90.73, 91.27), CVC-PolypHD (80.65, 79.11), and ETIS-Larib (76.43, 78.70). This model gives the best result i.e. (96.67, 96.10) on wireless capsule endoscopy dataset.

SHIN YOUNGHAK at all (2018) et al. created a system for the automatic identification of polyps in the images and videos that are taken in the colonoscopy examination by using faster region-based CNN [34]. In order to overcome the hurdles and limited number of images of the polyp, they use deep CNN. And also, they proposed two effective methods i.e. automatic false positive learning and offline learning which can be integrated with detection system for the automatic identification of polyps in the images and videos. This proposed system applied on two datasets i.e. CVC-CLINIC

and ETIS-LARIB and also applied on video datasets i.e. ASU-Mayo Clinic Colonoscopy Video and CVC-ClinicVideoDB.

Jia Xiao at all (2020) et al. proposed a system that is called PLP Net which is the two-stage approach for identification of the polyps in images of colonoscopy by using very deep Convolutional Neural Network (CNN) [49]. The word Polyp (PLP) stands for "Polyp". This system consists of many steps. First stage is the polyp proposal stage which is the faster R-CNN, it recognizes the area which is affected and work as the region-based polyp identifier. Second stage of PLPNet is to build the CNN to classify each pixel in the image which is called pixelwise segmentation. Thus, this system improves the recognition accuracy. The proposed system is applied on two public datasets i.e. CVC-ColonDB and CVC-ClinicDB.

Lee JiYoung at all (2020) et al. created a system that used YOLOv2 for the identification of the polyps. This system is applied on 8075 images and 503 polyps [50]. Four independent datasets are applied for this system. First dataset consists of 1338 images and 1349 polyps which is called dataset A. Second dataset consist of 612 images which is the public CVC-ClinicDB and give name as dataset B. The third dataset consist of 7 videos with 26 different polyps and give name as dataset C. Fourth dataset consist of 15 unchanged videos give name as dataset D. In order to reduce the false positive rate in video examination they applied the median filter. For the datasets A and B the polyps identification accuracy is 96.7% and 90.2% respectively. For the dataset C which is the video dataset polyp identification accuracy is 87.5% and false positive rate is 12.5% if the median filter is not applied and 6.3% if the median filter is applied. For the dataset D the polyps identification accuracy is 89.3% and false positive rate is 8.3%. This system identified all 38 polyps and also 7 additional polyps.

Wang Pu at all (2018) et al. developed a system for the detection of polyps using deep learning technique [35]. For this they used 1290 patient data and this data is validated on the 27113 images collected from 1138 patients with one identified polyp. For this dataset 94.38% is the per image sensitivity, 95.92% is the per image specificity. The second dataset is CVC-clinicDB which consists of 612 images and 88.24% is its per image sensitivity. And the 54 videos that are without the polyp and 95.40% is its per image sensitivity. With the latency of 76.80 $\pm$ 5.60 ms minimum 25 frames per second can be processed by this system in the video examination which examine in real time.

This system helps to get the difference between difference polyps and polyp detection.

Xu Jianwei at all (2021) et al. proposed a method for the detection of images and frames that combines the 2D CNN network which is the real time object identifier with spatiotemporal information of the polyps [66]. They introduced a module which is the two featured, the FP Relearning module (Feature Prediction Relearning Module (FPRM)) is used to increase the precision of the detector network by learning it about the different features of the FPs and the Image Style Transfer Module (Image Style Transfer Module (ISTM)) to inflate the different features of the polyps in order to improve the sensitivity. In the video examination they combine the spatiotemporal information that uses the Structural Similarity (Structural Similarity (SSIM)) for measuring the similarities between the frames of videos. For making the final decision, they proposed Inter-frame Similarity Correlation Unit (Inter-Frame Similarity Correlation Unit (ISCU)) which combine the results gathered from the detector network and the similarity of frames. The proposed system is applied on different four publicly available datasets i.e. CVCClinicDB (CVC-612), CVC-ColonDB, ETIS-LiribPolypDB and CVC-ClinicVideoDB. This system provides the better results in terms of sensitivity, precision and specificity.

Brandao Patrick at all (2018) et al. present a deep leaning algorithm for the identification of the affected area in colonoscopy [26]. They convert the VGG and Reset to fully convoluted neural network FCNs and enhance and study their capabilities for the segmentation and identification of the polyps. For the better representation of structure of the tissues in the images of the colonoscopy they used shape from shading (Shape from Shading (SfS)). They get better results as they included the depth into their model as additional input chancel to the information of the RGB. They model is applied on two different publicly available datasets i.e. e ETIS-Larib and CVC-Colon DB and get the mean segmentation of 47.78% and 56.95% respectively.

Zheng Yali at all (2018) et al. present a method for identification of the polyps in images of endoscopy with bounding box based on You Only Look Once (YOLO) CNN [38]. The model was fine-tuned using colonoscope images that gathered from the three distinct datasets and before this the model was pre-trained with non-medical images. Also including the dataset of images that are collected from the different 106 patenting by using Narrow Band NB imaging endoscopy. When YOLO tested on 196 images of

White Light White Light (WL) of dataset it gets precsion and sensitivity of 79.3% and 68.3% respectively. This model was tested on three public datasets (CVC-ColonDB, CVC-ClinicDB and ETIS-Larib) and one private dataset (PWH-ColonDB).

Li Kaidong at all (2021) et al. create a database of endoscopic images by utilizing different sources and then explain the location of the polyps and also describe the type of polyps by the contribution of different gastroenterologists [61]. This database can be used for training and evaluating the machine learning algorithm for the polyp classification. This compared with different eight previous deep learning methods and get the better results.

Li Kaidong at all (2015) et al. present a system that based on the CNN and three-way image representation. in order to localize the polyps in accurate manner this method learns the different features of the polyps in terms of color, shape, temporal information etc [16]. For the candidate of polyp, many CNN that each have one feature of polyp were applied on the candidate that give result to either accept or reject the candidate. They also proposed a performance curve in this system that tells the latency of the polyp identification. This defined as the time from polyp appears first in the video to the time of its first identification. This method gives the better performance as we compared to the state-of-the-art method and also reduce the false positive rate.

Zobel Pascal at all (2019) et al. proposed a method in which mask of R-CNN applied for the identification of the affected area in the colonoscopy screening. In order to train and test the data different datasets are used [45]. These datasets included three independent datasets that consist of 2484HD labeled images and two public datasets that consist of 612SD and 194HD labeled images. They get the best results for the three datasets when train deep CNN in the range of recall = 0.92, precision = 0.86, F1 = 0.89 for the dataset A, rec = 0.86, prec = 0.80, F1 = 0.82 for the dataset B and rec = 0.83, prec = 0.74, F1 = 0.79 for the dataset C.

Liua Xinyu at all (2021) et al. created the system in which they sum up the detection and localization of polyps in order to remove the gap between different datasets [62]. This system consists of two parts. First part is pixel level adaption in which they proposed a method called Gaussian Fourier Domain Adaptation (GFDA) method to reduce the domain gaps between the datasets without disturbing their contents. The second one is hierarchical feature-level adaptation that consists of two modules one is Hierarchical

Attentive Adaptation (HAA) and second is Iconic Concentrative Adaptation (ICA). These modules maintain the consistency in prediction and these are regularized by Generalized Consistency Regularizer (GCR). The experimental results shows that the system perform better than the previous systems and achieve the recall rate is 87.5%.

Mohammed Ahmed kedir at all (2018) et al. proposed a system to tackle the problem of limited data available and the variability of polyps in terms of shape, size and appearance [30]. In this system they used Y-Net which is the novel deep learning that consists of two encoders and one decoder block. This system depends on the pre-trained and un-trained data with novel deep learning. Every encoder has to be trained with the encoder specific learning rate along with the decoder. Comparison of previous method this system gives the better results in detection and segmentation of the polyps. This method gives good result as compared to sat-of-the-art methods in terms of detection rate with F1 score 7.3% and recall improvement is 13%.

Tajbakhsh Nima at all (2014) et al. present a very new method for the detection of the polyps [10]. This system originality depends on the integration of the polyp geometric constraints with the intensity variation patterns across the polyp boundary. In previous, the detector drives for the objects with the curvy boundary, and after this minimize the false rate of polyp-like structures. This system consists of three contributions i) a quick and accurate descriptor of patch for accurately explaining the variation patterns across the boundary. ii) a very new two-stage scheme of the classification of polyps efficiently removing the non-polyp edges from the complete edge map. iii) a creative voting system for different polyps from the retained edges that are robustly identified.

Sánchez-Peralta Luisa F. at all (2020) et al. present a dataset named PICCOLO [52]. This dataset consists of 3433 images that are manually collected and in these 3433 images 2131 are the white light images and 1303 images are the narrow band images. This dataset collected from the 40 different patients and 76 affected areas in which polyps present. Also, this dataset divided into 3 sets that consists of 2203 (training), 897 (validation) and 333 (test) different images, ensuring that the independence of the patients. Additionally, the data about the data of the affected area also provided. From this they obtained the four different models by combining the architectures of two encoder-decoder and two backbones. These models have trained by the PICCOLO dataset and two other publicly available datasets to make the comparisons between

them. They get the results of all the test set for every dataset. The models that are trained with the PICCOLO dataset gives the better results as compared to other two datasets. And give the best result for its own test set. This dataset is available at the Basque Biobank website, so that in future to get better results in the identification, localization and segmentation by using the different deep learning techniques.

Tian Yu at all (2019) et al. present research in which they adapt a previously one stage classification and detection method for the new five class polyp detection and classification problem [44]. In this research, they show that the one stage method not only feasible for the identification and classification of the polyps with respect to the two-stage method, but also this one stage method is performed good and faster for the testing and training. In this research they used MICCAI 2015 dataset for the detection and classification of the polyps. They also show that this one stage method gives the best results as compared to the state-of-the-art methods when apply on the MICCAI 2015 dataset.

Chapter 3

Dataset

The Kvasir-SEG dataset that is used in this research is basiclly originated from prviously dataset i.e. Kvasir[21]. For the identification and detection of the GI diseases, Kvasir dataset is firstly MC (Multi Class) dataset.

3.1 The Kvasir dataset

It consists of eight classes of polyps and 8000 images of gastrointestinal tract polyps. in each of the eight classes of polyps consists of 1000 different images. To improve the dataset quality, they replace the 13 new HD frames from the older polyp class. Vestre Viken Health Trust institute Norway were collect these frames and verified them from the different gastroenterologists which they are much experienced. The different classes of polyps in this dataset consists of procedures of endoscopic, different findings of pathological and also landmarks of anatomical. A much and more details and explanation about the different classes of the images, procedure for collection of the data and deep understanding of the dataset is present[21].

To compare and develop the methods to get the best result and performance on the classification of the multiclass in the large scale of endoscopic findings, the Kvasir dataset has been used for the Medico Task which is the medicine challenge at the MEBI (MediaEval Benchmarking Initiative) for the evaluation of multimedia in 2017 [22] and 2018 [32]. Kvasir dataset is restricted only for frames classifications because it contains frame-wise annotations. Thus, Pozdeev et al. [40] suggested a method in which CVC-ClinicDB dataset was used to train a model and then try to estimate the segmentation masks

for this dataset, but unfortunately they failed to present their report due to lack of the ground truths and experimental scores.

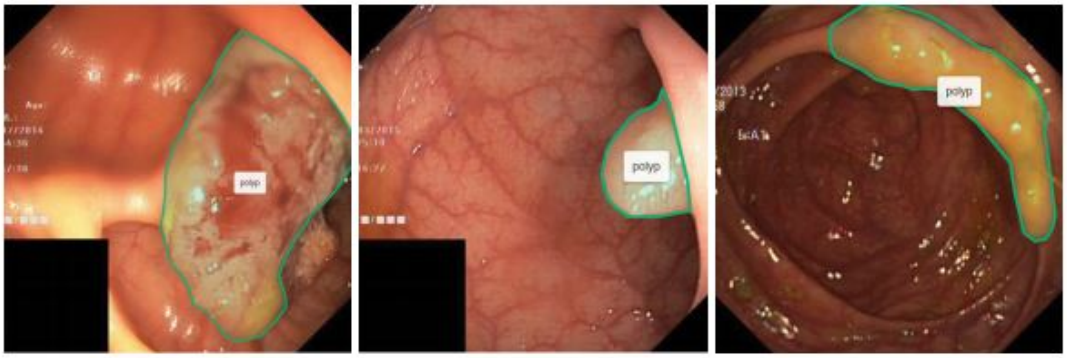

Figure 3.1: *Example of frames shows that the Kvasir-SEG dataset have been marked by the green outlined to the polyps*

3.2 The Kvasir-SEG Dataset

In this research, Kvasir-SEG dataset [48] has been used for the localization, segmentation and detection of GI polyps. Figure 3.2 reflects the original image, information about ground truth and their identification (the bounding boxes in purple). This dataset is used to develop recent as well as advance methods for the detection and localization of polyps in gestrointestine. The Kvasir-SEG dataset contains 1000 HD polyp images and also the information of their bounding boxes and corresponding masks. The original frames and their corresponding ground truths utilized for the segmentation purpose and the information about bounding box used for the identification purpose. The proposed dataset can download from the https://datasets.simula.no/kvasir-seg/. This dataset consists of 1071 different polyps in which 700 large polyps 323 medium polyps and 48 tiny polyps. So as a whole, the dataset contains of 1071 frames of different polyps and their corresponding segmentation mask and their bounding boxes.

3.3 Mask Extraction

The Kvasir-SEG has been uploaded to the Label Box [41] and after uploading the entire dataset to label box then by using this label box application all the segmentations have been created. The tool that is used for labelling the region of interest ROI in the frame

of images is called Label Box. The ROI in our case is polyp region. In all 1000 images in the dataset the margins of polyps outlined manually by a team that consists of a doctor and an engineer. After this the margins of polyps have been viewed by the experienced gastroenterologist.

In the Figure 3.1 the example of frames shows that the Kvasir-SEG dataset have been marked by the green outlined to the polyps additionally. After this, we have to export the file for the generation of the annotated masks. The file that is obtained in JSON format consists of all relevant and necessary data about the images and the coordination points for the creation of mask hence, for this purpose. For this purpose, ROIs coordination was used to create contours on the empty frames having black background and then these contours would be filled with white color. The obtained mask is white in color with balck background and have the depth of 1-bit. The Figure 3.2 reflects the examples of the images, segmentation mask and their bounding boxes from the dataset Kvasir-SEG.

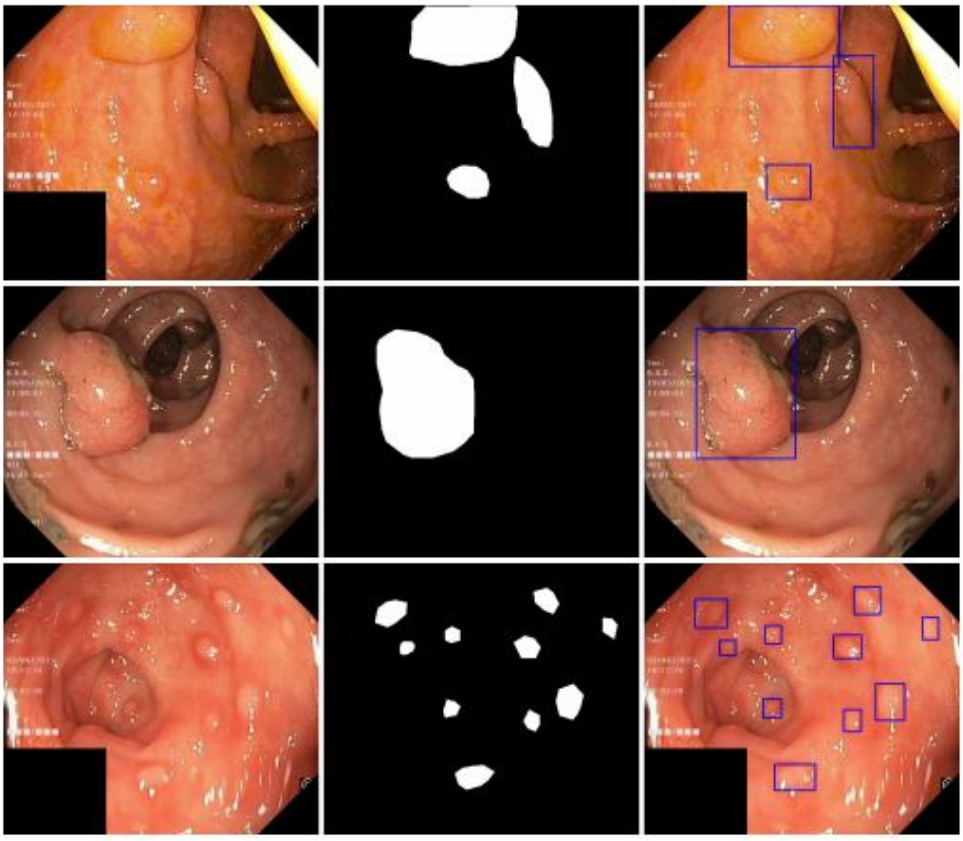

Figure 3.2: *Examples of the images, segmentation mask and their corresponding bounding boxes from the dataset Kvasir-SEG*

Chapter 4

Methodology

4.1 Method

For each pixel in the image, the segmentation method is used to classify the object class, while for localization the detection method is used to predict the object class and regress the bounding box. In Figure 1, in second column the ground truth mask has shown for the segmentation task while in third column corresponding bounding boxes shows for the detection task. For the segmentation and automatic detection of GI polyps in the Kvasir-SEG dataset, this section explains the basic methods for localization, segmentation and detection.

4.1.1 Localization and Identification Baseline Methods

The Detection methods include head, neck, input and backbone. The head is used for the management of the prediction boxes and for dense prediction it can be a one stage detector e.g. RetinaNet, YOLO, RPN and for sparse prediction it can be a two-stage detector e.g. Faster R-CNN. The neck is the portion of the backbone network which consists of Bi-FPN, PANet and FPN. The input can be patches, images etc. The backbone can be different CNN architectures that includes Darknet, Resnet50 etc. In later, due to the speed and the ability to get desirable accuracy the one stage method has much attention. This is possible because the previous networks utilize the feature pyramid networks for the prediction of bounding boxes shown in Figure 4.1.

In this research, EfficeintDet [53] has been used which uses BiFPN (bi-directional feature pyramid network) as the feature network, EfficientNet [43] as the backbone architecture

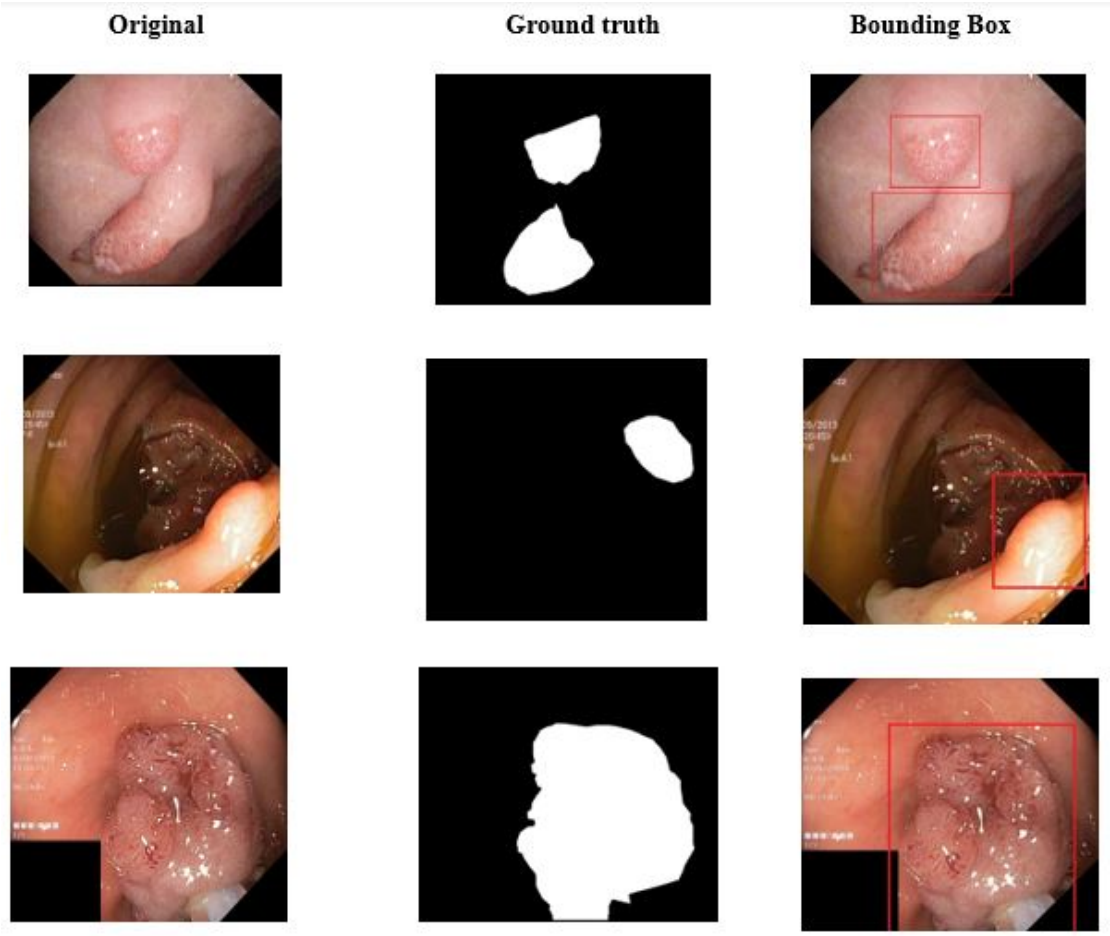

Figure 4.1: *Sample HD frames from Kvasir-SEG Dataset: second column (ground truth) third column (bounding box)*

and the network of shared box. In addition, we also used R-CNN [14] which is the faster Region based convolutional neural network and uses the RPN (Region proposal network) as proposal network and Faster R-CNN as detector network. Additionally, we employ YOLOv3 [33], which makes use of multi-class logistic loss modeled using regularizes like objectness prediction scores. In addition, we employed YOLOv4 [33], which makes use of a cross-stage partial connections and an extra bounding box regressor based on the Intersection over Union (Intersection over Union (IoU)) in their backbone design. Moreover, YOLOv4 allows on fly data augmentation, such as mosaic and cut matrix.

For the improvement of accuracy RetinaNet [19] consider the data driven property which allows the network to focusing on the hard samples. The architectures such as ResNet50, RetinaNet, YOLOv3, YOLOv4 and darknet53 provide the opportunity at the early stage of the network for the feature extraction. Optimal anchor boxes are found and searched

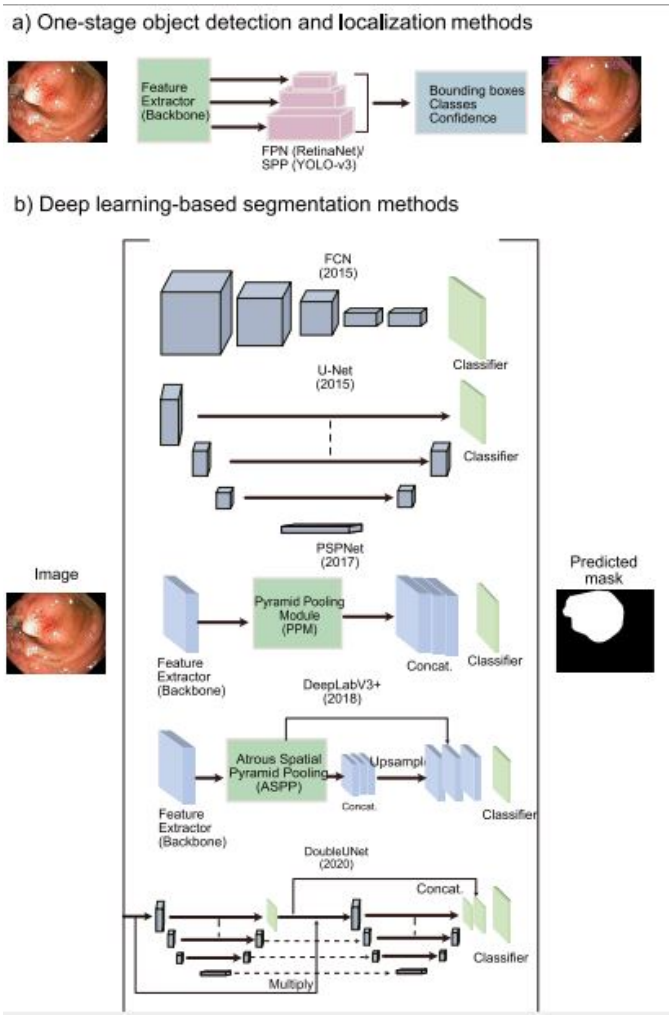

Figure 4.2: *Summary of Baseline Methods for Detection, Localization, and Semantic Segmentation*

to handle the different aspect ratio problems. In table 2 hyperparameter for the detection is shown.

4.1.2 Segmentation Baseline Methods

In the recent years, different approaches of data driven by using CNNs have changed the pattern of segmentation and computer visions methods. In order to get the feature maps, an input image is put directly to the convolutional layers, providing the object segmentation this image can be unsampled later on for the prediction of pixel wise classification. These models get information from the ground truth labels that are available and these can be used for the prediction of the labels from the other similar dataset. Long et al. [13] first introduced the FCN (Fully Convolutional Network)

based on segmentation which can be trained end-to-end. Roneberger et al. [15] has been modified the architecture of RCN and also extended the architecture to a UNet architecture. The UNet consists of two portions first is the analysis which is called encoder and second is the synthesis path which is called decoder. In the analysis path deep features have been extracted and, in the synthesis, path the segmentation is to be performed on the basis of these extracted features.

PSPNet which is called Pyramid Scene Network [24] proposed a pyramid pooling component that has aimed to sum up all the information that are gathered globally from the different regions and then unsampled and concatenated to make the final representation of feature. As shown in figure 3.2 after the convolutional layer a final per pixel prediction is obtained. In order to extract the features, we used the architecture of the ResNet50 that is pretrained on the imageNet. As similar to the architecture of the UNet, DeepLabV3+ [27] is a network of encoder and decoder. While it uses the atrous separable convolutional and spatial pyramid pooling that is shown the last in figure 3.2 for the improved accuracy and the fast inference. The atrous convolutional has to be used to control the resolution of features that are computed and also used to adjust the respective filed to get the multi-scale information efficiently. In this research, we have used for the encoder and decoder networks of DeepLabV3 the output stride 16 and have to be experimented on the ResNet50 and ResNet101 backbones.

ResUNet [37] is the improved version of the UNet with the influence of residual neural networks. ResUNet++ [60] and ResUNet are similar but the former has added layers like "squeeze-and-excite blocks, Atrous Spatial Pyramid Pooling (Atrous Spatial Pyramid Pooling (ASPP)), and attention blocks" which help in learning deep features required for improved pixel prediction for segmentation of objects. DoubleU-Net [47] includes two modified UNet architectures; the first encoder used is VGG-19 [4], pre-trained on ImageNet [9]. The extra modules that are incorporated in DoubleUNet are as follows; Squeeze-and-Excite module plus ASPP block. HRNet [54] preserves the representation with high-resolution convolution in parallel to other resolutions and transfers data between them constantly, making it one of the latest and most applied methods in recent research.

Moreover, we also utilized "UNet with ResNet34" as the network's backbone while providing training on the compared networks for semantic segmentation. Table 1 depicts

the hyperparameters set for each benchmark method In Table mA. The table shows that baseline methods have many trainable parameters, and therefore, this makes the network large and has low frame rates. Hence, developing the best configuration with lightweight, efficient architecture that comprises a higher frame rate and gives better performance is crucial. This research introduces ColonSegNet which needs fewer training parameters and therefore saves learning time as well as time for prediction. Further details of the architecture are explained in the following section, where the workings of each component are described.

4.1.3 ColonSegNet

In this research, we proposed a ColonSegNet method that is shown in the figure 3.3. ColonSegNet consists of encoder and decoder that uses the residual block [17] with its main component which is the squeeze and excitation network [29]. This method has to be designed with a very few and small parameters as compared the other state-of-the-art methods. This method is simpler and lighter weighted as compared to other methods like DeepLabV3 [24], PSPNet [27] and the other methods. The less trainable parameters lead to the methods becomes more lightweight and efficient for the detection and localization of polyps.

The proposed network consists of two encoder and two decoder blocks. All the necessary information has been extracted from the image through encoder. And then this passed to the decoder. From the encoder block there are two skip connections to each of the decoder block. The first skip connection is the simple concatenation and the second skip connection to pass the multi-scale features in the decoder through the transpose convolutional. To get the more semantic and fruitful information in form of segmentation mask the multi-scale features helps to the decoder.

ColonSegNet includes two residual blocks and a strided convolution of 3×3 in between them. The input image is given to the first encoder block. A max pooling of 2×2 is followed this layer. Then, the feature map of output reduced by ¼ to the given input image. Same as the first encoder block the second encoder block also consists of two residuals block and a 3×3 strided convolution between these two residuals block. This all show in the block diagram of the ColonSegNet.

The decoder block has to be start by the transpose convolution, where the first decoder

block gets the stride value of 4 which has to be increases the dimension of the feature map by the value of 4. As similar to the first decoder block the second decoder block get the stride value of 2 and increase the dimension of feature map by the value of 2. After all this the model follow the simple residual block and the concatenation. Furthermore, the second skip is concatenate this model and follows the residual block. After all this to generate the binary segmentation mask the output at the last decoder block passed through convolution of 1×1 and also passes through the sigmoid activation function.

4.1.4 Augmentation of Data

Supervised Learning Methods (SLM) which is called Supervised Learning Methods require a big and large data to obtain the better, reliable and efficient networks. SLM are also data acquisitive which means that these methods are data greedy. Attaining this training data through the dataset, it is the manual process that requires efficient resources and also requires man efforts from the computational scientists and also from the endoscopists.

In order to increase the number of training sample in the dataset we used data augmentation technique in this research. In this research we used the basic technique of the data augmentation., this basic technique includes the vertical and horizontal flipping, random rotation and scale etc. the images that we used in all the experiments in this research have to reseized by the standard and fixed by 512×512. In order to normalize the image, we subtract the image by its mean and divide by its standard deviation.

Chapter 5

Results

First, in this section, the performance measures and the experiment conducted will be described. After that, the quantitative and qualitative findings are described.

5.1 Evaluation metrics

We employed the conventional computer vision evaluation parameters in this study using the Kvasir-SEG dataset to assess, localization, the polyp detection segmentations of semantic techniques.

5.1.1 Localization and detection of task

For the object localization and detection of tasks, AP and IoU presented in the literature [11][8] were applied.

IoU: This therefore calculates the similarity of two skirting boxes A and B and is the ratio of the intersected area to the total union area.

$$\text{IoU}(A, B) = \frac{A \cap B}{A \cup B}$$

Regular Accuracy is calculated as the Part Below the Arc (AUC) of the accuracy-recollection arc, tested at entirely exclusive recollection principles (r1, r2, ...), specifically when there is a drop in the maximum precision value. The formula for AP is:

$$\text{AP} = \sum_{n} (r_{n+1} - r_n) \cdot p_{\text{interp}}(r_{n+1})$$

By $p_{\text{interp}}(r_{n+1}) = \max_{\tilde{r} \geq r_{n+1}} p(\tilde{r})$. Now, $p(r_n)$ represents the accuracy of a specific recall value. This definition guarantees that accuracy values do not increase as recall increases. AP is computed as an average over different Intersections over Union (IoU) thresholds, ranging from 0.25 to 0.75 with a step size of 0.05. This results in an average over 11 IoU stages (denoted as AP@[.25:.05:.75]).

5.1.2 Segmentation task

For the segmentation task of the polyp, in this research we used the CV (Computer Vision) metrices that are globally accepted. These includes Dice Coefficient (DSC) (Dice Coefficient), r (recall), p (precision) and IoU (Intersection over Union). It also includes Acc which is called the overall accuracy. IoU is also called Jaccard Coefficient (JC) (Jaccard Coefficient). In this research we have also used FPS which is called frames per second. FPS used for the evaluation of the efficiency, applicability for the segmentation purpose in terms of IT (inference time) during the endoscopy.

For the definition of each matric of computer vision, we used the different terms in this research. The terms include false positive, false negatives, true positive and true negatives and these terms are denoted by fp, fn, tp and tn respectively.

$$\text{DSC} = \frac{2 \cdot \text{tp}}{2 \cdot \text{tp} + \text{fp} + \text{fn}}$$

$$\text{IoU} = \frac{\text{tp}}{\text{tp} + \text{fp} + \text{fn}}$$

$$r = \frac{\text{tp}}{\text{tp} + \text{fn}}$$

$$p = \frac{\text{tp}}{\text{tp} + \text{fp}}$$

$$F_2 = \frac{5 \cdot p \cdot r}{4 \cdot p + r}$$

$$\text{Acc} = \frac{\text{tp} + \text{tn}}{\text{tp} + \text{tn} + \text{fp} + \text{fn}}$$

$$\text{FPS} = \frac{\text{No. of frames}}{\text{sec}}$$

5.2 Setup of experiment and configuration

The structures incorporated in the research involved ResUNet U.NET, DOUBLE U NET ResUNet++, End-to-End and HRNet and were expressed using Keras[58] along the tensor flow[72] and performed on volta100 Nvidia DGX-2 AI. FCN8, GPU PSPNet, DeepLabv3+, UNetResNet34, and ColonSegNet, the used framework for implementing the networks, are the PyTorch". Specifically, the training of these methods was done on a computer with "NVIDIA Quadro RTX 6000 while the inference testing of all the reported methods was done on a computer with NVIDIA GTX2080Ti".

For all the experiments, we employed a dataset of 880 images for training out of which 120 images were applied to validation. Since the sizes of these images in the database were different, we normalized all the images to a size of 512×512. Hyperparameters have a deciding utility in having deep learning algorithms search for the best solutions and yet, how to arrive at the best hyperparameters is not easy to determine because there is no well-defined way of doing it. Several ways of selecting the initial hyperparameters include the grid search method as well as the random search method others include the advanced methods such as the Bayesian search. On the other hand, the use of Bayesian optimization is quite expensive and often challenging to implement over several deep learning models.

We devoted extensive time to the optimization of the model's hyperparameters with regards to tasks we trained it on which included the exposure of localization and Subdivision of growth. These hyperparameters were adjusted using empirical validation and the specifics are stated in Tables 4.1 and 4.3 for the Kvasir-SEG dataset.

5.3 Quantitative Evaluation

5.3.1 Detection and localization

As presented in Table 4.2, the results of the comparison of various methods that are assigned to detect and localize the polyps using the data of the Kvasir-SEG dataset are shown. Thus, the comparison shows that RetinaNet provides a higher score in terms of mean Average Precision (mean Average Precision (mAP)) than the chosen competitors, YOLOv3 and YOLOv4, depending on Intersection over Union (IoU) threshold and Av-

erage Precision at 0. 25 (AP25) and 0. 50 (AP50). In particular, RetinaNet with a backbone based on ResNet101 achieved an AP indicator of 0. Thus, it obtained a value of 8.745, which is significantly higher than the scores of YOLOv4 which got only 0. 8513. But, at an IoU threshold of 0, 75, which exceeded the outcomes of the RetinaNet model with an AP75 of 0. 7594 as compared to RetinaNet which was 0. 7132.

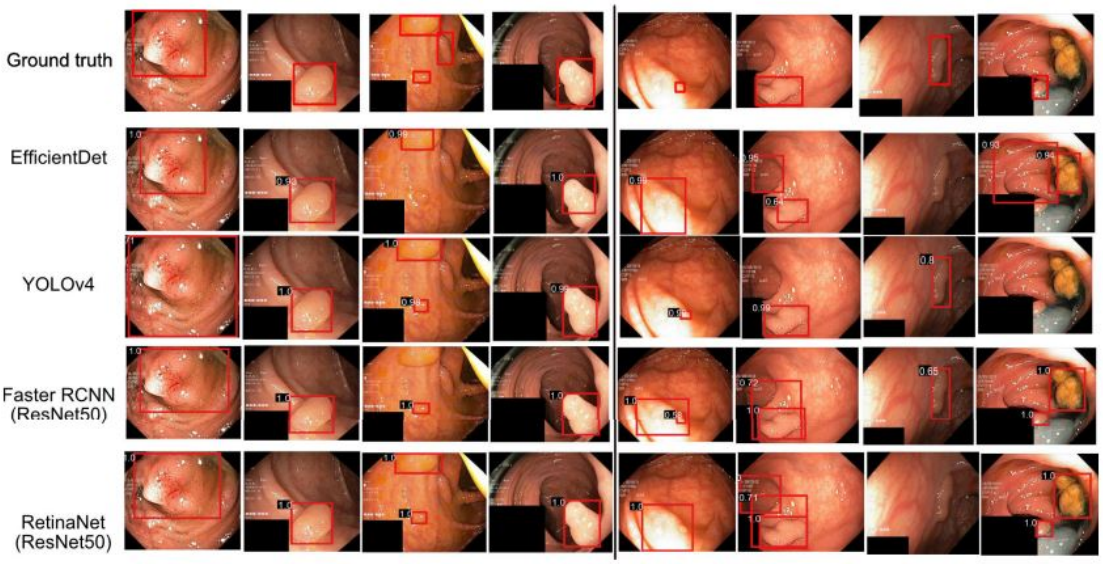

Figure 5.1: *Dataset test samples for detection and localization*

Method	Learning rate	Optimizer	Batch size	Anchors	Threshold
Faster R-CNN	$2.5e^{-4}$	Adam	8	256	0.4
RetinaNet	$1e^{-3}$	SGD	8	15 (pyramid)	0.3
YOLO v3-spp	$1e^{-3}$	SGD	16	8	0.25
YOLO v4	$1e^{-3}$	SGD	16	8	0.25
EfficientDet-D0	$1e^{-4}$	Adam	8	default	0.4

Table 5.1: *Kvasir-SEG dataset hyperparameters in the baselines of the polyp detection and localization task.*

In the Figure 4.1 on the the right side of black line images have different results with different techniques like EfficientDet-D0, YOLOv4, Faster R-CNN and RetinaNet and also in most cases get highest IoU. On the left side of the black line images are with false cases means that wrong localization. At the top left of the red boxes the confidence prediction scores has provided.

In terms of the AP, it must be noted that the proposed method reached an AP of 0.

8248, which is about 8% higher than the RetinaNet proposal. IoU provides the measures of bounding box localization precision. However, the detected AP of the architectures is as follows; the least AP is recorded to be 0 from the EfficientDetD0. 4756 and the IoU was equal to 0. 4322. Computing the AP of Faster R-CNN, the model resulted in an AP of 0. While it was able to decode 7866 it only resulted in a frame rate of only 8 FPS. On the other hand, YOLOv4 in this study using a Darknet 53 backbone provided results of 48 FPS, six more times than the Faster R-CNN. The real-time performance of YOLOv3 was obtained to be around 45 Msps on average. Unreal 01 FPS but the average precision was 5% less than YOLO v4.

Method	**AP**	**IoU**	**AP$_{25}$**	**AP$_{50}$**	**AP$_{75}$**	**FPS**
EfficientDet-D0	0.4756	0.4322	0.6846	0.5047	0.2280	35.00
Faster R-CNN	0.7866	0.5621	0.8947	0.8418	0.5660	8.00
RetinaNet	0.8697	0.7313	0.9395	0.9095	0.6967	16.20
RetinaNet	0.8745	0.7579	0.9483	0.9095	0.7132	16.80
YOLOv3+spp	0.8105	0.8248	0.8856	0.8532	0.7586	45.01
YOLOv4 , CSP	0.8513	0.8025	0.9123	0.8234	0.7594	48.00
ColonSegNet (Proposed)	0.8000	0.8100	0.9000	0.8166	0.6706	180.00

Table 5.2: *Performance in the Detection of polyps and Localization on a dataset of the Kvasir-SEG.*

As shown in the above quantitative outcomes, it is revealed the defined Darknet53 with YOLOv4 model may identify unusual polyps' types in actual-period detection at FPS 48 and a standard accuracy of 0. 8513. Based on these assessment systems of measurement, Darknet53 with YOLOv 4 become the finest network on behalf of detecting the segmenting polyps. The above model may help gastroenterologists reduce the growth neglect level and increase detection rates of endoscopically invisible neoplasia.

As stated, the ColonSegNet model outlined in this research focuses more on real-time polyp segmentation Nevertheless, we compared the proposed model's bounding box predictions with other existing traditional detection methods. The inference speed of ColonSegNet is approximately four times faster than YOLOv4 with a frame rate of 180 FPS while there is a small difference in AP scores and an IoU of 0. 81, associated with an AP of 0. 80. For this reason, it is also known as one of the most efficient techniques

Method	No. of parameters	Learning rate	Optimizer	Batch size
UNet	7,858,433	$1e^{-2}$	SGD	8
ResUNet	8,420,077	$1e^{-4}$	Adam	8
ResUNet++	16,242,785	$1e^{-4}$	Adam	8
HRNet	9,524,036	$1e^{-4}$	Adam	8
DoubleUNet	29,303,426	$1e^{-4}$	Adam	8
PSPNet	48,631,850	$1e^{-2}$	SGD	8
DeepLabv3+	39,756,962	$1e^{-8}$	SGD	8
DeepLabv3+	58,749,090	$1e^{-8}$	SGD	8
FCN8	134,270,278	$1e^{-2}$	SGD	8
UNet-ResNet34	33,509,098	$1e^{-5}$	Adam	8
ColonSegNet	5,014,049	$1e^{-4}$	Adam	8

Table 5.3: *Kvasir-SEG dataset baseline method for Segmentation of polyps Dataset: Selected Hyperparameters.*

of the diagnosis and identification of polyps' locations.

5.3.2 Segmentation Results

Table no 4.4 presents the consequences of the segmentation of the polyp's task. The proposed ColonSegNet outperformed additional condition-of-the-ability sector techniques significantly in expressions of the Similarity of Dice and Coefficient (DSC) and Connection above Association. Specifically, ColonSegNet processed colonoscopy frames approximately 4.5 years quicker than the UNet along a backbone, ResNet34 and had six times fewer parameters. While it was 0.75 times quicker than the standard UNet, its performance metrics were superior towards classic UNet and its modified versions, like ResUNet++ ResUNet.

Original UNet implementation had the lowest DSC score of 0.5969. In contrast, the UNet with a ResNet34 backbone achieved the uppermost DSC with 0.8757. score Second and third- uppermost were 0.8643 and 0.8554 OF DSC Score, respectively.

Additionally, DeepLabv3+ with ResNet101 attained the highest mean IoU (mIoU) of 0.8572, while DeepLabv3+ with ResNet50 achieved a mIoU of 0.8518. DeepLabv3+

Method	DSC	F2-score	Precision	Recall	Overall Acc.	FPS
UNet	0.5969	0.5980	0.6722	0.6171	0.8936	11.0161
ResUNet	0.6902	0.6986	0.7454	0.7248	0.9169	14.8204
ResUNet++	0.7143	0.7198	0.7836	0.7419	0.9172	7.0193
FCN8	0.8310	0.8248	0.8817	0.8346	0.9524	24.9100
HRNet	0.8446	0.8467	0.8778	0.8588	0.9524	11.6970
DoubleUNet	0.8129	0.8207	0.8611	0.8402	0.9489	7.4687
PSPNet	0.8406	0.8314	0.8901	0.8357	0.9525	16.8000
DeepLabv3+	0.8572	0.8545	0.8907	0.8616	0.9614	27.9000
DeepLabv3+	0.8643	0.8570	0.9064	0.8592	0.9608	16.7500
UNet	0.8757	0.8622	0.9435	0.8597	0.9681	35.0000
ColonSegNet	0.8206	0.8206	0.8435	0.8496	0.9493	182.3812

Table 5.4: *Kvasir-SEG Dataset baseline assessment for Polyp Segmentation*

with ResNet101 outperformed the version with ResNet50, likely due to the higher top-5 accuracy of ResNet101 in the ImageNet model. "Despite DeepLabv3+ with ResNet101 having over 11.5 times more trainable parameters and ResNet34 with Deep Labv3 standing almost 8 periods more computationally intensive, ColonSegNet delivered comparable DSC scores. Notably, ColonSegNet was 10.7 points quicker more than ResNet101 with DeepLabv3+ and 6.9 ages quicker more than ResNet34".With DeepLabv3+.

"The DSC scores for FCN8, HRNet, and DoubleUNet were 0.8310, 0.8446, and 0.8129, respectively, while ResUNet++ had a DSC of 0.7143. The F2-score trends were consistent across approaches. On behalf of accuracy, a ResNet34 with UNet support attained the highest count of 0.9435, while ResNet50 with the DeepLabv3+ backbone had the maximum recall score of 0.8616. The original UNet had the lowest precision (0.6722) and recall (0.6171). Overall, most methods performed well, with a ResNet34 with UNet support showing the maximum overall accuracy and a mean IoU (mIoU) of 0.8100. It also had the highest frames per second (FPS) rate of 35, surpassing ResNet50 (27.9 FPS) with the DeepLabv3+ and ResNet101 (16.75 FPS)". with the DeepLabv3+.

In terms of constraint count, the ResNet34 with UNet support utilized fewer parameters compared to FCN8 or DeepLabv3+. Due to its minimal trainable constraints and rapid implication period, ColonSegNet is well-organized and ideal for actual-period polyp

subdivision "(NVIDIA GTX2080Ti 182.38 FPS)", making it suitable for deployment on low-end hardware in clinical settings. However, based on DSC scores, the ResNet34 backbone with UNet remnants is the prime selection, despite its lower FPS of 35 on the NVIDIA GTX2080Ti.

5.4 Qualitative Evaluation

Results of polyp detection and the location as well as the confidence score are given in Figure 4.1. It is clear from the left side of the RetinaNet's vertical line and YOLOv4 in the figure that both algorithms have high confidence in identifying and outlining the position of the polyps regarding most of the images. However, the smearing techniques demonstrated in the third column do not fully characterize the areas of the polyps.

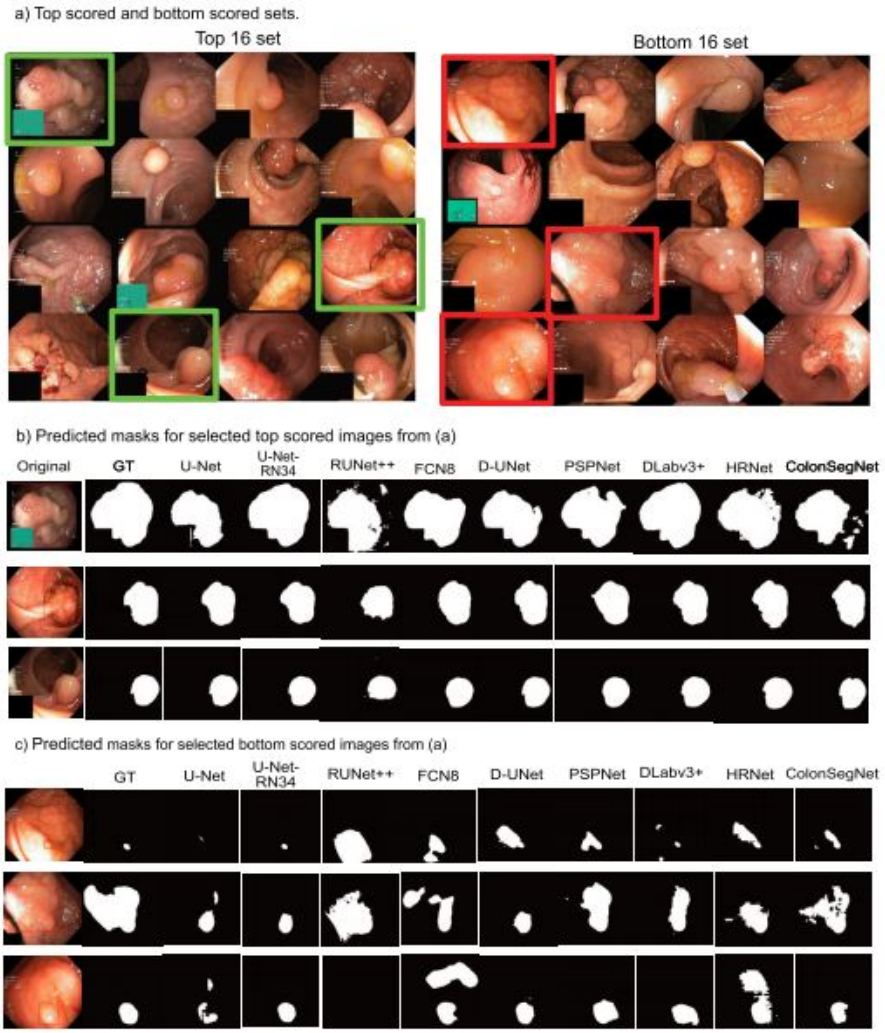

Figure 5.2: *Polyp Segmentation performance evaluation*

To the right of the vertical line, regarding the IoU, the images in the 5th and 6th columns show high scores in RetinaNet while YOLOv4 has good localization as shown by the bounding boxes. What is also observed is that both RetinaNet and the more complex efficient det D0 model are not able to detect a polyp in the seventh column. Both the YOLOv4 and EfficientDet D0 failed to detect the small Faster R-CNN polyp and the RetinaNet misidentified polyp and stool as one polyp in the eighth column.

Exploiting the Dice coefficient of similarity of the semantic methods of segmentation, the final heatmap was created and analyses informing about sets with the highest and lowest scores are presented in Figure. 4.2. All algorithms perform well in identifying large polyps and creating precise masks and this can be seen in Figure 4.2(b). Analyzing the results of the segmentation, it can be concluded that the highest accuracy is in DeepLabv3+ and UNet-ResNet34. Still, it can be noticed in Figure 4.2(c) that segmentation precision declines when dealing with small and flat polyps as well as inclined views and images with saturated areas. Figure 4.2(b) shows ColonSegNet, the proposed model that results in smooth segmentation shapes to meet the ground truth on the validation data. As depicted in Figure. 4.2(c), for the inferior images, ColonSegNet gives relatively better predictions in almost all cases.

Chapter 6

Discussion and Conclusion

6.1 Discussion

Investigating efficiency and accuracy of computational support techniques used to evaluate images from endoscopy is a field of interest now a days. For identification and localization of GI polyps Kvasir-SEG has been used for the first time and then State-of-the-Art (SOTA) methods has been utilized to compare segmentation methods. We have done repeatable validation of DL methods via metrics of computer vision in semantic segmentation, identification and localization of polyps. We selected those methods that are very popular in medical fields for example for image detection and segmentation Faster R-CNN and UNet , for speed YOLOv3, UNet along with ResNet34, for accuracy FCN8, DoubleUNet and collectively YOLOv4 and DeepLabv3+.

Given experimental results in table 4.2 YOLOv3 paired with Darknet53 as backbone is better than other methods with respect to mIoU which indicates that far better localization than RetinaNet. Similarly, trade-off b/w IoU and average precision of results is 3 times faster for YOLOv4 than RetinaNet. This is due to the regression in bounding box lost their CIoU and CSP (Cross-Stage-Partial-Connetions). Moreover, combination of RetinaNet with ResNet101 as backbone shows impressive results than all other methods in terms of AP but silightly less for IoU i.e. 5% less than both YOLOv4 and YOLOv3-spp. However, Faster R-CNN and systematic Det-D0 the standard methods that has been used showed lowest Average Precision and IoU.

To obtain performance of colonoscopy where data of video is available and have good speed as crucial element, identification and localization of object requires a selection

between precision, MIPS (computational speed) and accuracy as a necessary element. So, for localization and identification of polyps we can consider combination of YOLOv4 with Darknet53 and cross-stage-partial-connections as backbone is comparatively better than other techniques.

The given results also shows that ColonSegNet method used produce best results in comparison to all other methods. Highest Frame per second obtained was 182.38 by this method. Figure 4.2 (b) reflects the quantitative results i.e. that highest accuracy in segmentation of polyps has been achieved as compared to all other considered SOTA methods. Pairing of UNet with ResNet34 as backbone is the excellent method of ColonSegNet. As DeepLabc3+ has ability to navigate desired regions semantically with its SPP and atrous convolution process so it was also a good method.

Furthermore, maps with high accuracy for semantic representation of object leading to segmentation may also computed by concatenation of feature with former maps of feature. PSPNet, another competitor used same idea but instead of using dilated convolutions to collect context data from multiple regions globally. The computational speed associated with DeepLabv3+ coupled with ResNet50 as backbone which is the same backbone used for PSPNet is due to its spatial-pyramid pooling connections and 1dimensional dividable convolutions. HRNet [54] which is the most advanced and recent methods was evaluated which gave better results than all other SOTA methods. As, already discussed ColonSegNet outperforming regarding accuracy similarly, combination of UNet with ResNet34 as backbone and DeepLabv3+ gave far better result for most metrics evaluation than all other standard methods.

With regards of segmentation on dice similarity coefficient (DSC), 16 top scorer and also 16 bottom scorer images were placed in Figure 4.2. With reference to Figure 4.2 (a) and (c) for improvement in DSC and other segmentation metrics samples having different texture, variable lightening conditions and angular views should be included as Figure 4.2 (c) reflects that the polyps' samples have shown similar appearance to healthy texture of skin of GI under the provided lightening conditions. Flat or immobile polyps serving as limiting factors for algorithm validity was also observed. By considering small polyps with reference to size help to generalize better results of algorithm and also ease process for difficult to find polyps' detection.

So, here for improving robustness of used methods shape information or association

of artifacts that are context-aware processes and SPP to manage smaller polyps can be used. Retrospective or reminiscence design may consider as possible limitation of work. For real world study clinical setup and understanding is required. Image resizing was done in research work can lead to information lost and alter performance of algorithm [31]. Furthermore, empirical estimation-based algorithm was optimized. As standard hyper-variables were set after experimentation they can be calibrated further. Additionally, in order to optimize hyper-variables meta-learning approaches that can perform even in limited resources settings can be used.

6.2 Conclusion

In this thesis Kvasir-SEG dataset was used to benchmark various deep learning approaches. By conducting series of experiments for identification, segmentation and localization of polyps the obtained results shown the performance of different algorithms on various resolutions of images and different sizes of polyps.

Polyps were detected and localized at 180 FPS while segmentation of polyps were done at 182.38 FPS by ColonSegNet which was the proposed method. For algorithm of detection high AP, IoU, and frames per second while for algorithm of segmentation dice similarity coefficient (DSC), IoU, recall, frames per second, precision and F2-score served as evidence for excellent performance of algorithms of automatic identification, localization and segmentation of polyps. The investigations of algorithms related to identification, segmentation and localization done in this thesis will surely help gastroenterologists in clinical systems. The features like speed, robustness and accuracy of these methods can be further improved by scientist in computer field.

Moreover, the results obtained by these methods can also provide us information about limitations of methods and also provide details about obstacles relevant to dataset used i.e. Kvasir-SEG. Additionally, excellent performers of metrics together with their datasets has been also provided in this thesis to compare different techniques. Obtained results can also improved by using more advance approaches, data generation or augmentation and fine tuning.

Furthermore, we can also enhance detection, segmentation and localization of polyps by using different artifacts like bubbles, contrast, secularity and saturation[55].

6.3 Future Prospects

The application of deep learning techniques to the identification, localization, and segmentation of gastrointestinal (GI) polyps in colonoscopy images holds significant promise for improving colorectal cancer detection and treatment. In Future, I will focus to ensure enhanced detection accuracy through advanced algorithms, larger annotated datasets, and transfer learning, as well as real-time processing capabilities enabled by optimized models and hardware acceleration. Integration with clinical workflows, development of user-friendly interfaces, and multimodal approaches combining various data sources will further enhance diagnostic accuracy and help medical health professionals to take immediate decisions even during the process and also help them to prioritize the case and streamline their further process.

In Future, I will be more focused to consider many different parameters and algorithms to build even better models and perform tasks like identification, localization and segmentation of polyps with more accurate results.

Bibliography

[1] Steven J Stryker et al. "Natural history of untreated colonic polyps". In: *Gastroenterology* 93.5 (1987), pp. 1009–1013.

[2] Theodore R Levin et al. "Complications of screening flexible sigmoidoscopy". In: *Gastroenterology* 123.6 (2002), pp. 1786–1792.

[3] Sanjay Popat, R Hubner, and RS Houlston. "Systematic review of microsatellite instability and colorectal cancer prognosis". In: *Journal of clinical oncology* 23.3 (2005), pp. 609–618.

[4] Jia Deng et al. "Imagenet: A large-scale hierarchical image database". In: *2009 IEEE conference on computer vision and pattern recognition*. Ieee. 2009, pp. 248–255.

[5] Øyvind Holme et al. "Flexible sigmoidoscopy versus faecal occult blood testing for colorectal cancer screening in asymptomatic individuals". In: *Cochrane Database of Systematic Reviews* 9 (2013).

[6] Li Deng, Dong Yu, et al. "Deep learning: methods and applications". In: *Foundations and trends® in signal processing* 7.3–4 (2014), pp. 197–387.

[7] Ross Girshick et al. "Rich feature hierarchies for accurate object detection and semantic segmentation". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2014, pp. 580–587.

[8] Tsung-Yi Lin et al. "Microsoft coco: Common objects in context". In: *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V 13*. Springer. 2014, pp. 740–755.

[9] Karen Simonyan and Andrew Zisserman. "Very deep convolutional networks for large-scale image recognition". In: *arXiv preprint arXiv:1409.1556* (2014).

[10] Nima Tajbakhsh, Suryakanth R Gurudu, and Jianming Liang. "Automatic polyp detection using global geometric constraints and local intensity variation patterns". In: *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2014: 17th International Conference, Boston, MA, USA, September 14-18, 2014, Proceedings, Part II 17*. Springer. 2014, pp. 179–187.

[11] Mark Everingham et al. "The pascal visual object classes challenge: A retrospective". In: *International journal of computer vision* 111 (2015), pp. 98–136.

[12] Ross Girshick. "Fast r-cnn". In: *Proceedings of the IEEE international conference on computer vision*. 2015, pp. 1440–1448.

[13] Jonathan Long, Evan Shelhamer, and Trevor Darrell. "Fully convolutional networks for semantic segmentation". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2015, pp. 3431–3440.

[14] Shaoqing Ren et al. "Faster r-cnn: Towards real-time object detection with region proposal networks". In: *Advances in neural information processing systems* 28 (2015).

[15] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. "U-net: Convolutional networks for biomedical image segmentation". In: *Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18*. Springer. 2015, pp. 234–241.

[16] Nima Tajbakhsh, Suryakanth R Gurudu, and Jianming Liang. "Automatic polyp detection in colonoscopy videos using an ensemble of convolutional neural networks". In: *2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI)*. IEEE. 2015, pp. 79–83.

[17] Kaiming He et al. "Deep residual learning for image recognition". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016, pp. 770–778.

[18] Joseph Redmon et al. "You only look once: Unified, real-time object detection". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016, pp. 779–788.

[19] Tsung-Yi Lin et al. "Focal loss for dense object detection". In: *Proceedings of the IEEE international conference on computer vision*. 2017, pp. 2980–2988.

[20] Konstantin Pogorelov et al. “Efficient disease detection in gastrointestinal videos–global features versus neural networks”. In: *Multimedia Tools and Applications* 76 (2017), pp. 22493–22525.

[21] Konstantin Pogorelov et al. “Kvasir: A multi-class image dataset for computer aided gastrointestinal disease detection”. In: *Proceedings of the 8th ACM on Multimedia Systems Conference*. 2017, pp. 164–169.

[22] Michael Riegler et al. “Multimedia for medicine: the medico task at mediaeval 2017”. In: (2017).

[23] Arica White. “Cancer screening test use—United States, 2015”. In: *MMWR. Morbidity and mortality weekly report* 66 (2017).

[24] Hengshuang Zhao et al. “Pyramid scene parsing network”. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017, pp. 2881–2890.

[25] Johannes Asplund et al. “Survival trends in gastric adenocarcinoma: a population-based study in Sweden”. In: *Annals of surgical oncology* 25 (2018), pp. 2693–2702.

[26] Patrick Brandao et al. “Towards a computed-aided diagnosis system in colonoscopy: automatic polyp segmentation using convolution neural networks”. In: *Journal of Medical Robotics Research* 3.02 (2018), p. 1840002.

[27] Liang-Chieh Chen et al. “Encoder-decoder with atrous separable convolution for semantic image segmentation”. In: *Proceedings of the European conference on computer vision (ECCV)*. 2018, pp. 801–818.

[28] Jeff Heaton. “Ian goodfellow, yoshua bengio, and aaron courville: Deep learning: The mit press, 2016, 800 pp, isbn: 0262035618”. In: *Genetic programming and evolvable machines* 19.1 (2018), pp. 305–307.

[29] Jie Hu, Li Shen, and Gang Sun. “Squeeze-and-excitation networks”. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2018, pp. 7132–7141.

[30] Ahmed Mohammed et al. “Y-net: A deep convolutional neural network for polyp detection”. In: *arXiv preprint arXiv:1806.01907* (2018).

[31] Yuichi Mori et al. “Real-time use of artificial intelligence in identification of diminutive polyps during colonoscopy: a prospective study”. In: *Annals of internal medicine* 169.6 (2018), pp. 357–366.

[32] Konstantin Pogorelov et al. “Medico multimedia task at mediaeval 2018”. In: *CEUR workshop proceedings*. Vol. 2283. Technical University of Aachen. 2018, pp. 1–4.

[33] Joseph Redmon and Ali Farhadi. “Yolov3: An incremental improvement”. In: *arXiv preprint arXiv:1804.02767* (2018).

[34] Younghak Shin et al. “Automatic colon polyp detection using region based deep CNN and post learning approaches”. In: *IEEE Access* 6 (2018), pp. 40950–40962.

[35] Pu Wang et al. “Development and validation of a deep-learning algorithm for the detection of polyps during colonoscopy”. In: *Nature biomedical engineering* 2.10 (2018), pp. 741–748.

[36] Andrew MD Wolf et al. “Colorectal cancer screening for average-risk adults: 2018 guideline update from the American Cancer Society”. In: *CA: a cancer journal for clinicians* 68.4 (2018), pp. 250–281.

[37] Zhengxin Zhang, Qingjie Liu, and Yunhong Wang. “Road extraction by deep residual u-net”. In: *IEEE Geoscience and Remote Sensing Letters* 15.5 (2018), pp. 749–753.

[38] Yali Zheng et al. “Localisation of colorectal polyps by convolutional neural network features learnt from white light and narrow band endoscopic images of multiple databases”. In: *2018 40th annual international conference of the IEEE engineering in medicine and biology society (EMBC)*. IEEE. 2018, pp. 4142–4145.

[39] Debesh Jha et al. “Resunet++: An advanced architecture for medical image segmentation”. In: *2019 IEEE international symposium on multimedia (ISM)*. IEEE. 2019, pp. 225–2255.

[40] Alexandr A Pozdeev, Nataliia A Obukhova, and Alexandr A Motyko. “Automatic analysis of endoscopic images for polyps detection and segmentation”. In: *2019 IEEE conference of russian young researchers in electrical and electronic engineering (EIConRus)*. IEEE. 2019, pp. 1216–1220.

[41] M Sharma et al. “Labelbox: The best way to create and manage training data. software, LabelBox”. In: *Inc, https://www. labelbox. com* (2019).

[42] Sudhir Sornapudi, Frank Meng, and Steven Yi. “Region-based automated localization of colonoscopy and wireless capsule endoscopy polyps”. In: *Applied Sciences* 9.12 (2019), p. 2404.

[43] Mingxing Tan and Quoc Le. “Efficientnet: Rethinking model scaling for convolutional neural networks”. In: *International conference on machine learning*. PMLR. 2019, pp. 6105–6114.

[44] Yu Tian et al. “One-stage five-class polyp detection and classification”. In: *2019 IEEE 16th international symposium on biomedical imaging (ISBI 2019)*. IEEE. 2019, pp. 70–73.

[45] Thomas Wittenberg et al. “Computer aided detection of polyps in whitelight-colonoscopy images using deep neural networks”. In: *Current Directions in Biomedical Engineering* 5.1 (2019), pp. 231–234.

[46] Alexander P Abadir et al. “Artificial intelligence in gastrointestinal endoscopy”. In: *Clinical endoscopy* 53.2 (2020), pp. 132–141.

[47] Debesh Jha et al. “Doubleu-net: A deep convolutional neural network for medical image segmentation”. In: *2020 IEEE 33rd International symposium on computer-based medical systems (CBMS)*. IEEE. 2020, pp. 558–564.

[48] Debesh Jha et al. “Kvasir-seg: A segmented polyp dataset”. In: *MultiMedia modeling: 26th international conference, MMM 2020, Daejeon, South Korea, January 5–8, 2020, proceedings, part II 26*. Springer. 2020, pp. 451–462.

[49] Xiao Jia et al. “Automatic polyp recognition in colonoscopy images using deep learning and two-stage pyramidal feature prediction”. In: *IEEE Transactions on Automation Science and Engineering* 17.3 (2020), pp. 1570–1584.

[50] Ji Young Lee et al. “Real-time detection of colon polyps during colonoscopy using deep learning: systematic validation with four independent datasets”. In: *Scientific reports* 10.1 (2020), p. 8379.

[51] Rahul Pannala et al. “Artificial intelligence in gastrointestinal endoscopy”. In: *VideoGIE* 5.12 (2020), pp. 598–613.

[52] Luisa F Sánchez-Peralta et al. “Piccolo white-light and narrow-band imaging colonoscopic dataset: A performance comparative of models and datasets”. In: *Applied Sciences* 10.23 (2020), p. 8501.

[53] Mingxing Tan, Ruoming Pang, and Quoc V Le. "Efficientdet: Scalable and efficient object detection". In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2020, pp. 10781–10790.

[54] Jingdong Wang et al. "Deep high-resolution representation learning for visual recognition". In: *IEEE transactions on pattern analysis and machine intelligence* 43.10 (2020), pp. 3349–3364.

[55] Sharib Ali et al. "A deep learning framework for quality assessment and restoration in video endoscopy". In: *Medical image analysis* 68 (2021), p. 101900.

[56] Shaik Johny Basha et al. "A novel approach for optical character recognition (OCR) of handwritten Telugu alphabets using convolutional neural networks". In: *2021 Second International Conference on Electronics and Sustainable Communication Systems (ICESC)*. IEEE. 2021, pp. 1494–1500.

[57] Daniel J France et al. "Defining the epidemiology of safety risks in neonatal intensive care unit patients requiring surgery". In: *Journal of patient safety* 17.8 (2021), e694–e700.

[58] Daniele Grattarola and Cesare Alippi. "Graph neural networks in tensorflow and keras with spektral [application notes]". In: *IEEE Computational Intelligence Magazine* 16.1 (2021), pp. 99–106.

[59] Debesh Jha et al. "A comprehensive study on colorectal polyp segmentation with ResUNet++, conditional random field and test-time augmentation". In: *IEEE journal of biomedical and health informatics* 25.6 (2021), pp. 2029–2040.

[60] Debesh Jha et al. "Real-time polyp detection, localization and segmentation in colonoscopy using deep learning". In: *Ieee Access* 9 (2021), pp. 40496–40510.

[61] Kaidong Li et al. "Colonoscopy polyp detection and classification: Dataset creation and comparative evaluations". In: *Plos one* 16.8 (2021), e0255809.

[62] Xinyu Liu et al. "Consolidated domain adaptive detection and localization framework for cross-device colonoscopic images". In: *Medical image analysis* 71 (2021), p. 102052.

[63] Eleftheria Michalopoulou et al. "Impact of comorbidities at diagnosis on the 10-year colorectal cancer net survival: A population-based study". In: *Cancer Epidemiology* 73 (2021), p. 101962.

[64] Jeremi Podlasek et al. “Real-time deep learning-based colorectal polyp localization on clinical video footage achievable with a wide array of hardware configurations”. In: *Endoscopy International Open* 9.05 (2021), E741–E748.

[65] Hemin Ali Qadir et al. “Toward real-time polyp detection using fully CNNs for 2D Gaussian shapes prediction”. In: *Medical Image Analysis* 68 (2021), p. 101897.

[66] Jianwei Xu et al. “Real-time automatic polyp detection in colonoscopy using feature enhancement module and spatiotemporal similarity correlation unit”. In: *Biomedical Signal Processing and Control* 66 (2021), p. 102503.

[67] Umar Albalawi and Mohammed Mustafa. “Current artificial intelligence (AI) techniques, challenges, and approaches in controlling and fighting COVID-19: a review”. In: *International Journal of Environmental Research and Public Health* 19.10 (2022), p. 5901.

[68] Brandy Jo Branham. *Implementation of a Colorectal Screening Decision Aid in Rural Ohio*. Grand Canyon University, 2022.

[69] Eun Jeong Gong et al. “Impact of the volume and distribution of training datasets in the development of deep-learning models for the diagnosis of colorectal polyps in endoscopy images”. In: *Journal of Personalized Medicine* 12.9 (2022), p. 1361.

[70] Ruchika Kulshrestha et al. *Tourism, Event and Digital Media*. OrangeBooks Publication, 2022.

[71] Jarrett D Blair et al. “A gentle introduction to computer vision-based specimen classification in ecological datasets”. In: *Journal of Animal Ecology* 93.2 (2024), pp. 147–158.

[72] M Abadi et al. “1567 S. Ghemawat, G. Irving, M. Isard, et al., Tensorflow: A system for large-1568 scale machine learning”. In: *12th {USENIX} symposium on operating*. Vol. 1569, pp. 265–283.